



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN
INGEGNERIA BIOMEDICA, ELETTRICA E DEI SISTEMI

Ciclo 38

Settore Concorsuale: 09/E4 - MISURE

Settore Scientifico Disciplinare: ING-INF/07 - MISURE ELETTRICHE E ELETTRONICHE

AI-BASED MEASUREMENT ARCHITECTURES IN ELECTRICAL POWER
SYSTEMS

Presentata da: Virginia Negri

Coordinatore Dottorato

Michele Monaci

Supervisore

Roberto Tinarelli

Co-supervisore

Lorenzo Peretto

Esame finale anno 2026

To the future,
may it inspire joy and passion in those who embark upon it.

Contents

Abstract	6
Summary	7
List of Publications	9
List of Figures	13
List of Tables	16
List of Abbreviations and Acronyms	19
1 Introduction	1
1.1 Background and Motivation	1
1.2 Objectives of the Dissertation	4
2 Overview of Artificial Intelligence and Machine Learning	7
2.1 Introduction to Artificial Intelligence	7
2.2 Machine Learning Paradigms	8
2.3 Supervised Learning Methods	9
2.3.1 Classical Machine Learning Algorithms	9
2.3.2 Artificial Neural Networks	12
2.3.3 Deep Learning Models	13
2.3.3.1 Long Short-Term Memory Networks	16
2.3.3.2 Nonlinear Autoregressive with Exogenous Inputs	16
2.3.3.3 Convolutional Neural Networks	17
2.3.3.4 Physics-Informed Neural Networks	19
2.3.3.5 Graph Attention Networks	20
2.3.3.6 Training Parameters	21
2.3.4 Evaluation Metrics	22
2.3.4.1 Classification Metrics	22
2.3.4.2 Regression Metrics	24
2.4 Unsupervised Learning Methods	26
2.4.1 K-Means Clustering	26
2.4.2 Spectral Clustering	28
2.4.3 Evaluation Metrics	29
2.5 Summary	31

3	Artificial Intelligence for the Prediction of Electrical Quantities	32
3.1	Forecasting Electrical Quantities	32
3.1.1	Motivation and Proposed Approach	33
3.1.2	Evaluation Strategy	35
3.1.3	Results and Discussion	38
3.2	AI-Driven Distributed Measurement Systems	40
3.2.1	Building Overview and Data Collection	41
3.2.2	Artificial Intelligence Framework	42
3.2.3	Tests Description	43
3.2.4	Results and Discussion	45
3.2.4.1	Performance across Distributed Measurement System Nodes and Electrical Quantities	45
3.2.4.2	Correlation Analysis Results	49
3.2.4.3	Model Transferability across Correlated Nodes	50
3.2.4.4	Comparison with State-of-the-Art Methods	52
3.2.4.5	Discussion	53
3.3	Enhancing Prediction in Fuel Cell Hybrid Systems	55
3.3.1	Data Collection	56
3.3.2	Proposed Approach	57
3.3.3	Results and Discussion	58
4	Artificial Intelligence for Fault Detection and Predictive Maintenance	61
4.1	Machine Learning for Cable Joint Fault Diagnostics	61
4.1.1	Overview of Cable Joints	63
4.1.1.1	Classification of Cable Joints	64
4.1.1.2	Construction and Installation Methods	64
4.1.2	Structure	65
4.1.3	Failure Modes	66
4.1.4	Diagnostic Parameters of Cable Joints	66
4.1.4.1	Thermal Behavior	67
4.1.4.2	Partial Discharges	68
4.1.4.3	Other Relevant Parameters	69
4.1.5	Case Study	70
4.1.6	Test Description	74
4.1.6.1	Number of Samples for the Learning Stage	74
4.1.6.2	Repeatability	75
4.1.6.3	Fault Percentage Control	75
4.1.7	Results and Discussion	75
4.1.7.1	Number of Samples for the Learning Stage	75
4.1.7.2	Repeatability	77
4.1.7.3	Fault Percentage Control	78
4.1.7.4	Overall Discussion	78
4.2	A Novel Health Index for Cable Joint Aging Prediction	79
4.2.1	Related Literature	80
4.2.2	The Proposed Health Index	81
4.2.3	Case Study	84
4.2.3.1	Dataset Description	84

4.2.3.2	Data Preprocessing and AI Model	86
4.2.4	Training and Testing Details	89
4.2.5	Results	90
4.2.5.1	Health Index Forecasting	90
4.2.5.2	Comparison with State-of-the-Art Models	91
4.2.5.3	Ablation Study on Input Measurements	92
4.2.5.4	Direct Forecasting of Measurements	93
4.2.6	Final Remarks	94
4.3	Predictive Maintenance of Medium Voltage Switchgears	94
4.3.1	Operational Challenges and Degradation Mechanisms in Medium Voltage Switchgears	96
4.3.2	Experimental Setup	97
4.3.3	Supervised Learning Approach	100
4.3.3.1	Artificial Intelligence Model	101
4.3.3.2	Data Description	102
4.3.3.3	Results on Synthetic Data	103
4.3.3.4	Results on Experimental Data	104
4.3.3.5	Discussion	105
4.3.4	Unsupervised Learning Approach	106
4.3.4.1	Clustering Methods	107
4.3.4.2	Results on Synthetic Data	108
4.3.4.3	Results on Experimental Data	111
4.3.4.4	Results on Arduino Board	112
4.3.4.5	Discussion	114
4.3.5	Final Remarks	114
5	Artificial Intelligence and Metrology	116
5.1	The Influence of Measurement Uncertainty in Artificial Intelligence Models	117
5.1.1	The Cable Joint Test Case	118
5.1.1.1	Uncertainty of Input Data	119
5.1.1.2	Uncertainty of Fault Thresholds	119
5.1.1.3	Discussion	121
5.1.2	State of Charge Estimation in Lithium-Ion Batteries	122
5.1.2.1	Data Collection and Reference Setup	123
5.1.2.2	Deep Learning Model	125
5.1.2.3	Uncertainty Analysis	125
5.1.2.4	Discussion	128
5.2	Uncertainty-Driven Data Augmentation	129
5.2.1	Data Description	130
5.2.2	Measurement Uncertainty Quantification	131
5.2.3	Metrological Framework	132
5.2.4	Results and Discussion	136
5.2.4.1	Uncertainty-Based Training in a Single Substation	136
5.2.4.2	Cross-Substation Generalization Using Augmented Data	141
5.2.4.3	Uncertainty-Aware Testing on Different Substations	143
5.2.4.4	Model Ensemble for Training Variability Assessment	144
5.2.4.5	Dropout-Based Uncertainty Estimation during Testing	145

5.2.4.6	Discussion	146
5.3	Role of Temperature in Measurement Accuracy	147
5.3.1	Data Collection and Preprocessing	148
5.3.2	Artificial Intelligence Model	151
5.3.3	Results and Discussion	151
5.3.3.1	Instantaneous Accuracy Estimation Using Temperature Only	151
5.3.3.2	Instantaneous Accuracy Estimation Using Temperature and Voltage	152
5.3.3.3	Effect of Dataset Size on Instantaneous Accuracy Estimation	153
5.3.3.4	Long-Term Accuracy Estimation Using Temperature	153
5.3.3.5	Long-Term Accuracy Estimation Using Temperature and Voltage	154
5.3.3.6	Long-Term Accuracy Estimation vs. Dataset Size	154
5.3.3.7	Discussion	155
5.3.4	Microcontroller Implementation	155
5.3.5	Final Remarks	157
5.4	Advanced Modeling Approaches in Current Sensing	158
5.4.1	Rogowski Coils in Smart Sensing Applications	158
5.4.2	Devices Under Test: Commercial Rogowski Coils	159
5.4.3	Rogowski Coils Temperature Compensation via Physics-Informed Neural Networks	160
5.4.3.1	Scenario and Standard Tests for Rogowski Coils	161
5.4.3.2	Data Collection	162
5.4.3.3	Physics-Informed Learning	164
5.4.3.4	Results	166
5.4.3.5	Overall Discussion	174
5.4.4	Microcontroller-Based Digital Twin for Embedded Signal Verification	175
5.4.4.1	Microcontroller Implementation	176
5.4.4.2	Microcontroller Preliminary Characterization	177
5.4.4.3	Rogowski Coils Preliminary Characterization	179
5.4.4.4	Results	180
5.4.5	Final Remarks	182
5.5	Summary of AI and Metrology Integration	183
6	Conclusion	186

Abstract

This doctoral dissertation investigates the integration of Artificial Intelligence (AI) with measurement systems to improve the performance, reliability, and robustness of modern electrical power systems. The increasing penetration of Renewable Energy Sources (RES), distributed generation, and smart grid technologies poses significant challenges in monitoring, forecasting, and fault management. To address these challenges, the research develops and investigates AI-based measurement architectures based on Machine Learning (ML) techniques, providing scalable, accurate, and deployable solutions.

The work is structured around three main technical areas. The first addresses AI-based forecasting of electrical quantities within Distributed Measurement Systems (DMSs) deployed in smart buildings and hybrid power systems, improving prediction accuracy for monitoring purposes and enabling real-time energy management architectures and Digital Twins (DTs) in DMS-based environments. The second area focuses on fault detection and predictive maintenance in Medium Voltage (MV) distribution networks. ML-based diagnostic methods are developed for Cable Joints (CJs) and MV switchgears using both supervised and unsupervised approaches. A novel Health Index (HI) is introduced to quantify CJ aging and is estimated through Deep Learning (DL) models, supporting condition-based maintenance strategies.

The third area integrates metrological principles into AI-based architectures. The impact of measurement uncertainty on AI model performance is systematically analyzed, leading to uncertainty-aware evaluation frameworks and uncertainty-driven Data Augmentation (DA) strategies to improve robustness and generalization. In parallel, AI is applied to metrological estimation: Artificial Neural Networks (ANNs) are used to estimate temperature-induced accuracy deviations in Instrument Transformers (ITs), while physics-informed models are developed for Rogowski Coils (RCs) to compensate thermal effects. These solutions are also implemented on microcontrollers, enabling local and real-time modeling of sensor behavior.

Overall, this dissertation provides an interdisciplinary contribution by integrating advanced ML with domain-specific knowledge in power systems and metrology, supporting the development of future smart grids that are intelligent, reliable, and metrologically sound.

Keywords: artificial intelligence, electrical measurements, fault detection, machine learning, measurement systems, medium voltage networks, metrology, power forecasting, predictive maintenance, uncertainty modeling.

Summary

This dissertation investigates the integration of AI into measurement systems to enhance the intelligence, robustness, and accuracy of electrical power system monitoring and control. The work is structured into six chapters to develop a comprehensive and multifaceted approach to AI-based architectures in electrical metrology and diagnostics.

Chapter 1 – Introduction establishes the scientific and technological context of the dissertation. It discusses the evolution of modern power systems toward decentralized and intelligent architectures, emphasizing the challenges introduced by renewable integration, data-driven management, and measurement uncertainty. The enabling role of AI and ML is introduced across three key domains: forecasting of electrical quantities, fault detection and predictive maintenance, and uncertainty-aware metrological modeling. The chapter defines the main research objectives, centered on the development of AI-enhanced, explainable, and deployable solutions for measurement systems. These objectives form the foundation for the investigations presented in the subsequent chapters.

Chapter 2 – Overview of Artificial Intelligence and Machine Learning presents the theoretical foundations and computational principles underlying the AI techniques adopted in this dissertation. It introduces the main categories of ML and provides a concise description of the algorithms and DL architectures employed in subsequent chapters. The chapter serves as a methodological reference, providing a description of ML techniques and evaluation metrics.

Chapter 3 – Artificial Intelligence for the Prediction of Electrical Quantities focuses on the development of AI models for forecasting electrical quantities such as voltage, current, and frequency. These models are also evaluated within DMSs deployed in environments such as smart buildings and fuel cell hybrid systems. The chapter explores different ML approaches and compares their predictive performance with traditional statistical models. A key contribution is the design of an AI-enhanced DMS architecture, providing a foundational layer for the future implementation of DT frameworks.

Chapter 4 – Artificial Intelligence for Fault Detection and Predictive Maintenance addresses the application of ML to diagnostic tasks in MV power networks. This includes two main subsystems: CJs and MV switchgears. For CJs, supervised classification algorithms are developed to identify fault signatures, and a novel HI is introduced to quantify the aging status of the component. The HI is then estimated through advanced DL architectures, allowing the prediction of CJs' health status and enabling condition-based maintenance. For MV switchgears, both supervised and unsupervised learning techniques are used to identify anomalies. The models are validated on experimental datasets, demonstrating their applicability for condition monitoring and early fault prevention in power distribution systems.

Chapter 5 – Artificial Intelligence and Metrology explores the intersection of AI with metrological analysis. The chapter begins by analyzing the impact of measurement uncertainty on AI performance, a factor often neglected in traditional AI pipelines. It proposes different methods and a metrological framework to incorporate measurement uncertainty into the training, testing, and deployment phases of AI models. Techniques such as uncertainty-driven DA are developed to enhance robustness under incomplete or noisy data conditions. Subsequently, an ANN is employed to estimate key metrological parameters of ITs (i.e., ratio error and phase displacement) as functions of temperature. The AI model is also implemented on a microcontroller, enabling real-time and local estimation of accuracy parameters. In parallel, a Physics-Informed Neural Network (PINN) is developed to model the temperature-dependent behavior of RCs, allowing for accurate input-output estimation across varying thermal conditions. Finally, a DT of the RC is further implemented on an embedded platform, supporting continuous local monitoring and correction of sensor output.

Chapter 6 – Conclusion synthesizes the main contributions of the dissertation across the three core technical areas: AI for forecasting, AI for fault diagnostics, and metrologically sound AI systems. The chapter also outlines promising directions for future research, including the advancement of DTs and the standardization of uncertainty-aware AI frameworks in electrical power systems.

List of Publications

The following publications were co-authored during the Ph.D. program. In addition, the doctoral research was recognized with the **Una Europa PhD Impact Award 2025**.

Journal Papers

1. V. Negri, A. Mingotti, R. Tinarelli, L. Peretto, *Comparison of Algorithms for the AI-Based Fault Diagnostic of Cable Joints in MV Networks*, *Energies*, vol. 16, no. 1, p. 470, 2023.
[DOI: 10.3390/en16010470](https://doi.org/10.3390/en16010470)
2. V. Negri, A. Mingotti, L. Peretto, R. Tinarelli, G. Iadarola, S. Spinsante, *Predictive Maintenance Based on Artificial Neural Network for MV Switchgears*, *IEEE Sensors Journal*, vol. 24, no. 21, pp. 35448-35455, 1 Nov. 2024.
[DOI: 10.1109/JSEN.2024.3455755](https://doi.org/10.1109/JSEN.2024.3455755)
3. V. Negri, A. Mingotti, R. Tinarelli, L. Peretto, L. S. S. Ray, B. Zhou, P. Lukowicz, *A Novel Health Index for MV Cable Joint Aging Prediction Based on Dynamic Graph Attention Model*, *IEEE Transactions on Instrumentation and Measurement*, 2025, vol. 74, pp. 1-8.
[DOI: 10.1109/TIM.2025.3550602](https://doi.org/10.1109/TIM.2025.3550602)
4. V. Negri, A. Mingotti, R. Tinarelli, L. Peretto, *Comprehensive Forecasting of Electrical Quantities in an Educational Building via Artificial Intelligence-Driven Distributed Measurement System*, *Sensors*, vol. 25, no. 8, 2025, art. 2456.
[DOI: 10.3390/s25082456](https://doi.org/10.3390/s25082456)
5. G. Iadarola, A. Mingotti, V. Negri, S. Spinsante, *Fault Detection in MV Switchgears through Unsupervised Learning of Temperature Conditions*, *Sensors* 2025, vol. 25, no. 15, art. 4818.
[DOI: 10.3390/s25154818](https://doi.org/10.3390/s25154818)

Conference Papers

The research contributions presented in Papers [10–12] were acknowledged through the **IEEE MeMeA 2025 Student Travel Grant**. Furthermore, Paper [14] received the **IEEE AMPS 2025 2nd Best Paper Award**, while Paper [15] was selected as a **Finalist in the IEEE Industrial Electronics Society Generative AI Challenge 2025**.

1. V. Negri, A. Mingotti, L. Peretto, R. Tinarelli, *Artificial Intelligence and Smart Grids: the Cable Joint Test Case*, in: 2022 IEEE 12th International Workshop on Applied Measurements for Power Systems (AMPS), 2022, pp. 1-6.
DOI: [10.1109/AMPS55790.2022.9978846](https://doi.org/10.1109/AMPS55790.2022.9978846)
2. V. Negri, A. Mingotti, R. Tinarelli, L. Peretto, *Uncertainty and Lack of Information Affecting the Training of Machine Learning Algorithms for Fault Prediction of Cable-Joints*, in: 2023 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2023, pp. 1-5.
DOI: [10.1109/I2MTC53148.2023.10176110](https://doi.org/10.1109/I2MTC53148.2023.10176110)
3. V. Negri, A. Mingotti, R. Tinarelli, L. Peretto, *Comparison Between the Machine Learning and the Statistical Approach to the Forecasting of Voltage, Current, and Frequency*, in: 2023 IEEE 13th International Workshop on Applied Measurements for Power Systems (AMPS), 2023, pp. 1-6.
DOI: [10.1109/AMPS59207.2023.10297192](https://doi.org/10.1109/AMPS59207.2023.10297192)
4. V. Negri, A. Mingotti, R. Tinarelli, L. Peretto, L. Apa, L. D'Alvia, Z. Del Prete, E. Rizzuto, *Analyzing the Performance of AI-Based Battery SoC Estimation: A Metrological Point of View*, in: 2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2024, pp. 1-5.
DOI: [10.1109/I2MTC60896.2024.10560993](https://doi.org/10.1109/I2MTC60896.2024.10560993)
5. V. Negri, S. Zanella, A. Mingotti, R. Tinarelli, L. Peretto, F. Barchi, A. Acquaviva, *Towards the DT of an Educational Building: an AI-based Distributed Measurement System for Power Forecasting*, in: 2024 14th IEEE International Workshop on Applied Measurements for Power Systems (AMPS), 2024.
DOI: [10.1109/AMPS62611.2024.10706690](https://doi.org/10.1109/AMPS62611.2024.10706690)
6. G. Frigo, V. Negri, A. Mingotti, R. Tinarelli, L. Peretto, J. Rens, T. Joubert, *Validating Inertia Assessment Metrics in Distribution Systems with High RES Penetration*, in: 2024 IEEE 14th International Workshop on Applied Measurements for Power Systems (AMPS), 2024, pp. 1-5.
DOI: [10.1109/AMPS62611.2024.10706692](https://doi.org/10.1109/AMPS62611.2024.10706692)

7. V. Negri, A. Mingotti, R. Tinarelli, L. Peretto, *Temperature-Based Accuracy Estimation on Instrument Transformers via Neural Networks and Microcontrollers*, in: 2024 IEEE 14th International Workshop on Applied Measurements for Power Systems (AMPS), 2024, pp. 1-6.
[DOI: 10.1109/AMPS62611.2024.10706689](https://doi.org/10.1109/AMPS62611.2024.10706689)
8. V. Negri, L. D'Alvia, A. Mingotti, *Enhancing Efficiency Prediction Accuracy in PEM Fuel Cell Hybrid Systems Using AI*, in: 2025 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Chemnitz, Germany, 19-22 May, 2025.
[DOI: 10.1109/I2MTC62753.2025.11078974](https://doi.org/10.1109/I2MTC62753.2025.11078974)
9. V. Negri, S. Mari, A. Mingotti, F. Ciancetta, R. Tinarelli, L. Peretto, *Uncertainty-Driven Data Augmentation for Improved Load Time Series Forecasting*, in: 2025 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Chemnitz, Germany, 19-22 May, 2025.
[DOI: 10.1109/I2MTC62753.2025.11079054](https://doi.org/10.1109/I2MTC62753.2025.11079054)
10. V. Negri, G. Iadarola, A. Mingotti, R. Tinarelli, L. Peretto, *Bridging Supervised and Unsupervised Learning for Classification of Breast Tissue*, in: 2025 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Chania, Greece, 28-30 May, 2025, pp. 1-6.
[DOI: 10.1109/MeMeA65319.2025.11067980](https://doi.org/10.1109/MeMeA65319.2025.11067980)
11. V. Negri, A. Laffi, M. Raffi, A. Piras, A. Mingotti, R. Tinarelli, *Enhancing Diabetic Retinopathy Diagnosis with Machine Learning: A Random Forest Approach Using Muscle Response Data*, in: 2025 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Chania, Greece, 28-30 May, 2025, pp. 1-6.
[DOI: 10.1109/MeMeA65319.2025.11068030](https://doi.org/10.1109/MeMeA65319.2025.11068030)
12. V. Negri, A. Mingotti, R. Tinarelli, L. Peretto, *Uncertainty-Aware Human Activity Recognition: Investigating Sensor Impact in ML Models*, in: 2025 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Chania, Greece, 28-30 May, 2025.
[DOI: 10.1109/MeMeA65319.2025.11067967](https://doi.org/10.1109/MeMeA65319.2025.11067967)
13. V. Negri, A. Mingotti, R. Tinarelli, L. Peretto, *Lightweight Digital Twin for Current Sensing Using Microcontroller-Integrated Rogowski Coil Model*, in: 2025 IEEE 15th International Workshop on Applied Measurements for Power Systems (AMPS), Bucharest, Romania, 24-26 September, 2025.
[DOI: 10.1109/AMPS66841.2025.11219948](https://doi.org/10.1109/AMPS66841.2025.11219948)

14. V. Negri, A. Mingotti, R. Tinarelli, L. Peretto, L. S. S. Ray, B. Zhou, P. Lukowicz, *Mitigating Temperature Effects in Rogowski Coils' Accuracy Through Physics-Informed Model*, in: 2025 IEEE 15th International Workshop on Applied Measurements for Power Systems (AMPS), Bucharest, Romania, 24-26 September, 2025.
DOI: [10.1109/AMPS66841.2025.11219924](https://doi.org/10.1109/AMPS66841.2025.11219924)
15. V. Negri, F. Bazzani, A. Mingotti, R. Mandrioli, *AI-Driven Assistant for Optimal Design of Two-Level Three-Phase Voltage Source Inverters*, in: 2025 IEEE 51st Annual Conference of the Industrial Electronics Society (IECON), Madrid, Spain, 14-17 October, 2025.
DOI: [10.1109/IECON58223.2025.11221349](https://doi.org/10.1109/IECON58223.2025.11221349)

List of Figures

2.1	Workflow of an LSTM unit.	16
2.2	Series-parallel (open-loop) and parallel (closed-loop) architectures of the NARX model.	17
2.3	Flowchart of the K-means++ algorithm.	28
3.1	Voltage time series (RMS values) measured on the LV network.	37
3.2	Frequency time series (RMS values) measured on the LV network.	37
3.3	Simulated current time series following the same sampling logic.	37
3.4	Performance metrics of voltage ML-based forecasting models with direct and recursive implementations: (A) MAE values and (B) RMSE values. .	39
3.5	Performance metrics of frequency ML-based forecasting models with direct and recursive implementations: (A) MAE values and (B) RMSE values. .	39
3.6	Performance metrics of current ML-based forecasting models with direct and recursive implementations: (A) MAE values and (B) RMSE values. .	39
3.7	Performance metrics of statistical forecasting models: (A) MAE values and (B) RMSE values.	39
3.8	Diagram of the DMS reading points and served areas.	42
3.9	Actual (blue) vs. predicted (violet) values of power at LVS-A over the week of July 24–30 with 15 min-ahead forecasting and 2-hour input window. . .	48
3.10	Actual (blue) vs. predicted (violet) values of current at LVS-A over the week of July 24–30 with 15 min-ahead forecasting and 2-hour input window. .	49
3.11	Actual (blue) vs. predicted (violet) values of voltage at LVS-A over the week of July 24–30 with 15 min-ahead forecasting and 2-hour input window. .	49
3.12	Correlation matrices at (A) TR2 and (B) LVS-A.	51
3.13	Correlation matrices at (A) LVS-B and (B) LVS-D.	51
3.14	Correlation matrix between quantities measured at LVS-A and LVS-B. . .	51
3.15	MAPE values for power forecasting on TR2 measurement point for LSTM, Random Forest (RF), ARIMA, and CNN.	54
3.16	Actual (blue) vs. estimated (violet) engine efficiency values (100 data points from simple train trip, forward direction).	60
4.1	CJ generic multilayer structure.	65
4.2	Flow diagram of the considered case study.	74
4.3	Accuracy vs. number of training samples for the six algorithms under different input distributions: (A) uniform and (B) normal.	76
4.4	Precision vs. number of training samples for the six algorithms under different input distributions: (A) uniform and (B) normal.	77
4.5	Recall vs. number of training samples for the six algorithms under different input distributions: (A) uniform and (B) normal.	77

4.6	Block diagram for the proposed HI calculation based on the WSM.	84
4.7	<i>GraphHealthNet</i> architecture combining GAT and MLP modules.	87
4.8	Comparison between state-of-the-art methods: CNN, Conv-LSTM, and Transformer architectures with <i>GraphHealthNet</i> for HI forecasting.	92
4.9	Compartments of the MV switchgear and locations of the thermocouples (blue numbers).	99
4.10	Block diagram of the proposed ANN model with activation functions. . .	102
4.11	Comparison of ANN accuracy with state-of-the-art methods.	104
4.12	ARI metrics of the algorithms on datasets with ± 1.0 % and ± 1.5 % uncertainty for (A) thermocouple 3 and (B) thermocouple 5 data.	110
4.13	V_1 metrics of the algorithms on datasets with ± 1.0 % and ± 1.5 % uncertainty for (A) thermocouple 3 and (B) thermocouple 5 data.	111
4.14	Silhouette metrics of the algorithms on datasets with ± 1.0 % and ± 1.5 % uncertainty for (A) thermocouple 3 and (B) thermocouple 5 data.	111
4.15	Workflow of the K-means implementation on microcontroller.	112
5.1	Performance metrics of ML algorithms applied to dataset corrupted by (A) initial uncertainty and (B) doubled uncertainty.	120
5.2	Performance metrics of ML algorithms applied to dataset with and without uncertain threshold.	121
5.3	Performance metrics of ML algorithms applied to dataset with and without uncertain threshold (doubled uncertainty).	122
5.4	Actual (blue) vs. predicted (violet) SoC values.	127
5.5	Comparison between estimated and reference SoC values intervals.	128
5.6	Weekly power consumption pattern from the Beacon Hill substation dataset. .	131
5.7	Weekly trend of actual (blue) vs. predicted (violet) power values with DA 2 on Beacon Hill data (2017-2019).	137
5.8	Comparison of (A) MAPE and (B) times with respect to considered datasets of Beacon Hill data.	138
5.9	Distribution of original and augmented power values for Beacon Hill data – DA with 1.0 % uncertainty (2017-2021).	141
5.10	Weekly trend of actual (blue) vs. predicted (violet) with DA 2 on Beacon Hill data for training (2017-2021) – Tested on Auburn data (2022).	143
5.11	Error metrics (A) mean and (B) standard deviation values for Beacon Hill and Auburn substations with test data (2022) corrupted by 0.5 % and 1.0 % uncertainty – One-day-ahead forecasting.	145
5.12	Example of measurement result from the characterization process.	150
5.13	Real (blue) and inferred (violet) ratio error values result from the test of the ANN on DUT-B data with temperature input feature.	152
5.14	MAPE values of (A) ratio error and (B) phase displacement for all considered cases for DUT-A.	155
5.15	Temperature cycle accuracy test.	162
5.16	PINN estimation model (A) ratio error and (B) phase displacement values for R1 with different testing signals and temperatures.	168
5.17	PINN estimation model (A) ratio error and (B) phase displacement values for R2 with different testing signals and temperatures.	168
5.18	PINN estimation model (A) ratio error and (B) phase displacement values for R3 with different testing signals and temperatures.	168

5.19 PINN estimation model (A) ratio error and (B) phase displacement values for R1 with input signals at 150 Hz and different temperatures.	169
5.20 PINN estimation model (A) ratio error and (B) phase displacement values for R2 with input signals at 150 Hz and different temperatures.	169
5.21 PINN estimation model (A) ratio error and (B) phase displacement values for R3 with input signals at 150 Hz and different temperatures.	170
5.22 PINN estimation model (A) ratio error and (B) phase displacement values for R1 with input signals at 1250 Hz and different temperatures.	170
5.23 PINN estimation model (A) ratio error and (B) phase displacement values for R2 with input signals at 1250 Hz and different temperatures.	171
5.24 PINN estimation model (A) ratio error and (B) phase displacement values for R3 with input signals at 1250 Hz and different temperatures.	171
5.25 PINN estimation model (A) ratio error and (B) phase displacement values for R1 with different frequencies of input signal (100 A amplitude) and different temperatures.	172
5.26 PINN estimation model (A) ratio error and (B) phase displacement values for R1 with sinc-type input signals at -15°C and 50 A.	172
5.27 PINN estimation model (A) ratio error and (B) phase displacement values for R2 with sinc-type input signals at -15°C and 50 A.	173
5.28 PINN estimation model (A) ratio error and (B) phase displacement values for R3 with sinc-type input signals at -15°C and 50 A.	173
5.29 PINN estimation model (A) ratio error and (B) phase displacement values for R1 with sinc-type input signals at different temperatures and frequencies with 100 A amplitude.	173
5.30 PINN estimation model (A) ratio error and (B) phase displacement values for R2 with sinc-type input signals at different temperatures and frequencies with 100 A amplitude.	174
5.31 PINN estimation model (A) ratio error and (B) phase displacement values for R3 with sinc-type input signals at different temperatures and frequencies with 100 A amplitude.	174
5.32 Simple schematic of the microcontroller characterization.	178
5.33 Simple schematic of the RC characterization.	179
5.34 Measurement setup to validate the proposed procedure.	181

List of Tables

2.1	Overview of AI methods and their usage across dissertation chapters. . . .	31
3.1	LSTM metrics with different input sequence lengths and time horizons for power forecasting on TR2 measurement point.	46
3.2	LSTM performance on current and voltage forecasting at TR2 measurement point.	47
3.3	LSTM performance on electrical quantities forecasting at LVS-A measurement point.	47
3.4	LSTM performance on electrical quantities forecasting at LVS-B measurement point.	47
3.5	LSTM performance on electrical quantities forecasting at LVS-D measurement point.	48
3.6	LSTM metrics with different time horizons of electrical quantities on LVS-B measurement point with LSTM model trained on LVS-A data.	52
3.7	Random Forest metrics with different time horizons on TR2 measurement point.	53
3.8	ARIMA metrics with different time horizons on TR2 measurement point.	54
3.9	CNN metrics with different time horizons on TR2 measurement point.	54
3.10	Training and testing time for power forecasting on TR2 measurement point for methods – 4 h time horizon.	55
3.11	Random Forest metrics on simple train trip test set.	59
3.12	Random Forest metrics on train trip with braking test set.	59
3.13	Random Forest metrics on fuel cell powered train trip test set.	59
4.1	Typical failure modes and their causes.	67
4.2	Variation ranges of the quantities of interest.	71
4.3	Fault thresholds for input quantities.	73
4.4	Mean and standard deviation of accuracy for algorithms over 500 repetitions.	77
4.5	Accuracy (A), Precision (P), and Recall (R) metrics vs. percentage of faults in the training set.	78
4.6	Measurements of interest for device 1 across the years.	86
4.7	Weights assigned to each parameter of interest for the HI calculation.	86
4.8	Measurements of interest and HI values for devices from 1 to 10 in 2010.	86
4.9	Evaluation metrics of <i>GraphHealthNet</i> for HI forecasting with different input and output years.	91
4.10	Average evaluation metrics of <i>GraphHealthNet</i> for HI forecasting grouped by the number of input years.	91

4.11	Impact of excluding different input measurements on <i>GraphHealthNet</i> HI forecasting performance with various input and output years. The input measurements are abbreviated as follows: PD – Partial Discharge, TD – Tan Delta, VC – Visual Condition, and NC – Neutral Corrosion.	93
4.12	Evaluation metrics of <i>GraphHealthNet</i> for HI forecasting with different output types for various input and output years.	94
4.13	Minimum and maximum temperature readings acquired during the three sessions with CT fastening of 60 N m, 20 N m and 4 N m.	100
4.14	Overall performance and execution times for synthetic datasets.	103
4.15	Metrics for synthetic dataset with uncertainty $\pm 0.5^{\circ}\text{C}$	104
4.16	Metrics for synthetic dataset with uncertainty $\pm 1.5^{\circ}\text{C}$	104
4.17	Overall performance and execution times for experimental measurements.	105
4.18	Metrics for experimental measurements.	105
4.19	Initial and final temperature readings acquired during the three sessions with CT fastening of 60 N m, 40 N m and 4 N m.	107
4.20	Clustering metrics of the algorithms on datasets with $\pm 1.0\%$ uncertainty (thermocouple 3) in relation with the number of generated samples per class.	109
4.21	Clustering metrics of the algorithms on datasets with $\pm 1.0\%$ uncertainty (thermocouple 5) in relation with the number of generated samples per class.	109
4.22	Clustering metrics of K-means tested on the experimental measurements.	112
4.23	Execution times for fitting and prediction.	113
5.1	Uncertainty values applied to fault thresholds.	121
5.2	Specifications of the tested cell.	124
5.3	Metric values of the NARX model.	126
5.4	Percentage of overlapped SoC intervals.	128
5.5	Error metrics for LSTM model without DA – Beacon Hill (2017-2019).	137
5.6	Error metrics for LSTM model with DA – Beacon Hill (2017-2019).	137
5.7	Training time for LSTM model without and with DA – Beacon Hill (2017-2019).	138
5.8	Error metrics and training time for LSTM model – Beacon Hill (2017-2020).	138
5.9	Error metrics and training time for LSTM model – Beacon Hill (2017-2022).	138
5.10	Error metrics and training times for different strides – Beacon Hill (2017-2019).	139
5.11	Error metrics and training times with DA (one-Week Stride) – Beacon Hill (2017-2019).	139
5.12	K-S test results for original and augmented datasets – Beacon Hill data (2017-2021).	140
5.13	Paired t-test results for original vs. augmented datasets – Beacon Hill data (2017-2022) – 0.5% uncertainty.	141
5.14	Paired t-test results for original vs. augmented datasets – Beacon Hill data (2017-2022) – 1.0% uncertainty.	142
5.15	Error metrics and training time for LSTM model trained on Beacon Hill data (2017-2021) and tested on Auburn data (2022) – One-week-ahead forecasting.	142
5.16	Error metrics and training time for LSTM model trained on Beacon Hill data (2017-2021) and tested on Auburn data (2022) – One-day-ahead forecasting.	143

5.17 Mean and standard deviation of error metrics for LSTM model – Beacon Hill data – tested on 0.5 % uncertainty corrupted data – One-week-ahead forecasting.	144
5.18 Mean and standard deviation of error metrics for LSTM model – Beacon Hill data – tested on 1.0 % uncertainty corrupted data – One-week-ahead forecasting.	144
5.19 Mean and standard deviation of error metrics for LSTM model on Auburn data (2017–2022) – tested on uncertainty corrupted data – One-week-ahead forecasting.	144
5.20 Mean and standard deviation of error metrics for the LSTM ensemble model – Beacon Hill data (2017–2022) – One-week-ahead forecasting.	145
5.21 Mean and standard deviation of error metrics for the LSTM model using Monte Carlo dropout – Beacon Hill data (2017–2022) – One-week-ahead forecasting.	146
5.22 Characteristics of the LPVTs under test.	148
5.23 ANN metrics on DUT-A using temperature input feature.	152
5.24 ANN metrics on DUT-B using temperature input feature.	152
5.25 ANN metrics on DUT-A with temperature and voltage input features. . .	153
5.26 ANN metrics on DUT-B with temperature and voltage input features. . .	153
5.27 ANN metrics on DUT-A with temperature and voltage input features using 10-weeks dataset.	153
5.28 ANN metrics on DUT-A with temperature and weekly input features. . .	154
5.29 ANN metrics on DUT-B with temperature and weekly input features. . .	154
5.30 ANN metrics on DUT-A with temperature, voltage, and weekly input features.	154
5.31 ANN metrics on DUT-B with temperature, voltage, and weekly input features.	154
5.32 ANN metrics on DUT-A with temperature, voltage, and weekly input features using 10-weeks dataset.	155
5.33 ANN metrics on the DUT-B with temperature input feature executed on the Arduino board.	157
5.34 Characteristics of the RCs under test.	159
5.35 Ratio error of RCs at different temperatures.	163
5.36 Phase displacement of RCs at different temperatures.	163
5.37 Amplitude characterization results.	179
5.38 Phase characterization results.	179
5.39 RCs characterization results.	180
5.40 Actual transformation ratio obtained from the characterization.	181
5.41 Results from the implementation of the proposed procedure.	182
5.42 Summary of AI-based case studies in Chapter 5, highlighting metrological aspects and uncertainty handling strategies.	185

List of Abbreviations and Acronyms

AI	Artificial Intelligence
RES	Renewable Energy Sources
ML	Machine Learning
DMS	Distributed Measurement System
DT	Digital Twin
MV	Medium Voltage
CJ	Cable Joint
HI	Health Index
DL	Deep Learning
DA	Data Augmentation
ANN	Artificial Neural Network
IT	Instrument Transformer
RC	Rogowski Coil
PINN	Physics-Informed Neural Network
IoT	Internet of Things
GUM	Guide to the Expression of Uncertainty in Measurement
UQ	Uncertainty Quantification
LR	Logistic Regression
SVM	Support Vector Machine
SVR	Support Vector Regression
KNN	K-Nearest Neighbors
AdaBoost	Adaptive Boosting
XGBoost	eXtreme Gradient Boosting

LGBM	Light Gradient Boosting Machine
ReLU	Rectified Linear Unit
Tanh	Hyperbolic Tangent
MLP	Multi-Layer Perceptron
FC	Fully Connected
FNN	Feedforward Neural Network
RNN	Recurrent Neural Network
CNN	Convolutional Neural Network
GNN	Graph Neural Network
LSTM	Long Short-Term Memory
NARX	Nonlinear AutoRegressive with eXogenous inputs
TDL	Time Delay Layer
GAT	Graph Attention Network
Adam	Adaptive Moment Estimation
MAE	Mean Absolute Error
MSE	Mean Squared Error
RMSE	Root Mean Square Error
ARI	Adjusted Rand Index
PEM	Proton Exchange Membrane
LV	Low Voltage
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
SES	Simple Exponential Smoothing
SE	Exponential Smoothing
RMS	Root Mean Square
THD	Total Harmonic Distortion
CT	Current Transformer
EES	Electrical Energy System
SO	System Operator
PD	Partial Discharge
DOS	Distribution System Operator
WSM	Weighting and Scoring Method
AHP	Analytic Hierarchy Process

HV	High Voltage
PQ	Power Quality
NETA	International Electrical Testing Association
API	Application Programming Interface
SoC	State of Charge
BESS	Battery Energy Storage System
LFP	Lithium Iron Phosphate
BMS	Battery Management System
VT	Voltage Transformer
NER	National Electricity Rules
LPIT	Low-Power Instrument Transformer
LPVT	Low-Power Voltage Transformer
DUT	Device Under Test
IDE	Integrated Development Environment
LPCT	Low-Power Current Transformer
FFT	Fast Fourier Transform
DSP	Digital Signal Processor

Chapter 1

Introduction

This chapter introduces the research context, outlining the technological and scientific motivations that underpin the dissertation. It discusses the current challenges in modern power systems related to distributed monitoring, predictive intelligence, and measurement uncertainty. Finally, it defines the main objectives and research directions of the dissertation, highlighting the unifying role of AI across all investigated domains.

1.1 Background and Motivation

The electrical power grid has undergone radical transformations over the past decades. Traditionally, the power system consisted of centralized plants supplying energy radially to a limited number of loads. Today, RES are integrated at all voltage levels, creating decentralized generation nodes alongside legacy power plants. This evolution has shifted power flow from a unidirectional to a bidirectional paradigm, significantly complicating network management. Furthermore, the intermittency and unpredictability of RES generation impose additional challenges on system operators, who increasingly rely on weather forecasting and advanced control strategies to maintain system stability and reliability [1, 2].

In parallel, the emergence of the Internet of Things (IoT) has enabled the deployment of vast sensor networks, producing high-resolution data across the power grid. These technologies have accelerated the transition from conventional infrastructure to smart grids, enhancing real-time visibility, automation, and data-driven decision-making [3–5].

This growing digitalization has led to deeper integration of power systems with fields such as power electronics, computer science, and information technology. Among the most impactful developments is the application of AI and ML, which provide powerful tools for extracting insights from the massive datasets generated by DMSs, operational logs, and environmental sensors. The fusion of AI with DMS enables advanced condition monitoring, fault diagnosis, and predictive maintenance [6, 7], thereby fostering smarter and more adaptive electrical networks [8].

Condition monitoring plays a crucial role in ensuring the robustness and reliability of both transmission and distribution networks. By analyzing electrical variables (e.g. voltage, current) and environmental parameters (e.g., temperature, humidity), real-time monitoring helps operators detect anomalies early and prevent failures [9]. In transmission networks, monitoring ensures operational safety by tracking power flows and equipment status, while distribution networks require dense instrumentation to detect local issues such as voltage imbalances and load fluctuations [10, 11].

Predictive maintenance, empowered by AI and ML, goes beyond traditional time-based approaches by leveraging real-time data to predict faults before they occur. This results in optimized resource usage, minimized downtime, and cost-effective maintenance planning [12]. It is particularly impactful in renewable energy assets like wind turbines, where early fault detection drastically improves efficiency and reduces operational expenses [13, 14].

A further enabler of smart grid intelligence is the DT paradigm. DTs are dynamic, high-fidelity virtual replicas of physical assets or systems, continuously updated via real-time data to reflect their current states. Their application in the power sector supports enhanced monitoring, forecasting, and operational optimization, especially when combined with AI techniques [15–17]. AI-enhanced DTs enable robust anomaly detection, energy management, and decision support under uncertainty [4, 5, 18–21].

However, a critical and often underrepresented aspect in these developments is the role of measurement uncertainty in AI-powered systems. Metrology requires every measurement to be associated with uncertainty, as formalized in the Guide to the Expression of Uncertainty in Measurement (GUM) [22]. This ensures traceability, repeatability, and meaningful interpretation. Yet, most AI models output point predictions without uncertainty information, even when processing noisy or error-prone input data. This limits

their usability in contexts requiring trustworthy measurements [23]. Moreover, AI and ML models inherit and amplify these uncertainties, affecting the reliability and interpretability of their predictions [24, 25].

Two primary and widely adopted types of uncertainty can be identified in the context of AI:

- **Model-Related Uncertainty**, arising from the learning process itself, including model structure, parameter estimation, and training stochasticity, and manifested in the variability of the model outputs;
- **Input Data Uncertainty**, arising from measurement noise, sensor errors, and limited data precision, particularly relevant in electrical monitoring applications.

These categories encompass the two main sources of uncertainty commonly discussed in the ML literature, often referred to as epistemic uncertainty (model-related) and aleatoric uncertainty (data-related), respectively.

Several strategies have been proposed in the literature to manage these uncertainties. For example, input data uncertainty has been studied in image recognition [26], while Monte Carlo dropout and Bayesian approximation have been used to assess and propagate uncertainty in AI models [27, 28]. Other works analyze how noise and input perturbations impact model performance [29].

Systematic approaches to Uncertainty Quantification (UQ) are therefore essential for deploying AI in critical infrastructure such as power systems. They improve not only accuracy and robustness but also transparency and trustworthiness of AI-based decision-making [30]. This aspect becomes even more relevant in edge computing scenarios, where embedded AI systems must make autonomous decisions close to the physical assets, often under non-ideal or uncertain conditions [31, 32].

In conclusion, the convergence of renewable energy integration, AI and ML technologies, DMS, IoT, edge computing, and DT modeling is driving a fundamental transformation of power grids. However, addressing uncertainty – particularly from measurements – is crucial to ensuring these systems are accurate, interpretable, and resilient in real-world scenarios. These topics form the foundation of the present dissertation, which aims to

develop and evaluate advanced AI-based forecasting and monitoring solutions, along with built-in uncertainty handling mechanisms for modern power distribution systems.

1.2 Objectives of the Dissertation

The objective of this doctoral research is to investigate how AI, and in particular ML, can be effectively integrated into measurement systems to enhance the intelligence, resilience, and metrological reliability of modern electrical power networks. The dissertation moves across three technical domains – forecasting, diagnostics, and uncertainty modeling – which at first glance may appear distinct. However, they are tightly connected by a common goal: enabling the development of AI-based systems that are not only accurate, but also transparent, robust, and suitable for deployment in real-world, safety-critical energy infrastructures.

The core of the dissertation is the pursuit of intelligent measurement: an architecture where data acquisition, processing, and interpretation are not isolated steps, but part of a cohesive and adaptive system. In this vision, measurements are not merely collected but understood (i.e., contextualized, validated, and leveraged to make informed decisions). This is made possible by embedding AI directly within measurement chains, allowing systems to evolve from passive monitoring tools into proactive decision-making agents.

The first research direction focuses on AI-based forecasting. Here, the goal is to develop predictive models capable of estimating fundamental electrical quantities (e.g., voltage, current, and frequency) within distributed measurement environments. These include smart buildings, hybrid power systems, and other infrastructures. The models are designed to operate using data from sensors, enabling real-time forecasting and supporting intelligent grid management. A significant outcome is the design of an AI-based DMS architecture, which lays the foundation for the future implementation of DTs – virtual replicas of physical systems used for simulation, optimization, and control.

The second area addresses fault detection and predictive maintenance in MV power distribution systems. This part of the research explores how AI can identify emerging faults, classify abnormal behaviors, and predict degradation in key components such as CJs and switchgears. The integration of ML models with historical data enables the estimation of a novel HI, designed to serve as a reproducible and quantitative indicator

of asset condition. By supporting condition-based maintenance, this approach increases system availability and reduces the need for periodic inspections, aligning with the broader goals of smart grid automation and sustainability.

The third and final research domain delves into the intersection of AI and metrology. In traditional applications, AI models often neglect the role of measurement uncertainty: a cornerstone of any rigorous measurement process. However, in critical infrastructure, overlooking uncertainty can compromise the trustworthiness of AI-based decisions. This dissertation tackles this issue from two perspectives. First, it studies how measurement uncertainty affects model performance, proposing new methods to incorporate uncertainty into the training, evaluation, and deployment phases of AI systems. Second, it uses AI as a tool for metrological estimation, for instance, modeling thermal effects in electrical sensors such as ITs and RCs. These models are also deployed on embedded platforms, enabling real-time and localized metrological analysis.

A key novelty of the dissertation lies in the actual implementation and validation of the developed AI models on microcontroller-based platforms. Many of the proposed solutions, including forecasting architectures, fault detection algorithms, and metrological estimators, have been embedded into low-power and resource-constrained devices. This step transforms the research from a purely algorithmic contribution to a fully deployable technological solution, capable of operating at the edge, close to the measurement point. It enables real-time inference, local decision-making, and reduces reliance on centralized infrastructure. This embedded implementation aspect is crucial in distributed systems, where latency, bandwidth, and reliability constraints limit the feasibility of cloud-based approaches. The results confirm the technical feasibility and performance of AI-enhanced measurements in constrained environments, opening the path towards scalable, smart, and autonomous measurement infrastructures.

While the three technical areas address different layers of the electrical grid – from monitoring to diagnostics to metrological modeling – they converge towards a single, coherent vision: empowering the electrical infrastructure with intelligent, distributed, and trustworthy systems. This vision embraces the complexity of modern power systems, where massive data flows, distributed architectures, and high-performance requirements coexist with the need for reliability, explainability, and compliance with metrological standards.

In summary, the dissertation aims to:

- Develop **AI models for forecasting** electrical quantities in distributed environments, validated through real-world and synthetic data.
- Design and implement **diagnostic systems for MV network** components, combining supervised and unsupervised ML with health assessment indicators.
- Integrate the **quantification of uncertainty** into AI pipelines, and use AI to model and correct metrological deviations in sensor behavior.
- Validate **embedded implementation of AI algorithms** on microcontroller platforms, enabling edge intelligence in real-time, resource-constrained environments.

By addressing these objectives, this dissertation contributes to the development of a new class of AI-enhanced electrical systems – systems that are not only intelligent but also explainable, resilient, and metrologically sound.

Chapter 2

Overview of Artificial Intelligence and Machine Learning

This chapter provides a comprehensive overview of the AI and ML techniques employed throughout the dissertation, with a focus on their application to forecasting, diagnostics, and uncertainty modeling. After introducing the fundamental concepts of AI, the chapter presents the main ML paradigms – supervised and unsupervised learning – detailing the most relevant algorithms used in the dissertation, including ensemble methods, neural networks, and clustering techniques. DL models are discussed, along with key training concepts.

2.1 Introduction to Artificial Intelligence

AI refers to the ability of machines to perform tasks that typically require human intelligence, such as reasoning, learning, perception, and decision-making. In the last decade, the adoption of AI has expanded rapidly across diverse fields – from healthcare to autonomous vehicles – and is now gaining increasing relevance in electrical engineering and power system management [33, 34].

At its core, AI aims to simulate aspects of human cognition through algorithms capable of learning from data and adapting to new inputs. One of the most impactful branches of AI is ML, which focuses on developing statistical models that can automatically learn from data without being explicitly programmed for each task [35].

Unlike traditional statistical methods, which typically rely on fixed assumptions about data distributions and linear relationships, ML techniques are primarily data-driven, rather than based on explicit assumptions. This allows them to capture complex, non-linear patterns in high-dimensional datasets without requiring explicit prior modeling. While statistical models offer interpretability and are well-suited for systems with known behavior, AI methods provide greater flexibility and predictive power in dynamic or poorly understood environments, making them particularly effective for real-time applications in electrical measurement and diagnostics [36, 37].

A subdomain of ML, known as DL, uses multi-layered neural network architectures to model highly complex, nonlinear relationships. DL has proven particularly powerful in domains involving time series, image recognition, and signal processing, making it especially suitable for modern measurement and monitoring applications in power systems.

A particularly relevant emerging field is tinyML, which refers to the deployment of ML models on resource-constrained embedded systems such as microcontrollers [38]. This innovation is critical in distributed measurement contexts, where edge devices need to perform intelligent analysis locally, with limited computational power and energy consumption. By integrating ML directly into sensing units, tinyML enables real-time forecasting, diagnostics, and control in remote or bandwidth-limited environments, reducing the dependency on centralized processing and enhancing overall system autonomy.

This chapter presents the ML techniques underlying the applications developed throughout the dissertation, with particular attention to the main learning paradigms and to the specific algorithms adopted. The aim is to provide a clear and structured foundation for understanding the function of each method within the broader framework of AI-enabled DMSs.

2.2 Machine Learning Paradigms

ML methods are typically organized into three main paradigms, based on the type and availability of data, and the specific learning objectives: supervised learning, unsupervised learning, and reinforcement learning [39].

- **Supervised learning** refers to the class of algorithms trained on labeled datasets, where the input data is associated with known target outputs. The goal is to learn a mapping from inputs to outputs that generalizes well to unseen data. This approach is widely used in tasks such as classification and regression, where historical measurements and associated ground truth are available.

In classification, the output variable is categorical, and the task is to assign input samples to discrete classes. Classification problems can be further divided into binary classification, where there are only two possible classes (i.e., 0 and 1), and multiclass classification, where there are more than two classes.

In contrast, regression involves predicting continuous numerical values based on input features, such as forecasting house prices or temperature values.

- **Unsupervised learning** deals with unlabeled data. The aim is to discover hidden structures, such as patterns or groupings, within the data. Unlike supervised learning, no explicit feedback is given during training. This paradigm is suitable for applications such as anomaly detection, clustering, and dimensionality reduction, especially when labels are difficult or expensive to obtain.
- **Reinforcement learning** is based on the interaction between an agent and an environment, where the agent learns to make decisions by receiving feedback in the form of rewards or penalties. Reinforcement learning is typically employed in control systems and real-time optimization problems. Although it is not directly applied in this dissertation, it represents an emerging area in smart grid applications.

2.3 Supervised Learning Methods

2.3.1 Classical Machine Learning Algorithms

Linear Regression (LR): LR is a fundamental algorithm used for modeling the relationship between a dependent variable and one or more independent variables. It assumes a linear relationship and fits a line (or hyperplane) that minimizes the sum of squared errors between the predicted and actual values.

Support Vector Machines (SVM): SVMs are supervised learning models primarily used for classification tasks. They work by finding the optimal hyperplane that maximally

separates data points of different classes in the feature space, often using kernel functions to handle nonlinear separability.

Support Vector Regression (SVR): SVR is an extension of SVM to regression problems. Instead of classifying data points, SVR aims to find a continuous function that approximates the target values within a specified margin of tolerance (known as the epsilon-insensitive tube). The goal is to minimize prediction errors that fall outside this margin while keeping the model complexity low, resulting in robust and generalizable regression performance.

K-Nearest Neighbors (KNN): KNN is a nonparametric, instance-based learning algorithm that predicts the output of a new data point based on the outputs of its k closest neighbors in the feature space.

Decision Tree: A Decision Tree is a model that uses a tree-like structure composed of nodes and leaves to represent a sequence of decisions. Internal nodes correspond to tests on input features, branches represent the outcomes of these tests, and leaf nodes correspond to the predicted output values. Each node applies a conditional function to evaluate whether a particular property holds for the data point in question.

The decision-making process starts from the root node and proceeds down the tree, following the branches according to the outcomes of the tests, until reaching a leaf node that provides the prediction. The core objective of the algorithm is to identify the optimal split functions at each node that divide the dataset into subsets. These subsets should be as homogeneous as possible internally (i.e., data points within each group are similar) and as heterogeneous as possible between groups. This partitioning strategy reduces classification uncertainty and improves interpretability by forming simple, rule-based models.

Random Forest: Random Forest is a flexible ML algorithm based on the ensemble learning paradigm, which builds a collection of Decision Trees and aggregates their outputs to produce robust predictions.

The algorithm uses randomness during the training phase: each tree is built on a random subset of the training data (bagging), and at each split, only a random subset of features is considered. This process introduces diversity among the individual trees, reducing the risk of overfitting and improving generalization performance.

Several key hyperparameters influence the performance and complexity of a Random Forest model:

- **Number of Estimators:** Defines the number of trees in the forest. A larger number generally leads to more stable and accurate predictions but increases computational cost.
- **Max Depth:** Sets the maximum depth of each tree. Limiting depth helps control model complexity and prevents overfitting, especially in noisy datasets.
- **Min Samples Split:** Specifies the minimum number of samples required to split an internal node. Higher values result in shallower trees and can improve generalization.
- **Min Samples Leaf:** Determines the minimum number of samples required to be at a leaf node. Setting this parameter helps avoid leaves that capture noise from very small subsets of data.

By combining the predictions of multiple diverse Decision Trees, Random Forest achieves high predictive accuracy while maintaining robustness and interpretability, making it a suitable choice for a wide range of supervised learning tasks.

Adaptive Boosting (AdaBoost): AdaBoost is a boosting ensemble technique that sequentially trains weak learners (typically shallow Decision Trees), giving more weight to previously misclassified examples. The final model is a weighted sum of these weak learners, improving overall accuracy.

eXtreme Gradient Boosting (XGBoost): XGBoost is an advanced implementation of gradient boosting designed for speed and performance. Gradient boosting is an ensemble method in which models are added sequentially, each one trained to correct the errors of the previous ones. This produces an additive model, where the final predictor is the sum of many weak learners, each minimizing a shared differentiable loss function.

XGBoost extends this approach with regularization techniques, efficient tree construction, and optimized handling of sparse data, enabling superior performance and control over model complexity.

Light Gradient Boosting Machine (LGBM): LGBM is a gradient boosting framework optimized for efficiency and scalability. It uses a leaf-wise tree growth strategy with

depth limitations and histogram-based algorithms to reduce memory usage and training time while maintaining high accuracy.

Direct and Recursive Forecasting Strategies

Regression-based ML models can be applied to forecasting by adopting one of two different strategies: Single-step forecaster and multi-step forecaster. The former predicts only the next value in the time series, whereas the latter aims to estimate a sequence of future values or a distant horizon.

For multi-step forecasters, two implementation approaches exist:

- **Direct Strategy:** A separate regression model is trained for each forecasted time step, producing all required predictions in a single stage. This method avoids error propagation but requires more training resources.
- **Recursive Strategy:** The model predicts the next value, then feeds this prediction back as input to generate subsequent forecasts. This approach is computationally efficient but may accumulate errors over long horizons.

Models such as Decision Tree, Random Forest, SVR, AdaBoost, XGBoost, and LGBM can be used under both strategies, offering flexible adaptation to the temporal forecasting requirements of the application.

2.3.2 Artificial Neural Networks

ANNs are computational models inspired by biological neural systems. They consist of interconnected processing units (neurons), each computing a weighted sum of its inputs followed by a nonlinear activation. Shallow ANNs are traditionally considered part of classical ML. DL can be viewed as an extension in which many such layers are stacked, enabling hierarchical feature extraction.

A neuron computes a *pre-activation* value:

$$z = \mathbf{w}^\top \mathbf{x} + b,$$

where $\mathbf{x}$ is the input vector, $\mathbf{w}$ the weight vector, and b a bias term. An activation function $f(\cdot)$ is then applied to produce the output:

$$a = f(z).$$

Training is typically performed using backpropagation and gradient-based optimization, even in shallow networks.

2.3.3 Deep Learning Models

DL models are a class of ML algorithms based on ANNs with multiple processing layers. These models are capable of automatically learning hierarchical representations of data, making them highly effective for tasks such as image recognition, time series forecasting, and structured data modeling [40].

Neural Network Architecture: The basic structure of an ANN consists of an input layer, which receives the raw data; one or more hidden layers, which perform nonlinear transformations using weighted connections and activation functions; and an output layer, which generates the final prediction based on the learned features. The depth (number of hidden layers) and width (number of neurons per layer) determine the model's capacity to approximate complex mappings between inputs and outputs.

Each connection between neurons is associated with a weight, which quantifies the strength and direction of the influence between neurons. These weights are learned during training.

Loss Function: To evaluate how well a neural network performs, a loss function is defined. The loss function quantifies the difference between the predicted output of the model and the actual target value. The goal of training is to minimize this loss.

Training Process: Training a neural network involves the iterative optimization of weights to minimize the loss function. This is typically done using gradient-based optimization algorithms. The training loop consists of the following steps:

1. **Forward Pass:** Input data is propagated through the network to produce a prediction.

2. **Loss Computation:** The loss function computes the error between the predicted and true values.
3. **Backward Pass (Backpropagation):** Gradients of the loss with respect to the weights are computed using the chain rule.
4. **Weight Update:** The optimizer updates the weights based on the computed gradients and the chosen learning rate.

This process is repeated over many epochs, gradually improving the model's performance on the training data and ideally leading to good generalization on unseen data.

Activation Functions: Activation functions introduce nonlinearity into the model. Without them, even deep networks would reduce to a single linear transformation. Activation functions operate on the neuron's pre-activation value $z = \mathbf{w}^\top \mathbf{x} + b$.

The most commonly used activation functions are:

- **Linear (Identity):**

$$\text{Linear}(z) = z. \quad (2.1)$$

- **Rectified Linear Unit (ReLU):**

$$\text{ReLU}(z) = \max(0, z). \quad (2.2)$$

- **Leaky ReLU:**

$$\text{Leaky ReLU}(z) = \begin{cases} \alpha z, & \text{if } z < 0, \\ z, & \text{if } z \geq 0, \end{cases} \quad (2.3)$$

where α is a small positive constant (e.g., 0.01).

- **Sigmoid:**

$$\text{Sigmoid}(z) = \frac{1}{1 + e^{-z}}. \quad (2.4)$$

- **Hyperbolic Tangent (Tanh):**

$$\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}. \quad (2.5)$$

- **Softmax:** Typically used in the output layer for multiclass classification. Given N classes, the Softmax function converts a vector of raw class scores $\mathbf{z}$ into a probability distribution:

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^N e^{z_j}}, \quad i = 1, \dots, N, \quad (2.6)$$

where N denotes the number of output classes.

Multilayer Perceptron (MLP): The MLP is the foundational architecture in DL, consisting of a sequence of Fully Connected (FC) layers, where each neuron in a layer is connected to all neurons in the previous and subsequent layers.

An MLP typically includes an input layer, one or more hidden layers with nonlinear activation functions (e.g., ReLU, Equation 2.2), and an output layer tailored to the specific task – for example, the Softmax activation (Equation 2.6) for classification tasks or the Linear function (Equation 2.1) for regression. Despite their simple structure, MLPs are powerful universal function approximators and can model a wide range of complex relationships when given sufficient depth and data. However, they may struggle with spatial or temporal data, where more specialized architectures are preferred.

Network Types:

- **Feedforward Neural Networks (FNNs):** Data flows in one direction from input to output without cycles. These networks are foundational for most DL tasks.
- **Recurrent Neural Networks (RNNs):** Designed to process sequential data by maintaining a hidden state that captures temporal dependencies.
- **Convolutional Neural Networks (CNNs):** Well-suited for spatial data such as images, using convolutional layers to capture local patterns.
- **Graph Neural Networks (GNNs):** These networks operate on data represented as graphs, where relationships between elements (nodes and edges) are explicitly modeled.

2.3.3.1 Long Short-Term Memory Networks

Long Short-Term Memory (LSTM) networks are a type of RNN specifically designed to capture long-term dependencies in sequential data [41]. The LSTM architecture incorporates memory cells controlled by gating mechanisms (i.e., forget, input, and output gates) which regulate the flow of information, selectively retaining or discarding data to improve time series modeling capabilities. Figure 2.1 illustrates the internal structure of an LSTM cell, showing previous and current cell states C_{t-1} , C_t , hidden states h_{t-1} , h_t , and the current input x_t [42]. The gates utilize the Sigmoid (Equation 2.4) and Tanh (Equation 2.5) activation functions to manage information flow effectively [43].

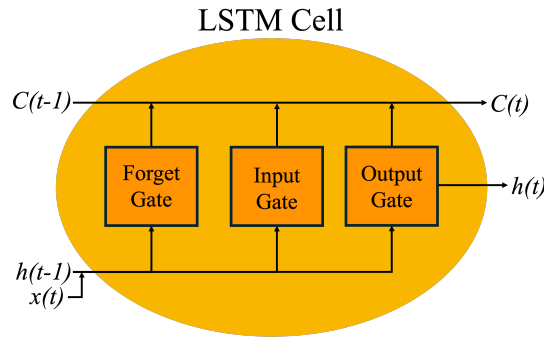


FIGURE 2.1: Workflow of an LSTM unit.

2.3.3.2 Nonlinear Autoregressive with Exogenous Inputs

Among RNN architectures, the Nonlinear AutoRegressive with eXogenous inputs (NARX) model stands out for its ability to capture dynamic nonlinear behaviors and incorporate both past inputs and outputs in the prediction process [44].

Conceptually, a NARX model is a specialized RNN where the input at time step t includes not only the current and past exogenous inputs, but also a number of previous output values. This recurrent structure enables the network to memorize temporal dependencies and accurately model system dynamics. The number of previous time steps considered defines the network delay, typically implemented using Time Delay Layers (TDLs). The activation functions used in the NARX model include the Tanh (Equation 2.5) in the hidden layer, which enables the network to capture complex nonlinear relationships, and the ReLU (Equation 2.2) in the output layer, promoting numerical stability and computational efficiency.

Depending on how past outputs are fed into the network, two main architectural configurations are used (see Figure 2.2):

- **Series-Parallel NARX (Open-Loop):** During training, the network uses the true output values from previous time steps as inputs. This architecture improves learning stability and reduces the accumulation of errors.
- **Parallel NARX (Closed-Loop):** In this configuration, the network uses its own previous predictions as feedback during inference. While more realistic for deployment, it may suffer from error propagation over time.

These architectures make NARX models highly suitable for applications requiring robust dynamic modeling with exogenous signals, offering a valuable compromise between learning stability and real-time applicability.

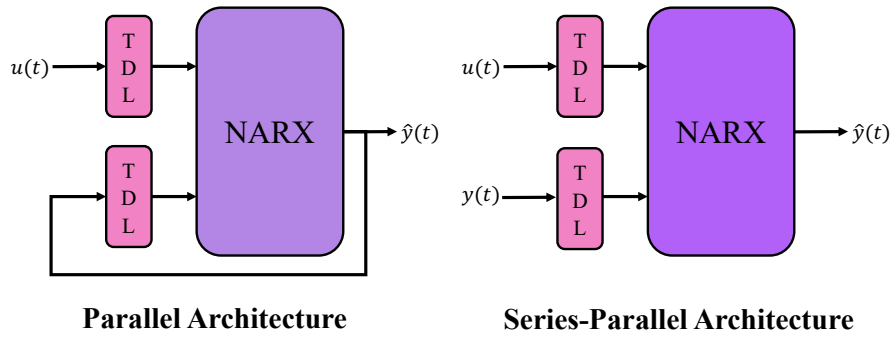


FIGURE 2.2: Series-parallel (open-loop) and parallel (closed-loop) architectures of the NARX model.

2.3.3.3 Convolutional Neural Networks

CNNs are a specialized class of DL models particularly effective in processing data with a grid-like topology, such as images (2D grids of pixels) or time series (1D sequences). Unlike FC networks, CNNs leverage spatial locality and parameter sharing to reduce computational cost and capture local patterns effectively [45].

The core idea behind CNNs is the use of convolutional layers, where learnable filters (or kernels) are applied across the input data to extract local features. These layers exploit the fact that nearby pixels or values often carry related information, allowing

the network to recognize edges, textures, or temporal patterns in early layers, and more abstract representations in deeper layers.

Key components of CNNs include:

- **Kernel (or Filter) Size:** Defines the dimensions of the filter applied to the input. For instance, a 3×3 kernel in image data inspects 3 rows and 3 columns of pixels at a time. Each filter performs an element-wise multiplication (dot product) followed by a sum, producing a single value in the output feature map. Multiple filters are used to capture different features.
- **Stride:** Indicates how much the filter shifts at each step across the input. A stride of 1 means the filter moves one unit at a time, preserving spatial resolution. Larger strides result in a reduced output size, effectively downsampling the data.
- **Padding:** Controls how the input borders are treated. Without padding, the output size is reduced after each convolution (valid padding). By adding zero-padding around the input (same padding), the output size can be preserved, which is useful in deeper networks to maintain consistent spatial dimensions.
- **Activation Function:** After each convolution operation, a nonlinear activation function such as ReLU (Equation 2.2) is applied to introduce nonlinearity, allowing the network to learn complex patterns.
- **Max-Pooling:** A downsampling operation that reduces the spatial dimensions of the feature maps by selecting the maximum value within a defined window (e.g., 2×2). Pooling helps reduce the number of parameters, control overfitting, and introduce translation invariance, making the model more robust to small shifts in the input.
- **Feature Maps and Depth:** Each filter produces a 2D feature map, and stacking multiple filters in a layer increases the depth of the output. This enables the network to learn a diverse set of features at each spatial location.

CNNs are particularly suited for image classification, object detection, and time series analysis tasks, where spatial or temporal correlations in the data are important. By learning spatial hierarchies through stacked layers, CNNs efficiently capture both low-level (edges, textures) and high-level (shapes, objects) features.

2.3.3.4 Physics-Informed Neural Networks

PINNs are a class of DL models that incorporate prior knowledge of the underlying physical systems into the learning process by embedding physical laws – typically expressed as partial differential equations, ordinary differential equations, or algebraic constraints – directly into the network’s loss function. This approach allows the network to learn solutions that are not only data-consistent but also physically plausible [46].

Unlike traditional purely data-driven models that rely on large quantities of labeled training data, PINNs leverage available physical principles (e.g., conservation laws, thermodynamics, mechanics) to guide the learning process, making them particularly valuable in scenarios where data is limited, noisy, or expensive to obtain.

The training of a PINN involves minimizing a composite loss function that typically consists of two components:

- **Data Loss:** Measures the discrepancy between the network’s prediction and available observational or experimental data.
- **Physics Loss:** Quantifies the violation of the governing equations evaluated at selected points within the problem domain, where physical constraints are explicitly enforced.

Formally, given a partial differential equation of the form:

$$\mathcal{N}[u(x, t)] = 0, \quad (x, t) \in \Omega, \quad (2.7)$$

where $\mathcal{N}[\cdot]$ is a nonlinear differential operator, PINNs seek an approximate solution $u_\theta(x, t)$, represented by a neural network with parameters θ , such that the total loss:

$$\mathcal{L}(\theta) = \mathcal{L}_{\text{data}} + \lambda \mathcal{L}_{\text{physics}} \quad (2.8)$$

is minimized, where λ is a weighting parameter that balances data fidelity and physical consistency.

PINNs are particularly well-suited for problems in fluid dynamics, heat transfer, structural mechanics, and energy systems, where the governing equations are well known but

data is often sparse or incomplete. Their ability to interpolate and extrapolate based on physical laws makes them a powerful tool for simulation, forecasting, and system identification in scientific and engineering applications.

Moreover, the flexibility of neural networks allows PINNs to be extended to handle more complex systems, including multi-physics problems, inverse problems (e.g., parameter estimation), and spatiotemporal modeling of high-dimensional systems.

2.3.3.5 Graph Attention Networks

Graphs are a fundamental data structure used to represent entities and their relationships. In a graph, the basic components are:

- **Nodes (or Vertices):** represent the entities in the system, such as people in a social network, sensors in a network, or atoms in a molecule.
- **Edges:** represent the connections or relationships between nodes. These can be directed or undirected, and may also carry weights or features.

Each node can be associated with a feature vector that describes its attributes. In many real-world applications, the structure of the graph is irregular and sparse, making traditional DL methods unsuitable. To address this, GNNs have been developed to perform learning directly on graph structures by iteratively aggregating information from a node's neighbors.

Graph Attention Networks (GATs) are an extension of GNNs that incorporate attention mechanisms to improve the learning process. Instead of treating all neighbors equally during message passing, GATs learn to assign different levels of importance to each neighbor by computing attention scores. This allows the model to focus on the most relevant or informative connections and to reduce the influence of noisy or less useful neighbors.

The key building block of GATs is the GATConv layer, which implements this attention aggregation. The GATConv layer computes attention coefficients for each edge dynamically during training, using the feature vectors of the connected nodes. These coefficients weigh the contribution of each neighbor when updating the node's representation, enabling the network to adaptively prioritize information flow. By stacking multiple

GATConv layers, the model can capture higher-order neighborhood dependencies while maintaining interpretability through learned attention weights.

GATs also support multi-head attention, where multiple independent attention mechanisms are applied in parallel. The outputs of each attention head are then concatenated or averaged, which increases the model's expressiveness and robustness. This architecture is particularly effective for heterogeneous or non-Euclidean data where the connectivity pattern is irregular and data relationships are complex.

GATs are particularly effective in handling:

- **Heterogeneous graphs**, where node connections are irregular or of varying significance.
- **Inductive learning tasks**, where the model needs to generalize to previously unseen nodes or subgraphs.
- **Applications with noisy or incomplete topologies**, as the attention mechanism can adaptively suppress irrelevant neighbors.

This flexibility makes GATs suitable for a wide range of applications, including citation networks, recommender systems, traffic forecasting, molecular property prediction, and more [47]. Their ability to learn task-specific node interactions without requiring prior knowledge of the graph structure gives them a significant advantage over traditional graph models.

2.3.3.6 Training Parameters

The performance, convergence speed, and generalization ability of DL models are strongly influenced by several training parameters. Proper tuning of these parameters is crucial to ensure effective learning and to avoid issues such as underfitting, overfitting, or divergence.

- **Batch Size:** Refers to the number of training samples used to compute a single gradient update. Smaller batch sizes (e.g., 16 or 32) can introduce more noise into the gradient estimation, potentially leading to better generalization. Larger batch sizes (e.g., 128 or more) produce more stable updates and benefit from parallelism,

but may require more memory and lead to poorer generalization if not handled carefully.

- **Stride:** Indicates the step size with which the input data are traversed during the application of an operation (e.g., convolution, sampling, or sequence processing). A stride of 1 means that the operation is applied at every position or timestep, preserving the maximum amount of information. Larger strides skip over portions of the input, thereby reducing the dimensionality of the output and the computational cost, but at the risk of losing fine-grained details.
- **Epochs:** An epoch is defined as one complete pass through the entire training dataset. The total number of epochs determines how many times the model will see each training example. Too few epochs may result in underfitting, while too many may cause overfitting. Early stopping techniques can be used to mitigate this risk.
- **Learning Rate:** This parameter controls the size of the steps taken during gradient descent to update the model's weights. A learning rate that is too high can lead to unstable training or divergence, while a value too low can slow convergence or cause the model to get stuck in local minima. Learning rate schedules or adaptive learning techniques are often used to improve performance.
- **Optimizer:** An optimizer is a key component of the training process that updates the weights of the ANN based on the computed gradients from the loss function. Its goal is to minimize the loss by efficiently navigating the loss landscape. Common optimizers include Stochastic Gradient Descent and Adaptive Moment Estimation (Adam). Adam computes adaptive learning rates for each parameter, leading to faster convergence.

2.3.4 Evaluation Metrics

2.3.4.1 Classification Metrics

In classification tasks, the model's performance is evaluated by comparing predicted classes to true labels. Common metrics include:

- **Accuracy:** The ratio of correctly predicted samples to the total number of samples.

Defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (2.9)$$

where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives. Accuracy measures overall correctness but can be misleading for imbalanced datasets. The value of accuracy ranges from 0 to 1, where 1 indicates perfect prediction and 0 indicates that all predictions are incorrect.

- **Precision:** The ratio of correctly predicted positive observations to the total predicted positives. Defined as:

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (2.10)$$

Precision indicates the reliability of positive predictions. Precision can vary between 0 and 1, with 1 meaning that all predicted positives are correct and 0 meaning that none are correct.

- **Recall (or Sensitivity):** The ratio of correctly predicted positive observations to all actual positives. Defined as:

$$\text{Recall} = \frac{TP}{TP + FN}. \quad (2.11)$$

Recall measures the ability to detect all positive samples. Recall ranges from 0 to 1, where 1 means all actual positives are correctly detected and 0 means none are detected.

- **F₁-Score:** The harmonic mean of precision and recall, providing a balance between the two. Defined as:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (2.12)$$

F₁-score is especially useful when the class distribution is imbalanced. The F₁-score also varies between 0 and 1, where 1 indicates perfect precision and recall, and 0 indicates that either precision or recall is 0.

In the case of multi-class classification, evaluation metrics such as Precision, Recall, and the F_1 -score can be computed for each class individually or aggregated across all classes using various averaging methods.

One commonly used method is macro-averaging, which calculates the metric independently for each class and then computes the unweighted mean across all classes. The macro-average Precision, Recall, and F_1 -score are defined as:

$$\text{Macro-average}_{Precision} = \frac{1}{K} \sum_{k=1}^K \text{Precision}(k), \quad (2.13)$$

$$\text{Macro-average}_{Recall} = \frac{1}{K} \sum_{k=1}^K \text{Recall}(k), \quad (2.14)$$

$$\text{Macro-average}_{F_1} = \frac{1}{K} \sum_{k=1}^K F_1(k), \quad (2.15)$$

where K is the number of classes, and $\text{Precision}(k)$, $\text{Recall}(k)$, and $F_1(k)$ denote the respective metric computed for the k -th class. All macro-averaged metrics take values between 0 and 1, with 1 indicating perfect performance and 0 indicating the worst performance.

Alternative aggregation methods include micro-averaging and weighted averaging. The micro-averaging method aggregates all true positives, false negatives, and false positives across classes before computing the metric, often resulting in values equivalent to overall accuracy in multi-class problems. Weighted averaging calculates a weighted mean of class-wise metrics based on class support (number of instances per class). In balanced datasets, weighted averaging is equivalent to macro-averaging, as all class weights are equal.

2.3.4.2 Regression Metrics

For regression tasks, where the model predicts continuous values, evaluation metrics measure the difference between predicted and true values:

- **Mean Absolute Error (MAE)**: The average of absolute differences between predicted and actual values. Defined as:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (2.16)$$

where n is the total number of samples, y_i is the actual value for the i -th sample, $\hat{y}_i$ is the predicted value for the i -th sample. MAE provides a straightforward measure of average error magnitude. MAE can take any non-negative value, with 0 indicating perfect predictions.

- **Mean Squared Error (MSE)**: The average of squared differences between predicted and actual values. Defined as:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (2.17)$$

MSE penalizes larger errors more than MAE, making it sensitive to outliers. MSE can take any non-negative value, and 0 corresponds to perfect predictions.

- **Root Mean Squared Error (RMSE)**: The square root of MSE, providing error magnitude in the original units. Defined as:

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}. \quad (2.18)$$

RMSE is often preferred for interpretability. RMSE can take any non-negative value, with 0 indicating perfect predictions.

- **Mean Absolute Percentage Error (MAPE)**: The average absolute relative deviation between predicted and true values. Defined as:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|. \quad (2.19)$$

This formula returns a value in the range $[0, +\infty)$, where values closer to 0 indicate better predictive performance. MAPE can be reported either:

- **as a fraction** (dimensionless, using Equation 2.19), or
- **as a percentage**, obtained by multiplying Equation 2.19 by 100.

- **Coefficient of Determination (R^2):** The proportion of variance in the dependent variable explained by the model. Defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (2.20)$$

where $\bar{y}$ is the mean of observed values. An R^2 close to 1 indicates a good fit. The R^2 value ranges from $-\infty$ to 1, where 1 indicates perfect predictions and values less than 0 indicate that the model performs worse than simply predicting the mean.

2.4 Unsupervised Learning Methods

2.4.1 K-Means Clustering

Among unsupervised learning techniques, the K-means algorithm stands out for its computational efficiency and ease of implementation [48]. Its scalability makes it particularly suitable for large datasets, as it can handle growing data volumes without significant performance degradation. The algorithm partitions the dataset into K clusters by assigning each data point to the nearest cluster centroid and iteratively updating these centroids to minimize intra-cluster variance.

The procedure begins with the initialization of K centroids μ_k , typically selected from the dataset. Each data point x_i is then assigned to the cluster with the closest centroid, with distance usually measured via the Euclidean norm:

$$d(x_i, \mu_k) = \|x_i - \mu_k\|. \quad (2.21)$$

Once all data points have been assigned, the centroid of each cluster is recomputed as the mean of the points within the cluster:

$$\nu_k = \frac{1}{N_k} \sum_{x_i \in C_k} x_i, \quad (2.22)$$

where N_k is the number of elements in cluster k , and C_k denotes the set of points in the same cluster. This process of assignment and centroid update is repeated until convergence, which typically occurs when the centroids change minimally or after a fixed number of iterations [49].

The algorithm aims to minimize the following cost function, which represents the total within-cluster sum of squared distances:

$$J = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \nu_k\|^2. \quad (2.23)$$

Despite its strengths, K-means has certain limitations. It assumes that clusters are spherical, equally sized, and well-separated in the feature space, which may not always reflect the structure of real-world data. Moreover, it is sensitive to outliers, as extreme values can heavily distort the position of centroids. Another major drawback is its sensitivity to the initial placement of centroids, which can lead to convergence towards local minima rather than the global optimum [50].

To mitigate this issue, the K-means++ initialization method was introduced [51]. Instead of randomly selecting all centroids, the process begins with one randomly chosen point. Each subsequent centroid is selected from the remaining data points with a probability proportional to the squared distance $D(x)^2$ from the closest already chosen centroid:

$$D(x) = \min_j \|x - \mu_j\|^2, \quad (2.24)$$

$$p(x) = \frac{D(x)^2}{\sum_{x' \in X} D(x')^2}, \quad (2.25)$$

where μ_j denotes a previously selected centroid and X is the set of all data points. This strategy promotes the selection of well-separated initial centroids, improving both convergence speed and clustering quality.

Figure 2.3 illustrates the overall flowchart of the K-means algorithm using the K-means++ initialization strategy.

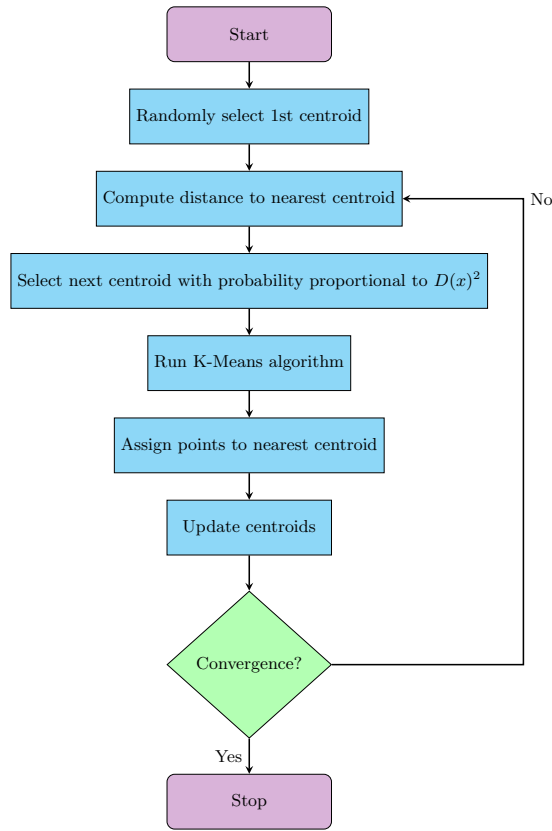


FIGURE 2.3: Flowchart of the K-means++ algorithm.

2.4.2 Spectral Clustering

Spectral Clustering is an unsupervised learning technique that leverages graph theory and linear algebra to identify clusters in data, particularly when they are non-convex or not linearly separable. Unlike traditional clustering algorithms such as K-means, which rely on geometric assumptions about the data, Spectral Clustering exploits the structure of the data represented as a graph, making it suitable for detecting complex patterns.

The first step involves constructing a similarity graph in which each data point is treated as a node, and edges represent the similarity between pairs of points. A common choice is the Gaussian kernel to define the similarity between points x_i and x_j :

$$w_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \quad (2.26)$$

where σ is a scaling parameter and w_{ij} forms the weighted adjacency matrix W of the graph.

From the similarity matrix, the graph Laplacian L is constructed. The unnormalized Laplacian is defined as:

$$L = D - W, \quad (2.27)$$

where D is the degree matrix with diagonal elements $d_{ii} = \sum_j w_{ij}$.

The core idea of Spectral Clustering is to project the original data into a lower-dimensional space using the eigenvectors of the Laplacian matrix. Specifically, the first K eigenvectors corresponding to the smallest eigenvalues (excluding the zero eigenvalue in the case of the unnormalized Laplacian) are selected to form a new feature space:

- Each data point is represented as a point in this reduced eigenvector space.
- Traditional clustering (e.g., K-means) is then applied to group the transformed points.

This technique effectively separates points based on the connectivity structure of the similarity graph rather than their direct geometric location in the original space, making it robust to irregular cluster shapes and noisy boundaries.

Spectral Clustering has proven effective in a variety of real-world tasks, including image segmentation, social network analysis, and anomaly detection, where cluster boundaries are complex or nonlinearly separable.

2.4.3 Evaluation Metrics

In clustering tasks, the evaluation of performance can rely on two main categories of metrics: supervised and unsupervised. Supervised metrics are used when ground truth labels are available, allowing a direct comparison between the predicted clusters and the actual class distribution. Unsupervised metrics, instead, do not require labels and assess the quality of the clustering based on intrinsic properties of the data, such as cohesion and separation among clusters. In practice, the choice between supervised and unsupervised metrics depends on the availability of labeled data and the specific goals of the analysis, and the two approaches are often used in a complementary way. Typical metrics are:

- **Adjusted Rand Index (ARI)**: A supervised metric that measures the similarity between predicted clusters and true labels, adjusted for chance. Defined as:

$$ARI = \frac{\sum_{ij} \binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}{0.5 \left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2} \right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}, \quad (2.28)$$

where n_{ij} is the number of samples in both cluster i and true class j , $a_i = \sum_j n_{ij}$, $b_j = \sum_i n_{ij}$, and n is the total number of samples. ARI ranges from -1 to $+1$: values near $+1$ indicate high agreement with the ground truth; 0 corresponds to random labeling; negative values indicate disagreement.

- **V-Measure (V_1)**: An unsupervised metric, equal to the harmonic mean of two complementary scores: homogeneity h , which ensures each cluster contains only members of a single class, and completeness c , where all members of a class belong to the same cluster. Defined as:

$$V_1 = 2 \cdot \frac{h \cdot c}{h + c}. \quad (2.29)$$

The score ranges from 0 to 1 , with higher values indicating better clustering quality with respect to class consistency and class coverage.

- **Silhouette**: An unsupervised metric that evaluates how well each point lies within its cluster compared to other clusters. Defined as:

$$\text{Silhouette} = \frac{b - a}{\max(a, b)}, \quad (2.30)$$

where:

$$a = \frac{1}{|C_i| - 1} \sum_{\substack{x' \in C_i \\ x' \neq x}} d(x, x') \quad (2.31)$$

is the average intra-cluster distance, and:

$$b = \min_{j \neq i} \frac{1}{|C_j|} \sum_{x' \in C_j} d(x, x') \quad (2.32)$$

is the smallest average inter-cluster distance. Values range from -1 to $+1$: a value near $+1$ indicates well-separated, compact clusters; around 0 indicates overlapping clusters; near -1 suggests incorrect clustering.

2.5 Summary

Table 2.1 provides an overview of the ML and DL methods addressed throughout the dissertation. Given the variety of tasks – ranging from classification and regression to time series forecasting and clustering – multiple models have been explored and applied. The following summary indicates where each method has been used, referencing the corresponding chapters in which they are discussed and evaluated.

TABLE 2.1: Overview of AI methods and their usage across dissertation chapters.

Method	Chapter 3	Chapter 4	Chapter 5
AdaBoost	✓		
Decision Tree	✓	✓	✓
LGBM	✓		
Random Forest	✓	✓	✓
SVR	✓		
XGBoost	✓		
LSTM	✓		
CNN	✓		
LR		✓	
SVM		✓	✓
KNN		✓	✓
ANN		✓	✓
GAT		✓	
K-means		✓	
Spectral Clustering		✓	
NARX			✓
PINN			✓

Chapter 3

Artificial Intelligence for the Prediction of Electrical Quantities

This chapter explores the use of AI and, in particular ML, for the prediction of electrical quantities. The first part compares ML-based forecasting techniques with traditional statistical models for the prediction of voltage, current, and frequency, evaluating their respective advantages and limitations in practical applications [52]. The second part focuses on the development and implementation of an AI-driven DMS, highlighting its role in real-time monitoring and prediction in smart environments [53, 54]. Finally, the chapter presents a case study on the application of an ML algorithm to predict the efficiency of Proton Exchange Membrane (PEM) fuel cell hybrid systems, demonstrating how AI-based models can enhance performance evaluation and system optimization [55].

3.1 Forecasting Electrical Quantities

One of the most promising applications of AI in power systems is the forecasting of electrical quantities (e.g., voltage, current, and frequency) which plays a critical role in ensuring optimal system operation and planning. Two main paradigms are typically adopted for this task: statistical methods and ML techniques. Statistical models rely on historical data and predefined mathematical formulations to generate forecasts, offering simplicity and interpretability. However, their ability to capture complex nonlinear patterns is often limited, and they can struggle when faced with distributional shifts in the data.

Conversely, ML approaches leverage data-driven algorithms capable of learning intricate relationships within the data, providing greater flexibility and adaptability to changing system dynamics [37, 56]. Nevertheless, ML models often demand larger datasets, greater computational resources, and may suffer from reduced interpretability compared to their statistical counterparts.

In this context, both statistical and ML models are employed to forecast electrical quantities, and their performance is systematically compared. A thorough understanding of the strengths and limitations of each approach is essential to improve forecasting reliability and, consequently, the efficiency and resilience of power systems. While numerous studies have evaluated both paradigms for power system forecasting, no definitive consensus has been reached regarding their superiority in specific scenarios. For instance, ML-based methods have been applied to voltage forecasting in Low Voltage (LV) networks [57], and to frequency prediction tasks [58, 59], while statistical approaches for frequency forecasting are discussed in [60, 61].

The contribution of this analysis lies in the systematic comparison between statistical and ML methods for the forecasting of voltage, current, and frequency. Rather than aiming to declare a superior approach, the goal is to highlight how the intrinsic characteristics of each electrical quantity influence the effectiveness of the predictive models. This comparison provides valuable insights into the advantages and limitations of both paradigms when applied to real-world scenarios.

By analyzing the performance of several methods, the study lays a foundation for the informed selection of forecasting methodologies in the development of future data-driven monitoring and control systems. In particular, it supports the design of AI-enhanced DMSs and DTs, where forecasting plays a central role in system reliability, adaptability, and energy optimization.

3.1.1 Motivation and Proposed Approach

Forecasting electrical quantities such as voltage, current, and frequency represents a key application of AI in power systems, particularly for enhancing monitoring, control, and

planning functionalities. These quantities, however, differ significantly in terms of behavior, variability, and operational relevance, factors that directly impact the effectiveness of forecasting models.

Voltage and frequency typically exhibit stable trends, with random oscillations and limited variation under normal operating conditions. Their fluctuations are generally constrained by regulatory standards (i.e., $\pm 10\%$ for voltage and $\pm 1\%$ for frequency as per [62]) and significant deviations tend to occur only during faults or abnormal events. In contrast, current shows higher variability, directly influenced by load profiles. It can undergo variations of up to $\pm 100\%$ or more of its nominal value and often displays cyclical patterns aligned with daily, weekly, or monthly usage.

Due to this diverse behavior, the selection of an appropriate forecasting method becomes context-dependent. The explored forecasting approaches include both single-step and multi-step prediction strategies. Single-step methods estimate the immediate next value in a time series, while multi-step approaches forecast multiple future instances. The latter can be implemented using either recursive methods – where each predicted value feeds into the next prediction – or direct methods, which train separate models for each future time point.

To evaluate the ML approach, six regression algorithms were selected based on their widespread use in time series prediction: AdaBoost, Decision Tree, LGBM, Random Forest, SVR, and XGBoost (as described in Section 2.3.1). These models vary in complexity and learning strategies, from ensemble techniques like boosting and bagging to kernel-based regression. Their adoption reflects a balance between performance, interpretability, and computational cost.

On the statistical side, four classical models were considered: Autoregressive (AR), Autoregressive Integrated Moving Average (ARIMA), Simple Exponential Smoothing (SES), and Holt’s method. AR models are based on regression techniques that exploit the autocorrelation structure of time series data. In these models, each new value is estimated as a linear combination of past observations and their associated errors, weighted appropriately.

Owing to their simplicity, AR models have been extended into several variants. One of the most widely used is the ARIMA model. ARIMA enhances the basic AR model

by incorporating differencing operations – subtracting the current value from its previous one – to transform a non-stationary time series into a stationary one. Stationary time series exhibit stable statistical properties over time, such as constant mean and variance, which makes them more suitable for reliable forecasting. In addition to AR and differencing components, ARIMA also includes a moving average term that accounts for the autocorrelation of past forecast errors.

Exponential Smoothing (ES) models also generate predictions by forming weighted linear combinations of past observations. Unlike AR models, ES assigns exponentially decreasing weights to older data points, reducing their influence on future predictions. Depending on the number of smoothing parameters, three main types of ES models are typically used:

- **SES:** Utilizes a single smoothing parameter, α , which determines the weight assigned to past observations.
- **Double Exponential Smoothing (Holt’s Method):** Introduces a second parameter, β , to model trends in the data.

Overall, these techniques are grounded in the analysis of temporal dependencies and trends, often requiring smaller datasets and offering greater transparency in their assumptions.

The selection of these methods was guided by their proven relevance in the literature and their complementarity in modeling different time series behaviors. By applying and comparing them across different electrical quantities, this research aims to provide a structured understanding of when and why one approach may be preferred over the other, ultimately informing the design of robust forecasting frameworks for real-world energy systems.

3.1.2 Evaluation Strategy

The proposed forecasting approaches were evaluated using a dataset composed of real and simulated electrical measurements. Specifically, Root Mean Square (RMS) voltage values and frequency values were sampled at one-minute intervals using a KEYSIGHT

3458A digital multimeter [63] connected to the LV network. Each entry represents an instantaneous measurement at the sampling time. A synthetic current signal was generated following the same sampling logic, to ensure consistency in the comparative analysis. The complete dataset spans from 16:07 on 20/02/2023 to 15:32 on 22/02/2023, for a total of 2846 data points. All electrical quantities are expressed in per unit (p.u.) values. The voltage, frequency, and current time series are shown in Figures 3.1, 3.2, and 3.3, respectively.

To assess the performance of the forecasting models, the dataset was split chronologically into a training set (80 %) and a test set (20 %). This chronological split was adopted to preserve causal ordering in the forecasts; to verify that this choice did not bias the results basic statistical properties and temporal patterns have been compared between training and test partitions, confirming that the test set is representative of the signal dynamics observed during training. Six consecutive past observations are used as input to predict the subsequent value, employing a sliding-window approach that spans the entire training set. At each time index, the six preceding values of the target series constitute the input vector, while the next value serves as the prediction target, thus generating one instance for every step in the training portion.

The ML models were evaluated in both direct and recursive multi-step configurations. All implementations were carried out in Python using the **Scikit-learn** and **Skforecast** libraries [64, 65]. Model hyperparameters (e.g., the number of trees, maximum depth, number of estimators, and the lag order) were selected on the training set by means of a grid search procedure (i.e., systematic testing of multiple hyperparameter configurations to identify the best-performing model) with MSE (Equation 2.17) as the optimization criterion.

In parallel, statistical forecasting methods were applied to the same train and test sets using the Time Series Analysis module of the **Statsmodels** library [66]. In the case of ES, a simple (non-seasonal) formulation was adopted, where the level component is updated recursively through an optimized smoothing coefficient. This allows the model to adapt to gradual changes in the signal while avoiding unnecessary complexity. Statistical forecasts were generated for the entire test horizon in a multi-step configuration, and no re-estimation was performed within the forecasting window to preserve the chronological

integrity of the evaluation. All hyperparameters were automatically estimated by maximum likelihood (i.e., by identifying the parameter values under which the data have the highest likelihood of being forecasted).

All models were evaluated using normalized MAE (Equation 2.16) and RMSE (Equation 2.18), calculated over the test portion of the dataset.

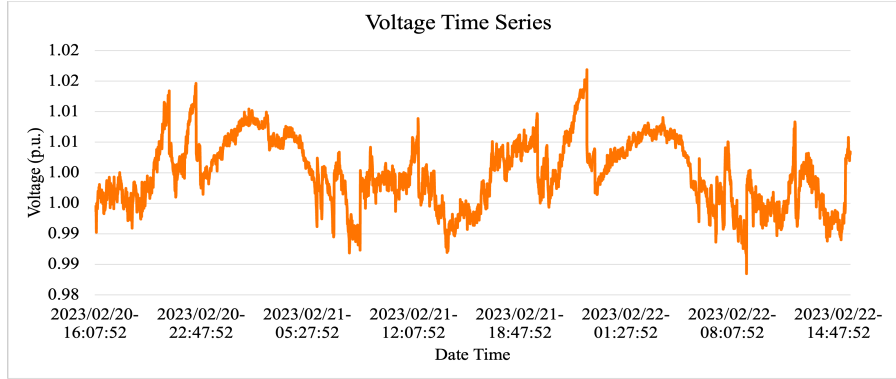


FIGURE 3.1: Voltage time series (RMS values) measured on the LV network.

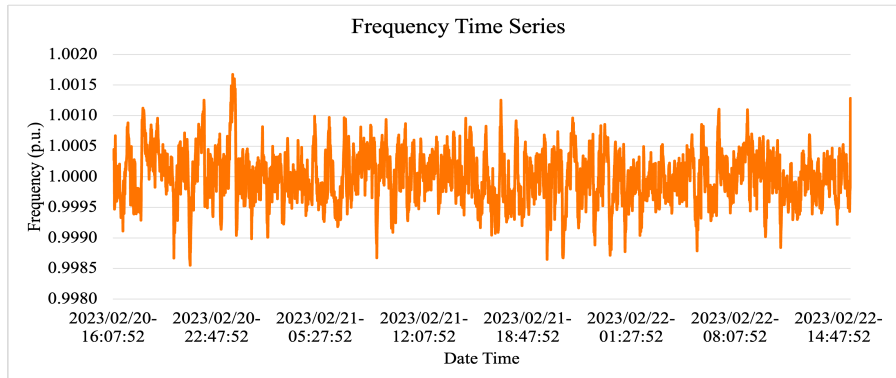


FIGURE 3.2: Frequency time series (RMS values) measured on the LV network.

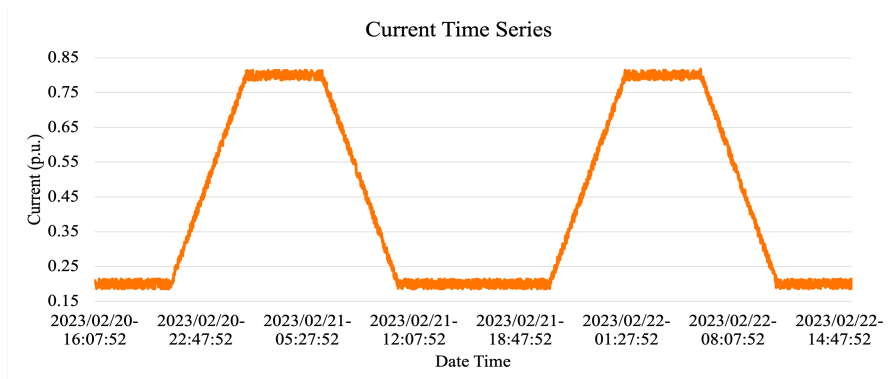


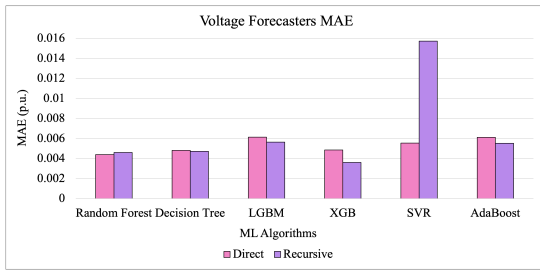
FIGURE 3.3: Simulated current time series following the same sampling logic.

3.1.3 Results and Discussion

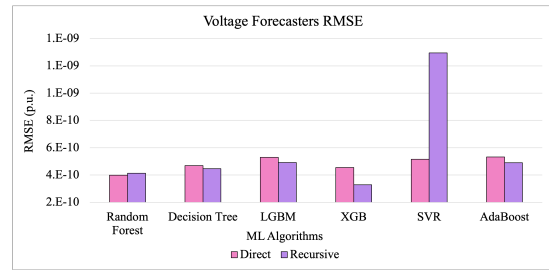
The MAE and RMSE metric results obtained from the direct and recursive implementations of the ML algorithms applied to voltage, frequency and current forecasting are shown in Figures 3.4, 3.5, and 3.6, respectively. For voltage forecasting, there is no significant disparity between direct and recursive methods; however, the recursive implementation generally achieves better performance across all algorithms except for SVR, where the recursive model exhibits errors more than twice those of the direct implementation. Similar trends emerge in frequency forecasting, with recursive approaches outperforming direct ones except in the case of the Decision Tree algorithm, where differences are minimal or negligible. On the other hand, when dealing with current forecasting, direct methods generally outperform recursive methods, with the exception of the Random Forest algorithm, which exhibits better results under the recursive approach.

The performance of the statistical forecasting methods for voltage, frequency, and current is presented in Figure 3.7. When applied to voltage and frequency, the AR and ARIMA models consistently emerge as the most accurate, demonstrating reliable results across both quantities. In the case of current forecasting, all methods show a noticeable degradation in accuracy, with higher error values observed overall. Nevertheless, AR and ARIMA continue to outperform the other statistical approaches, confirming their robustness even under more dynamic input conditions.

These results offer valuable insights into the comparative effectiveness of forecasting strategies. While both statistical and ML models were evaluated, the analysis highlights the superior adaptability and accuracy of ML-based approaches, particularly when dealing with complex or cyclical patterns such as those found in current signals. The distinction between aleatory and seasonal behaviors proves to be a crucial factor in model selection. Recursive methods are generally better suited for random, slowly varying quantities like voltage and frequency, whereas direct approaches perform better with more structured, repetitive signals such as current. Therefore, the optimal forecasting strategy must be chosen not only based on the model architecture, but also by carefully considering the temporal characteristics of the target quantity.

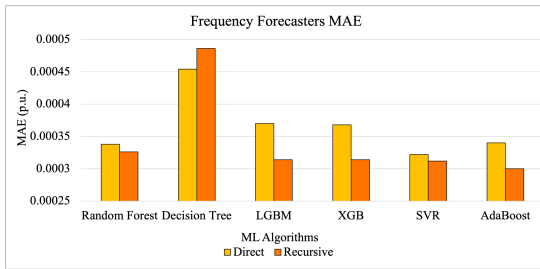


(A) MAE for voltage forecasting.

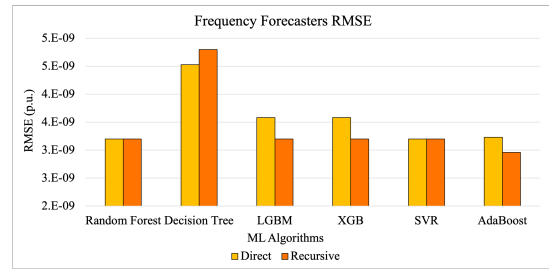


(B) RMSE for voltage forecasting.

FIGURE 3.4: Performance metrics of voltage ML-based forecasting models with direct and recursive implementations: (A) MAE values and (B) RMSE values.

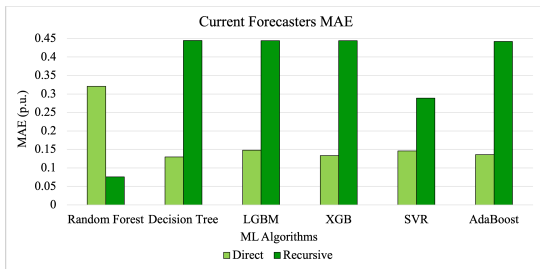


(A) MAE for frequency forecasting.

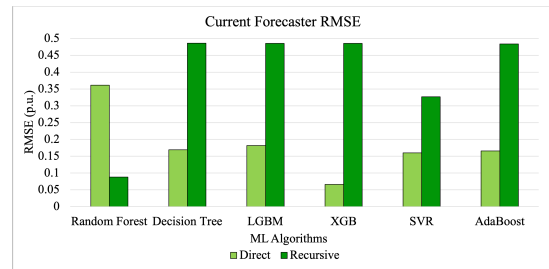


(B) RMSE for frequency forecasting.

FIGURE 3.5: Performance metrics of frequency ML-based forecasting models with direct and recursive implementations: (A) MAE values and (B) RMSE values.

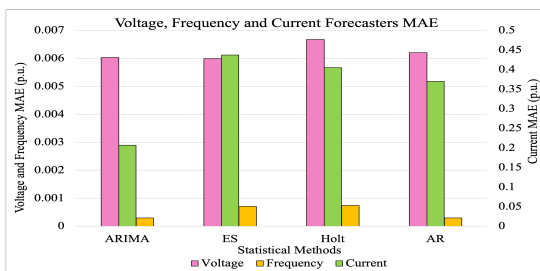


(A) MAE for current forecasting.

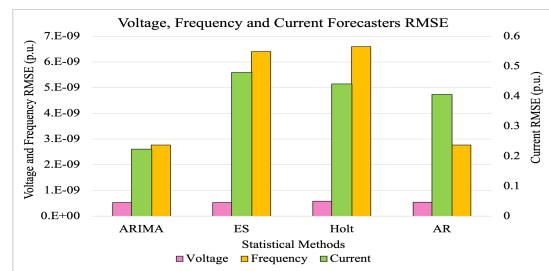


(B) RMSE for current forecasting.

FIGURE 3.6: Performance metrics of current ML-based forecasting models with direct and recursive implementations: (A) MAE values and (B) RMSE values.



(A) MAE for statistical forecasting.



(B) RMSE for statistical forecasting.

FIGURE 3.7: Performance metrics of statistical forecasting models: (A) MAE values and (B) RMSE values.

3.2 AI-Driven Distributed Measurement Systems

The integration of AI with DMSs represents a transformative paradigm for the development of smart, efficient, and sustainable energy infrastructures [67]. Accurate forecasting of electrical quantities is critical in modern power systems, enabling optimized energy management, reduced operational costs, and enhanced grid stability [68].

Time series forecasting is central in this context, with AI models demonstrating superior performance in energy management systems [69, 70]. ML and DL algorithms, such as Random Forest, SVR, and LSTM networks, are widely employed due to their capability to capture complex, nonlinear relationships in energy consumption data [71–73]. CNNs have also been applied to predict electricity consumption in buildings, successfully extracting spatial and temporal features from raw time series data [74].

DTs are a key enabler in this ecosystem, providing real-time digital replicas of physical assets. Their predictive accuracy is heavily dependent on the underlying forecasting models, highlighting the importance of integrating AI within DT frameworks. Moreover, coupling DL models with IoT-based DMSs allows real-time monitoring and control of distributed energy resources, further enhancing operational efficiency and occupant comfort in residential and commercial buildings [5, 21].

Compared to prior literature, which often addresses isolated aspects of distributed monitoring [75–77], this study presents a comprehensive AI-driven framework that unifies multiple features: time series forecasting, DT integration, high-quality DMS data handling, and model transferability across network nodes. This approach lays the foundation for scalable and interoperable AI-driven energy monitoring within DMS environments.

This study focuses on an AI-driven forecasting framework deployed in an educational building, where the DMS collects voltage, current, and power consumption data from multiple nodes. At the core of the proposed system lies an LSTM network, chosen for its proven effectiveness in capturing temporal dependencies and enhancing forecasting accuracy. The LSTM-based model is designed to operate across various nodes, thereby providing reliable and scalable monitoring throughout the building.

A distinguishing aspect of this research is the in-depth experimental evaluation, which includes model fine-tuning via extensive testing, data quality assessment, and correlation

analysis. Importantly, the framework investigates the generalization capability of the forecasting model within the DMS. This is achieved by training the LSTM model on data from one node and validating it on another, allowing the system to compensate for missing or incomplete data at individual nodes by leveraging information from other locations. Such capability supports the scalability and robustness of the monitoring system, critical for practical deployment.

Beyond forecasting accuracy, the study also addresses the practical challenges related to AI integration in DMSs. These challenges encompass sensor installation, optimization of AI models for real-world data conditions, and constraints associated with data acquisition. Moreover, the trade-offs between forecasting accuracy, computational complexity, and ease of deployment are analyzed to ensure adaptability of the solution across diverse infrastructure scenarios.

While none of these individual contributions may represent a breakthrough in isolation, their combined integration and application constitute a significant advancement, offering valuable insights to the scientific community.

This research also contributes to the broader development of DTs for energy systems. Although DT technology originated in manufacturing and aerospace sectors, its adoption is rapidly expanding into power systems [78, 79]. The full digitalization of power grids remains a complex challenge. In this context, this research demonstrates a step towards energy DT implementation by forecasting power consumption and supporting smart energy management.

3.2.1 Building Overview and Data Collection

The framework is implemented at the “Bertalia-Lazzaretto” Engineering Faculty of the University of Bologna, Italy, a large 230,000 m² complex comprising classrooms, offices, and laboratories. The building features a rooftop photovoltaic system with a capacity of 440 kW, making it an ideal environment to integrate energy consumption forecasting with solar production for enhanced energy efficiency and sustainability.

Data collection is enabled by a DMS supported by the iOtto interface from Onit [80]. The DMS monitors multiple electrical parameters (i.e., voltage, current, active and reactive power, Total Harmonic Distortion (THD), power factor, and system frequency)

across several measurement points distributed throughout the building. These points, marked in the system's block diagram represented in Figure 3.8, include two buildings powered by transformer TR2, with LV switchboards A, B, C and D (LVS-A, -B, -C, -D) supplying departments, laboratories, lecture halls, and common areas like the cafeteria.

Measurement instrumentation consists of ENERGYMID EM2389 energy meters [81] (with 1 % accuracy on active power) and Lovato Electric DM3TA series Current Transformers (CTs) [82] of class 0.5. Data are recorded every 15 minutes and have been continuously collected since June 6th, 2023. Each entry represents an instantaneous measurement at the sampling time.

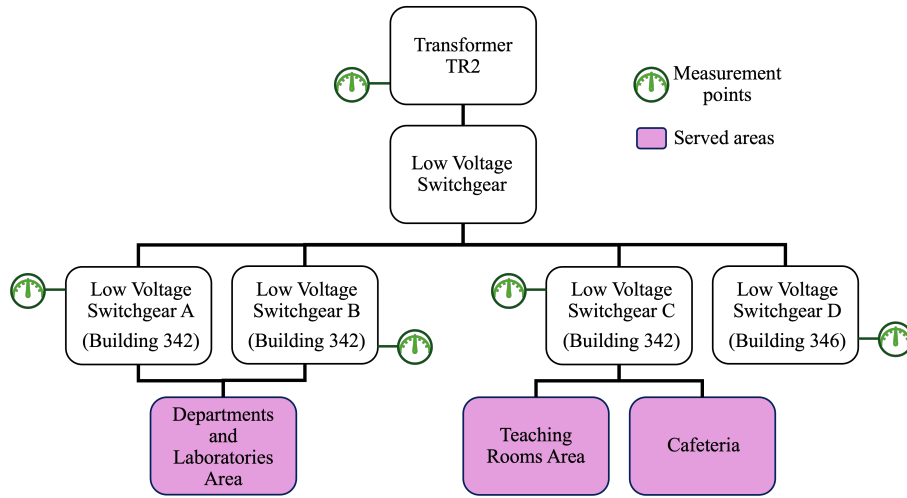


FIGURE 3.8: Diagram of the DMS reading points and served areas.

3.2.2 Artificial Intelligence Framework

The AI forecasting framework employs an LSTM network (as described in Section 2.3.3.1). The model has been implemented in Python using the `PyTorch` library [83] and comprises two LSTM layers with a hidden size of 64 units, balancing learning capacity and computational efficiency. The Adam optimizer with a learning rate of 0.001 ensures stable convergence, training the model over 50 epochs with a batch size of 256 and MSE (Equation 2.17) as the loss function.

The model is trained and tested on data collected from the DMS described previously, including timestamps and electrical parameters such as active power, current, and voltage. Data preprocessing involves circular encoding of the timestamp to properly represent cyclical temporal features via sine and cosine transformations. This circular encoding

explicitly captures seasonality at the sampling granularity (e.g., hour-of-day, day-of-week or month-of-year as appropriate), enabling the model to discriminate recurring patterns such as weekday vs weekend behavior or daily cycles without introducing discontinuities at cycle boundaries. Missing data points are interpolated, and all features are normalized to the $[0, 1]$ range. For model evaluation, the last week of each month is reserved for testing, with the remaining weeks used for training. A sliding window approach segments the data into input-output sequences, enabling the model to learn temporal dependencies; further details on sequence lengths are provided in following sections.

This AI framework aligns with the principles of DT, requiring a modular and scalable infrastructure that integrates diverse data sources and supports extensions such as AI forecasting and visualization. The DMS provides continuous, high-resolution data, while the LSTM model delivers accurate consumption forecasts. To support these requirements, the ExamonX framework [84] has been adopted. ExamonX is designed for IoT monitoring and predictive maintenance and includes components for data processing and flow management.

3.2.3 Tests Description

To evaluate the proposed forecasting framework, a set of tests has been conducted, each addressing a specific aspect of model performance and applicability within the DMS. The key test cases are the following:

1. **Performance across DMS Nodes and Electrical Quantities:** Forecasting performance was assessed at four DMS points: TR2, LVS-A, LVS-B, and LVS-D. The selection of measurement points was based on the distribution of electrical loads across the building. TR2 was chosen as a measurement point because it is the upstream transformer, providing power to the entire system. LVS-A and LVS-B were selected because they feed the same section of the building, specifically, the departments and laboratories, which include heavy machinery that can significantly influence energy consumption patterns. Lastly, LVS-D was included, as it feeds a separate building, allowing for an analysis of energy consumption in a different part of the facility. The tested quantities include the total active power across three phases, the current of the phase 1, and the voltage between phase 1 and neutral.

The considered dataset includes measurements sampled every 15 minutes, from June 6, 2023, to October 17, 2024, for a total of 47485 data points per variable per node. Model performance was first tuned on TR2, optimizing input sequence length and forecast horizon, then extended to the other nodes.

2. **Correlation Analysis of Electrical Quantities:** To investigate inter-variable dependencies, the Pearson correlation coefficient was computed between power, current, voltage, and frequency at each measurement point. The purpose of this analysis is not limited to describing statistical associations, but rather to provide a structured criterion for interpreting variable interdependence within the DMS and for supporting subsequent modeling choices. This analysis highlights how variations in one electrical quantity may be associated with changes in others, offering insights into the system's operational behavior and potential sources of redundancy among measured signals. This analysis serves three main objectives: (i) to verify the physical consistency of the acquired measurements with respect to known electrical relationships, (ii) to identify strongly correlated variables that may carry overlapping information, and (iii) to preliminarily assess whether certain quantities or measurement points exhibit sufficient coupling to justify cross-variable or cross-node modeling strategies. The Pearson coefficient r ranges from -1 (perfect negative correlation) to 1 (perfect positive correlation), with $r = 0$ indicating no linear relationship [85]. Although Pearson correlation captures only linear dependencies and does not imply causality or predictive relevance, it represents a computationally efficient and interpretable first-order tool for exploring dependencies in multivariate electrical time series. For this reason, correlation values are interpreted in conjunction with physical and operational knowledge of the electrical system, rather than being used as a strict selection or exclusion criterion. This combined approach ensures that the analysis remains informative, reproducible, and grounded in the underlying system topology and load characteristics.
3. **Model Transferability across Correlated Nodes:** To assess model generalization, the LSTM network was trained on LVS-A data and then used to forecast the data at LVS-B. This test evaluates whether the model can be transferred across spatially correlated points, a crucial capability for scaling the system or compensating for missing data in under-instrumented areas.

4. **Comparison with State-of-the-Art Methods:** The LSTM model was compared with three alternative approaches – Random Forest (described in Section 2.3.1), ARIMA, and CNN (described in Section 2.3.3.3) – using data from TR2 measurement point (power, current, and voltage readings):

- **ARIMA:** Implemented via `statsmodels` [66], with $(p = 5, d = 1, q = 0)$ configuration. Here, p denotes the order of the autoregressive component, which specifies how many past observations are used to predict the current value. d represents the degree of differencing applied to the series to achieve stationarity. q corresponds to the order of the moving average component, indicating how many past forecast errors are incorporated into the model. Seasonal components, which account for repeating patterns over fixed periods, were excluded due to excessive training times and computational complexity.
- **Random Forest:** Implemented using `Scikit-learn` [64] with 100 estimators.
- **CNN:** A custom model implemented in `PyTorch` [83] with two 1D convolutional layers (32 and 64 channels, kernel size 3), max-pooling, two FC layers (64 hidden units), ReLU activations, and the Adam optimizer (learning rate 0.001). The model was trained for 50 epochs with batch size 256 and MSE loss.

The performance of the AI-based models has been assessed using MAE (Equation 2.16), RMSE (Equation 2.18), MAPE (Equation 2.19), and R^2 (Equation 2.20) as evaluation metrics.

3.2.4 Results and Discussion

3.2.4.1 Performance across Distributed Measurement System Nodes and Electrical Quantities

The forecasting performance of the LSTM network has been evaluated on active power, current, and voltage across different measurement points within the DMS (TR2, LVS-A, LVS-B, and LVS-D). A preliminary test was conducted on the TR2 node to identify the optimal input sequence length and forecast horizon for active power prediction. Table 3.1 reports the results for input sequences of 1, 2, 4, and 24 hours (corresponding to 4, 8, 16, and 96 samples), and forecast horizons of 15 minutes, 1, 4, and 24 hours.

The results suggest that using a 2-hour input sequence (8 samples) yields a favorable trade-off between model accuracy and computational efficiency, achieving high R^2 values and low MAE and RMSE values across all time horizons. This configuration has been adopted for all subsequent evaluations.

TABLE 3.1: LSTM metrics with different input sequence lengths and time horizons for power forecasting on TR2 measurement point.

Input sequence	Time horizon	MAE (kW)	RMSE (kW)	MAPE (-)	R^2 (-)
1 h	15 min	4.06	5.24	0.04	0.98
1 h	1 h	6.19	8.28	0.05	0.94
1 h	4 h	8.26	11.33	0.07	0.88
1 h	24 h	11.60	17.37	0.10	0.73
2 h	15 min	4.50	5.77	0.03	0.99
2 h	1 h	6.45	8.50	0.04	0.93
2 h	4 h	8.03	11.11	0.05	0.90
2 h	24 h	11.67	17.19	0.08	0.76
4 h	15 min	4.83	6.13	0.04	0.97
4 h	1 h	6.41	8.44	0.06	0.94
4 h	4 h	7.96	10.94	0.07	0.89
4 h	24 h	11.64	17.23	0.10	0.73
24 h	15 min	4.60	5.96	0.04	0.97
24 h	1 h	6.30	8.29	0.06	0.94
24 h	4 h	7.84	10.72	0.07	0.90
24 h	24 h	12.21	20.12	0.11	0.64

Tables 3.2, 3.3, 3.4, and 3.5 present the metrics obtained for each quantity across different forecast horizons for TR2, LVS-A, -B, and -D, respectively.

The LSTM model demonstrated strong predictive performance across different forecast horizons, electrical quantities, and measurement points. Short-term forecasts (e.g., 15 minutes) consistently yielded the lowest error values and highest R^2 scores, confirming the model's accuracy for near-future predictions. As expected, a gradual decline in performance was observed for longer horizons due to increased uncertainty. Among the measurement points, TR2 and LVS-A achieved the most accurate results, with low MAPE values and high R^2 across both short and extended horizons. It is also worth noting that in some cases the reported MAPE appears as zero, due to extremely low absolute errors and rounding based on significant digits.

Figures 3.9, 3.10, and 3.11 respectively illustrate the actual and predicted values of power, current, and voltage at the LVS-A node over a one-week period, from July 24 to July 30. The model performs a 15-minute ahead forecast using a 2-hour input window.

TABLE 3.2: LSTM performance on current and voltage forecasting at TR2 measurement point.

Quantity	Time horizon	MAE	RMSE	MAPE (-)	R ² (-)
Current	15 min	7.89 A	10.18 A	0.05	0.96
Current	1 h	11.22 A	14.86 A	0.06	0.92
Current	4 h	13.90 A	19.15 A	0.08	0.87
Current	24 h	19.60 A	28.91 A	0.11	0.71
Voltage	15 min	0.25 V	0.35 V	0.00	0.92
Voltage	1 h	0.60 V	0.80 V	0.00	0.57
Voltage	4 h	0.81 V	1.02 V	0.00	0.40
Voltage	24 h	0.90 V	1.12 V	0.00	0.16

TABLE 3.3: LSTM performance on electrical quantities forecasting at LVS-A measurement point.

Quantity	Time horizon	MAE	RMSE	MAPE (-)	R ² (-)
Power	15 min	0.97 kW	1.29 kW	0.05	0.97
Power	1 h	1.58 kW	2.18 kW	0.08	0.91
Power	4 h	1.97 kW	2.78 kW	0.09	0.86
Power	24 h	2.47 kW	3.58 kW	0.12	0.77
Current	15 min	2.50 A	3.24 A	0.07	0.90
Current	1 h	3.31 A	4.39 A	0.09	0.82
Current	4 h	3.86 A	5.15 A	0.11	0.75
Current	24 h	4.50 A	6.26 A	0.12	0.63
Voltage	15 min	0.22 V	0.75 V	0.00	0.93
Voltage	1 h	0.62 V	0.80 V	0.00	0.57
Voltage	4 h	0.82 V	1.01 V	0.00	0.41
Voltage	24 h	0.92 V	1.14 V	0.00	0.14

TABLE 3.4: LSTM performance on electrical quantities forecasting at LVS-B measurement point.

Quantity	Time horizon	MAE	RMSE	MAPE (-)	R ² (-)
Power	15 min	0.80 kW	1.06 kW	0.06	0.96
Power	1 h	1.17 kW	1.61 kW	0.08	0.91
Power	4 h	1.48 kW	2.07 kW	0.11	0.85
Power	24 h	2.03 kW	2.92 kW	0.15	0.70
Current	15 min	2.11 A	2.89 A	0.08	0.92
Current	1 h	2.81 A	3.90 A	0.11	0.85
Current	4 h	3.58 A	4.84 A	0.14	0.77
Current	24 h	4.35 A	6.07 A	0.17	0.64
Voltage	15 min	0.21 V	0.74 V	0.00	0.92
Voltage	1 h	0.60 V	0.81 V	0.00	0.58
Voltage	4 h	0.61 V	1.03 V	0.00	0.40
Voltage	24 h	0.93 V	1.18 V	0.00	0.15

TABLE 3.5: LSTM performance on electrical quantities forecasting at LVS-D measurement point.

Quantity	Time horizon	MAE	RMSE	MAPE (-)	R ² (-)
Power	15 min	3.06 kW	4.30 kW	0.12	0.87
Power	1 h	4.16 kW	5.62 kW	0.18	0.78
Power	4 h	4.66 kW	6.47 kW	0.19	0.70
Power	24 h	5.49 kW	7.62 kW	0.23	0.59
Current	15 min	5.45 A	7.47 A	0.15	0.88
Current	1 h	7.28 A	10.24 A	0.19	0.77
Current	4 h	8.44 A	12.12 A	0.22	0.67
Current	24 h	9.96 A	14.07 A	0.27	0.56
Voltage	15 min	0.23 V	0.78 V	0.00	0.90
Voltage	1 h	0.48 V	0.79 V	0.00	0.57
Voltage	4 h	0.62 V	1.05 V	0.00	0.42
Voltage	24 h	0.95 V	1.20 V	0.00	0.16

Overall, the predictions effectively track the actual signal trends, successfully capturing recurring patterns and temporal dynamics. Slight discrepancies are observed during abrupt load variations, where the model tends to smooth sharp peaks. Nonetheless, these deviations remain within acceptable tolerance levels defined by the application context. The visual results confirm the model's good predictive capability, although additional improvements could be considered for highly dynamic conditions, which fall beyond the scope of this study.

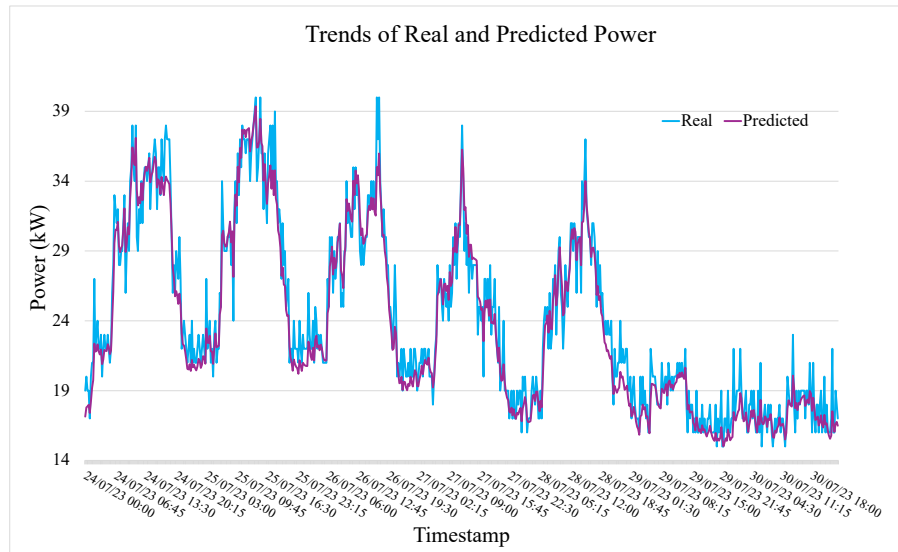


FIGURE 3.9: Actual (blue) vs. predicted (violet) values of power at LVS-A over the week of July 24–30 with 15 min-ahead forecasting and 2-hour input window.

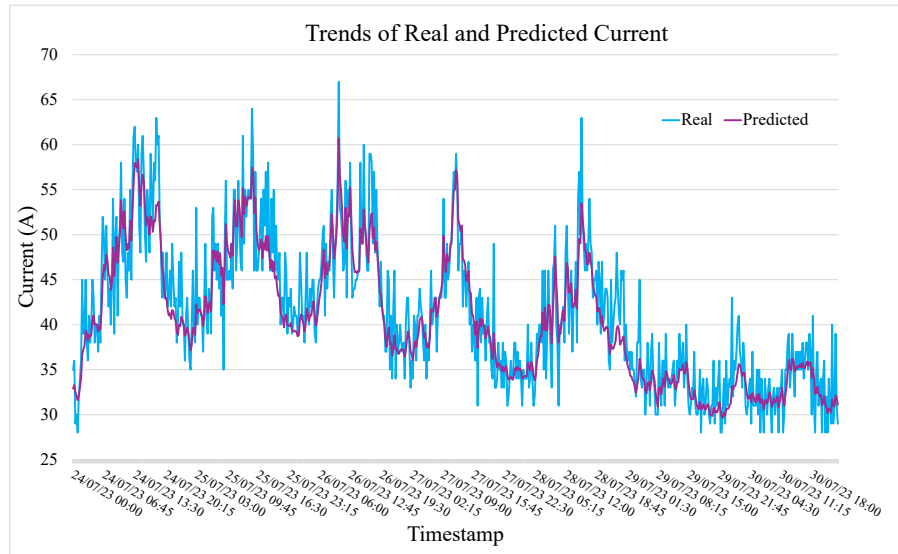


FIGURE 3.10: Actual (blue) vs. predicted (violet) values of current at LVS-A over the week of July 24–30 with 15 min-ahead forecasting and 2-hour input window.

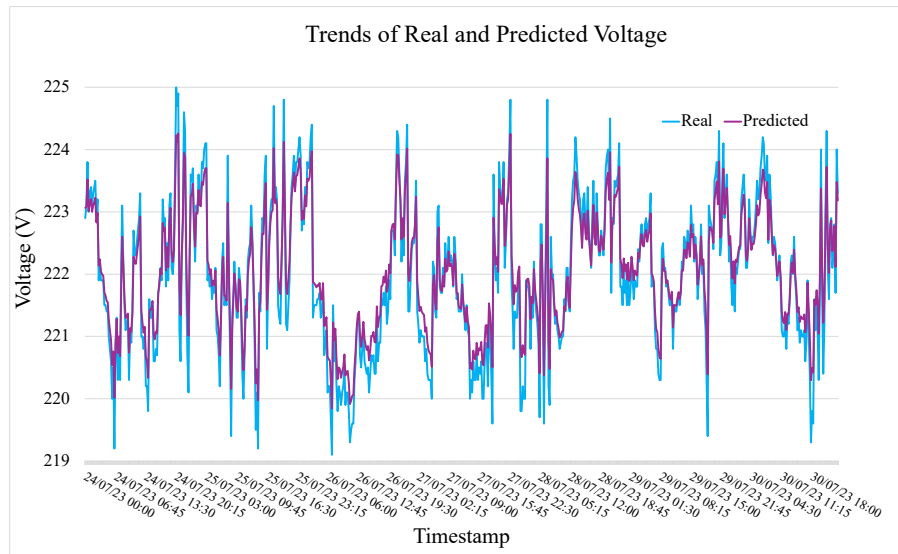


FIGURE 3.11: Actual (blue) vs. predicted (violet) values of voltage at LVS-A over the week of July 24–30 with 15 min-ahead forecasting and 2-hour input window.

3.2.4.2 Correlation Analysis Results

A correlation analysis was conducted to assess the relationships among the main electrical quantities (i.e., active power, current, voltage, and frequency) at various measurement points in the DMS. The correlation matrices for TR2, LVS-A, LVS-B, and LVS-D are shown in Figures 3.12 and 3.13.

Across all measurement points, a consistently strong positive correlation is observed between active power and current, reflecting their direct physical coupling and confirming

the coherence of the measured signals throughout the system. Frequency, on the other hand, shows negligible correlation with the remaining quantities, indicating that it is largely decoupled from local load variations and remains stable due to upstream regulation mechanisms.

Voltage exhibits a weak to moderate negative correlation with both power and current, which can be attributed to load-dependent voltage drops along the distribution lines. This behavior is observed consistently across all nodes, although with minor variations that can be ascribed to differences in local load composition and electrical distance from the upstream transformer.

To investigate spatial relationships within the DMS, the correlation between measurements at LVS-A and LVS-B was also analyzed. As illustrated in Figure 3.14, corresponding electrical quantities measured at the two nodes display a high degree of mutual correlation, particularly for active power and current. This finding is consistent with the physical layout of the system, as both nodes supply adjacent areas of the same building and are therefore influenced by similar operational schedules and load dynamics.

The observed correlation patterns indicate that certain measurement points share comparable electrical behavior, while others exhibit more independent dynamics. In this context, the results support the interpretation that LVS-A and LVS-B form a pair of strongly related nodes, whereas LVS-D shows more distinct characteristics due to its role in supplying a separate building. These outcomes provide a quantitative basis for interpreting inter-node relationships within the DMS and directly motivate the subsequent analysis of model transferability across correlated measurement points.

3.2.4.3 Model Transferability across Correlated Nodes

Table 3.6 reports the forecasting performance of the LSTM model trained on data from the LVS-A measurement point and evaluated on LVS-B data. The model was tested across several forecasting horizons – 15 minutes, 1 hour, 4 hours, and 24 hours – using 8 time steps as input.

The error metrics reveal a distinct behavior across the electrical quantities. Power and current achieve the highest R^2 values, especially for short-term forecasts, indicating that the model is able to retain a substantial portion of the variance when transferred from

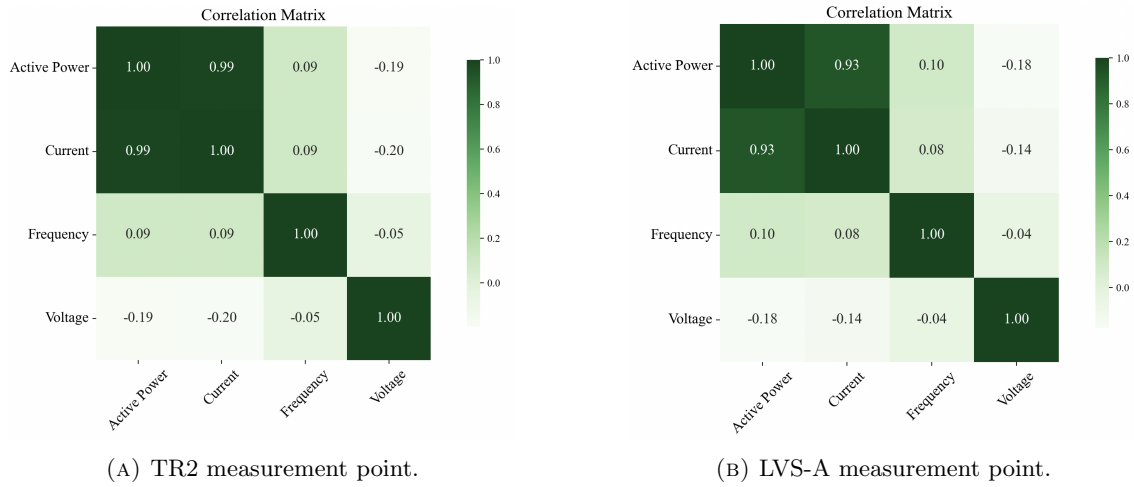


FIGURE 3.12: Correlation matrices at (A) TR2 and (B) LVS-A.

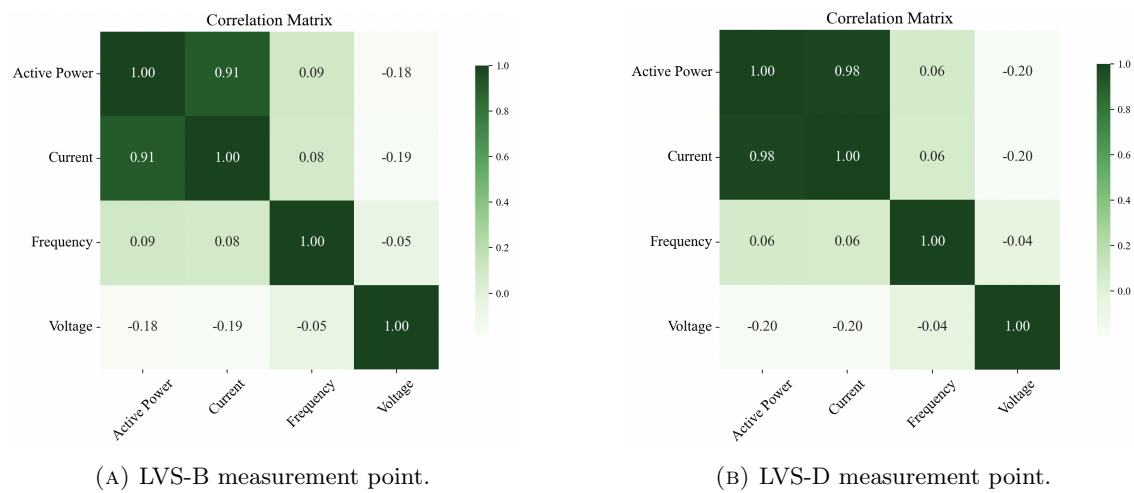


FIGURE 3.13: Correlation matrices at (A) LVS-B and (B) LVS-D.

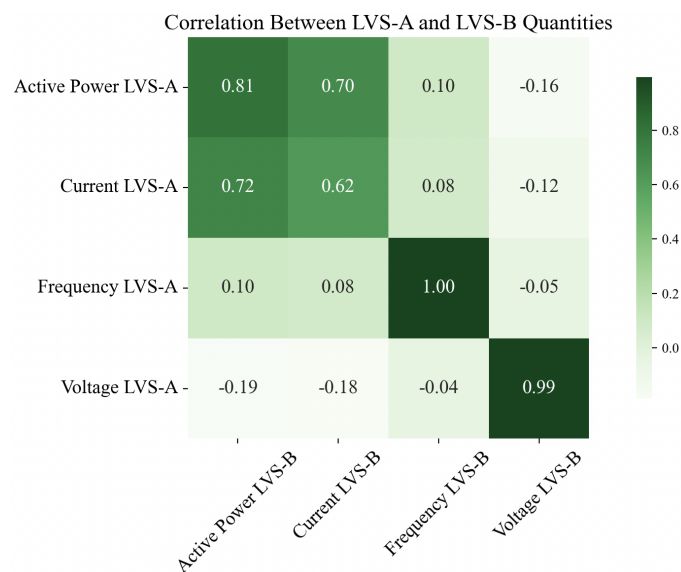


FIGURE 3.14: Correlation matrix between quantities measured at LVS-A and LVS-B.

LVS-A to LVS-B. The R^2 values, however, decrease with longer horizons, suggesting a reduced ability to capture the dynamic patterns over extended time spans.

Voltage exhibits the smallest absolute errors (MAE, RMSE and MAPE), confirming that the model produces highly accurate point predictions for this variable. Nonetheless, the R^2 values remain lower than those of power and current. This apparent discrepancy is explained by the very limited variability of voltage measurements: when the variance of the target is low, even small absolute errors translate into modest R^2 values. In this context, R^2 becomes less informative, whereas MAE and RMSE more reliably reflect predictive accuracy.

Overall, despite the degradation in performance over longer forecasting horizons – particularly for power and current – the model maintains acceptable accuracy and demonstrates a clear degree of robustness and transferability across different nodes of the distribution system. The results suggest that strong inter-node correlation facilitates model transfer, while weaker correlations limit long-horizon predictive capability but do not entirely compromise performance.

TABLE 3.6: LSTM metrics with different time horizons of electrical quantities on LVS-B measurement point with LSTM model trained on LVS-A data.

Electrical Quantity	Time Horizon	MAE	RMSE	MAPE (-)	R^2 (-)
Power	15 min	1.96 kW	2.58 kW	0.12	0.76
Power	1 h	2.78 kW	3.70 kW	0.15	0.60
Power	4 h	4.22 kW	5.50 kW	0.25	0.46
Power	24 h	5.13 kW	6.60 kW	0.34	0.32
Current	15 min	3.40 A	4.46 A	0.11	0.85
Current	1 h	4.08 A	5.65 A	0.17	0.59
Current	4 h	7.51 A	8.08 A	0.24	0.43
Current	24 h	8.98 A	10.09 A	0.31	0.30
Voltage	15 min	0.95 V	1.09 V	0.00	0.46
Voltage	1 h	1.15 V	1.75 V	0.01	0.38
Voltage	4 h	2.02 V	2.59 V	0.01	0.30
Voltage	24 h	2.82 V	3.01 V	0.02	0.26

3.2.4.4 Comparison with State-of-the-Art Methods

To benchmark the LSTM model, three widespread forecasting techniques were evaluated: Random Forest, ARIMA, and CNN. The methods were applied to power forecasting at the TR2 measurement point using a 2-hour input window and different prediction horizons.

Tables 3.7, 3.8, and 3.9 summarize the forecasting performance for each model. Among the approaches, CNN provides the most accurate results overall, particularly for shorter horizons. Random Forest offers reasonable accuracy, especially for short-term forecasts, while ARIMA yields the poorest performance, particularly for long-term forecasts.

Figure 3.15 visualizes the MAPE values for power forecasting at TR2 across all methods. It highlights the superiority of LSTM over traditional methods across all time horizons.

Table 3.10 compares the training and inference times for the three models in the scenario of a 4-hour-ahead forecasting task. While ARIMA is the most computationally efficient, its forecasting accuracy is substantially lower. LSTM and CNN, despite longer training durations, provide significantly better predictive performance.

TABLE 3.7: Random Forest metrics with different time horizons on TR2 measurement point.

Electrical Quantity	Time Horizon	MAE	RMSE	MAPE (-)	R ² (-)
Power	15 min	6.23 kW	9.24 kW	0.06	0.93
Power	1 h	7.23 kW	10.96 kW	0.07	0.90
Power	4 h	10.41 kW	13.11 kW	0.09	0.84
Power	24 h	13.85 kW	22.12 kW	0.11	0.66
Current	15 min	11.35 A	14.99 A	0.07	0.91
Current	1 h	12.99 A	17.40 A	0.08	0.88
Current	4 h	16.30 A	23.02 A	0.10	0.82
Current	24 h	22.23 A	32.04 A	0.14	0.62
Voltage	15 min	0.52 V	0.74 V	0.00	0.63
Voltage	1 h	0.77 V	0.99 V	0.01	0.34
Voltage	4 h	0.93 V	1.15 V	0.01	0.12
Voltage	24 h	1.05 V	1.32 V	0.01	0.12

3.2.4.5 Discussion

The proposed AI-based approach for short-term forecasting within the DMS demonstrates strong potential for improving energy management in complex infrastructures such as educational buildings. The integration of AI, specifically LSTM models, plays a key role in enabling the development of an effective DT. This approach not only simplifies implementation by reducing the need for detailed system models and asset-specific knowledge, but also opens opportunities for broader functionalities, such as energy quality assessment, anomaly detection, and system health monitoring. Employing a single model across multiple measurement points further enhances scalability and efficiency. Expanding the DT

TABLE 3.8: ARIMA metrics with different time horizons on TR2 measurement point.

Electrical Quantity	Time Horizon	MAE	RMSE	MAPE (-)	R ² (-)
Power	15 min	18.50 kW	25.31 kW	0.20	0.10
Power	1 h	22.60 kW	28.18 kW	0.23	0.04
Power	4 h	25.07 kW	32.92 kW	0.24	0.00
Power	24 h	28.42 kW	35.38 kW	0.26	0.00
Current	15 min	35.20 A	49.73 A	0.23	0.10
Current	1 h	36.22 A	51.01 A	0.24	0.08
Current	4 h	40.02 A	55.05 A	0.25	0.04
Current	24 h	44.25 A	58.73 A	0.25	0.00
Voltage	15 min	1.39 V	1.69 V	0.01	0.05
Voltage	1 h	1.39 V	1.69 V	0.01	0.05
Voltage	4 h	1.39 V	1.69 V	0.01	0.05
Voltage	24 h	1.39 V	1.69 V	0.01	0.05

TABLE 3.9: CNN metrics with different time horizons on TR2 measurement point.

Electrical Quantity	Time Horizon	MAE	RMSE	MAPE (-)	R ² (-)
Power	15 min	4.80 kW	6.35 kW	0.05	0.96
Power	1 h	4.33 kW	9.13 kW	0.06	0.92
Power	4 h	8.88 kW	13.01 kW	0.08	0.88
Power	24 h	13.28 kW	19.83 kW	0.10	0.73
Current	15 min	8.98 A	11.96 A	0.06	0.95
Current	1 h	13.01 A	15.93 A	0.07	0.90
Current	4 h	14.88 A	20.04 A	0.09	0.84
Current	24 h	21.20 A	30.12 A	0.13	0.70
Voltage	15 min	0.30 V	0.45 V	0.00	0.87
Voltage	1 h	0.65 V	0.84 V	0.00	0.51
Voltage	4 h	0.85 V	1.07 V	0.00	0.29
Voltage	24 h	0.98 V	1.18 V	0.00	0.15

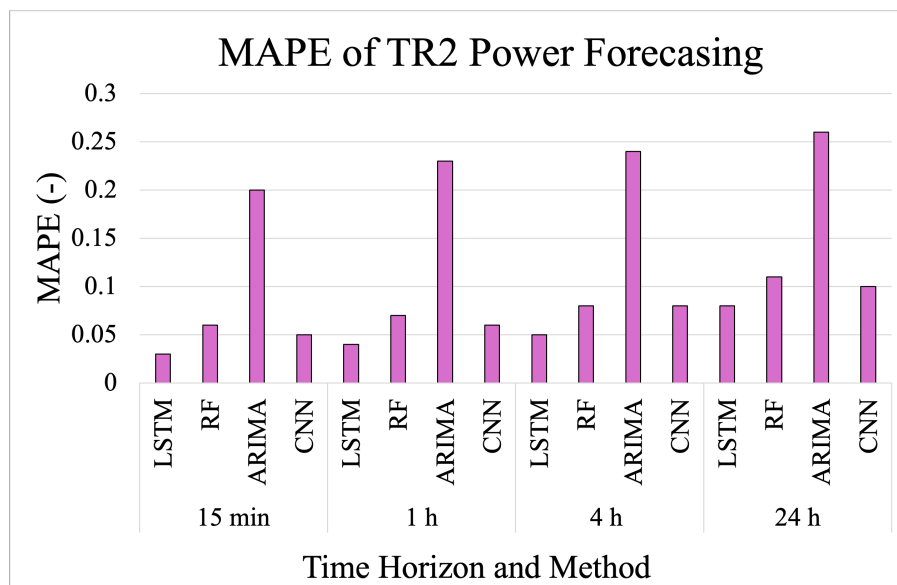


FIGURE 3.15: MAPE values for power forecasting on TR2 measurement point for LSTM, Random Forest (RF), ARIMA, and CNN.

TABLE 3.10: Training and testing time for power forecasting on TR2 measurement point for methods – 4 h time horizon.

Method	Training Time (s)	Testing Time (s)
LSTM	37.03	0.18
Random Forest	6.75	0.07
ARIMA	1.23	0.07
CNN	31.46	0.13

to include contextual data (e.g., room occupancy, schedules, holidays, and CO₂ levels) can significantly enrich its capabilities, albeit at the cost of increased system complexity. Ultimately, the combination of AI, distributed sensing, and data management in a cloud-based infrastructure offers a robust, scalable, and resilient framework for sustainable energy optimization. The system’s demonstrated ability to forecast effectively, even with incomplete data, confirms its suitability for dynamic operational environments and underscores its value as a foundation for future DT applications in energy management and predictive maintenance.

3.3 Enhancing Prediction in Fuel Cell Hybrid Systems

Sustainable mobility presents a multifaceted challenge, particularly within rapidly urbanizing contexts where up to 80 % of the global population is expected to reside in urban areas [86]. Contemporary urban planning research emphasizes the importance of city designs that promote environmental sustainability, such as medium-density, mixed-use developments integrated along efficient public transit corridors [87, 88]. These urban configurations reduce trip lengths and encourage walking, cycling, and the use of shared or public transportation, thereby minimizing car dependency without restricting personal freedom [89].

In line with these principles, modern mobility policies increasingly prioritize alternatives to private vehicle use, advocating for intermodal transport systems, cleaner propulsion technologies, and intelligent transport infrastructures [90, 91]. Among these alternatives, hydrogen-powered fuel cell systems are gaining traction, particularly for non-electrified regional rail lines, which remain prevalent in countries like Italy. Electrifying these routes is often economically unfeasible, making diesel-powered trains the default solution despite their environmental drawbacks [92, 93].

To address this, recent studies have investigated the integration of hydrogen fuel cell technologies and hybrid energy storage systems within rail vehicles [94, 95]. A promising architecture involves distributing fuel cells, batteries, and electric traction motors across powered railcars, managed via a centralized control system. This hybrid configuration enables dynamic switching between power sources to optimize efficiency.

The present research aims to explore this optimization challenge. Specifically, it investigates when each power source (fuel cell or battery) should be activated to maximize system efficiency. To this end, preliminary experimental data collected from a scaled test bench were used to train a ML model capable of predicting energy conversion efficiency under varying operating conditions. This study addresses a gap in current literature by focusing on predictive strategies for intelligent energy source management in fuel cell hybrid railway systems.

3.3.1 Data Collection

The experimental setup consists of a scaled test bench where an electric motor is powered independently by a lithium iron phosphate (LiFePO_4) battery and a PEM fuel cell. The battery is a 25.6 V, 12.0 Ah Revolithium LiFePO_4 unit, chosen for its longevity, energy density, and safety. The fuel cell system includes a Horizon H-200 PEM stack [96] (200 W, 24 V, 8.3 A) with 40 Membrane Electrode Assemblies, supplied by hydrogen stored in a hybrid metals tank and fed by an external electrolyzer.

Current and voltage from both Electrical Energy Systems (EESs) are measured with precision instruments: a $0.1\ \Omega$ shunt resistor for current and a voltage divider to scale the voltage within the acquisition board range (NI-USB 6002 [97]). A DC/DC converter stabilizes the voltage at 24 V before powering the electric motor (a MAXON EC-45 brushless motor with 200 W max power), which is controlled via a dedicated amplifier.

Additional instrumentation includes a torque meter mounted on the crankshaft for torque and rpm measurements and an electromagnetic brake for load control. The data acquisition and control are managed with a LabVIEW 2018 program, which can operate the motor at manual or automatic speed profiles to simulate load cycles.

Key parameters recorded include shaft speed (rpm), torque (Nm), voltage, current, power from EES, mechanical power, and hydrogen flow rate. From the collected data, engine and fuel cell efficiencies are calculated as in Equation 3.1 and 3.2, respectively.

$$\eta_{\text{engine}} = \frac{P_m}{P_e}, \quad (3.1)$$

where P_m is the mechanical power (W), calculated as the product of torque and angular velocity, and P_e is the electrical power (W).

$$\eta_{\text{fuel cell}} = \frac{P_e}{Q_{H_2} \cdot LHV_{H_2}}, \quad (3.2)$$

where Q_{H_2} is the hydrogen flow rate (kg/s) and $LHV_{H_2} = 120 \text{ MJ/kg}$ is the lower heating value of hydrogen.

To replicate real vehicle operation, dynamic load cycles based on EN standards [98, 99] were applied, involving speed profiles with maximum speeds of 60 km/h and 90 km/h featuring acceleration and deceleration ramps. The speed curves were adapted to the test bench, converting vehicle speed to motor rpm for feedback control. Both battery and fuel cell systems underwent five repetitions of the combined urban and highway cycles, ensuring reproducibility and consistent data for modeling and analysis.

3.3.2 Proposed Approach

The application of ML techniques to hybrid powertrain systems offers a data-driven alternative to traditional modeling methods, which often struggle to capture the complex and nonlinear relationships that govern electric motor performance when powered by multiple energy sources [100]. In this study, a Random Forest regression model (as described in Section 2.3.1) was developed to predict engine efficiency based on experimental data acquired from the described scaled test bench.

The Random Forest model was implemented using the `Scikit-learn` library [64] in Python. Hyperparameters (i.e., the number of estimators, maximum depth, and minimum samples required to split or remain in a node) were optimized through a grid search with cross-validation, ensuring the best performance on unseen data.

Key variables include acquisition time, shaft angular speed, torque, and engine efficiency (expressed as a percentage). After removing invalid efficiency values (less than or equal to 0 or greater than 100) using linear interpolation with forward/backward filling, a correlation analysis was performed using Pearson's coefficient [85]. Torque was excluded from the final feature set due to its low correlation (below 0.1) with engine efficiency. The final input features therefore consisted of: (i) acquisition time, (ii) shaft angular speed, and (iii) the categorical variable identifying the test configuration (simple trip, trip with braking, or fuel cell trip). The target variable for all experiments was the engine efficiency, expressed as a percentage. The categorical configuration variable was one-hot encoded, while shaft angular speed was scaled to the $[0,1]$ range to ensure numerical stability during training.

Each trip, both in the forward and backward directions, comprises 948 measurements per cycle, repeated over 10 cycles. For testing purposes, 20 % of the total measurements (equivalent to 1896 data points) were reserved for evaluation.

The model's performance was evaluated using MAE (Equation 2.16), RMSE (Equation 2.18), and R^2 (Equation 2.20) metrics.

3.3.3 Results and Discussion

Table 3.11 reports the model's performance for the simple train trip configuration. In the forward trip, the MAE is 14.39 %, a value that may initially appear high. However, it aligns with the distribution of engine efficiency in the dataset, which is concentrated around 50 %, with a secondary peak at 0 % and very few values near 100 %. This reflects the nature of the system, where mid-range efficiencies dominate, and extreme values are rare. The RMSE of 18.88 % is slightly higher than the MAE, indicating the presence of some outliers with greater prediction error. The coefficient of determination ($R^2 = 0.46$) suggests that the model explains nearly half of the variance in engine efficiency, indicating a moderate level of predictive accuracy. Results for the backward trip are consistent with the forward case, though with slightly higher errors, confirming the model's general robustness but also revealing potential for refinement.

Figure 3.16 shows the comparison between real and estimated values for 100 points from the test set of the simple forward trip. The model captures the trend in engine

efficiency values well, especially between 30 % and 70 %, while deviations increase near the extremes (0 % and 100 %).

Table 3.12 summarizes the metrics for the configuration involving braking. The presence of braking introduces dynamic complexity, which affects prediction performance. The backward direction shows better performance in this setup, with a higher R^2 and lower MAE and RMSE. This indicates that the model may generalize better under braking conditions in one direction, possibly due to more stable patterns in the dataset.

Finally, Table 3.13 presents the model's performance for the fuel cell-powered configuration. The results are in line with previous tests, further validating the model's moderate ability to capture engine efficiency trends under various operating scenarios.

Overall, the Random Forest model demonstrates a consistent ability to approximate engine efficiency under different powertrain configurations and operating directions. Although prediction accuracy is moderate, these results highlight the promise of data-driven methods for analyzing hybrid railway systems, while also indicating the need for further enhancements, such as incorporating additional features or exploring more advanced ML algorithms.

TABLE 3.11: Random Forest metrics on simple train trip test set.

Direction	MAE (%)	RMSE (%)	R^2 (-)
Forward	14.39	18.88	0.46
Backward	16.67	20.73	0.38

TABLE 3.12: Random Forest metrics on train trip with braking test set.

Direction	MAE (%)	RMSE (%)	R^2 (-)
Forward	17.38	21.68	0.25
Backward	15.39	19.66	0.40

TABLE 3.13: Random Forest metrics on fuel cell powered train trip test set.

Direction	MAE (%)	RMSE (%)	R^2 (-)
Forward	17.50	20.94	0.26
Backward	16.10	20.23	0.33

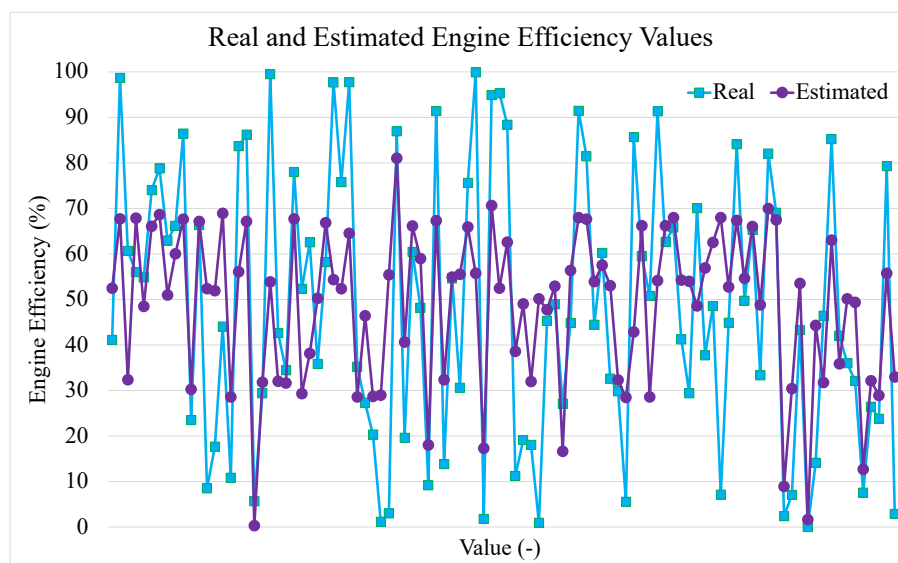


FIGURE 3.16: Actual (blue) vs. estimated (violet) engine efficiency values (100 data points from simple train trip, forward direction).

Chapter 4

Artificial Intelligence for Fault Detection and Predictive Maintenance

This chapter investigates the application of AI, with a focus on ML algorithms, for fault detection and predictive maintenance in MV networks. The first section presents supervised classification algorithms developed for CJ fault diagnostics, enabling accurate identification of fault signatures [101, 102]. The second section introduces a novel HI to quantify CJ aging, which is then estimated through advanced DL architectures to predict component health and support condition-based maintenance strategies [103]. The final section explores both supervised and unsupervised learning methods applied to MV switchgear predictive maintenance [104, 105]. Notably, the unsupervised approach is implemented on a microcontroller to evaluate the feasibility and effectiveness of an embedded, low-power diagnostic system.

4.1 Machine Learning for Cable Joint Fault Diagnostics

The expansion of cable networks is continually growing to support the electrical grid development, particularly in urban areas where overhead lines are not feasible. Moreover, compared to overhead lines, cable networks offer greater reliability and reduce visual pollution. CJs are crucial for the installation of new underground lines or the maintenance

of existing cables. However, these components are prone to failure and are considered weak points in the network [106]. CJs are challenging to maintain due to their exposure to various strains, such as mechanical, thermal, electrical, and environmental stresses. Furthermore, these components are often located several meters underground, making their maintenance even more difficult.

To address this issue, condition monitoring of CJs is essential for the early detection of potential faults [107]. This proactive approach helps utilities anticipate failures, schedule timely interventions, and optimize maintenance activities. As a result, it minimizes unplanned outages and extends the lifespan of the infrastructure. Current studies aim to measure PD to classify them [108], locate their source to identify the failure point [109], and estimate the aging state of the insulation [110]. Other research focuses on detecting faults or defects based on measurements of current [111], pressure and temperature [112], and electric field [113]. In [114] the dielectric quality is assessed by calculating the dielectric loss factor (Tan Delta). Furthermore, the relationships between these interconnected parameters are investigated in [115–117]. Finally, ambient temperature strongly affects the CJ performance, with higher failure rates in summer when temperatures are elevated [118].

As for the main topic of this research, AI has not yet been widely adopted for predictive maintenance of MV CJs. However, some examples can be found in the literature. The aim of [119] is to classify PD; while in [120] the goal is to reduce the fault false alarms. In [121], a method is introduced to assess which electrical asset is responsible for the fault, based on historical data. Finally, a way to predict the CJ temperature, 30 min in advance, has been described in [122].

Therefore, the added value of this research is the application of AI to analyze the measurements obtained from the CJs to warn the System Operator (SO) of a potential fault. Such a goal is achieved starting from the evaluation of the fault modes of a CJ and the correlated quantities (electrical or not). Afterwards, a realistic scenario is designed to be the learning platform for ML algorithms. Particularly, the problem is formulated as a binary classification task based on synthetic measurements, which are generated from structural considerations and the characteristics of the CJs. The goal is to classify each condition as faulty or non-faulty, using some of the most popular algorithms: LR, SVM, KNN, Decision Tree, Random Forest, and ANN.

4.1.1 Overview of Cable Joints

Power cables represent one of the two standard methods for expanding electric power networks. In densely populated or urban areas, they are generally the preferred choice, as the construction of overhead lines is often infeasible due to space constraints or regulatory limitations.

Because of manufacturing constraints associated with reel size, weight, and transportation limits, MV cables must be segmented. CJs are therefore required approximately every 500 – 800 m to connect consecutive cable sections. These joints, however, constitute critical points of vulnerability in MV networks, exhibiting significantly higher failure rates than the cables themselves. Replacement is particularly challenging, as joints are often buried several meters underground [123, 124].

While the expected service life of MV cables typically ranges from 30 to 40 years, CJs may begin to fail after only 7 to 8 years. Consequently, the long-term reliability and performance of MV distribution networks are strongly dependent on the durability of their joints.

Two main types of CJs are commonly employed: heat-shrinkable and cold-shrinkable joints. Heat-shrinkable joints are made of cross-linked polymeric materials that contract upon heating, ensuring adhesion to the cable insulation. In contrast, cold-shrinkable joints are composed of highly elastic silicone rubber, pre-expanded over a removable spiral core. During installation, the core is extracted, causing the joint to contract and adhere to the cable without the need for external heat.

Cold-shrinkable technology offers several advantages over its heat-shrinkable counterpart. Because cold-shrinkable joints are prefabricated under controlled factory conditions, the risk of installation errors is significantly reduced. Their high elasticity improves resistance to mechanical stress and thermal cycling, minimizing the risk of void formation and moisture ingress. Moreover, since no open flames or hot-air tools are required during installation, worksite safety is enhanced. For these reasons, cold-shrinkable joints have become the preferred solution among Distribution System Operators (DSOs) for new MV cable installations.

4.1.1.1 Classification of Cable Joints

CJs are generally categorized according to their insulation type and structural configuration. These categories serve mainly for classification convenience and do not cover all possible designs. Some joints may combine features from multiple categories. The primary classifications include:

- **Extruded Dielectric Joints:** Designed for cables insulated with extruded dielectric materials, typically rated from 2.5 kV to 500 kV.
- **Laminated Dielectric Joints:** Used with cables insulated using impregnated paper, paper/synthetic laminate tapes, or varnished fabric.
- **Transition Joints:** Enable the connection between a cable with extruded insulation and one with laminated insulation, ensuring compatibility between different dielectric types.

4.1.1.2 Construction and Installation Methods

CJs can also be classified based on their construction method and installation process. The most common types include:

- **Field-Vulcanised Joints:** Fabricated on-site using heat and pressure to cross-link the polymeric materials, resulting in a monolithic joint.
- **Filled Joints:** Consist of an external shell filled with insulating compound that occupies the space around the conductor insulation.
- **Heat-Shrinkable Joints:** Supplied as expanded polymeric components that shrink upon heating. Electrical testing prior to installation is not possible, as the stress control, insulation, and shielding layers are not integrated during molding or extrusion.
- **Cold-Shrinkable Multi-Component Joints:** Composed of two or more pre-expanded components that contract onto the cable when their internal supports are removed. Like heat-shrinkable types, these joints cannot be electrically tested before installation.

- **Pre-Molded Joints:** Factory-molded to their final shape and slid onto the prepared cable. Unlike heat- and cold-shrinkable multi-component joints, these allow for pre-installation electrical testing, as all key layers are integrated during manufacturing.
- **Cold-Shrinkable Single-Component Joints:** Delivered as a single pre-expanded component. Unlike multi-component versions, these joints permit factory electrical testing because the stress control, insulation, and shielding layers are integrated.
- **Taped Joints:** Constructed on-site by wrapping one or more layers of tape around the cable insulation. Heating may or may not be used during installation. These joints cannot be electrically tested before installation.

4.1.2 Structure

Figure 4.1 shows the generic structure of a cold-shrinkable joint, consisting of five concentric layers, numbered from the inside to the outside. At the core, (1) represents one of the two previously prepared cables to be jointed. As for (2), it is a metallic connection between the conductors of the cables. This component is subjected to the current flowing in the cable. The component (3) summarizes the first layer of insulation and semiconductive materials used to cover the conductive part. This layer is responsible for ensuring the proper insulation of the cable. Moving to (4), the metallic mesh restores the ground connection between the two pieces of cables, but it also protects the joint from external mechanical stresses. Finally, (5) is the cold shrink, as it represents the last insulation layer and its installation does not require any heat or external flame.

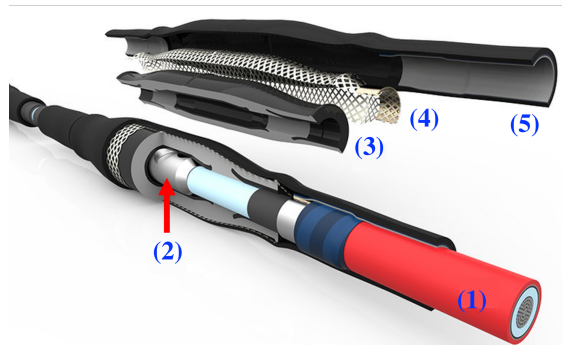


FIGURE 4.1: CJ generic multilayer structure.

4.1.3 Failure Modes

The most common failure mode affecting these accessories is the degradation of electrical insulation, typically recognized as the primary consequence of insulation wear and/or aging. However, insulation degradation is a complex phenomenon that can arise from multiple causes and lead to a variety of failure outcomes.

Other significant failure modes include: (i) the breakdown of the main or cable insulation, (ii) thermal degradation within the dielectric or failure at the conductor connection point, and (iii) damage to the sheath or at the cable-joint interface.

The primary causes of these failures are associated with the onset of PDs and the development of electrical or water treeing, phenomena that are closely interrelated. These may be triggered by several stress factors such as overvoltages, excessive current, overheating, insulation defects or voids, thermal cycling, mechanical stress, moisture ingress due to sheath damage, and thermal runaway. Additional contributing factors include the mechanical failure or loosening of components, conductor breakage or short-circuiting, and direct physical damage [111, 123, 124].

Moreover, failures can stem from long-term aging, poor manufacturing processes, the presence of manufacturing defects, insufficient mechanical strength of the external coatings, improper laying procedures, or incorrect installation.

Table 4.1 summarizes the main failure modes and typical root causes.

4.1.4 Diagnostic Parameters of Cable Joints

To ensure the long-term reliability of MV CJs and prevent failures, various diagnostic parameters can be monitored. These include environmental factors such as temperature, pressure, and humidity, as well as electrical parameters as current, voltage, electric field, PD, and Tan Delta. Among these, temperature and PD are particularly significant due to their direct link to thermal and insulation integrity issues.

TABLE 4.1: Typical failure modes and their causes.

Failure Mode	Causes of Failure
Electric breakdown of the main insulation	Electrical treeing Water treeing Overvoltage Overheating Breakdown at the interfaces
Electric breakdown of the cable insulation	Electrical treeing Water treeing Thermal cycles Mechanical stress Structural defects
Thermal breakdown	Overcurrent Overheating Water treeing Soil conditions Operational cycles
Fault at the conductors' connection point	Wrong installation Voids Design errors Overcurrent

4.1.4.1 Thermal Behavior

Overheating of MV joints during normal operation is a well-documented phenomenon [125] especially in mechanically joined aluminum conductors. The causes of overheating are not always well understood and are influenced by joint assembly methods, materials used, and quality of installation [126–128].

Studies have focused on the contact resistance at joint interfaces, particularly for aluminum and copper conductors, using both experimental analysis and finite element simulations [129–131]. These investigations aim to improve thermal performance and minimize energy loss at connection points.

Temperature increase within a joint is often correlated with internal pressure buildup during thermal cycling [112, 117, 132]. This interaction can further alter contact resistance and compromise thermal efficiency over time.

Modern MV cables commonly incorporate water-blocking materials (e.g., fillers, powders, and tapes) within their stranded conductors to prevent moisture ingress and water treeing [126, 133, 134]. However, the presence of these materials inside joints has been

shown to increase contact resistance and cause thermal degradation [135, 136]. Proper removal of water-blocking agents during conductor preparation is thus essential, although difficult in constrained installation environments such as manholes or narrow ducts.

These challenges often lead to suboptimal joint assembly, resulting in localized heating that can exceed the thermal design limits of the cable.

4.1.4.2 Partial Discharges

Partial Discharges (PDs) are localized dielectric breakdowns that occur within small voids or defects in the insulation, without bridging the entire insulation thickness [137, 138]. PD activity is a key indicator of insulation degradation and is particularly relevant in field-installed joints, which are more prone to imperfections.

PD detection is widely regarded as one of the most effective diagnostic tools for assessing insulation condition in cables and accessories [108–110]. Detection techniques are generally categorized into conventional and non-conventional methods.

The conventional method, defined by IEC 60270 [139], typically employs narrowband detection systems with upper frequency limits below 500 kHz [140]. Non-conventional methods utilize radio-frequency techniques, capturing PD-induced electromagnetic transients at frequencies up to several GHz [141–143]. These are particularly suitable for on-site applications due to better signal-to-noise ratios and the ability to locate the source of discharges.

Non-conventional PD detection methods can be further classified based on the type of coupling sensor:

- **Inductive Approach:** Uses high-frequency CTs installed around the ground or phase conductor to capture PD signals in the tens of kHz to tens of MHz range [144–146].
- **Capacitive Approach:** Involves coupling electrodes placed between the cable screen and the semiconductive outer layer, typically effective up to 100 MHz [147, 148].

- **Radiometric Approach:** Employs sensors like transient earth voltage detectors or spherical antennas to measure PD activity in the [0.1, 100] MHz range without direct electrical contact [149–152].

For long MV cable systems, distributed PD monitoring is often implemented by installing sensors at each joint and terminal [153–157]. Economic feasibility is crucial, especially for systems longer than 1 km, where cross-bonding at joint locations can facilitate PD signal extraction through intentional screen discontinuities [158, 159].

4.1.4.3 Other Relevant Parameters

In addition to temperature and PDs, several other parameters play a crucial role in the diagnostics of electrical joints. These include current, pressure, electric field, and Tan Delta, each of which provides specific information about the insulation condition and the overall integrity of the joint.

Current measurements [111], for example, can detect operational anomalies or imbalances in electrical flow that may indicate the presence of localized faults. Similarly, pressure and electric field [112, 113] are highly sensitive to discontinuities or geometric distortions within the joint, and can therefore be used to identify mechanical deformations or contact issues between insulating surfaces.

A widely used diagnostic parameter in insulation systems is Tan Delta, which expresses the ratio between the power dissipated and the power stored in a dielectric material. This parameter is strongly correlated with the quality of the insulation and is particularly affected by aging or contamination [114, 151]. High values of Tan Delta indicate increased dielectric losses, reflecting a degraded condition of the insulating material. Studies in the literature [115, 116] have also highlighted a correlation between Tan Delta, temperature, and pressure.

A crucial aspect is that all these parameters are strongly interrelated. Research has increasingly focused on understanding the mutual relationships among these quantities, with the objective of providing a comprehensive picture of the physical and degradation processes occurring within CJs. The analysis of the temporal evolution of these parameters serves as a fundamental tool for fault diagnosis and prevention, and a deep understanding

of their interdependencies is essential for improving the reliability of joints and the systems in which they are installed.

In this context, it is important to emphasize that installation conditions significantly affect not only the initiation and behavior of PDs, but also the evolution of all other diagnostic parameters. Factors such as mechanical stress, electrical contact quality, ambient humidity, and installation geometry can influence the distribution of electric fields, current response, thermal conditions, and even the value of Tan Delta. Furthermore, improper joint preparation – due to residual impurities, inaccurate surface calibration, misalignment, or defects in the coupling of insulating materials – can severely compromise the homogeneity of the dielectric system and create favorable conditions for the development of localized defects. These imperfections, often undetectable at the time of installation, may later manifest through changes in multiple diagnostic parameters simultaneously, thereby making an integrated and multi-parameter assessment of the joint even more essential.

4.1.5 Case Study

This section presents the case study designed to assess the performance of various ML algorithms in the context of CJ health monitoring.

A total of six supervised learning algorithms have been selected for comparative analysis and evaluation: LR, SVM, KNN, Decision Tree, Random Forest, and ANN, as described in Sections 2.3.1 and 2.3.3. All algorithms have been implemented using the `Scikit-learn` library [64] in Python. No hyperparameter tuning was performed, in order to assess the algorithms under baseline conditions. The ANN model was implemented as a FC MLP with a single hidden layer of 100 neurons with ReLU activation function (Equation 2.2), and a single output neuron with Sigmoid activation function (Equation 2.6) for binary classification.

CJ health diagnostics rely on the monitoring of specific quantities, which may include both electrical and environmental parameters. These measurements are typically obtained through DMSs installed at or near the CJ. Since these systems provide localized data, each measurement reflects the condition of a specific segment of the network.

In this case study, the data acquisition is simulated to represent a scenario in which a SO periodically receives measurements from the field. The four quantities considered are:

- **CJ Age:** Although not a measured quantity, it is included as a crucial diagnostic factor.
- **Current:** The current flowing through the cable, directly correlated with both temperature and load.
- **Tan Delta:** A widely recognized diagnostic electrical parameter related to insulation performance.
- **Temperature:** The temperature at the surface of the CJ, an important indicator of thermal stress.

Considering installation complexity and costs, it is realistic to assume that a DMS may monitor up to three quantities per joint. Systems designed to capture more than three or four parameters are rare due to the increased system complexity and deployment cost.

Once the quantities of interest are selected, their corresponding ranges of variation must be defined to enable coherent data generation during simulation. Table 4.2 reports the minimum and maximum values considered for each parameter.

TABLE 4.2: Variation ranges of the quantities of interest.

Quantity	Min Value	Max Value
Age (years)	0	30
Current (A)	0	300
Tan Delta (-)	10^{-5}	10^{-1}
Temperature (°C)	5	45

The rationale behind these choices is as follows:

- **CJ Age:** The expected lifespan of most electrical assets, including CJs, is approximately 30 years.
- **Current:** This parameter is directly derived from commercial datasheets of generic MV cables. The maximum value corresponds to the cable ampacity, while the minimum is set to zero to reflect no-load conditions.

- **Tan Delta:** The selected range captures values representative of CJs exhibiting very good, medium, or even poor insulation properties.
- **Temperature:** Given that MV cables are generally installed underground, soil temperature variation is relatively moderate, especially compared to ambient air temperature.

To generate data for the experimental scenarios, two types of probability distributions are employed: uniform and normal. According to the GUM [22], when no additional information is available, a uniform distribution can be reasonably assumed to represent a worst-case scenario. This rationale underpins its selection as the baseline distribution. However, when prior knowledge or operational experience is available, alternative distributions may be considered. In this research, the normal distribution has also been considered to compare the results obtained with uniformly distributed data. The same learning and testing sample sizes were employed in both cases to ensure a consistent comparison.

For the normal distribution, the mean was set to the midpoint of each interval defined in Table 4.2, while the standard deviation was calculated assuming that the interval bounds represent three standard deviations (i.e., 3σ limits).

Before assigning class labels, fault thresholds must be defined for the physical quantities under consideration. These thresholds, reported in Table 4.3, are chosen on the basis of practical engineering experience and diagnostic ranges commonly adopted for thermal and insulation monitoring. Their purpose is to identify operating conditions that may indicate abnormal behavior or incipient failure. It is important to note that the CJ age is treated differently: instead of being associated with fixed thresholds, it is linked to a probabilistic failure model based on a bathtub hazard rate curve, which reflects the typical lifetime profile of electrical assets, characterized by early-life failures, a long period of stable operation, and a final wear-out phase.

It is worth clarifying that the aging behavior is not modeled through a full parametric Weibull estimation based on field failure data, which would require long-term and statistically representative historical records that are typically unavailable for MV cable joints.

Instead, a simplified approximation of the bathtub curve is adopted to introduce a time-dependent failure probability within the synthetic data generation process. The curve is defined through a piecewise-linear interpolation of characteristic points representing the three classical aging phases: early-life failures, useful life, and wear-out. The corresponding values are selected to qualitatively reproduce the expected evolution of failure likelihood over the asset lifetime, rather than to fit a specific statistical distribution.

In practice, the aging contribution is implemented as a probabilistic gating function that modulates the occurrence of a fault condition based on the joint age. This approach enables the incorporation of temporal degradation effects into the dataset while maintaining full control over the generation process and avoiding unjustified assumptions on unavailable failure statistics. Although inspired by the Weibull-based bathtub concept commonly adopted in reliability engineering, the proposed implementation should therefore be interpreted as a phenomenological aging model aimed at supporting comparative analysis and algorithm validation, rather than as a predictive lifetime model.

A sample is labeled as healthy (0) unless the age-based failure probability exceeds 0.5. When this condition is satisfied, the sample is classified as faulty (1) if at least one variable exceeds Threshold 1, or any pair of variables exceeds Threshold 2, or all three variables exceed Threshold 3. This hierarchical scheme models increasingly permissive fault scenarios depending on how many parameters simultaneously indicate abnormal behavior. Finally, to emulate realistic measurement uncertainty, a controlled perturbation is introduced: each output label has a 10 % probability of being flipped, replicating typical sensor imperfections and avoiding the creation of an unrealistically deterministic dataset.

TABLE 4.3: Fault thresholds for input quantities.

Quantity	Threshold 1	Threshold 2	Threshold 3
Current (A)	300	150	50
Temperature (°C)	40	30	25
Tangent delta (-)	10^{-2}	10^{-3}	10^{-4}
Age (years)	Bathtub curve		

The complete flow of the case study is visually summarized in Figure 4.2.

To assess the performance of the ML models, Accuracy (Equation 2.9), Precision (Equation 2.10), and Recall (Equation 2.11) metrics are adopted. These metrics provide a comprehensive and interpretable characterization of the classifiers' behavior within the

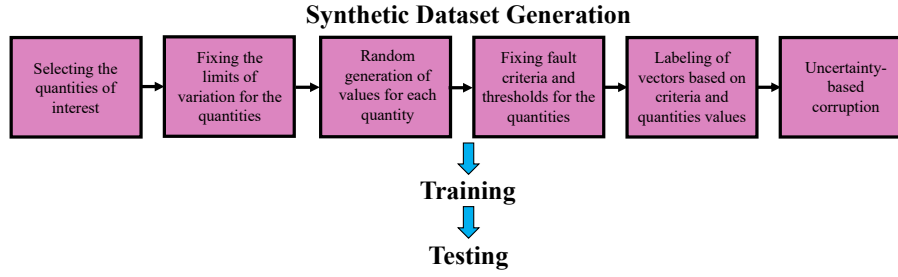


FIGURE 4.2: Flow diagram of the considered case study.

case study, supporting a reliable comparison of their effectiveness in identifying CJ health conditions.

4.1.6 Test Description

This section provides an overview of the experimental procedures adopted to evaluate the performance and reliability of the selected ML algorithms. Three different tests have been carried out: optimization of the learning sample size, repeatability analysis, and fault percentage variation analysis.

4.1.6.1 Number of Samples for the Learning Stage

Determining an appropriate number of samples for the training phase is a critical task in any machine learning-based application. An optimal training size ensures both computational efficiency and satisfactory performance. The aim of this first test is to identify, for each algorithm, the minimum number of learning samples that achieves the highest accuracy, or at least a stable one, when evaluated from a practical and realistic perspective.

To this end, each of the six algorithms was trained using increasing sizes of training datasets: 100, 200, 500, 1000, 2000, 5000, and 10000 samples. For consistency, each model was tested using a fixed testing dataset of 100000 samples. The evaluation focused on detecting when the performance gain begins to plateau, in order to define an efficient trade-off between sample size and computational effort.

4.1.6.2 Repeatability

As with physical measurements, repeatability is a fundamental requirement in AI applications. In this context, repeatability refers to the consistency of algorithmic outputs when the same configuration and input conditions are applied multiple times.

To assess repeatability, all six algorithms were executed 500 times under a fixed experimental setup: 1000 samples for the learning phase, 100000 samples for the testing phase, and a uniform distribution for data generation. The aim was to quantify the degree of variability in the results and ensure that the models provide stable performance despite the intrinsic randomness of the training process.

4.1.6.3 Fault Percentage Control

Given the structure of the data generation process, the proportion of fault cases within the dataset can vary depending on the randomly generated inputs. To better understand how this class imbalance affects algorithmic performance, an additional test was conducted by explicitly controlling the proportion of fault cases in the simulated datasets.

Specifically, data generation was constrained to stop once a predefined percentage of fault cases (i.e., 5 %, 10 %, 25 %, or 50 %) was reached. This approach allows for a controlled analysis of the sensitivity of each algorithm to different class distributions. It is worth noting that, in this particular test, the 10 % uncertainty was not applied. The exclusion of measurement uncertainty ensures that the variation in algorithmic performance is solely attributed to the fault distribution, avoiding unintended bias in the results.

4.1.7 Results and Discussion

4.1.7.1 Number of Samples for the Learning Stage

Figures 4.3, 4.4, and 4.5 compare the performance of six ML algorithms in terms of Accuracy, Precision, and Recall, respectively. Each figure displays two subplots: the left one corresponds to input data generated with a uniform distribution, and the right one to normally distributed data, allowing direct comparison of the impact of input distribution on performance.

From the accuracy results (Figure 4.3), three distinct trends are evident. LR and SVM show low and nearly flat accuracy across all training sizes, indicating limited dependence on the number of training samples. ANN and KNN exhibit gradual improvement as training samples increase, while Decision Tree and Random Forest show a steep increase that saturates beyond a certain sample size. Among all models, Decision Tree and Random Forest consistently achieve the highest accuracy.

These trends hold for both uniform and normal input distributions. Slightly better performance is observed with normally distributed data between 100 and 500 samples, suggesting a minor advantage in low-data regimes. Overall, however, accuracy is largely independent of input distribution.

Similar observations apply to precision and recall metrics (Figures 4.4 and 4.5). Random Forest and ANN lead in performance, followed closely by KNN and Decision Tree. SVM performs poorest, often resulting in zero precision and recall due to inability to detect positive cases. LR's precision fluctuates, influenced by the ratio of faulty cases during training.

Recall trends confirm the same performance hierarchy. While KNN, ANN, Decision Tree, and Random Forest converge to comparable recall values as sample size increases, LR and SVM remain significantly lower, with SVM consistently at zero recall.

In summary, the input data distribution – uniform or normal – has negligible impact on learning behavior for most algorithms. Therefore, selecting 1000 training samples remains the optimal balance between accuracy, computational efficiency, and data requirements.

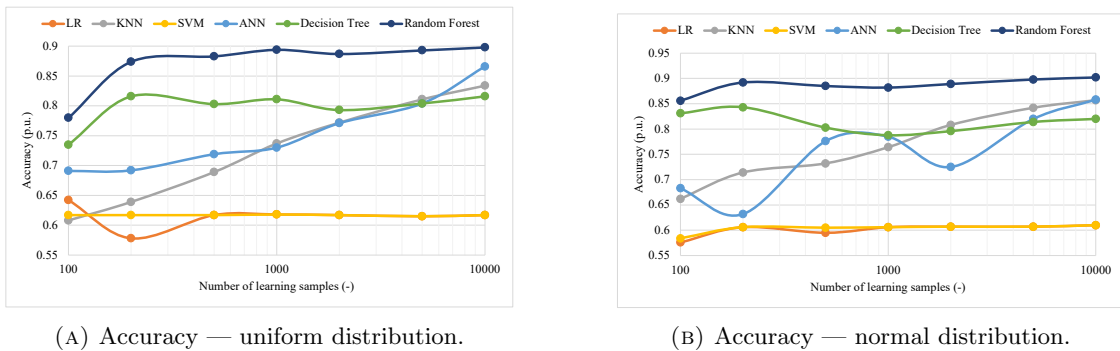
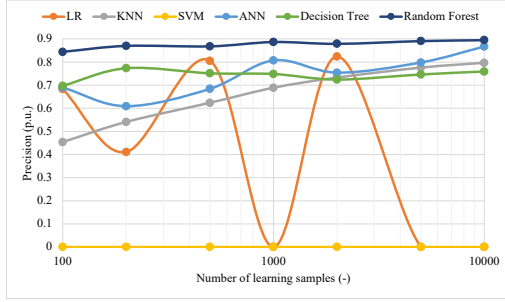
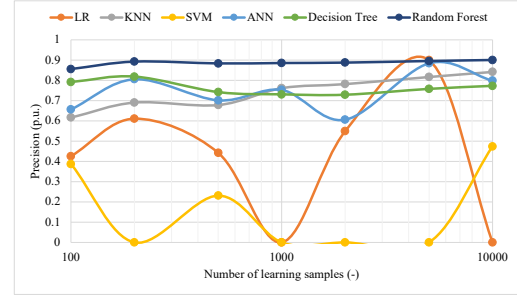


FIGURE 4.3: Accuracy vs. number of training samples for the six algorithms under different input distributions: (A) uniform and (B) normal.

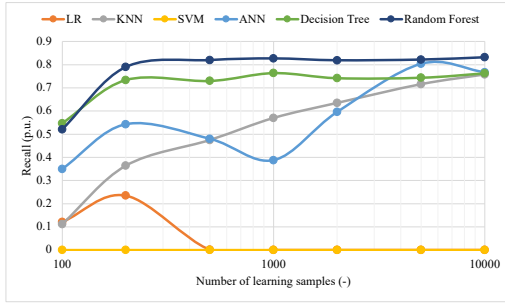


(A) Precision — uniform distribution.

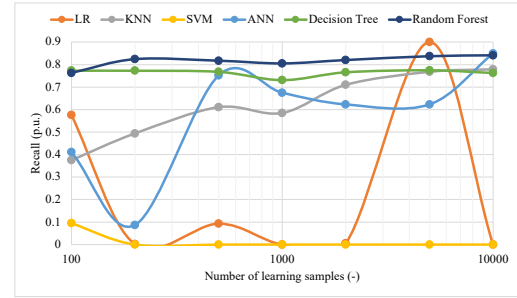


(B) Precision — normal distribution.

FIGURE 4.4: Precision vs. number of training samples for the six algorithms under different input distributions: (A) uniform and (B) normal.



(A) Recall — uniform distribution.



(B) Recall — normal distribution.

FIGURE 4.5: Recall vs. number of training samples for the six algorithms under different input distributions: (A) uniform and (B) normal.

4.1.7.2 Repeatability

Table 4.4 reports the results of the repeatability test. For each algorithm, the mean accuracy obtained over 500 repetitions and the corresponding standard deviation are presented. The tests were conducted using 1000 samples for the learning stage and uniformly distributed input data. The minimal variations observed confirm the high repeatability of the evaluated algorithms.

TABLE 4.4: Mean and standard deviation of accuracy for algorithms over 500 repetitions.

Algorithm	Mean (p.u.)	Standard Deviation (p.u.)
LR	0.62	0.01
KNN	0.730	0.007
SVM	0.616	0.002
ANN	0.75	0.02
Decision Tree	0.80	0.01
Random Forest	0.886	0.004

4.1.7.3 Fault Percentage Control

Table 4.5 summarizes the Accuracy, Precision, and Recall metrics for each algorithm at fault rates of 5 %, 10 %, 25 %, and 50 %.

TABLE 4.5: Accuracy (A), Precision (P), and Recall (R) metrics vs. percentage of faults in the training set.

Algorithm	5 %			10 %			25 %			50 %		
	A	P	R	A	P	R	A	P	R	A	P	R
LR	0.649	0	0	0.645	0	0	0.646	0	0	0.517	0.369	0.507
KNN	0.654	0.973	0.015	0.669	0.899	0.074	0.769	0.853	0.421	0.818	0.712	0.818
SVM	0.649	0	0	0.645	0	0	0.646	0	0	0.419	0.355	0.531
ANN	0.731	0.982	0.238	0.767	0.970	0.355	0.779	0.971	0.386	0.843	0.756	0.822
Decision Tree	0.982	0.999	0.950	0.982	0.999	0.959	0.991	0.980	0.994	0.992	0.980	0.970
Random Forest	0.972	0.995	0.924	0.985	0.992	0.964	0.991	0.975	0.999	0.991	0.997	0.999

From the results, Decision Tree and Random Forest show robust performance across all fault percentages, maintaining high values in all three metrics. ANN and KNN improve slightly as the percentage of faulty cases increases, achieving their best results at 50 % faults. In contrast, LR and SVM exhibit poor performance across all conditions, with marginal improvement only in precision and recall at the highest fault rate, likely due to underfitting issues identified in earlier tests.

4.1.7.4 Overall Discussion

The results demonstrate that ML algorithms offer a promising approach for fault prediction in MV CJs. Among the tested methods, Decision Tree and Random Forest consistently show superior performance, balancing accuracy and computational efficiency. However, the choice of algorithm should be tailored to the specific operational context of the SO, considering factors such as computational resources and data availability. It is also clear that the quality and characteristics of input data have a critical impact on the algorithms' effectiveness, often more so than the algorithmic complexity itself. The study highlights the importance of realistic data simulation and fault representation to ensure reliable model training. Overall, rather than seeking a perfect algorithm, operators should focus on understanding the strengths and limitations of each method to select the most appropriate solution for their needs. This research lays the groundwork for future research, particularly in integrating real-world measurement data and exploring adaptive learning strategies.

4.2 A Novel Health Index for Cable Joint Aging Prediction

Despite the growing interest in asset health assessment across power systems, the development of a reliable HI specifically tailored to CJs remains limited. As discussed in the previous sections, the degradation of CJs is a multifactorial process involving thermal, electrical, mechanical, and environmental stresses. A composite indicator that integrates these aspects is essential for accurate aging prediction and risk assessment.

The precise relationship between individual CJ parameters and their corresponding health status remains poorly defined. Assessing CJ aging through a single diagnostic metric offers only a partial understanding of the joint's condition. In contrast, an HI provides a more comprehensive evaluation by aggregating multiple physical and operational indicators. Previous studies [160, 161] have shown that techniques such as fuzzy logic can be used to assign weights to individual parameters, resulting in a single synthetic score that reflects the overall health state of the component.

Over the past 15 years, HI-based methodologies have gained widespread adoption in asset management for power utilities, enabling more effective maintenance planning and improving network reliability [162]. These indices typically include a combination of inputs such as PD activity, Tan Delta, temperature, electrical and mechanical stress, and environmental exposure [163]. Some frameworks, such as five-level HI models [164], classify asset condition based on reliability, degradation state, and associated environmental risk.

However, there is still no universally accepted method for computing the HI of power cables, largely due to variations in the selection and weighting of diagnostic parameters. These variations arise from differences in utility practices, regional operating conditions, and data availability [165]. The inconsistency across studies also hinders reproducibility and benchmarking of proposed models. When it comes to CJs, the gap is even more pronounced: the literature offers only limited insight, and no established HI model is currently dedicated to CJs.

This study aims to fill this critical gap by introducing a novel, adaptable HI specifically designed for CJs. The proposed HI formulation is flexible, allowing it to accommodate different sets of diagnostic data available to SOs, thus ensuring practical applicability across various operational contexts.

To complement the HI, a ML-based approach is adopted to estimate both the HI and the underlying CJ parameters. A dynamic graph attention model, named *GraphHealthNet*, is introduced to capture the complex, nonlinear interdependencies among parameters influencing CJ aging. Unlike static models, *GraphHealthNet* adapts to changing network conditions and data distributions, offering enhanced accuracy and robustness even under sparse or evolving datasets.

The proposed approach is validated using a publicly available dataset for MV cable condition monitoring. Emphasis is placed not on perfect modeling but on flexibility and applicability, providing SOs with a practical, customizable, and efficient tool for predictive maintenance. By combining a dynamic AI-driven inference model with an adaptable HI, this research contributes to advancing reliability-oriented diagnostics and fault prevention strategies in MV distribution networks.

4.2.1 Related Literature

The HI is a quantitative metric used to evaluate the condition and operational performance of electrical assets such as power cables and CJs. It integrates a wide range of indicators (e.g., physical, electrical, and environmental parameters, as well as historical operating conditions and maintenance records) to provide a comprehensive assessment of asset health [166].

HIIs can be expressed either as continuous values or as discrete categorical levels, akin to a traffic light system. When represented as a continuous value, the HI enables fine-grained assessment and tracking of asset deterioration over time, supporting detailed diagnostics and condition-based maintenance. Alternatively, the traffic light model classifies assets into intuitive categories such as “Green” (healthy), “Yellow” (attention required), and “Red” (critical), sometimes with additional gradations to enhance interpretability [167]. The choice of representation typically depends on the complexity of the monitoring infrastructure and the decision-making requirements of the utility or SO.

Due to its value in monitoring and decision support, the HI concept has been the subject of extensive research. Various methodologies have been proposed to translate raw inspection data into health scores. Among these, the Weighting and Scoring Method (WSM) and the Analytic Hierarchy Process (AHP) are frequently used to assign weights

to relevant indicators (e.g., PD, Tan Delta, cable age, and failure history) and compute a composite score that reflects the overall condition of an asset [168].

For instance, [169] presents an HI model for MV cables based on PD measurements, Tan Delta values, operational age, and historical failures. Similarly, [170] refines HI models for underground cables by combining WSM and AHP with actual degradation indicators, transforming static assessments into dynamic ones. Other studies, such as [171], incorporate real-time operational conditions into the HI framework to improve its practical relevance, while [172] introduces a probabilistic method for estimating the remaining life of cables, joints, and terminations.

More recently, AI techniques have been increasingly applied to HI prediction. ML and DL algorithms are capable of analyzing large volumes of historical data to model degradation trends and forecast asset conditions. For example, [173] utilizes an XGBoost algorithm, optimized via Particle Swarm optimization, to assess cable aging based on parameters such as PD, service life, corrosion levels, and load conditions.

Despite these advancements, a standardized and universally accepted approach to HI computation has yet to emerge. This is largely due to the variability of data sources, environmental conditions, and maintenance practices across different utilities and geographic regions. Moreover, while HIs are widely used to evaluate the present health of cables and joints, their predictive potential, particularly for forecasting future degradation, remains underexploited.

This gap highlights the need for advanced, AI-driven HI models that can effectively integrate heterogeneous data and adapt to different operating contexts. Such models would support more accurate aging predictions, enable proactive maintenance strategies, and ultimately improve the reliability of power distribution networks.

4.2.2 The Proposed Health Index

As previously discussed, there is a significant lack of a universally accepted and easily reproducible HI in the existing literature. This research seeks to bridge this gap by proposing an HI formulation that can be calculated by any utility or expert, based on the available data. The method relies on the widely adopted WSM, chosen for its simplicity,

adaptability, and proven effectiveness in aggregating heterogeneous parameters into a single, meaningful score. These characteristics make WSM particularly suitable for practical scenarios, where diverse and incomplete datasets must be handled efficiently.

To construct the HI, four key criteria have been identified for assigning weights to each available measurement:

1. **Measurement Availability:** Higher values are assigned to measurements that are more easily obtainable by utilities. This criterion captures the practical accessibility of the data, recognizing that some variables may not be readily available across all SOs.
2. **Independence from Operator Intervention:** A higher value corresponds to less need for manual involvement in acquiring the data. For instance, intrinsic characteristics like component age require no active measurement.
3. **Lack of Instrumentation Requirements:** This factor assigns greater importance to measurements that do not require specialized instrumentation, thereby emphasising low-cost and scalable monitoring.
4. **Diagnostic Significance:** This reflects the relative importance of the measurement in assessing the actual health status of the CJ.

These criteria were selected for their general applicability across different asset types and measurement contexts. Their use ensures that the proposed HI framework remains broadly adaptable, easily implementable, and robust in varied operational scenarios. In particular, the first three criteria (measurement availability, need for operator intervention, and instrumentation requirements) are included because a primary objective of the proposed HI is reproducibility across utilities with heterogeneous monitoring infrastructures. Different SOs may have access to different sensing technologies, measurement frequencies, or diagnostic tools; therefore, an HI that depends too strongly on hard-to-acquire or highly specialized measurements would not be comparable or scalable. Weighting these aspects allows the HI to reflect not only the diagnostic quality of each measurement, but also its practical attainability, ensuring that the index remains meaningful even when monitoring capabilities differ. By focusing on attributes that are consistently identifiable

and widely relevant, the method promotes reproducibility and scalability across utilities with differing levels of data availability and infrastructure.

Each of the four characteristics is rated on a scale from 1 to 3, and the overall HI is computed as a weighted sum of the measurement values:

$$HI = \sum_{i=1}^n (m_i \cdot w_{A,i} \cdot w_{O,i} \cdot w_{I,i} \cdot w_{S,i}), \quad (4.1)$$

where:

- m_i : normalized value of measurement i ,
- n : Total number of available measurements,
- $w_{A,i}$: Weight related to the availability of measurement i ,
- $w_{O,i}$: Weight reflecting the need for operator presence for measurement i ,
- $w_{I,i}$: Weight representing the instrumentation requirement for measurement i ,
- $w_{S,i}$: Weight for the diagnostic significance of measurement i .

Although the proposed HI depends on expert-defined weights, it is conceived as a relative indicator for asset ranking rather than an absolute measure; moderate variations of the weights or the temporary absence of individual parameters do not affect the comparability of the index over time and across assets, provided that a consistent normalization and weighting scheme is preserved.

The scale from 1 to 3 was selected as a balance between simplicity and effectiveness, based on preliminary testing. The methodology emphasizes practicality over theoretical optimization and is grounded in real-world engineering experience. The classification scheme for weights and parameters was defined based on domain expertise and is validated using actual monitoring data from cable joints. This helps ensure that the HI is not only implementable, but also meaningful in reflecting asset condition under real operating conditions.

The overall process is illustrated in Figure 4.6, which outlines the sequential steps: selecting relevant quantities, normalizing them to a common scale, assigning weights based

on the four criteria, and finally computing the HI using Equation 4.1. A major advantage of this framework is its inherent flexibility. The HI can be recalculated and adapted to different datasets and operational requirements, maintaining consistency and interpretability across contexts. Notably, a higher HI value corresponds to a worse health condition of the CJ, providing a straightforward and actionable metric for maintenance planning.

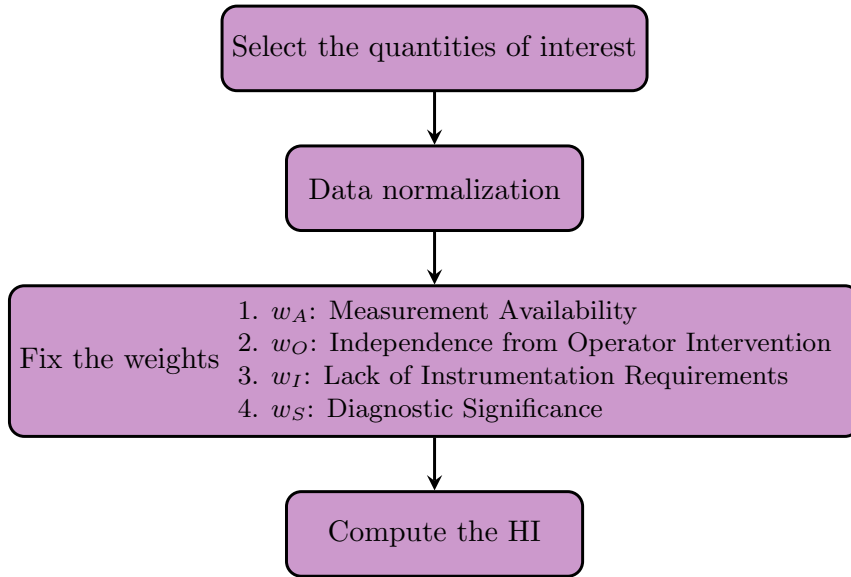


FIGURE 4.6: Block diagram for the proposed HI calculation based on the WSM.

4.2.3 Case Study

This case study demonstrates the application of the proposed HI methodology and the *GraphHealthNet* model using a real-world dataset of underground power cables. The objective is to validate the framework's ability to estimate and predict component degradation over time.

4.2.3.1 Dataset Description

To validate the proposed HI, the 20 kV XLPE Cable Historical Inspection Data public dataset [174] has been used. The dataset includes measurements of several key metrics:

- **Age:** Age of the cable.
- **PD:** Measured in picocoulombs (pC) according to IEC 60270 [175].

- **Tan Delta Metrics:** Following IEEE 400.2-2013 [176], and including:
 - **Tan Delta Stability:** Reflecting the stability of Tan Delta values over time.
 - **Tan Delta:** The differential Tan Delta, indicating changes in Tan Delta values.
 - **Mean Tan Delta:** The mean Tan Delta value representing the average condition of the insulation.
- **Visual Condition:** A categorical measure assessed on a scale from 0 to 4, where lower values indicate a worse condition.
- **Neutral Corrosion:** Measured in % using a time domain reflectometer during cable inspections.

The measurements were taken at three-year intervals in 2010, 2013, 2016, and 2019, with a total of 5489 cables evaluated each year. This dataset, which includes both electrical and physical metrics, is essential for evaluating cable aging and validating the proposed HI methodology. Specifically, cable age, PD, Mean Tan Delta, Visual Condition, and Neutral Corrosion have been selected as key parameters for HI calculation. As an example, Table 4.6 provides the values of these parameters for one cable (asset ID 1) across all inspection years (2010, 2013, 2016, and 2019).

It is important to note that, as designed, the proposed HI interprets higher values as indicative of worse conditions. Therefore, an adjustment was made to the Visual Condition feature, which originally followed an inverse logic. Specifically, the Visual Condition values have been inverted so that higher values now correspond to poorer conditions. Table 4.6 displays the adjusted values for Visual Condition. Note that, although the original dataset refers to cable inspections, the same inspection criteria, measurement procedures, and visual assessment protocols are commonly applied to CJs as well. For this reason, the dataset has been used in this work as if it were associated with CJs. This choice is further motivated by the fact that CJs are widely recognized as one of the most vulnerable components in MV networks. Since the available inspection parameters are equally meaningful for both cables and joints, and given the practical relevance of joint diagnostics, the dataset provides a suitable and representative basis for evaluating the proposed HI framework.

Following the block diagram in Figure 4.6, after selecting and normalizing the quantities of interest, the next step is to fix the respective weights. Table 4.7 shows the values

of the four weights assigned to each quantity of interest. These weights, ranging from 1 to 3, were assigned based on the four characteristics previously explained, reflecting the authors' experience and judgment. Note that these weights are somewhat arbitrary and were determined based on practical considerations rather than strict empirical data. The weights are intended to represent the relative importance of each characteristic from different points of view in the overall assessment. However, other values could potentially be used depending on different contexts or expert opinions. Finally, the HI is computed using the formula given in Equation 4.1. Table 4.8 lists the measurements and the corresponding HI for the first 10 assets in 2010.

TABLE 4.6: Measurements of interest for device 1 across the years.

Age (years)	PD (pC)	Mean Tan Delta (-)	Visual Condition (-)	Neutral Corrosion (%)
27	1237.0	3.20	2	58
30	1578.1	3.74	2	61
33	2126.1	4.15	3	65
36	2613.5	4.42	3	76

TABLE 4.7: Weights assigned to each parameter of interest for the HI calculation.

Parameter	w_A (-)	w_O (-)	w_I (-)	w_S (-)
Age	3	3	3	1
PD	1	2	1	2
Mean Tan Delta	1	2	1	2
Visual Condition	2	1	3	3
Neutral Corrosion	2	1	3	3

TABLE 4.8: Measurements of interest and HI values for devices from 1 to 10 in 2010.

ID (-)	Age (years)	PD (pC)	Mean Tan Delta (-)	Visual Condition (-)	Neutral Corrosion (%)	HI (-)
1	27	1237.0	3.20	2	58	18
2	25	1053.2	2.87	2	55	17
3	39	3089.4	5.11	3	81	27
4	41	3811.6	5.24	4	83	30
5	16	304.7	1.53	1	40	10
6	6	24.0	0.46	1	22	5
7	10	88.8	0.89	1	29	7
8	14	219.8	1.26	1	40	9
9	8	52.3	0.65	1	25	6
10	33	2167.7	4.07	3	68	23

4.2.3.2 Data Preprocessing and AI Model

To enable effective learning from historical inspection data, the dataset is preprocessed by constructing geometric graphs for each unique cable (i.e., asset ID). Each graph encodes

the temporal sequence of CJ measurements and the corresponding estimated HI. The procedure includes:

1. **Node Construction:** For each CJ, graphs are generated based on the available year combinations. Each node corresponds to a specific year, and its feature vector includes PD, Tan Delta, Visual Condition, Neutral Corrosion, and HI values.
2. **Edge Construction:** Directed edges are established according to chronological order (e.g., from 2010 to 2013, 2013 to 2016), forming a temporally structured graph.
3. **Normalization:** All features are normalized to the $[0, 1]$ range using global minimum and maximum values, ensuring comparability across features while preserving their relative variation.

The proposed *GraphHealthNet* model combines a GAT (as described in Section 2.3.3.5) and a MLP (as described in Section 2.3.3) to predict the HI and CJ condition for a future target year. The model processes geometric graphs containing measurements from previous years (e.g., 2010, 2013, 2016).

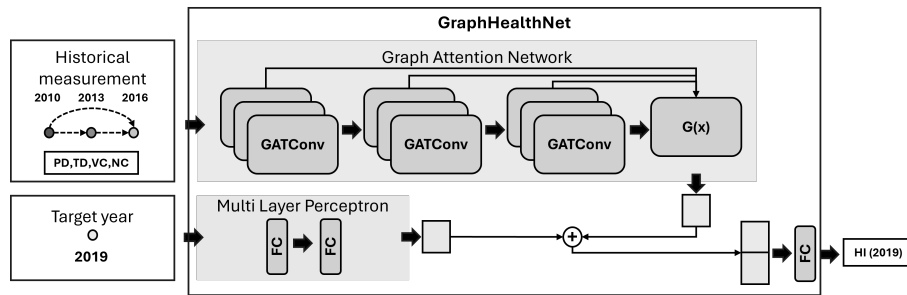


FIGURE 4.7: *GraphHealthNet* architecture combining GAT and MLP modules.

Historical data are processed using three GATConv layers followed by a FC layer. Each node represents a specific year, and edges encode the temporal order of measurements. Although each graph is temporally structured and typically contains only one or two neighbors per node, the attention mechanism still plays a meaningful role. In conventional graph convolutions, information from neighboring nodes is aggregated uniformly or according to fixed weights. In contrast, the GAT layers used here learn data-dependent attention coefficients that modulate the relative importance of each neighboring year when updating node representations. This allows the model to emphasize or down-weight specific historical measurements depending on their relevance for predicting the target year's

condition, even in graphs with simple linear connectivity. Consequently, the attention mechanism does not rely on having many neighbors, but rather enables a more flexible and adaptive temporal aggregation than standard GNNs.

Each GATConv block includes:

- **Graph Convolution:** Captures the structural dependencies between years.
- **Attention Mechanism:** Computes attention coefficients α_{ij} via:

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(a^\top [Wh_i \parallel Wh_j]))}{\sum_{k \in \mathcal{N}(i)} \exp(\text{LeakyReLU}(a^\top [Wh_i \parallel Wh_k]))}, \quad (4.2)$$

where h_i and h_j are the feature vectors of nodes i and j , respectively, and W is a learnable weight matrix that linearly transforms these features. The concatenation operator $\parallel$ joins the transformed features of the source node i and its neighbor j . This concatenated vector is then projected by a learnable attention vector a and passed through a Leaky ReLU activation function to introduce non-linearity. The exponential ensures positivity of the scores, and the denominator normalizes the coefficients over all neighbors $\mathcal{N}(i)$ of node i via a softmax function. As a result, the attention coefficient α_{ij} represents the relative importance of node j in updating the representation of node i .

- **Aggregation:** Updates node features by weighting neighbor information using the attention coefficients.

In parallel, the target year's input is processed through an MLP composed of two FC layers with LeakyReLU (Equation 2.3) activations. The outputs of GAT and MLP are concatenated and passed through additional FC layers to predict the HI and CJ measurements for the target year. The architecture supports variable input history lengths (from 1 to 3 years) and enables forecasting for multiple target years using the same historical input, ensuring flexibility and robustness.

The model optimizes a composite loss function combining MSE and Cosine Similarity Loss:

$$L_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2, \quad (4.3)$$

where y_i is the true value, $\hat{y}_i$ is the predicted value, and N is the number of data points.

$$L_{\text{Cosine}} = 1 - \frac{y \cdot \hat{y}}{\|y\| \|\hat{y}\|}, \quad (4.4)$$

where y and $\hat{y}$ represent the true and predicted vectors respectively, $y \cdot \hat{y}$ denotes their dot product, and $\|y\|$ and $\|\hat{y}\|$ are the Euclidean norms (magnitudes) of these vectors.

The total loss is defined as:

$$L_{\text{Total}} = L_{\text{MSE}} + \lambda \cdot L_{\text{Cosine}}, \quad (4.5)$$

where λ is a hyperparameter that controls the balance between magnitude accuracy and temporal trend alignment.

The model was implemented using PyTorch Geometric library [177].

4.2.4 Training and Testing Details

The dataset was split based on Asset IDs, allocating 20 % of the data (Asset IDs 1 to 1098) for testing and the remaining 80 % (Asset IDs 1099 to 5489) for training. This division ensured that both training and testing sets were representative and balanced, which is essential for robust model evaluation.

Additionally, the training set was randomly partitioned in a 1:9 ratio to create a validation set and the final training subset. Training was conducted over 500 epochs, with early stopping applied if no improvement was observed for 25 consecutive epochs, thereby preventing overfitting.

Several tests were carried out to evaluate the performance of the proposed model on the test set, which includes measurements from the years 2010, 2013, 2016, and 2019 for CJs with IDs ranging from 1 to 1098. The main tests are summarized as follows:

- **HI Forecasting:** Evaluating *GraphHealthNet* ability to predict the HI for future years using different sequences of historical data as input (e.g., only 2010, sequences 2010-2013, and 2010-2013-2016).

- **Comparison with State-of-the-Art Models:** Benchmarking *GraphHealthNet* against a baseline CNN, a state-of-the-art Conv-LSTM model [178], and an architecture Transformer-based [179].
- **Ablation Study on Input Measurements:** Assessing the importance of each measurement (PD, Tan Delta, Visual Condition, and Neutral Corrosion) by training the model while excluding one measurement at a time and observing the impact on HI forecast accuracy.
- **Direct Forecasting of Measurements:** Testing the model's ability to directly forecast future measurement values (PD, Tan Delta, Visual Condition, and Neutral Corrosion) instead of the HI, and then computing the HI from these forecasts to evaluate accuracy.

To evaluate the accuracy and performance of the model, three key metrics have been selected: R^2 (Equation 2.20), RMSE (Equation 2.18), and MAPE (Equation 2.19).

4.2.5 Results

4.2.5.1 Health Index Forecasting

Firstly, the proposed model was evaluated for HI forecasting using various input-output year sequences. Specifically, using data from 2010 as input, the model predicted the HI for subsequent years 2013, 2016, and 2019. Additionally, input sequences spanning multiple years (e.g., 2010-2013 and 2010-2013-2016) were also tested for the same forecasting task. Table 4.9 summarizes the corresponding R^2 , RMSE, and MAPE metrics for these scenarios. When using only 2010 as input, forecasts for the near future (+3 years) were more accurate than those for the farther future (+6 or +9 years), as indicated by the highlighted values. Overall, the results demonstrate that the model's performance improves as more historical input data is included. This trend is especially pronounced for the 2010-2013-2016 sequence, where the R^2 approaches 1 and both RMSE and MAPE significantly decrease, reflecting the model's enhanced robustness with richer input information. These findings are further supported by Table 4.10, which reports average metrics grouped by the number of input years. As expected, incorporating three years of historical data yields the best forecasting accuracy, followed by two years.

TABLE 4.9: Evaluation metrics of *GraphHealthNet* for HI forecasting with different input and output years.

Input Year(s) $\Rightarrow$ Forecasted Year	R^2 (-)	RMSE (-)	MAPE (-)
2010 $\Rightarrow$ 2013	0.883	0.914	0.126
2010 $\Rightarrow$ 2016	0.848	1.045	0.143
2010 $\Rightarrow$ 2019	0.814	1.207	0.176
2010, 2013 $\Rightarrow$ 2019	0.953	0.625	0.088
2010, 2013, 2016 $\Rightarrow$ 2019	0.999	0.401	0.058

TABLE 4.10: Average evaluation metrics of *GraphHealthNet* for HI forecasting grouped by the number of input years.

Number of Input Years	Input Combinations	R^2 (-)	RMSE (-)	MAPE (-)
1	2010 $\Rightarrow$ 2013; 2010 $\Rightarrow$ 2016	0.850	1.055	0.144
	2010 $\Rightarrow$ 2019; 2013 $\Rightarrow$ 2016			
	2013 $\Rightarrow$ 2019; 2016 $\Rightarrow$ 2019			
2	2010, 2013 $\Rightarrow$ 2019	0.966	0.618	0.081
	2010, 2016 $\Rightarrow$ 2019			
	2013, 2016 $\Rightarrow$ 2019			
3	2010, 2013, 2016 $\Rightarrow$ 2019	0.999	0.401	0.058

4.2.5.2 Comparison with State-of-the-Art Models

To demonstrate the advantages of *GraphHealthNet* over traditional neural network models that require fixed input shapes, it was compared against a baseline CNN (described in 2.3.3.3), a Conv-LSTM architecture [178], and a Transformer-based architecture [179].

The baseline CNN model consists of three 2D convolutional blocks with 128, 64, and 16 filters respectively, each using a (3,3) kernel and stride of (1,1), followed by batch normalization. After the convolutional layers, the output is flattened and passed through a FC layer to produce a single scalar value for forecasting. This design progressively reduces the number of filters to enable efficient feature extraction and stable training. The Conv-LSTM architecture incorporates a 1D CNN block, followed by a spatial attention module and a bidirectional LSTM for forecasting the HI. Meanwhile, the Transformer-based model employs a denoising auto-encoder followed by a transformer block with multi-headed attention to perform the forecasting task.

For *GraphHealthNet*, two architectures were evaluated: one based on simple graph convolutional layers (GCN), and the other, the proposed model, which integrates the graph attention mechanism (GATConv). These models were tested on forecasting scenarios

using data from a single year (2010). The results, shown in Figure 4.8, clearly demonstrate that *GraphHealthNet* with GATConv layers achieves superior forecasting accuracy across all metrics and forecasted years.

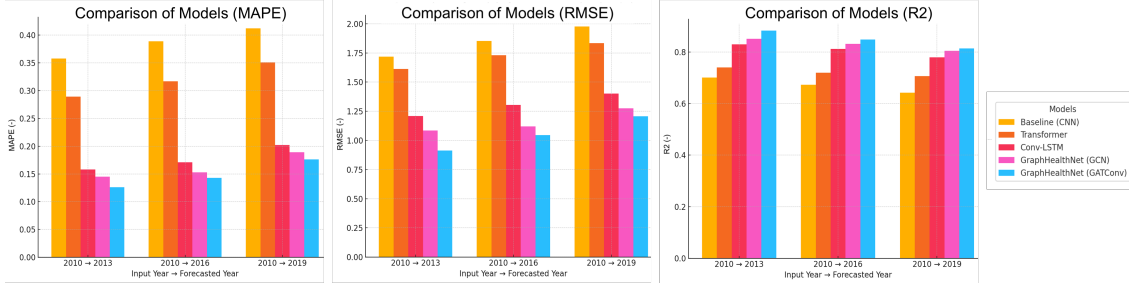


FIGURE 4.8: Comparison between state-of-the-art methods: CNN, Conv-LSTM, and Transformer architectures with *GraphHealthNet* for HI forecasting.

4.2.5.3 Ablation Study on Input Measurements

Next, the model was trained to assess the importance of each of the four measurements used to compute the HI by systematically excluding one measurement at a time. This evaluation is relevant since not all SOs have access to the full set of measurements for assessing their network’s health. The forecasting performance was quantified using the R^2 and MAPE metrics, as reported in Table 4.11. RMSE was excluded from this comparison due to its sensitivity to the measurement value ranges.

The results show that omitting any measurement consistently reduces the R^2 score and increases the MAPE across different forecasting horizons. Among all quantities, Tan Delta is the most critical: its exclusion invariably leads to the largest degradation in performance, regardless of the specific input–output year combination. For example, when predicting the HI for 2019 using historical data from 2010, 2013, and 2016, removing Tan Delta lowers the R^2 from 0.991 to 0.937 and increases the MAPE from 0.065 to 0.119, confirming its central role in accurate HI forecasting. While excluding PD also degrades performance, its impact is consistently smaller compared to Tan Delta, indicating that PD contributes meaningfully but is less influential than Tan Delta in capturing long-term degradation trends.

TABLE 4.11: Impact of excluding different input measurements on *GraphHealthNet* HI forecasting performance with various input and output years. The input measurements are abbreviated as follows: PD – Partial Discharge, TD – Tan Delta, VC – Visual Condition, and NC – Neutral Corrosion.

Input Year/s $\Rightarrow$ Forecasted Year	Input Measurements	Excluded Measurement	R^2 (-)	MAPE (-)
2010 $\Rightarrow$ 2013	All	-	0.818	0.255
	TD, VC, NC	PD	0.771	0.312
	PD, VC, NC	TD	0.736	0.347
	PD, TD, NC	VC	0.759	0.324
	PD, TD, VC	NC	0.765	0.318
2010 $\Rightarrow$ 2016	All	-	0.783	0.281
	TD, VC, NC	PD	0.741	0.323
	PD, VC, NC	TD	0.708	0.356
	PD, TD, NC	VC	0.728	0.336
	PD, TD, VC	NC	0.734	0.329
2010 $\Rightarrow$ 2019	All	-	0.759	0.324
	TD, VC, NC	PD	0.715	0.368
	PD, VC, NC	TD	0.681	0.402
	PD, TD, NC	VC	0.702	0.381
	PD, TD, VC	NC	0.707	0.376
2010, 2013 $\Rightarrow$ 2019	All	-	0.926	0.105
	TD, VC, NC	PD	0.885	0.239
	PD, VC, NC	TD	0.811	0.321
	PD, TD, NC	VC	0.849	0.255
	PD, TD, VC	NC	0.858	0.250
2010, 2013, 2016 $\Rightarrow$ 2019	All	-	0.991	0.065
	TD, VC, NC	PD	0.961	0.095
	PD, VC, NC	TD	0.937	0.119
	PD, TD, NC	VC	0.950	0.108
	PD, TD, VC	NC	0.954	0.102

4.2.5.4 Direct Forecasting of Measurements

Finally, the model was modified to directly forecast the future values of the individual measurements (PD, Tan Delta, Visual Condition, and Neutral Corrosion) instead of predicting the HI itself. The predicted measurements were subsequently used to compute the HI according to the formula in Equation 4.1, with weights specified in Table 4.7. Table 4.12 reports the error metrics for the HI obtained both from direct model prediction and from the calculated values based on the forecasted measurements.

The results indicate that direct prediction of the HI yields higher accuracy than deriving it from individually forecasted measurements. Nonetheless, the accuracy achieved in forecasting each measurement is still satisfactory, demonstrating the model's flexibility. This analysis confirms the robustness of the approach by showing its capability to predict individual measurements, which enables the calculation of a tailored HI. Overall, this highlights the model's reliability and strong performance in aging prediction tasks for distribution networks in practical applications.

TABLE 4.12: Evaluation metrics of *GraphHealthNet* for HI forecasting with different output types for various input and output years.

Input Year/s $\Rightarrow$ Forecasted Year	Output Type	R^2 (-)	RMSE (-)	MAPE (-)
2010 $\Rightarrow$ 2013	HI	0.883	0.914	0.126
	Measurements	0.764	1.045	0.158
2010 $\Rightarrow$ 2016	HI	0.848	0.848	0.143
	Measurements	0.721	1.003	0.172
2010 $\Rightarrow$ 2019	HI	0.814	0.814	0.176
	Measurements	0.693	0.956	0.194
2010, 2013 $\Rightarrow$ 2019	HI	0.953	0.625	0.088
	Measurements	0.897	0.775	0.112
2010, 2013, 2016 $\Rightarrow$ 2019	HI	0.999	0.401	0.058
	Measurements	0.988	0.584	0.072

4.2.6 Final Remarks

In this section, a novel HI for predicting the aging of CJs in power distribution networks has been introduced. This index integrates multiple relevant factors to provide a comprehensive and flexible assessment of CJ condition. Unlike conventional approaches, the proposed HI adapts to the data availability specific to SOs, enabling a tailored evaluation suited to diverse real-world scenarios.

The validation performed on a publicly available dataset has confirmed the robustness and reliability of the HI across varying data densities and operational conditions. These results highlight the practical applicability of the method and its potential to improve predictive maintenance strategies.

Overall, this research fills a significant gap in the current state-of-the-art by offering a flexible, adaptable, and data-driven tool for CJ health monitoring. The proposed HI not only enhances the understanding of asset aging but also contributes to extending the operational lifespan and reliability of power distribution networks.

4.3 Predictive Maintenance of Medium Voltage Switchgears

MV switchgears are fundamental components in power distribution systems, ensuring the safe switching, protection, and isolation of electrical circuits during both normal operation and fault conditions. Their reliable functioning is crucial for the safety, efficiency, and overall stability of the electrical grid. Due to their critical role and the potentially severe

consequences of failures, condition monitoring and predictive maintenance strategies have become focal points in both industrial practice and academic research.

Advances in sensor technologies, DMSs, and ML techniques have enabled continuous and more accurate monitoring of MV switchgear conditions. Key signals such as thermal, mechanical, and PD measurements have proven effective for early detection of degradation and incipient faults [180–186]. Among sensing options, wireless and acoustic sensors stand out for their minimally invasive installation and high sensitivity [183, 187]. Complementary diagnostic techniques such as arc detection and gas monitoring (e.g., sulfur hexafluoride alternatives) further enrich the condition assessment toolkit [188–190].

From a data analysis perspective, AI and ML offer powerful methods to interpret sensor data and predict faults before they lead to failures. Supervised learning models, which rely on labeled datasets, have been successfully applied to fault classification and regression tasks across various energy domains, including photovoltaic system monitoring, boiler emissions control, and High Voltage (HV) fault diagnosis [191–193]. Conversely, unsupervised learning excels in anomaly detection and pattern discovery without requiring labeled data, making it suitable for scenarios where fault examples are scarce or unknown [194–197].

Despite these advances, the specific application of unsupervised learning techniques for real-time fault detection and localization in MV switchgear is still limited. Practical challenges such as the scarcity of labeled fault data and computational constraints of embedded devices hinder broader adoption. Moreover, predictive maintenance approaches based solely on temperature monitoring of MV switchgear have not been extensively explored. Additionally, unsupervised learning methods have not been adequately applied for fault detection and diagnosis in this specific context.

This research aims to address the existing gap by exploring two complementary data-driven approaches for predictive maintenance of MV switchgear based on temperature data acquired from the same DMS:

- A **supervised learning approach** employing an ANN to map temperature patterns to known mechanical states using labeled data, enabling accurate fault classification;

- An **unsupervised learning approach** leveraging clustering techniques to detect anomalies and abnormal operating conditions without prior labeling, thus supporting early fault detection under limited data availability.

By evaluating these methodologies, the study seeks to enhance fault diagnosis accuracy and robustness, while accommodating real-world constraints such as limited annotations and computational resources, ultimately contributing to improved reliability and safety in MV switchgear operation.

4.3.1 Operational Challenges and Degradation Mechanisms in Medium Voltage Switchgears

MV switchgears face a variety of operational challenges that can significantly affect their reliability, safety, and service life [198]. These challenges stem from a complex interplay of electrical, mechanical, environmental, and human factors. Among these, electrical issues are particularly critical: conditions such as overloading, overheating, and short circuits may lead to failures ranging from minor faults to catastrophic events. Insulation failures, in particular, play a key role in switchgear reliability. The insulation system prevents electrical faults but can degrade over time due to moisture ingress, temperature variations, and contamination, potentially resulting in short circuits, arc flash hazards, or even fires [199].

Overheating can occur due to overcurrent conditions when currents exceed rated limits, or from unstable Power Quality (PQ), including voltage fluctuations and harmonic distortions. PD, localized dielectric breakdowns within the insulation under HV stress, often initiate from defects such as voids, cracks, or contaminants. Though initially small and difficult to detect, PDs progressively deteriorate insulation, ultimately risking flashover and equipment failure [200].

Mechanical failures also contribute substantially to switchgear issues and can account for a large fraction of total failures. Poor or loose electrical connections increase resistance, leading to overheating, arcing, and damage to key components like circuit breakers and busbars [187]. Furthermore, mechanical parts including moving mechanisms degrade over time due to normal wear, mechanical stresses, and environmental influences. Environmental factors include vibrations from adjacent machinery or seismic activity that

loosen connections and induce stress, as well as accumulation of dust, dirt, and airborne pollutants (e.g., industrial chemicals, salt in coastal areas, exhaust fumes), which degrade insulation, cause tracking, and corrode metal parts, thereby reducing performance and increasing fault risks. Additionally, high humidity and moisture cause condensation, while extreme temperatures – both high and low – affect switchgear performance. Elevated temperatures accelerate insulation aging and may cause overheating, whereas low temperatures can impair mechanical and control systems [201].

Temperature is a dominant degradation factor, significantly shortening insulation life and accelerating the aging process of both electrical and mechanical components. This inherent vulnerability to thermal stress makes temperature monitoring and management essential in switchgear design and maintenance. Maintaining temperatures within prescribed limits is not just a regulatory compliance issue but a fundamental requirement for ensuring long-term operational efficiency and safety. Standards from bodies like the IEEE Standards Association [202] emphasize temperature rise limits and the importance of thermal monitoring, while industry guidelines such as those from the International Electrical Testing Association (NETA) [203] provide protocols for thermal image interpretation and maintenance decision-making based on temperature differentials. Continuous monitoring systems that track temperature alongside other relevant parameters directly contribute to extending switchgear service life, reducing costly repairs and unexpected outages.

Addressing these multifaceted challenges necessitates a comprehensive approach that combines regular preventive maintenance based on condition monitoring, environmental control measures, robust equipment design, and thorough training of operating and maintenance personnel [204].

4.3.2 Experimental Setup

The predictive maintenance approach proposed here is based on temperature measurements carried out inside MV switchgears, which comply with the International Standard IEC 62271-200 [205]. The experimental setup used in this case study consists of an MV switchgear and a data acquisition system for temperature measurement.

The MV switchgear employed is an Imesa MINIVER/C, a metal-enclosed switchboard designed in accordance with IEC 62271-200 [205] for voltages above 1 kV and up to 52 kV.

It is composed of four main compartments, shown in Figure 4.9: the busbar compartment (a), the line compartment (b), the circuit breaker compartment (c), and the LV compartment (d). The busbar compartment houses two copper busbars connected by a joint, linked to the fixed contact of the upper interrupting device enclosure. The busbar compartment is supplied by a three-phase system that branches out through all the MV compartments. In the line compartment, the current is stepped down by a factor of 750 through a WATTSUD IWR10K CT operating at 50 Hz. This compartment also accommodates the cables connecting the power lines, located at the rear, to the fixed contact of the lower interrupting device enclosure. Finally, the switchgear includes an interrupting device and LV equipment, located in the circuit breaker compartment (c) and the LV compartment (d), respectively. The switchgear is supplied with a current of 630 A at 50 Hz.

Under proper operating conditions, the joint between the busbars and the connections to the power cables are tightened using a torque wrench with M10 bolts tightened to 45 N m [206]. The fixed contacts with the interrupting device enclosures – both in the busbar and line compartments – as well as the upper and lower tulips of the circuit breaker, are fastened using M16 bolts tightened to 165 N m [206]. The CT is secured with M12 bolts tightened to 60 N m [206].

Separate temperature acquisition sessions were conducted, simulating different operating conditions of the MV switchgear. The first session corresponded to the correct operating conditions as described above (CT tightened to 60 N m). In the subsequent sessions, the fastening torque on the CT was progressively reduced, while all other fastenings remained at the correct operating values. Specifically, in the subsequent acquisitions, the CT was tightened to 40 N m, 20 N m, and 4 N m. Details on the acquisition times are given in the following subsections, depending on the used approach (supervised or unsupervised).

Temperature measurements were performed using seven SR30KX K-type thermocouples placed inside the switchgear. These thermocouples were positioned to monitor differential temperatures at seven locations indicated in Figure 4.9: (1) joint between the busbars; (2) fixed contact with the upper interrupting device enclosure in the busbar compartment; (3) top contact of the CT; (4) fixed contact with the lower interrupting

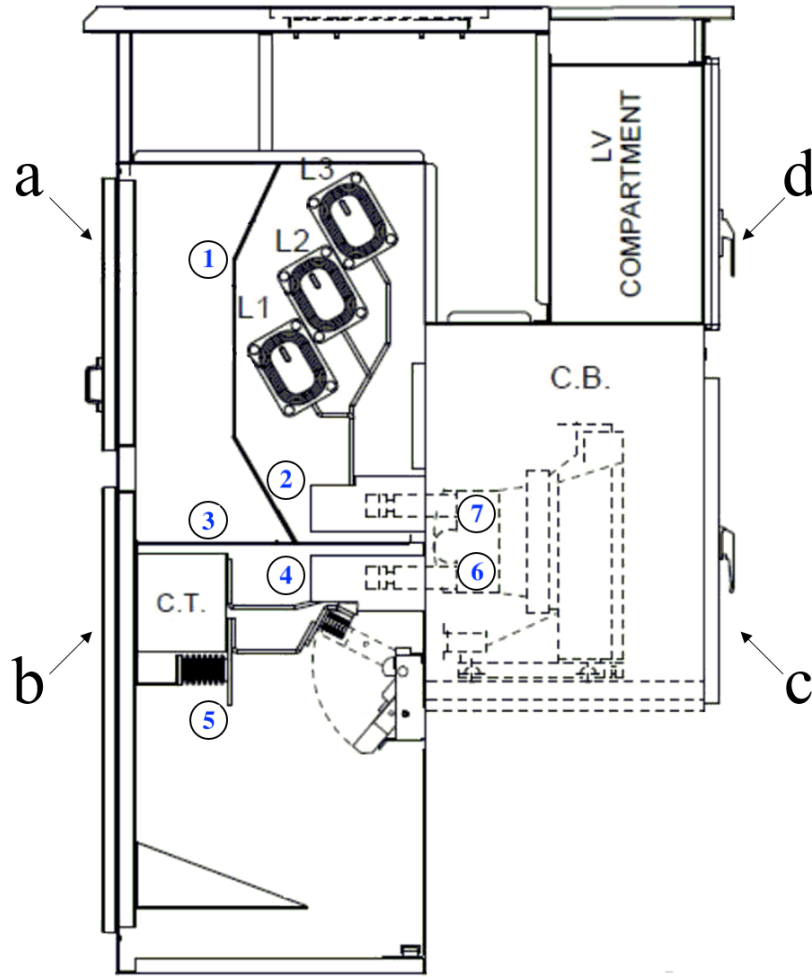


FIGURE 4.9: Compartments of the MV switchgear and locations of the thermocouples (blue numbers).

device enclosure in the line compartment; (5) connection to the power cables; (6) lower tulip; and (7) upper tulip of the circuit breaker.

The thermocouples were connected to a data acquisition system composed of a National Instruments cDAQ-9178 CompactDAQ chassis [207] equipped with a National Instruments NI-9214 C Series Temperature Input Module [208]. The NI-9214 module [208] converts the temperature-dependent voltage signals from the thermocouples into temperature readings. Since it supports up to sixteen thermocouple inputs, a single module sufficed to simultaneously acquire data from all seven thermocouples, occupying only one of the eight slots in the cDAQ chassis.

According to IEC 62271-200 [205], MV switchgears operate under normal conditions if the ambient air temperature lies within the range $[-5.0^{\circ}\text{C}, 40.0^{\circ}\text{C}]$. To monitor this,

three K-type thermocouples were placed outside the switchgear during all sessions, recording average air temperatures of 18.9°C, 19.1°C, and 18.2°C, respectively. Hence, the ambient conditions during the tests remained within the acceptable limits. It is crucial to control the environmental conditions of MV switchgears to ensure the external temperature remains within the specified range. Any deviations beyond these limits, as defined by IEC 62271-200 [205], could affect the expected temperature behavior and undermine predictive maintenance accuracy.

TABLE 4.13: Minimum and maximum temperature readings acquired during the three sessions with CT fastening of 60 N m, 20 N m and 4 N m.

Thermocouple	60 Nm		20 Nm		4 Nm	
	Min (°C)	Max (°C)	Min (°C)	Max (°C)	Min (°C)	Max (°C)
1	42.1	44.2	43.2	45.3	42.7	45.1
2	51.1	53.5	52.3	54.6	52.7	55.2
3	52.0	55.0	66.8	69.3	81.9	86.9
4	53.7	56.2	59.0	61.3	63.3	66.4
5	43.9	46.4	54.4	56.8	62.4	65.2
6	53.4	55.8	57.5	59.8	60.8	63.7
7	51.3	53.7	52.9	55.3	52.9	55.5

4.3.3 Supervised Learning Approach

In this study, a supervised learning strategy is employed to support predictive maintenance in MV switchgears. The approach is based on the temperature data collected from multiple locations within the switchgear, under different mechanical fastening conditions of the CT. Specifically, three torque levels (i.e., 60 N m, 20 N m, and 4 N m) represent three distinct health states: correct, intermediate, and critical. To capture the temperature changes resulting from the different fastening torques, each acquisition session started 5 h and 35 min after the switchgear was powered on. Then, 50 temperature readings were recorded by each thermocouple at intervals of 2 min and 30 s from the start of the session. Table 4.13 reports the minimum and maximum temperature values measured during the acquisition sessions, corresponding to fastening torques of 60 N m, 20 N m, and 4 N m on the CT.

An ANN is developed to classify the conditions based on the temperature patterns. The input to the ANN consists of temperature measurements recorded by seven thermocouples positioned at key points inside the MV switchgear. The main novelty of the proposed method lies in leveraging the correlation between CT fastening and thermal behavior to

assess the health condition of the equipment. While this study focuses on temperature as the input feature, the proposed methodology can be extended to include other relevant physical quantities for switchgear diagnostics.

4.3.3.1 Artificial Intelligence Model

The adopted AI model is a feed-forward ANN (as described in Section 2.3.3) composed of three layers: input, hidden, and output. The hidden layer contains 20 neurons and uses the ReLU (Equation 2.2) activation function, while the output layer employs the Softmax function (Equation 2.6) to perform multi-class classification.

ReLU is selected due to its computational efficiency and ability to accelerate model convergence during training. It activates only for positive inputs, thus improving the network's learning capacity [209]. Softmax, on the other hand, is commonly used in classification tasks, as it transforms the network output into a normalized probability distribution, with each element interpretable as the likelihood of a specific class.

The model is implemented in Python using the `Keras` Application Programming Interface (API) [210] of the `TensorFlow` library [211]. For training, the Adam optimizer is employed to adjust network weights efficiently.

Figure 4.10 illustrates the block diagram of the ANN architecture, including input features, activation functions, and output labels. The input features consist of the temperature measurement and the corresponding thermocouple identifier, resulting in 8 total input features. To improve learning performance, proper data preprocessing is applied. Temperature values are normalized within the range $[0, 1]$, while the thermocouple number (ranging from 1 to 7) is encoded using one-hot encoding. This encoding converts the categorical thermocouple ID into a binary vector; for example, thermocouple 1 becomes $[1, 0, 0, 0, 0, 0, 0]$, and thermocouple 7 becomes $[0, 0, 0, 0, 0, 0, 1]$. The output layer classifies input samples into three classes, each representing a different torque level applied to the CT: 60 N m (label 0 – correct), 20 N m (label 1 – intermediate), and 4 N m (label 2 – critical). This classification reflects the health status of the MV switchgear as inferred from thermal behavior under different fastening conditions.

Concerning the evaluation of the ANN model performance, Accuracy (Equation 2.9), Precision (Equation 2.10), Recall (Equation 2.11), and F_1 score (Equation 2.12). Since

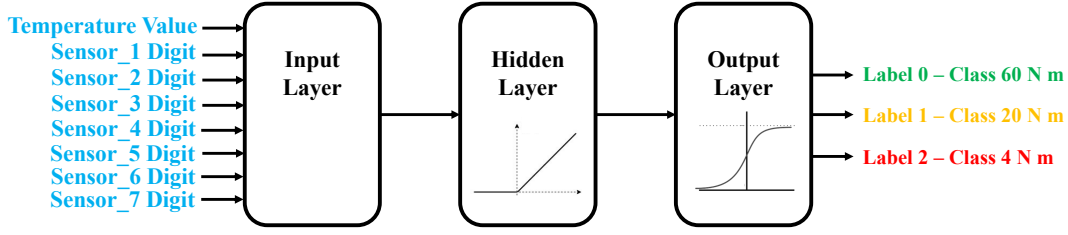


FIGURE 4.10: Block diagram of the proposed ANN model with activation functions.

the proposed case study is formulated as a multi-class classification task, Precision, Recall, and F_1 score have been calculated using the macro-averaging approach (Equations 2.13, 2.14, 2.15, respectively).

4.3.3.2 Data Description

To train and test the ANN, representative datasets for each fastening condition are required. Given that ANNs generally demand large datasets to achieve robust performance, synthetic data are generated based on the real temperature measurements acquired in the field.

Specifically, for each thermocouple and each torque level (i.e., 60 N m, 20 N m, and 4 N m), a synthetic dataset of 1000 temperature values is generated using a uniform distribution. This choice aligns with the GUM [22], which advises uniform distributions when the true distribution is unknown. The distribution is centered on the mean of the 50 measured temperature values per thermocouple per fastening, as described in Section 4.3.2. The range of the uniform distribution is defined according to the expected sensor uncertainty, in compliance with the International Standard IEC 60751 [212].

Three uncertainty scenarios are considered to generate synthetic data: $\pm 0.5^\circ\text{C}$, $\pm 1.0^\circ\text{C}$, and $\pm 1.5^\circ\text{C}$. The lower and upper bounds are selected to represent best-case and worst-case scenarios, respectively, while the intermediate value of $\pm 1.0^\circ\text{C}$ corresponds to the actual uncertainty of the employed data acquisition system.

As a result, each fastening condition yields a dataset of 7000 synthetic temperature values (1000 per each of the 7 thermocouples), leading to a total dataset size of 21000 temperature values across all three torque levels. It is worth noting that, in this first analysis based on the synthetic dataset, only the generated temperature data are used, while the original 50 experimental measurements are intentionally excluded. This choice is

motivated by the need to ensure a sufficiently large and statistically homogeneous dataset for the training and testing of the ANN model. Given the data-driven nature of ANN-based diagnostic approaches, the limited number of experimental measurements would not be adequate to properly capture the variability of the thermal behavior under different fastening conditions, nor to support a reliable separation between training and testing sets. The synthetic data, generated to be consistent with the statistical characteristics of the experimental measurements, allow for a controlled assessment of the ANN's capability to learn temperature-based diagnostic patterns in the switchgear, before moving to validation on real experimental data in subsequent analyses.

4.3.3.3 Results on Synthetic Data

Firstly, the model's performance was assessed through a numerical evaluation using the synthetic datasets generated with two uncertainty levels, $\pm 0.5^\circ\text{C}$ and $\pm 1.5^\circ\text{C}$, representing the best-case and worst-case scenarios, respectively. This preliminary analysis aims to provide an initial understanding of the algorithm's robustness under varying measurement uncertainties.

The two synthetic datasets, each consisting of 21000 temperature values, have been split using 80 % of the data for training, and the remaining 20 % for testing.

Table 4.14 reports the Accuracy values and the execution times for both training and testing phases. Detailed results of the ANN performance, including Precision, Recall, and F_1 metrics for each output class, are provided in Tables 4.15 and 4.16 for the datasets with $\pm 0.5^\circ\text{C}$ and $\pm 1.5^\circ\text{C}$ uncertainty, respectively.

For the dataset with $\pm 0.5^\circ\text{C}$ uncertainty, the highest metric values are achieved for the 60 N m class, while the 4 N m class shows the lowest performance. Overall, the average metric values exceed 83 %. Increasing the measurement uncertainty to $\pm 1.5^\circ\text{C}$ causes a performance decline, with average metric values above 77 %. Notably, the 20 N m class exhibits the poorest classification results in this scenario.

TABLE 4.14: Overall performance and execution times for synthetic datasets.

Uncertainty value	Accuracy (%)	Training time (s)	Testing time (s)
$\pm 0.5^\circ\text{C}$	83.38	21.92	0.35
$\pm 1.5^\circ\text{C}$	77.48	20.66	0.33

TABLE 4.15: Metrics for synthetic dataset with uncertainty $\pm 0.5^\circ\text{C}$.

	Precision (%)	Recall (%)	F ₁ (%)
60 N m Class	100.00	91.79	95.72
20 N m Class	66.73	100.00	80.05
4 N m Class	100.00	58.36	73.70
macro-averages	88.91	83.38	83.16

TABLE 4.16: Metrics for synthetic dataset with uncertainty $\pm 1.5^\circ\text{C}$.

	Precision (%)	Recall (%)	F ₁ (%)
60 N m Class	86.89	81.93	84.34
20 N m Class	67.72	76.57	71.87
4 N m Class	79.80	73.93	76.75
macro-averages	78.14	77.48	77.65

Furthermore, the proposed ANN model is compared with other ML algorithms commonly used in predictive maintenance: SVM, Random Forest, Decision Tree, and KNN. All models are trained using 80 % of the datasets with uncertainties $\pm 0.5^\circ\text{C}$ and $\pm 1.5^\circ\text{C}$, and tested on the remaining 20 %. As shown in Figure 4.11, the ANN outperforms these alternative methods across both uncertainty scenarios.

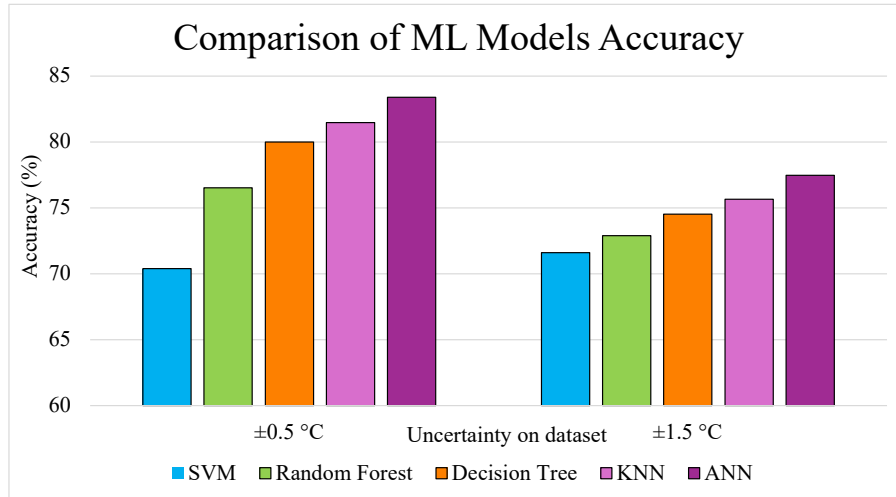


FIGURE 4.11: Comparison of ANN accuracy with state-of-the-art methods.

4.3.3.4 Results on Experimental Data

The experimental evaluation is intended to assess the effectiveness of the ANN in practical scenarios. In this case, the synthetic dataset with uncertainty $\pm 1.0^\circ\text{C}$ is employed for the training phase, while the testing is performed using the actual temperature measurements

described in Section 4.3.2. The experimental testing dataset consists of 350 temperature values per fastening, for a total of 1050 measurements.

Table 4.17 presents the achieved Accuracy value and the execution times for both training and testing phases. The detailed performance in terms of Precision, Recall, and F_1 metrics is shown in Table 4.18. All the metrics report values exceeding 78 %, confirming the ANN model’s robustness when applied to real data.

TABLE 4.17: Overall performance and execution times for experimental measurements.

Uncertainty value	Accuracy (%)	Training time (s)	Testing time (s)
$\pm 1.0^\circ\text{C}$	78.38	22.18	0.28

TABLE 4.18: Metrics for experimental measurements.

	Precision (%)	Recall (%)	F_1 (%)
60 N m Class	87.54	84.29	85.88
20 N m Class	68.93	75.43	72.03
4 N m Class	80.00	75.43	77.65
macro-averages	78.82	78.38	78.52

4.3.3.5 Discussion

The objective of this research is to contribute to the improvement of fault detection and predictive maintenance strategies for MV switchgears. To this end, a method based on the integration of an ANN with a distributed temperature measurement system has been proposed. The novelty of the approach lies in exploiting the correlation between temperature variations – caused by different levels of component fastening – and the actual health condition of the switchgear. This correlation enables the detection of potentially critical conditions before a fault occurs, thus supporting condition-based maintenance actions.

The ANN has been trained and tested on both synthetically generated datasets – characterized by controlled uncertainty levels – and real measurement data acquired from the field. The results obtained in both numerical and experimental evaluations demonstrate the effectiveness of the model, confirming its suitability for this type of application.

In general, the findings of this research support the adoption of predictive maintenance strategies based on the combination of distributed sensing and ML algorithms. This integrated approach can improve the reliability and safety of MV switchgears while reducing the risk of unexpected failures and associated downtime.

4.3.4 Unsupervised Learning Approach

In this approach, an unsupervised learning strategy is adopted to assess the health condition of MV switchgears using temperature data collected under different CT fastening conditions. In this case study, three torque levels (i.e., 60 N m, 40 N m, and 4 N m) represent distinct mechanical states of the switchgear (i.e., correct, intermediate, and critical). For each torque level, the switchgear was powered on for 3 h and 30 min before starting the acquisition, allowing a stable thermal profile to develop. Then, 100 temperature samples were collected from each thermocouple, recorded every 2 min and 30 s from the start of the session. The initial and final temperature values observed during the acquisition sessions are reported in Table 4.19.

Unlike the supervised approach, this study focuses on a different sensor deployment strategy. Specifically, only thermocouples 3 and 5 (see Figure 4.9) are considered. This targeted configuration allows for a more localized and representative monitoring of the switchgear's thermal behavior.

The main novelty of this approach lies in the application of clustering techniques to unlabeled temperature measurements. In practical industrial environments, labeled datasets are often unavailable or too costly to obtain. Therefore, unsupervised methods are investigated to assess whether meaningful fault categories can be inferred directly from the temperature data, without prior knowledge of the system state. In this case, labeled data are only used to evaluate the classification performance of clustering algorithms. To address this task, K-means and Spectral Clustering methods are employed. While K-means serves as the primary clustering method, Spectral Clustering is chosen as a comparison method.

Finally, the unsupervised inference pipeline has been implemented on a low-power microcontroller embedded in the DMS. This deployment enables fully real-time, on-board

processing without the need for data transmission to remote servers. The proposed approach represents a cost-effective and scalable solution for predictive maintenance, highlighting the feasibility of using unsupervised learning for autonomous condition monitoring in MV switchgears.

TABLE 4.19: Initial and final temperature readings acquired during the three sessions with CT fastening of 60 N m, 40 N m and 4 N m.

Thermocouple	60 Nm		40 Nm		4 Nm	
	Min (°C)	Max (°C)	Min (°C)	Max (°C)	Min (°C)	Max (°C)
3	47.5	55.0	59.3	66.6	74.5	86.9
5	40.3	46.4	47.0	53.2	57.8	65.2

4.3.4.1 Clustering Methods

Clustering techniques are fundamental in modern power systems to support automated decision-making, especially when managing large volumes of unlabeled data. In this study, the K-means algorithm was employed due to its computational efficiency, scalability, and simplicity. Specifically, the K-means++ initialization (described in Section 2.4.1) was adopted to improve convergence and reduce sensitivity to initial centroid placement.

However, K-means presents some limitations. It assumes clusters to be spherical, equally sized, and well-separated, which may not reflect the structure of real-world datasets. Furthermore, it relies on Euclidean distance, making it sensitive to outliers and ineffective when clusters exhibit irregular shapes or varying densities. To address these issues and provide a robustness check, Spectral Clustering (described in Section 2.4.2) was also applied. Unlike K-means, Spectral Clustering does not depend on centroid-based partitioning and is capable of identifying non-convex and arbitrarily shaped clusters. This makes it a valuable benchmark to validate and support the outcomes obtained with the K-means algorithm. To evaluate the clustering performance, ARI (Equation 2.28), V_1 (Equation 2.29), and Silhouette (Equation 2.30) metrics have been selected.

Both algorithms were implemented using the `Scikit-learn` library [64], specifically through the `Cluster` module.

4.3.4.2 Results on Synthetic Data

The experimental dataset comprises temperature readings collected from thermocouples 3 and 5 under three different fastening conditions, labeled as classes 0, 1, and 2, corresponding to tightening torques of 60 N m, 40 N m, and 4 N m, respectively. For each class and thermocouple, 100 measurements were recorded, totaling 600 samples.

From this base, synthetic datasets were generated by computing the mean of the 100 original measurements for each thermocouple and class. New samples were then drawn from a uniform distribution centered on the mean, with bounds of $\pm 1.0\%$ and $\pm 1.5\%$ of the mean value. As in the supervised approach, the uncertainty levels were selected based on typical accuracy figures in temperature DMSs.

To assess the impact of uncertainty and data availability, five datasets were created for each uncertainty level and mean value, with $N = \{50, 100, 500, 1000, 2000\}$ samples per class. Each dataset, therefore, includes measurements from both thermocouples and all three fastening conditions. Note that the listed samples are not the total amount of instances in the dataset. For example, the dataset with 50 samples comprises 300 instances, 150 for thermocouple 1 and 150 for thermocouple 2 (i.e., 50 for each fastening class).

The adoption of a distinct dataset in this unsupervised learning study is intentional and serves a twofold purpose. First, it enables the explicit assessment of measurement uncertainty and data variability, which are critical aspects when clustering unlabeled temperature data and when generating additional samples representative of realistic operating conditions. Second, this dataset supports an exploration of the minimum number of temperature samples required to obtain stable and meaningful clusters, as the unsupervised approach inherently involves higher-dimensional feature spaces and lacks prior class information. This analysis therefore complements the supervised studies by evaluating data sufficiency and robustness in a fully data-driven, label-free diagnostic framework.

Initially, results were evaluated on datasets with $\pm 1.0\%$ uncertainty. Clustering algorithms were applied separately to data from thermocouples 3 and 5. Tables 4.20 and 4.21 report the clustering performance for both sensors in relation with the number of generated samples per class. For thermocouple 3, the K-means algorithm consistently achieves

perfect scores in both ARI and V_1 , demonstrating its ability to accurately capture well-separated and compact clusters. This is further confirmed by the high Silhouette values. Conversely, Spectral Clustering yields lower ARI and Silhouette scores, indicating greater difficulty in distinguishing clear cluster boundaries, particularly as the dataset size increases. A similar behavior is observed for thermocouple 5: K-means maintains flawless ARI and V_1 scores across all sample sizes, along with stable Silhouette values around 0.96. Spectral Clustering, on the other hand, shows noticeably poorer results, with ARI scores ranging between 0.50 and 0.80 and generally low or even negative Silhouette values – especially in larger datasets – suggesting weak cluster separability. Overall, K-means proves to be a robust and stable solution across different data volumes, while Spectral Clustering demonstrates considerable variability in its performance.

TABLE 4.20: Clustering metrics of the algorithms on datasets with ± 1.0 % uncertainty (thermocouple 3) in relation with the number of generated samples per class.

Algorithm	Metric	50	100	500	1000	2000
K-means	ARI (-)	1.00	1.00	1.00	1.00	1.00
	V_1 (-)	1.00	1.00	1.00	1.00	1.00
	Silhouette (-)	0.97	0.97	0.97	0.97	0.97
Spectral Clustering	ARI(-)	0.65	0.59	0.63	0.59	0.64
	V_1 (-)	0.77	0.74	0.76	0.74	0.76
	Silhouette (-)	-0.02	0.02	-0.02	0.02	-0.02

TABLE 4.21: Clustering metrics of the algorithms on datasets with ± 1.0 % uncertainty (thermocouple 5) in relation with the number of generated samples per class.

Algorithm	Metric	50	100	500	1000	2000
K-means	ARI (-)	1.00	1.00	1.00	1.00	1.00
	V_1 (-)	1.00	1.00	1.00	1.00	1.00
	Silhouette (-)	0.96	0.96	0.96	0.96	0.96
Spectral Clustering	ARI (-)	0.77	0.80	0.50	0.60	0.58
	V_1 (-)	0.82	0.84	0.71	0.74	0.74
	Silhouette (-)	-0.11	0.27	-0.02	-0.10	-0.01

Secondly, the same models were applied to synthetic datasets generated with ± 1.5 % uncertainty. Figures 4.12, 4.13, and 4.14 respectively represent ARI, V_1 , and Silhouette scores for both thermocouple data and the two levels of applied uncertainty. The results confirm the robustness of the K-means algorithm, which maintains high metric values across all dataset sizes and uncertainty levels. In particular, for both thermocouples 3 and 5, K-means achieves $ARI = 1.00$, $V_1 = 1.00$, and $Silhouette = 0.94 - 0.95$, independently of sample size or uncertainty level. Spectral Clustering, on the other hand, only

shows comparable performance for thermocouple 3 when using at least 500 samples and uncertainty $\pm 1.5\%$, achieving $\text{ARI} = 1.00$, $V_1 = 1.00$, and $\text{Silhouette} = 0.94$. However, with fewer samples or lower uncertainty $\pm 1.0\%$, its performance worsens significantly. For thermocouple 5 in particular, performance drops sharply in smaller datasets, with $\text{ARI} = 0.59$ (50 samples) and 0.46 (100 samples); $V_1 = 0.72$ and 0.65 ; and $\text{Silhouette} = 0.37$ and 0.70 , respectively.

The observed improvement of Spectral Clustering performance at higher uncertainty levels can be attributed to the regularizing effect introduced by the added variability. In low-uncertainty or small-sample regimes, the temperature data tend to form compact and partially overlapping distributions, which limits the ability of graph-based methods to construct a well-separated affinity matrix. The introduction of moderate uncertainty $\pm 1.5\%$ increases the spread of the data, enhancing inter-cluster separation in the spectral domain and leading to more informative eigenstructures of the similarity graph. As a result, cluster boundaries become more distinguishable, particularly when a sufficient number of samples is available. This behavior is consistent with the known sensitivity of Spectral Clustering to data density, noise distribution, and affinity matrix construction, and does not indicate a degradation of measurement quality but rather an improved discriminative structure for clustering.

Overall, this behavior indicates that spectral clustering is sensitive to data sparsity. Nonetheless, K-means remains the most reliable algorithm, consistently delivering perfect clustering results across different uncertainty levels and dataset sizes, due to its local distance-based approach and efficient centroid optimization.

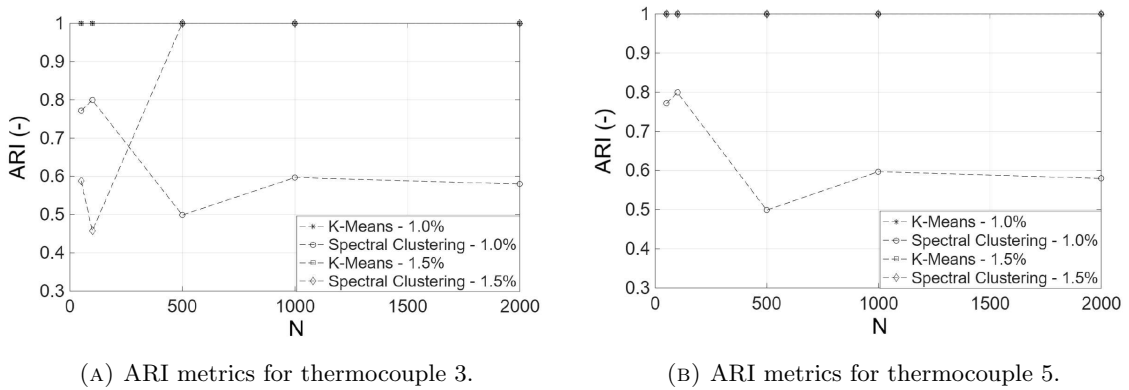


FIGURE 4.12: ARI metrics of the algorithms on datasets with $\pm 1.0\%$ and $\pm 1.5\%$ uncertainty for (A) thermocouple 3 and (B) thermocouple 5 data.

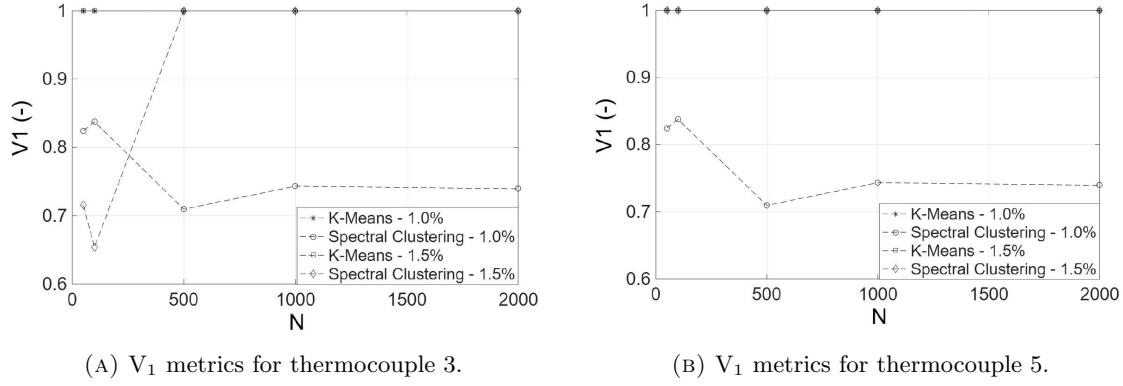


FIGURE 4.13: V_1 metrics of the algorithms on datasets with $\pm 1.0\%$ and $\pm 1.5\%$ uncertainty for (A) thermocouple 3 and (B) thermocouple 5 data.

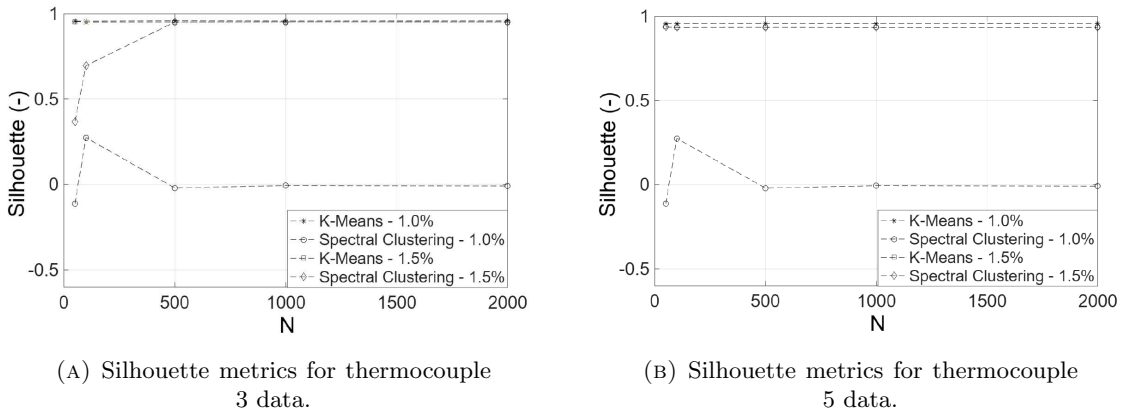


FIGURE 4.14: Silhouette metrics of the algorithms on datasets with $\pm 1.0\%$ and $\pm 1.5\%$ uncertainty for (A) thermocouple 3 and (B) thermocouple 5 data.

4.3.4.3 Results on Experimental Data

To assess the effectiveness of the proposed approach under real conditions, the algorithms were tested on the experimental measurements collected from the DMS. Table 4.22 reports the values of the clustering metrics obtained by applying the K-means algorithm to the real data. Only K-means results are presented, as this algorithm consistently outperformed spectral clustering during preliminary evaluations.

The results show that K-means achieves high performance on both datasets. For thermocouple 3, it reaches perfect scores of 1.00 in both ARI and V_1 , indicating excellent clustering accuracy. For thermocouple 5, the values are slightly lower but still strong, with $ARI = 0.98$ and $V_1 = 0.97$.

These results are supported by the Silhouette metric, which reflects the degree of separation between clusters. In particular, the Silhouette value is 0.77 for thermocouple 3

and 0.70 for thermocouple 5, suggesting a moderate but acceptable level of separability between the three physical regimes. This confirms the ability of K-means to effectively identify underlying structures, even in the presence of borderline cases that may otherwise lead to false alarms or undetected faults in real-world scenarios. Overall, K-means demonstrates reliable clustering performance on real measurements, with good cluster coherence and separation.

TABLE 4.22: Clustering metrics of K-means tested on the experimental measurements.

Data	Metric	Value
Thermocouple 3	ARI (-)	1.00
	V_1 (-)	1.00
	Silhouette (-)	0.77
Thermocouple 5	ARI (-)	0.98
	V_1 (-)	0.97
	Silhouette (-)	0.70

4.3.4.4 Results on Arduino Board

The deployment of ML algorithms on microcontrollers (i.e., TinyML [213]), aims to enable on-device intelligence in resource-constrained environments. Due to limited memory and processing power, such devices require lightweight and efficient models to ensure real-time performance. In this context, K-means proves to be a suitable choice thanks to its low memory footprint and computational simplicity compared to other unsupervised algorithms [214]. The proposed solution was implemented on an Arduino Nano 33 BLE Sense Rev2 board [215], which integrates sensing capabilities with a low-power microcontroller, making it ideal for edge-level applications. An overview of the on-board implementation is shown in Figure 4.15.

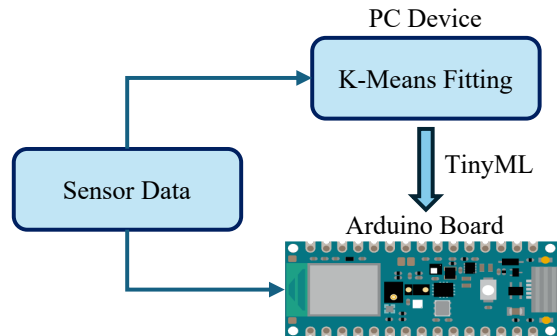


FIGURE 4.15: Workflow of the K-means implementation on microcontroller.

The implementation begins by fitting the data offline on a computer using the K-means algorithm, which determines the optimal cluster centroids. These centroids are then exported and uploaded to the microcontroller. On the Arduino board, classification of new temperature readings is performed by computing the Euclidean distance from each measurement to the predefined centroids. Each reading is then assigned to the closest cluster. This approach ensures that the clustering results obtained on the microcontroller are consistent with those produced by the PC-based implementation. The correctness of the method was verified by comparing the output of the Arduino with that of the Python-based algorithm, using identical input datasets. The two implementations yielded identical clustering results, confirming the robustness and portability of the approach.

To quantitatively assess the computational performance, both fitting and prediction times were measured on the PC, while only the prediction time was recorded on the Arduino. The results are summarized in Table 4.23. As expected, the PC performs both fitting and prediction very efficiently, with fitting times ranging from 18.36 to 20.07 ms and prediction times under 1 ms. On the other hand, prediction on the Arduino requires more time – between 3.36 and 4.28 ms – due to its limited processing capabilities. Despite this, the prediction remains fast enough for real-time applications.

TABLE 4.23: Execution times for fitting and prediction.

Thermocouple	Time (ms)		
	Fitting (PC)	Prediction (PC)	Prediction (Arduino)
3	20.07	0.82	4.28
5	18.36	0.77	3.36

In addition to performance, memory usage was evaluated. Despite the Arduino's constrained resources, the implementation uses only a small portion of the available memory. The algorithm requires storing a limited number of centroids and performing basic distance calculations, making the code lightweight and efficient. Flash and RAM consumption remain well below the board's limits, leaving space for potential future expansions such as multi-dimensional input, additional clustering features, or sensor integration. These results confirm that the proposed K-means-based approach is highly suitable for embedded applications on resource-constrained platforms like the Arduino Nano 33 BLE Sense Rev2.

4.3.4.5 Discussion

This research demonstrates the feasibility of deploying unsupervised ML algorithms on low-cost microcontrollers for real-time monitoring of MV switchgears. The proposed system, based on distributed temperature sensing and local processing, enables anomaly detection without requiring labeled data or cloud-based computation. Validation on both synthetic and real datasets confirms robust clustering performance and reliable classification, even under noisy conditions and limited hardware capabilities. The implementation on an Arduino board highlights the method's efficiency, with minimal memory usage and acceptable prediction times, making it suitable for embedded applications in real-world scenarios. This decentralized approach not only reduces communication overhead but also supports scalable, low-latency monitoring architectures, promoting energy efficiency and predictive maintenance in smart grid environments.

4.3.5 Final Remarks

This section presented and analyzed two complementary data-driven strategies for the predictive maintenance of MV switchgear, both relying on temperature measurements collected from a DMS.

On one side, a supervised learning approach based on an ANN was developed to accurately classify known fault conditions using labeled data. On the other, an unsupervised learning approach employing K-means enabled anomaly detection in the absence of annotated datasets, offering a flexible alternative when prior labeling is not feasible.

By integrating these methods, the research addresses both condition discrimination and early anomaly detection challenges, focusing on the identification of different CT fastening levels corresponding to normal, potentially degraded, and critical operating states, while explicitly accounting for real-world constraints such as low computational power, communication limitations, and data scarcity. The successful implementation of the unsupervised method on a microcontroller further demonstrates the practical viability of lightweight, decentralized monitoring solutions.

Together, these approaches offer a robust and scalable framework for predictive maintenance, supporting improved reliability, safety, and operational efficiency in MV switchgear

systems. This dual-path strategy lays the groundwork for future developments in intelligent condition monitoring within smart grid infrastructures.

Chapter 5

Artificial Intelligence and Metrology

This chapter explores the bidirectional integration between metrology and AI: on the one hand, metrological principles are embedded within AI-based systems to enhance their reliability and robustness; on the other hand, AI techniques are employed to improve measurement estimation and modeling. In the first part, the influence of measurement uncertainty on the learning process of AI algorithms is investigated [216]. Moreover, an approach to evaluate the performance of ML is presented from a metrological perspective [217]. Finally, a novel metrological framework for DL models is introduced, comprising uncertainty-driven DA strategies to improve performance in load forecasting [218]. In the second part, AI is applied to metrological estimation, where an ANN is used to infer accuracy-related parameters of ITs, such as temperature-induced deviations, enabling local real-time estimation on embedded devices [219]. Additionally, a PINN is proposed for accurate modeling of RCs under varying thermal conditions [220]. Finally, an RC model is implemented on a microcontroller, representing a foundational step towards the development of DTs for metrological instruments [221].

5.1 The Influence of Measurement Uncertainty in Artificial Intelligence Models

While AI and ML algorithms offer powerful capabilities for solving complex problems, their decision-making processes – especially those based on DL architectures – often lack transparency. This limited interpretability places a strong emphasis on the quality and reliability of the data used for training and evaluation. Inaccuracies or uncertainties in the input data can severely impair model performance, potentially leading to erroneous predictions and suboptimal decisions. To ensure the robustness and scientific validity of AI-based systems, it is essential to train and test models using datasets that not only capture the available information but also explicitly account for measurement uncertainty.

To investigate the role of uncertainty in AI systems, this section considers two case studies within the energy sector, each focusing on a different type of measurement and model uncertainty.

The first case study concerns the fault prediction of MV CJs. Two distinct sources of uncertainty affect this task: the uncertainty in the acquired temperature data, and the intrinsic unpredictability of fault events. Since real-world datasets often provide incomplete or indirect indications of impending failures, this case study critically evaluates how such uncertainties influence the predictive maintenance of CJs. To this end, five ML algorithms are applied to the test case and compared using conventional performance metrics.

The second case study focuses on the estimation of the State of Charge (SoC) of lithium iron phosphate (LiFePO_4) batteries. While the accurate determination of SoC is essential for ensuring the reliable operation of battery-powered systems, SoC predictions derived from AI models often lack a quantifiable measure of confidence. This study advances the field by introducing a rigorous uncertainty evaluation framework, elevating AI-based SoC estimations to metrological-grade assessments. During the model testing phase, a quasi-Monte Carlo method is applied to analyze the influence of uncertainty and to derive confidence intervals associated with SoC estimates. These results are then benchmarked against laboratory-grade reference instrumentation. The core contribution lies not only in improving SoC estimation accuracy, but in establishing a principled approach for evaluating and communicating the reliability of AI-generated outputs.

Together, these two studies highlight the necessity of integrating metrological principles into AI workflows. By quantifying and managing uncertainty at both the data and model levels, AI solutions become more trustworthy, interpretable, and suitable for critical applications in power systems. In this chapter, uncertainty is not limited to the dispersion of aggregate performance metrics, but is explicitly interpreted as uncertainty associated with individual model predictions. To this end, stochastic inference and input perturbation strategies are employed to generate predictive distributions, enabling the estimation of confidence intervals and coverage at the single-prediction level.

5.1.1 The Cable Joint Test Case

The case study is the one presented in Section 4.1.5, where a synthetic dataset was considered to train and test several ML models applied to the classification of CJ health status. This research aims to investigate in greater depth how the characteristics of input data influence the performance of ML algorithms. Previously, measurement uncertainty was introduced in a simplified manner: during the final stage of synthetic generation, a 10 % error probability was simulated.

To enhance this analysis, two additional aspects are addressed. First, the measurement uncertainty of the physical quantities used as input features is explicitly evaluated. Indeed, the field instrumentation used for data acquisition introduces its own uncertainty, which must be quantified and incorporated into the overall analysis [114, 222]. Second, attention is given to the uncertainty associated with the definition of failure thresholds for these quantities. In practical scenarios, such thresholds are typically determined based on historical records or expert knowledge, both of which may be imprecise and inherently uncertain. It is important to note that, in real-world applications, SOs have access to measurements acquired prior to fault events, which indirectly inform their decision thresholds. Conversely, in the simulated environment considered in this research, fixed thresholds must be defined a priori to evaluate algorithm performance.

The ML models selected for this study are among the most widely adopted supervised algorithms for binary classification. The implemented methods include SVM, KNN, ANN, Decision Tree, and Random Forest (as described in Section 2.3.1). To evaluate the models, Accuracy (Equation 2.9), Precision (Equation 2.10), and Recall (Equation 2.11) metrics have been selected.

5.1.1.1 Uncertainty of Input Data

As is well known, in-field measurements are inherently affected by uncertainty. This section investigates the impact of such uncertainty on the performance of ML algorithms. To this end, each algorithm was tested using two synthetic datasets: one representing ideal measurement conditions, and the other incorporating realistic uncertainty in selected input features.

Specifically, temperature and current values were altered to reflect typical measurement uncertainties found in MV distribution networks: $\pm 1.0^\circ\text{C}$ for temperature and $\pm 0.5\%$ for current. Tan Delta and CJ age were excluded from this perturbation: CJ age is measured in years with negligible uncertainty, and Tan Delta, ranging from 10^{-5} to 10^{-1} , is assumed to be measured with sufficient accuracy (uncertainty below 10^{-5}). A uniform probability distribution was assumed for all uncertainty contributions. A second test doubled these uncertainty levels to $\pm 2.0^\circ\text{C}$ and $\pm 1.0\%$.

Each algorithm was trained using 1000 samples and tested on 100000 samples, both synthetically generated with uniformly distributed inputs (as suggested by the GUM [22]). The methodology involves simulating measurement uncertainty by replacing each original value with a new one drawn from a uniform distribution centered on the true value and bounded by the specified relative uncertainty.

The performance metrics (Accuracy, Precision, Recall) for both tests are reported in Figure 5.1. Results show that input uncertainty has minimal impact on model performance. Decision Tree and Random Forest consistently achieve the highest scores across all metrics, while SVM fails to identify any positive cases, resulting in zero precision and recall. KNN and ANN display moderate performance. The stability of metrics despite increased input uncertainty suggests that the models are robust to the levels of measurement noise typical of low-cost sensors, supporting their deployment in cost-sensitive environments such as MV distribution networks.

5.1.1.2 Uncertainty of Fault Thresholds

In a synthetic dataset, class labels (faulty or non-faulty) are typically assigned in a deterministic manner. However, in real-world scenarios, the thresholds defining fault conditions for CJs are not strictly defined. Due to the complexity and variability of CJ aging and

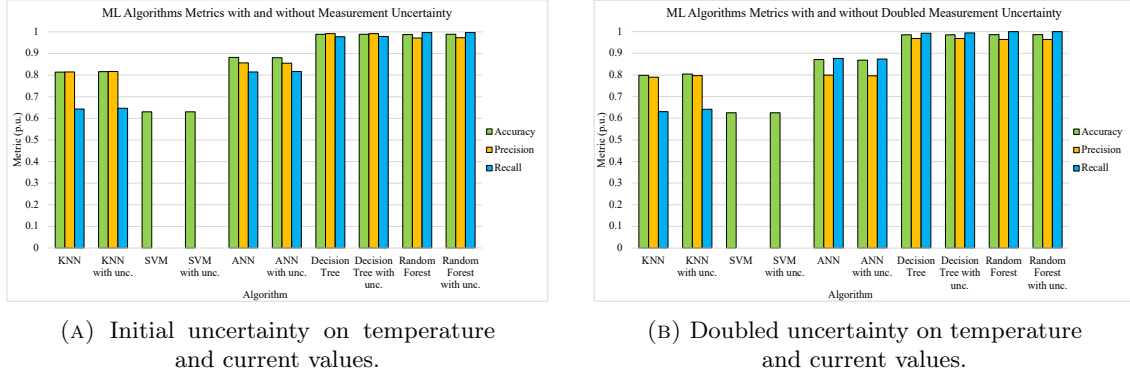


FIGURE 5.1: Performance metrics of ML algorithms applied to dataset corrupted by (A) initial uncertainty and (B) doubled uncertainty.

failure mechanisms, multiple interacting factors may lead to a fault, and identical input conditions may not always yield the same outcome.

To assess how threshold variability affects algorithm performance, a dedicated test was conducted. The same five algorithms were trained and tested on two datasets: one with fixed thresholds and one with thresholds varying according to a defined uncertainty model.

Three distinct thresholds were assigned to each monitored quantity, as listed in Table 4.3. These define fault conditions depending on whether one, two, or all three quantities simultaneously exceed their respective limits. The thresholds are generally derived from operator experience or historical data, although pre-fault measurements are rarely available and often non-generalizable across different CJs. Therefore, such thresholds inherently carry uncertainty.

To simulate this, thresholds were randomly varied following a uniform distribution centered on the nominal values in Table 4.3, with ranges specified in Table 5.1. Two distinct datasets were thus constructed: one labeled using fixed thresholds and one labeled using variable thresholds. The same models were trained and tested on both datasets, allowing a direct evaluation of the impact of threshold uncertainty on classification performance. Each input instance in the variable-threshold dataset was associated with a set of thresholds drawn from this distribution, simulating real-world ambiguity in defining fault conditions.

The resulting performance metrics are shown in Figure 5.2. While the SVM remains ineffective, the other models show varying degrees of degradation. ANN and KNN exhibit

TABLE 5.1: Uncertainty values applied to fault thresholds.

Quantity	Uncertainty 1	Uncertainty 2
Current (A)	± 10	± 20
Temperature ($^{\circ}\text{C}$)	± 2	± 4
Tangent delta (-)	$\pm 10\%$	$\pm 20\%$
Age (years)	± 2	± 4

moderate performance drops, while DT and RF, which previously performed best, are more sensitive to threshold uncertainty.

Figure 5.3 presents the same analysis with doubled uncertainty ranges. As expected, performance further deteriorates, especially for DT and RF. These results highlight the critical role that fault threshold uncertainty plays in ML-based fault prediction and suggest that algorithms must be rigorously validated under realistic uncertainty conditions.

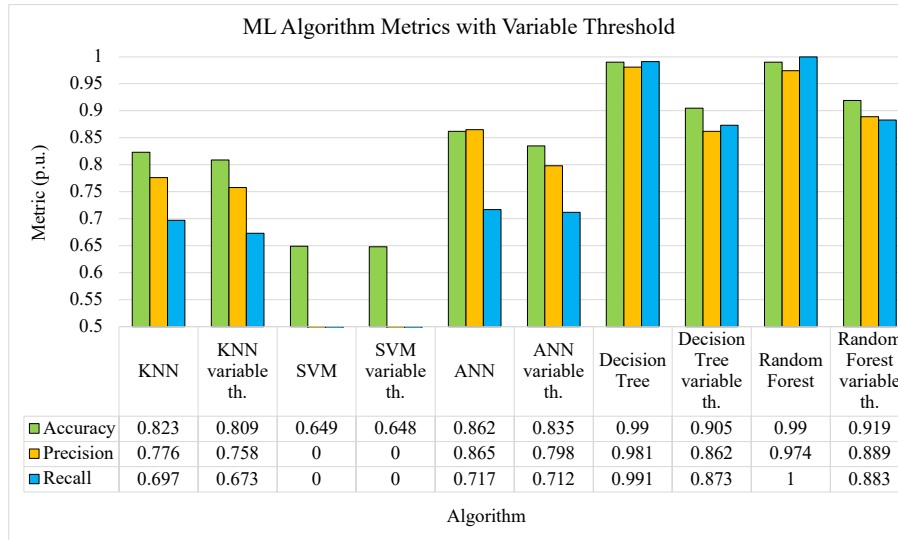


FIGURE 5.2: Performance metrics of ML algorithms applied to dataset with and without uncertain threshold.

5.1.1.3 Discussion

This research investigates specific aspects of ML algorithms application, focusing in particular on how input data uncertainty and limited domain knowledge impact algorithm performance. The analysis reveals that uncertainty in the input data has a negligible effect on the performance of the tested algorithms. Moreover, such uncertainty can be mitigated through careful design of the measurement system and repeated in-field data acquisition. In contrast, uncertainty related to fault thresholds – arising from limited

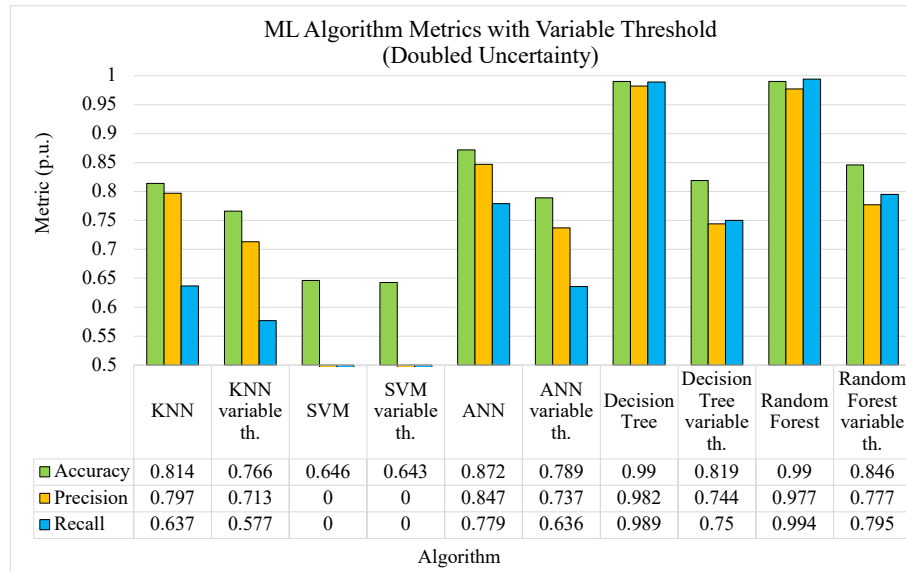


FIGURE 5.3: Performance metrics of ML algorithms applied to dataset with and without uncertain threshold (doubled uncertainty).

experience or insufficient historical data – has a significantly stronger influence on model behavior.

This highlights the critical importance of thoroughly assessing ML models within the context of their intended application. Different algorithms may respond differently to uncertainties, particularly when key domain-specific knowledge is missing or difficult to formalize. Consequently, while AI offers promising tools for fault detection, its effectiveness strongly depends on the quality and completeness of the available information.

5.1.2 State of Charge Estimation in Lithium-Ion Batteries

Battery Energy Storage Systems (BESSs) are playing an increasingly central role in modern power systems, supporting renewable integration, grid stability, and the electrification of transport [223]. In particular, BESSs enable the storage of surplus energy generated during periods of high renewable production and its release during demand peaks, thereby ensuring supply reliability [224]. Additionally, they contribute to peak shaving and grid resilience during outages.

Lithium-ion batteries currently represent the most promising BESS technology due to their high energy density, long life cycle, low self-discharge, and fast charging capability [225, 226]. However, their performance and health are affected by several factors, including

operating temperature, aging, and the number of charge/discharge cycles. Consequently, effective battery management and parameter monitoring are crucial.

A key indicator of battery status is the SoC, which indicates the remaining charge relative to full capacity. Since SoC cannot be directly measured, various estimation methods have been developed. Traditional approaches include open circuit voltage, Coulomb Counting, and model-based methods such as Kalman Filters. While widely used, each method has trade-offs depending on the required accuracy, computational complexity, and battery characteristics.

More recently, AI techniques have gained traction for SoC estimation due to their ability to model complex, nonlinear battery behavior. For example, a NARX neural network with LSTM units has been employed to mitigate issues like gradient explosion [227]. NARX-based regression has also been adopted to enhance estimation accuracy [228]. A forecasting model combining NARX with optimized Levenberg-Marquardt and Bayesian-Regularization training has been proposed [229]. ANNs have further been used to estimate SoC in hybrid electric vehicles, demonstrating low error across different battery sizes and validating robustness to input current offsets through comparison with Coulomb Counting [44].

These studies confirm the growing reliability of data-driven methods for SoC estimation, particularly when high accuracy and adaptability are required.

5.1.2.1 Data Collection and Reference Setup

The case study focuses on battery SoC estimation, selected as a representative application for uncertainty analysis in ML-based predictions. Reference measurements were obtained in a highly controlled laboratory environment using high-accuracy instrumentation. These reference values are considered the benchmark for evaluating the accuracy of the ML estimations, which are treated as the measured values subject to uncertainty analysis. Comparing the confidence intervals of estimated and reference SoC values allows the evaluation of when and under which conditions the ML predictions can be considered reliable.

The experimental setup includes a TDK Lambda GEN 10-240 power supply [230] (rated 10 V, 240 A) and two Agilent N3306A DC electronic load modules [231] (rated 60 V,

240 A). The setup is controlled via a custom LabVIEW 2018 application interfaced with a National Instruments PCIe-6353 data acquisition board [232]. Voltage measurements were recorded directly at the battery poles with an accuracy of $981 \mu\text{V}$.

Temperature monitoring was performed using three copper disc K-type thermocouples placed on the battery poles and shell, along with a standard thermocouple for ambient temperature. These signals were acquired using NI 9211 modules [233], although only cell and ambient temperatures were used as model inputs.

For the uncertainty evaluation, a single lithium iron phosphate (LFP) cell was discharged at a constant current of 2C, in accordance with manufacturer specifications. Key parameters of the tested battery are reported in Table 5.2. Measurements were acquired every 0.1 s, resulting in a total of 18432 samples.

The reference SoC values were obtained through coulomb counting under constant-current discharge conditions. Specifically, the instantaneous extracted charge was computed by integrating the measured current over time, and the remaining charge was expressed as a percentage of the nominal battery capacity. The initial condition was set to 100 % SoC, corresponding to the fully charged state, while the discharge was limited to a depth of discharge of 80 %. Voltage measurements were used to identify the initial fully charged condition and to ensure consistency with the selected discharge limits, rather than for direct SoC estimation.

Although a small-capacity battery was tested in a highly controlled environment, the insights gained from this experimental campaign are scalable and applicable to larger battery systems [44].

TABLE 5.2: Specifications of the tested cell.

Feature	Lithium iron phosphate battery
Shape (-)	Cylindrical
Weight (g)	86
Nominal capacity (Ah)	3.2
Set-point voltage (V)	3.65
Minimum discharge voltage (V)	2.5
Cut-off current (A)	0.032
Temperature limits ($^{\circ}\text{C}$)	-20 to 45

5.1.2.2 Deep Learning Model

To model the battery SoC estimation, a NARX model was adopted (as described in Section 2.3.3.2), structured in an open-loop configuration. This choice was motivated by the NARX model's suitability for sequence-based forecasting tasks and its effectiveness in modeling dynamic systems such as batteries [227, 228].

The NARX network was implemented in Python using the `Keras` module [210] within the `TensorFlow` framework [211]. After defining the network architecture, the training and testing phases were performed using the dataset generated from the laboratory discharge test. Input features included voltage, quantity of charge, current, and temperatures (cell and ambient), while the target output was the SoC percentage.

This architecture enabled the learning of the nonlinear relationship between the measurable battery parameters and the corresponding SoC, forming the basis for the subsequent uncertainty evaluation.

5.1.2.3 Uncertainty Analysis

As previously introduced, the uncertainty analysis has been carried out through a metrological comparison between the reference and the estimated SoC values. The reference SoC is obtained at each time step from the voltage measurements, as described in Section 5.1.2.1. Its confidence interval is derived from the accuracy specifications of the voltage sensor, following a type B uncertainty evaluation.

For the estimated SoC, a quasi-Monte Carlo approach has been adopted to propagate input measurement uncertainty through the ML models. Specifically, multiple synthetic datasets were generated by perturbing the original voltage and current measurements, which constitute the model input features. Each synthetic dataset represents a plausible realization of the measured signals within the specified sensor accuracy bounds.

At each time step, synthetic input samples were generated by drawing random values from a uniform distribution centered on the original measurement. The perturbation bounds were set to $\pm 0.1\%$ and $\pm 1\%$ of the nominal voltage and current values, reflecting low-accuracy and high-accuracy sensing scenarios typically encountered in Battery Management Systems (BMSs) with different cost and performance constraints.

For each uncertainty level, sets of 10, 100, and 1000 synthetic datasets were generated. Each dataset was independently processed by the trained ML model, yielding a corresponding set of SoC estimates. The resulting ensemble of outputs was then used to approximate the output probability distribution of the estimated SoC at each time step.

The lowest number of realizations (10) represents a realistic upper bound for the number of repeated measurements or re-evaluations that can be performed in real time by a resource-constrained BMS, where computational power and data acquisition rates are limited. Conversely, the highest number of realizations (1000) is not intended for real-time deployment but is used as a reference to assess the convergence and stability of the estimated SoC uncertainty bounds. Intermediate values (100) provide insight into the trade-off between computational effort and uncertainty estimation accuracy.

For each combination of input uncertainty level and number of realizations, the mean estimated SoC and the corresponding 95 % confidence interval were computed. These intervals were then compared with the confidence interval of the reference SoC values to evaluate the consistency and reliability of the ML-based estimations under different uncertainty and computational constraints. Particularly, for each time step, the estimated SoC is considered metrologically consistent if the 95 % confidence interval of the estimated value overlaps with the corresponding 95 % confidence interval of the reference SoC. The percentage of overlapping intervals is then computed as the ratio between the number of time steps exhibiting overlap and the total number of time steps in the discharge profile, for each combination of input uncertainty level and number of synthetic realizations.

Importantly, the NARX network was first trained and validated on the original dataset, composed of input feature measurements (voltage, current, charge, cell and ambient temperatures) and reference SoC values. The dataset was split 80 % for training and 20 % for testing. Figure 5.4 shows the resulting SoC estimation compared to the reference, and Table 5.3 reports the achieved MAE (Equation 2.16) and MSE (Equation 2.17) metrics.

TABLE 5.3: Metric values of the NARX model.

Metric (-)	Value (-)
MAE	9.6×10^{-6}
MSE	3.8×10^{-10}

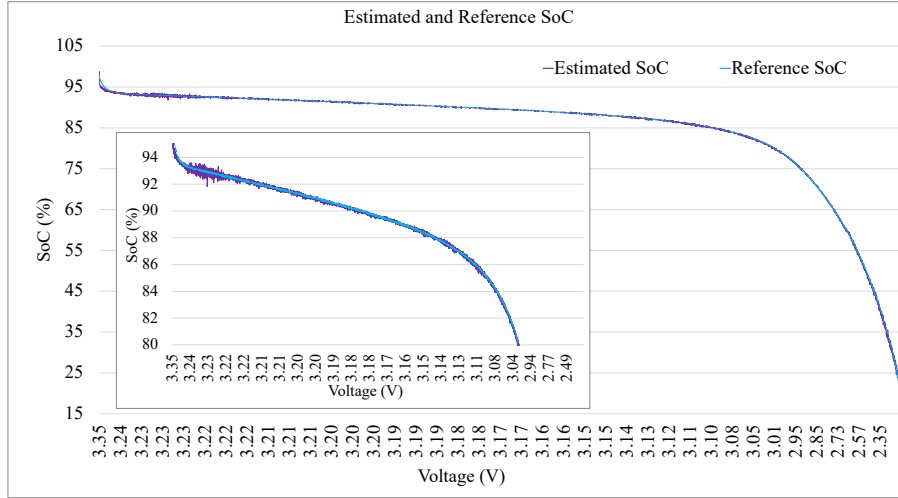


FIGURE 5.4: Actual (blue) vs. predicted (violet) SoC values.

It is worth noting that these metrics reflect the model's fitting performance and not its measurement uncertainty. The uncertainty analysis was instead conducted via the aforementioned quasi-Monte Carlo simulation.

Figure 5.5 presents a comparison between the 95 %-confidence intervals of the estimated and reference SoC values for three selected time instants under 1 % input uncertainty and 100 iterations. In each pair, the circle indicates the interval around the reference value, and the square around the estimated SoC. All displayed intervals overlap, indicating good metrological consistency. Notably, the reference intervals are not visible in the plot due to their significantly smaller width compared to those of the estimated SoC.

The overall percentage of overlapping confidence intervals represents the fraction of time instants for which the 95 % confidence interval of the estimated SoC overlaps with the corresponding 95 % confidence interval of the reference SoC. This percentage is computed over all time steps of the discharge profile and for each combination of input uncertainty level and number of synthetic dataset realizations. The resulting values, averaged over the entire dataset, are summarized in Table 5.4. As expected, the overlap increases with both greater input uncertainty and a higher number of dataset generations. The results suggest that even highly accurate AI estimators may fail to reliably match reference values when based on extremely precise input data, a condition rarely met in practical deployments.

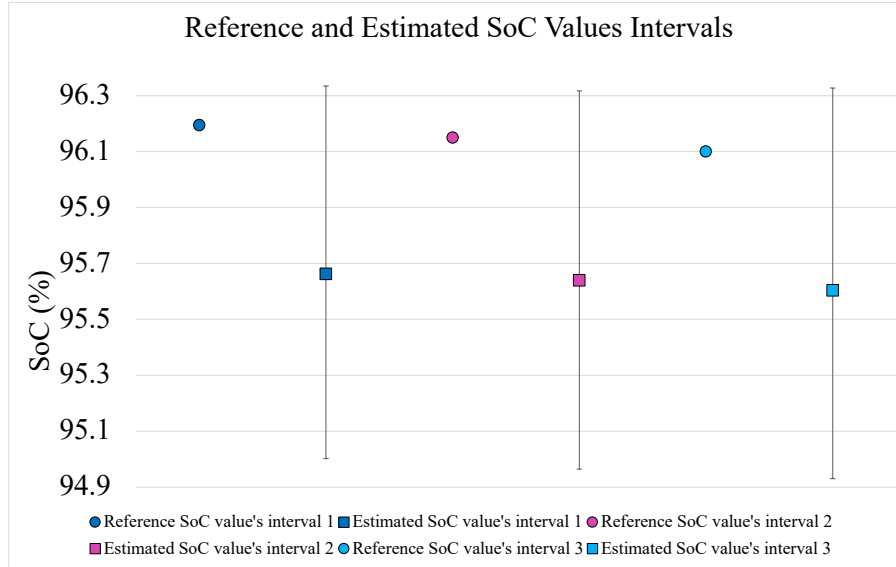


FIGURE 5.5: Comparison between estimated and reference SoC values intervals.

TABLE 5.4: Percentage of overlapped SoC intervals.

Iterations (-)	0.1 % Uncertainty	1.0 % Uncertainty
10	39.31 %	95.74 %
100	58.25 %	99.66 %
1000	58.49 %	99.69 %

5.1.2.4 Discussion

This research presented a novel procedure for handling measurement uncertainty in AI-based estimation systems, emphasising the importance of evaluating the performance of AI and ML models from a metrological perspective. The objective was to validate the application of the NARX model in estimating the SoC of batteries, treating the ML algorithm as a measurement instrument. To this end, a quasi-Monte Carlo method was applied and proved to be a valid strategy for the metrological validation of AI-based SoC estimators. The results showed that AI-generated estimates could be regarded as measurements and subjected to uncertainty evaluation using conventional metrological techniques. Furthermore, it was found crucial to define the target uncertainty a priori – taking into account the performance of the employed sensors – in order to realistically assess the algorithm’s accuracy in specific applications. The findings provided valuable metrological insights that could guide the selection of measurement instruments integrated within BMS.

5.2 Uncertainty-Driven Data Augmentation

DL models have become central in forecasting and measurement tasks due to their high accuracy and flexibility. However, they typically produce point predictions without explicitly quantifying uncertainty, even when processing noisy or error-prone sensor data. This limitation reduces their trustworthiness in applications where measurement uncertainty is critical [23].

The theoretical foundation for treating uncertainty as a measurable quantity has long been established [29, 234], yet its integration into AI systems trained on real data remains limited. This gap is especially significant in energy systems, where measurements are acquired through standardized instrumentation characterized by well-defined error margins. Despite this, uncertainty is rarely modeled or propagated during the DL training and inference processes, regardless of the model architecture.

Current literature reveals several critical gaps: (i) DL workflows rarely incorporate input measurement uncertainty during training or inference, despite its known presence in real-world data; (ii) existing UQ methods are often post-hoc, model-specific, and require access to model internals, lacking scalable, black-box-compatible frameworks applicable across different DL architectures; (iii) the impact of measurement uncertainty on model generalization and robustness remains underexplored; (iv) most UQ techniques focus on epistemic uncertainty via model stochasticity but neglect aleatory uncertainty inherent to sensor measurements, which is crucial for metrology-based applications; (v) there is no unified approach to interpreting AI outputs as uncertainty-bearing estimates, even when sensor accuracy and calibration data are available.

While several DL-based methods (e.g., Bayesian neural networks, Monte Carlo dropout, and deep ensembles [27, 28]) attempt to model epistemic uncertainty, they seldom integrate the aleatory and instrumental uncertainties intrinsic to measurement devices like CTs, Voltage Transformers (VTs), and temperature sensors, whose accuracy is well characterized [235, 236].

This research addresses these gaps by proposing a data-centric, metrology-aware framework for uncertainty-aware DL. Instead of modifying model architectures, we explicitly model and inject known measurement uncertainty into input data during training via

DA and testing, through quasi-Monte Carlo simulation. This enables construction of robust models that generalize across measurement points and substations, while allowing evaluation of prediction variability with traceable statistical descriptors.

Treating DL models as estimators acting on uncertain inputs, the framework interprets outputs as distributions of plausible values rather than fixed predictions. Validated using an LSTM model on the Ausgrid dataset [237], the method improves generalization and supports uncertainty assessment under realistic noise conditions. Its architecture-independent and non-intrusive design aligns with metrological principles, bridging the gap between advanced AI tools and the rigorous requirements of measurement science.

5.2.1 Data Description

This study utilizes the publicly available Ausgrid dataset [237], which provides detailed information on power consumption patterns across Australia. The dataset covers a 12-month period with measurements recorded at 15-minute intervals and includes data from over 180 zone substations. Key variables include year, substation name, date, and power demand. Ausgrid's metering infrastructure employs CTs and VTs to ensure accurate energy distribution measurements, complying with strict standards. These measurement systems undergo regular testing, including primary injection tests, to guarantee data integrity and conformity with the National Electricity Rules (NER) [238].

For the purposes of this research, data from the Beacon Hill and Auburn substations in Sydney were selected, spanning the years 2017 to 2022. This subset provides a consistent and largely complete record of power consumption trends. Typical weekly fluctuations from the dataset are illustrated in Figure 5.6. Although generally reliable, the data may contain occasional errors or gaps caused by metering inaccuracies, switching events, or measurement inconsistencies. Substations serving large individual customers were excluded to preserve confidentiality.

To prepare the data for the LSTM model, missing power values were initially imputed using interpolation. Subsequently, the timestamp feature was transformed using circular encoding. Finally, power readings were normalized to the range $[0,1]$, standardizing the data to improve model convergence and performance.

Data were then fed to an LSTM (described in Section 2.3.3.1), implemented in PyTorch library [83] to forecast future power demand. Training is performed using the Adam optimizer (learning rate 0.001, $\text{eps} = 1\text{e-}7$) and MSE (Equation 2.17) as loss function, with a batch size of 12. The model uses a fixed input window of one week and predicts the different horizon lengths (e.g., 1-week or 1-day ahead forecasting).

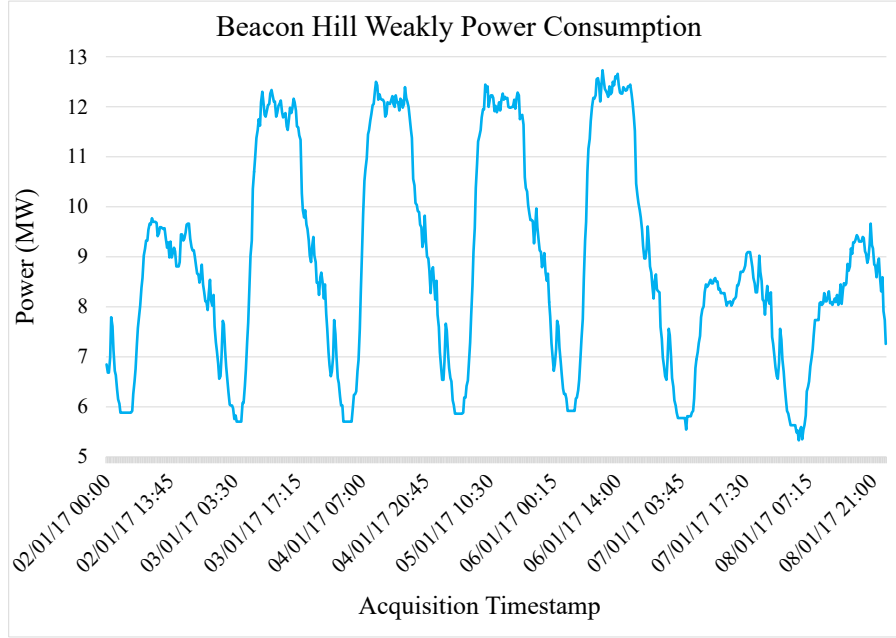


FIGURE 5.6: Weekly power consumption pattern from the Beacon Hill substation dataset.

5.2.2 Measurement Uncertainty Quantification

The uncertainty in active power, $P = VI \cos \phi$, derived from the voltage V , current I , and the cosine of the phase difference ϕ , is typically calculated using the law of uncertainty propagation. However, since typical smart meter specifications do not always provide information about the uncertainty on ϕ , a simplifying assumption is made. Instead of calculating the uncertainty for active power, it is computed for apparent power $S = VI$. Note that this assumption does not offer any methodological advantage and serves purely as a simplification.

The general formula for the propagation of uncertainty in a function of multiple variables, as outlined in the GUM [22], is given by:

$$u(f) = \sqrt{\left(\frac{\partial f}{\partial x_1}u(x_1)\right)^2 + \left(\frac{\partial f}{\partial x_2}u(x_2)\right)^2 + \cdots + \left(\frac{\partial f}{\partial x_n}u(x_n)\right)^2}, \quad (5.1)$$

where $u(f)$ is the uncertainty of the function f , $x_1, x_2, \dots, x_n$ are the variables, $u(x_i)$ represents the uncertainty of the i -th variable, and $\frac{\partial f}{\partial x_i}$ is the partial derivative of f with respect to x_i .

For the apparent power, the uncertainty is derived from the uncertainties of voltage and current:

$$u(S) = \sqrt{(I u(V))^2 + (V u(I))^2}, \quad (5.2)$$

where $u(V)$ and $u(I)$ are the uncertainties of voltage and current, respectively. Since the measurements are taken using CTs and VTs of class 0.5 [238], indicating a maximum uncertainty of 0.5 %, the relative uncertainty of the apparent power is:

$$\frac{u(S)}{S} = \frac{\sqrt{(I \cdot 0.5 \% \cdot V)^2 + (V \cdot 0.5 \% \cdot I)^2}}{VI} = 0.707 \%. \quad (5.3)$$

This value is then considered in the proposed metrological framework, described in 5.2.3.

5.2.3 Metrological Framework

The proposed metrological framework is designed to evaluate the role of measurement uncertainty in both the training and testing phases of the forecasting model. The analysis is structured into three main parts, followed by two complementary experiments aimed at assessing model robustness under different sources of variability.

First, uncertainty is used to generate augmented data within a single substation to enrich the training set. Although the augmented dataset is larger in absolute size than the original one, the same order of magnitude is maintained to ensure a fair comparison of model performance with and without augmentation. Second, the model's ability to generalize is evaluated across a different substation using both original and augmented

data. Finally, uncertainty is introduced directly into the test data to simulate realistic operating conditions and assess the model's robustness to measurement variability.

Two uncertainty levels, 0.5 % and 1.0 %, are considered based on the relative uncertainty of apparent power computed in Section 5.2.2. The 0.5 % level represents an optimistic scenario of minimal uncertainty, while 1.0 % accounts for a more realistic condition, including the typically neglected uncertainty in $\cos \phi$. If precise uncertainty values were available for installed field devices, the framework could be tailored to those specific values, thereby improving the fidelity of the evaluation.

The metrological framework comprises the following components:

1. **Uncertainty-Based Training in a Single Substation:** To mitigate the limitations of small datasets, a DA strategy incorporating measurement uncertainty is introduced. Starting from the original data, new samples are generated by perturbing each measurement using a uniform distribution centered on the original value and bounded by $\pm$ the relative uncertainty (0.5 % or 1.0 %). This approach aligns with GUM [22] recommendations, which advocate a uniform distribution when the exact error distribution is unknown. Data from the Beacon Hill substation are used for this purpose. By explicitly modeling measurement uncertainty, the DA approach enables the model to better capture inherent variability in the data, enhancing generalization capability.

A two-fold statistical validation was performed to assess the soundness of the proposed DA method:

- A **Kolmogorov–Smirnov (K–S) test** was applied to the original and augmented datasets to verify whether the augmented data distribution differs statistically from the original one, confirming that the introduced perturbations enrich the training data in a meaningful way [239].
- A **two-sample t-test** was conducted on the model prediction errors under different experimental configurations (with and without DA) to determine whether performance improvements are statistically significant [240].

This dual statistical analysis ensures that the proposed uncertainty-based DA is validated both at the data and performance levels, reinforcing the robustness of the framework.

2. **Cross-Substation Generalization Using Augmented Data:** To evaluate the model's robustness under different operational conditions, a cross-substation validation is carried out. The model is trained on data from the Beacon Hill substation and tested on the Auburn substation, which was not seen during training. Two training configurations are compared: one using only the original dataset and another using the uncertainty-augmented dataset. This setup allows assessment of whether uncertainty-aware augmentation enhances the model's generalization capability across substations with distinct load patterns and operating characteristics.
3. **Uncertainty-Aware Testing on Different Substations:** In this configuration, the model is trained solely on the original, unperturbed dataset. Measurement uncertainty is introduced exclusively during testing by replacing each test value with samples drawn from a uniform distribution centered on the original measurement and bounded by $\pm$ the selected uncertainty level (0.5 % or 1.0 %). This simulates realistic measurement noise typically encountered in field conditions. The procedure is applied to both substations, and each test instance is perturbed 1000 times to obtain statistically reliable estimates of the model's mean and standard deviation of the performance metrics, providing a robust evaluation of its resilience to uncertainty in real-world deployments.
4. **Model Ensemble for Training Variability Assessment:** To isolate the intrinsic variability due to stochastic model training (e.g., random weight initialization, batch sampling), an ensemble of identical LSTM architectures is trained independently on the original dataset. The ensemble results are aggregated by computing the mean and standard deviation of the evaluation metrics. This experiment quantifies model-driven variability and provides a baseline for comparison with the variability induced by uncertainty-based DA.
5. **Dropout-Based Uncertainty Estimation during Testing:** As a complementary analysis, dropout layers are kept active during inference (i.e., Monte Carlo Dropout) to estimate epistemic uncertainty and assess prediction stability under stochastic model behavior. The model, trained on the original dataset without DA, is evaluated 1000 times with dropout enabled at test time. The resulting mean and standard deviation of the metrics quantify the model's internal uncertainty, complementing the measurement-uncertainty experiments and providing a comprehensive view of the forecasting framework's robustness.

The performance of all configurations is evaluated using Mean Absolute Error (MAE) (Equation 2.16), Root Mean Square Error (RMSE) (Equation 2.18), and Mean Absolute Percentage Error (MAPE) (Equation 2.19). These metrics, along with the statistical tests and ensemble-based analyses, provide a rigorous and comprehensive assessment of the proposed metrology-aware DL framework.

Algorithm 1 provides a schematic pseudocode representation of the proposed framework, highlighting the main phases and their interconnections.

Algorithm 1: Metrology-aware training and testing framework.

Input: Original dataset D , sensor specifications
Output: Trained LSTM model, uncertainty-aware predictions

Step 1: Compute Measurement Uncertainty
 Extract relative uncertainty u_{sensor} (%) from specifications

Step 2: Select Uncertainty Levels
 Use $u = \{0.5\%, 1.0\%\}$ or a single known value

Step A: Uncertainty-Based Training
foreach training point $x_i \in D_{\text{train}}$ **do**
 | Generate $x'_i \sim \text{Uniform}[x_i - ux_i, x_i + ux_i]$
end
 Train LSTM on augmented dataset D'
 Evaluate on reserved test year (MAE, RMSE, MAPE)
 Apply K-S test and t-test

Step B: Cross-Substation Evaluation
 Train on Beacon Hill, test on Auburn
 Compare performance with/without DA (MAE, RMSE, MAPE)

Step C: Uncertainty-Aware Testing
foreach test sample $x_j \in D_{\text{test}}$ **do**
 for $k = 1$ to N **do**
 | Sample $x'_j \sim \text{Uniform}[x_j - ux_j, x_j + ux_j]$
 | Evaluate model on x'_j
 end
end
 Compute mean and Std. Dev. of (MAE, RMSE, MAPE)

Step D: Model Ensemble for Training Variability
for $s = 1$ to M **do**
 | Initialize LSTM with seed s , train on D , evaluate on test data
end
 Compute mean and Std. Dev. of metrics

Step E: Dropout-Based Uncertainty Estimation
 Enable dropout at inference
for $k = 1$ to N_{MC} **do**
 | Evaluate trained LSTM on test data with dropout
end
 Compute mean and Std. Dev. of metrics

return Trained LSTM, performance metrics, robustness assessment

5.2.4 Results and Discussion

5.2.4.1 Uncertainty-Based Training in a Single Substation

This subsection presents the performance of the LSTM model trained on data from the Beacon Hill substation. Initially, the model was trained on the original dataset, and subsequently on datasets augmented via the proposed DA technique that incorporates measurement uncertainty. The forecasting task focused on one-week-ahead predictions in a sequence-to-sequence setting, where the model predicts the entire load profile of the following week based on measurements from the previous week. Data from 2017–2019 were used, with 2019 reserved for testing, to emphasize the impact of DA on a relatively small training set.

Training employed a sliding window approach (one week input, one week prediction) with a stride of one week, comparing results after 50 and 100 epochs. Table 5.5 reports the error metrics for these baseline configurations. Table 5.6 shows improvements achieved by applying DA during training with uncertainty levels of 0.5 % and 1.0 %, and different augmentation factors e.g., “DA 1” indicates that, for each real training year, one additional synthetic year is generated by perturbing all measurements of that year according to the selected uncertainty level, while preserving the original temporal structure). DA was applied exclusively to the training set, whereas the test year remained unchanged and purely real.

Results demonstrate that DA consistently improves accuracy over the baseline, even outperforming the 100-epoch baseline despite training for only 50 epochs. Differences between 0.5 % and 1.0 % uncertainty levels are minimal, so 0.5 % was selected for subsequent experiments. Figure 5.7 illustrates predicted versus actual power for a sample week using DA 2 at 0.5 % uncertainty, showing good agreement and a slight conservative bias in predictions.

Table 5.7 compares training times with and without DA, revealing increased computational cost due to augmentation but justified by the accuracy gains. To further validate these findings, Tables 5.8 and 5.9 report error metrics and training times for extended datasets covering 2017–2020 and 2017–2022, respectively, with 2020 and 2022 as test years. In all cases, DA consistently improves performance across all metrics, with larger augmentation levels yielding greater gains but also higher training times.

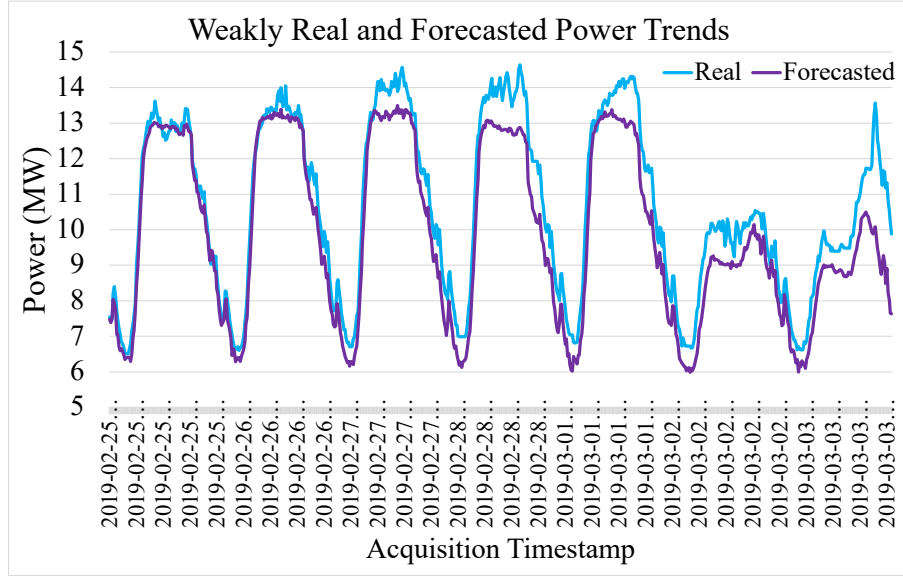


FIGURE 5.7: Weekly trend of actual (blue) vs. predicted (violet) power values with DA 2 on Beacon Hill data (2017-2019).

Figures 5.8 summarize MAPE values and training times across all tested configurations. Notably, for each considered test year, models trained on uncertainty-augmented datasets consistently outperform models trained on longer but unaugmented datasets when evaluated on their corresponding test sets. This highlights the benefit of synthetic data diversity introduced by DA, compared to simply increasing the amount of real training data, under otherwise comparable evaluation conditions. The best accuracy (lowest MAPE) is achieved with DA 2 on the 2017-2020 dataset, although this comes with increased training time, illustrating the trade-off between accuracy and computational cost.

TABLE 5.5: Error metrics for LSTM model without DA – Beacon Hill (2017-2019).

Epochs	MAE (MW)	RMSE (MW)	MAPE (-)
50	1.20	1.56	0.12
100	0.99	1.30	0.10

TABLE 5.6: Error metrics for LSTM model with DA – Beacon Hill (2017-2019).

Uncertainty	DA	MAE (MW)	RMSE (MW)	MAPE (-)
0.5 %	DA 1	0.91	1.25	0.10
	DA 2	0.86	1.19	0.09
	DA 4	0.80	1.13	0.08
1.0 %	DA 1	0.90	1.20	0.09
	DA 2	0.89	1.18	0.09
	DA 4	0.88	1.18	0.09

TABLE 5.7: Training time for LSTM model without and with DA – Beacon Hill (2017-2019).

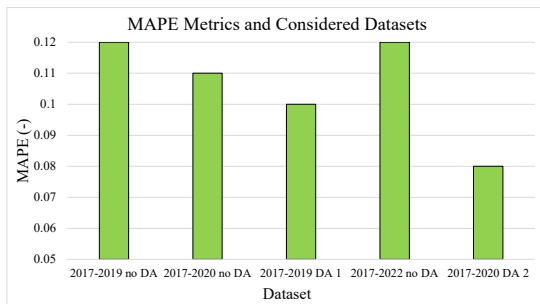
Training Dataset	Time (s)
No DA - 50 epochs	57.45
No DA - 100 epochs	113.87
DA 1 - 50 epochs	101.79
DA 2 - 50 epochs	148.22
DA 4 - 50 epochs	240.95

TABLE 5.8: Error metrics and training time for LSTM model – Beacon Hill (2017-2020).

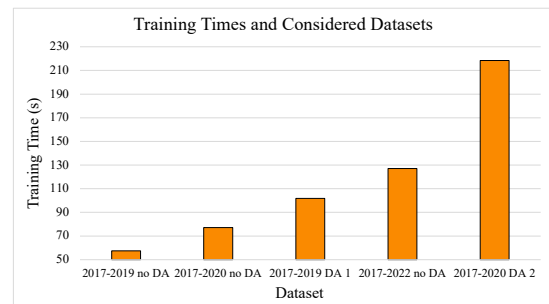
Dataset	MAE (MW)	RMSE (MW)	MAPE (-)	Training Time (s)
No DA	1.11	1.48	0.11	77.16
DA 1	0.90	1.40	0.09	149.23
DA 2	0.84	1.23	0.08	218.43
DA 4	0.80	1.09	0.07	358.85

TABLE 5.9: Error metrics and training time for LSTM model – Beacon Hill (2017-2022).

Dataset	MAE (MW)	RMSE (MW)	MAPE (-)	Training Time (s)
No DA	1.07	1.51	0.12	127.11
DA 1	1.03	1.41	0.11	244.22
DA 2	0.99	1.28	0.09	364.91
DA 4	0.90	1.11	0.09	593.78



(A) MAPE values.



(B) Training and testing times.

FIGURE 5.8: Comparison of (A) MAPE and (B) times with respect to considered datasets of Beacon Hill data.

To further assess the effectiveness of DA in forecasting power measurements, the LSTM model was tested using a one-week input and a one-day output horizon. This shorter prediction window is designed to support operational decisions by focusing on near-term power demand. Two stride configurations were evaluated: one-week and one-day strides, both with 50 training epochs, using the 2017-2019 dataset (2019 as test year). In the 1-week stride configuration, the input/output windows are advanced by exactly one week, resulting in non-overlapping weekly samples that span the entire test year (2019). Therefore, the model does not predict a single fixed weekday; rather, it produces consecutive weekly forecasts over the whole year, with each forecast conditioned on the immediately preceding week. The main difference with the 1-day stride is the number of training/testing windows (and thus the computational cost), not the temporal coverage of the test set.

Table 5.10 presents the results for both configurations. As expected, the one-day stride yields significantly better forecasting accuracy across all metrics, but at the cost of a sixfold increase in training time. To maintain computational efficiency in later DA experiments, the one-week stride was selected as a baseline.

Table 5.11 reports the performance of the model trained with DA under the one-week stride configuration. Despite the longer stride, DA significantly improves model accuracy, nearing the performance of the more computationally expensive one-day stride while keeping training time much lower. These findings confirm that DA not only enhances prediction quality but also allows for computationally efficient training, offering a practical alternative to reducing stride.

TABLE 5.10: Error metrics and training times for different strides – Beacon Hill (2017-2019).

Stride	MAE (MW)	RMSE (MW)	MAPE (-)	Training Time (s)
1 week	1.13	1.51	0.11	57.87
1 day	0.64	0.93	0.06	378.30

TABLE 5.11: Error metrics and training times with DA (one-Week Stride) – Beacon Hill (2017-2019).

DA	MAE (MW)	RMSE (MW)	MAPE (-)	Training Time (s)
DA 1	0.70	0.92	0.07	105.05
DA 2	0.67	0.89	0.07	148.08
DA 4	0.65	0.90	0.07	244.05

Regarding the statistical assessment, Table 5.12 reports the K–S test results for the original and augmented datasets at the Beacon Hill substation, considering the years 2017–2021. The K–S test is applied to the empirical distributions of the power measurements in the original dataset and in the corresponding uncertainty-augmented dataset (same substation and same years). In practice, for each DA configuration the set of original power samples is compared against the augmented samples generated via uncertainty perturbation, to verify whether DA alters the global data-generating distribution. The year 2022 was excluded, as it was reserved exclusively for model testing and not subjected to any DA. In all cases, the K–S statistics remain very small, and the associated p-values are close to 1.0, indicating that the augmented datasets maintain the same underlying distribution as the original measurements. Slight increases in the K–S statistic for higher augmentation levels reflect minor, controlled variability introduced by uncertainty perturbations – sufficient to enhance model robustness without altering the global statistical characteristics of the data.

TABLE 5.12: K–S test results for original and augmented datasets – Beacon Hill data (2017–2021).

Uncertainty (%)	DA	K-S Statistic (-)	P-Value (-)
0.5	DA 1	0.0008	1.0000
	DA 2	0.0009	0.9999
	DA 4	0.0011	0.9947
1.0	DA 1	0.0007	1.0000
	DA 2	0.0009	0.9999
	DA 4	0.0012	0.9908

Figure 5.9 illustrates the distributions of the original and augmented datasets for the years 2017–2021 at 1.0 % uncertainty. The histogram demonstrates that the augmented data closely follow the original distribution, with slight, localized variability around each measurement – consistent with the applied measurement uncertainty.

Tables 5.13 and 5.14 summarize the paired t-test results comparing the baseline configuration (no DA) with uncertainty-driven augmentation at 0.5 % and 1.0 % uncertainty levels, respectively. The very low p-values confirm that DA introduces statistically significant variations in model performance. At 0.5 % uncertainty, DA 1 and DA 4 yield clear improvements in MAE and MAPE, while DA 2 shows limited effects, particularly for RMSE. At 1.0 % uncertainty, all DA configurations significantly affect the results. The sign of the t-statistics indicates the direction of impact relative to the baseline, revealing

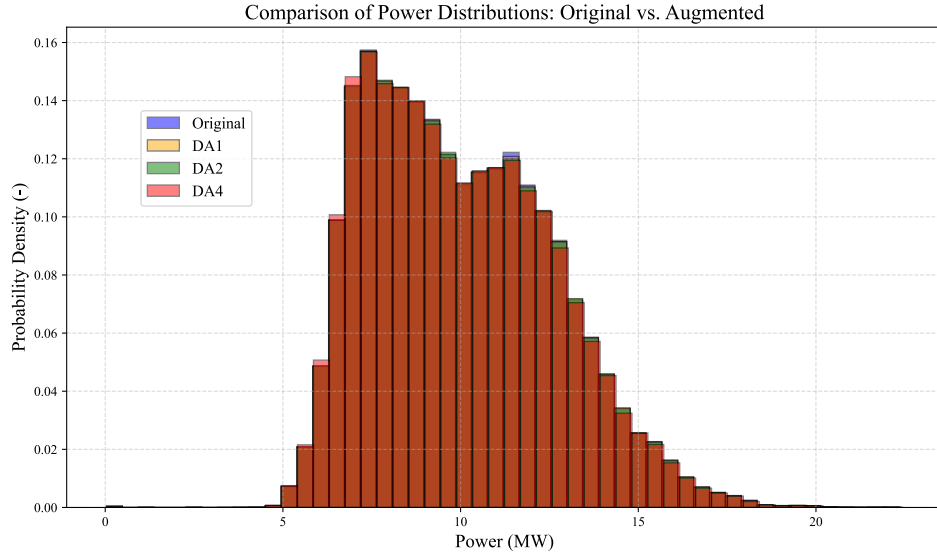


FIGURE 5.9: Distribution of original and augmented power values for Beacon Hill data – DA with 1.0 % uncertainty (2017-2021).

that DA consistently modifies model behavior – although not always in the same direction – thus confirming its influence on model learning dynamics.

TABLE 5.13: Paired t-test results for original vs. augmented datasets – Beacon Hill data (2017–2022) – 0.5 % uncertainty.

DA	Metric	T-Statistic (-)	P-Value (-)
DA 1	MAE	22.7516	0.0000
	RMSE	2.6812	0.0073
	MAPE	18.0610	0.0000
DA 2	MAE	26.5337	0.0000
	RMSE	0.2702	0.7870
	MAPE	18.8624	0.0000
DA 4	MAE	35.2978	0.0000
	RMSE	6.1012	0.0000
	MAPE	28.2244	0.0000

Overall, the results confirm that the proposed uncertainty-based DA strategy introduces statistically relevant yet controlled variability, improving the model’s robustness and its ability to generalize under realistic measurement conditions.

5.2.4.2 Cross-Substation Generalization Using Augmented Data

To evaluate the model’s ability to generalize across substations, the LSTM was trained on Beacon Hill data (2017-2021) and tested on Auburn data (2022), a substation not seen

TABLE 5.14: Paired t-test results for original vs. augmented datasets – Beacon Hill data (2017–2022) – 1.0 % uncertainty.

DA	Metric	T-Statistic (-)	P-Value (-)
DA 1	MAE	18.4911	0.0000
	RMSE	−9.7327	0.0073
	MAPE	13.8740	0.0000
DA 2	MAE	17.2273	0.0000
	RMSE	−3.0604	0.0010
	MAPE	10.1304	0.0000
DA 4	MAE	20.4356	0.0000
	RMSE	−1.9383	0.0000
	MAPE	9.9984	0.0000

during training. This setup helps assess how well the model handles geographically and operationally distinct conditions. Two forecasting horizons were tested: one-week ahead and one-day ahead, both using a one-week input window and stride.

Table 5.15 presents the results for one-week-ahead forecasts, while Table 5.16 shows the one-day-ahead case. In both tests, training was repeated with and without DA using 0.5 % uncertainty, and models were trained for either 50 or 100 epochs (baseline) or 50 epochs with varying DA levels. As expected, applying DA generally improves accuracy metrics, particularly in the one-day-ahead configuration, and consistently outperforms or matches the 100-epoch non-augmented baseline.

Figure 5.10 illustrates a typical prediction week using DA 2, showing a good match between predicted and actual load trends. These results confirm that DA enhances model generalization even when moving across substations with different load profiles, reinforcing its robustness in heterogeneous scenarios.

TABLE 5.15: Error metrics and training time for LSTM model trained on Beacon Hill data (2017-2021) and tested on Auburn data (2022) – One-week-ahead forecasting.

Configuration	MAE (MW)	RMSE (MW)	MAPE (-)	Training Time (s)
No DA - 50 epochs	1.80	2.40	0.15	125.26
No DA - 100 epochs	1.67	2.22	0.14	289.02
DA 1 - 50 epochs	1.79	2.31	0.15	222.98
DA 2 - 50 epochs	1.68	2.15	0.15	361.42
DA 4 - 50 epochs	1.59	2.05	0.14	599.11

TABLE 5.16: Error metrics and training time for LSTM model trained on Beacon Hill data (2017-2021) and tested on Auburn data (2022) – One-day-ahead forecasting.

Configuration	MAE (MW)	RMSE (MW)	MAPE (-)	Training Time (s)
No DA - 50 epochs	1.80	2.40	0.16	125.26
No DA - 100 epochs	1.67	2.22	0.17	289.02
DA 1 - 50 epochs	1.79	2.31	0.15	222.98
DA 2 - 50 epochs	1.68	2.15	0.15	361.42
DA 4 - 50 epochs	1.59	2.05	0.16	599.11

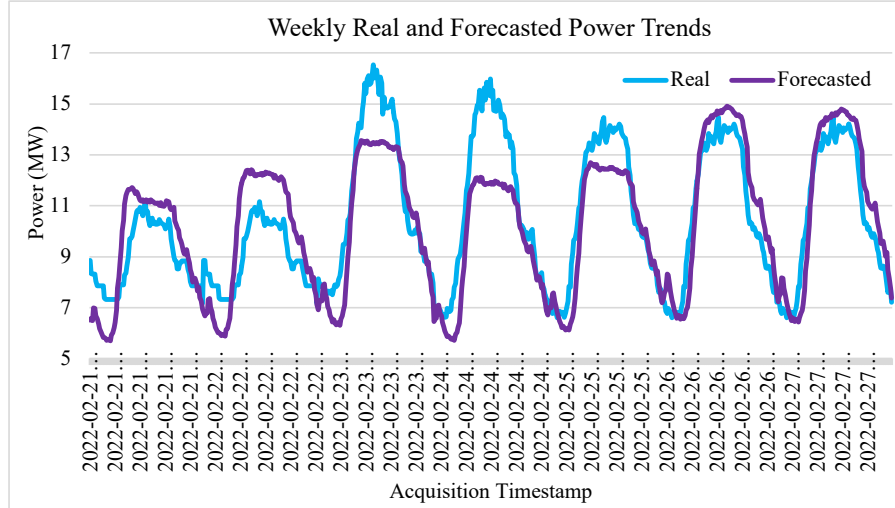


FIGURE 5.10: Weekly trend of actual (blue) vs. predicted (violet) with DA 2 on Beacon Hill data for training (2017-2021) – Tested on Auburn data (2022).

5.2.4.3 Uncertainty-Aware Testing on Different Substations

This section analyzes the robustness of the LSTM model by performing 1000 inference runs on corrupted test data from the Beacon Hill and Auburn substations. Input uncertainty levels of 0.5 % and 1.0 % were considered, with results reported in terms of the mean and standard deviation of the error metrics computed across runs.

Unlike previous sections, metrics here are reported with four decimal digits to highlight subtle but consistent variations across runs. The low standard deviation values indicate minimal sensitivity to the stochastic nature of the input noise.

Tables 5.17 and 5.18 report results for Beacon Hill under 0.5 % and 1.0 % input uncertainty, respectively, using one-week-ahead forecasting and one-week stride. The model shows consistent performance under noisy conditions, with a slight decline in accuracy for the 2017-2022 case, especially under 1 % uncertainty. This may be due to increased variability in the longer training set, which slightly hampers generalization.

Table 5.19 presents the same analysis for Auburn over 2017-2022. Similar trends are observed, reaffirming the model’s resilience to corrupted inputs across both locations.

TABLE 5.17: Mean and standard deviation of error metrics for LSTM model – Beacon Hill data – tested on 0.5 % uncertainty corrupted data – One-week-ahead forecasting.

Years	MAE (MW)	Std. Dev. (MW)	RMSE (MW)	Std. Dev. (MW)	MAPE (-)	Std. Dev. (-)
2017-2019	1.0207	0.0002	1.3645	0.0002	0.1020	0.0002
2017-2020	0.8400	0.0002	1.2294	0.0002	0.0832	0.0000
2017-2022	1.0052	0.0002	1.4359	0.0002	0.1083	0.0000

TABLE 5.18: Mean and standard deviation of error metrics for LSTM model – Beacon Hill data – tested on 1.0 % uncertainty corrupted data – One-week-ahead forecasting.

Years	MAE (MW)	Std. Dev. (MW)	RMSE (MW)	Std. Dev. (MW)	MAPE (-)	Std. Dev. (-)
2017-2019	0.8901	0.0003	1.2223	0.0004	0.0896	0.0000
2017-2020	0.8453	0.0003	1.2298	0.0004	0.0840	0.0000
2017-2022	1.0542	0.0003	1.4636	0.0003	0.1156	0.0000

TABLE 5.19: Mean and standard deviation of error metrics for LSTM model on Auburn data (2017-2022) – tested on uncertainty corrupted data – One-week-ahead forecasting.

Uncertainty (%)	MAE (MW)	Std. Dev. (MW)	RMSE (MW)	Std. Dev. (MW)	MAPE (-)	Std. Dev. (-)
0.5	0.8901	0.0003	1.2223	0.0004	0.0896	0.0000
1.0	0.8453	0.0003	1.2298	0.0004	0.0840	0.0000

Finally, Figure 5.11 illustrates the mean and standard deviation of the error metrics for the one-day-ahead prediction scenario, respectively. The configuration parameters are consistent with those used in the previous experiments. Results reveal a clear trend: the accuracy of the LSTM model decreases as the level of input uncertainty increases, for both the Beacon Hill and Auburn substations. In particular, the transition from 0.5 % to 1.0 % uncertainty leads to higher error values, with a more significant degradation observed at the Auburn site. Despite this, the standard deviation values remain consistently low across all settings, confirming that the model’s performance is stable and reproducible even when exposed to noisy input data.

5.2.4.4 Model Ensemble for Training Variability Assessment

To evaluate the variability induced by random initialization during training, an ensemble of ten LSTM models was trained using data from the Beacon Hill substation. All models shared the same architecture, hyperparameters, and data partitions (2017–2021 for training and 2022 for testing), differing only in their random seeds. A sliding-window

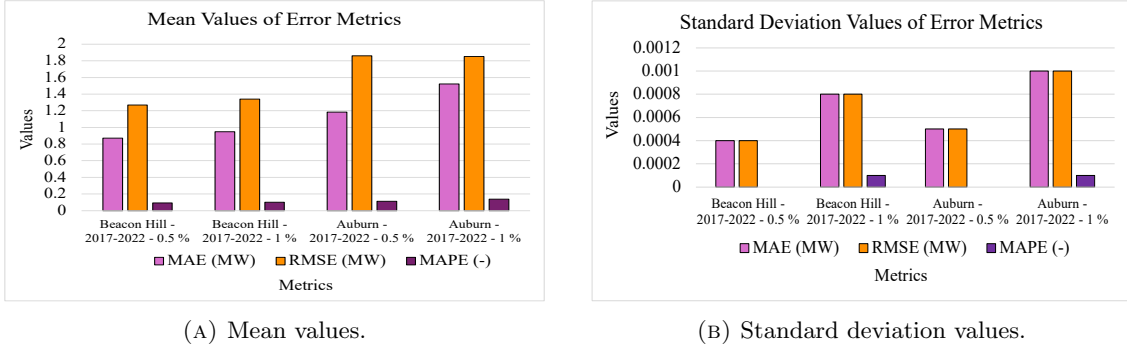


FIGURE 5.11: Error metrics (A) mean and (B) standard deviation values for Beacon Hill and Auburn substations with test data (2022) corrupted by 0.5 % and 1.0 % uncertainty – One-day-ahead forecasting.

configuration was adopted, where one week of input data was used to predict the subsequent week, with a stride of one week between samples. Each model was trained for 50 epochs.

Table 5.20 summarizes the ensemble results in terms of the mean and standard deviation of MAE, RMSE, and MAPE. The ensemble achieved a mean MAE of approximately 1.10 MW and RMSE of 1.67 MW, both with low standard deviations. The MAPE value of 0.15 ± 0.02 confirms consistent and stable one-week-ahead forecasting performance across all runs. These results indicate that stochastic factors related to weight initialization have a minimal effect on predictive accuracy and overall model behavior.

TABLE 5.20: Mean and standard deviation of error metrics for the LSTM ensemble model – Beacon Hill data (2017–2022) – One-week-ahead forecasting.

	MAE (MW)		RMSE (MW)		MAPE (-)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
LSTM Ensemble	1.10	0.18	1.67	0.22	0.15	0.02

5.2.4.5 Dropout-Based Uncertainty Estimation during Testing

A complementary experiment was performed to quantify epistemic uncertainty at inference time through Monte Carlo dropout. The LSTM model was trained on the same dataset configuration (Beacon Hill data, 2017–2021 for training, and 2022 for testing) using a one-week sliding window, identical to the previous setup, and trained for 50 epochs. During testing, dropout was activated to perform multiple stochastic forward passes, allowing the estimation of prediction variability due to inference randomness.

Table 5.21 presents the mean and standard deviation of MAE, RMSE, and MAPE computed across all Monte Carlo iterations. The model achieved accurate predictions with low variability, confirming its robustness and stable predictive behavior under stochastic inference conditions.

TABLE 5.21: Mean and standard deviation of error metrics for the LSTM model using Monte Carlo dropout – Beacon Hill data (2017–2022) – One-week-ahead forecasting.

	MAE (MW)		RMSE (MW)		MAPE (-)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Monte Carlo Dropout	1.17	0.03	1.76	0.04	0.161	0.003

5.2.4.6 Discussion

The results obtained in this study highlight the relevance of integrating metrological considerations into data-driven forecasting models, particularly in the energy domain. The proposed metrology-aware framework leverages measurement uncertainty as a resource rather than treating it as a nuisance, providing a more realistic training environment through DA techniques. This approach proved effective not only in improving predictive performance but also in enhancing the robustness of DL models against variability inherent to real-world sensor data. Notably, the cross-substation generalization observed in the experiments indicates that the model is capable of adapting to diverse operational conditions without the need for extensive retraining, which is a key advantage in scalable energy forecasting applications.

The analyses presented in the previous subsections provide a comprehensive assessment of the various sources of uncertainty that influence the LSTM model’s predictive performance. The standard deviation values obtained through Monte Carlo dropout are smaller than those from the ensemble-based analysis, suggesting that inference stochasticity introduces less variability than training randomness. However, when compared with the uncertainty-aware testing performed on input-perturbed data with 0.5 % and 1.0 % uncertainty levels, the variability associated with dropout remains notably higher.

This observation implies that measurement uncertainty – representing sensor accuracy and acquisition precision – has a negligible effect on the model’s stability, whereas epistemic uncertainty, originating from stochastic elements in training and inference, plays a

dominant role. Quantitatively, expressing the standard deviation as a percentage of the corresponding mean error reveals that:

- Input perturbations cause relative variability of approximately 0.02-0.03 %.
- Ensemble-based training variability is around 16 %.
- Monte Carlo dropout inference variability is about 2.6 %.

Hence, the contribution of instrumental uncertainty to the overall predictive variability is negligible compared with the epistemic components arising from the learning process. These findings confirm that the LSTM model is robust to realistic levels of measurement noise, and that the main source of uncertainty in its forecasts stems from the stochastic nature of model training and inference rather than from input data quality.

However, the method also introduces certain trade-offs. The injection of uncertainty during training and testing can lead to slightly more conservative predictions which, while advantageous in safety-critical contexts, may reduce responsiveness in highly dynamic environments. Moreover, the reliance on publicly available datasets, though valuable for benchmarking and reproducibility, may not fully capture the heterogeneity of sensor characteristics encountered in real-world installations. Future research could extend the framework to account for correlated uncertainties, sensor drift over time, and other degradation mechanisms that reflect operational aging and environmental influences.

Overall, this study reinforces the importance of bridging AI methodologies with metrological rigor. By embedding uncertainty directly into the learning pipeline, the proposed framework not only aligns with measurement science standards but also paves the way for deploying more trustworthy, explainable, and field-ready AI systems in power grid applications.

5.3 Role of Temperature in Measurement Accuracy

The evolution of smart grid infrastructures has been significantly driven by the integration of ITs, which enable accurate and continuous monitoring of electrical networks. Extensive research has addressed various aspects of ITs, including their design [241], performance under distorted conditions [242], sensitivity to external influences [243, 244], and overall

characterization [245]. Standardization efforts, particularly the IEC 61869 series, provide the reference framework for both conventional and Low-Power Instrument Transformers (LPITs). The most recent update of IEC 61869-1 [246] consolidates key requirements for reliable IT operation.

Despite this, assessing the real-time accuracy of ITs remains a critical and unresolved challenge, due to the complex and variable environmental conditions – especially temperature – experienced in field deployments. Prior studies have proposed methods such as synchrophasor-based correction factors [247] and mobile calibration setups [248], yet these solutions often face practical limitations.

This research addresses the problem by proposing a novel, data-driven method for estimating key IT accuracy parameters (i.e., ratio error (ε) and phase displacement ($\Delta\varphi$)) during actual operating conditions. The method leverages: (i) preliminary off-site characterizations across varying temperature conditions; (ii) ANNs to model the temperature-dependent behavior of ITs; and (iii) embedded microcontroller platforms to enable real-time or near-real-time deployment. This approach aims to improve grid measurement reliability through a scalable and cost-effective solution.

5.3.1 Data Collection and Preprocessing

The dataset used to train the proposed ANN-based method (as described in Section 2.3.3) was obtained through the systematic characterization of two MV resistive Low-Power Voltage Transformers (LPVTs). Both Devices Under Test (DUTs-A and -B) belong to accuracy class 0.5 and fall within the $-5/40^\circ\text{C}$ temperature category, as defined by IEC 61869-1 [246]. Their technical specifications are summarized in Table 5.22.

TABLE 5.22: Characteristics of the LPVTs under test.

Rated Voltage (kV)	Dimensions (cm)	Type	Nominal Ratio (V/V)
20	30x10x10	Resistive	20000/3.25

The characterization campaign lasted 25 weeks and involved an aging process in which the LPVTs were exposed to controlled thermal variations ranging from 20°C to 50°C . These values were selected to stress the devices without inducing irreversible degradation. During the test period, the transformers were energized at nominal voltage, while the

following quantities were monitored and recorded: ambient temperature, device temperature, LPVT output voltage, and the output voltage of a reference transformer.

From these measurements, the ratio error ε and the phase displacement $\Delta\varphi$ were computed and stored. The ratio error ε (in percentage) is defined as:

$$\varepsilon = \left(\frac{K_n I_{\text{out}} - I_{\text{in}}}{I_{\text{in}}} \right) \times 100, \quad (5.4)$$

where K_n is the nominal transformation ratio, I_{out} the measured output, and I_{in} the input. This metric reflects the amplitude deviation between the actual and ideal sensor response.

The phase displacement $\Delta\varphi$ is computed as:

$$\Delta\varphi = \varphi_{\text{out}} - \varphi_{\text{in}}, \quad (5.5)$$

where φ_{out} and φ_{in} denote the phases of the output and input waveforms, respectively.

For ANN training, 70 representative measurement samples were extracted each week from the collected dataset, resulting in a rich and temporally diverse input space. The dataset was split into 60 % for training, 20 % for validation, and 20 % for testing. Regarding thermal data, only the internal temperature of the DUTs was retained as a predictive input feature for the neural network, as it better reflects the operating conditions affecting transformer accuracy. An illustrative subset of the collected data is shown in Figure 5.12, where the ratio error trends for DUT-A and DUT-B are plotted in violet and pink, respectively, alongside the corresponding device temperature (orange curve). It is worth noting that other temperature acquisition strategies – such as those defined in IEC 61869-6 [249] – could be equally applied in similar characterization workflows.

Starting from this dataset, multiple data configurations were designed to explore the ANN model's performance under different operating assumptions:

- **Input Variable Sets:** Two configurations were considered. The first uses only the device temperature, while the second incorporates both temperature and secondary voltage as input features.

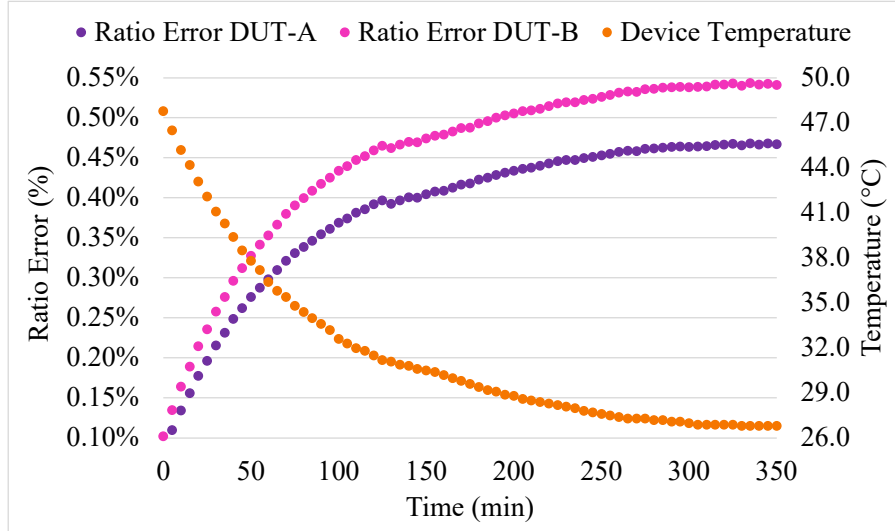


FIGURE 5.12: Example of measurement result from the characterization process.

- **Dataset Size:** Two subsets of data were evaluated. One includes all 25 weeks of acquisition, and the other is limited to the first 10 weeks. This choice allows analysis of the model's sensitivity to the amount of available training data.
- **Estimation Mode:** A distinction was made between instantaneous and long-term estimation. Instantaneous estimation treats all measurements as if acquired during a generic time window, omitting temporal information (e.g., the week of acquisition). In contrast, long-term estimation includes the week index (from 1 to 25) as an input to model gradual accuracy drift over time, supporting aging analysis.

Depending on the configuration, the ANN's input features consisted of either temperature alone, temperature and voltage, or temperature, voltage, and week of acquisition. In all cases, the output values to be inferred by the model were ε and $\Delta\varphi$.

To enhance model performance, input features were preprocessed. Specifically, temperature and voltage values were normalized to the $[0, 1]$ range. For the week of acquisition, one-hot encoding was applied to transform the categorical variable into a binary vector where a single element is set to 1 (indicating the current week), and all others are 0. This representation improves the ANN's ability to learn from categorical inputs in numerical form.

To evaluate the accuracy of the model, three performance metrics commonly used in regression tasks were employed: MAE (Equation 2.16), RMSE (Equation 2.18), and MAPE (Equation 2.19).

5.3.2 Artificial Intelligence Model

In this research, a FNN with four layers is proposed: an input layer, two hidden layers, and an output layer. The activation functions chosen for these layers are ReLU (Equation 2.2) for the input and first hidden layers, and Tanh (Equation 2.5) for the second hidden layer. The selection of ReLU for the input and first hidden layers is motivated by its computational efficiency and its ability to mitigate the vanishing gradient problem, thereby accelerating the model's convergence during training. Conversely, the Tanh activation is employed in the second hidden layer due to its capacity to handle a broader range of input values and produce outputs within the $[-1, 1]$ interval. This characteristic helps prevent neuron saturation and enhances the network's ability to capture complex patterns in the data. For the output layer, no activation function is applied, effectively using a linear activation. This is a common practice in regression tasks, as it allows the network to generate any real-valued output without constraints imposed by nonlinear activation functions, enabling direct prediction of continuous target variables.

The network is trained using the Adam optimizer and the MAE (Equation 2.16) loss function, both recognized for their effectiveness in training ANNs. The implementation is carried out using the Keras API [210] within the TensorFlow library [211].

5.3.3 Results and Discussion

The results obtained from the testing phase are presented below, organised according to the dataset utilized, as described in Section 5.3.1.

5.3.3.1 Instantaneous Accuracy Estimation Using Temperature Only

Tables 5.23 and 5.24 report the performance metrics achieved by DUT-A and DUT-B, respectively, when only temperature is used as the input feature for estimating the ratio error (ε) and phase displacement ($\Delta\varphi$). In all tables, the number of decimal places for error metrics has been set in accordance with the standard deviation of ratio error and phase displacement measurements.

The results indicate strong estimation capability, particularly for the ratio error in DUT-B, which exhibits the lowest error values. This can be visually confirmed in Figure 5.13, where the true and predicted values of ε for DUT-B are compared.

TABLE 5.23: ANN metrics on DUT-A using temperature input feature.

Metric	MAE	RMSE	MAPE (-)
ε	0.02 %	0.03 %	0.14
$\Delta\varphi$	1.2 mrad	1.5 mrad	0.5

TABLE 5.24: ANN metrics on DUT-B using temperature input feature.

Metric	MAE	RMSE	MAPE (-)
ε	0.01 %	0.01 %	0.04
$\Delta\varphi$	1.2 mrad	1.5 mrad	0.2

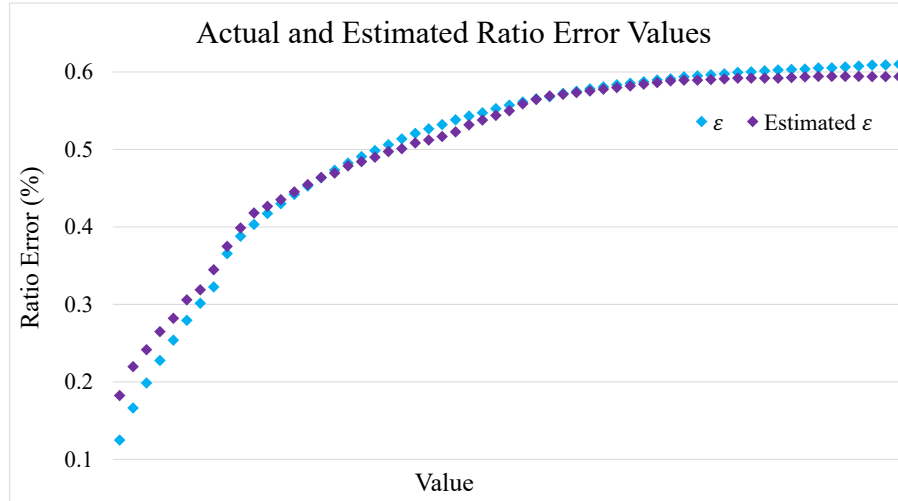


FIGURE 5.13: Real (blue) and inferred (violet) ratio error values result from the test of the ANN on DUT-B data with temperature input feature.

5.3.3.2 Instantaneous Accuracy Estimation Using Temperature and Voltage

Tables 5.25 and 5.26 show the metrics for DUT-A and DUT-B, respectively, when both temperature and voltage are provided as input features to the ANN model. The inclusion of voltage generally reduces the errors in ratio estimation, while phase displacement errors show a slight increase. However, these differences are minimal, suggesting that the decision to include voltage should consider factors such as sensor availability and installation constraints.

TABLE 5.25: ANN metrics on DUT-A with temperature and voltage input features.

Metric	MAE	RMSE	MAPE (-)
ε	0.02 %	0.03 %	0.14
$\Delta\varphi$	1.2 mrad	1.5 mrad	0.5

TABLE 5.26: ANN metrics on DUT-B with temperature and voltage input features.

Metric	MAE	RMSE	MAPE (-)
ε	0.01 %	0.01 %	0.04
$\Delta\varphi$	1.2 mrad	1.5 mrad	0.2

5.3.3.3 Effect of Dataset Size on Instantaneous Accuracy Estimation

To assess how the quantity of training data influences model performance, Table 5.27 presents results obtained using only the first 10 weeks of data, compared to 25 weeks in previous experiments. This analysis was conducted on DUT-A data with temperature and voltage as inputs.

The results reveal a degradation in performance, especially for phase displacement estimation. Notably, the model trained on 10 weeks of data shows signs of overfitting after epoch 70, despite being trained for the same 150 epochs as the 25-week model. This suggests that limiting training duration may improve results in smaller datasets.

TABLE 5.27: ANN metrics on DUT-A with temperature and voltage input features using 10-weeks dataset.

Metric	MAE	RMSE	MAPE (-)
ε	0.04 %	0.05 %	0.09
$\Delta\varphi$	2.5 mrad	3.0 mrad	0.9

5.3.3.4 Long-Term Accuracy Estimation Using Temperature

Tables 5.28 and 5.29 summarize results of the ANN model trained on weekly aggregated data combined with temperature for DUT-A and DUT-B, respectively. Overfitting is observed starting from epoch 40 in both cases.

TABLE 5.28: ANN metrics on DUT-A with temperature and weekly input features.

Metric	MAE	RMSE	MAPE (-)
ε	0.10 %	0.13 %	0.29
$\Delta\varphi$	2.1 mrad	2.5 mrad	0.7

TABLE 5.29: ANN metrics on DUT-B with temperature and weekly input features.

Metric	MAE	RMSE	MAPE (-)
ε	0.18 %	0.23 %	0.39
$\Delta\varphi$	2.9 mrad	3.3 mrad	0.4

5.3.3.5 Long-Term Accuracy Estimation Using Temperature and Voltage

Tables 5.30 and 5.31 report the metrics when voltage is added as an input along with temperature and weekly information. In general, including voltage reduces the errors across all cases except for the ratio error estimation on DUT-A.

TABLE 5.30: ANN metrics on DUT-A with temperature, voltage, and weekly input features.

Metric	MAE	RMSE	MAPE (-)
ε	0.15 %	0.16 %	0.41
$\Delta\varphi$	1.5 mrad	1.8 mrad	0.5

TABLE 5.31: ANN metrics on DUT-B with temperature, voltage, and weekly input features.

Metric	MAE	RMSE	MAPE (-)
ε	0.06 %	0.07 %	0.30
$\Delta\varphi$	2.5 mrad	2.9 mrad	0.3

5.3.3.6 Long-Term Accuracy Estimation vs. Dataset Size

Table 5.32 shows the ANN performance considering only the first 10 weeks of data with temperature, voltage, and weekly input features on DUT-A.

Here, the estimation errors for phase and ratio show different trends. Specifically, while the ratio error MAPE decreases, the phase error significantly increases. This suggests that dataset size has a stronger impact on phase displacement estimation than on ratio error. It is also important to consider the different units of measurement when comparing these metrics.

TABLE 5.32: ANN metrics on DUT-A with temperature, voltage, and weekly input features using 10-weeks dataset.

Metric	MAE	RMSE	MAPE (-)
ε	0.10 %	0.12 %	0.32
$\Delta\varphi$	2.0 mrad	2.2 mrad	0.8

5.3.3.7 Discussion

Figures 5.14 illustrates the MAPE values for ratio error and phase displacement estimation, respectively, considering DUT-A data for brevity. These results are reported for all the considered cases: 1) Instantaneous accuracy estimation with temperature, 2) Instantaneous accuracy estimation with temperature and voltage, 3) Instantaneous accuracy estimation versus dataset size, 4) Long-term accuracy estimation with temperature, 5) Long-term accuracy estimation with temperature and voltage, and 6) Long-term accuracy estimation versus dataset dimension.

The results indicate that the proposed approach achieves superior performance in instantaneous accuracy estimation. Nevertheless, the model maintains satisfactory accuracy even in the context of long-term estimation, which may be viewed as a form of device aging.

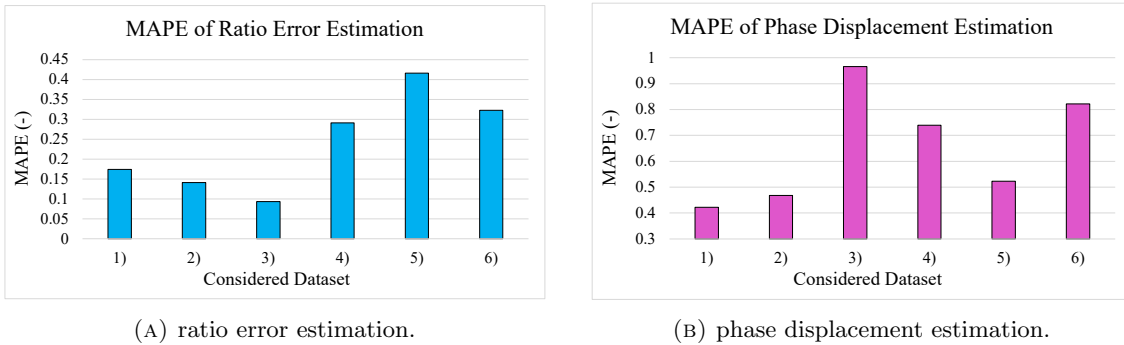


FIGURE 5.14: MAPE values of (A) ratio error and (B) phase displacement for all considered cases for DUT-A.

5.3.4 Microcontroller Implementation

After completing the training and testing phases, the developed ANN model was embedded into a microcontroller to evaluate its performance in a real-world application. For this purpose, the Arduino Nano 33 BLE Sense Rev 2 [215] was selected as the development platform. This compact board integrates a variety of features, including Bluetooth

connectivity and multiple onboard sensors such as an accelerometer, gyroscope, magnetometer, temperature, humidity, and pressure sensors. These characteristics make it highly suitable for IoT prototyping and experimentation. The board is powered by an Arm Cortex-M4 processor running at 64 MHz and is equipped with 1 MB of Flash memory and 256 KB of RAM, providing sufficient resources for running complex applications. The cost of the board is approximately 50 euros.

Implementing an ANN on a microcontroller involves several key steps: model preparation, conversion, and integration. Initially, the model is developed and trained within a generic Integrated Development Environment (IDE) on a computer. Subsequently, the model undergoes conversion to compress and optimize it for execution on embedded hardware, reducing resource demands. Finally, the optimized model is integrated into the microcontroller project, which also manages input and output data handling. In this case study, the **TensorFlow Lite** framework [250] was utilized for model conversion and integration into the Arduino IDE.

A primary challenge when deploying AI models on microcontrollers is limited memory capacity. Consequently, model compression is essential to enable deployment; however, highly complex ANNs may still require memory beyond what embedded devices can provide, even after compression.

The target application scenario involves collecting temperature data via the onboard sensor, which is then fed to the embedded ANN model to estimate the ratio error and phase displacement of the examined LPVT. The acquired data can either be stored locally in the Arduino memory or continuously transmitted to a central data concentrator.

To validate the embedded model's performance, the same ANN trained for instantaneous accuracy estimation with temperature input (see Section 5.3.3.1) was deployed on the Arduino board and tested with DUT-B data. The results are summarized in Table 5.33. The error metrics obtained on the Arduino closely match those from the computer-based implementation, with minor discrepancies (on the order of tenths) attributed to the limited precision of data representation on the microcontroller, which supports only two decimal digits.

Regarding execution time, the estimation of 358 test samples (pairs of ε and $\Delta\varphi$) was performed in roughly 100 ms on both the computer and the Arduino board, demonstrating

TABLE 5.33: ANN metrics on the DUT-B with temperature input feature executed on the Arduino board.

	MAE	RMSE	MAPE (-)
ε	0.01%	0.02%	0.05
$\Delta\varphi$	1.2 mrad	1.6 mrad	0.1

comparable processing speed.

5.3.5 Final Remarks

In conclusion, this research proposes a novel and effective methodology for the real-time evaluation of accuracy parameters in LPVTs under field operating conditions. By combining ANNs with microcontroller-based implementation, a simple, low-cost, and deployable solution is developed to estimate the ratio error and phase displacement directly from sensor inputs.

The ANN model, trained using a limited set of preliminary characterizations, demonstrated high accuracy in estimating the target parameters across different configurations and datasets. The inclusion of temperature, voltage, and time-related features further improved the model's generalization capability under various operating conditions. The successful deployment of the trained ANN on a resource-constrained device, namely the Arduino Nano 33 BLE Sense Rev 2 [215], confirms the practical applicability of the approach. The embedded model achieves estimation accuracy comparable to that of the desktop implementation, with negligible loss due to numerical precision limits and memory constraints. Moreover, the integration of the model within a microcontroller platform opens new opportunities for in-situ monitoring and decentralized control in power systems, especially in smart grid and IoT contexts. The approach offers an attractive alternative to traditional laboratory-based calibration methods, enabling continuous, real-time monitoring of IT accuracy without interrupting network operation.

Overall, the presented solution addresses a critical need for real-time, low-cost accuracy evaluation of LPVTs, and lays the groundwork for future developments in intelligent power system diagnostics, asset management, and quality assurance.

5.4 Advanced Modeling Approaches in Current Sensing

The integration of LPITs into modern power networks represents a key advancement in the pursuit of more accurate, reliable, and intelligent measurement infrastructures. Unlike legacy inductive ITs, LPITs offer improved safety, reduced secondary energy output, compact design, and extended measurement ranges. Technologies such as RCs, optical sensors, and capacitive voltage dividers are being increasingly adopted due to their compatibility with modern process bus protocols, enabling seamless digital communication with intelligent devices and merging units [251, 252].

Within the evolving context of smart grids, these sensing elements are essential for enabling real-time monitoring, adaptive protection, and data-driven control strategies. Their measurements, when reliably acquired across dynamic and harsh environments, feed into advanced grid applications such as fault detection, asset health monitoring, and predictive load forecasting. Moreover, their integration with cutting-edge digital technologies (e.g., AI, DTs, and IoT frameworks) positions LPITs at the forefront of next-generation metrological solutions [253, 254].

5.4.1 Rogowski Coils in Smart Sensing Applications

The RC is classified as a Low-Power Current Transformer (LPCT) and primarily used for current measurement. Its design and testing standards are defined in IEC61869-10 [255]. RCs offer several advantages over traditional inductive CTs: they are compact, lightweight, iron-free, and generally feature an openable structure that allows for convenient installation without service interruption.

Structurally, an RC consists of a uniform copper winding with N turns placed around a non-magnetic insulating core that encircles the primary conductor. The coil produces an output voltage $u_s(t)$ that is proportional to the time derivative of the primary current I_p :

$$u_s(t) = -M \frac{dI_p}{dt}, \quad (5.6)$$

where M is the mutual inductance between the conductor and the coil. While this derivative output simplifies analog-to-digital interfacing, it also necessitates an integration stage – either analog or digital – to reconstruct the original current waveform. In this study,

the integration stage is excluded to isolate and model the coil behavior under varying environmental conditions using AI-based methods.

5.4.2 Devices Under Test: Commercial Rogowski Coils

To ensure consistency across experimental analyses, both modeling approaches presented in this section utilize the same three off-the-shelf RCs, labeled R1, R2, and R3. Their physical and electrical characteristics are summarized in Table 5.34. The transformation ratio at 50 Hz (TR_{50}), cross-sectional diameter d , inner radius a , and winding resistance R_w provide insight into their geometric and functional configurations. These parameters influence not only the dynamic behavior of the coils but also their thermal and frequency responses.

TABLE 5.34: Characteristics of the RCs under test.

Feature	R1	R2	R3
TR_{50} (mV/kA)	100	100	100
d (mm)	8	8	12
a (mm)	50	57	21.5
R_w (Ω)	256	381	22
Cross-Section	Circular	Circular	Oval
Accuracy Class	0.5	1	1
Bandwidth	[1 Hz, 100 kHz]	[50 Hz, 5 kHz]	[20 Hz, 5 kHz]
Operating Temperature ($^{\circ}\text{C}$)	[-30, 80]	[-20, 85]	[-20, 70]
Rated Current (A)	100	500	1000

The following subsections introduce two complementary strategies aimed at improving the accuracy, interpretability, and intelligence of RC-based measurement systems:

1. A **PINN-based model** for thermal compensation, trained on standard-compliant test data, which allows accurate estimation of ratio and phase errors across temperature variations.
2. A **lightweight behavioral model** implemented directly on a low-power microcontroller, enabling real-time validation of RC outputs and paving the way for embedded digital twins in edge devices.

5.4.3 Rogowski Coils Temperature Compensation via Physics-Informed Neural Networks

Recent developments in the deployment of LPITs demonstrate that these devices not only coexist with traditional inductive ITs during grid modernization phases but also potentially outperform them in terms of transient response, bandwidth, and immunity to magnetic saturation. Nonetheless, several challenges persist, including ensuring electromagnetic compatibility, calibration traceability, and maintaining accuracy under variable environmental conditions. Standardization efforts, notably IEC 61869-1 [246], establish regulatory frameworks that facilitate the widespread adoption of LPITs in MV and HV applications, particularly within smart grids and renewable energy integration contexts.

As utilities pursue cost-effective and scalable grid modernization strategies, LPITs are increasingly regarded as fundamental components enabling accurate, real-time monitoring, protection, and control across decentralized and complex infrastructures. Among LPIT technologies, the RC stands out due to its compact size, ease of installation, and measurement accuracy, making it a key sensor in smart grid applications. However, its sensitivity to environmental influences, especially temperature, remains a critical limitation affecting its operational reliability.

Temperature is a critical factor for many IT technologies. For example, optical CTS have been widely investigated for temperature-induced errors, with studies proposing thermal drift models and real-time compensation algorithms to ensure precision in varying environments [256]. Advanced modeling and simulation, including temperature-dependent birefringence and thermal expansion, have further improved accuracy under dynamic conditions [257]. Hall-effect sensors have also been addressed, with adaptive neural networks [258] and real-time drift correction mechanisms [259] demonstrating effectiveness in preserving measurement fidelity across temperature changes. Sensor design efforts highlight thermally stable materials and packaging strategies to reduce ratio error and phase displacement [260]. Finally, calibration and metrological research have assessed the impact of temperature on uncertainty, proposing advanced calibration setups and traceability frameworks to mitigate these effects [261].

Hybrid and mechanical sensor designs employ strategies such as dual-path magnetic coupling and 3D-printed compensation blocks to regulate thermal variations [262]. For

RCs, which are especially sensitive to temperature, recent advances include firmware-based corrections, optimized winding layouts, and reference thermal models to mitigate gain errors [263, 264]. In this context, this work focuses on exploiting standardized temperature tests to improve RC accuracy.

In the case of RCs, several innovations have been proposed in the literature, including firmware-based compensation, winding optimizations, and thermal modeling to reduce gain errors and improve stability under thermal stress. This research aims to estimate the output waveforms of off-the-shelf RCs by combining mandatory temperature tests prescribed in standards with PINN models. Training the PINNs on data acquired at three different temperatures enables the estimation of coil outputs across a continuous temperature range, improving in-field performance assessment and compensation. Moreover, signals containing frequency components beyond 50 Hz have been considered, thus covering the full PQ frequency range. As widely recognized, PQ is a crucial topic in the literature, with instrumentation playing a central role in its assessment [265].

Overall, the main contribution of this research lies in the integration of metrological principles with data-driven modeling to enhance the temperature compensation of RCs. The proposed approach incorporates standardized temperature tests and leveraging PINNs to predict the coil response across a wide temperature range. This enables a more accurate estimation of the RC output under varying environmental conditions. Furthermore, the study broadens the analysis to include complex excitation signals up to 3050 Hz, both single-tone and multi-tone, thus encompassing the full PQ frequency spectrum. The methodology is experimentally validated on commercial devices, demonstrating its practical applicability and robustness. In addition, a comprehensive performance evaluation is conducted, supported by metrological considerations that ensure the traceability and reliability of the obtained results. Finally, the research discusses implementation aspects and provides guidelines for applying the proposed compensation technique in real-world power measurement systems.

5.4.3.1 Scenario and Standard Tests for Rogowski Coils

According to IEC 61869-1 [246], ITs must undergo a temperature cycle accuracy test to validate their performance under specified environmental conditions. The test involves exposing the device to a thermal cycle encompassing minimum and maximum ambient

temperatures defined by its operating category (e.g., from -5°C to 40°C , -25°C to 40°C , or -40°C to 40°C), plus a reference temperature of 20°C , following the thermal profile depicted in 5.15. Each temperature plateau is maintained for at least three times the device's thermal time constant, as declared by the manufacturer, with accuracy measurements taken at the end of each stabilised period.

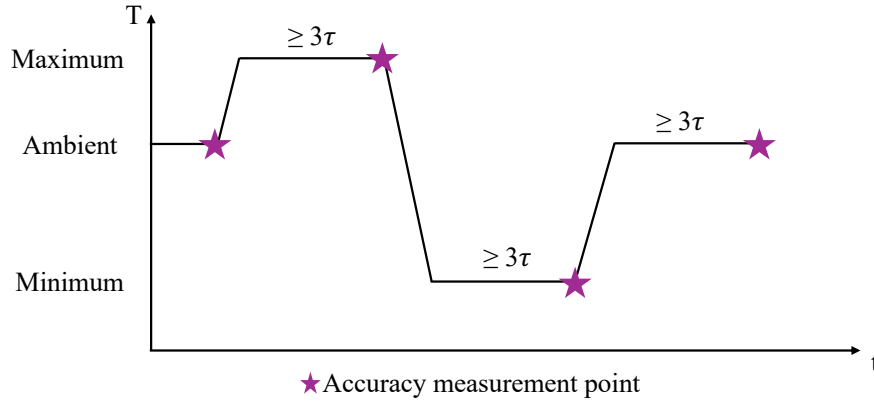


FIGURE 5.15: Temperature cycle accuracy test.

The measured errors at each temperature must comply with the assigned accuracy class limits under normal environmental conditions. If linear behavior can be proven above certain thresholds, testing may be performed with reduced primary current or voltage [246].

The motivation of this research is to utilize these mandatory temperature tests, combined with a physics-informed learning approach, to accurately predict and compensate for temperature-dependent ratio and phase errors in RCs. This approach supports improved accuracy and reliability in field operations, where temperature variations can cause accuracy deviations exceeding 100 %, thus necessitating effective compensation techniques [219, 263].

5.4.3.2 Data Collection

To assess the effectiveness and practical applicability of the proposed method, an experimental validation was conducted using three commercial RCs, labeled R1, R2, and R3, which specifications are listed in Table 5.34.

The RCs were initially characterized in terms of Fast Fourier Transform (FFT)-based ratio error (ε) and phase displacement ($\Delta\varphi$) (Equations 5.4 and 5.5). These metrics

quantify how accurately the sensor reproduces the magnitude and phase of the primary current signal. Data were acquired using the measurement setup described in [251]. In summary, a Fluke 6105A electrical calibrator [266] was employed to inject a sinusoidal current, while a precision reference shunt resistor was used to measure the primary current for phase comparison.

The characterization was performed by applying sinusoidal currents at 50 Hz and 100 A nominal amplitude while subjecting the coils to three standardized accuracy test temperatures: -5°C , 20°C , and 40°C , as prescribed in IEC 61869-1 [246]. This preliminary analysis, conducted in a prior measurement campaign, was intended to evaluate the temperature-induced variations in accuracy and stability of the coils.

The results are presented in Tables 5.35 and 5.36. The ratio error data reveal a pronounced temperature dependence. In particular, R2 and R3 show substantial variations – exceeding 100 % and 80 %, respectively – across the tested range, while R1 maintains relatively stable performance. Conversely, the phase displacement generally decreases with temperature for R1 and R2, whereas R3 displays a non-linear trend. These findings underscore the influence of temperature on sensor accuracy, which appears to be design-dependent.

TABLE 5.35: Ratio error of RCs at different temperatures.

Temperature ($^{\circ}\text{C}$)	R1 (%)	R2 (%)	R3 (%)
-5	-1.91	-1.31	-3.06
20	-1.90	-2.58	-3.27
40	-1.83	-3.53	-3.45

TABLE 5.36: Phase displacement of RCs at different temperatures.

Temperature ($^{\circ}\text{C}$)	R1 (mrad)	R2 (mrad)	R3 (mrad)
-5	0.71	0.75	0.87
20	0.69	0.70	0.79
40	0.68	0.66	0.82

After the characterization process, the data required for the implementation of the PINN model were collected. In particular, input and output waveforms were acquired from the three RCs under test, considering several operating temperatures. The dataset includes both the reference temperatures defined by the IEC 61869-1 [246] accuracy test

(-5°C , 20°C , and 40°C) and additional measurement points at -15°C , -10°C , 0°C , 5°C , 10°C , 15°C , 25°C , 30°C , 35°C , 45°C , 50°C , and 55°C .

Different types of excitation signals were employed during the acquisition campaign. These include:

- **Sinusoids** at 50 Hz with amplitudes of 50, 90, and 100 A.
- **Single-frequency sinusoids** at multiple harmonic frequencies (150 Hz, 250 Hz, 350 Hz, 550 Hz, 650 Hz, 850 Hz, 1250 Hz, and 2500 Hz) with amplitudes of 10 and 20 A.
- **Sinc-type signals** containing harmonic components from 150 Hz up to 3050 Hz, with a frequency step of 50 Hz, generated at two current amplitudes (50 A and 100 A).

The experimental setup employed a Fluke 6105A calibrator [266] in combination with its transconductance amplifier Fluke 52120 [267] to generate the test currents. The primary current was monitored through a precision reference shunt resistor (uncertainty of 2×10^{-6}) for phase and amplitude verification. For the generation of sinc-type signals, a general-purpose function generator was used to provide the broadband voltage excitation to the transconductance module, enabling accurate voltage-to-current conversion over the desired frequency range.

All the collected data were used for both the training and validation of the PINN model. Specifically, the model was trained using all input–output waveform pairs recorded at the three reference temperatures defined in the IEC 61869-1 [246] standard (-5°C , 20°C , and 40°C), and subsequently tested using the same signals measured at the additional temperatures listed above. This strategy allows assessing the model’s capability to generalize to unseen operating conditions and to interpolate the coil’s behavior across the entire temperature range.

5.4.3.3 Physics-Informed Learning

The implemented model leverages a PINN (as described in Section 2.3.3.4) to predict the output voltage of a RC under varying temperature conditions. The framework is structured in two main stages:

1. **Stage 1: Inverse PINN – Learning Discrete M :** In the first stage, an inverse PINN is trained to estimate the temperature-dependent mutual inductance M , based on the governing relationship 5.6.

Training data consists of measurements collected at three reference temperatures: -5°C , 20°C , and 40°C . For each temperature, the training samples consist of time-domain waveform pairs $\{t, I_p(t), u_m(t)\}$, where t denotes the continuous time variable, $I_p(t)$ is the measured primary current waveform, and $u_m(t)$ is the corresponding measured RC output voltage. The time dimension is treated explicitly: the PINN operates on individual time samples extracted from the acquired waveforms, rather than on aggregated features or frequency-domain representations. Automatic differentiation is used to compute $\frac{dI_p(t)}{dt}$ directly with respect to the time input t , ensuring consistency with the governing physical equation.

For each temperature, the network minimizes the following composite loss:

$$\text{Loss}_{\text{data}} = \|u_{\text{NN}}(t) - u_m(t)\|^2, \quad (5.7)$$

$$\text{Loss}_{\text{physics}} = \left\| u_{\text{NN}}(t) + M \frac{dI_p(t)}{dt} \cdot x \right\|^2, \quad (5.8)$$

where $u_{\text{NN}}(t)$ is the model-predicted output, $u_m(t)$ is the measured coil output, and x denotes the network input. The inverse PINN learns discrete values of M corresponding to the known temperatures.

In the inverse PINN formulation, the network inputs are the time variable t and the measured primary current $I_p(t)$, while the network output is the predicted secondary voltage $u_{\text{NN}}(t)$.

To enable generalization across unseen temperatures, a secondary feedforward neural network, denoted as $MNet$, is trained to model the continuous mapping $T \mapsto M$. Once trained, $MNet$ can interpolate and extrapolate M within a reasonable thermal range and is embedded into the subsequent forward PINN as a dynamic component.

2. **Stage 2: Forward PINN – Predicting Output Voltage $u_s(t)$:** In the second stage, a forward PINN is trained to predict the RC output voltage $u_s(t)$ based on three inputs: time t , input current $I_p(t)$, and temperature T . The learned M from $MNet$ is integrated internally. The total loss optimized by the forward PINN is:

$$\text{Loss}_{\text{total}} = \lambda_{\text{data}} \overline{\text{Loss}}_{\text{data}} + \lambda_{\text{physics}} \overline{\text{Loss}}_{\text{physics}}, \quad (5.9)$$

where λ_{data} and λ_{physics} are weighting coefficients used to balance data fidelity and physical consistency.

This two-stage approach ensures that the network remains consistent with both physical laws and empirical data, while dynamically adapting to variations depending on temperature in sensor behavior via the learned relationship.

5.4.3.4 Results

The prediction performance of the PINN model was assessed in comparing the output signals acquired from RCs and the estimated ones. Model accuracy was evaluated in terms of ratio error (ε) and phase displacement ($\Delta\varphi$) (Equations 5.4 and 5.5).

Results with Sinusoidal Inputs at 50 Hz

The results for device R1 are presented in Figure 5.16. The estimated ratio error and phase displacement are compared against the reference values for several current amplitudes and temperatures. Note that, for the sake of brevity, only few temperature levels are displayed.

The ratio error exhibits a strong dependence on current amplitude. At lower amplitudes (e.g., 50 A), errors range between approximately 11 % and 13 %. As the amplitude increases to 100 A, the error progressively reduces, reaching values between 7 % and 9 %. Temperature has a more limited impact, although slightly higher errors are observed at the lowest test point (-15°C), suggesting minor thermal sensitivity in the model's predictions.

Phase displacement remains consistently low, ranging between 0.08 mrad and 0.13 mrad across all tested conditions. A slight increase is noted with rising temperature, whereas the effect of amplitude on phase accuracy is negligible. These results indicate strong generalization of the model in capturing phase behavior, with ratio accuracy being more amplitude-sensitive. Overall, the model shows robust performance in estimating both ratio error and phase displacement, consistent with the preliminary characterization.

Figure 5.17 presents the results for device R2. Compared to R1, the model exhibits higher ratio error, particularly at lower amplitudes, where values reach up to 19.6 %. At 100 A, the error reduces to a range between 12.8 % and 14.5 %, confirming an improved accuracy trend with increasing amplitude.

Phase displacement is also generally larger than that observed for R1, especially at 50 A and high temperatures (e.g., 0.38 mrad at 55 °C). The error significantly decreases at higher currents and lower temperatures, reaching a minimum of 0.02 mrad at 100 A and −15 °C. These findings suggest that while the model maintains acceptable accuracy, its generalization ability is more sensitive to sensor-specific parameters under thermal and low-signal conditions.

The error trends for R3 are reported in Figure 5.18. The ratio error follows the same amplitude-dependent trend observed in R1 and R2. At 5 A, values range from 16.1 % to 18.4 %, while at 100 A, they drop to approximately 12.2-13.6 %.

As shown in the phase displacement results, the value remains relatively small, ranging from −0.16 mrad to 0.07 mrad. Some sign changes are observed, with negative values being more frequent at lower amplitudes and temperatures, suggesting a mild phase lag in the model predictions. Despite these minor deviations, phase prediction stays within acceptable bounds, indicating good stability across the third device configuration.

Across all three devices, the PINN model exhibits consistent behavior. The ratio error decreases with increasing current amplitude, while phase displacement remains low across all test conditions. Slightly higher errors are observed at low currents, mainly due to the reduced voltage output of the coils at these amplitudes.

R1, designed for higher accuracy and with balanced electrical characteristics, shows the most stable performance. R2, with its higher winding resistance and wide current range, produces larger errors. R3, characterized by an oval cross-section and low resistance, performs well in phase prediction but shows greater variability in ratio error under low-current conditions.

These results highlight the influence of coil design and electrical parameters on model performance. Importantly, the estimation errors obtained from the model are significantly lower than the deviations induced by temperature alone in practical RC operation –

variations which can exceed 100 % of the accuracy class limits. This confirms the model's potential for improving current sensing reliability under realistic conditions.

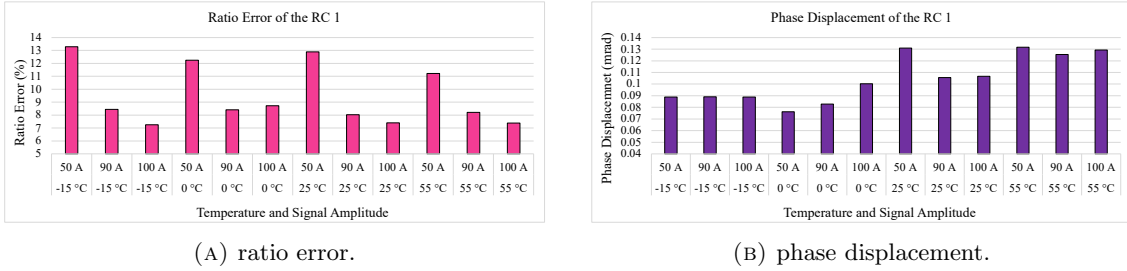


FIGURE 5.16: PINN estimation model (A) ratio error and (B) phase displacement values for R1 with different testing signals and temperatures.

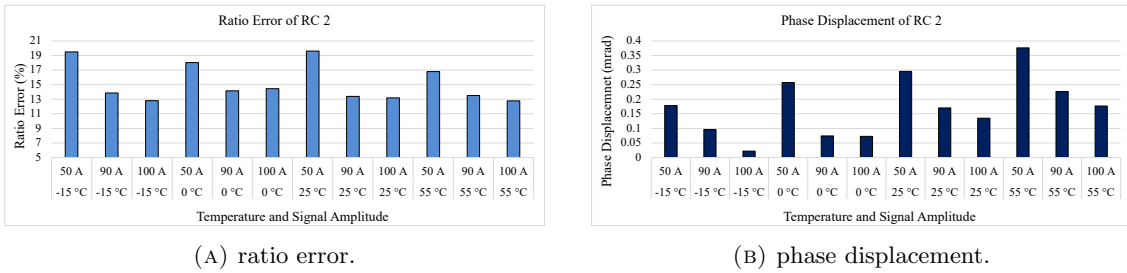


FIGURE 5.17: PINN estimation model (A) ratio error and (B) phase displacement values for R2 with different testing signals and temperatures.

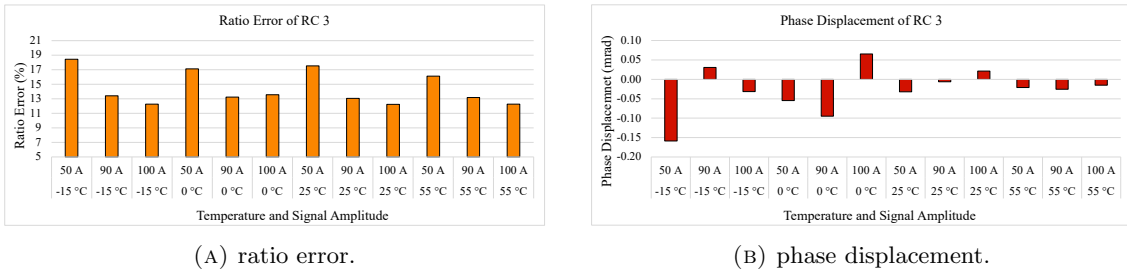


FIGURE 5.18: PINN estimation model (A) ratio error and (B) phase displacement values for R3 with different testing signals and temperatures.

Results with Single-Frequency Sinusoid Inputs at Multiple Frequencies

Figures 5.19, 5.20, and 5.21 present the results obtained for R1, R2, and R3 when tested with sinusoidal input signals at 150 Hz and different amplitudes (10 A and 20 A) across the full temperature range from -15°C to 55°C . Each figure reports both ratio error and phase displacement between the reference and the PINN-predicted signals.

For R1, the ratio error remains low across all operating conditions, ranging from approximately 3.17 % to 4.18 %. The model thus maintains good accuracy even at higher frequencies and reduced amplitudes. Slight variations are observed with temperature and amplitude: lower current levels (10 A) generally exhibit marginally higher errors

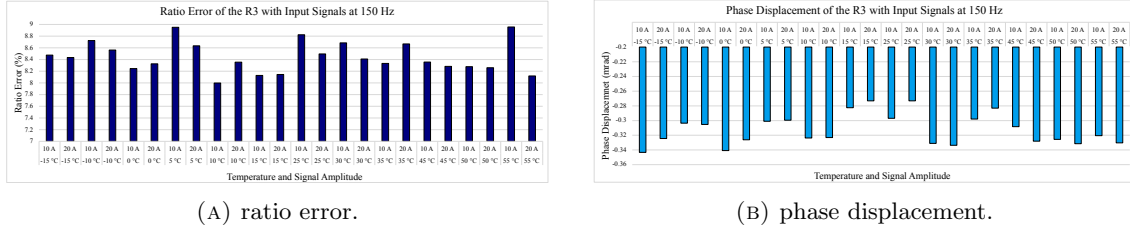


FIGURE 5.21: PINN estimation model (A) ratio error and (B) phase displacement values for R3 with input signals at 150 Hz and different temperatures.

To further validate the model under more demanding conditions, Figures 5.22, 5.23, and 5.24 respectively present the results for R1, R2, and R3, for sinusoidal inputs at 1250 Hz, again for amplitudes of 10 A and 20 A and the full temperature range.

For R1, the ratio error remains between 4.21 % and 5.32 %, with minimal dependence on temperature or amplitude. As previously observed, 10 A signals tend to yield slightly higher errors. The phase displacement remains low, between 3.83 mrad and 4.49 mrad, showing only a slight increase at intermediate temperatures (25-35 °C), while amplitude effects are negligible.

For R2, the ratio error ranges from approximately 9.18 % to 10.48 %, with small oscillations (within ± 0.5 %) across temperatures from -15 °C to 55 °C. The phase displacement values are higher than for R1, between 14.6 mrad and 17.8 mrad, showing a mild upward trend at higher temperatures. Amplitude influence remains negligible, with 10 A and 20 A curves nearly overlapping.

For R3, the ratio error remains stable and low, between 9.19 % and 10.31 %, with no clear temperature trend. The phase displacement remains negative, between -0.82 mrad and -0.39 mrad, again indicating a slight phase lag. Variations with temperature are minimal, and amplitude effects are negligible.

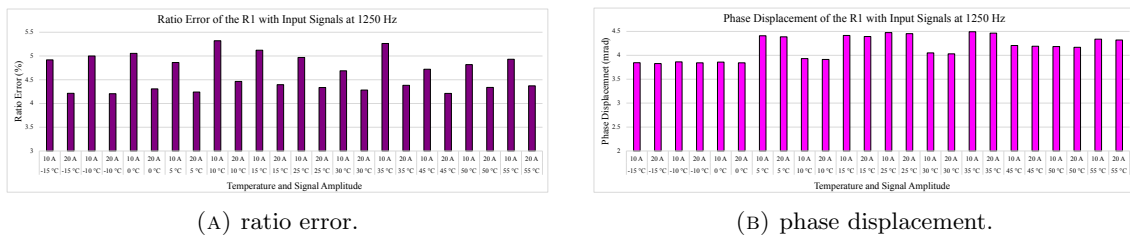


FIGURE 5.22: PINN estimation model (A) ratio error and (B) phase displacement values for R1 with input signals at 1250 Hz and different temperatures.

Finally, Figure 5.25 present the heatmaps of the ratio error and phase displacement for all the considered frequencies and operating temperatures, respectively. The ratio

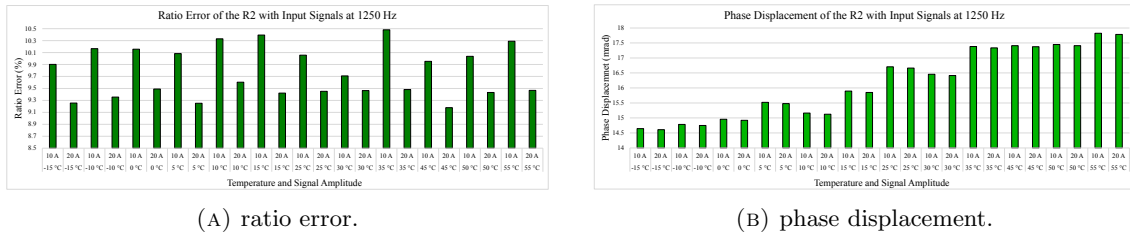


FIGURE 5.23: PINN estimation model (A) ratio error and (B) phase displacement values for R2 with input signals at 1250 Hz and different temperatures.

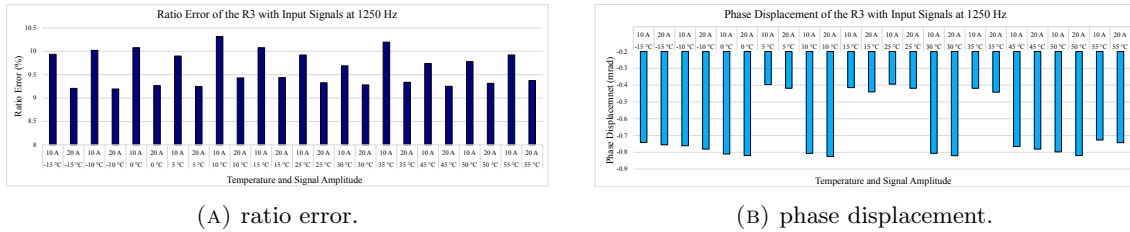


FIGURE 5.24: PINN estimation model (A) ratio error and (B) phase displacement values for R3 with input signals at 1250 Hz and different temperatures.

error distribution demonstrates that the proposed model maintains stable and accurate amplitude estimations across most operating conditions, with only a slight increase in error at higher temperatures and frequencies. This trend indicates a good generalization capability of the model within the typical operating range. Likewise, the phase displacement map exhibits low and consistent phase estimation errors across the central region of the operating domain, with moderate degradation becoming evident at high frequency operating conditions. From a metrological standpoint, the magnitude and stability of both errors indicate that the proposed PINN exhibits consistent behavior, ensuring the level of reliability expected for practical current measurement applications.

Overall, the experimental results confirm that the proposed PINN-based framework can accurately reproduce the dynamic behavior of RCs across a broad range of frequencies (150-1250 Hz) and temperatures (-15°C to 55°C), maintaining low ratio error and minimal phase distortion. This demonstrates the model's robustness, precision, and applicability for broadband current measurement in realistic and thermally variable environments.

Results with Sinc-Type Inputs

Figures 5.26 report the ratio error and phase displacement values for R1, with the PINN model tested on sinc-type signals of 50 A amplitude at $-15\text{ }^{\circ}\text{C}$. The ratio error decreases with increasing frequency, from approximately 5.17 % at 150 Hz to 2.61 % at

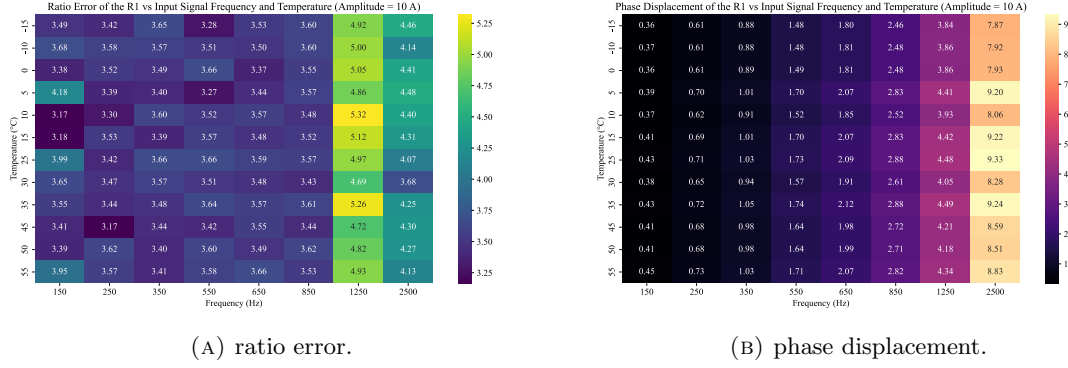


FIGURE 5.25: PINN estimation model (A) ratio error and (B) phase displacement values for R1 with different frequencies of input signal (100 A amplitude) and different temperatures.

3050 Hz, while the phase displacement increases from 0.57 mrad to 9.31 mrad, indicating a slight lag of the output signal relative to the input at higher frequencies.

Figures 5.27 show the results for R2 under the same conditions. The ratio error decreases from 10.40 % to 7.53 % across the frequency range, while the phase error increases from 1.29 mrad to 36.42 mrad. These results highlight the frequency-dependent behavior of the coil and the PINN model, with minor overestimation at lower frequencies and larger phase lag at higher harmonics.

Figures 5.28 illustrate the outcomes for R3 at 50 A and -15°C . The ratio error gradually declines from 10.24 % to 7.58 %, and the phase error remains negative, ranging from -0.01 mrad to -1.87 mrad. R3 is the only coil exhibiting consistently negative phase deviations, indicating a slight lag of the output signal.

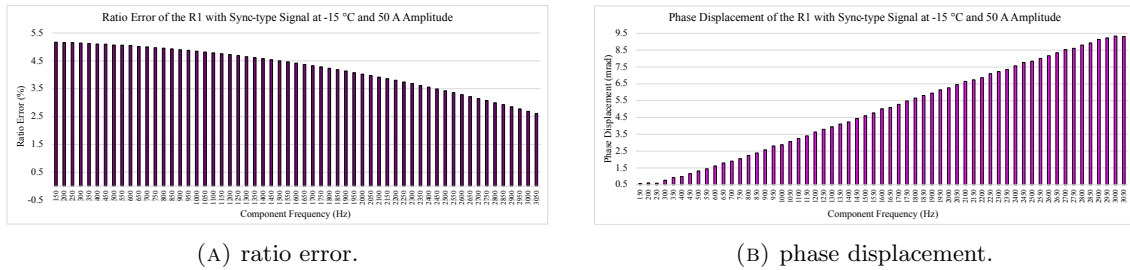


FIGURE 5.26: PINN estimation model (A) ratio error and (B) phase displacement values for R1 with sinc-type input signals at -15°C and 50 A.

For higher amplitude sinc signals (100 A), Figures 5.29, 5.30, and 5.31 present the corresponding results. Only a representative subset of harmonic components and temperatures are shown for clarity. Overall, the ratio error decreases with frequency, while

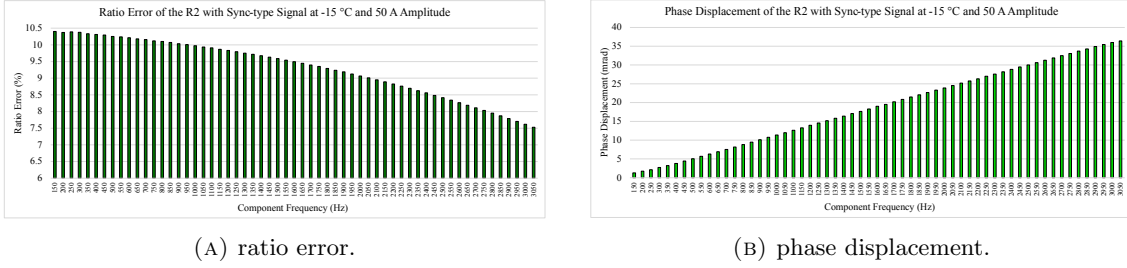


FIGURE 5.27: PINN estimation model (A) ratio error and (B) phase displacement values for R2 with sinc-type input signals at -15°C and 50 A.

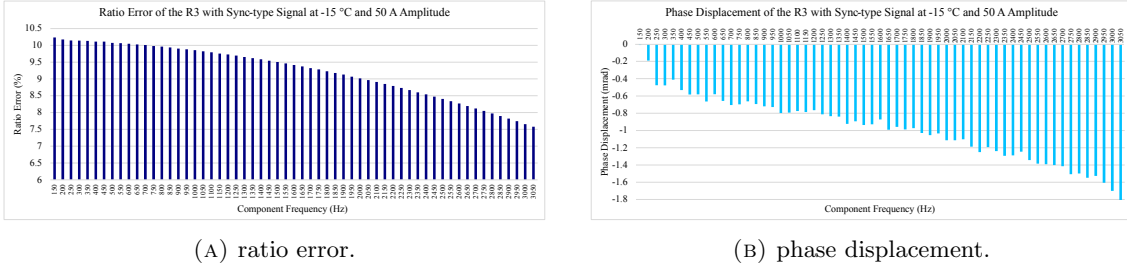


FIGURE 5.28: PINN estimation model (A) ratio error and (B) phase displacement values for R3 with sinc-type input signals at -15°C and 50 A.

the phase error increases, showing minimal influence from temperature. R3 again shows consistently negative phase deviations, although small in magnitude.

These results confirm that the PINN model accurately predicts both amplitude and phase response of all three RCs across a broad frequency range. The ratio error generally decreases at higher frequencies, improving amplitude estimation, whereas the phase error increases, reflecting a growing lag between predicted and actual outputs. Notably, R3 is the only coil with systematically negative phase errors. These findings demonstrate the model's ability to generalize from sinusoidal training signals to broadband, non-sinusoidal excitations, highlighting its potential for power quality monitoring and advanced diagnostics.

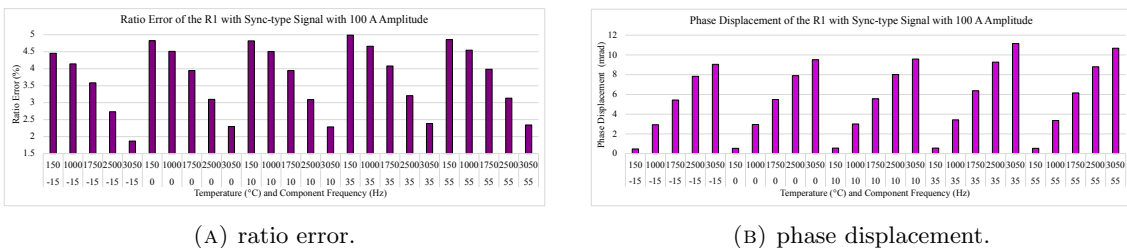
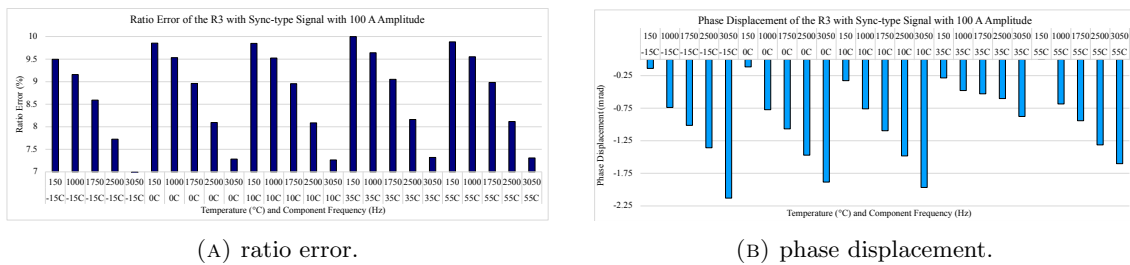
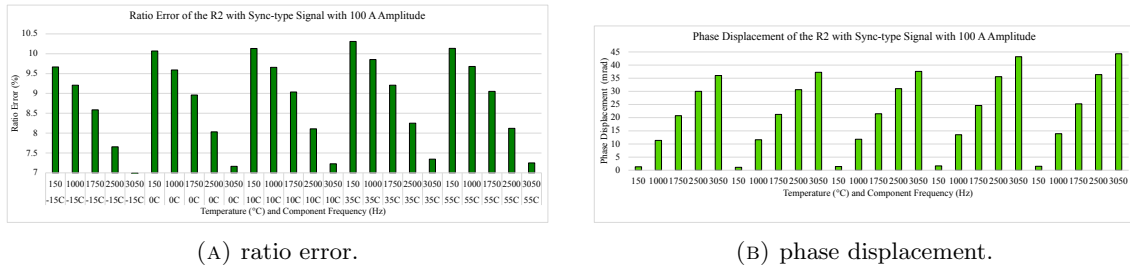


FIGURE 5.29: PINN estimation model (A) ratio error and (B) phase displacement values for R1 with sinc-type input signals at different temperatures and frequencies with 100 A amplitude.



The results showed that the model can accurately estimate both the ratio error and the phase displacement over a wide temperature range. It was able to generalize effectively across different coil designs, adapting to the specific electrical and geometrical characteristics of each device. Prediction errors were consistently smaller than the natural

variability of the measurements caused by temperature, confirming the model's ability to compensate for thermal effects and improve overall measurement accuracy.

From a metrological perspective, the proposed approach contributes to reducing the uncertainty associated with temperature influences, increasing the confidence and traceability of current measurements. More generally, the study demonstrates that integrating physical knowledge with data-driven modeling can effectively enhance the stability and reliability of measurement systems.

Finally, these findings highlight the potential of this method for developing new generations of ITs with self-compensation capabilities, supporting more accurate and robust monitoring of modern power systems, including complex scenarios relevant for power quality evaluation.

5.4.4 Microcontroller-Based Digital Twin for Embedded Signal Verification

Recent advances in smart grid monitoring highlight the central role of sensor systems in enabling high-performance, cost-effective, and intelligent diagnostics. Several contributions in the literature showcase how signal acquisition and embedded intelligence can enhance fault detection, self-diagnostics, and virtualization in power systems.

For instance, a novel method has been proposed for estimating switchgear bushing temperatures using built-in capacitive couplers, enabling fault detection through simulation data and virtual testbeds, while minimizing additional hardware requirements [268]. An ANN-based technique for fault detection in RC systems has also been developed, combining microcontroller-based signal acquisition with a fully CNN for robust self-diagnostics in HV scenarios [269].

The next evolutionary step involves the development of a DT for the smart grid. A DT facilitates real-time interaction between the physical and virtual layers of the system, enabling enhanced monitoring, diagnostics, and predictive maintenance. A DT-based protective relay for transformers has been introduced, comparing real-time measurements to detect internal faults [270]. Similarly, a DT system integrated with DL has been demonstrated for power transformer fault detection [271]. DT applications have further been extended to intelligent substations, enabling 3D visualization and health diagnostics [272].

Finally, DT integration in digital substations has been explored, leveraging communication protocols for virtual testing [273]. These efforts highlight how embedded systems can play a critical role in real-time data acquisition and feedback to the DT.

Building upon recent progress in microcontroller and sensor technologies, this research also aims to achieve three key goals. First, it implements a simplified and computationally efficient RC model within an embedded platform for real-time signal estimation. Second, it introduces a custom measurement setup designed to assess the performance of the proposed solution. Finally, by combining local measurement and model-based estimation, the system enhances monitoring capabilities and supports early detection of anomalies in both the network and the sensing hardware.

5.4.4.1 Microcontroller Implementation

To implement the RC model on a microcontroller, the continuous-time transfer function formulated in [274] was first converted into a discrete-time form suitable for real-time digital computation. The original Laplace-domain transfer function is defined as:

$$H(s) = \frac{Ks}{Xs^2 + Ys + W}, \quad (5.10)$$

where K , X , Y , and W are constants derived from the coil's physical parameters. The transfer function models the dynamic relationship between the primary current and the output voltage of the RC. Through a discretization process based on the system's sampling frequency, the continuous-time model was transformed into a Z-domain rational function. This transformation yielded a second-order linear difference equation:

$$y[n] = a_1y[n-1] + a_2y[n-2] + b_1x[n-1] + b_2x[n-2], \quad (5.11)$$

where $y[n]$ represents the estimated output at the current time step, while $x[n-1]$ and $x[n-2]$ are the previous input samples. The coefficients a_1 , a_2 , b_1 , and b_2 are derived from the discrete Z-transform. This recursive formulation is ideal for real-time embedded applications, as it requires only basic arithmetic operations on a small set of past samples.

The model was deployed on a Microchip dsPIC33CK Curiosity Development Board [275], which integrates a dsPIC33 family digital signal controller optimized for real-time control and signal processing tasks. This platform was selected for its high-performance Digital Signal Processor (DSP) core, integrated 12-bit ADCs, and efficient support for both fixed- and floating-point operations.

The input signal – corresponding to the time derivative of the current sensed by the RC – is sampled at a fixed frequency and processed in real-time using Equation 5.11. Two synchronised 12-bit ADC channels acquire the necessary analog signals, providing high-resolution measurements. The recursive model requires only four multiplications and four additions per sample, enabling fast and reliable execution within each sampling interval.

The sampling frequency was chosen to ensure a trade-off between signal fidelity and computational headroom. The selected rate allows accurate signal reconstruction while leaving sufficient processing time to compute the output without overloading the system.

In terms of resource usage, the implementation is lightweight: it occupies approximately 7 % of the available FLASH memory (6229 words out of 89728) and about 43 % of RAM (10618 bytes out of 13958). These figures confirm that the solution is scalable and leaves space for future enhancements.

Although full integration with a supervisory DT platform lies outside the scope of this study, the implemented embedded model establishes a solid foundation for future developments. The microcontroller can transmit both measured and estimated outputs via standard communication protocols (e.g., Ethernet or wireless) to upper-layer systems responsible for data aggregation, global analytics, fault diagnosis, and decision-making.

5.4.4.2 Microcontroller Preliminary Characterization

Before running the main tests, some preliminary characterizations were necessary to ensure the accuracy of the microcontroller-based system. The microcontroller features an analog filtering stage that adapts the RC's output to the $[0, 3.3]$ V input range. After conditioning, the signal is transformed into a voltage range of $[-1.5, 1.5]$ V.

To characterize the system's acquisition chain, a series of sinusoidal test signals with varying amplitudes and phase shifts were generated using the Fluke Calibration 6105A

calibrator [266]. The signals were injected into both analog input channels of the microcontroller. A schematic representation of the characterization setup is shown in Figure 5.32. For each test condition, 100 samples were acquired to ensure statistical reliability. The digital outputs from the ADCs were converted to voltages, and their amplitude and phase were estimated using FFT-based methods and a time-domain phase projection technique.

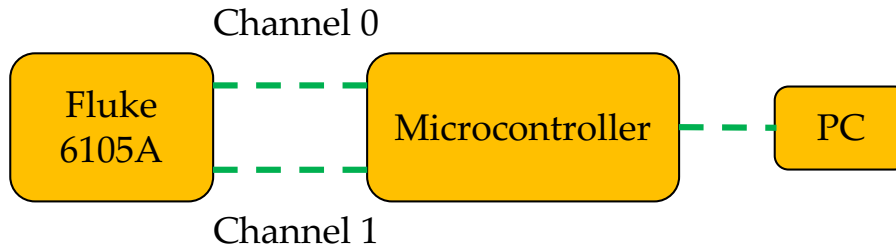


FIGURE 5.32: Simple schematic of the microcontroller characterization.

To quantify the system response, two gain correction factors, k_0 and k_1 , were computed as the ratio between the measured and reference signal amplitudes for each channel. These were calculated across five amplitude levels. Furthermore, the inter-channel phase shift $\Delta\varphi$ was estimated under three different phase settings.

In this characterization, the same reference sinusoidal voltage generated by the calibrator is applied to both analog input channels of the microcontroller. ADC 0 acquires the signal after the full analog conditioning chain (including attenuation and filtering applied to the RC output), whereas ADC 1 acquires the reference signal without attenuation, serving as a high-level reference channel. As a consequence, the measured amplitudes at ADC 0 are significantly lower than those at ADC 1.

For each test amplitude, the gain correction factors k_0 and k_1 are computed as the ratio between the reference signal amplitude and the corresponding amplitude estimated from the ADC measurements. These coefficients quantify the overall gain of each acquisition channel, including the analog front-end and the ADC scaling. The reported mean values are obtained by averaging the gain estimates over the five tested amplitude levels, confirming the linearity and repeatability of the acquisition chain.

The results are reported in Tables 5.37 and 5.38, confirming the correct functionality of the analog front-end and the ADC acquisition chain, and ensuring a reliable basis for subsequent tests.

TABLE 5.37: Amplitude characterization results.

Reference Amplitude (V)	ADC 0 (V)	ADC 1 (V)	Mean k_0 (-)	Mean k_1 (-)
0.035	0.004	1.42	1.42	34.02
0.070	0.010	1.42	1.42	34.04
0.200	0.015	1.41	1.41	34.33
0.350	0.020	1.42	1.42	34.24
0.500	0.025	1.42	1.42	34.26

TABLE 5.38: Phase characterization results.

Phase Shift Setting (°)	Mean $\Delta\varphi_0$ (°)
0	0.68
30	30.67
60	59.79

5.4.4.3 Rogowski Coils Preliminary Characterization

To validate the proposed measurement and estimation procedure, a preliminary experimental characterization of the RCs under test was performed using the setup shown in Figure 5.33. Although it is not standard practice to characterize each device installed in the field, this step was necessary to ensure the accuracy and repeatability of the methodology during validation.

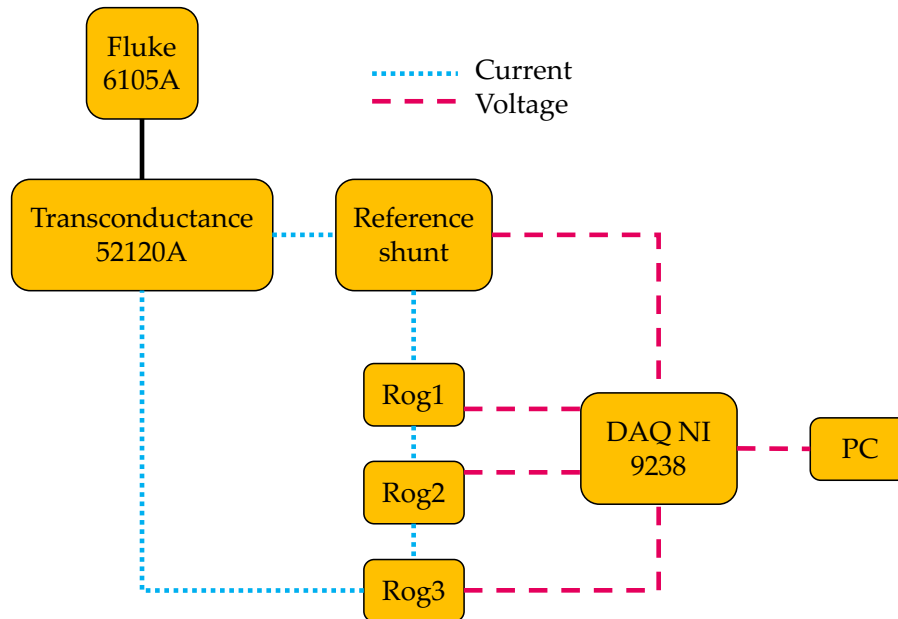


FIGURE 5.33: Simple schematic of the RC characterization.

The tests were carried out on three commercial RCs, denoted as R1, R2, and R3, which parameters are summarized in Table 5.34.

The characterization procedure involved injecting sinusoidal currents at 60, 90, and 120 A, all at 50 Hz, using a Fluke Calibration 6105A calibrator [266] in combination with its 52120A transconductance amplifier [267]. The true current values were measured using a precision shunt resistor with an uncertainty of 2×10^{-6} , serving as a reference. The RC outputs were compared against this reference under three phase shift conditions: 0° , 30° , and 60° .

For each test condition, a 1 second acquisition window was recorded at 50 kSa/s using a NI-9238 DAQ module [276]. Each configuration was repeated 100 times to ensure statistical reliability and to assess repeatability.

The raw measurement results (i.e., the current values detected by each coil) are reported in Table 5.39. From these, the actual transformation ratios were calculated and expressed as a percentage of the rated value. These values are presented in Table 5.40.

The results confirm the stability and reliability of the RCs across a range of operating conditions, supporting their use in the subsequent validation and deployment of the estimation algorithm.

TABLE 5.39: RCs characterization results.

Current (A)	Phase ($^\circ$)	Shunt (A)	R1 (A)	R2 (A)	R3 (A)
60	0	61.37	60.15	59.22	58.40
60	30	61.43	60.15	59.22	58.39
60	60	61.42	60.15	59.22	58.39
90	0	92.08	90.23	88.83	87.59
90	30	92.19	90.23	88.84	87.60
90	60	92.18	90.23	88.83	87.59
120	0	122.83	120.31	118.45	116.78
120	30	122.90	120.31	118.45	116.79
120	60	122.87	120.31	118.45	116.78

5.4.4.4 Results

To validate the proposed modeling approach, each RC was tested individually in a controlled laboratory environment. A sinusoidal current at 50 Hz was injected using the Fluke 6105A calibrator [266] and its transconductance [267], and both the input and output signals were acquired using the microcontroller-based data acquisition system.

TABLE 5.40: Actual transformation ratio obtained from the characterization.

Current (A)	Phase (°)	TR-R1 (%)	TR-R2 (%)	TR-R3 (%)
60	0	98.01	96.50	95.15
60	30	97.93	96.41	95.06
60	60	97.93	96.41	95.06
90	0	97.99	96.47	95.12
90	30	97.87	96.36	95.02
90	60	97.88	96.37	95.02
120	0	97.95	96.43	95.07
120	30	97.89	96.38	95.03
120	60	97.91	96.40	95.04

Results have been evaluated in terms of ratio error (ε) and phase displacement ($\Delta\varphi$) between the measured and the estimated RC output (Equations 5.4 and 5.5). Channel 0 of the microcontroller captured the reference input current signal, while Channel 1 measured the corresponding voltage output from the RC under test. A dual ± 5 V power supply was used to power the acquisition platform. A photograph of the experimental setup is provided in Figure 5.34.

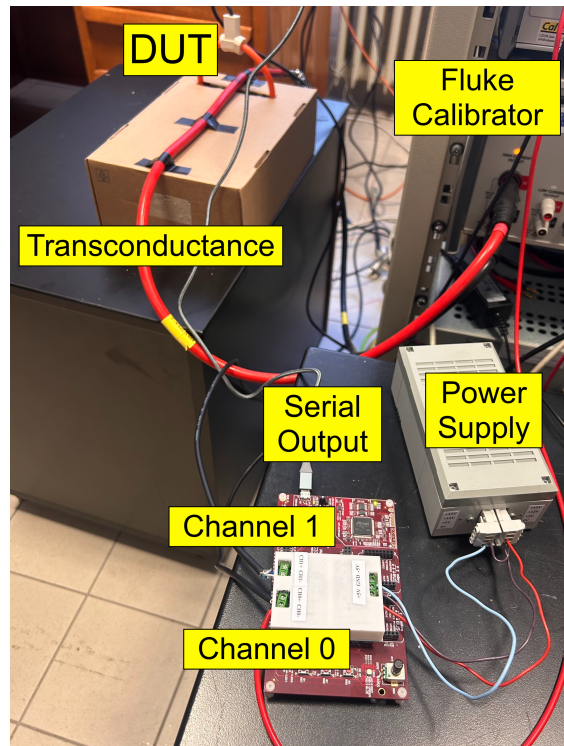


FIGURE 5.34: Measurement setup to validate the proposed procedure.

During testing, three amplitude levels (60 A, 90 A and 120 A) were applied, each with a fixed phase of 0° , based on earlier results of the phase sensitivity. For each test, 10

complete cycles of the waveform were recorded, sampled at 128 samples per cycle (i.e., 6.4 kSa/s).

The microcontroller concurrently computed the estimated RC output based on the discrete-time model in Equation 5.11, using the input signal as a reference. All acquired and computed data – including raw input, measured output, and model-based estimate – were transmitted via serial communication and analyzed offline using a Python script.

Table 5.41 presents the results for each RC across the three current levels. The RMS voltage values from Channel 1 were corrected using previously determined transformation ratios, ensuring accurate amplitude comparison.

TABLE 5.41: Results from the implementation of the proposed procedure.

Current (A)	RC	ε (%)	$\Delta\varphi$ (mrad)
60	R1	−1.81	−0.46
60	R2	−2.18	0.64
60	R3	−2.40	−0.57
90	R1	−1.77	−0.62
90	R2	−2.15	0.48
90	R3	−2.38	−0.28
120	R1	−1.87	−0.85
120	R2	−2.25	0.29
120	R3	−2.47	1.29

Overall, the results confirm the effectiveness of the proposed modeling approach. The ratio error remains within ± 2.5 % in all conditions, with coil R1 consistently performing slightly better than R2 and R3. Similarly, phase displacement values stay well below ± 1 mrad, ensuring that the model reliably preserves the phase characteristics of the signal. This is particularly relevant for power measurement and protection applications.

These findings demonstrate that the discrete-time model implemented on the microcontroller is capable of accurately reproducing both the amplitude and phase behavior of the tested RCs under real-world conditions, thereby validating the proposed method.

5.4.5 Final Remarks

This research has explored two complementary strategies to enhance the performance and intelligence of RC-based measurement systems. On one hand, a PINN was developed to address one of the most critical limitations of LPITs: their sensitivity to temperature

variations. By leveraging standardized experimental data and DL techniques, the model accurately predicted RCs output waveforms across varying thermal conditions. The results confirmed its ability to generalize across different RCs designs while adapting to individual geometric and electrical characteristics. Most importantly, the predictive accuracy consistently outperformed the intrinsic measurement variability caused by temperature, demonstrating the model's practical effectiveness.

On the other hand, a lightweight microcontroller-based implementation of a DT was presented. By embedding a discrete-time analytical model directly onto the embedded hardware, the system enabled local, real-time estimation of the RC output, offering a means to verify and enhance current measurements. Experimental validations across multiple devices and signal amplitudes confirmed the reliability and adaptability of the approach. Despite the resource constraints of the platform, the model-based estimation achieved low errors, underscoring the feasibility of integrating such verification mechanisms in edge-level applications.

Together, these results emphasize the potential of combining data-driven modeling and embedded DTs to enhance the accuracy, robustness, and intelligence of distributed current sensing. The approaches are complementary: while the DL model provides high-level compensation for environmental effects, the embedded DT supports real-time diagnostics and local autonomy. From an architectural perspective, the two solutions could also be conceived in a cascaded configuration, where the embedded DT performs local signal verification and pre-conditioning, and the PINN-based model operates at a higher level to compensate for temperature-dependent effects. Although such a cascaded implementation is not explored in this work, the presented results demonstrate that the two methodologies are inherently compatible and could be integrated within a unified multi-layer sensing framework.

5.5 Summary of AI and Metrology Integration

This chapter has explored the interaction between AI and metrology from multiple and complementary perspectives, positioning uncertainty as a central element of the analysis. Across the different application domains, uncertainty is modeled, propagated, or exploited

depending on the measurement objective, the available information, and the intended operational context of the AI system.

Rather than enforcing a single uncertainty paradigm, the chapter adopts a metrology-aware viewpoint in which AI models are interpreted as integral components of measurement systems. From this perspective, their reliability, validity, and applicability must be assessed using criteria consistent with measurement science.

Table 5.42 provides a structured synthesis of the case studies discussed in this chapter, highlighting how uncertainty in AI-based systems can be addressed at different levels, ranging from data acquisition and labeling to model inference and embedded deployment. In some applications, uncertainty is explicitly propagated to derive confidence information associated with individual predictions; in others, it is leveraged to enhance robustness, generalization capability, or real-time accuracy estimation.

Overall, the chapter demonstrates that AI can be rigorously integrated within a metrological framework, provided that uncertainty sources are explicitly identified and coherently managed. This metrology-aware perspective represents a key contribution of the thesis, bridging data-driven methodologies with metrology and laying the groundwork for trustworthy, deployable, and interpretable AI-based measurement systems.

TABLE 5.42: Summary of AI-based case studies in Chapter 5, highlighting metrological aspects and uncertainty handling strategies.

Application Domain	Dominant Uncertainty Source	AI-Metrology Strategy	Key Metrological Contribution
CJ fault prediction	Measurement uncertainty (temperature, current)	Input perturbation of CJ features in supervised ML classifiers	Demonstrated robustness of ML-based diagnosis to realistic sensor noise
	Knowledge uncertainty (fault thresholds)	Stochastic modeling of expert-defined fault thresholds	Sensitivity analysis of diagnostic decisions to threshold variability
SoC estimation	Measurement uncertainty (voltage, current)	Quasi-Monte Carlo propagation through ML-based SoC estimator	Prediction-level confidence intervals and metrological consistency assessment
	Model uncertainty	Stochastic inference via repeated realizations	Prediction-level uncertainty evaluation of ML outputs
Load forecasting in power substations	Measurement uncertainty (power)	Uncertainty-driven DA for DL models	Improved generalization under realistic measurement variability
	Model uncertainty	Ensemble learning and Monte Carlo dropout	Quantification of model-driven variability in DL forecasting
IT accuracy estimation	Environment-related uncertainty (temperature)	ANN-based regression with embedded implementation	Real-time estimation of IT accuracy parameters under field conditions
RC modeling	Environment- and operating-condition-related uncertainty (temperature, frequency)	PINNs model	Physically consistent modeling across extended operating ranges
	Implementation-related reliability aspects	Embedded DT on microcontroller	Local signal verification and support to measurement traceability

Chapter 6

Conclusion

This dissertation has explored the integration of AI, particularly ML, into modern electrical power systems with the aim of improving their intelligence, resilience, and metrological reliability. The research spans three technical domains: forecasting of electrical quantities, fault diagnostics in MV networks, and uncertainty-aware modeling in measurement systems. Although these areas may appear distinct at first glance, they are unified by a common objective: developing AI-based solutions that are accurate, robust, interpretable, and suitable for deployment in real-world, safety-critical energy infrastructures.

The first part of the research focuses on forecasting models capable of predicting key electrical parameters such as voltage, current, and frequency within distributed environments. These models leverage data from sensor networks deployed in smart buildings and hybrid power systems and are designed for real-time inference. This lays the foundation for advanced applications such as DTs, which can enable simulation and optimization of power systems in operation.

The second area addresses fault detection and predictive maintenance in MV distribution systems. By combining supervised and unsupervised learning techniques with historical data, the dissertation introduced methods to detect emerging faults, classify abnormal conditions, and assess component health through a reproducible HI. This approach supports condition-based maintenance strategies, enhancing system availability while reducing operational costs and environmental impact.

The third domain investigates the interplay between AI and metrology, with particular attention to the role of measurement uncertainty. Recognizing that uncertainty is

fundamental in ensuring trustworthy decisions, the research developed approaches to incorporate UQ into AI training and evaluation processes. Moreover, AI was used to model and compensate for metrological deviations in electrical sensors, with implementations validated on embedded devices for localized, real-time correction.

A significant contribution of this dissertation is demonstrating that these AI models can be effectively embedded on microcontroller-based platforms. This capability enables intelligence at the edge, reducing dependence on centralized infrastructures and addressing constraints related to latency, bandwidth, and reliability in DMSs. This transition from algorithmic development to practical deployment highlights the potential for scalable and autonomous measurement infrastructures within modern power grids.

A distinctive strength of this dissertation lies in its interdisciplinary approach. The research bridges AI, electrical engineering, and metrology to address complex challenges in modern power systems. This integration has been crucial for developing AI solutions that not only provide high accuracy but also comply with strict metrological requirements in critical energy infrastructures. By combining domain expertise from these areas, the dissertation ensures that AI models are grounded in physical realities, operational constraints, and rigorous measurement principles, leading to more reliable, transparent, and deployable systems.

The three research areas, while addressing different layers of power system operation – from monitoring and forecasting to diagnostics and metrological calibration – all converge towards a vision of electrical networks enhanced by intelligent, distributed, and reliable AI systems. This vision embraces the complexity of modern energy infrastructures, which require solutions that are interdisciplinary by nature.

Looking ahead, several avenues for future research emerge. Advancing the implementation of DTs for power systems remains a promising direction, especially in refining their integration with real-time forecasting and control. Additionally, further standardization of uncertainty-aware AI frameworks is necessary to ensure the broad adoption of trustworthy AI in critical infrastructure. Addressing challenges such as explainability, robustness against adversarial conditions, and the seamless integration of heterogeneous data sources will be key to the continued evolution of AI-enhanced power systems. Finally, expanding embedded AI implementations to increasingly constrained hardware platforms and diverse

sensor types will further push the boundaries of autonomous, real-time decision-making in smart grids.

In conclusion, this dissertation contributes to laying the groundwork for a new generation of AI-enhanced electrical power systems that are not only intelligent but also resilient, transparent, and metrologically sound, setting the stage for continued interdisciplinary innovation in this vital field.

Bibliography

- [1] E. Judson, O. Fitch-Roy, T. Pownall, R. Bray, H. Poulter, I. Soutar, and C. Mitchell. The centre cannot (always) hold: Examining pathways towards energy system decentralisation. *Renewable and Sustainable Energy Reviews*, 118, 2020. doi: 10.1016/j.rser.2019.109499. URL <https://doi.org/10.1016/j.rser.2019.109499>.
- [2] A. Dadkhah, B. Vahidi, M. Shafie-Khah, and J. P. S. Catalão. Power system flexibility improvement with a focus on demand response and wind power variability. *IET Renewable Power Generation*, 14(6):1095–1103, 2020. doi: 10.1049/iet-rpg.2019.0123. URL <https://doi.org/10.1049/iet-rpg.2019.0123>.
- [3] S. D. Panjaitan, J. Tjen, B. W. Sanjaya, F. T. P. Wigyantanto, and S. Khouw. A forecasting approach for iot-based energy and power quality monitoring in buildings. *IEEE Transactions on Automation Science and Engineering*, 2022. doi: 10.1109/TASE.2022.3171561. URL <https://doi.org/10.1109/TASE.2022.3171561>.
- [4] Abiodun E. Onile, Eduard Petlenkov, Yoash Levron, and Juri Belikov. Smart grid-based hybrid digital twins framework for demand side recommendation service provision in distributed power systems. *Future Generation Computer Systems*, 156:142–156, 2024. doi: 10.1016/j.future.2024.03.018. URL <https://doi.org/10.1016/j.future.2024.03.018>.
- [5] Juergen Doellner, Salvador Merino Cordoba, Francisco Guzman Navarro, Javier Martinez, Juan De Dios Lara, and Rafael Guzman. Towards concepts for climate and energy-oriented digital twins for buildings. In *Proceedings of the 28th International ACM Conference on 3D Web Technology (Web3D '23)*, pages 1–9, New York, NY, USA, 2023. Association for Computing Machinery. doi: 10.1145/3611314.3616066. URL <https://doi.org/10.1145/3611314.3616066>.

- [6] H. M. Hashemian and Wendell C. Bean. State-of-the-Art Predictive Maintenance Techniques. *IEEE Transactions on Instrumentation and Measurement*, 60(10):3480–3492, 2011. doi: 10.1109/TIM.2009.2036347. URL <https://doi.org/10.1109/TIM.2009.2036347>.
- [7] A. Mingotti, A. Ghaderi, G. Mazzanti, L. Peretto, R. Tinarelli, G. Valtorta, G. Amoroso, and S. Danesi. Low-Cost Monitoring Unit for MV Cable Joints Diagnostics. In *2018 IEEE 9th International Workshop on Applied Measurements for Power Systems (AMPS)*, pages 1–5, 2018. doi: 10.1109/AMPS.2018.8494848. URL <https://doi.org/10.1109/AMPS.2018.8494848>.
- [8] C. Chen, T. Xu, Z. Piao, and Y. Yuan. Power quality disturbances classification based on multi-class classification svm. In *PEITS 2009 - 2009 2nd Conference on Power Electronics and Intelligent Transportation System*, pages 1290–1294, 2009. doi: 10.1109/PEITS.2009.5407015. URL <https://doi.org/10.1109/PEITS.2009.5407015>.
- [9] Koji Kawakita. The role of condition monitoring and diagnostic technologies in electric energy systems to support the creation of a sustainable society. In *2022 9th International Conference on Condition Monitoring and Diagnosis (CMD)*, pages 1–4, 2022. doi: 10.23919/CMD54214.2022.9991728. URL <https://doi.org/10.23919/CMD54214.2022.9991728>.
- [10] Chao Zhang, Huaiyu Wang, Zhiming Wang, and Yudun Li. Active detection fault diagnosis and fault location technology for lvdc distribution networks. *International Journal of Electrical Power and Energy Systems*, 148, 2023. doi: 10.1016/j.ijepes.2022.108921. URL <https://doi.org/10.1016/j.ijepes.2022.108921>.
- [11] Rajarshi Dutta, Saikat Chakrabarti, and Ankush Sharma. Topology Tracking for Active Distribution Networks. *IEEE Transactions on Power Systems*, 36(4):2855–2865, 2021. doi: 10.1109/TPWRS.2020.3045855. URL <https://doi.org/10.1109/TPWRS.2020.3045855>.
- [12] Yuri Merizalde, Luis Hernández-Callejo, Oscar Duque-Perez, and Víctor Alonso-Gómez. Maintenance Models Applied to Wind Turbines. A Comprehensive Overview. *Energies*, 12, 1 2019. doi: 10.3390/en12020225. URL <https://doi.org/10.3390/en12020225>.

- [13] Sara Barja-Martinez, Mònica Aragüés-Peñalba, Íngrid Munné-Collado, Pau Lloret-Gallego, Eduard Bullich-Massagué, and Roberto Villafafila-Robles. Artificial intelligence techniques for enabling Big Data services in distribution networks: A review. *Renewable and Sustainable Energy Reviews*, 150:111459, 2021. ISSN 1364-0321. doi: 10.1016/j.rser.2021.111459. URL <https://doi.org/10.1016/j.rser.2021.111459>.
- [14] Emin Mammadov, Mostafa Farrokhhabadi, and Claudio A. Cañizares. AI-enabled Predictive Maintenance of Wind Generators. In *2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, pages 1–5, 2021. doi: 10.1109/ISGTEurope52324.2021.9640162. URL <https://doi.org/10.1109/ISGTEurope52324.2021.9640162>.
- [15] Matthew Deakin, Marta Vanin, Zhong Fan, and Dirk Van Hertem. Smart energy network digital twins: Findings from a uk-based demonstrator project. *International Journal of Electrical Power & Energy Systems*, 162:110302, 2024. ISSN 0142-0615. doi: 10.1016/j.ijepes.2024.110302. URL <https://doi.org/10.1016/j.ijepes.2024.110302>.
- [16] Y. Fu, Y. Huang, F. Hou, and K. Li. A brief review of digital twin in electric power industry. In *Proceedings of 2022 IEEE 5th International Electrical and Energy Conference, CIEEC 2022*, pages 2314–2318, 2022. doi: 10.1109/CIEEC54735.2022.9846081. URL <https://doi.org/10.1109/CIEEC54735.2022.9846081>.
- [17] Panayiotis Moutis and Omid Alizadeh-Mousavi. Digital twin of distribution power transformer for real-time monitoring of medium voltage from low voltage measurements. *IEEE Transactions on Power Delivery*, 36(4):1952–1963, 2021. doi: 10.1109/TPWRD.2020.3017355. URL <https://doi.org/10.1109/TPWRD.2020.3017355>.
- [18] C. Li, P. Lu, W. Zhu, H. Zhu, and X. Zhang. Intelligent monitoring platform and application for building energy using information based on digital twin. *Energies*, 16(19):6839, 2023. doi: 10.3390/en16196839. URL <https://doi.org/10.3390/en16196839>.
- [19] J. Feng, N. Zhao, and B. Wang. Digital twin-driven hot spot capture for power transformer interturn short-circuit fault diagnosis based on microthermal field.

- IEEE Transactions on Instrumentation and Measurement*, 73:1–13, 2024. doi: 10.1109/TIM.2024.3457945. URL <https://doi.org/10.1109/TIM.2024.3457945>.
- [20] Herman Zahid, Adil Zulfiqar, Muhammad Adnan, Sajid Iqbal, and Salah Eldeen Gasim Mohamed. A review on socio-technical transition pathway to european super smart grid: Trends, challenges and way forward via enabling technologies. *Results in Engineering*, 25:104155, 2025. doi: 10.1016/j.rineng.2025.104155. URL <https://doi.org/10.1016/j.rineng.2025.104155>.
- [21] Weixi Wang, Han Guo, Xiaoming Li, Shengjun Tang, Jizhe Xia, and Zhihan Lv. Deep learning for assessment of environmental satisfaction using bim big data in energy efficient building digital twins. *Sustainable Energy Technologies and Assessments*, 50:101897, 2022. doi: 10.1016/j.seta.2021.101897. URL <https://doi.org/10.1016/j.seta.2021.101897>.
- [22] JCGM. Evaluation of measurement data — guide to the expression of uncertainty in measurement, 2008. JCGM 100:2008.
- [23] P. Singh and L. K. Singh. Improved measurement accuracy in critical parameters of safety-critical systems with multisensor data fusion. *IEEE Transactions on Instrumentation and Measurement*, 70:1–8, 2021. doi: 10.1109/TIM.2021.3124852. URL <https://doi.org/10.1109/TIM.2021.3124852>.
- [24] H. M. Dipu Kabir, Abbas Khosravi, Mohammad Anwar Hosen, and Saeid Nahavandi. Neural network-based uncertainty quantification: A survey of methodologies and applications. *IEEE Access*, 6:36218–36234, 2018. doi: 10.1109/ACCESS.2018.2836917. URL <https://doi.org/10.1109/ACCESS.2018.2836917>.
- [25] S. Shirmohammadi. Measurement methodology: Visualizing uncertainty in machine learning-assisted measurements. *IEEE Instrumentation & Measurement Magazine*, 26(7):20–27, October 2023. doi: 10.1109/MIM.2023.10238361. URL <https://doi.org/10.1109/MIM.2023.10238361>.
- [26] Moloud Abdar, Farhad Pourpanah, Saddam Hussain, Dana Rezazadegan, Lei Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U. Rajendra Acharya, Vladimir Makarenekov, and Saeid Nahavandi. A review of uncertainty quantification in deep learning: Techniques, applications and challenges.

- Information Fusion*, 76:243–297, 2021. doi: 10.1016/j.inffus.2021.05.008. URL <https://doi.org/10.1016/j.inffus.2021.05.008>.
- [27] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1050–1059, New York, New York, USA, 20–22 Jun 2016. PMLR.
- [28] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. 2017. URL <https://arxiv.org/abs/1612.01474>.
- [29] Moloud Abdar, Abbas Khosravi, Sheikh Mohammed Shariful Islam, U. Rajendra Acharya, and Athanasios V. Vasilakos. The need for quantification of uncertainty in artificial intelligence for clinical data analysis: increasing the level of trust in the decision-making process. *IEEE Systems, Man, and Cybernetics Magazine*, 8(3): 28–40, 2022. doi: 10.1109/MSMC.2022.3150144. URL <https://doi.org/10.1109/MSMC.2022.3150144>.
- [30] S. Shirmohammadi, M. H. Amiri, and H. Al Osman. Measurement methodology: Uncertainty as a predictor of classification accuracy in machine learning-assisted measurements. *IEEE Instrumentation & Measurement Magazine*, 27(7):37–45, October 2024. doi: 10.1109/MIM.2024.10700740. URL <https://doi.org/10.1109/MIM.2024.10700740>.
- [31] Khumbulani Sithole. Employing internet of things (iot) devices for monitoring and controlling energy management systems-a review. *Journal of Electrical Systems*, 20: 753–760, 11 2024. doi: 10.52783/jes.7273. URL <https://doi.org/10.52783/jes.7273>.
- [32] Aditya Joshi, Skieler Capezza, Ahmad Alhaji, and Mo-Yuen Chow. Survey on ai and machine learning techniques for microgrid energy management systems. *IEEE/CAA Journal of Automatica Sinica*, 10(7):1513–1529, 2023. doi: 10.1109/JAS.2023.123657. URL <https://doi.org/10.1109/JAS.2023.123657>.
- [33] Ibrahim Alhamrouni, Nor Hidayah Abdul Kahar, Mohaned Salem, Mahmood Swadi, Younes Zahroui, Dheyaa Jasim Kadhim, Faisal A. Mohamed, and Mohammad

- Alhuyi Nazari. A comprehensive review on the role of artificial intelligence in power system stability, control, and protection: Insights and future directions. *Applied Sciences*, 14(14), 2024. ISSN 2076-3417. doi: 10.3390/app14146214. URL <https://doi.org/10.3390/app14146214>.
- [34] Felipe Henao, Rick Edgell, Anurag Sharma, Kalp Parmar, Matthew Cook, James May, Cameron Faulkner, Xiang Zhang, Molly O’Keefe, Kameshwar Poolla, and Pravin Varaiya. Ai in power systems: a systematic review of key matters of concern. *Energy Informatics*, 8:76, 2025. doi: 10.1186/s42162-025-00529-1. URL <https://doi.org/10.1186/s42162-025-00529-1>.
- [35] R. Martinez. Artificial intelligence: Distinguishing between types & definitions. *Nevada Law Journal*, 19:1015–1042, 2019.
- [36] Vitor Cerqueira, Luís Torgo, and Carlos Soares. A case study comparing machine learning with statistical methods for time series forecasting: size matters. *Journal of Intelligent Information Systems*, 59:415–433, 2022. doi: 10.1007/s10844-022-00713-9. URL <https://doi.org/10.1007/s10844-022-00713-9>.
- [37] F. Leite Coelho da Silva, K. da Costa, P. Canas Rodrigues, R. Salas, and J. L. López-Gonzales. Statistical and artificial neural networks models for electricity consumption forecasting in the brazilian industrial sector. *Energies*, 15(2):588, 2022. doi: 10.3390/en15020588. URL <https://doi.org/10.3390/en15020588>.
- [38] Partha Pratim Ray. A review on tinyml: State-of-the-art and prospects. *Journal of King Saud University - Computer and Information Sciences*, 34(4):1595–1623, 2022. ISSN 1319-1578. doi: 10.1016/j.jksuci.2021.11.019. URL <https://doi.org/10.1016/j.jksuci.2021.11.019>.
- [39] M. El Mrabet, K. El Makkaoui, and A. Faize. Supervised machine learning: A survey. In *Proceedings of the 2021 4th International Conference on Advanced Communication Technologies and Networking (CommNet)*, pages 1–10, 2021. doi: 10.1109/CommNet52204.2021.9641998. URL <https://doi.org/10.1109/CommNet52204.2021.9641998>.
- [40] Srivatsa Joshi, Vishwas Kumar, Vishaka Venkataramanan, and Kaliprasad C S. A review on neural networks and its applications. *Journal of Computer Technology*

- E Applications*, 14:2023, 09 2023. doi: 10.37591/jocta.v14i2.1062. URL <https://doi.org/10.37591/jocta.v14i2.1062>.
- [41] Benjamin Lindemann, Timo Müller, Hannes Vietz, Nasser Jazdi, and Michael Weyrich. A survey on long short-term memory networks for time series prediction. *Procedia CIRP*, 99:650–655, 05 2021. doi: 10.1016/j.procir.2021.03.088. URL <https://doi.org/10.1016/j.procir.2021.03.088>.
- [42] Josh Patterson and Adam Gibson. *Deep Learning: A Practitioner’s Approach*. O’Reilly Media, Inc, 1st edition, 2017.
- [43] Benyamin Ghogh and Ali Ghodsi. Recurrent neural networks and long short-term memory networks: Tutorial and survey, 2023. URL <https://arxiv.org/abs/2304.11461>.
- [44] Stefano Feraco, Pier Giuseppe Anselma, Angelo Bonfitto, and Phil Kollmeyer. Robust data-driven battery state of charge estimation for hybrid electric vehicles. *SAE International Journal of Alternative Powertrains*, 11(2):18–27, 2021. doi: 10.4271/14-11-02-0017. URL <https://doi.org/10.4271/14-11-02-0017>.
- [45] Arohan Ajit, Koustav Acharya, and Abhishek Samanta. A review of convolutional neural networks. In *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, pages 1–5, 2020. doi: 10.1109/ic-ETITE47903.2020.049. URL <https://doi.org/10.1109/ic-ETITE47903.2020.049>.
- [46] Amer Farea, Olli Yli-Harja, and Frank Emmert-Streib. Understanding physics-informed neural networks: Techniques, applications, trends, and challenges. *AI*, 5(3):1534–1557, 2024. ISSN 2673-2688. doi: 10.3390/ai5030074. URL <https://doi.org/10.3390/ai5030074>.
- [47] Aristidis G. Vrahatis, Konstantinos Lazaros, and Sotiris Kotsiantis. Graph attention networks: A comprehensive review of methods and applications. *Future Internet*, 16(9), 2024. ISSN 1999-5903. doi: 10.3390/fi16090318. URL <https://doi.org/10.3390/fi16090318>.

- [48] Seyed Mahdi Miraftebzadeh, Cristian Giovanni Colombo, Michela Longo, and Federica Foiadelli. K-means and alternative clustering methods in modern power systems. *IEEE Access*, 11:119596–119633, 2023. doi: 10.1109/ACCESS.2023.3327640. URL <https://doi.org/10.1109/ACCESS.2023.3327640>.
- [49] Jiawei Han, Micheline Kamber, and Jian Pei. Cluster analysis: Basic concepts and methods. In *Data Mining*, The Morgan Kaufmann Series in Data Management Systems, pages 443–495. Morgan Kaufmann, Boston, 3rd edition, 2012. ISBN 978-0-12-381479-1. doi: 10.1016/B978-0-12-381479-1.00010-1. URL <https://doi.org/10.1016/B978-0-12-381479-1.00010-1>.
- [50] Mohiuddin Ahmed, Raihan Seraj, and Syed Mohammed Shamsul Islam. The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9(8), 2020. ISSN 2079-9292. doi: 10.3390/electronics9081295. URL <https://doi.org/10.3390/electronics9081295>.
- [51] Mahnoor Chaudhry, Imran Shafi, Mahnoor Mahnoor, Debora Libertad Ramírez Vargas, Ernesto Bautista Thompson, and Imran Ashraf. A systematic literature review on identifying patterns using unsupervised clustering algorithms: A data mining perspective. *Symmetry*, 15(9), 2023. ISSN 2073-8994. doi: 10.3390/sym15091679. URL <https://doi.org/10.3390/sym15091679>.
- [52] Virginia Negri, Alessandro Mingotti, Roberto Tinarelli, and Lorenzo Peretto. Comparison between the machine learning and the statistical approach to the forecasting of voltage, current, and frequency. In *2023 IEEE 13th International Workshop on Applied Measurements for Power Systems (AMPS)*, pages 01–06, 2023. doi: 10.1109/AMPS59207.2023.10297192. URL <https://doi.org/10.1109/AMPS59207.2023.10297192>.
- [53] Virginia Negri, Samuel Zanella, Alessandro Mingotti, Roberto Tinarelli, Lorenzo Peretto, Francesco Barchi, and Andrea Acquaviva. Towards the dt of an educational building: An ai-based distributed measurement system for the power forecasting. In *2024 IEEE 14th International Workshop on Applied Measurements for Power Systems (AMPS)*, pages 1–6, 2024. doi: 10.1109/AMPS62611.2024.10706690. URL <https://doi.org/10.1109/AMPS62611.2024.10706690>.

- [54] Virginia Negri, Roberto Tinarelli, Lorenzo Peretto, and Alessandro Mingotti. Comprehensive forecasting of electrical quantities in an educational building via artificial intelligence-driven distributed measurement system. *Sensors*, 25(8), 2025. ISSN 1424-8220. doi: 10.3390/s25082456. URL <https://doi.org/10.3390/s25082456>.
- [55] Virginia Negri, Livio D’Alvia, and Alessandro Mingotti. Enhancing efficiency prediction accuracy in pem fuel cell hybrid systems using ai. In *2025 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1–6, 2025. doi: 10.1109/I2MTC62753.2025.11078974. URL <https://doi.org/10.1109/I2MTC62753.2025.11078974>.
- [56] Shuli Wen, Yu Wang, Yi Tang, Yan Xu, and Pengfei Li. Proactive frequency control based on ultra-short-term power fluctuation forecasting for high renewables penetrated power systems. *IET Renewable Power Generation*, 13, 07 2019. doi: 10.1049/iet-rpg.2019.0234. URL <https://doi.org/10.1049/iet-rpg.2019.0234>.
- [57] Maizura Mokhtar, Valentin Robu, David Flynn, Ciaran Higgins, Jim Whyte, Caroline Loughran, and Fiona Fulton. Predicting the voltage distribution for low voltage networks using deep learning, 2019. URL <https://doi.org/10.1109/ISGTEurope.2019.8905434>.
- [58] Iroshani Jayawardene and Ganesh Venayagamoorthy. Cellular computational extreme learning machine network based frequency predictions in a power system. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 3377–3384, 05 2017. doi: 10.1109/IJCNN.2017.7966280. URL <https://doi.org/10.1109/IJCNN.2017.7966280>.
- [59] Yichao Zhang, Xiaoru Wang, and Lijie Ding. Lstm-based dynamic frequency prediction. In *2020 IEEE Power & Energy Society General Meeting (PESGM)*, pages 1–5, 08 2020. doi: 10.1109/PESGM41954.2020.9281602. URL <https://doi.org/10.1109/PESGM41954.2020.9281602>.
- [60] Jiaming Chang, Yang Du, Xiaoyang Chen, Enggee Lim, Huiqing Wen, Xingshuo Li, and Lin Jiang. Frequency regulation in adaptive virtual inertia and power reserve control with high pv penetration by probabilistic forecasting. *Frontiers in Energy Research*, 10, 2022. ISSN 2296-598X. doi: 10.3389/fenrg.2022.929113. URL <https://doi.org/10.3389/fenrg.2022.929113>.

- [61] Jiabao Du, Chang Yue, Ying Shi, Jicheng Yu, Fan Sun, Changjun Xie, and Tao Su. A frequency decomposition-based hybrid forecasting algorithm for short-term reactive power. *Energies*, 14:6606, 10 2021. doi: 10.3390/en14206606. URL <https://doi.org/10.3390/en14206606>.
- [62] Public Electricity Networks. BS EN 50160:2010+A1. Voltage Characteristics of Electricity Supplied by Public Electricity Networks. BSI Standards Publication, 2015. British Standards Institution.
- [63] Keysight 3458a digital multimeter, 8.5 digit. URL <https://www.keysight.com/de/de/product/3458A/digital-multimeter-8-5-digit.html>. Accessed: 2025-07-22.
- [64] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011.
- [65] Joaquin Amat Rodrigo and Javier Escobar Ortiz. skforecast, 8 2025.
- [66] Skipper Seabold and Josef Perktold. Statsmodels: Econometric and Statistical Modeling with Python. pages 92 – 96, 2010. doi: 10.25080/Majora-92bf1922-011. URL <https://doi.org/10.25080/Majora-92bf1922-011>.
- [67] A. Benchekroun, A. Davigny, K. Hassam-Ouari, V. Courtecuisse, and B. Robyns. Grid-aware energy management system for distribution grids based on a co-simulation approach. *IEEE Transactions on Power Delivery*, 38(5):3571–3581, 2023. doi: 10.1109/TPWRD.2023.3281984. URL <https://doi.org/10.1109/TPWRD.2023.3281984>.
- [68] Naqash Ahmad, Yazeed Ghadi, Muhammad Adnan, and Mansoor Ali. Load forecasting techniques for power system: Research challenges and survey. *IEEE Access*, 10:71054–71090, 2022. doi: 10.1109/ACCESS.2022.3187839. URL <https://doi.org/10.1109/ACCESS.2022.3187839>.
- [69] Weicong Kong, Zhao Yang Dong, David J. Hill, Fengji Luo, and Yan Xu. Short-term residential load forecasting based on resident behaviour learning. *IEEE Transactions*

- on Power Systems*, 33(1):1087–1088, 2018. doi: 10.1109/TPWRS.2017.2688178. URL <https://doi.org/10.1109/TPWRS.2017.2688178>.
- [70] Tomás Vantuch, Aurora González Vidal, Alfonso P. Ramallo-González, Antonio F. Skarmeta, and Stanislav Misák. Machine learning based electric load forecasting for short and long-term period. In *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*, pages 511–516, 2018. doi: 10.1109/WF-IoT.2018.8355123. URL <https://doi.org/10.1109/WF-IoT.2018.8355123>.
- [71] Seyed Mahdi Miraftabzadeh, Michela Longo, Federica Foiadelli, Marco Pasetti, and Raul Igual. Advances in the application of machine learning techniques for power system analytics: A survey. *Energies*, 14(16), 2021. ISSN 1996-1073. doi: 10.3390/en14164776. URL <https://doi.org/10.3390/en14164776>.
- [72] Weicong Kong, Zhao Yang Dong, Youwei Jia, David J. Hill, Yan Xu, and Yuan Zhang. Short-term residential load forecasting based on lstm recurrent neural network. *IEEE Transactions on Smart Grid*, 10(1):841–851, 2019. doi: 10.1109/TSG.2017.2753802. URL <https://doi.org/10.1109/TSG.2017.2753802>.
- [73] Jason Runge and Etienne Saloux. A comparison of prediction and forecasting artificial intelligence models to estimate the future energy demand in a district heating system. *Energy*, 269:126661, 2023. ISSN 0360-5442. doi: 10.1016/j.energy.2023.126661. URL <https://doi.org/10.1016/j.energy.2023.126661>.
- [74] Nehal Elshaboury, Eslam Mohammed Abdelkader, Abobakr Al-Sakkaf, and Ashutosh Bagchi. A deep convolutional neural network for predicting electricity consumption at grey nuns building in canada. *Construction Innovation*, 25(2):270–289, 04 2023. ISSN 1471-4175. doi: 10.1108/CI-01-2023-0005. URL <https://doi.org/10.1108/CI-01-2023-0005>.
- [75] Mohammad Azhar Mat Daut, Mohammad Yusri Hassan, Hayati Abdullah, Hasimah Abdul Rahman, Md Pauzi Abdullah, and Faridah Hussin. Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review. *Renewable and Sustainable Energy Reviews*, 70(C):1108–1118, 2017. doi: 10.1016/j.rser.2016.12.015. URL <https://doi.org/10.1016/j.rser.2016.12.015>.

- [76] Jason Runge and Radu Zmeureanu. Forecasting energy use in buildings using artificial neural networks: A review. *Energies*, 12(17), 2019. ISSN 1996-1073. doi: 10.3390/en12173254. URL <https://doi.org/10.3390/en12173254>.
- [77] Mathieu Bourdeau, Xiao-Qiang Zhai, Elyes Nefzaoui, Xiaofeng Guo, and Patrice Chatellier. Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustainable Cities and Society*, 48, 07 2019. doi: 10.1016/j.scs.2019.101533. URL <https://doi.org/10.1016/j.scs.2019.101533>.
- [78] Peng Yu, Lihong Ma, Rong Fu, Yafeng Liang, Dan Qin, Jie Yu, and Siyang Liao. Framework design and application perspectives of digital twin microgrid. *Energy Reports*, 9:669–678, sep 2023. doi: 10.1016/j.egyr.2023.04.253. URL <https://doi.org/10.1016/j.egyr.2023.04.253>.
- [79] Chaoqun Song, Xiaohui Shen, and Junchao Xia. A digital twin model for automatic width control of hot rolling mill. *IEEE Access*, PP:1–1, 01 2023. doi: 10.1109/ACCESS.2023.3306782. URL <https://doi.org/10.1109/ACCESS.2023.3306782>.
- [80] iotto. URL <https://www.iotto.it/>. Accessed: 2025-07-16.
- [81] Energymid em2389. URL <https://camillebauer.com/en/product/energymid-em2389/>. Accessed: 2025-07-24.
- [82] Lovato electric dm3ta. URL https://catalogue.lovatoelectric.com/gl_en/Products/p?search=dm3ta. Accessed: 2025-07-21.
- [83] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. 2019.
- [84] Andrea Borghesi, Alessio Burrello, and Andrea Bartolini. Examon-x: A predictive maintenance framework for automatic monitoring in industrial iot systems. *IEEE Internet of Things Journal*, PP:1–1, 11 2021. doi: 10.1109/JIOT.2021.3125885. URL <https://doi.org/10.1109/JIOT.2021.3125885>.

- [85] Joe Rodgers and Alan Nicewander. Thirteen ways to look at the correlation coefficient. *American Statistician - AMER STATIST*, 42:59–66, 02 1988. doi: 10.1080/00031305.1988.10475524. URL <https://doi.org/10.1080/00031305.1988.10475524>.
- [86] Hana Brůhová Foltýnová, Eliška Vejchodská, Kristýna Rybová, and Viktor Květoň. Sustainable urban mobility: One definition, different stakeholders' opinions. *Transportation Research Part D: Transport and Environment*, 87:102465, 2020. ISSN 1361-9209. doi: 10.1016/j.trd.2020.102465. URL <https://doi.org/10.1016/j.trd.2020.102465>.
- [87] David Banister. The sustainable mobility paradigm. *Transport Policy*, 15(2):73–80, 2008. ISSN 0967-070X. doi: 10.1016/j.tranpol.2007.10.005. URL <https://doi.org/10.1016/j.tranpol.2007.10.005>. New Developments in Urban Transportation Planning.
- [88] Peter Hall and Kathy Pain. *The Polycentric Metropolis: Learning from Mega-City Regions in Europe*. Routledge, 1st edition, 2006. doi: 10.4324/9781849773911. URL <https://doi.org/10.4324/9781849773911>.
- [89] D. Ding and X.-Y. Wu. Hydrogen fuel cell electric trains: Technologies, current status, and future. *Applications in Energy and Combustion Science*, 17:100255, Mar 2024. doi: 10.1016/J.JAECS.2024.100255. URL <https://doi.org/10.1016/J.JAECS.2024.100255>.
- [90] M. A. Hannan, M. S. H. Lipu, A. Hussain, and A. Mohamed. A review of lithium-ion battery state of charge estimation and management system in electric vehicle applications: Challenges and recommendations. *Renewable and Sustainable Energy Reviews*, 78:834–854, Oct 2017. doi: 10.1016/j.rser.2017.05.001. URL <https://doi.org/10.1016/j.rser.2017.05.001>.
- [91] F. Leach, G. Kalghatgi, R. Stone, and P. Miles. The scope for improving the efficiency and environmental impact of internal combustion engines. *Transportation Engineering*, 1:100005, Jun 2020. doi: 10.1016/j.treng.2020.100005. URL <https://doi.org/10.1016/j.treng.2020.100005>.
- [92] M. R. Khalid, I. A. Khan, S. Hameed, M. S. J. Asghar, and J.-S. Ro. A comprehensive review on structural topologies, power levels, energy storage systems,

- and standards for electric vehicle charging stations and their impacts on grid. *IEEE Access*, 9:128069–128094, 2021. doi: 10.1109/ACCESS.2021.3112189. URL <https://doi.org/10.1109/ACCESS.2021.3112189>.
- [93] S. F. Tie and C. W. Tan. A review of energy sources and energy management system in electric vehicles. *Renewable and Sustainable Energy Reviews*, 20:82–102, Apr 2013. doi: 10.1016/j.rser.2012.11.077. URL <https://doi.org/10.1016/j.rser.2012.11.077>.
- [94] G. D’Ovidio, A. Ometto, and O. Valentini. A novel predictive power flow control strategy for hydrogen city rail train. *International Journal of Hydrogen Energy*, 45 (7):4922–4931, Feb 2020. doi: 10.1016/j.ijhydene.2019.12.067. URL <https://doi.org/10.1016/j.ijhydene.2019.12.067>.
- [95] M. Mykhalevych, S. Shuklinov, V. Dvadnenko, and O. Yaryta. Prospects of ‘mild hybrid’ technology for creating a hybridization system of vehicles. *Automobile Transport*, (50):29–39, Jul 2022. doi: 10.30977/AT.2019-8342.2022.50.0.04. URL <https://doi.org/10.30977/AT.2019-8342.2022.50.0.04>.
- [96] Horizon h-200 pem fuel cell - 200w. URL <https://www.horizoneducational.com/h-200-pem-fuel-cell-200w/p1349?currency=usd>. Accessed: 2025-07-23.
- [97] National Instruments. *NI USB-6002 Specifications*, 2025. URL <https://www.ni.com/docs/en-US/bundle/usb-6002-specs/resource/374371a.pdf>. Accessed: 2025-07-16.
- [98] Uni en 1986-1 electrically propelled road vehicles - measurement of energy performances - pure electric vehicles, n.d. Standard.
- [99] En 1821-1:1996 electrically propelled road vehicles - measurement of road operating ability - pure electric vehicles, 1996. Standard.
- [100] Julakha Jahan Jui, Mohd Ashraf Ahmad, M. M. Imran Molla, and Muhammad Ikram Mohd Rashid. Optimal energy management strategies for hybrid electric vehicles: A recent survey of machine learning approaches. *Journal of Engineering Research*, 12(3):454–467, 2024. ISSN 2307-1877. doi: 10.1016/j.jer.2024.01.016. URL <https://doi.org/10.1016/j.jer.2024.01.016>.

- [101] V. Negri, A. Mingotti, L. Peretto, and R. Tinarelli. Artificial intelligence and smart grids: the cable joint test case. In *2022 IEEE 12th International Workshop on Applied Measurements for Power Systems (AMPS)*, pages 1–6, 2022. doi: 10.1109/AMPS55790.2022.9978846. URL <https://doi.org/10.1109/AMPS55790.2022.9978846>.
- [102] Virginia Negri, Alessandro Mingotti, Roberto Tinarelli, and Lorenzo Peretto. Comparison of algorithms for the ai-based fault diagnostic of cable joints in mv networks. *Energies*, 16(1), 2023. ISSN 1996-1073. doi: 10.3390/en16010470. URL <https://doi.org/10.3390/en16010470>.
- [103] Virginia Negri, Alessandro Mingotti, Roberto Tinarelli, Lorenzo Peretto, Lala Shakti Swarup Ray, Bo Zhou, and Paul Lukowicz. A novel health index for mv cable joint aging prediction based on dynamic graph attention model. *IEEE Transactions on Instrumentation and Measurement*, 74:1–8, 2025. doi: 10.1109/TIM.2025.3550602. URL <https://doi.org/10.1109/TIM.2025.3550602>.
- [104] Virginia Negri, Grazia Iadarola, Alessandro Mingotti, Susanna Spinsante, Roberto Tinarelli, and Lorenzo Peretto. Predictive maintenance based on artificial neural network for mv switchgears. *IEEE Sensors Journal*, 24(21):35448–35455, 2024. doi: 10.1109/JSEN.2024.3455755. URL <https://doi.org/10.1109/JSEN.2024.3455755>.
- [105] Grazia Iadarola, Alessandro Mingotti, Virginia Negri, and Susanna Spinsante. Fault detection in mv switchgears through unsupervised learning of temperature conditions. *Sensors*, 25(15), 2025. ISSN 1424-8220. doi: 10.3390/s25154818. URL <https://doi.org/10.3390/s25154818>.
- [106] Massimo Pompili, Luigi Calcara, and Silvia Sangiovanni. Heatwaves and underground mv cable joints failures. In *2021 AEIT International Annual Conference (AEIT)*, pages 1–5, 2021. doi: 10.23919/AEIT53387.2021.9627070. URL <https://doi.org/10.23919/AEIT53387.2021.9627070>.
- [107] Alessandro Mingotti, Farshid Babaei, Roberto Tinarelli, and Lorenzo Peretto. Medium-voltage ac cable joints: A review of testing methods, standards, and emerging trends. *Sensors*, 25(13), 2025. doi: 10.3390/s25133843. URL <https://doi.org/10.3390/s25133843>.

- [108] Yun Jiang, Hong Min, Junhua Luo, Yu Li, Xiaojuan Jiang, Rong Xia, and Wenjie Li. Partial discharge pattern characteristic of hv cable joints with typical artificial defect. In *2010 Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pages 1–4, May 2010. doi: 10.1109/APPEEC.2010.5449349. URL <https://doi.org/10.1109/APPEEC.2010.5449349>.
- [109] Luigi Testa, Andrea Cavallini, Gian Montanari, and A. Makovoz. On-line partial discharges monitoring of high voltage xlpe / fluid-filled transition joints. In *2011 Electrical Insulation Conference (EIC)*, pages 354–358, Jun 2011. doi: 10.1109/EIC.2011.5996177. URL <https://doi.org/10.1109/EIC.2011.5996177>.
- [110] Ruay-Nan Wu and Chien-Kuo Chang. The use of partial discharges as an on-line monitoring system for underground cable joints. *IEEE Transactions on Power Delivery*, 26:1585–1591, Aug 2011. doi: 10.1109/TPWRD.2011.2124474. URL <https://doi.org/10.1109/TPWRD.2011.2124474>.
- [111] Xiang Dong, Yanling Yuan, Zhongqiang Gao, Chengke Zhou, Peter Wallace, Babakalli Alkali, Bojie Sheng, and Hao Zhou. Analysis of cable failure modes and cable joint failure detection via sheath circulating current. In *2014 IEEE Electrical Insulation Conference (EIC)*, pages 294–298, 2014. doi: 10.1109/EIC.2014.6869395. URL <https://doi.org/10.1109/EIC.2014.6869395>.
- [112] D. Fournier and L. Lamarre. Effect of pressure and length on interfacial breakdown between two dielectric surfaces. In *1992 IEEE International Symposium on Electrical Insulation*, pages 270–272, 1992. doi: 10.1109/ELINSL.1992.247004. URL <https://doi.org/10.1109/ELINSL.1992.247004>.
- [113] Xiangxian Zhou, Junping Cao, Shaohua Wang, Yukuan Jiang, Te Li, and Yang Zou. Simulation of electric field around typical defects in 110kv xlpe power cable joints. In *2017 IEEE International Conference on Condition Monitoring and Diagnosis (ICCMD)*, pages 21–24, Sep 2017. doi: 10.1109/ICCD.2017.8120443. URL <https://doi.org/10.1109/ICCD.2017.8120443>.
- [114] Alessandro Mingotti, Abbas Ghaderi, Lorenzo Peretto, Roberto Tinarelli, and Filippo Lama. Test setup design, and calibration for tan delta measurements on mv cable joints. In *2018 IEEE 9th International Workshop on Applied Measurements for*

- Power Systems (AMPS)*, pages 1–5, Sep 2018. doi: 10.1109/AMPS.2018.8494860. URL <https://doi.org/10.1109/AMPS.2018.8494860>.
- [115] Abbas Ghaderi, Alessandro Mingotti, Filippo Lama, Lorenzo Peretto, and Roberto Tinarelli. Effects of temperature on mv cable joints tan delta measurements. *IEEE Transactions on Instrumentation and Measurement*, 68(10):3892–3898, 2019. doi: 10.1109/TIM.2019.2900131. URL <https://doi.org/10.1109/TIM.2019.2900131>.
- [116] Abbas Ghaderi, Alessandro Mingotti, Lorenzo Peretto, and Roberto Tinarelli. Effects of mechanical pressure on the tangent delta of mv cable joints. *IEEE Transactions on Instrumentation and Measurement*, 68, Apr 2019. doi: 10.1109/TIM.2019.2910001. URL <https://doi.org/10.1109/TIM.2019.2910001>.
- [117] Raffaella Sante, Abbas Ghaderi, Alessandro Mingotti, Lorenzo Peretto, and Roberto Tinarelli. Test bed characterization for the interfacial pressure vs. temperature measurements in mv cable-joints. In *2019 IEEE Metro Industry Research Conference (MetroInd4)*, pages 186–190, Jun 2019. doi: 10.1109/METROI4.2019.8792867. URL <https://doi.org/10.1109/METROI4.2019.8792867>.
- [118] Rogier Jongen, Edward Gulski, and J. Smit. Failures of medium voltage cable joints in relation to the ambient temperature. In *IET Conference Publications*, pages 1–4, Jul 2009. doi: 10.1049/cp.2009.0832. URL <https://doi.org/10.1049/cp.2009.0832>.
- [119] Chien-Kuo Chang, Chang-Sing Lai, and Ruay-Nan Wu. Decision tree rules for insulation condition assessment of pre-molded power cable joints with artificial defects. *IEEE Transactions on Dielectrics and Electrical Insulation*, 26(5):1636–1644, 2019. doi: 10.1109/TDEI.2019.008208. URL <https://doi.org/10.1109/TDEI.2019.008208>.
- [120] Lei Tan, Peng Li, Fei Tao, Aimin Miao, and Min Cao. Cable joint fault detection for the ring main unit based on an adaptive tnpe algorithm. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10, Oct 2019. doi: 10.1002/widm.1336. URL <https://doi.org/10.1002/widm.1336>.
- [121] Jacco Heres, Roel Stijl, and Frank Reinders. Machine learning methods for the health-indexing and ranking of underground distribution cables and joints. In

- CIREC Workshop 2016*, pages 1–4, 2016. doi: 10.1049/cp.2016.0751. URL <https://doi.org/10.1049/cp.2016.0751>.
- [122] P. L. Lewin, T. Coleman, N. Koumbari, Y. Liu, and S. Christou. Avoiding medium voltage cable joint failure: Development of a real-time prognostic tool. In *2021 IEEE Electrical Insulation Conference (EIC)*, pages 181–184, 2021. doi: 10.1109/EIC49891.2021.9612392. URL <https://doi.org/10.1109/EIC49891.2021.9612392>.
- [123] K. Rana and B. Dasgupta. Analysis of ht cable and joint failures and associated design modifications. In CIGRE Working Group B1.21, editor, *Third-Party Damage to Underground and Submarine Cables*. CIGRE, Paris, France, 2009.
- [124] Analysis of joint and termination failures on extruded dielectric cables a report of the accessories subcommittee of the insulated conductors committee. *IEEE Power Engineering Review*, PER-4(12):30–31, 1984. doi: 10.1109/MPER.1984.5526216. URL <https://doi.org/10.1109/MPER.1984.5526216>.
- [125] S. Moon. Implications of thermal characteristics on joint processing and criteria for mv cables. *IEEE Transactions on Power Delivery*, 39(5):2810–2817, Oct 2024. doi: 10.1109/TPWRD.2024.3435415. URL <https://doi.org/10.1109/TPWRD.2024.3435415>.
- [126] C. Dang. Performance of compression connectors on strand-filled conductors. *IEEE Transactions on Power Delivery*, 19(3):943–946, 2004. doi: 10.1109/TPWRD.2004.829917. URL <https://doi.org/10.1109/TPWRD.2004.829917>.
- [127] C. Dang and D. Fournier. Dielectric performance of interfaces in premolded cable joints. *IEEE Transactions on Power Delivery*, 12(1):29–32, 1997. doi: 10.1109/61.568221. URL <https://doi.org/10.1109/61.568221>.
- [128] EPRI. Mv cable accessory research: 35-kv termination investigation, 2019. EPRI Report, no DOI or URL available.
- [129] Wanbin Ren, Chao Zhang, and Xilai Sun. Electrical contact resistance of contact bodies with cambered surface. *IEEE Access*, 8:93857–93867, 2020. doi: 10.1109/ACCESS.2020.2994085. URL <https://doi.org/10.1109/ACCESS.2020.2994085>.

- [130] Peng Zhang, Y. Y. Lau, and R. M. Gilgenbach. Thin film contact resistance with dissimilar materials. *Journal of Applied Physics*, 109(12):124910, Jun 2011. doi: 10.1063/1.3596759. URL <https://doi.org/10.1063/1.3596759>.
- [131] Wanbin Ren, Yu Chen, Zhaobin Wang, Shengjun Xue, and Xu Zhang. Electrical contact resistance of coated spherical contacts. *IEEE Transactions on Electron Devices*, 63(11):4373–4379, 2016. doi: 10.1109/TED.2016.2612545. URL <https://doi.org/10.1109/TED.2016.2612545>.
- [132] Raffaella Sante, Abbas Ghaderi, Alessandro Mingotti, Lorenzo Peretto, and Roberto Tinarelli. Effects of thermal cycles on interfacial pressure in mv cable joints. *Sensors*, 20:169, Dec 2019. doi: 10.3390/s20010169. URL <https://doi.org/10.3390/s20010169>.
- [133] Standard test method for relative resistance to vented water-tree growth in solid dielectric insulating material, 2016. ASTM Standard D6097 - 16, West Conshohocken, PA, USA.
- [134] Specifications for crosslinked polyethylene insulated shielded power cables rated 5 through 46 kv, 1994. AEIC Standard CS5-94, Melbourne VIC, Australia.
- [135] Compression and mechanical connectors for power cables for rated voltages up to 30 kv (um= 36 kv) - part 1: Test methods and requirements, 2003. IEC Standard 61238-1, Geneva, Switzerland.
- [136] Cable accessory workmanship on extruded high voltage câbles, 2011. Standard TB 476, Paris, France.
- [137] J. Densley. Ageing mechanisms and diagnostics for power cables - an overview. *IEEE Electrical Insulation Magazine*, 17(1):14–22, 2001. doi: 10.1109/57.901613. URL <https://doi.org/10.1109/57.901613>.
- [138] Hanyu Ye, Tobias Fechner, Xianzhang Lei, Yi Luo, Mingyu Zhou, Zhengyi Han, Haitian Wang, Qikai Zhuang, Ruoyu Xu, and Duo Li. Review on hvdc cable terminations. *High Voltage*, 3, 2018. doi: 10.1049/hve.2017.0144. URL <https://doi.org/10.1049/hve.2017.0144>.
- [139] International Electrotechnical Commission. High-voltage test techniques - partial discharge measurements, 2000. IEC Standard 60270.

- [140] High-voltage test techniques - partial discharge measurements, 2000. IEC Standard 60270.
- [141] High voltage test techniques - measurement of partial discharges by electromagnetic and acoustic methods, 2016. IEC TS 62478.
- [142] Martin D. Judd. Radiometric partial discharge detection. In *2008 International Conference on Condition Monitoring and Diagnosis*, pages 1025–1030, 2008. doi: 10.1109/CMD.2008.4580457. URL <https://doi.org/10.1109/CMD.2008.4580457>.
- [143] Sinda Kaziz, Mohamed Hadj Said, Antonino Imburgia, Bilel Maamer, Denis Flandre, Pietro Romano, and Farès Tounsi. Radiometric partial discharge detection: A review. *Energies*, 16, Feb 2023. doi: 10.3390/en16041978. URL <https://doi.org/10.3390/en16041978>.
- [144] Jiayang Wu, Armando Rodrigo Mor, Paul V.M. van Nes, and Johan J. Smit. Measuring method for partial discharges in a high voltage cable system subjected to impulse and superimposed voltage under laboratory conditions. *International Journal of Electrical Power & Energy Systems*, 115:105489, 2020. doi: 10.1016/j.ijepes.2019.105489. URL <https://doi.org/10.1016/j.ijepes.2019.105489>.
- [145] Martin Fritsch and Martin Wolter. High-frequency current transformer design and construction guide. *IEEE Transactions on Instrumentation and Measurement*, 71: 1–9, 2022. doi: 10.1109/TIM.2022.3177189. URL <https://doi.org/10.1109/TIM.2022.3177189>.
- [146] Meng-Hui Wang, Shiue-Der Lu, and Rui-Min Liao. Fault diagnosis for power cables based on convolutional neural network with chaotic system and discrete wavelet transform. *IEEE Transactions on Power Delivery*, 37(1):582–590, 2022. doi: 10.1109/TPWRD.2021.3065342. URL <https://doi.org/10.1109/TPWRD.2021.3065342>.
- [147] D. Passow, M. Beltle, S. Tenbohlen, and J. Hohloch. Bandwidth and detection sensitivity analysis of integrated capacitive pd sensors for pre-molded cable joints. *IEEE Transactions on Dielectrics and Electrical Insulation*, 27(6):2156–2164, 2020. doi: 10.1109/TDEI.2020.009078. URL <https://doi.org/10.1109/TDEI.2020.009078>.
- [148] Guo-Ming Ma, Cheng-Rong Li, Xue-Wei Chen, Jun Jiang, Zhen-Dong Ge, and Wen-Zhi Chang. Numerical sensor design for partial discharge detection on power cable

- joint. *IEEE Transactions on Dielectrics and Electrical Insulation*, 22(4):2311–2319, 2015. doi: 10.1109/TDEI.2015.004957. URL <https://doi.org/10.1109/TDEI.2015.004957>.
- [149] Mingshu Zhao, Xiaoyan Cao, Kai Zhou, Yao Fu, Xutao Li, and Li Wan. Flexible sensor array based on transient earth voltage for online partial discharge monitoring of cable termination. *Sensors*, 20:6646, Nov 2020. doi: 10.3390/s20226646. URL <https://doi.org/10.3390/s20226646>.
- [150] Pietro Romano, Antonino Imburgia, and Guido Ala. Partial discharge detection using a spherical electromagnetic sensor. *Sensors*, 19:1014, Feb 2019. doi: 10.3390/s19051014. URL <https://doi.org/10.3390/s19051014>.
- [151] Antonino Madonia, Eleonora Riva Sanseverino, Pietro Romano, Ivan Troia, Stefano Franchi Bononi, Marco Albertini, Simone Giannini, and Giovanni Mazzanti. Wireless partial discharge tracking on cross-linked polyethylene mv and hv cables. *IEEE Electrical Insulation Magazine*, 34(6):8–17, 2018. doi: 10.1109/MEI.2018.8507713. URL <https://doi.org/10.1109/MEI.2018.8507713>.
- [152] Joel Yeo and Low Chi Kin. On-line partial discharge detection on transformer cable sealing ends in singapore’s transmission network. *IEEE Electrical Insulation Magazine*, 39(3):22–30, 2023. doi: 10.1109/MEI.2023.10101706. URL <https://doi.org/10.1109/MEI.2023.10101706>.
- [153] Wojciech Koltunowicz and Ronald Plath. Synchronous multi-channel pd measurements. *IEEE Transactions on Dielectrics and Electrical Insulation*, 15(6):1715–1723, 2008. doi: 10.1109/TDEI.2008.4712676. URL <https://doi.org/10.1109/TDEI.2008.4712676>.
- [154] Tianxin Zhuang, Ming Ren, Xuze Gao, Ming Dong, Wenguang Huang, and Chongxing Zhang. Insulation condition monitoring in distribution power grid via iot-based sensing network. *IEEE Transactions on Power Delivery*, 34(4):1706–1714, 2019. doi: 10.1109/TPWRD.2019.2918289. URL <https://doi.org/10.1109/TPWRD.2019.2918289>.
- [155] Jan Fulneček and Stanislav Mišák. A simple method for tree fall detection on medium voltage overhead lines with covered conductors. *IEEE Transactions on*

- Power Delivery*, 36(3):1411–1417, 2021. doi: 10.1109/TPWRD.2020.3008482. URL <https://doi.org/10.1109/TPWRD.2020.3008482>.
- [156] Wei-qi Qin, Guo-ming Ma, Jing Hu, Tengjun Guo, Sihan Wang, Hao Liu, and Yuan Wang. Distributed detection and acoustic emission waveform retrieval of cable joint partial discharge. *IEEE Transactions on Power Delivery*, 38(4):2977–2980, 2023. doi: 10.1109/TPWRD.2023.3276197. URL <https://doi.org/10.1109/TPWRD.2023.3276197>.
- [157] Mirosław Parol, Jacek Wasilewski, and Janusz Jakubowski. Assessment of electric shock hazard coming from earth continuity conductors in 110 kv cable lines. *IEEE Transactions on Power Delivery*, 35(2):600–608, 2020. doi: 10.1109/TPWRD.2019.2916890. URL <https://doi.org/10.1109/TPWRD.2019.2916890>.
- [158] Mingzhen Li, Chengke Zhou, Wenjun Zhou, Jun Zhang, Liang Zhang, and Leiming Yao. Feasibility study on lengthening the high voltage cable section and reducing the number of cable joints via alternative bonding methods. *High Voltage*, 4, Dec 2019. doi: 10.1049/hve.2019.0011. URL <https://doi.org/10.1049/hve.2019.0011>.
- [159] Erik Winkelmann, Iaroslav Shevchenko, Christoph Steiner, Christian Kleiner, Uwe Kaltenborn, Peter Birkholz, Harald Schwarz, and Thomas Steiner. Monitoring of partial discharges in hvdc power cables. *IEEE Electrical Insulation Magazine*, 38(1):7–18, 2022. doi: 10.1109/MEI.2022.9648269. URL <https://doi.org/10.1109/MEI.2022.9648269>.
- [160] A. Naderian, S. Cress, R. Piercy, F. Wang, and J. Service. An approach to determine the health index of power transformers. In *Conference Record of the 2008 IEEE International Symposium on Electrical Insulation*, pages 192–196, 2008. doi: 10.1109/ELINSL.2008.4570308. URL <https://doi.org/10.1109/ELINSL.2008.4570308>.
- [161] Tanachai Somsak, Cattareeya Suwanasri, and Thanapong Suwanasri. Asset management of underground cable system for industrial estate in thailand. In *2018 International Electrical Engineering Congress (iEECON)*, pages 1–4, 2018. doi: 10.1109/IEECON.2018.8712331. URL <https://doi.org/10.1109/IEECON.2018.8712331>.
- [162] Pranav Pattabi, Ali Naderian Jahromi, Shanon Lo, Kurtis Martin-Sturmey, and Nirvaan Bhagwandass. Practical approach to underground distribution power cable

- fleet management. *Energies*, 16(8):3526, 2023. doi: 10.3390/en16083526. URL <https://doi.org/10.3390/en16083526>.
- [163] Yuli Wang, Ge Wang, Yifan He, and Cheng Xie. Status evaluation of distribution cable based on combined weighting method. In *2023 International Conference on Power System Technology (PowerCon)*, pages 1–5, 2023. doi: 10.1109/PowerCon58120.2023.10331096. URL <https://doi.org/10.1109/PowerCon58120.2023.10331096>.
- [164] Gen Zhao, Haibo Di, Hao Bai, and Yuxuan Lin. A cable health assessment method based on multi-agent and matter-element extension model. *Sustainable Energy Technologies and Assessments*, 56:103108, 2023. doi: 10.1016/j.seta.2023.103108. URL <https://doi.org/10.1016/j.seta.2023.103108>.
- [165] Ming Dong, Wenyuan Li, and Alexandre B. Nassif. Long-term health index prediction for power asset classes based on sequence learning. *IEEE Transactions on Power Delivery*, 37(1):197–207, 2022. doi: 10.1109/TPWRD.2021.3055622. URL <https://doi.org/10.1109/TPWRD.2021.3055622>.
- [166] Gian Montanari. Condition monitoring and dynamic health index in electrical grids. pages 82–85, Sep 2016. doi: 10.1109/CMD.2016.7757787. URL <https://doi.org/10.1109/CMD.2016.7757787>.
- [167] Rakesh Sahoo, Subrata Karmakar, and Satyajit Panigrahy. Health index analysis of XLPE cable insulation using machine learning technique. pages 1–6, Nov 2020. doi: 10.1109/UPCON50219.2020.9376573. URL <https://doi.org/10.1109/UPCON50219.2020.9376573>.
- [168] Dineenath Romphoyen, Nattapon Panmala, Thanapong Suwanasri, and Cattareeya Suwanasri. Condition assessment of underground cable system using fuzzy logic and scoring and weighting average methods. In *2023 International Conference on Power, Energy and Innovations (ICPEI)*, pages 108–111, 2023. doi: 10.1109/ICPEI58931.2023.10473896. URL <https://doi.org/10.1109/ICPEI58931.2023.10473896>.
- [169] Krzysztof Siodla, Aleksandra Rakowska, and Slawomir Noske. The proposal of a new tool for condition assessment of medium voltage power cable lines. *Energies*, 14(14):4116, 2021. doi: 10.3390/en14144116. URL <https://doi.org/10.3390/en14144116>.

- [170] Tanachai Somsak, Thanapong Suwanasri, and Cattareeya Suwanasri. Lifetime estimation based health index and conditional factor for underground cable system. *Energies*, 14(23):8113, 2021. doi: 10.3390/en14238113. URL <https://doi.org/10.3390/en14238113>.
- [171] Tanachai Somsak, Thanapong Suwanasri, and Cattareeya Suwanasri. Remaining useful life estimation for underground cable systems based on historical health index. *Energies*, 15:9447, Dec 2022. doi: 10.3390/en15249447. URL <https://doi.org/10.3390/en15249447>.
- [172] Gian Carlo Montanari, Paolo Seri, and R.E. Hebner. A scheme for the health index and residual life of cables based on measurement and monitoring of diagnostic quantities. In *2018 IEEE Power & Energy Society General Meeting (PESGM)*, pages 1–5, 2018. doi: 10.1109/PESGM.2018.8585858. URL <https://doi.org/10.1109/PESGM.2018.8585858>.
- [173] Qiaosheng Pan, Chi Zhang, Xinyuan Wei, Aode Wan, and Zeping Wei. Assessment of MV XLPE cable aging state based on PSO-xgboost algorithm. *Electric Power Systems Research*, 221:109427, 2023. doi: 10.1016/j.epsr.2023.109427. URL <https://doi.org/10.1016/j.epsr.2023.109427>.
- [174] Riley Alden. 20kv XLPE cable historical inspection data, 2020. URL <https://dx.doi.org/10.21227/jyrf-mk47>. Accessed: 2025-07-01.
- [175] IEC. International standard iec 60270 high-voltage test techniques partial discharge measurements.
- [176] IEEE. IEEE guide for field testing of shielded power cable systems using very low frequency (vlf)(less than 1 hz). *IEEE Std 400.2-2013*, pages 1–60, 2013. doi: 10.1109/IEEESTD.2013.6517854. URL <https://doi.org/10.1109/IEEESTD.2013.6517854>.
- [177] Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric. 2019.
- [178] Sungho Suh, Dhruv Aditya Mittal, Hymalai Bello, Bo Zhou, Mayank Shekhar Jha, and Paul Lukowicz. Remaining useful life prediction of lithium-ion batteries using spatio-temporal multimodal attention networks. *Heliyon*, 10(16):e36236, 2024.

- doi: 10.1016/j.heliyon.2024.e36236. URL <https://doi.org/10.1016/j.heliyon.2024.e36236>.
- [179] Daoquan Chen, Weicong Hong, and Xiuze Zhou. Transformer network for remaining useful life prediction of lithium-ion batteries. *IEEE Access*, 10:19621–19628, 2022. doi: 10.1109/ACCESS.2022.3151975. URL <https://doi.org/10.1109/ACCESS.2022.3151975>.
- [180] Ahmet Efe Seker, Egemen Aslan Sakaci, Arif Deniz, Bayram Celik, and Deniz Yildirim. Thermal analyses for a simplified medium-voltage switchgear: Numerical and experimental benchmark studies. In *2019 11th International Conference on Electrical and Electronics Engineering (ELECO)*, pages 116–120, 2019. doi: 10.23919/ELECO47770.2019.8990614. URL <https://doi.org/10.23919/ELECO47770.2019.8990614>.
- [181] Sung Won Park and Hyunsu Cho. A practical study on electrical contact resistance and temperature rise at the connections of the copper busbars in switchgears. In *2014 IEEE 60th Holm Conference on Electrical Contacts (Holm)*, pages 1–7, 2014. doi: 10.1109/HOLM.2014.7031066. URL <https://doi.org/10.1109/HOLM.2014.7031066>.
- [182] Xiaorui Wang, Haotian Xia, Zhanqing Yu, Yonggang Guan, Biao Zhao, Yulong Huang, Lu Qu, and Jiudong Ding. Precise multi-dimensional temperature-rise characterisation of switchgear based on multi-conditional experiments and lptn model for high-capacity application. *High Voltage*, 6(1):138–148, 2021. doi: 10.1049/hve.2019.0280. URL <https://doi.org/10.1049/hve.2019.0280>.
- [183] Guo-Ming Ma, Zhen Wu, Hong-Yang Zhou, Jun Jiang, Wei-Xue Chen, Shu-Sheng Zheng, Cheng-Rong Li, Xiong Li, and Zeng-Bin Wang. A wireless and passive online temperature monitoring system for gis based on surface-acoustic-wave sensor. *IEEE Transactions on Power Delivery*, 31(3):1270–1280, 2016. doi: 10.1109/TPWRD.2015.2482985. URL <https://doi.org/10.1109/TPWRD.2015.2482985>.
- [184] Ghulam Amjad Hussain, Ashraf A. Zaher, Detlef Hummes, Madia Safdar, and Matti Lehtonen. Hybrid sensing of internal and surface partial discharges in air-insulated medium voltage switchgear. *Energies*, 13(7), 2020. doi: 10.3390/en13071738. URL <https://doi.org/10.3390/en13071738>.

- [185] Gian Carlo Montanari, Riddhi Ghosh, Leonardo Cirioni, Giuseppe Galvagno, and Salvatore Mastroeni. Partial discharge monitoring of medium voltage switchgears: Self-condition assessment using an embedded bushing sensor. *IEEE Transactions on Power Delivery*, 37(1):85–92, 2022. doi: 10.1109/TPWRD.2021.3053658. URL <https://doi.org/10.1109/TPWRD.2021.3053658>.
- [186] Yasutomo Otake and Hiroshi Murase. Improving the estimation of partial discharge direction using a four-terminal surface current sensor. *Energies*, 16(21), 2023. doi: 10.3390/en16217389. URL <https://doi.org/10.3390/en16217389>.
- [187] Xin Zhou and Thomas Schoepf. Characteristics of Overheated Electrical Joints Due to Loose Connection. In *2011 IEEE 57th Holm Conference on Electrical Contacts (Holm)*, pages 1–7, 2011. doi: 10.1109/HOLM.2011.6034795. URL <https://doi.org/10.1109/HOLM.2011.6034795>.
- [188] Hajime Kabashima and Fumio Kasahara. Experimental study of high energy arcing faults using medium voltage metalclad switchgears. *Nuclear Technology*, 205(5): 694–707, 2019. doi: 10.1080/00295450.2018.1518556. URL <https://doi.org/10.1080/00295450.2018.1518556>.
- [189] Frank Tre Ore and Matti Lehtonen. Novel approach for the monitoring of single phase-to-ground fault associated with power arc. *International Review of Electrical Engineering*, 17(6):544–555, 2022. doi: 10.15866/iree.v17i6.22764. URL <https://doi.org/10.15866/iree.v17i6.22764>.
- [190] Pieter Billen, Ben Maes, Macarena Larrain, and Johan Braet. Replacing sf6 in electrical gas-insulated switchgear: Technological alternatives and potential life cycle greenhouse gas savings in an eu-28 perspective. *Energies*, 13(7), 2020. doi: 10.3390/en13071807. URL <https://doi.org/10.3390/en13071807>.
- [191] Lars Maaløe, Ole Winther, Sergiu Spataru, and Dezso Sera. Conditional monitoring in photovoltaic systems by semi-supervised machine learning. *Energies*, 13(3), 2020. doi: 10.3390/en13030584. URL <https://doi.org/10.3390/en13030584>.
- [192] Dušan Strušnik, Marko Agrež, Juri Avsec, and Marjan Golob. Optimisation of an old 200 mw coal-fired boiler with urea injection through the use of supervised machine learning algorithms to achieve cleaner power generation. *Journal of Cleaner*

- Production*, 290, 2021. doi: 10.1016/j.jclepro.2020.125200. URL <https://doi.org/10.1016/j.jclepro.2020.125200>.
- [193] Cheng Fan, Yutian Lei, Yongjun Sun, and Like Mo. Novel transformer-based self-supervised learning methods for improved hvac fault diagnosis performance with limited labeled data. *Energy*, 278, 2023. doi: 10.1016/j.energy.2023.127972. URL <https://doi.org/10.1016/j.energy.2023.127972>.
- [194] Lian Wang, Yuedong Yao, Kongjie Wang, Caspar Daniel Adenutsi, Guoxiang Zhao, and Fengpeng Lai. Hybrid application of unsupervised and supervised learning in forecasting absolute open flow potential for shale gas reservoirs. *Energy*, 243, 2022. doi: 10.1016/j.energy.2021.122747. URL <https://doi.org/10.1016/j.energy.2021.122747>.
- [195] Chih-Chiang Wei and Yen-Chen Yang. A global solar radiation forecasting system using combined supervised and unsupervised learning models. *Energies*, 16(23), 2023. doi: 10.3390/en16237693. URL <https://doi.org/10.3390/en16237693>.
- [196] Wanjun Huang, Minghua Chen, and Steven H. Low. Unsupervised learning for solving ac optimal power flows: Design, analysis, and experiment. *IEEE Transactions on Power Systems*, page 1–13, 2024. doi: 10.1109/TPWRS.2024.3373399. URL <https://doi.org/10.1109/TPWRS.2024.3373399>.
- [197] Younes Mohammadi, Seyed Mahdi Miraftebadeh, Math H.J. Bollen, and Michela Longo. Voltage-sag source detection: Developing supervised methods and proposing a new unsupervised learning. *Sustainable Energy, Grids and Networks*, 32, 2022. doi: 10.1016/j.segan.2022.100855. URL <https://doi.org/10.1016/j.segan.2022.100855>.
- [198] Manisha Gupta and Behin Elahi. Prioritizing Failure-Contributing Causes in Switchgear Systems. In *2025 Annual Reliability and Maintainability Symposium (RAMS)*, pages 1–6, 2025. doi: 10.1109/RAMS48127.2025.10935263. URL <https://doi.org/10.1109/RAMS48127.2025.10935263>.
- [199] Dev Paul and Kor Yan. Cause of Arcing Inside a Medium-Voltage Arc-Resistant Metal-Clad Switchgear Compartment on a Sunny Day Without Any Evidence of Switching Surge: Recommendations for avoiding future incidents. *IEEE Industry*

- Applications Magazine*, 30(4):27–36, 2024. doi: 10.1109/MIAS.2023.3338589. URL <https://doi.org/10.1109/MIAS.2023.3338589>.
- [200] D. Alberto, D. El Khoury, E. Centenaro, V. Ferraro, and A. Cavallini. Advanced switchgear diagnostics through PD monitoring correlated with environmental and operating parameters. In *27th International Conference on Electricity Distribution (CIRED 2023)*, volume 2023, pages 3655–3659, 2023. doi: 10.1049/icp.2023.0780. URL <https://doi.org/10.1049/icp.2023.0780>.
- [201] Keyur Tandel and Thierry Cormenier. Extreme weather has become the norm – is your mv switchgear ready? *CIRED*, 2017:166–169, 2017. doi: 10.1049/oap-cired.2017.0372. URL <https://doi.org/10.1049/oap-cired.2017.0372>.
- [202] IEEE Standard for Metal-Enclosed Interrupter Switchgear Rated above 1 kV AC up to and Including 48.3 kV AC. *IEEE Std C37.20.3-2023 (Revision of IEEE Std C37.20.3-2013)*, pages 1–92, 2023. doi: 10.1109/IEEESTD.2023.10264853. URL <https://doi.org/10.1109/IEEESTD.2023.10264853>.
- [203] InterNational Electrical Testing Association, . URL <https://www.netaworld.org/home>. Accessed: 2025-07-16.
- [204] Nediljko Kaštelan, Igor Vujović, Maja Krčum, and Nur Assani. Switchgear digitalization—research path, status, and future work. *Sensors*, 22(20), 2022. ISSN 1424-8220. doi: 10.3390/s22207922. URL <https://doi.org/10.3390/s22207922>.
- [205] International Electrotechnical Commission (IEC). High-voltage switchgear and controlgear – part 200: Ac metal-enclosed switchgear and controlgear for rated voltages above 1 kv and up to and including 52 kv, May 2011. International Standard IEC 62271-200.
- [206] Susanna Spinsante, Grazia Iadarola, Gianmarco Mazzocchi, and Claudio Romagnoli. Temperature rise in mv switchgears: the role of loose busbar joints. pages 51–56, 01 2022. doi: 10.21014/tc4-2022.10. URL <https://doi.org/10.21014/tc4-2022.10>.
- [207] National instruments cdaq-9178 compactdaq chassis. URL https://www.ni.com/it-it/shop/model/cdaq-9178.html?srsltid=AfmB0ooZgU2N8nkrwwMxeWDzzqSd63K02GncD5FeQ__STalSWtGxY4rd. Accessed: 2025-07-25.

- [208] National instruments: Ni-9214 specifications, . URL <https://www.ni.com/docs/en-US/bundle/ni-9214-specs/page/specs.html>. Accessed: 2025-07-20.
- [209] David C. Marcu and Cristian Grava. The impact of activation functions on training and performance of a deep neural network. In *2021 16th International Conference on Engineering of Modern Electric Systems (EMES)*, pages 1–4, 2021. doi: 10.1109/EMES52337.2021.9484108. URL <https://doi.org/10.1109/EMES52337.2021.9484108>.
- [210] François Chollet et al. Keras, 2015.
- [211] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mane, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viegas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. 2016.
- [212] International Electrotechnical Commission. IEC 60751:2008 - Industrial platinum resistance thermometers and platinum temperature sensors, 2008.
- [213] Robert David, Jared Duke, Advait Jain, Vijay Janapa Reddi, Nat Jeffries, Jian Li, Nick Kreeger, Ian Nappier, Meghna Natraj, Shlomi Regev, Rocky Rhodes, Tiezheng Wang, and Pete Warden. Tensorflow lite micro: Embedded machine learning on tinymml systems. 2021. URL <https://arxiv.org/abs/2010.08678>.
- [214] Abiodun M. Ikotun, Absalom E. Ezugwu, Laith Abualigah, Belal Abuhaija, and Jia Heming. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622:178–210, 2023. ISSN 0020-0255. doi: 10.1016/j.ins.2022.11.139. URL <https://doi.org/10.1016/j.ins.2022.11.139>.
- [215] Arduino. Nano 33 BLE Sense Rev2, 2023. URL <https://docs.arduino.cc/hardware/nano-33-ble-sense-rev2/>. Accessed: 2025-06-29.

- [216] Virginia Negri, Alessandro Mingotti, Roberto Tinarelli, and Lorenzo Peretto. Uncertainty and lack of information affecting the training of machine learning algorithms for fault prediction of cable-joints. In *2023 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1–5, 2023. doi: 10.1109/I2MTC53148.2023.10176110. URL <https://doi.org/10.1109/I2MTC53148.2023.10176110>.
- [217] Virginia Negri, Alessandro Mingotti, Roberto Tinarelli, Lorenzo Peretto, Ludovica Apa, Livio D’Alvia, Zaccaria Del Prete, and Emanuele Rizzuto. Analyzing the performance of ai-based battery soc estimation: A metrological point of view. In *2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1–5, 2024. doi: 10.1109/I2MTC60896.2024.10560993. URL <https://doi.org/10.1109/I2MTC60896.2024.10560993>.
- [218] Virginia Negri, Simone Mari, Alessandro Mingotti, Fabrizio Ciancetta, Roberto Tinarelli, and Lorenzo Peretto. Uncertainty-driven data augmentation for improved load time series forecasting. In *2025 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1–6, 2025. doi: 10.1109/I2MTC62753.2025.11079054. URL <https://doi.org/10.1109/I2MTC62753.2025.11079054>.
- [219] Virginia Negri, Alessandro Mingotti, Roberto Tinarelli, and Lorenzo Peretto. Temperature-based accuracy estimation on instrument transformers via neural networks and microcontrollers. In *2024 IEEE 14th International Workshop on Applied Measurements for Power Systems (AMPS)*, pages 1–6, 2024. doi: 10.1109/AMPS62611.2024.10706689. URL <https://doi.org/10.1109/AMPS62611.2024.10706689>.
- [220] Virginia Negri, Alessandro Mingotti, Roberto Tinarelli, Lorenzo Peretto, Lala Shakti Swarup Ray, Bo Zhou, and Paul Lukowicz. Mitigating temperature effects in rogowski coils’ accuracy through physics-informed model. In *2025 IEEE 15th International Workshop on Applied Measurements for Power Systems (AMPS)*, pages 1–5, 2025. doi: 10.1109/AMPS66841.2025.11219924. URL <https://doi.org/10.1109/AMPS66841.2025.11219924>.

- [221] Virginia Negri, Roberto Tinarelli, Lorenzo Peretto, and Alessandro Mingotti. Lightweight digital twin for current sensing using microcontroller-integrated rogowski coil model. In *2025 IEEE 15th International Workshop on Applied Measurements for Power Systems (AMPS)*, pages 1–5, 2025. doi: 10.1109/AMPS66841.2025.11219948. URL <https://doi.org/10.1109/AMPS66841.2025.11219948>.
- [222] L. Peretto, R. Tinarelli, A. Ghaderi, A. Mingotti, G. Mazzanti, G. Valtorta, and S. Danesi. Monitoring cable current and laying environment parameters for assessing the aging rate of mv cable joint insulation. In *2018 Annual Report - Conference on Electrical Insulation and Dielectric Phenomena (CEIDP)*, pages 390–393, 2018. doi: 10.1109/CEIDP.2018.8544904. URL <https://doi.org/10.1109/CEIDP.2018.8544904>.
- [223] G. Vishnu, D. Kaliyaperumal, R. Jayaprakash, A. Karthick, V. Kumar Chinnaiyan, and A. Ghosh. Review of challenges and opportunities in the integration of electric vehicles to the grid. *World Electric Vehicle Journal*, 14(9):259, 2023. doi: 10.3390/wevj14090259. URL <https://doi.org/10.3390/wevj14090259>.
- [224] P. Tsafack, S. E. Fru, A. V. Nghemachi, and E. Tanyi. Impact of high constant charging current rates on the charge/discharge efficiency in lead acid batteries for residential photovoltaic system applications. *Journal of Energy Storage*, 63:107013, 2023. doi: 10.1016/j.est.2023.107013. URL <https://doi.org/10.1016/j.est.2023.107013>.
- [225] Xiaopeng Chen, Weixiang Shen, Thanh Tu Vo, Zhenwei Cao, and Ajay Kapoor. An overview of lithium-ion batteries for electric vehicles. In *2012 10th International Power & Energy Conference (IPEC)*, pages 230–235, 2012. doi: 10.1109/ASSCC.2012.6523269. URL <https://doi.org/10.1109/ASSCC.2012.6523269>.
- [226] Yuan-Li Ding, Zachary Cano, Aiping Yu, Jun Lu, and Zhongwei Chen. Automotive li-ion batteries: Current status and future perspectives. *Electrochemical Energy Reviews*, 2:1–28, 2019. doi: 10.1007/s41918-018-0022-z. URL <https://doi.org/10.1007/s41918-018-0022-z>.

- [227] M. Wei, M. Ye, J. B. Li, Q. Wang, and X. Xu. State of charge estimation of lithium-ion batteries using lstm and narx neural networks. *IEEE Access*, 8:189236–189245, 2020. doi: 10.1109/ACCESS.2020.3031340. URL <https://doi.org/10.1109/ACCESS.2020.3031340>.
- [228] Qiao Wang, Min Ye, Meng Wei, Chenguang Wu, and Xinxin Xu. A novel state of charge estimation method of lithium-ion battery based on narx neural network model. In *2021 IEEE International Conference on Mechatronics and Automation (ICMA)*, pages 381–385, Takamatsu, Japan, 2021. doi: 10.1109/ICMA52036.2021.9512693. URL <https://doi.org/10.1109/ICMA52036.2021.9512693>.
- [229] Amal Nefraoui, Khalid Kandoussi, Mohamed Louzazni, Abderrahim Boutahar, Rabie Elotmani, and Abdelmajid Daya. Optimal battery state of charge parameter estimation and forecasting using non-linear autoregressive exogenous. *Materials Science for Energy Technologies*, 6:522–532, 2023. doi: 10.1016/j.mset.2023.05.003. URL <https://doi.org/10.1016/j.mset.2023.05.003>.
- [230] Tdk lambda gen 10-240. URL https://product.tdk.com/en/search/power/switching-power/prg-power/info?part_no=GEN10-240-1P230. Accessed: 2025-07-20.
- [231] Agilent n3306a dc electronic load. URL <https://www.keysight.com/de/de/product/N3306A/600-watt-electronic-load-module.html>. Accessed: 2025-07-26.
- [232] National Instruments. *NI PCIe/USB-6353 Specifications*, 2025. URL <https://www.ni.com/docs/en-US/bundle/pcie-usb-6353-specs/page/specs.html>. Accessed: 2025-07-16.
- [233] National Instruments. *NI-9211 Specifications*, 2025. URL <https://www.ni.com/docs/en-US/bundle/ni-9211-specs/page/specs.html>. Accessed: 2025-07-16.
- [234] V. Carratù, V. Gallo, V. Laino, C. Liguori, and A. Pietrosanto. A survey on uncertainty assessment in ann-based measurements. In *Proc. 2023 IEEE Int. Conf. on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRINE)*, pages 1128–1132, Benevento, Italy, 2023. doi: 10.1109/METROXRINE58569.2023.10405788. URL <https://doi.org/10.1109/METROXRINE58569.2023.10405788>.

- [235] A. Mingotti, A. Baldi, L. Peretto, and R. Tinarelli. A general easy-to-use expression for uncertainty evaluation in residual voltage measurement. *IEEE Transactions on Instrumentation and Measurement*, 69(4):1576–1584, 2020. doi: 10.1109/TIM.2019.2917694. URL <https://doi.org/10.1109/TIM.2019.2917694>.
- [236] M. Faifer, C. Laurano, R. Ottoboni, and S. Toscani. A new method to represent the harmonic measurement accuracy of current transformers. *IEEE Transactions on Instrumentation and Measurement*, 73:1–11, 2024. doi: 10.1109/TIM.2024.3353857. URL <https://doi.org/10.1109/TIM.2024.3353857>. Art no. 9001611.
- [237] Ausgrid. Distribution zone substation data. URL <https://www.ausgrid.com.au/Industry/Our-Research/Data-to-Help-Industry/Distribution-zone-substation-data>. Accessed: 2025-06-15.
- [238] Ausgrid. Network standard ns178: Secondary system requirements for major substations (metering only), September 2012. URL <https://www.ausgrid.com.au>. Accessed: 2025-07-01.
- [239] Alberta Dadeboe, Farzaneh Mansourifard, and Shridatt Sugrim. Uncertainty quantification and data provenance for data pipeline security analysis. In *Proceedings of the 7th Workshop on Design Automation for CPS and IoT (DESTION '25)*, pages 1–6, New York, NY, USA, 2025. Association for Computing Machinery. doi: 10.1145/3722573.3727831. URL <https://doi.org/10.1145/3722573.3727831>.
- [240] Davide Chicco, Alessandra Sichenze, and Giuseppe Jurman. A simple guide to the use of student’s t-test, mann-whitney u test, chi-squared test, and kruskal-wallis test in biostatistics. *BioData Mining*, 18(1):56, Aug 2025. doi: 10.1186/s13040-025-00465-6.
- [241] E. Lesniewska. Modern methods of construction problem solving in designing various types of instrument transformers. *Energies*, 15(21):8199, 2022. doi: 10.3390/en15218199. URL <https://doi.org/10.3390/en15218199>.
- [242] M. Kaczmarek and E. Stano. Review of measuring methods, setups and conditions for evaluation of the inductive instrument transformers accuracy for transformation of distorted waveforms. *Energies*, 16(11):4360, 2023. doi: 10.3390/en16114360. URL <https://doi.org/10.3390/en16114360>.

- [243] A. Mingotti, L. Peretto, R. Tinarelli, F. Mauri, and I. Gentilini. Assessment of metrological characteristics of calibration systems for accuracy vs. temperature verification of voltage transformer. In *AMPS 2017 - IEEE International Workshop on Applied Measurements for Power Systems*, 2017. doi: 10.1109/AMPS.2017.8078325. URL <https://doi.org/10.1109/AMPS.2017.8078325>.
- [244] M. Agazar, D. Istrate, and P. Pradayrol. Evaluation of the accuracy and frequency response of medium-voltage instrument transformers under the combined influence factors of temperature and vibration. *Energies*, 16(13):5012, 2023. doi: 10.3390/en16135012. URL <https://doi.org/10.3390/en16135012>.
- [245] M. Faifer, C. Laurano, R. Ottoboni, and S. Toscani. Compensating the harmonic distortion introduced by instrument transformers: An improved method based on frequency-domain polynomials. In *AMPS 2021 - 11th IEEE International Workshop on Applied Measurements for Power Systems*, 2021. doi: 10.1109/AMPS50177.2021.9586015. URL <https://doi.org/10.1109/AMPS50177.2021.9586015>.
- [246] International Electrotechnical Commission. IEC 61869-1:2021 – Instrument Transformers – Part 1: General Requirements, 2021. International Standard.
- [247] B. Verhelst, S. Rens, J. Rens, J. Knockaert, and J. Desmet. On the remote calibration of instrumentation transformers: Influence of temperature. *Energies*, 16(12):4744, 2023. doi: 10.3390/en16124744. URL <https://doi.org/10.3390/en16124744>.
- [248] D. Chen, C. Fu, Y. Tong, and Y. Tian. The online calibration system for transformers based on hybrid convolution window technology. In *2022 China International Conference on Electricity Distribution (CICED)*, pages 1233–1237, September 2022. doi: 10.1109/CICED56215.2022.9928957. URL <https://doi.org/10.1109/CICED56215.2022.9928957>.
- [249] International Standardization Organization. Iec 61869-6: Part 6: Additional general requirements for low-power instrument transformers, 2016.
- [250] Avid Farhoodfar. Machine learning for mobile developers: Tensorflow lite framework. 04 2019.
- [251] Alessandro Mingotti, Christian Betti, Lorenzo Peretto, and Roberto Tinarelli. Low-power instrument transformers and energy meters: Opportunities and obstacles.

- volume 26, pages 44–48. 2023. doi: 10.1109/MIM.2023.10146561. URL <https://doi.org/10.1109/MIM.2023.10146561>.
- [252] M. Daboul, J. Orságová, V. Jurák, and M. Vrtal. Testing rogowski coils with merging units for smart grids. *Energies*, 16:7323, 2023. doi: 10.3390/en16217323. URL <https://doi.org/10.3390/en16217323>.
- [253] Alessandro Mingotti, Federica Costa, Lorenzo Peretto, and Roberto Tinarelli. Accuracy type test for rogowski coils subjected to distorted signals, temperature, humidity, and position variations. *Sensors*, 22:1397, 02 2022. doi: 10.3390/s22041397. URL <https://doi.org/10.3390/s22041397>.
- [254] Lukas M.N. Gabriel, John A. Adebisi, Leokadia N.P. Ndjuluwa, and Dickson K. Chembe. Investigation of smart grid technologies deployment for energy reliability enhancement in electricity distribution networks. *Franklin Open*, 10:100227, 2025. doi: 10.1016/j.fraope.2025.100227. URL <https://doi.org/10.1016/j.fraope.2025.100227>.
- [255] International Electrotechnical Commission (IEC). Instrument transformers – part 10: Digital interface for instrument transformers, 2016. Edition 1.0.
- [256] J. Zhao, L. Shi, and X. Sun. Design and performance study of a temperature compensated ± 1100 -kv uhvdc all fiber current transformer. *IEEE Transactions on Instrumentation and Measurement*, 70:1–6, 2021. doi: 10.1109/TIM.2020.3029364. URL <https://doi.org/10.1109/TIM.2020.3029364>.
- [257] Li T., Chen Q., Zhang X., and Lu H. Analysis of the influence of ambient temperature on the optical fiber characteristics of fiber optic current transformer sensor. In *Proceedings of 2024 IEEE 7th International Electrical and Energy Conference (CIEEC 2024)*, pages 1692–1696, 2024. doi: 10.1109/CIEEC60922.2024.10583697. URL <https://doi.org/10.1109/CIEEC60922.2024.10583697>.
- [258] Zhang C., Kang W., Xia X., Zhou Z., Liu Y., and Peng Q. Research on improving the measurement accuracy of the hall current sensor based on the extreme learning machine. *2023 IEEE 7th Conference on Energy Internet and Energy System Integration (EI2 2023)*, pages 2174–2178, 2023. doi: 10.1109/EI259745.2023.10512708. URL <https://doi.org/10.1109/EI259745.2023.10512708>.

- [259] Wei D., Wu J., Wang K., and Chen R. Analysis and adjust of temperature drift in current measurement based on hall sensor. In *Proceedings of 2022 IEEE 5th International Electrical and Energy Conference (CIEEC 2022)*, pages 1929–1934, 2022. doi: 10.1109/CIEEC54735.2022.9846645. URL <https://doi.org/10.1109/CIEEC54735.2022.9846645>.
- [260] Li H., Xie P., Li C., and Zheng Y. A high accuracy and wide input voltage range current sensor with adaptive amplification and temperature compensation. *2019 IEEE International Conference on Electron Devices and Solid-State Circuits (EDSSC 2019)*, 2019. doi: 10.1109/EDSSC.2019.8753921. URL <https://doi.org/10.1109/EDSSC.2019.8753921>.
- [261] A. Mingotti, L. Peretto, and R. Tinarelli. Effects of multiple influence quantities on rogowski-coil-type current transformers. *IEEE Transactions on Instrumentation and Measurement*, 69(7):4827–4834, 2020. doi: 10.1109/TIM.2019.2953419. URL <https://doi.org/10.1109/TIM.2019.2953419>.
- [262] Tang X., Pan P., Zhang F., Yan L., Lv Y., Wu Y., Rong M., and Teklu B.G. Self-correction sensor for measuring medium dc voltage based on resistance voltage divider. *CSEE Journal of Power and Energy Systems*, 2022. doi: 10.17775/CSEEJPES.2021.00790. URL <https://doi.org/10.17775/CSEEJPES.2021.00790>.
- [263] D Lehtonen T.A. and Hallstrom J. Optimizing temperature coefficient and frequency response of rogowski coils. *IEEE Sensors Journal*, 17(20):6646–6652, 2017. doi: 10.1109/JSEN.2017.2743259. URL <https://doi.org/10.1109/JSEN.2017.2743259>.
- [264] J. Wang, S. Li, X. Yu, R. Zhou, Z. Zhu, and W. Chen. Analysis of influencing factors on current measurement accuracy of pcb rogowski coil. *IEEE Sensors Journal*, 25(6):9531–9539, 2025. doi: 10.1109/JSEN.2025.3530966. URL <https://doi.org/10.1109/JSEN.2025.3530966>.
- [265] N. M. Rodrigues, F. M. Janeiro, and P. M. Ramos. Deep learning for power quality event detection and classification based on measured grid data. *IEEE Transactions on Instrumentation and Measurement*, 72:1–11, 2023. doi: 10.1109/TIM.2023.3293555. URL <https://doi.org/10.1109/TIM.2023.3293555>.

- [266] Fluke Corporation. *6100B/6105A Electrical Power Standards*, 2025. URL https://assets.fluke.com/manuals/6100B___gseng0000.pdf. Accessed: 2025-07-20.
- [267] Fluke Corporation. *Fluke 52120A Transconductance Amplifier*, 2025. URL <https://www.fluke.com/it-it/prodotto/strumenti-per-la-calibrazione/calibrazione-elettrica/calibratori-elettrici/52120a>. Accessed: 2025-10-01.
- [268] E. Aumann, E. Buchta, D. Westermann, and S. Niessen. Novel temperature sensing for switchgear bushings based on the built-in capacitive coupler. *IEEE Transactions on Power Delivery*, 39(3):1975–1982, 2024. doi: 10.1109/TPWRD.2024.3386605. URL <https://doi.org/10.1109/TPWRD.2024.3386605>.
- [269] P. Dopierała. Fault detection method for energy measurement systems equipped with a rogowski coil using the coil’s response to a unit voltage jump and a fully convolutional neural network. *Measurement: Journal of the International Measurement Confederation*, 190:110749, 2022. doi: 10.1016/j.measurement.2022.110749. URL <https://doi.org/10.1016/j.measurement.2022.110749>.
- [270] R. J. Hamidi. Protective relay for power transformers based on digital twin systems. In *2022 IEEE Kansas Power and Energy Conference (KPEC)*, 2022. doi: 10.1109/KPEC54747.2022.9814786. URL <https://doi.org/10.1109/KPEC54747.2022.9814786>.
- [271] W. Zhou, C. Han, Z. Xue, J. Yin, Z. Wang, X. Zhang, and L. Wang. Application of digital twin system in power transformer fault detection. In *2023 IEEE 6th International Electrical and Energy Conference (CIEEC)*, pages 2123–2127, 2023. doi: 10.1109/CIEEC58067.2023.10167284. URL <https://doi.org/10.1109/CIEEC58067.2023.10167284>.
- [272] K. Shi and Y. She. Application of digital twin technology in health assessment and fault diagnosis of intelligent substation. In *2022 International Conference on Wireless Communications, Electrical Engineering and Automation (WCEEA)*, pages 179–182, 2022. doi: 10.1109/WCEEA56458.2022.00043. URL <https://doi.org/10.1109/WCEEA56458.2022.00043>.
- [273] P. Y. Lak, S. Key, S.-M. Yoon, and S.-R. Nam. Digital twin application for the evaluation of protection performance in iec-61850-based digital substations. In *2023*

- International Conference on Advanced Power System Automation and Protection (APAP)*, pages 68–72, 2023. doi: 10.1109/APAP59666.2023.10348444. URL <https://doi.org/10.1109/APAP59666.2023.10348444>.
- [274] A. Mingotti, C. Betti, R. Tinarelli, and L. Peretto. Simplifying rogowski coil modeling: Simulation and experimental verification. *Sensors*, 23(19):8032, 2023. doi: 10.3390/s23198032. URL <https://doi.org/10.3390/s23198032>.
- [275] Microchip Technology Inc. dsPIC33CK Curiosity Development Board, 2025. URL <https://www.microchip.com/en-us/development-tool/dm330030>. Accessed: 2025-06-30.
- [276] National Instruments. NI-9238 C Series Current Input Module. URL <https://www.ni.com/docs/en-US/bundle/ni-9238-specs/page/specs.html?srsltid=AfmBOoqLOS0JegB1rugSaDa0u7T3Qru5635hELqaUfPw5dW5yoNH091t>. Accessed: 2025-07-16.