Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

SCIENZE CARDIO-NEFRO-TORACICHE

Ciclo 31°

Settore Concorsuale: 06/D1

Settore Scientifico Disciplinare: MED/11

TITOLO TESI

AFICILL: a single-cohort, retrospective study on Atrial Fibrillation In Critically ILL patients admitted to a medical sub-intensive care unit: implications for clinical management, outcomes and elaboration of new data-driven models

Presentata da: Lorenzo Falsetti

Coordinatore Dottorato
Prof. Gaetano Domenico Gargiulo

Supervisore
Prof. Francesco Grigioni

Co-Supervisore
Prof. Alessandro Capucci

Esame finale anno 2019

Summary

- **Abstract**

1. **Introduction**
    a. Atrial fibrillation in the critically-ill patient
    b. Medical population in a subintensive care unit
    c. Current use of big data and machine learning in medicine
    d. Aims of the study
2. **Patients and Methods**
    a. Power analysis
    b. Ethics committee approval
    c. Electronic database structure
    d. Definitions
    e. Statistical analysis
    f. Data-driven modelling of new scoring systems
        i. Visualizing multi-dimensional datasets: the t-SNE approach
        ii. Mining insights from multi-dimensional datasets: computational topology and Mapper algorithm
        iii. Statistical methods adopted in topological data analysis
            1. Chi-squared test (χ^2)
            2. Receiver-operating characteristic (ROC)
            3. Jaccard similarities
            4. Machine learning for data-driven modelling
3. **Results**
4. **Discussion**
    a. Study limitations

Abstract

Introduction: atrial fibrillation (AF) is common among critically-ill patients, who are considered at increased cardioembolic and haemorragic risk. Consequently, anticoagulant therapy might be ineffective or harmful for an excess of haemorragic events which could not be counterbalanced by an adequate reduction of cardioembolic occurrences.

Aims: main outcome (MO) was the composite of death or intensive care unit (ICU) transfer in a population of critically-ill subjects admitted to a medical subintensive care unit (sICU); we assessed (i) thromboembolic events (TEE) and major haemorrhages (MH); (ii) current guidelines (GL) adherence and related outcomes; (iii) performance of validated risk scores for TEE and MH; we engineered (iv) new scores adopting machine learning (ML) predicting MO, TEE, MH.

Patients and Methods: single-center, retrospective study enrolling all the consecutive AF-affected patients admitted to a sICU for critical illness. Demographic, clinical, therapeutic and laboratoristic data were collected. Performance of CHA_2DS_2-VASc and HAS-BLED scores was evaluated. GL-adherence and its relationship with outcomes was studied. ML was used to engineer new predictive models.

Results: we enrolled 1430 subjects; CHA_2DS_2-VASc (AUC:0.516;95%CI:0.472-0.560) and HAS-BLED (AUC:0.493;95%CI:0.443-0.543) did not predict TEE or MH; in-hospital warfarin use was associated to increased MO risk (OR:1.73;95%CI:1.06-2.83; $p<0.05$); low-molecular-weight-heparin use was not associated to an increased MO risk; antiplatelet drugs use was associated to MO risk reduction (OR:0.51;95%CI:0.34-0.78;$p<0.002$). GL-adherent treatment was associated to TEE risk reduction and MH and MO risk increase; ML identified specific features for MO, TEE, MH: ML-based classifiers outperformed CHA_2DS_2-VASc (AUC: from 0.516 to 0.90, $p<0.0001$) and HAS-BLED (AUC: from 0.493 to 0.82, $p<0.0001$).

Discussion: AF-related outcomes cannot be predicted in critically-ill patients with currently validated methods. GL-adherence is associated to a significant TEE reduction, but also to

MH and MO increase. ML algorithms can identify the most important features and shape specific scores able to outperform the classical models.

1. Introduction

Atrial fibrillation in the critically-ill patient

Atrial fibrillation (AF) is the most common sustained arrhythmia observed both in general population[1] and in several groups of hospitalized patients[2]. Among critically-ill subjects admitted in intensive care units (ICU), pre-existing and new-onset forms of AF can be observed in 1 out of 3 admitted patients[3].

New-onset AF has a prevalence ranging between 5 and 46%[4] and represents 47.4-61% of all the arrhythmias and 52% of the atrial arrhythmias observed in ICU[2,5,6]. Pre-existing AF follows the same prevalence of the general population[7], and is present in 9% of the patients admitted to ICU for critical illness[8]. Moreover, pre-existing AF has a better-known pathophysiology: it is strongly linked to ageing[9,10] and its associated comorbidities, as chronic heart failure (CHF), diabetes, valvular diseases, acute coronary syndromes (ACS) and hypertension (HYP)[11]. These disorders favour atrial structural and electrical remodelling, offering an ideal arrhythmogenic substrate[12]. Several factors, as electrolytic and volume disturbances, sympathetic and parasympathetic activity alterations are common AF triggers. The association of a substrate and a trigger is able to initiate and maintain AF[12,13].

The classical risk factors and triggers, however, show a weaker association with AF when it occurs during a critical illness[14], and other features seem to be implied in triggering and maintaining new-onset AF[4]. Acute pathologies, presence of organ failure and the activation of the inflammatory systemic response are supposed to induce atrial structural and electric remodelling[3,14,15]. Beta-agonist and vasopressor drugs, sustained tachycardia, bacterial toxins, neuro-hormonal and electrolyte disturbances, myocardial ischemia and volume overload can trigger a new-onset AF[3,4]. Moreover, proinflammatory cytokines have a direct arrhythmogenic effect of on atrial myocardium[16].

Acute occurrence of AF in the setting of a critical illness is often associated to a

deterioration of global haemodynamic due to fast and irregular ventricular response rate and to the loss of atrial systole[3,17], but also to an increased risk of stroke[18], acute heart failure (AHF)[19], and death[3].

Several authors showed that AF increased the risk of in-hospital mortality in specific pathologies, as in sepsis[20], trauma[21], ACS[22,23] and AHF[24], but also in generic cohorts of critically-ill patients admitted both in medical[25] and surgical[26] ICU. In the long-term, these patients had a greater risk of rehospitalisation for AHF and stroke and an increased risk of death, which remained high up to 5 years after hospital discharge[27].

Previous reports underlined the association between poorer clinical outcomes and new-onset AF[14]. However, new-onset AF did not always result independently predictive for in-hospital mortality after adjusting for disease severity, suggesting that this type of AF could be a marker of prognosis rather than an independent risk factor[28,29]. Pre-existing AF has been identified as an independent risk factor for in-hospital mortality and worse functional outcomes in ACS, AHF[22,30,31] and in generic ICU populations[8].

Despite the epidemiologic and prognostic relevance of the problem, the clinical management of AF during a critical illness is still object of debate. Currently, studies underline no benefit of a rhythm control over a rate control strategy in the critically-ill patient, and do not allow to generate any recommendation for a standard treatment[32], except for the cases of cardiogenic shock due to elevated cardiac frequency, where urgent electric cardioversion is mandatory[1]. Medications adopted for both rate and rhythm control are poorly evaluated in the setting of a critical illness and several studies underlined a significant practice variability in the prescription of different rate-control drugs[3]. Even the correct dose of commonly used medications, as amiodarone or magnesium sulphate, is currently under investigation in the specific clinical setting of ICU and in severe sepsis (clinicaltrials.gov ID NCT01049464; clinicaltrials.gov ID: NCT02668432).

Critical illness itself represents a procoagulant state, and the coexistence of AF implies a markedly increased thromboembolic risk[33]. This increased risk has been demonstrated for new-onset AF in severe sepsis[18], ACS[34], AHF[35] and acute respiratory failure (ARF)[36]. Similarly, pre-existing AF has been associated to an increased stroke risk in ICU patients[8]. Anticoagulation in this subset of patients, however, has not been related to a significant reduction of stroke risk, but to a significant increase of bleeding risk[36,37].

The accuracy of the currently adopted clinical prediction scores in critically-ill subjects is still object of debate: while some authors recommend the use of CHA_2DS_2-VASc score at a different cutoff to stratify the thromboembolic risk in the critically-ill patients[38], others underline its low predictive value[36,39] and emphasise the presence of a very high haemorragic risk which cannot be accurately quantified by HAS-BLED score[36,40].

Medical population of a semi-intensive care unit

The number of critically-ill patients (and their comorbidities) is increasing along with ageing of the population. In order to reduce ICU overcrowding and optimize resources, this subset of patients is often managed in specific Internal Medicine departments, named semi-intensive care units, or sICU. The sICU population usually differs from ICU for a more advanced age of patients but also for an increased number of comorbidities as chronic respiratory failure, chronic kidney disease, dementia, CHF and cancer, with a worse prognostic profile and an even higher AF prevalence due to age and associated pathologies.

Current use of big data and machine learning in medicine

Medical informations are now collected continuously at the bedside: demographic data, clinical informations, pharmacological therapy, physiological signs, laboratory analysis and radiologic data can now be easily collected, stored and analysed. Thus, a *big data* repository is usually defined by five “Vs”: volume (large quantity of data), velocity (high speed of acquisition), variety (difference of the data sources), veracity (uncertainty of data quality) and

value (possible valorisation of the data)[41]. Both ICU and sICU, for their technological implementation, represent the ideal environment where to collect and analyse this type of informations.

Due to its multidimensionality, big data analysis cannot be adequately performed with the classical statistical methods: several machine learning (ML) techniques are currently used to explore hidden relationships between different variables. This process is done automatically, but a human supervision is often necessary to clarify the results and avoid spurious interpretations.

Techniques based on a specific set of algorithms, named *topological data analysis* (TDA), are commonly used to explain relationships between variables in large datasets, especially in critical bio-medical and medical phenomena. TDA has been successfully applied in medical studies regarding cancer[42], simulated human immune systems dynamics[43] and pulmonary embolism (PE)[44]. In the setting of TDA, our group already used an hypernetwork approach and Q-analysis to identify informative medical features and instruct an artificial neural network to predict automatically the pretest probability of PE[44]. Hypernetworks provide a significant generalization of network theory, enabling the integration of relational structure, logic and analytic dynamics. With this novel approach, the resulting neural hypernetwork correctly recognized 94% of the patients affected by PE before the CT-scan. In other studies in the same dataset, we identified key features which were best associated to PE diagnosis[45,46] to engineer a ML algorithm which was able to outperform the classical methods, represented by Wells and Geneve scores.

Aims of the study

Objective of Atrial Fibrillation In Critically ILL (AFICILL) study was to evaluate the occurrence of the main outcome (MO), defined as death or ICU transfer, in a single-cohort, retrospective study of critically-ill patients affected by AF and admitted to a medical sICU. We also aimed to:

- evaluate risk factors, comorbidities and concurrent clinical events significantly and independently associated to MO;
- assess the prevalence of cardioembolic events (TEE) and major haemorrhage (MH) as main clinical adverse events associated to AF in this setting;
- appraise the association and the predictive capacity of CHA_2DS_2-VASc for TEE and HAS-BLED for MH in this cohort;
- estimate the adherence to the indications of the European Society of Cardiology (ESC) AF guidelines[1] for anticoagulant therapy and assess the association between guidelines adherence (GL) and adverse clinical events;
- identify risk factors for MO, TEE and MH in this population of patients and generate new predictive models adopting a TDA-based, ML algorithm.

2. Patients and Methods

Study cohort and baseline characteristics

In order to evaluate the study objectives, we retrospectively analyzed a cohort of critically-ill patients with AF admitted to the internal and sub-intensive medicine department of the Azienda Ospedaliero-Universitaria "Ospedali Riuniti" in Ancona, Italy. Since January 01st 2002 the department adopted an electronic medical record (eMR) system for inpatients' management, that allows to interrogate the main database to select patients characterized by a specific diagnosis. All diagnoses in the eMR are coded according to ICD-9 system. In the aim of the study, we selected all patients admitted to the sICU with a concurrent diagnosis of AF (ICD-9: 427.31) from inception to 31/03/2018, then we randomly decided "a priori" to select the first consecutive 25% of the entire AF cohort to include in the study, in order to keep the data collection timely and effective. Afterward, we excluded all patients admitted performing a planned cardioversion procedure for AF rhythm control, stable patients without acute organ failure and patients admitted for trauma or surgical pathologies in order to keep only the medical, critically-ill patients. We then obtained from the eMR and from the detailed examination of discharge reports all the data regarding demographics, history of risk factors and comorbidities, admission diagnoses, concurrent clinical events and use of antithrombotic drugs.

Power Analysis

According to literature, mortality due to critical illness in ICU is estimated at 15% of the population. Setting *a priori* an alfa of 0.050 in a two-tailed test, we estimated that a sample size of 1430 patients was able to establish this outcome with a precision of ±2% (95%CI).

Ethics Committee Approval

The local ethics committee (CERM, Comitato Etico Regione Marche, Azienda Ospedaliero-Universitaria "Ospedali Riuniti", Ancona, Italy) reviewed the protocol and approved the study (protocol number: 2018/168, 21/06/2018, see Appendix 1).

Electronic Database Structure

The database structure, the format of the collected variables, their content, their names and abbreviations are synthesized in Table 1. We collected the unique identifier, the number of admissions and the patient's age at the admission as continuous variables. Main outcome, sex, electric cardioversion (CVE), pharmacologic cardioversion (CVF), the reason of sUTI admission (syncope, trauma, acute coronary syndrome, acute heart failure, cardiogenic shock, haemorragic shock, septic shock, acute kidney injury and acute respiratory failure), comorbidities (chronic heart failure, chronic obstructive lung disease, peripheral artery disease, previous stroke/TIA, chronic hepatic pathology, chronic kidney disease, chronic ischemic cardiopathy, diabetes mellitus, chronic anaemia, hypertension, active cancer, alcohol abuse, mitral valve disease and aortic valve disease), previous gastrointestinal (GI) bleeding, low time in therapeutic range (TTR), acetylsalicylic acid/clopidogrel use were categorized as binary variables.

Type of AF, type of MH, type of TEE, anticoagulant therapy at admission, anticoagulant therapy at discharge, acute neurologic syndromes, infections and the number of comorbidities were collected as categorical variables.

Table 1: Database structure and type of collected variables

Variable Name	Content	Format
DATE	Date of admission	dd/mm/yyyy
ID	Unique identifier	continuous
ADM	Number of admissions	continuous

SEX	Patient's sex	binary
AGE	Patient's age at the admission	continuous
MO	In-hospital death or ICU transfer 0: Discharged 1: Death in sUTI 1: UTI transfer	binary
AF_TYPE	Atrial fibrillation type 1: Paroxysmal 2: Persistent 3: Permanent	categorial
MH_TYPE	Major haemorrhage type 1: ICH/ESA 2: Gastrointestinal bleeding 3: Urinary tract bleeding 4: Intramuscular bleeding 5: Other	categorial
TEE_TYPE	Cardioembolic event type 1: Stroke/TIA 2: Atrial appendage thrombus 3: Systemic embolization 4: Lower limb ischemia (embolic)	categorial
CVE	Electric cardioversion	binary
CVF	Pharmacologic cardioversion	binary
AC_ING	Anticoagulant therapy at admission	categorial

	0: Warfarin 1: LMWH 3: No Anticoagulant	
AC_DISMISS	Anticoagulant therapy at discharge 0: Warfarin 1: LMWH 3: No Anticoagulant	categorial
Critical Illnesses at sICU admission		
SYN	Syncope	binary
AC_NEUR	Acute neurologic syndromes	binary
TRAUMA	Trauma	binary
ACS	Acute coronary syndromes	binary
AHF	Acute heart failure	binary
CS	Cardiogenic shock	binary
HS	Haemorragic shock	binary
SS	Septic shock	binary
AKI	Acute kidney injury	binary
ARF	Acute respiratory failure	binary
INF	Infection 1: Pneumonia 2: Abdominal infections 3: Urinary tract infections 4: Other	categorial

Comorbidities		
CHF	Chronic heart failure	binary
COPD	Chronic obstructive lung disease	binary
PAD	Peripheral artery disease	binary
STROKE_TIA	Previous stroke or TIA	binary
CHP	Chronic hepatic pathologies	binary
CKD	Chronic kidney disease	binary
CCS	Chronic ischemic cardiopathy	binary
T2DM	Type 2 diabetes mellitus	binary
CA	Chronic anaemia	binary
PREVIOUS_BLEED	Previous gastrointestinal bleeding	binary
HYP	Hypertension	binary
AC	Active cancer	binary
COMORBIDITIES	Number of comorbidities	categorial
LOW_TTR	Low time in therapeutic range	binary

ASA_CLOP	Acetylsalicylic acid or Clopidogrel use	binary
ALC_ABUSE	Chronic alcohol abuse	binary
MV_PAT	Mitral Valve Disease	binary
AO_PAT	Aortic valve disease	binary
CHA2DS2-VASc	CHA_2DS_2-VASc Score	categorial
HASBLED	HAS-BLED Score	categorial
CHADS2	$CHADS_2$ Score	categorial

Legend: ICU=intensive-care unit; sICU= subintensive-care unit; ICH= intracranial hemorrhage; ESA= subarachnoid hemorrhage; TIA= transient ischemic attack; LMWH= low molecular weight heparin.

We calculated $CHADS_2$, CHA_2DS_2-VASc and HAS-BLED score following their original definitions [47–49], as shown in Table 2 and Table 3 and collected these scores as categorical variables.

Table 2: $CHADS_2$ and CHA2DS2-VASc score

$CHADS_2$ Score		
Item	Meaning	Score
C	CHF History	+1
H	Hypertension History	+1

A	Age >75	+1
D	Diabetes Mellitus	+1
S_2	Previous Stroke / TIA / Thromboembolism	+2
CHA_2DS_2-VASc Score		
Item	Meaning	Score
C	CHF History	+1
H	Hypertension History	+1
A_2	Age: >75	+2
D	Diabetes Mellitus	+1
S_2	Previous Stroke / TIA / Thromboembolism	+2
V	Vascular disease History	+1
A	Age: 65-74	+1
Sc	Sex Category (i.e. Female Sex)	+ 1

Legend: CHF= chronic heart failure; TIA= transient ischemic attack.

Table 3: HAS-BLED score

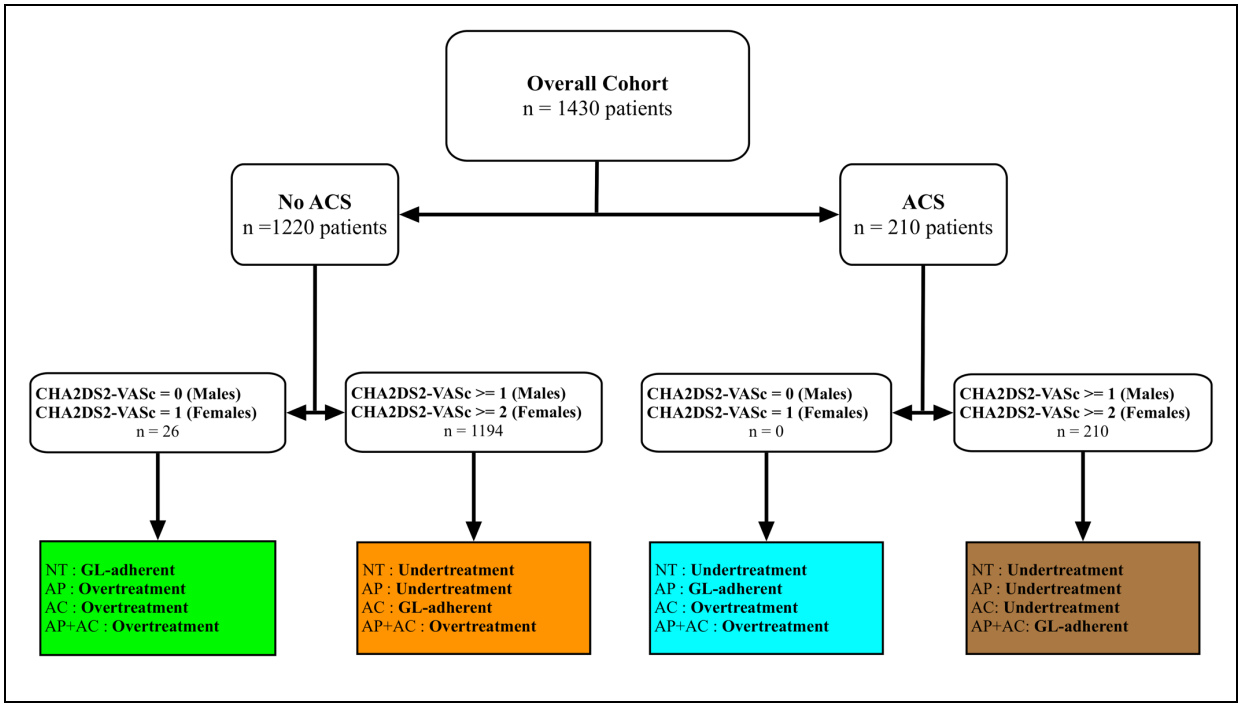
Item	Meaning	Score
H	Hypertension	+1
A	Age > 65	+1
S	Stroke History	+1

	Labile INR (TTR < 60%)	+1
B	Prior Major Bleeding or Predisposition	+1
L	Liver Disease	+1
	Renal Disease	+1
E	Alcohol Use	+1
D	Drugs (ASA, NSAIDS)	+1

Legend: INR= international normalized ratio; TTR= time in therapeutic range; ASA= acetyl salicylic acid; NSAIDS= non-steroidal anti-inflammatory drugs.

Last, we subdivided the overall sample according to GL-adherence. For patients without ACS we considered the indications in the 2016 ESC guidelines on AF[1]. For patients admitted with ACS, we also considered the indications in latest ESC GL on ACS[50,51]. Patients were divided into GL-adherent, overtreated or undertreated according their admission diagnosis and therapy, as shown in Figure 1.

Figure 1: Subdivision of the sample according to current GL adherence



Legend: NT= no treatment; AP= antiplatelet therapy; AC= anticoagulant therapy; AP+AC= antiplatelet and anticoagulant therapy; GL= guidelines

Definitions

- *critically-ill patient:* we defined as "*critically-ill*" all the subjects who – at the admission – had one or more medical conditions at high risk of death, following the MeSH definition, and who were admitted with at least one acute organ dysfunction.
- *main outcome (MO)*: we intended to evaluate the rate of therapeutic failure, defined as in-hospital mortality or ICU transfer.
- *major bleeding (MH)*: we collected all the major haemorragic events defined according ISTH definition of MH in non-surgical patients[52], intended as (i) fatal bleeding, and/or (ii) symptomatic bleeding in a critical area or organ (intracranial, intraspinal, intraocular, retroperitoneal, intra-articular or pericardial, or intramuscular with compartment syndrome), and/or (iii)

bleeding causing a fall in haemoglobin level ≥ 20 g/L, or leading to transfusion of ≥ 2 units of whole blood or red cells. Particularly, we classified MH into the following subclasses:

  - *intracranial haemorrhage (ICH) or subarachnoid haemorrhage (ESA)*
  - *gastrointestinal bleeding*
  - *urinary tract bleeding*
  - *intramuscular or retroperitoneal bleeding*
  - *other sites*

- *thromboembolic event (TEE)*: we recorded all the ischemic events with a presumable cardioembolic source during AF as stroke, TIA or systemic embolization.
  - *stroke/TIA:* we enrolled all subjects with stroke or transient ischemic attack (TIA) where AF was the most probable source of embolism, after evaluating all other causes of non-AF cardioembolic stroke[53].
  - *atrial appendage thrombus:* enrolled subjects undergoing to urgent electric or pharmacologic cardioversion were submitted to trans-oesophageal echocardiography, and the presence of thrombi in the atrial appendages was recorded.
  - *embolic limb ischemia:* acute ischemia of embolic origin appearing in lower or upper limbs.
  - *systemic embolization:* presence of synchronous embolization in multiple sites (visceral, limb and cerebral).
- *atrial fibrillation (AF):* all patients underwent electrocardiogram at the admission in our sICU; we admitted all patients showing the typical AF pattern, as defined by ESC Guidelines 2016: “absolutely irregular RR intervals and no

discernible, distinct P waves"[1]. We excluded from the study all the cases where the cause of the arrhythmia was deemed to be associated to a valvular disease (such as mechanic valves or severe mitral disease). We deemed as diagnostic the documented episodes lasting at least 30 seconds. According to clinical history, we classified AF into three subclasses[1]:

  - *paroxysmal:* events self-terminating within 48 hours after admission or cardioverted within 7 days from onset.
  - *persistent:* events lasting more than 7 days or undergoing cardioversion after 7 or more days.
  - *permanent:* events in which a rate-control approach was preferred to a rhythm-control approach.

- *electrical cardioversion (CVE), pharmacologic cardioversion (CVF):* patients who were selected for rhythm-control strategies underwent to CVE or CVF, according current guidelines[1]. A treatment strategy, drug or procedure was not preferred over another.
- *anticoagulant therapy at admission (ADM), anticoagulant therapy at discharge (DIS):* we recorded all the anticoagulant therapies at admission and at the discharge of each subject. Direct oral anticoagulants (DOACs) were not available at the time of the study. Particularly, we evaluated the use of:
  - *warfarin*
  - *low molecular weight heparin (LMWH)*
  - *no anticoagulant therapy*
- *reason of admission in sICU:* the acute pathology for sICU admission was synthesized and collected. Of note, more than one acute pathology could be present in the same patient. Particularly, we assessed:

- *syncope (SYN):* patients assessed for transient loss of consciousness in whom a definite diagnosis of cardiogenic cause was ascertained according to the 2009 ESC guidelines on management of syncope[54], which were adopted at the time of the study.
- *acute neurological syndromes (ANS):* subjects admitted for status epilepticus or other life-threatening neurological pathologies, except stroke/TIA, which was recorded as part of TEE.
- *trauma:* patients admitted for major head trauma, blunt or open thoracic trauma, blunt or open abdominal trauma were gathered.
- *acute coronary syndrome (ACS):* ST-elevated or non-ST elevated ACS were diagnosed according current guidelines[50,51] and classified according to the third universal definition of myocardial infarction[55].
- *acute heart failure (AHF)* was defined and diagnosed according current ESC guidelines[56] as a rapid onset or worsening of typical symptoms, as breathlessness, ankle swelling and fatigue, accompanied by signs, as elevated jugular venous pressure, pulmonary crackles and peripheral oedema[56].
- *shock:* was defined as a failure to perfuse or oxygenate vital organs, according MeSH definition, and was subdivided, upon the aetiology, into:
    - *cardiogenic shock (CS)* was identified as a state of end-organ hypoperfusion due to cardiac failure[57].
    - *haemorragic shock (HS)* was defined as a form of hypovolemic shock in which severe blood loss leads to inadequate oxygen delivery at the cellular level[58].

        - *septic shock (SS)* was diagnosed according to 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference definition[59].
    - *acute kidney injury (AKI)* was diagnosed, adopting the modified RIFLE criteria, as a serum creatinine increase ≥ 0.3 mg/dl occurring within a 48-hour period[60].
    - *acute respiratory failure (ARF):* was defined as the acute inadequacy of the lungs to maintain either acceptable blood oxygenation, or to allow a normal arterial blood carbon dioxide levels or both[61] and categorized in type 1 (hypoxemic) or type 2 (hypercapnic).
    - *infection (INF)* was defined by the clinical, radiologic and cultural detection of an infection in a specific organ, and was subdivided into:
        - *thoracic infections (pneumonia, mediastinitis)*
        - *non-surgical abdominal infections (appendicitis, cholecystitis, diverticulitis, peritonitis)*
        - *symptomatic urinary tract infections*
        - *other*
- *comorbidities*: we also investigated the presence of one or more associated chronic pathologies in each enrolled patient.
    - *chronic heart failure (CHF)* was diagnosed according current guidelines[56] and defined as "a clinical syndrome characterized by typical symptoms (e.g. breathlessness, ankle swelling and fatigue) that may be accompanied by signs (e.g. elevated jugular venous pressure, pulmonary crackles and peripheral oedema) caused by a structural

and/or functional cardiac abnormality, resulting in a reduced cardiac output and/or elevated intracardiac pressures at rest or during stress".

- *chronic obstructive lung disease (COPD)* was diagnosed according current guidelines[62] and defined as persistent respiratory symptoms and airflow limitation due to airway and/or alveolar abnormalities.
- *peripheral artery disease (PAD)* was diagnosed by physical examination, history and echo-colour doppler and defined as a chronic tissue hypoperfusion due to atherosclerosis of extracranial carotid and vertebral, mesenteric, renal, upper and lower extremity arteries[63].
- *previous stroke/TIA*: was defined as a history of stroke or TIA.
- *chronic hepatic pathologies (CHP)* were defined as the presence of cirrhosis of any cause or chronic infection by HBV, HCV or other hepatotropic viruses.
- *chronic kidney disease (CKD)* was defined as kidney damage or glomerular filtration rate (GFR) <60 mL/min/1.73 m2 for 3 months or more, irrespective of cause[64]. eGFR was estimated with Cockroft-Gault formula.
- *chronic ischemic cardiopathy (CCS)* was diagnosed in presence of a history of myocardial infarction or chest discomfort (angina pectoris)[65].
- *type 2 diabetes mellitus (T2DM)* was diagnosed in presence of a history of T2DM and/or anti-diabetic therapies at the admission.
- *chronic anaemia (CA)* was diagnosed when a chronic reduction of haemoglobin below 13.7 g/dl in men and below 12.2 g/dl in women[66].
- *previous gastrointestinal (GI) bleeding:* was defined as a history of upper or lower GI bleeding confirmed by appropriate endoscopic studies.

- *hypertension (HYP)* was identified in presence of a history of poorly controlled hypertension and/or anti-hypertensive therapy use at the admission with an history of poor blood pressure control. Due to the frequent alterations of blood pressure values in the critically-ill patient, we did not consider the blood pressure values during hospitalization for diagnostic purposes.
- *active cancer (AC)* was diagnosed in presence of history, physical examination and laboratoristic/instrumental exams suggestive for active cancer at the admission of the patient.
- *low time in therapeutic range (TTR)* the quality of anticoagulation in the 12 months preceding the hospitalization in patients anticoagulated with warfarin was evaluated with TTR, calculated with the Rosendaal method[67]: TTR was defined labile for values < 60%, as required by the HAS-BLED score.
- *use of acetylsalicylic acid or clopidogrel (ASA_CLOP)* was defined by the use, at the time of admission, of *acetylsalicylic acid*, clopidogrel or other antiplatelet drugs.
- *alcohol abuse (ALC)* was defined as chronic abuse of alcoholic substances.
- *mitral valve disease (MVP) or aortic valve disease (AVP):* were diagnosed if at echocardiogram were present mitral valve stenosis and/or insufficiency and/or aortic valve stenosis and/or insufficiency. We excluded from the present study all the patients in whom the cause of AF was attributable to the valvular pathology.

Statistical analysis

Continuous variables were reported as mean and standard deviation (SD) or median and interquartile range (IQR), according to a normal or non-normal distribution of values. The statistical difference between subgroups has been evaluated, respectively, with t-test (2 groups, normal distribution), ANOVA test (more than 2 groups) or with Mann-Whitney U test (2 groups, non-normal distribution) and Kruskal-Wallis ANOVA test (more than two groups). Categorial variables were reported as number of subjects and their proportion. Differences between subgroups have been evaluated with χ^2 test.

The association of the clinical factors with the main outcome was evaluated with logistic regression analysis. Demographic and clinical variables was assessed first with univariate analysis: all the variables associated with the main outcome with a p value equal or less than 0.010 were included in the multivariate analysis. The association between risk scores and adverse clinical events was assessed with a logistic regression model developed with the same methodology adopted to study the relationship between clinical factors and the main outcome.

Predictivity of risk scores was evaluated adopting the c-statistics. We calculated also sensitivity (Se), specificity (Sp), positive predictive value (PPV) and negative predictive value (NPV) adopting the standard cut-offs. Receiver-operator curves (ROC) were calculated in 2D graphs considering the true positive rate (Se) in the y-axis and the true negative rate (1-Sp) in the x-axis. For each curve, the area under the curve (AUC) was calculated. Differences between curves were calculated accordingly to DeLong method[68].

We considered as statistically significant a p-value less or equal to 0.05 for two-tailed tests. Statistical analysis has been performed with SPSS 25.0 and Microsoft Excel for Mac OSx.

Data-driven modelling of new scoring systems

Visualizing multi-dimensional datasets: the t-SNE approach. The dataset under examination described each patient with 45 categorical and numerical variables synthesizing different clinical aspects of the enrolled patients.

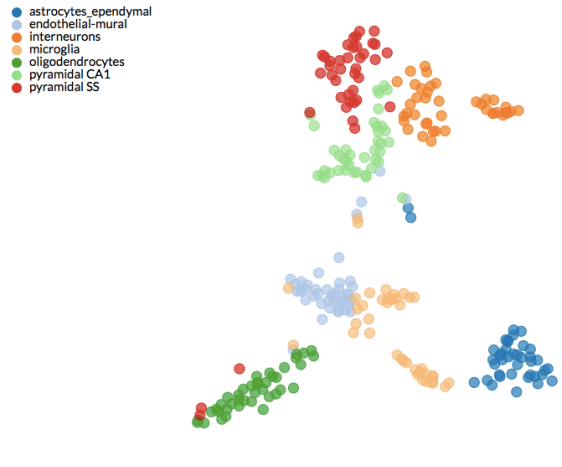
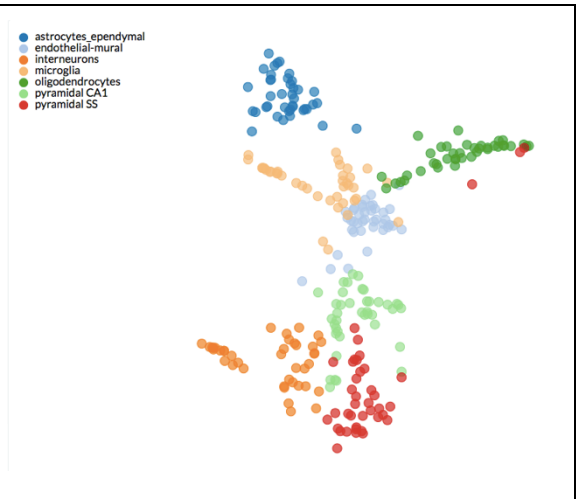
Because of this high dimensionality, the dataset could not be immediately visualized by human brain, which is capable to shape only data belonging to 3 dimensions. A first attempt to visualize the dataset could be of producing a huge amount of 3d scatter plots made by plotting 3 variables at time. This brute-force approach, however, is able to discern only a few and incomplete set of insights. Thus, is important to instruct a more complex but suitable approach for dataset visualization.

Literature suggests different techniques for visualizing high dimensional datasets: for a complete review we referred to [69]. Among these, t-Distributed Stochastic Neighbour Embedding (t-SNE) is one of the mostly used techniques, which is able to visualize datasets up to 50 dimensions. t-SNE is defined as a "non-linear dimensionality reduction algorithm", which finds patterns in the data by identifying observed clusters based on similarity of data points with multiple features. It is important to underline the concept that t-SNE is not a clustering algorithm, but a "dimensionality reduction algorithm", which maps the multi-dimensional data to a lower dimensional space, where the input features are no longer identifiable. Thus, it is not possible to make any inference using t-SNE output only which is conceptually a data exploration and visualization technique. t-SNE maps multi-dimensional data to two or more dimensions, making it suitable for human observation. The algorithm for computing t-SNE considers 4 main steps. Step 1 and Step 2 calculate the conditional probability of similarity between a pair of points in high dimensional space and then in low dimensional space. In Step 3 and Step 4 t-SNE tries to minimize the sum of the difference in conditional probabilities and the algorithm finds the best parameters for retaining the local structure of the data in the map. This algorithm optimizes the search of the "hyper-parameter" in regard to the so-called perplexity. Perplexity is inputted by the user: this value can be

interpreted as a smooth measure of the effective number of neighbours. Typical values are between 5 and 50. For a complete technical overview of this methodology we refer to [70,71]. In Table 4 we report a comparison by using different perplexity values on the same dataset[72].

Medical studies often use principal component analysis (PCA), which is a long-standing technique for data visualization and reduction. However, PCA has some limitations: it is a linear algorithm, and it will not be able to interpret complex non-linear relationship between features. A major problem with linear dimensionality reduction algorithms is that they concentrate on placing dissimilar data points far apart in a lower dimension representation. However, in order to represent high dimensional data on lower dimensions it is important that similar data points must be represented close together, which is not what linear dimensionality reduction algorithms do. For a complete technical review of t-SNE we refer to [73].

Table 4: example of t-SNE: visualization of cell types in the mouse cortex and hippocampus.

	
Perplexity 30: classes are well grouped except for the endothelial	Perplexity 50: the endothelial is now more grouped

Mining insights from multi-dimensional datasets: computational topology and Mapper algorithm: t-SNE retains both the local and global structure of the data at the same

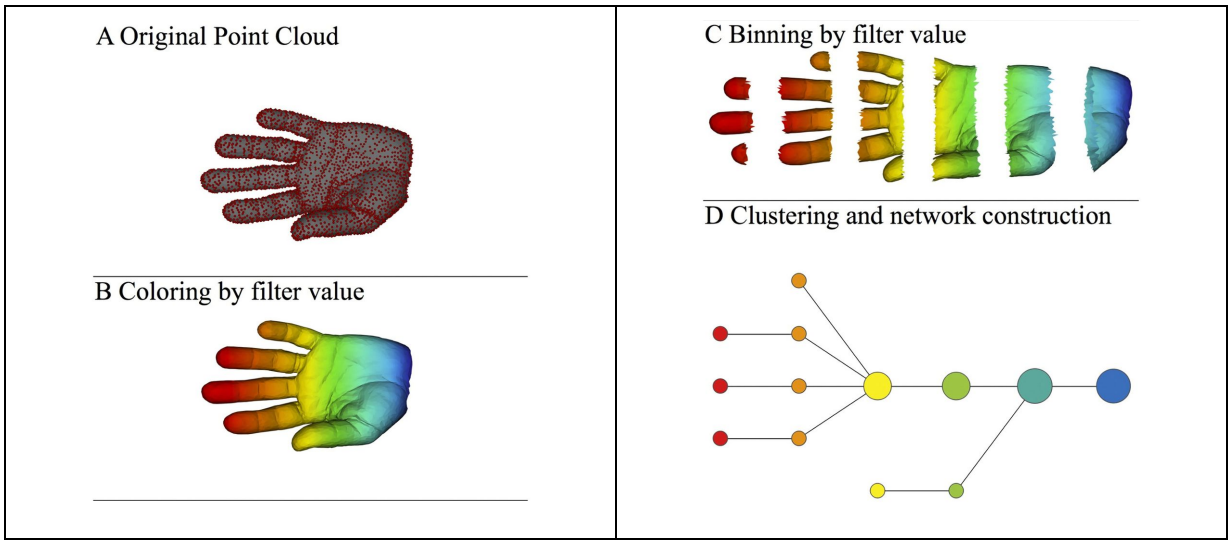
time but it does not care of the meta-scale proximities among data. Meta-scale is the dimension between micro and macro. Topology is the branch of mathematics that aims to study the shapes and the maps among them. A topological space is an abstract space equipped with some notions of similarities. There are several ways for building a topological space, we are interested to the ones obtained by using the so-called simplicial complex.

Simplicial complex is the most suitable construction of topological space even they are combinatorial objects that can be easily constructed and studied by software systems. An abstract simplicial complex is the subset of the power set of a vertex set. For example, given the vertex set V = {0,1,2}, the power set 2V of V is 2V ={{⌀},{0},{1},{2},{0,1},{0,2},{1,2},{0,1,2}}, and one simplicial complex can be {0,1,2}.

Instead of dealing with an abstract simplicial complex, the researcher could be interested in geometrical construction: a simplicial complex is obtained by nesting together small pieces, known as simplices. The most common simplices are labelled as follows: 0-simplex is represented by a vertex, 1-simplex is represented by an edge, 2-simplex is represented by a filled triangle, 3-simplex is represented by a filled tetrahedron. During the construction of the final simplicial complex only a constrain must be respected: the intersection between two simplices must be the empty set or must be proper, meaning that they must share all their simplices or at least one simplices of dimension less the dimension of the whole simplices. A new set of algorithms for the construction of simplicial complexes and their analysis has been derived from algebraic topology and they are known as *topological data analysis*. TDA is sensitive to both large- and small-scale patterns that often fail to be detected by other analysis methods, such as principal component analysis, (PCA), multidimensional scaling, (MDS), and cluster analysis.

This technique is able to explore and synthesize the relationships between large sets of data and is nowadays largely used for exploratory data mining in big data studies. An example of TDA workflow is shown in Figure 3.

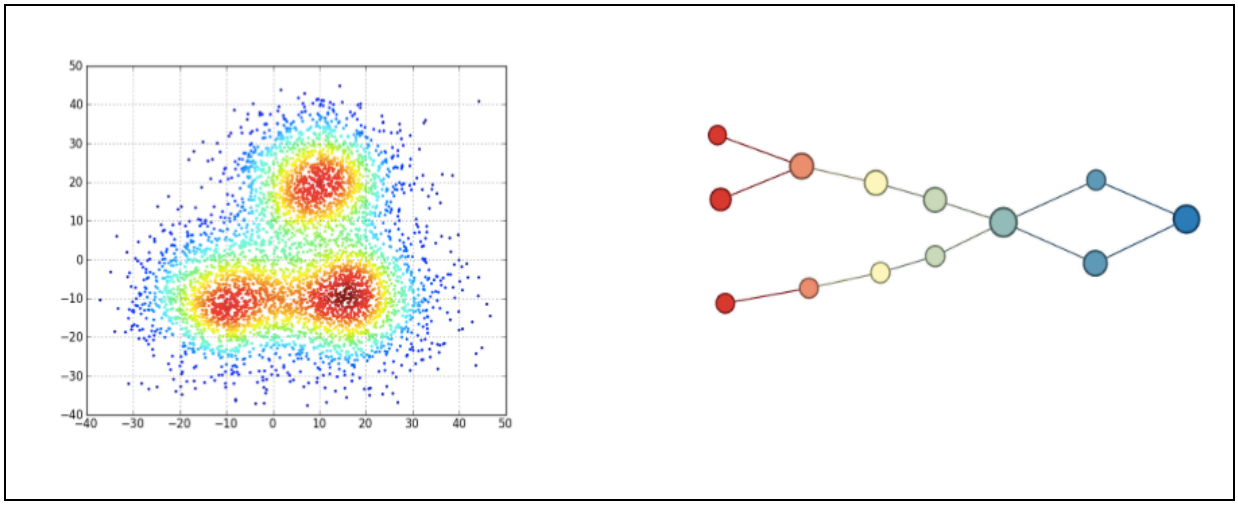
Figure 3: (A) a 3D object (hand) is represented as a point cloud. (B) a filter value is applied to the point cloud and the object is now coloured by the values of the filter function. (C) the data set is binned into overlapping groups. (D) each bin is clustered and a network is built. Picture and caption from [74].



TDA can be derived in three main classes of algorithms: *persistent homology*, *hypernetwork* and *mapper*. *Homology* is an algebraic machinery that counts the number of holes in a simplicial complex. *Persistent homology* is the computational implementation of homology. Persistent homology builds simplicial complexes from data in an iterative fashion. If the input dataset is a point cloud data, a Vietoris-Rips algorithm or equivalent is used to build the simplicial complex. While, if the input dataset is a network then the completion to a simplicial complex is obtained by clique weight rank persistent homology. Once the simplicial complex is obtained its homology is studied. Persistent homology takes as input the list of simplices within a simplicial complex and iterates over. At each iteration simplicies are added to the topological space and the number of n-dimensional holes is computed. Holes that are found at the end of the process are labelled as persistent, the other are classified as noise.

Mapper builds a 1-dimensional simplicial complex from data. A 1-dimensional simplicial complex is obtained connecting together vertices (nodes) with edges. This structure coincides with a graph. A graph G is a set of nodes V and a set of edges E: G=(V,E) where E⊆VxV. Before recalling the technical details of the Mapper algorithm, we provide an example in Figure 4.

Figure 4: An example of Mapper with artificially generated 2D point cloud data



The example in Figure 4 shows Mapper output from an artificially-generated point cloud data. The data in the example consists of 5000 points randomly generated from a Gaussian distribution surrounding three centroids at [x, y] coordinates: [10, 20], [-10, -17], [17, -10]; with a standard deviation of 9. The simplicial complex (right) contains a flare. The top arm ends with another flare made by two nodes indicating the two clusters in the bottom of the picture. The second arm ends with a node indicating the upper cluster in the picture. The method consists of a number of steps, given a point cloud with N points $x \in X$ (Figures 3 and 4):

1. We start with a function $f: X \rightarrow \Re$ whose value is known for the N data points. We call this function a *filter*. The function should convey some interesting geometric or other, properties of the data, relevant for the task at hand.

2. Citing from [75]: "Finding the range (I) of the filter f restricted to the set X and creating a cover of X by dividing I into a set of smaller intervals (S) which overlap. This gives us two parameters which can be used to control resolution namely the length of the smaller intervals (l) and the percentage overlap between successive intervals (p)".
3. Citing from [75]: "Now, for each interval $I_j \in S$, we find the set $X_j = \{x | f(x) \in I_j\}$ of points which form its domain. The set Xj forms a cover of X, and $X \subseteq \bigcup_j X_j$".
4. Choosing a metric d(-,-) to get the set of all interpoint distances

$$Dj = \{d(xa; xb) | xa; xb \in Xj\}$$

5. For each Xj together with the set of distances Dj we find clusters $\{X_{jk}\}$.
6. Each cluster then becomes a vertex in our complex and an edge is created between vertices if $X_{jk} \cap X_{lm} = \emptyset$ meaning that two clusters share a common point.

For a review of the technical details of the algorithm we refer to [75]. In this work we have used the Python language implementation of the Mapper algorithm called *Kepler Mapper* [76].

Statistical methods adopted in TDA

- *Chi-squared test (χ^2)*: feature reduction is the step of reducing the number of features to improve model construction. There are two main approaches for feature reduction: feature selection and feature combination. They can be used together. Feature selection is the process of selecting a subset of relevant and informative variables to be used in model construction. In feature combination the features are combined together (linearly or not) by building a new set of artificial features. Usually in feature combinations the features are weighted by coefficients reflecting features' relevance. Several approaches for feature selection are available, however they rely mainly on statistical tests. χ^2 statistical

tests are widely used for selecting features that form the input space of classifiers. χ^2 is used in statistics to test the independence of two events. Given dataset about two events, we can get the observed count $\boldsymbol{O}$ and the expected count $\boldsymbol{E}$. χ^2 measures how much the expected counts $\boldsymbol{E}$ and observed Count $\boldsymbol{O}$ derivate from each other. n feature selection, the two events are occurrence of the feature and occurrence of the class. If the two events are dependent, we can use the occurrence of the feature to predict the occurrence of the class. We aim to select the features, of which the occurrence is highly dependent on the occurrence of the class: the higher value of the χ^2 score, the more likelihood the feature is correlated with the class, thus it should be selected for model training [77].

- *Receiver-operating characteristic (ROC) in TDA*: in modern medicine, the validity of a dichotomous diagnostic test is determined by sensitivity and specificity. ROC curve is the plot that displays the full picture of trade-off between the true positive rate (Se) and false positive rate (1-Sp) across a series of inherent validity of a diagnostic test. This curve is useful in evaluating the discriminatory ability of a test to correctly pick up diseased and non-diseased subjects and finding the optimal cut-off point to least misclassify diseased and non-diseased subjects. The AUC is a single index for measuring the performance a test. The larger the AUC, the better is the overall performance of the medical test to correctly identify healthy and non-unhealthy subjects. Equal AUCs of two tests represent similar overall performances, however this does not necessary mean that both the curves are identical. We calculated the optimal cutoff value in each ROC curve adopting Youden's J statistic. ROC curves were compared adopting Delong method[68].

- *Jaccard similarities.* The Jaccard Similarity coefficient is a statistic used to compare the similarity and diversity of sample sets. The Jaccard coefficient measures similarity between sample sets, and it is defined as:

$$J(MD, AD) = \frac{|MD \cap AD|}{|MD \cup AD|}$$

Where:

  - MD is the medical doctor diagnosis (0 or 1)
  - AD is the diagnosis obtained with some other approach, e.g. algorithm (0 or 1)

- *Machine learning for data-driven modelling.* ML is often defined as a "field of study that gives computers the ability to learn without being explicitly programmed". It means that a ML algorithm learns from the data a set of parameters necessary for adapting the algorithm to the dataset under analysis. As an example, we will use a simple linear equation, defined as: $y = ax + b$: given a set of pairs of x and y (the so-called *training set*) a ML *regressor* algorithm will be able to extract automatically the proper values of a and b so that, given a new unseen x, it will approximate the equation. Of course, ML is used for more complicated problems, where it is quite impossible to explicitly write the analytical form. With this example we have introduced the concept of "*regressor*" that is an algorithm able for predicting continuous numbers. A second class of algorithms is called *classifier*: the ML algorithm learns from a *training set* to automatically *classify* a new unseen input in a set of classes. ML algorithms learn from data a set of parameters for better approximating the map projecting the inputs with the output(s). However, the selection of which ML algorithm to be used and its partial initial configuration (e.g. number of layers and nodes of an artificial neural network) is demanded to the data scientist experiences. Moreover, a ML pipeline is often composed by a collection of algorithms interacting each other.

Thus, given a dataset the engineering of a ML pipeline in terms of algorithms down-selection and their configuration becomes a daunting task. Recently, a new class of framework called *AutoML* is emerging. *AutoML* systems aim to assist data scientist in the selection of the proper algorithms and their automatic tuning. Among the other, *Tree-based Pipeline Optimization Tool (TPOT)* framework is receiving interest from the data science community. TPOT automatically designs and optimizes ML pipelines for a given problem domain, without any need for human intervention. In short, TPOT optimizes ML pipelines using a version of genetic programming, a well-known evolutionary computation technique for automatically constructing computer. TPOT can deal with both regression and classification problems. In the following, we list the main algorithms handled by TPOT used as methods of our data analysis:

- *Supervised Classification Operators.* Decision Tree, Random Forest, eXtreme Gradient Boosting Classifier, Logistic Regression, and K-Nearest Neighbor Classifier.
- *Feature Preprocessing Operators.* StandardScaler, RobustScaler, MinMaxScaler, MaxAbsScaler, RandomizedPCA, Binarizer, and Polynomial Features.
- *Feature Selection Operators.* VarianceThreshold, SelectKBest, SelectPercentile, Select and Recursive Feature Elimination (RFE).

TPOT developers are very active and new algorithms are added frequently [78].

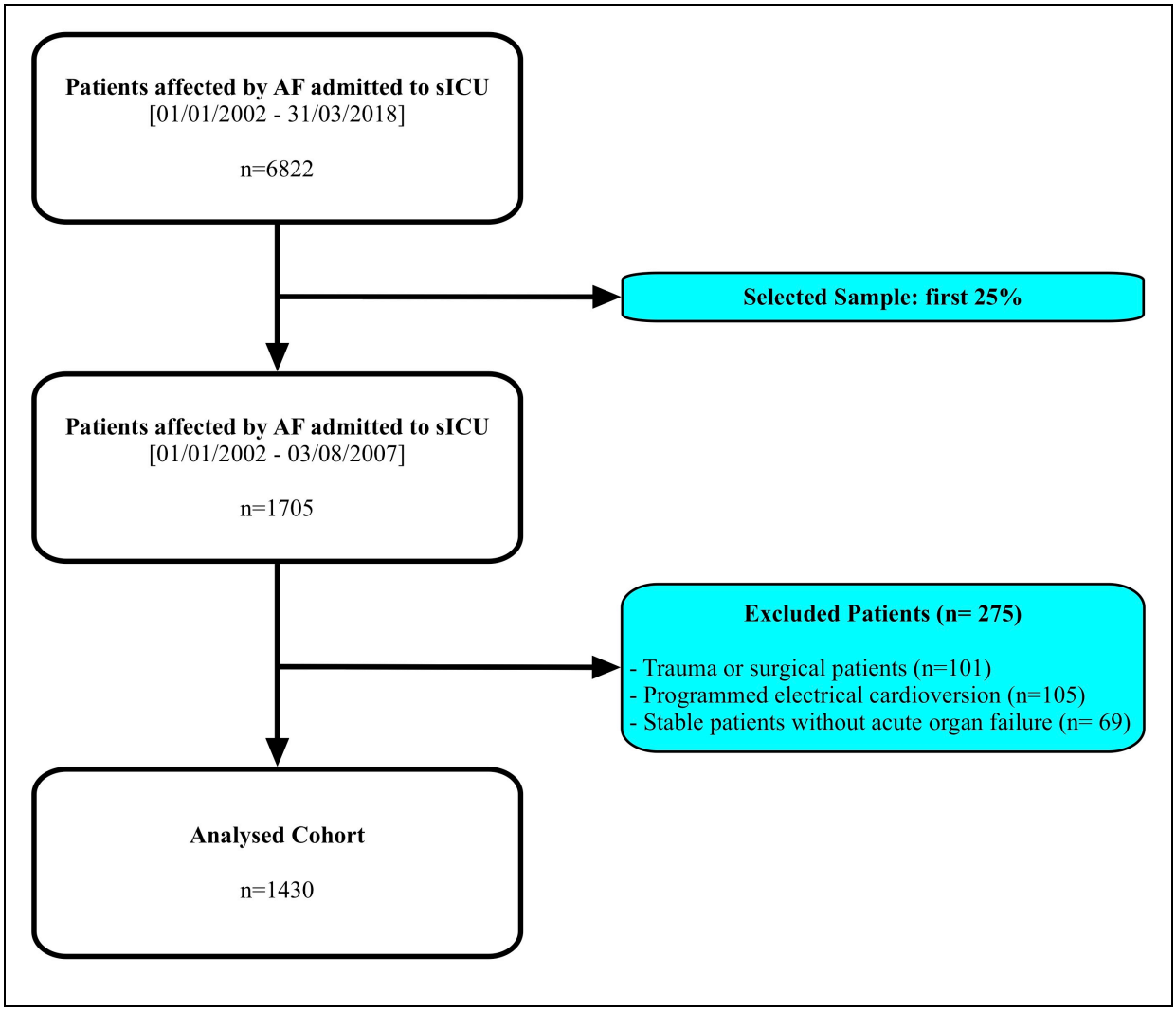
3. Results

Study cohort

The Internal and Subintensive Medicine department of an 800-beds teaching hospital (Azienda Ospedaliero-Universitaria "Ospedali Riuniti", Ancona, Italy) adopted an electronic database for inpatients' management since 01/01/2002; from that date to 31/03/2018, 6822 critically-ill patients affected by AF have been admitted to this unit.

According to the study design, we selected the first quarter, equal to 1705 patients. After excluding those admitted performing a planned cardioversion procedure for AF rhythm control, stable patients without acute organ failure and patients admitted for trauma or surgical pathologies, we obtained a total of 1430 patients, included in the analysis. We synthesized the criteria for patients' selection in Figure 2.

Figure 2: criteria for selection of the analysed sample



Baseline characteristics of the sample

Main outcome was met in 13.6% of the sample (194 subjects). TEE occurred in 14.8% of the cohort (212 patients). We observed MH in 9.30% of the analyzed group (133 subjects). Baseline characteristics of the sample at the admission according MO are synthesized in Table 5.

Table 5: Baseline Characteristics at Admission in Sub-Intensive Unit (sICU)

	MO		**p**
	No	Yes	

	N= 1236	N= 194	
Age, *years* median [IQR]	81 [75-85]	83 [77-89]	<0.001
Female Sex, n (%)	753 (49.9)	90 (46.2)	0.329
Type of AF, n (%)			0.077
Paroxysmal	220 (18.6)	24 (12.8)	
Persistent	292 (24.7)	57 (30.3)	
Permanent	668 (56.6)	107 (56.9)	
Previous Clinical History			
Hypertension, n (%)	637 (51.5)	64 (33.0)	<0.001
Diabetes Mellitus, n (%)	226 (18.3)	34 (17.5)	0.799
Chronic Anaemia, n (%)	114 (9.2)	17 (8.8)	0.836
Coronary Artery Disease, n (%)	531 (43.0)	78 (40.2)	0.471
Peripheral Artery Disease, n (%)	131 (10.6)	16 (8.2)	0.316
Chronic Heart Failure, n (%)	581 (47.0)	88 (45.4)	0.669
CVF, n (%)	183 (14.8)	27 (13.9)	0.745
CVE, n (%)	29 (2.3)	1 (0.5)	0.098
Previous stroke/TIA, n (%)	258 (20.9)	34 (17.5)	0.282
Mitral Valve Disease, n (%)	195 (15.8)	15 (7.7)	0.003
Aortic Valve Disease, n (%)	151 (12.2)	16 (8.2)	0.109
COPD, n (%)	340 (27.5)	54 (27.8)	0.925
Chronic hepatic pathologies, n (%)	35 (2.8)	7 (3.6)	0.551
CKD, n (%)	230 (18.6)	39 (20.1)	0.620
Previous GI Bleeding, n (%)	66 (5.3)	5 (2.6)	0.100

Active Cancer, n (%)	210 (17.0)	44 (22.7)	0.054
Comorbidities, *n* median [IQR]	3 [2-4]	2 [2-3]	0.114
Concurrent Clinical Events			
TEE, n (%)	177 (14.3)	35 (18.0)	0.175
MH, n (%)	110 (8.9)	23 (11.9)	0.188
Syncope, n (%)	68 (5.5)	2 (1.0)	0.007
Acute Neurologic Disorders, n (%)	47 (3.8)	6 (3.1)	0.627
Acute Coronary Syndrome, n (%)	160 (12.9)	50 (25.8)	<0.001
Acute Heart Failure, n (%)	669 (54.1)	99 (51.0)	0.421
Cardiogenic Shock, n (%)	28 (2.3)	43 (22.2)	<0.001
Septic Shock, n (%)	106 (8.6)	74 (38.1)	<0.001
AKI, n (%)	58 (4.7)	10 (5.2)	0.779
Acute Respiratory Failure, n (%)	319 (25.8)	77 (39.7)	<0.001
Infections, n (%)	330 (26.7)	102 (52.6)	<0.001

Legend: AF= atrial fibrillation; ACS= acute coronary syndrome; AKI= acute kidney injury; CKD= chronic kidney disease; COPD= chronic obstructive pulmonary disease; CVE= electrical cardioversion procedure; CVF= pharmacological cardioversion procedure; IQR= interquartile range; TEE= thromboembolic events; TIA= transient ischemic attack.

Thromboembolic and bleeding risk and relationships with outcomes

At baseline, there was no difference in terms of thromboembolic risk according to CHA_2DS_2-VASc score between patients that experience the composite outcome and those that did not experience it. Conversely, HAS-BLED score, as well the proportion of patients with high bleeding risk were lower (both $p<0.001$) patients that reported the main outcome than in those that did not report the outcome (Table 6). Examining the entire spectrum of the

two scores, no difference was found in the distribution of the MO according to CHA_2DS_2-VASc score points (p=0.501). Conversely, the MO occurred more frequently in patients with a lower HAS-BLED score (p<0.001).

Further, considering the occurrence of concurrent clinical events we examined the prevalence of TEE according to CHA_2DS_2-VASc score and the prevalence of major bleeding according to HAS-BLED score, finding no significant differences across the two scores' points and the occurrence of events (respectively p=0.641 and p=0.479).

Also, we found no association between CHA_2DS_2-VASc score and TEE occurrence and between HAS-BLED score and major bleeding occurrence (Table 10). Similarly, we found no predictive ability of the two scores regarding the respective events (CHA_2DS_2-VASc c-index for stroke/TIA: 0.545, 95% CI: 0.489-0.601; HAS-BLED c-index for major bleeding: 0.503, 95% CI: 0.453-0.554).

We also considered $CHADS_2$ score, which is deemed to be less age and vascular comorbidities dependent than CHA_2DS_2-VASc: however, when tested against TEE occurrence with ROC curve analysis, $CHADS_2$ did not result significantly predictive of events (AUC: 0.513; 95%CI: 0.487-0.539; p >0.05), with performances similar to CHA_2DS_2-VASc (AUC: 0.516; 95%CI: 0.472-0.560; p >0.05). Moreover, when comparing the two scores, the difference between AUCs did not result significantly different in predicting thromboembolic events (difference between areas: 0.00326; p= 0.7108). For this reason, we continued the analyses adopting only the most currently adopted scores, CHA_2DS_2-VASc and HAS-BLED.

Considering high thromboembolic risk, we found a high sensitivity (93.4%, 95% CI: 90.9-99.0%) and a high NPV (95.0%, 95% CI: 87.6-98.1%) of CHA_2DS_2-VASc ≥2 for stroke/TIA, with a low specificity and PPV (Table 10). Regarding the high bleeding risk, intermediate values for both sensitivity (46.6%, 95% CI: 37.9-55.5%) and specificity (57.1%,

95% CI: 54.4-59.8%) were found, while a high NPV (91.3%, 95% CI: 89.8-92.5%) and a very low PPV were reported (Table 10).

Table 6: Thromboembolic and Bleeding Risk at Baseline

	MO		**P**
	No	Yes	
	N= 1236	N= 194	
CHA_2DS_2-VASc, median [IQR]	4 [3-5]	4 [3-5]	0.057
CHA_2DS_2-VASc, mean (SD)	4.28 (1.68)	4.04 (1.72)	0.774
CHA_2DS_2-VASc ≥2, n (%)	1170 (94.7)	180 (92.8)	0.290
HAS-BLED, median [IQR]	2 [2-3]	2 [1-3]	<0.001
HAS-BLED, mean (SD)	2.38 (1.08)	2.07 (0.95)	<0.001
HAS-BLED ≥3, n (%)	563 (45.6)	55 (28.4)	<0.001

Legend: IQR= interquartile range; SD= standard deviation.

Factors affecting the main outcome

Regarding medical therapy, we observed that patients undergoing to the main composite outcome were more likely treated with anticoagulants and less likely treated with antiplatelet drugs at the admission in sICU, as shown in Table 7.

Table 7: Antithrombotic Therapies at Admission in Sub-Intensive Unit

	MO		**P**
	No	Yes	
	N= 1236	N= 194	

Anticoagulant Drugs, n (%)			<0.001
None	454 (36.7)	46 (23.7)	
Any Anticoagulant	782 (63.3)	148 (76.3)	
Type of Anticoagulant, n (%)			0.361
LMWH	312 (39.9)	65 (43.9)	
OAC	470 (60.1)	83 (56.1)	
Antiplatelet Drugs, n (%)	515 (41.7)	55 (28.4)	<0.001

Legend: LMWH= low-molecular weight heparin; OAC= oral anticoagulant.

When analysing at the multivariate logistic analysis the risk factors for the occurrence of MO, we identified - among the reasons for sICU admission - ACS, CS, SS and ARF. Increasing age was positively associated with the outcome, while hypertension and mitral valve disease were negatively associated with MO, as shown in Table 8.

Table 8: Multivariate Logistic Analysis for Composite Outcome Occurrence

	OR	**95% CI**	**P**
Age (per year)	1.04	1.02-1.06	0.001
Hypertension	0.52	0.36-0.76	0.001
Mitral Valve Disease	0.49	0.25-0.94	0.033
ACS	3.40	2.06-5.60	<0.001
Cardiogenic Shock	20.68	11.03-38.78	<0.001
Septic Shock	7.66	4.67-12.56	<0.001
Acute Respiratory Failure	2.34	1.57-3.50	<0.001
Anticoagulant Drugs			

None (reference)	-	-	-
LMWH	1.13	0.68-1.87	0.640
OAC	1.73	1.06-2.83	0.030
Antiplatelet Drugs	0.51	0.34-0.78	0.002

Legend: ACS= acute coronary syndrome; CI= confidence interval; LMWH= low-molecular weight heparin; OAC= oral anticoagulant; OR= odds ratio; TEE= thromboembolic events; TIA= transient ischemic attack.

CHA_2DS_2-VASc and HAS-BLED in the prediction of thrombotic and haemorragic events

The rate of TEE according to CHA_2DS_2-VASc score and MH according to HAS-BLED score are synthesized in Table 9. The distribution of TEE does not differ significantly among CHA_2DS_2-VASc classes and, similarly, the distribution of MH does not significantly differ among HAS-BLED classes.

Table 9: Major Adverse Events Rate according to Risk Scores

CHA_2DS_2-VASc	**TEE [n (%)]**	**HAS-BLED**	**MH [n (%)]**
0	0 (0.0)	0	3 (7.3)
1	4 (6.1)	1	27 (9.9)
2	10 (7.4)	2	41 (8.2)
3	15 (6.1)	3	49 (11.5)
4	27 (8.1)	4	10 (6.2)
5	24 (7.7)	5	3 (11.1)
6	14 (7.5)	6	0 (0.0)
7	13 (12.6)		

8	3 (11.5)
9	0 (0.0)

Legend: TEE= thromboembolic events; MH= major haemorrhage.

We did not observe a significant association between CHA_2DS_2-VASc and the occurrence of stroke/TIA, nor a significant association between HAS-BLED and MH adopting c-statistic, as shown in Table 10.

We adopted the currently suggested cut-offs for both scores. CHA_2DS_2-VASc, when analysed at a cutoff ≥2, had a high sensitivity and a good negative predictive value. HAS-BLED, adopting a cutoff ≥3 had a low sensitivity and specificity, but a good negative predictive value, as shown in Table 10.

We also evaluated the quality of the clinical scores, CHA_2DS_2-VASc and HAS-BLED, used for identifying high-risk patients by comparing them with the ground truth TEE and MH, respectively.

Results are represented and summarized in Figure 5. The classification error, that is 1-accuracy, reports the ration between the number of incorrect classified patients when compared with the overall population.

In our sample, the accuracy of each score was between 49 and 52% (CHA_2DS_2-VASc = 52%, HAS-BLED = 49%), meaning that the two scoring systems had the same probability of classifying or misclassifying TEE or MH in this sample of critically-ill patients.

Table 10: Association between Risk Scores, Major Adverse Events and Predictive Analysis

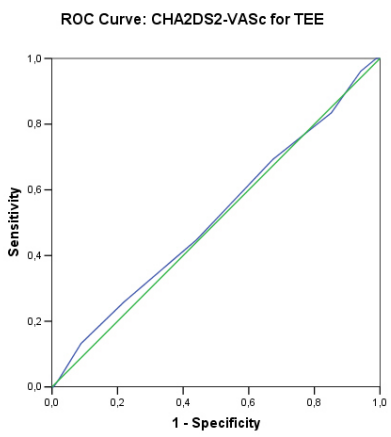
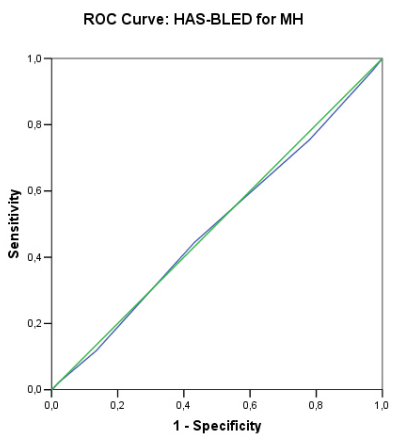
	OR (95 %)*	**P**	**c-index (95%)**	**P**
CHA_2DS_2-VASc	1.09	0.175	0.545	0.117
for TEE	(0.96-1.22)		(0.489-0.601)	

HAS-BLED	1.07	0.477	0.503	0.900
for MH	(0.90-1.27)		(0.453-0.554)	

	Se (95% CI)	**Sp (95% CI)**	**PPV (95% CI)**	**NPV (95% CI)**
CHA_2DS_2-VASc ≥2	93.4%	5.8%	7.8%	95.0%
for TEE	(90.9-99.0%)	(4.6-7.1%)	(6.4-9.2%)	(87.6-98.1%)
HAS-BLED ≥3	46.6%	57.1%	10.0%	91.3%
for MH	(37.9-55.5%)	(54.4-59.8%)	(8.4-11.9%)	(89.8-92.5%)

Legend: **adjusted for type of AF and anticoagulant treatment*; CI= confidence interval; OR= odds ratio; NPV= negative predictive value; PPV= positive predictive value; TIA= transient ischemic attack.

Figure 5: ROC Curve Analysis for CHA_2DS_2-VASc, HAS-BLED and Classification Error

	
ROC curve for CHA_2DS_2-VASC score	ROC curve for HAS-BLED score
	AUC
CHA_2DS_2-VASC	0.516 [95%CI: 0.472-0.560]
HAS-BLED	0.493 [95%CI: 0.443-0.543]

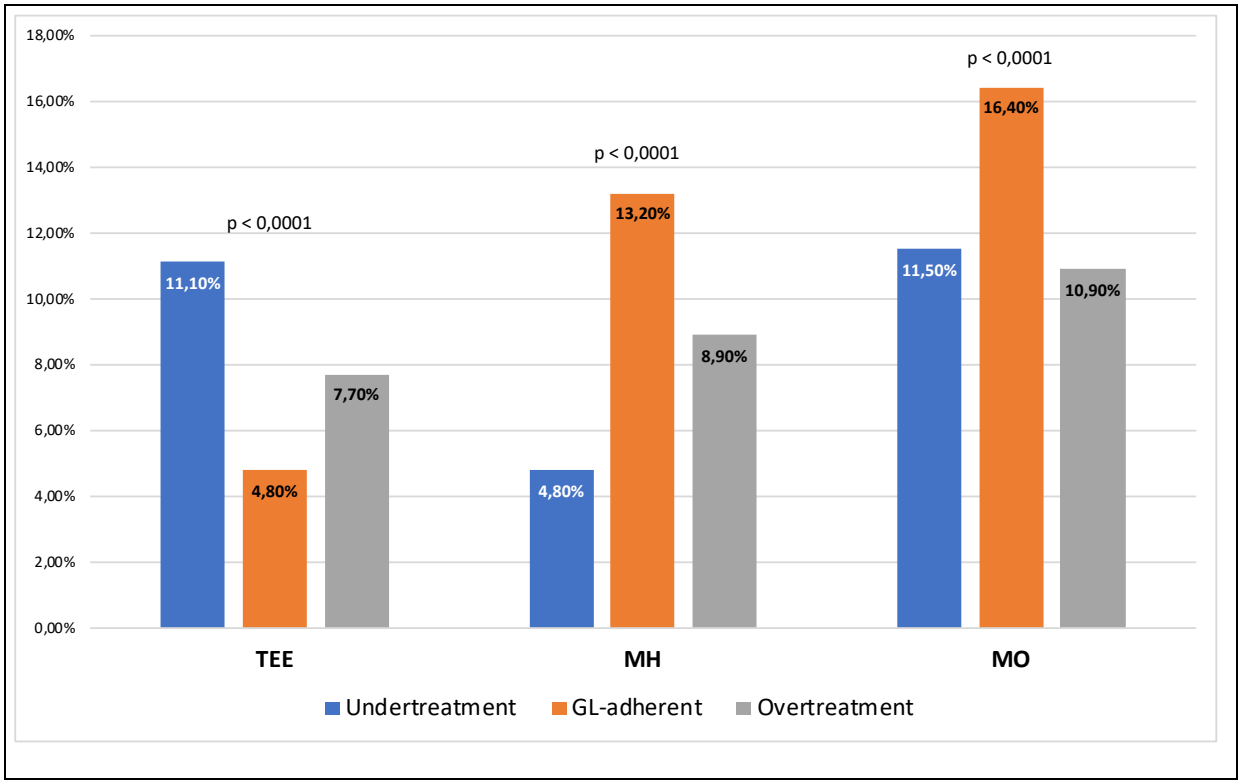
	Classification Error
CHA_2DS_2-VASC	0.963 [95%CI:0.955-0.973]
HAS-BLED	0.923 [95%CI: 0.910-0.936]

Legend: ROC= receiver operating curve; AUC= area under the curve.

Medical treatment and outcomes in the critically-ill patient

On the basis of the ESC 2016 guidelines, we found out that 642 (44.9%) were treated as adherent to the current recommendations, while 540 (37.8%) were undertreated and 248 (17.3%) were overtreated. Analyzing the rate of major adverse outcomes according to guidelines' adherence, while we found that the TEE rate was the lowest in those patients treated as adherent ($p<0.001$) (Figure 6), in the same patients the prevalence of both major bleeding and composite outcome was the higher ($p<0.001$ and $p=0.020$, respectively) (Figure 6). The final multivariable model (Table 11) found out that while undertreatment was associated with an increased risk of TEE, an inverse association with both major bleeding and composite outcome was found out. Conversely, overtreatment only showed a trend with occurrence of TEE, even though did not reach the statistical significance (Table 11).

Figure 6: Events distribution according to current guidelines adherence



Legend: TEE= thromboembolic events; MH= major bleeding; MO= main outcome; GL= guidelines.

Table 11: Multivariable-Adjusted Association between Guidelines Adherence and Major Adverse Events

	TEE		MH		MO	
	OR (95% CI)	**p**	**OR (95% CI)**	**P**	**OR (95% CI)**	**P**
GLs Adherent (ref.)	-	-	-	-	-	-
Undertreated	2.38 (1.45.-3.91)	0.001	0.30 (0.18-0.48)	<0.001	0.63 (0.42-0.97)	0.034
Overtreated	1.75	0.097	0.67	0.143	0.83	0.481

(0.90-3.39)	(0.39-1.15)	(0.48-1.41)

Legend: CI= confidence interval; GLs= guidelines; TIA= transient ischemic attack.

Towards new predictive models in the critically-ill patient

After observing a poor efficiency of CHA_2DS_2-VASc and HAS-BLED scores in predicting TEE as well as MH (at least for the considered dataset), we focused on instructing a new data-driven supervised solution for automatically predicting the three outcomes: TEE, MH and MO.

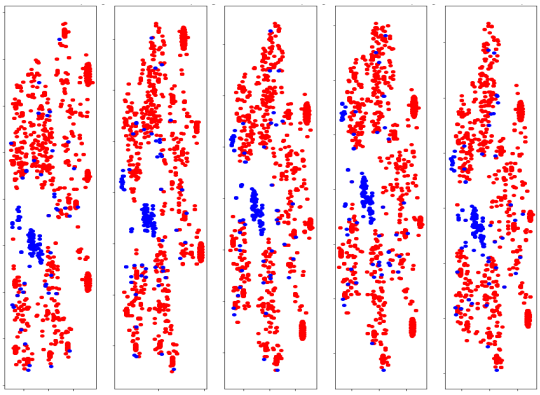
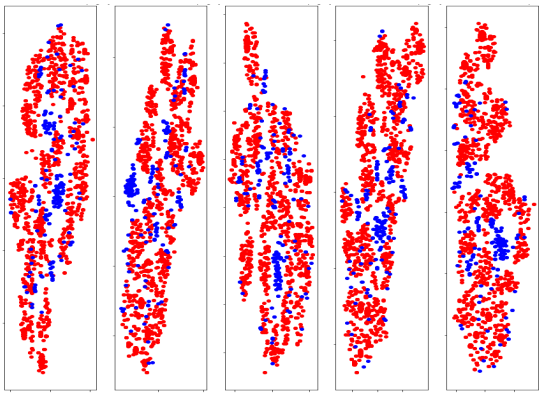
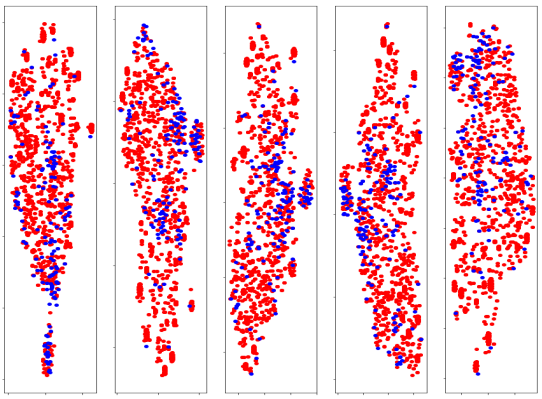
For each target variable, we dropped the other two scores and evaluated the correlation among the clinical variables and target under modeling. Before executing the methodology, we have transformed all the categorical variables in their dummy representation. A dummy variable is an artificial variable created to represent an attribute with two or more distinct categories/levels. The dummy variable represents the original value as a tuple of binary values. For example, if the original categorical variable has n values it will be represented by n-1 new binary variables.

t-SNE results: we evaluated the effect of the perplexities spanning between the values 30-50 with step 5 for the t-SNE algorithm. The t-SNE visualization of the three subsets is reported in Table 12. The red points are patients with target variable equal to 0, while the blue points are the patients with target variable equal to 1. In the plots for the MH and TEE the clusters formed by the blue points are well evident. In the MH plot the blue group is also quite distant from the red cloud indicating a good separation between the two clouds. While, in the last plot representing MO, the blue points are more scattered: at perplexity 35 it is possible to observe two blue subgroups overlapping with red points.

These results encourage the possibility to train a ML classifier to predict automatically the target variables. Due to the scarce separation on the red-blue points in MO,

we hypothesize that for this target variable the ML pipeline will require more steps than the others.

Table 12: Comparison of t-SNE for different target variables with perplexity values from 30 to 50 with step 5.

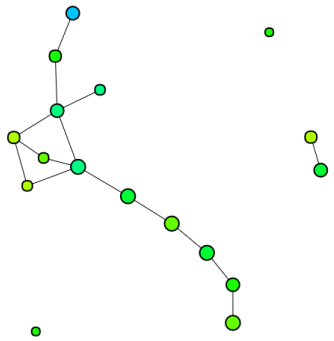
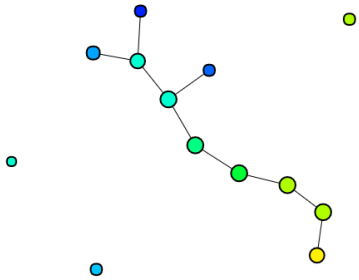
Outcome	t-SNE plot with perplexities from 30 to 50
MH	
TEE	
MO	

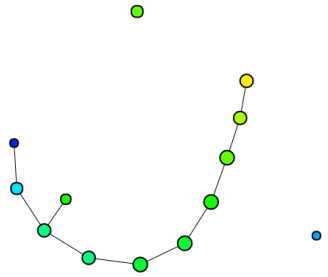
Legend: MH= major hemorrhage; TEE= thromboembolic events; MO= main outcome.

Mapper results

In order to mine other insights from the raw data before their manipulation we executed the Mapper algorithms. The inputs to the algorithm are the clinical variables plus their dummy representation, when necessary, plus the target variable under analysis. For Mapper algorithm, we used the Jaccard coefficient as metric and the DBscan as clustering algorithm, overlap percentage equal to 10%. The lens is the sum of the entries plus their MinMax scaling.

Table 13: Comparison of Mapper for different target variables

Outcome	**Topological Data Analysis (Mapper)**
MH	
TEE	

MO	

Legend: MH= major hemorrhage; TEE= thromboembolic events; MO= main outcome.

The Mapper analysis identified interesting features and relationships in the dataset confirming t-SNE results and extended previous results by highlighting interesting topological features.

We remark that, from a topological perspective, we focus on forks (flares) and big loops. In our case, each simplicial complex contained at least one fork. The main characteristics of the topological features are shown in Table 13:

- MH: The simplicial complex contains two small loop and a fork. The longest flare contains two nodes for a total of 80 patients, while the other node contains 9 samples.
- TEE: The simplicial complex contains 2 forks. The nodes with largest diameter in the bottom flare contain 27 samples, while the other two nodes contain 7 samples each.
- MO: The simplicial complex contains only one fork. The longest flare contains two nodes with totally 39 samples, while the other node contains 8 samples.

Mapper returns also the index of the patients belonging to each node and - driven by this information – we subsampled the dataset by selecting the subjects in the forks and compared them by using chi-squared tests. Results are synthesized in Table 14, Table 15 and Table 16. We report only the features with p-value < 0.05.

Table 14: statistically relevant features that differentiate the nodes in the fork of the MH simplicial complex

Features	χ^2	p-value
MH type	376.8212	0
Anticoagulant type at discharge	44.89189	0
MH	389	0
Acute Heart Failure	36.35367	0
Haemorragic Shock	92.71738	0
ASA or Clopidogrel use	20.71988	0.000005
Previous GI Bleeding	13.06815	0.0003
Age	10.86397	0.001071
Number of comorbidities	17.96861	0.021463
Diabetes Mellitus	4.287005	0.038405
CVE	4.265441	0.038895
Active Cancer	4.134792	0.04201

Legend: MH= major hemorrhage; ASA= acetylsalicylic acid; GI= gastrointestinal; CVE= electrical cardioversion.

Table 15: statistically relevant features that differentiate the nodes in the fork of the TEE simplicial complex

Features	χ^2	p-value
TEE type	42	0
TEE	42	0
AF Type	23.55294	0.000031
CVE	10.65613	0.001097

Anticoagulant type at admission	7.285045	0.006953
Sex	5.164035	0.023059
Previous stroke/TIA	5.164035	0.023059
Aortic valve disease	5.164035	0.023059

Legend: TEE= thromboembolic event; AF= atrial fibrillation; TIA= transient ischemic attack; CVE= electrical cardioversion.

Table 16: Statistically relevant features that differentiate the nodes in the fork of the main outcome simplicial complex

Features	χ^2	**p-value**
Septic Shock	11.84188	0.000579
CKD	5.9538	0.014686
Infection	6.880842	0.032051

Legend: CKD= chronic kidney disease.

The joined analysis of topological structures and statistical tests provided a first significant indication on which features should be used for training the ML classifiers: we observed that the features in the MH simplicial complex with highest χ^2 score reflected hemorrhage-related issues, such as bleeding predisposition, active cancer and drugs (Table 14). Features with the highest score in the TEE simplicial complex were related to cardioembolism and AF type (Table 15). Features with the highest score for MO in this analysis are synthesized in Table 16.

Feature selection

Mapper TDA was able to enlighten the importance of some variables. However, the evaluation of the simplicial forks provides only informations useful to understand some clusters and their statistical differences. In order to instruct automatic classifiers with ML we

extended the local results - provided by Mapper - to global analysis. In order to capture global statistical insights, we evaluated the dependency among the clinical variables and the target variable under modeling by performing both the χ^2 test with the Yates correction for continuity to evaluate the dependency among categorical features and target variable.

The F-value was used to study the dependencies of the discrete variables on the target variable. Results of this analysis are reported in Table 17, where we report only the features that received a p-value < 0.05. We obtained a reduced number of variables related with the target variables and this was useful on modeling the classifiers. The features identified by Mapper were partially included among the features indicated above in Table 17.

Table 17: Relevant features related to the target variables

Outcome	**Topological Data Analysis via Mapper**		
	Feature	**χ^2/F-value**	**p-value**
MH	MH type	1595.4561	0
	Anticoagulant therapy at admission	37.241885	0
	Anticoagulant therapy at discharge	79.376273	0
	MH	1655	0
	Haemorragic Shock	293.64996	0
	Chronic Anaemia	39.094657	0
	Low TTR	27.003168	0
	Previous GI Bleeding	18.03138	0.000022
	CVE	13.756998	0.000208
	ASA or Clopidogrel use	11.396544	0.000736
	Acute Heart Failure	11.093365	0.000866
	Age	10.006409	0.001588

	TEE Type	48.293552	0.007119
	Acute Respiratory Failure	5.248561	0.021965
	Active Cancer	4.629148	0.031433
	CKD	4.233212	0.03964
	Syncope	4.053498	0.04408
TEE	**Feature**	**χ^2/F-value**	**p-value**
	TEE Type	1613.2040	0
	CVE	28.056792	0
	Anticoagulant therapy at admission	65.840332	0
	Anticoagulant therapy at discharge	37.435286	0
	TEE	1646.1671	0
	Acute Heart Failure	37.235441	0
	Previous stroke/TIA	32.398337	0
	Low TTR	25.672567	0
	Age	16.45507	0.00005
	CHF	11.197515	0.00081
	ACS	9.988791	0.00157
	Mitral Valve Disease	8.180653	0.004234
	Sex	6.34187	0.01179
	AF Type	10.794743	0.01288
	Syncope	5.534671	0.01864
	CVF	5.237731	0.02210
	COPD	4.581552	0.03231
	Aortic Valve Disease	3.96779	0.04637

	Feature	χ^2/F-value	p-value
MO	Age	22.723161	0
	Anticoagulant therapy at discharge	71.900145	0
	ACS	31.465911	0
	Cardiogenic Shock	178.662335	0
	Septic Shock	161.661497	0
	Acute Respiratory Failure	39.144258	0
	Infection	101.341306	0
	Hypertension	22.521132	0.000013
	CVE	21.94512	0.000017
	Hemorrhagic Shock	21.932328	0.000017
	COPD	14.544708	0.000694
	TEE type	91.295819	0.00114
	Mitral Valve Disease	9.939222	0.006946
	ASA or Clopidogrel use	8.233689	0.016296
	Sex	7.700486	0.021275
	TEE	23.053895	0.027272
	MH	20.076694	0.028536
	Peripheral Artery Disease	6.524466	0.038303

Legend: MH= major haemorrhage; TEE= thromboembolic event; TTR= time in therapeutic range; GI= gastrointestinal; CVE= electric cardioversion; CVF= pharmacological cardioversion; ASA= acetyl salicylic acid; CKD= chronic kidney disease; TIA= transient ischemic attack; CHF= chronic heart failure; AF= atrial fibrillation; COPD= chronic obstructive pulmonary disease.

Training automatic classifiers with machine learning

The features identified with the statistical test were used as input for the TPOT framework. The modeling experiment was executed twice by changing the number of generations: respectively 10 and 20. The population size was fixed and equal to 20 while the number of k-folds for the cross validation on the training set was k=5. Models performance was evaluated using the classification error, defined as the percent of incorrect classifications, with a minimum possible score equal to 0.

The performances of the selected pipelines are reported in terms of average accuracy (1 – classification error), average classification error and 95% of confidence intervals (CI) on the training set and ROC and AUC for the test set (Figure 7). The results are reported in Table 18 and Table 19.

Table 18: Best Pipeline(s) fitted after 10 generations

Target	Accuracy %	Classification Error	95%CI	Best Pipeline
MH	100	0	0	LinearSVC (input_matrix, C=25.0, dual = Tur, loss = squared_hinge, penalty = 12, tol = 0.001)
TEE	100	0	0	GradientBoostingClassfier (input_matrix, learning_rate = 1.0, max_depth = 9, max_features = 0.8, min_samples_leaf = 1, min_samples_split = 3, n_estimators = 100, subsample = 0.65)
MO	87.68	0.123	0.09153-0.154847	RandomForestClassfier(OneHotEncoder(input_matrix, minimum_fraction = 0.15, sparse = False), bootstrap = False, criterion = entropy, max_features = 0.25, min_samples_leaf = 2, min_samples_split = 13, n_estimators = 100)

Legend: MH= major hemorrhage; TEE= thromboembolic events; MO= main outcome.

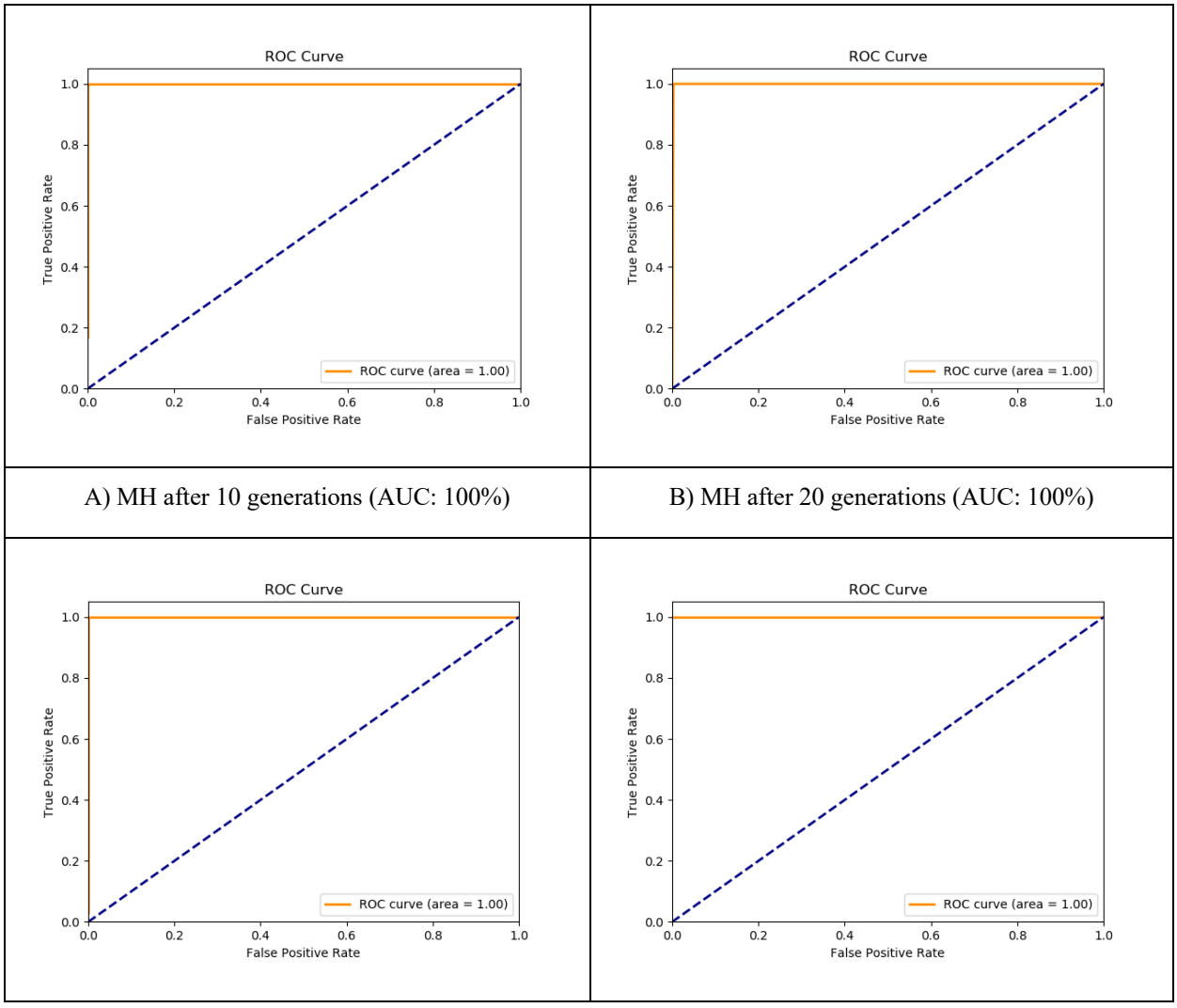
Table 19: Best Pipeline(s) fitted after 20 generations

Target	Accuracy %	Classification Error	95%CI	Best Pipeline

MH	100	0	0	RandomForestClassfier(input_matrix, bootstrap = True, criterion = entropy, max_features = 0.95, min_samples_leaf = 2, min_samples_split = 16, n_estimators = 100)
TEE	99.75	0. 2415	0.2313 – 0.3144	DecisionTreeClassifier(input_matrix, criterion = gini, max_depth = 9, min_samples_leaf = 1, min_samples_split = 11)
MO	90.33	0.09	0.068159 – 0.125077	KNeighborsClassifier(MaxAbsScaler(RFE(Normalizer(input_matrix), norm = max), criterion = gini, max_features = 0.95, n_estimators = 100, step = 0.95)), n_neighbors = 34, p =1, weights = distance)

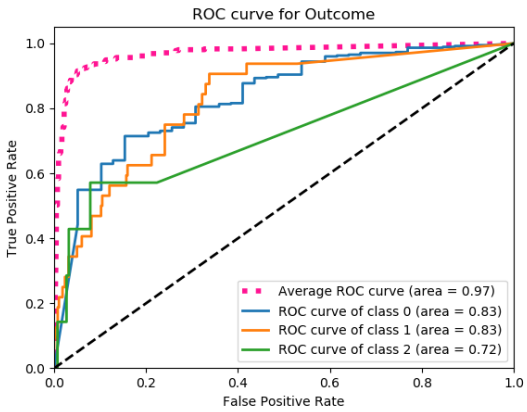
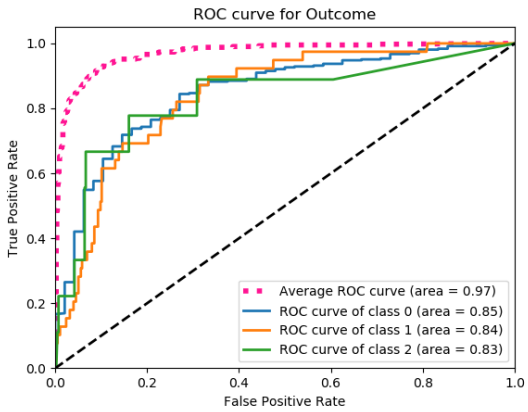
Legend: MH= major hemorrhage; TEE= thromboembolic events; MO= main outcome.

Figure 7: Results of classifiers



A) MH after 10 generations (AUC: 100%)

B) MH after 20 generations (AUC: 100%)

C) TEE after 10 generations (AUC: 100%)	D) TEE after 20 generations (AUC: 99.75%)
	
E) MO after 10 generations (AUC: 87,68%)	F) MO after 20 generations (AUC: 90,33%)

Legend: MH= major hemorrhage; TEE= thromboembolic events; MO= main outcome; AUC= area under the curve.

The obtained classifiers outperformed the CHA_2DS_2-VASC and HAS-BLED scoring systems in the prediction of TEE and MH, respectively. The small fluctuation from accuracy = 100% to accuracy = 99.75% in the TEE classifier was due to random initialization during the splitting of the training set but is not relevant since the classification error remains extremely low. The pipelines for predicting both MH as well as the TEE were relatively simple: they train decision-tree based algorithms for modelling the scores and this means the populations forming healthy and unhealthy patients are approximately linearly or polynomial separable.

The MO classifier showed good performances, and its accuracy raised from 87.68% to 90.33% by increasing the number of generations. These performances were not as excellent as the other two outcomes, since the numbers of patients belonging to class 2 is relatively low (few tens) and there are several overlaps among patients as pinpointed out by the t-SNE plot. This indicated that other features should be added to the dataset.

Mapper analysis suggested looking for features regarding septic shock, infections and kidney disease. Conversely, the pipelines fitted to predict the MO counted a number of intermediate steps to improve the separation among populations belonging to different classes.

In details, the best-fitted pipelines perform a normalization of the data by setting mean to 0 and standard deviation to 1 (Normalize). Then, it performs a recursive features elimination for reducing the number of features (RFE), the it executes a scaling of each feature such that the maximal absolute value of each feature in the training set will be 1.0. Eventually, the pipeline trains a proximity-based classifier (KNeighborsClassifier).

The first results were critically reviewed: since variable selection and model generation was machine-driven, the clinical role of each variable was discussed. The extremely impressive high accuracy of the TEE and MH classifier was – at least in part – motivated by the presence of some features that are synonyms of the outcomes we aimed to predict: the training set used for instructing the classifier for the MH contained among the other the features MH type, MH, HS and TEE type with highest χ^2: these features map exactly the outcome itself. Similarly, the training set for TEE contained two features (TEE type and TEE), that represented the outcome itself.

Thus, we removed those features from the list and retrained the classifiers. Moreover, in order to estimate how features impact the overall quality, we adopted the following procedure:

1) Sort in ascending order the features according their p-value (from 0 to n).
2) Span on the feature set and pick up one feature at time, accordingly to its p-value, and re-train the classifier.

The output of this algorithm is a plot, for each target variable, where horizontal axis is the number of features and the vertical axis is the corresponding AUC. The features used in this evaluation are reported in Table 20. The corresponding plots are reported in Table 21.

Table 20: reduced feature set by removing the features that are the images of the target variables

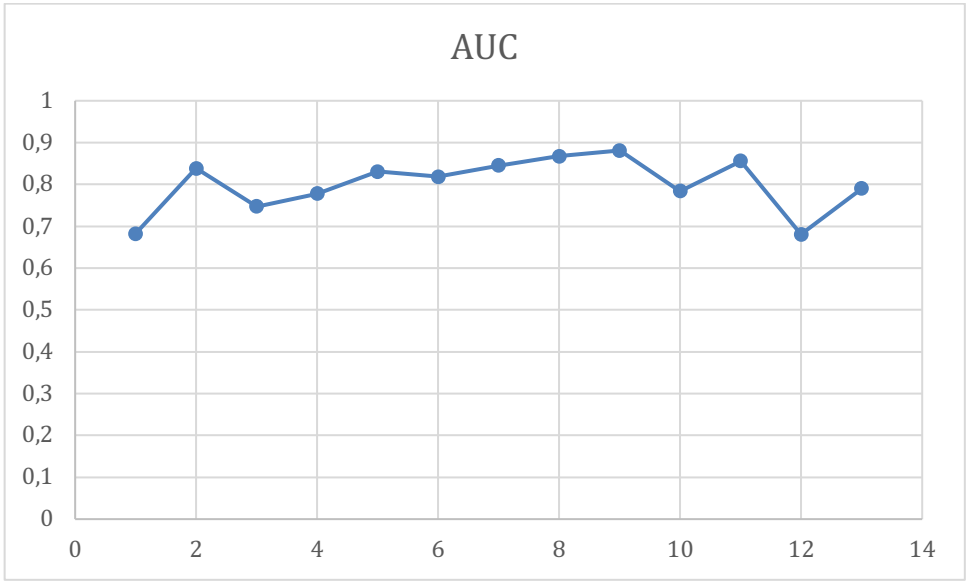
Outcome	Topological Data Analysis via Mapper		
MH	**Feature**	**χ^2/F-value**	**p-value**
	Anticoagulant therapy at admission	37.241885	0
	Anticoagulant therapy at discharge	79.376273	0
	Chronic Anaemia	39.094657	0
	Low TTR	27.003168	0
	Previous GI Bleeding	18.03138	0.000022
	CVE	13.756998	0.000208
	ASA or Clopidogrel use	11.396544	0.000736
	Acute Heart Failure	11.093365	0.000866
	Age	10.006409	0.001588
	Acute Respiratory Failure	5.248561	0.021965
	Active Cancer	4.629148	0.031433
	CKD	4.233212	0.03964
	Syncope	4.053498	0.04408
TEE	**Feature**	**χ^2/F-value**	**p-value**
	CVE	28.056792	0
	Anticoagulant therapy at admission	65.840332	0
	Anticoagulant therapy at discharge	37.435286	0
	Acute Heart Failure	37.235441	0
	Previous stroke/TIA	32.398337	0
	Low TTR	25.672567	0

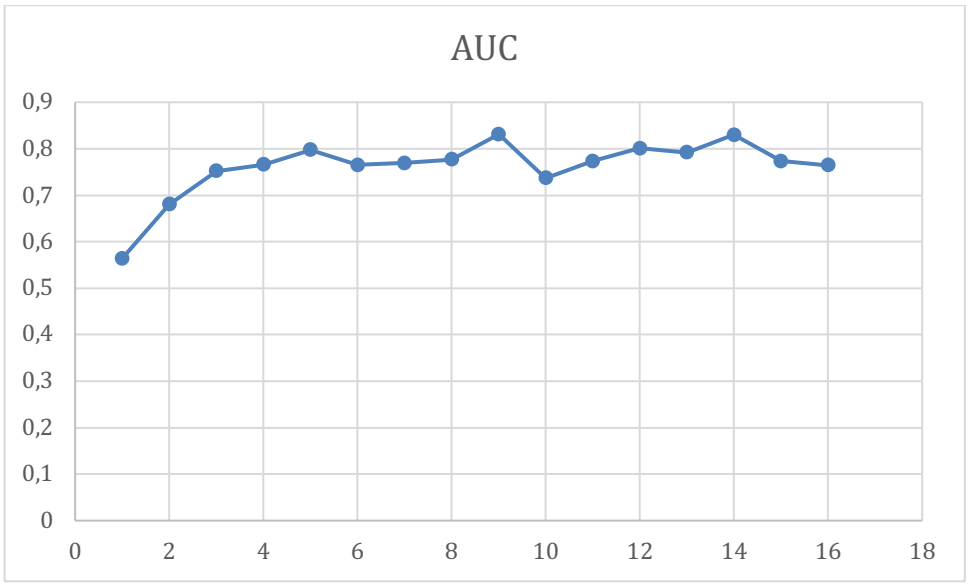
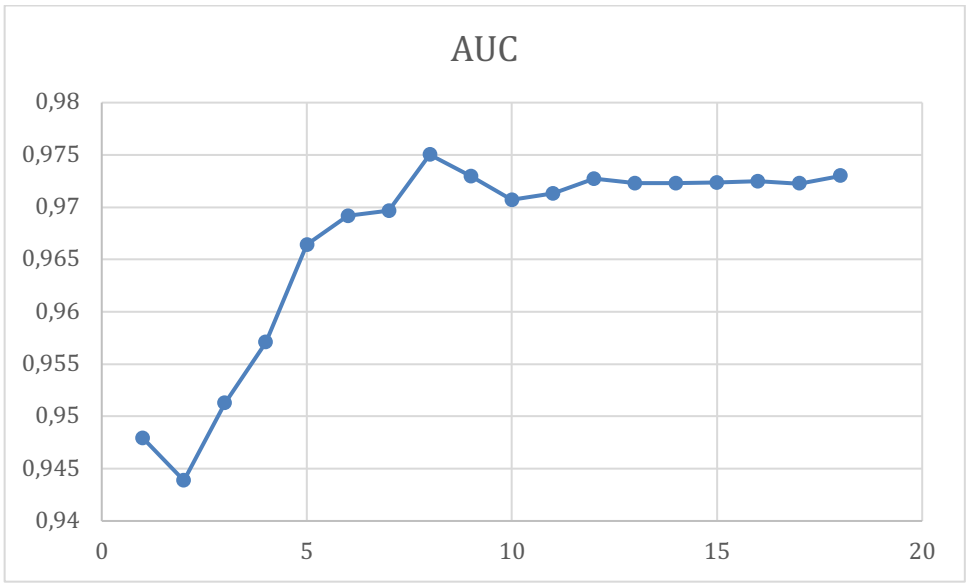
	Age	16.455079	0.00005
	CHF	11.197515	0.00081
	ACS	9.988791	0.00157
	Mitral Valve Disease	8.180653	0.004234
	Sex	6.34187	0.01179
	AF Type	10.794743	0.01288
	Syncope	5.534671	0.01864
	CVF	5.237731	0.02210
	COPD	4.581552	0.03231
	Aortic Valve Disease	3.96779	0.04637
MO	**Feature**	**χ^2/F-value**	**p-value**
	Age	22.723161	0
	Anticoagulant therapy at discharge	71.900145	0
	ACS	31.465911	0
	Cardiogenic Shock	178.662335	0
	Septic Shock	161.661497	0
	Acute Respiratory Failure	39.144258	0
	Infection	101.341306	0
	Hypertension	22.521132	0.000013
	CVE	21.94512	0.000017
	Hemorrhagic Shock	21.932328	0.000017
	COPD	14.544708	0.000694
	TEE type	91.295819	0.00114
	Mitral Valve Disease	9.939222	0.006946
	ASA or Clopidogrel use	8.233689	0.016296
	Sex	7.700486	0.021275

	TEE	23.053895	0.027272
	MH	20.076694	0.028536
	Peripheral Artery Disease	6.524466	0.038303

Legend: MH= major haemorrhage; TEE= thromboembolic event; TTR= time in therapeutic range; GI= gastrointestinal; CVE= electric cardioversion; CVF= pharmacological cardioversion; ASA= acetyl salicylic acid; CKD= chronic kidney disease; TIA= transient ischemic attack; CHF= chronic heart failure; AF= atrial fibrillation; COPD= chronic obstructive pulmonary disease.

Table 21: Evaluation of classifiers' performances trained with a different number of features. The x-axis is the feature number, while the y-axis is the AUC

MH	

TEE	
MO	

Legend: MH= major haemorrhage; TEE= thromboembolic events; MO= main outcome.

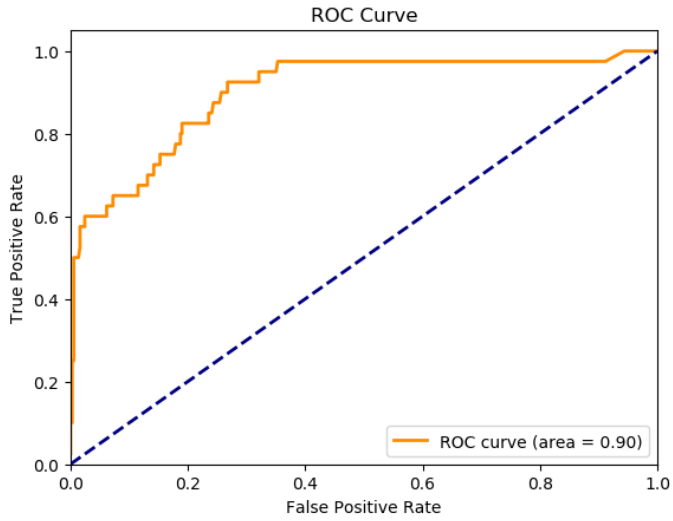
For MH, the classifier with highest accuracy (AUC: 90%) was obtained by using 9 features: anticoagulant therapy at admission, anticoagulant therapy at discharge, CA, low TTR, previous GI bleeding, CVE, ASA or Clopidogrel use, AHF and age, as shown in Table 21 and Table 22. At the best cutoff, selected with the Youden's index (optimal cut-off: 0.092056), the new classifier had a Sensitivity = 0.80, a 1-Specificity = 0.18984, a positive likelihood ratio of 4.21 and a negative likelihood ratio of 0.25.

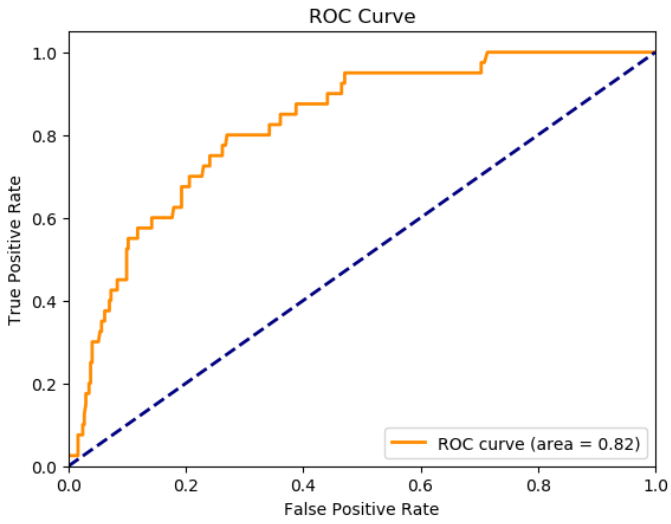
The best classifier for TEE (AUC: 82%) was obtained by using the following 9 features: CVE, anticoagulant therapy at admission, anticoagulant therapy at discharge, AHF, previous stroke/TIA, low TTR, age, CHF and ACS, as shown in Table 21 and Table 22. At the best cutoff, selected with the Youden's index (optimal cut-off: 0.137663), the new classifier had a Sensitivity = 0.75, a 1-Specificity = 0.245899, a positive likelihood ratio of 3.05 and a negative likelihood ratio of 0.33.

For MO, the highest accuracy (AUC: 97.5%) was obtained with the following variables: age, anticoagulant therapy at discharge, ACS, CS, SS, ARF, infection and HYP, as shown in Table 21.

This analysis highlighted that the 100% accuracy for MH and TEE classifiers decreased by reducing the number of features. However, the new scores outperformed the accuracy of HAS-BLED and CHA_2DS_2-VASC. We also observed that the average accuracy for the main outcome classifier slightly increased, from 97.2% to 97.5% by reducing the number of variables from 18 to 8.

Table 22: Evaluation of trained classifiers performances with ROC curve analysis.

MH	

TEE	

Legend: MH= major hemorrhage; TEE= thromboembolic events.

4. Discussion

Pre-existing or new-onset AF are common in an ICU population. In this cohort, critically-ill patients coaffected by AF represented 19% of the sample in the same time-period. This prevalence is slightly lower than the one reported in literature[5], however this observation could be due to the common effect of under-reporting of this type of arrhythmia in ICU[79], which is described especially when adopting administrative databases and retrospective models.

MO was observed in 194 (13.6%) subjects of the sample: this group was significantly older and more often affected by ACS, CS, SS, ARF or infectious diseases, as shown in Table 5. TEE were diagnosed in 212 patients (14.8% of the sample). In this subpopulation, CHA_2DS_2-VASc score had a median of 4 [IQR:3]. According to the original validation cohort of CHA_2DS_2-VASc, a similar embolic risk could be observed only in patients with a CHA_2DS_2-VASc ≥8[48]. The older age of the patients, the concomitant critical illnesses and comorbidities could at least partially explain the excess of TEE observed in this population.

MH was present in 133 subjects (9.30% of the sample). In this group, HAS-BLED score had a median of 2 [IQR:1]. Data from the overall SPORTIF cohorts underlined that a similar risk of bleeding was present only in subjects HAS-BLED scores ≥4[49]. In our cohort, 90.2% of MH were observed in patients with HAS-BLED <4. The excess in the haemorragic risk could be explained by several aspects: first, some critical illness, as severe sepsis or septic shock, especially when complicated by diffuse intravascular coagulation or atypical uremic-hemolytic syndromes, are often associated with platelet or coagulation abnormalities; second, patients with ACS are often treated with antiplatelet agents which could at least facilitate bleeding; third, the high prevalence of chronic or acute kidney dysfunction are recognized risk factors for MH, particularly among subjects undergoing parenteral anticoagulation.

Of note, both CHA_2DS_2-VASc and HAS-BLED showed a good negative predictive value, being able to exclude TEE and MH. CHA_2DS_2-VASc values <2 were associated to a NPV of 95.0% (87.6-98.1%) for TEE, while HAS-BLED scores <3 had a NPV of 91.3% (89.8-92.5%) for MH, as shown in Table 10. Thus, a possible role for the "classical" score systems in critically ill subjects could be of identifying clusters at "very low risk" of TEE and MH. However, this sample should be only used to suggest this hypothesis, since the number of "low-risk" patients is very scarce, due to advanced age, number of comorbidities and higher prevalence of cardiovascular diseases. This should not be interpreted as a selection bias, but as a clinical difference of the sICU population from the commonly studied cohorts. In fact, even adopting a less age- and vascular-dependent score, as $CHADS_2$, we obtained results similar to the ones observed with CHA_2DS_2-VASc.

These observations synthesize the difficulty of the emergency physician in managing anticoagulation in the elderly, critically-ill patients affected by AF: independently from the cause of admission, this group of subjects shows a very high risk of both cardioembolic and haemorragic events. Moreover, we observed in CHA_2DS_2-VASc and HAS-BLED a non-significant difference in the distribution of the subjects according different results, which translated into a non-significant predictive capacity of each score, as shown in Table 10 and Figure 5.

In this sample, the predictive performance of CHA_2DS_2-VASc score was non-significant and showed a very low accuracy. The role of CHA_2DS_2-VASc score in predicting TEE in patients admitted to an ICU/sICU is still object of debate: only a single perspective cohort[38] study of patients found a moderately predictive value of this score, but suggested a different cut-off value in order to improve its performance. Other studies observed a very poor or non-significant prediction of the events[19,37] in this setting. The ability of CHA_2DS_2-VASc in stratifying patient's thromboembolic risk during AF has been shown to be very low also in specific clinical settings, as AHF[80], ACS[36,39], sepsis[36,37] and

ARF[36]. This suggests that a correct risk stratification using this score could be performed in populations similar to the original validation cohort. Moreover, it is necessary to underline that, while pre-existing AF shares at least the same risk factors with primary AF observed in outpatients, being the critical illness a modifier of events, new-onset AF should be considered as a completely different entity, with different risk factors and a different cardioembolic and haemorragic risk profile[14].

Similarly, HAS-BLED score was not able to predict accurately MH in this cohort, despite a significantly increased prevalence of serious bleeding and haemorragic shock. Literature regarding the evaluation of bleeding risk in critical illness is poor, and previously published studies underlined that this risk is not accurately evaluated by HAS-BLED score[36]. Moreover, HAS-BLED has not been validated in such populations.

In the context of critically-ill patients with AF, we noted that, while both CHA_2DS_2-VASc and HAS-BLED scores were not associated with the occurrence of TEE and MH and did not show any predictive ability for these events, these scores demonstrated the ability of identifying the patients with a very low risk, who less likely would experience the outcome.

Moreover, in absence of more accurate and validated stratification tools, the latest European Heart Rhythm Association (EHRA) consensus document on the management of AF in critically-ill patients still suggests stratifying both thromboembolic and haemorragic risk with CHA_2DS_2-VASc and HAS-BLED[81] and to treat AF accordingly to the more recent ESC guidelines[1].

However, a large study performed in a retrospective cohort of septic patients underlined that the high risk of bleeding in critically-ill patients receiving parenteral anticoagulation was not counterbalanced by a significant reduction of ischemic stroke rate[37]. A $CHADS_2$-based anticoagulation strategy was also associated to an increased risk of bleeding in absence of a statistically significant increase in survival rate during the hospitalization in ICU[40]. Moreover, anticoagulation was not associated to a reduced in-

hospital and long-term stroke risk in ACS, ARF and sepsis[36], but it was clearly associated to a significantly increased risk of bleeding.

In a recently published work, we have already underlined the discrepancies between different guidelines in the topic of AF[82], which are mainly due to the absence of studies or to the poor quality of the studies adopted to elaborate specific indications.

In this study, when stratifying our sample according to current guidelines adherence, we observed that a GL-adherent approach was significantly associated to an overall reduction of TEE (4,8%), but also correlated to a significantly increased risk of adverse events. Particularly, GL-adherent patients had the highest probability of both MH (13,2%) and MO (16,4%), as shown in Figure 6. When compared to GL-adherent patients, undertreated patients had significantly increased risk of TEE, a reduced risk of MH and a reduced risk of MO, as shown in Table 11. Overtreatment did not confer significantly to a further increased risk of TEE, MH or MO when compared to a GL-adherent approach.

Moreover, we observed that warfarin use at the admission was associated to a two-fold increased risk of death or ICU transfer when compared to no anticoagulant use, as shown in Table 8. LMWH use, however, did not confer an increased risk of MO, and antiplatelet drugs use was associated to a reduction of the risk of MO.

In this sample, nor CHA_2DS_2-VASc nor HAS-BLED were able to predict events in critically-ill subjects. Moreover, the application of current ESC guidelines was effective in reducing TEE but was associated to an increased risk of both MO and MH. To date, according to our data and current literature, both risk stratification of ischemic and haemorragic risk and medical management of AF in critically-ill patients still represent a "gray zone" of evidence.

As in this case, ICU represents an especially compelling case for clinical data analysis: the value of many treatments and interventions is still unproven, and high-quality data supporting or discouraging specific practices are embarrassingly sparse[83,84].

Guidelines developed to standardize practice are dependent on an evidence base that is surprisingly thin considering the copious data generated in ICU. Big data analysis and ML represent a new and promising technologies which are deemed to be highly efficient in exploring patterns and hidden relationships between clinical variables[41], allowing to develop more specific clinical prediction tools and open the road to personalized medicine, especially in ICU/sICU where large volumes of data are readily available.

Predictive models adopting these new technologies have already outperformed their gold standard in several acute pathologies, as AKI[85] or PE[44]. With this study, we applied these new techniques in the clinical setting critically-ill patients affected by AF, aiming to highlight the most relevant features associated to the most important adverse events, that are death or ICU transfer, major bleeding and cardioembolism. We then engineered new predictive scores to improve the prediction of AF-related events.

The strongest association with major haemorragic events was observed with age, specific critical illnesses (AHF, ARF, SYN), specific chronic conditions (CKD, AC and CA), anticoagulant and antiplatelet therapy, procedures performed in sICU (CVE), labile TTR and a history of previous gastrointestinal bleeding. Some of these factors (age, antiplatelet drugs use, CKD, labile TTR and previous gastrointestinal bleeding) have already been identified and included in the HAS-BLED score[86]. Other factors (AC and CA) are recognized by literature as risk factors for MH, and have already been considered in other risk scores, as HEMORR$_2$HAGES[87]. Some items, however (AHF, ARF, SYN), are specific for the critically-ill patient. Adopting this 9-item score with a TDA-based computation, we obtained a significant increase in the accuracy of prediction of MH [AUC from 0.52 to 0.90, $p<0.001$], as shown in Table 20, Table 21 and Table 22.

The association with TEE was robust for age, sex, specific critical illnesses (AHF, ACS, SYN), procedures performed in sICU (CVE, CVF), AF type, specific chronic conditions (CHF, COPD, mitral and aortic valve pathologies), the anticoagulant strategy, the

time in therapeutic range and a previous stroke/TIA. Some of these factors are already considered in CHA_2DS_2-VASc score (CHF, age, sex and previous stroke/TIA). Other features, as anticoagulant therapy, time in therapeutic range, CVE and CVF are directly related to the acute management of the arrhythmia. Last, some items (AHF, ACS, SYN) refer directly to the critical pathology which caused the sICU admission. Thus, we engineered a TDA-based model which allowed a significant increase in the prediction of TEE [AUC from 0.493 to 0.82, $p<0.001$], as shown in Table 20, Table 21 and Table 22.

While HAS-BLED and CHA_2DS_2-VASc are still not validated for ICU, intensive-care physicians already have several validated tools to predict in-hospital mortality among critically-ill patients, as SAPS-II[88] and APACHE-II[89] scores which are able to predict the outcome with an area under the ROC curve greater than 80%[90]. Some predictors of MO extracted from our population were similar to the items considered in both scores, as age, COPD, ARF and shock. Other items were already associated in previous studies to a worse outcome in this clinical setting[91–93]. However, it is interesting to underline that, in a selected population of critically-ill patients affected by AF, we were able to identify AF-specific items, as anticoagulant strategy, antiplatelet drugs use, CVE, TEE, MH and HS, as strong determinants of in-hospital death or ICU transfer.

This last observation underlines also the importance of AF-related events in the determination of major outcomes of the critically-ill patient. The absence of robust evidence both in risk stratification and in the consequent anticoagulant strategy poses the critically-ill patient affected by AF at risk of adverse events even if treated according the current guidelines. This preliminary work underlines the urgent need of specific trials for the management of new-onset or pre-existing AF in critically-illness.

Study limitations

This is a single-center, retrospective study, thus generalizability of results is limited and require further confirmations with larger, multicentric and prospective studies. The

peculiar TDA-based approach adopted for feature selection and event prediction of events is robust but experimental and, despite the encouraging results, it will require further validations: in particular, a larger prospective validation cohort is necessary to confirm our preliminary results. The validation cohort should enroll particularly patients treated accordingly to ESC 2016 guidelines. Last, this retrospective cohort considered subjects treated with warfarin or LMWH but did not involve patients treated with direct oral anticoagulants: these drugs have been introduced later and seem to have a safer bleeding profile: future studies will have to consider also this new class of medications in this specific setting of patients.

Acknowledgments: we thank Dott. Matteo Rucco, PhD, for the topological data analysis and machine learning experiment; Dott. Marco Proietti, MD, FESC, for the clinical analysis; Dott. Aldo Salvi, for paper review, patients' data and review of single patient record and diagnosis; Dott. Giovanna Viticchi for clinical analysis and paper review; Prof. Mauro Silvestrini, for paper review; Dott. Francesco Grigioni, for the review and supervision of this study; Prof. Alessandro Capucci, for the review and supervision of this study.

References

[1] Kirchhof P, Benussi S, Kotecha D, Ahlsson A, Atar D, Casadei B, Castella M, Diener H-C, Heidbuchel H, Hendriks J, Hindricks G, Manolis AS, Oldgren J, Popescu BA, Schotten U, Van Putte B, Vardas P, Agewall S, Camm J, Baron Esquivias G, Budts W, Carerj S, Casselman F, Coca A, De Caterina R, Deftereos S, Dobrev D, Ferro JM, Filippatos G, Fitzsimons D, Gorenek B, Guenoun M, Hohnloser SH, Kolh P, Lip GYH, Manolis A, McMurray J, Ponikowski P, Rosenhek R, Ruschitzka F, Savelieva I, Sharma S, Suwalski P, Tamargo JL, Taylor CJ, Van Gelder IC, Voors AA, Windecker S, Zamorano JL, Zeppenfeld K (2016) 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *Eur. Heart J.* **37**, 2893–2962.

[2] Reinelt P, Karth GD, Geppert A, Heinz G (2001) Incidence and type of cardiac arrhythmias in critically ill patients: a single center experience in a medical-cardiological ICU. *Intensive Care Med.* **27**, 1466–73.

[3] Bosch NA, Cimini J, Walkey AJ (2018) Contemporary Reviews in Critical Care Medicine: Atrial Fibrillation in the Intensive Care Unit. *Chest* **0**,.

[4] Sibley S, Muscedere J (2015) New-onset atrial fibrillation in critically ill patients. *Can. Respir. J.* **22**, 179–82.

[5] Artucio H, Pereira M (1990) Cardiac arrhythmias in critically ill patients: epidemiologic study. *Crit. Care Med.* **18**, 1383–8.

[6] Knotzer H, Mayr A, Ulmer H, Lederer W, Schobersberger W, Mutz N, Hasibeder W (2000) Tachyarrhythmias in a surgical intensive care unit: a case-controlled epidemiologic study. *Intensive Care Med.* **26**, 908–14.

[7] Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ, Gillum RF, Kim Y-H, McAnulty JH, Zheng Z-J, Forouzanfar MH, Naghavi M, Mensah GA, Ezzati M, Murray CJL (2014) Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. *Circulation* **129**, 837–47.

[8] Gamst J, Christiansen CF, Rasmussen BS, Rasmussen LH, Thomsen RW (2015) Pre-existing atrial fibrillation and risk of arterial thromboembolism and death in intensive care

unit patients: a population-based cohort study. *Crit. Care* **19**, 299.

[9] Kannel WB, Abbott RD, Savage DD, McNamara PM (1982) Epidemiologic Features of Chronic Atrial Fibrillation. *N. Engl. J. Med.* **306**, 1018–1022.

[10] Lloyd-Jones DM, Wang TJ, Leip EP, Larson MG, Levy D, Vasan RS, D'Agostino RB, Massaro JM, Beiser A, Wolf PA, Benjamin EJ (2004) Lifetime Risk for Development of Atrial Fibrillation: The Framingham Heart Study. *Circulation* **110**, 1042–1046.

[11] Benjamin EJ, Levy D, Vaziri SM, D'Agostino RB, Belanger AJ, Wolf PA (1994) Independent risk factors for atrial fibrillation in a population-based cohort. The Framingham Heart Study. *JAMA* **271**, 840–4.

[12] Nattel S, Harada M (2014) Atrial remodeling and atrial fibrillation: recent advances and translational perspectives. *J. Am. Coll. Cardiol.* **63**, 2335–45.

[13] Nattel S, Burstein B, Dobrev D (2008) Atrial remodeling and atrial fibrillation: mechanisms and implications. *Circ. Arrhythm. Electrophysiol.* **1**, 62–73.

[14] Kuipers S, Klouwenberg PMK, Cremer OL (2014) Incidence, risk factors and outcomes of new-onset atrial fibrillation in patients with sepsis: a systematic review. *Crit. Care* **18**, 688.

[15] Heinz G (2006) Atrial fibrillation in the intensive care unit. *Intensive Care Med.* **32**, 345–348.

[16] Tselentakis EV, Woodford E, Chandy J, Gaudette GR, Saltman AE (2006) Inflammation Effects on the Electrical Properties of Atrial Tissue and Inducibility of Postoperative Atrial Fibrillation. *J. Surg. Res.* **135**, 68–75.

[17] Clark DM, Plumb VJ, Epstein AE, Kay GN (1997) Hemodynamic effects of an irregular sequence of ventricular cycle lengths during atrial fibrillation. *J. Am. Coll. Cardiol.* **30**, 1039–45.

[18] Walkey AJ, Wiener RS, Ghobrial JM, Curtis LH, Benjamin EJ (2011) Incident Stroke and Mortality Associated With New-Onset Atrial Fibrillation in Patients Hospitalized With Severe Sepsis. *JAMA* **306**, 2248–2254.

[19] Kanji S, Williamson DR, Yaghchi BM, Albert M, McIntyre L, Canadian Critical Care

Trials Group (2012) Epidemiology and management of atrial fibrillation in medical and noncardiac surgical adult intensive care unit patients. *J. Crit. Care* **27**, 326.e1-326.e8.

[20] Meierhenrich R, Steinhilber E, Eggermann C, Weiss M, Voglic S, Bögelein D, Gauss A, Georgieff M, Stahl W (2010) Incidence and prognostic impact of new-onset atrial fibrillation in patients with septic shock: a prospective observational study. *Crit. Care* **14**, R108.

[21] Seguin P, Laviolle B, Maurice A, Leclercq C, Mallédant Y (2006) Atrial fibrillation in trauma patients requiring intensive care. *Intensive Care Med.* **32**, 398–404.

[22] Biasco L, Radovanovic D, Moccetti M, Rickli H, Roffi M, Eberli F, Jeger R, Moccetti T, Erne P, Pedrazzini G (2018) New-onset or Pre-existing Atrial Fibrillation in Acute Coronary Syndromes: Two Distinct Phenomena With a Similar Prognosis. *Rev. Española Cardiol. (English Ed.*

[23] Lau DH, Huynh LT, Chew DP, Astley CM, Soman A, Sanders P (2009) Prognostic Impact of Types of Atrial Fibrillation in Acute Coronary Syndromes. *Am. J. Cardiol.* **104**, 1317–1323.

[24] Wang TJ, Larson MG, Levy D, Vasan RS, Leip EP, Wolf PA, D'Agostino RB, Murabito JM, Kannel WB, Benjamin EJ (2003) Temporal relations of atrial fibrillation and congestive heart failure and their joint influence on mortality: the Framingham Heart Study. *Circulation* **107**, 2920–5.

[25] Shaver CM, Chen W, Janz DR, May AK, Darbar D, Bernard GR, Bastarache JA, Ware LB (2015) Atrial Fibrillation Is an Independent Predictor of Mortality in Critically Ill Patients*. *Crit. Care Med.* **43**, 2104–2111.

[26] Chen AY, Sokol SS, Kress JP, Lat I (2015) New-Onset Atrial Fibrillation Is an Independent Predictor of Mortality in Medical Intensive Care Unit Patients. *Ann. Pharmacother.* **49**, 523–527.

[27] Walkey AJ, Hammill BG, Curtis LH, Benjamin EJ (2014) Long-term Outcomes Following Development of New-Onset Atrial Fibrillation During Sepsis. *Chest* **146**, 1187–1195.

[28] Makrygiannis SS, Rizikou D, Patsourakos NG, Lampakis M, Margariti A, Ampartzidou OS, Sakellaridis K, Tselioti P, Pipilis A, Prekates AA (2017) New-onset atrial fibrillation and clinical outcome in non-cardiac intensive care unit patients. *Aust. Crit. Care*.

[29] Carrera P, Thongprayoon C, Cheungpasitporn W, Iyer VN, Moua T (2016) Epidemiology and outcome of new-onset atrial fibrillation in the medical intensive care unit. *J. Crit. Care* **36**, 102–106.

[30] DiMarco JP (2009) Atrial fibrillation and acute decompensated heart failure. *Circ. Heart Fail.* **2**, 72–3.

[31] Kotecha D, Piccini JP (2015) Atrial fibrillation in heart failure: what should we do? *Eur. Heart J.* **36**, ehv513.

[32] Kanji S, Stewart R, Fergusson DA, McIntyre L, Turgeon AF, Hébert PC (2008) Treatment of new-onset atrial fibrillation in noncardiac intensive care unit patients: A systematic review of randomized controlled trials*. *Crit. Care Med.* **36**, 1620–1624.

[33] Watson T, Shantsila E, Lip GYH (2009) Mechanisms of thrombogenesis in atrial fibrillation: Virchow's triad revisited. *Lancet (London, England)* **373**, 155–66.

[34] Luo J, Li H, Qin X, Liu B, Zhao J, Maihe G, Li Z, Wei Y (2018) Increased risk of ischemic stroke associated with new-onset atrial fibrillation complicating acute coronary syndrome: A systematic review and meta-analysis. *Int. J. Cardiol.* **265**, 125–131.

[35] Massera D, Wang D, Vorchheimer DA, Negassa A, Garcia MJ (2016) Increased risk of stroke and mortality following new-onset atrial fibrillation during hospitalization. *Europace* **19**, euw110.

[36] Quon, MJ; MD, Hassan, B; Pilote L (2017) Anticoagulant Use and Risk of Ischemic Stroke and Bleeding in Patients With Secondary Atrial Fibrillation Associated With Acute Coronary Syndromes, Acute Pulmonary Disease, or Sepsis. *JACC Clin. Electrophysiol.*

[37] Walkey AJ, Quinn EK, Winter MR, McManus DD, Benjamin EJ (2016) Practice Patterns and Outcomes Associated With Use of Anticoagulation Among Patients With Atrial Fibrillation During Sepsis. *JAMA Cardiol.* **1**, 682–90.

[38] Champion S, Lefort Y, Gaüzère B-A, Drouet D, Bouchet BJ, Bossard G, Djouhri S,

Vandroux D, Mayaram K, Mégarbane B (2014) CHADS2 and CHA2DS2-VASc scores can predict thromboembolic events after supraventricular arrhythmia in the critically ill patients. *J. Crit. Care* **29**, 854–8.

[39] Guerra F, Scappini L, Maolo A, Campo G, Pavasini R, Shkoza M, Capucci A (2016) CHA2DS2-VASc risk factors as predictors of stroke after acute coronary syndrome: A systematic review and meta-analysis. *Eur. Hear. J. Acute Cardiovasc. Care*.

[40] Darwish OS, Strube S, Nguyen HM, Tanios MA (2013) Challenges of Anticoagulation for Atrial Fibrillation in Patients With Severe Sepsis. *Ann. Pharmacother.* **47**, 1266–1271.

[41] Bailly S, Meyfroidt G, Timsit J-F (2018) What's new in ICU in 2050: big data and machine learning. *Intensive Care Med.* **44**, 1524–1527.

[42] Cámara PG (2017) Topological methods for genomics: Present and future directions. *Curr. Opin. Syst. Biol.* **1**, 95–101.

[43] Merelli E, Rucco M, Sloot P, Tesei L, Merelli E, Rucco M, Sloot P, Tesei L (2015) Topological Characterization of Complex Systems: Using Persistent Entropy. *Entropy* **17**, 6872–6892.

[44] Rucco M, Sousa-Rodrigues D, Merelli E, Johnson JH, Falsetti L, Nitti C, Salvi A (2015) Neural hypernetwork approach for pulmonary embolism diagnosis. *BMC Res. Notes* **8**, 617.

[45] Falsetti LG, Merelli E, Rucco M, Nitti C, Gentili T, Pennacchioni M, Salvi A (2013) A data-driven clinical prediction rule for pulmonary embolism. *Eur. Heart J.* **34**, P243–P243.

[46] Rucco M, Merelli E, Herman D, Ramanan D, Petrossian T, Falsetti L, Nitti C, Salvi A (2015) Using topological data analysis for diagnosis pulmonary embolism. *J. Theor. Appl. Comput. Sci.* **9**, 41–55.

[47] Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ (2001) Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. *JAMA* **285**, 2864–70.

[48] Lip GYH, Nieuwlaat R, Pisters R, Lane DA, Crijns HJGM (2010) Refining Clinical Risk

Stratification for Predicting Stroke and Thromboembolism in Atrial Fibrillation Using a Novel Risk Factor-Based Approach. *Chest* **137**, 263–272.

[49] Lip GYH, Frison L, Halperin JL, Lane DA (2011) Comparative Validation of a Novel Risk Score for Predicting Bleeding Risk in Anticoagulated Patients With Atrial Fibrillation. *JAC* **57**, 173–180.

[50] Ibanez B, James S, Agewall S, Antunes MJ, Bucciarelli-Ducci C, Bueno H, Caforio ALP, Crea F, Goudevenos JA, Halvorsen S, Hindricks G, Kastrati A, Lenzen MJ, Prescott E, Roffi M, Valgimigli M, Varenhorst C, Vranckx P, Widimský P, Collet J-P, Kristensen SD, Aboyans V, Baumbach A, Bugiardini R, Coman IM, Delgado V, Fitzsimons D, Gaemperli O, Gershlick AH, Gielen S, Harjola V-P, Katus HA, Knuuti J, Kolh P, Leclercq C, Lip GYH, Morais J, Neskovic AN, Neumann F-J, Niessner A, Piepoli MF, Richter DJ, Shlyakhto E, Simpson IA, Steg PG, Terkelsen CJ, Thygesen K, Windecker S, Zamorano JL, Zeymer U, Windecker S, Aboyans V, Agewall S, Barbato E, Bueno H, Coca A, Collet J-P, Coman IM, Dean V, Delgado V, Fitzsimons D, Gaemperli O, Hindricks G, Iung B, Jüni P, Katus HA, Knuuti J, Lancellotti P, Leclercq C, McDonagh T, Piepoli MF, Ponikowski P, Richter DJ, Roffi M, Shlyakhto E, Simpson IA, Zamorano JL, Chettibi M, Hayrapetyan HG, Metzler B, Ibrahimov F, Sujayeva V, Beauloye C, Dizdarevic-Hudic L, Karamfiloff K, Skoric B, Antoniades L, Tousek P, Terkelsen PJ, Shaheen SM, Marandi T, Niemelä M, Kedev S, Gilard M, Aladashvili A, Elsaesser A, Kanakakis IG, Merkely B, Gudnason T, Iakobishvili Z, Bolognese L, Berkinbayev S, Bajraktari G, Beishenkulov M, Zake I, Lamin H Ben, Gustiene O, Pereira B, Xuereb RG, Ztot S, Juliebø V, Legutko J, Timóteo AT, Tatu-Chiţoiu G, Yakovlev A, Bertelli L, Nedeljkovic M, Studenčan M, Bunc M, García de Castro AM, Petursson P, Jeger R, Mourali MS, Yildirir A, Parkhomenko A, Gale CP (2018) 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation. *Eur. Heart J.* **39**, 119–177.

[51] Roffi M, Patrono C, Collet J-P, Mueller C, Valgimigli M, Andreotti F, Bax JJ, Borger MA, Brotons C, Chew DP, Gencer B, Hasenfuss G, Kjeldsen K, Lancellotti P,

Landmesser U, Mehilli J, Mukherjee D, Storey RF, Windecker S, Baumgartner H, Gaemperli O, Achenbach S, Agewall S, Badimon L, Baigent C, Bueno H, Bugiardini R, Carerj S, Casselman F, Cuisset T, Erol Ç, Fitzsimons D, Halle M, Hamm C, Hildick-Smith D, Huber K, Iliodromitis E, James S, Lewis BS, Lip GYH, Piepoli MF, Richter D, Rosemann T, Sechtem U, Steg PG, Vrints C, Luis Zamorano J, Zamorano JL, Aboyans V, Achenbach S, Agewall S, Badimon L, Barón-Esquivias G, Baumgartner H, Bax JJ, Bueno H, Carerj S, Dean V, Erol Ç, Fitzsimons D, Gaemperli O, Kirchhof P, Kolh P, Lancellotti P, Lip GY, Nihoyannopoulos P, Piepoli MF, Ponikowski P, Roffi M, Torbicki A, Carneiro AV, Windecker S, Chilingaryan A, Weidinger F, Najafov R, Sinnaeve PR, Terzic I, Postadzhiyan A, Milicic D, Eftychiou C, Widimsky P, Bang L, El Etriby A, Marandi T, Pietilä M, Kedev S, Koning R, Aladashvili A, Neumann F-J, Tsioufis K, Becker D, Guðnason T, Matetzky S, Bolognese L, Mussagaliyeva A, Beishenkulov M, Latkovskis G, Serpytis P, Pereira B, Magri CJ, Grosu A, Abir-Khalil S, Larsen AI, Budaj A, Mimoso JMV, Ginghina C, Averkov O, Nedeljkovic MA, Studencan M, Barrabés JA, Held C, Rickli H, Peters RJ, Mourali MS, Atalar E, Swanson N, Parkhomenko A (2016) 2015 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation. *Eur. Heart J.* **37**, 267–315.

[52] SCHULMAN S, KEARON C (2005) Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients. *J. Thromb. Haemost.* **3**, 692–694.

[53] Hart RG, Diener H-C, Coutts SB, Easton JD, Granger CB, O'Donnell MJ, Sacco RL, Connolly SJ, Cryptogenic Stroke/ESUS International Working Group (2014) Embolic strokes of undetermined source: the case for a new clinical construct. *Lancet Neurol.* **13**, 429–438.

[54] Task Force for the Diagnosis and Management of Syncope D in collaboration, European Society of Cardiology (ESC) EHRA, European Heart Rhythm Association (EHRA) HFA, Heart Failure Association (HFA) and HRS, Heart Rhythm Society (HRS) E by the following, Moya A, Sutton R, Ammirati F, Blanc J-J, Brignole M, Dahm JB, Deharo J-C,

Gajek J, Gjesdal K, Krahn A, Massin M, Pepi M, Pezawas T, Ruiz Granell R, Sarasin F, Ungar A, van Dijk JG, Walma EP, Wieling W, Pepi M, Pezawas T, Granell RR, Sarasin F, Ungar A, Dijk JG van, Walma EP, Wieling W, Contributors E, Abe H, Benditt DG, Decker WW, Grubb BP, Kaufmann H, Morillo C, Olshansky B, Parry SW, Sheldon R, Shen WK, (CPG) EC for PG, Vahanian A, Auricchio A, Bax J, Ceconi C, Dean V, Filippatos G, Funck-Brentano C, Hobbs R, Kearney P, McDonagh T, McGregor K, Popescu BA, Reiner Z, Sechtem U, Sirnes PA, Tendera M, Vardas P, Widimsky P, Reviewers D, Auricchio A, Acarturk E, Andreotti F, Asteggiano R, Bauersfeld U, Bellou A, Benetos A, Brandt J, Chung MK, Cortelli P, Costa A Da, Extramiana F, Ferro J, Gorenek B, Hedman A, Hirsch R, Kaliska G, Kenny RA, Kjeldsen KP, Lampert R, Mølgard H, Paju R, Puodziukynas A, Raviele A, Roman P, Scherer M, Schondorf R, Sicari R, Vanbrabant P, Wolpert C, Zamorano JL (2009) Guidelines for the diagnosis and management of syncope (version 2009). *Eur. Heart J.* **30**, 2631–71.

[55] Thygesen K, Alpert JS, Jaffe AS, Simoons ML, Chaitman BR, White HD, Thygesen K, Alpert JS, White HD, Jaffe AS, Katus HA, Apple FS, Lindahl B, Morrow DA, Chaitman BR, Clemmensen PM, Johanson P, Hod H, Underwood R, Bax JJ, Bonow RO, Pinto F, Gibbons RJ, Fox KA, Atar D, Newby LK, Galvani M, Hamm CW, Uretsky BF, Steg PG, Wijns W, Bassand J-P, Menasche P, Ravkilde J, Ohman EM, Antman EM, Wallentin LC, Armstrong PW, Simoons ML, Januzzi JL, Nieminen MS, Gheorghiade M, Filippatos G, Luepker R V., Fortmann SP, Rosamond WD, Levy D, Wood D, Smith SC, Hu D, Lopez-Sendon J-L, Robertson RM, Weaver D, Tendera M, Bove AA, Parkhomenko AN, Vasilieva EJ, Mendis S, Thygesen K, Alpert JS, White HD, Jaffe AS, Katus HA, Apple FS, Lindahl B, Morrow DA, Chaitman BR, Clemmensen PM, Johanson P, Hod H, Underwood R, Bax JJ, Bonow RO, Pinto F, Gibbons RJ, Fox KA, Atar D, Newby LK, Galvani M, Hamm CW, Uretsky BF, Steg PG, Wijns W, Bassand J-P, Menasche P, Ravkilde J, Ohman EM, Antman EM, Wallentin LC, Armstrong PW, Simoons ML, Januzzi JL, Nieminen MS, Gheorghiade M, Filippatos G, Luepker R V., Fortmann SP, Rosamond WD, Levy D, Wood D, Smith SC, Hu D, Lopez-Sendon J-L, Robertson RM,

Weaver D, Tendera M, Bove AA, Parkhomenko AN, Vasilieva EJ, Mendis S (2012) Third Universal Definition of Myocardial Infarction. *J. Am. Coll. Cardiol.* **60**, 1581–1598.

[56] Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, Falk V, González-Juanatey JR, Harjola V-P, Jankowska EA, Jessup M, Linde C, Nihoyannopoulos P, Parissis JT, Pieske B, Riley JP, Rosano GMC, Ruilope LM, Ruschitzka F, Rutten FH, van der Meer P, Filippatos G, McMurray JJ V, Aboyans V, Achenbach S, Agewall S, Al-Attar N, Atherton JJ, Bauersachs J, John Camm A, Carerj S, Ceconi C, Coca A, Elliott P, Erol Ç, Ezekowitz J, Fernández-Golfín C, Fitzsimons D, Guazzi M, Guenoun M, Hasenfuss G, Hindricks G, Hoes AW, Iung B, Jaarsma T, Kirchhof P, Knuuti J, Kolh P, Konstantinides S, Lainscak M, Lancellotti P, Lip GYH, Maisano F, Mueller C, Petrie MC, Piepoli MF, Priori SG, Torbicki A, Tsutsui H, van Veldhuisen DJ, Windecker S, Yancy C, Zamorano JL, Zamorano JL, Aboyans V, Achenbach S, Agewall S, Badimon L, Barón-Esquivias G, Baumgartner H, Bax JJ, Bueno H, Carerj S, Dean V, Erol Ç, Fitzsimons D, Gaemperli O, Kirchhof P, Kolh P, Lancellotti P, Lip GYH, Nihoyannopoulos P, Piepoli MF, Ponikowski P, Roffi M, Torbicki A, Vaz Carneiro A, Windecker S, Sisakian HS, Isayev E, Kurlianskaya A, Mullens W, Tokmakova M, Agathangelou P, Melenovsky V, Wiggers H, Hassanein M, Uuetoa T, Lommi J, Kostovska ES, Juillière Y, Aladashvili A, Luchner A, Chrysohoou C, Nyolczas N, Thorgeirsson G, Marc Weinstein J, Di Lenarda A, Aidargaliyeva N, Bajraktari G, Beishenkulov M, Kamzola G, Abdel-Massih T, Celutkiene J, Noppe S, Cassar A, Vataman E, Abir-Khalil S, van Pol P, Mo R, Straburzynska-Migaj E, Fonseca C, Chioncel O, Shlyakhto E, Otasevic P, Goncalvesová E, Lainscak M, Díaz Molina B, Schaufelberger M, Suter T, Yilmaz MB, Voronkov L, Davies C (2016) 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur. Heart J.* **37**, 2129–2200.

[57] Reynolds HR, Hochman JS (2008) Cardiogenic Shock Current Concepts and Improving Outcomes.

[58] Cannon JW (2018) Hemorrhagic Shock. *N. Engl. J. Med.* **378**, 370–379.

[59] Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, Cohen J, Opal SM,

Vincent J-L, Ramsay G, Conference for the ISD (2003) 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Intensive Care Med.* **29**, 530–538.

[60] Kellum JA, Lameire N, KDIGO AKI Guideline Work Group (2013) Diagnosis, evaluation, and management of acute kidney injury: a KDIGO summary (Part 1). *Crit. Care* **17**, 204.

[61] Burt CC, Arrowsmith JE (2009) Respiratory failure. *Surg.* **27**, 475–479.

[62] Mannino DM, Doherty DE, Sonia Buist A (2006) Global Initiative on Obstructive Lung Disease (GOLD) classification of lung disease and mortality: findings from the Atherosclerosis Risk in Communities (ARIC) study. *Respir. Med.* **100**, 115–22.

[63] Aboyans V, Ricco J-B, Bartelink M-LEL, Björck M, Brodmann M, Cohnert T, Collet J-P, Czerny M, De Carlo M, Debus S, Espinola-Klein C, Kahan T, Kownator S, Mazzolai L, Naylor AR, Roffi M, Röther J, Sprynger M, Tendera M, Tepe G, Venermo M, Vlachopoulos C, Desormais I, Widimsky P, Kolh P, Agewall S, Bueno H, Coca A, De Borst GJ, Delgado V, Dick F, Erol C, Ferrini M, Kakkos S, Katus HA, Knuuti J, Lindholt J, Mattle H, Pieniazek P, Piepoli MF, Scheinert D, Sievert H, Simpson I, Sulzenko J, Tamargo J, Tokgozoglu L, Torbicki A, Tsakountakis N, Tuñón J, de Ceniga MV, Windecker S, Zamorano JL, Windecker S, Aboyans V, Agewall S, Barbato E, Bueno H, Coca A, Collet J-P, Coman IM, Dean V, Delgado V, Fitzsimons D, Gaemperli O, Hindricks G, Iung B, Juni P, Katus HA, Knuuti J, Lancellotti P, Leclercq C, McDonagh T, Piepoli MF, Ponikowski P, Richter DJ, Roffi M, Shlyakhto E, Simpson IA, Zamorano JL, Zelveian PH, Haumer M, Isachkin D, De Backer T, Dilic M, Petrov I, Kirhmajer MV, Karetova D, Prescott E, Soliman H, Paapstel A, Makinen K, Tosev S, Messas E, Pagava Z, Müller OJ, Naka KK, Járai Z, Gudjonsson T, Jonas M, Novo S, Ibrahimi P, Lunegova O, Dzerve V, Misonis N, Beissel J, Pllaha E, Taberkant M, Bakken T, Teles R, Lighezan D, Konradi A, Zavatta M, Madaric J, Fras Z, Melchor LS, Näslund U, Amann-Vesti B, Obiekezie A (2018) 2017 ESC Guidelines on the Diagnosis and Treatment of Peripheral Arterial Diseases, in collaboration with the European Society for Vascular Surgery

(ESVS). *Eur. Heart J.* **39**, 763–816.

[64] Levey AS, Eckardt K-U, Tsukamoto Y, Levin A, Coresh J, Rossert J, Zeeuw DDE, Hostetter TH, Lameire N, Eknoyan G (2005) Definition and classification of chronic kidney disease: A position statement from Kidney Disease: Improving Global Outcomes (KDIGO). *Kidney Int.* **67**, 2089–2100.

[65] ESC GUIDELINES 2013 ESC guidelines on the management of stable coronary artery disease.

[66] Beutler E, Waalen J (2006) The definition of anemia: what is the lower limit of normal of the blood hemoglobin concentration? *Blood* **107**, 1747–50.

[67] Rosendaal FR, Cannegieter SC, van der Meer FJ, Briët E (1993) A method to determine the optimal intensity of oral anticoagulant therapy. *Thromb. Haemost.* **69**, 236–9.

[68] Hanley JA, McNeil BJ (1983) A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* **148**, 839–43.

[69] de Oliveira MCF, Levkowitz H (2003) From visual data exploration to visual data mining: A survey. *IEEE Trans. Vis. Comput. Graph.* **9**, 378–394.

[70] de Oliveira MCF, Levkowitz H (2003) From visual data exploration to visual data mining: A survey. *IEEE Trans. Vis. Comput. Graph.* **9**, 378–394.

[71] Maaten L van der, Hinton G (2008) Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605.

[72] https://jef.works/tsne-online/, Interactive t-SNE Online, Last updated 2018, Accessed on 2018.

[73] Maaten L van der, Hinton G (2008) Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605.

[74] Lum PY, Singh G, Lehman A, Ishkanov T, Vejdemo-Johansson M, Alagappan M, Carlsson J, Carlsson G (2013) Extracting insights from the shape of complex data using topology. *Sci. Rep.* **3**, 1236.

[75] Singh G, Mémoli F, Carlsson G (2007) *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition*.

[76] Saul N, Veen HJ van (2017) MLWave/kepler-mapper: 186f. https://kepler-mapper.scikit-org/index.html#.

[77] Chi Square test for feature selection.

[78] Olson, Randal S; Urbanowicz, Ryan J; Andrews, Peter C; Lavender, Nicole A; La Creis, Kidd; Moore JH Automating biomedical data science through tree-based pipeline optimization (https://epistasislab.github.io/tpot/). *Appl. Evol. Comput.* 123–137.

[79] Kuipers S, Klouwenberg PMK, Cremer OL (2014) Incidence, risk factors and outcomes of new-onset atrial fibrillation in patients with sepsis: a systematic review. *Crit. Care* **18**, 688.

[80] Melgaard L, Gorst-Rasmussen A, Lane DA, Rasmussen LH, Larsen TB, Lip GYH (2015) Assessment of the CHA $_2$ DS $_2$ -VASc Score in Predicting Ischemic Stroke, Thromboembolism, and Death in Patients With Heart Failure With and Without Atrial Fibrillation. *JAMA* **314**, 1030.

[81] Boriani G, Fauchier L, Aguinaga L, Beattie JM, Blomstrom Lundqvist C, Cohen A, Dan G-A, Genovesi S, Israel C, Joung B, Kalarus Z, Lampert R, Malavasi VL, Mansourati J, Mont L, Potpara T, Thornton A, Lip GYH, Gorenek B, Marin F, Dagres N, Ozcan EE, Lenarczyk R, Crijns HJ, Guo Y, Proietti M, Sticherling C, Huang D, Daubert JP, Pokorney SD, Cabrera Ortega M, Chin A (2018) European Heart Rhythm Association (EHRA) consensus document on management of arrhythmias and cardiac electronic devices in the critically ill and post-surgery patient, endorsed by Heart Rhythm Society (HRS), Asia Pacific Heart Rhythm Society (APHRS), Cardiac Arrhythmia Society of Southern Africa (CASSA), and Latin American Heart Rhythm Society (LAHRS). *EP Eur.*

[82] Costantino G, Podda GM, Falsetti L, Iannone P, Lages A, Marra AM, Masala M, Reiakvam OM, Savva F, Schovanek J, van Bree S, da Silva Chora IJ, Privitera G, Ragozzino S, von Rotz M, Woittiez L, Davidson C, Montano N (2017) Guidelines on the management of atrial fibrillation in the emergency department: a critical appraisal. *Intern. Emerg. Med.* **12**, 693–703.

[83] Vincent J-L, Singer M (2010) Critical care: advances and future perspectives. *Lancet*

(London, England) **376**, 1354–61.

[84] Anthony Celi L, Mark RG, Stone DJ, Montgomery RA (2013) “Big Data” in the Intensive Care Unit. Closing the Data Loop. *Am. J. Respir. Crit. Care Med.* **187**, 1157–1160.

[85] Flechet M, Güiza F, Schetz M, Wouters P, Vanhorebeek I, Derese I, Gunst J, Spriet I, Casaer M, Van den Berghe G, Meyfroidt G (2017) AKIpredictor, an online prognostic calculator for acute kidney injury in adult critically ill patients: development, validation and comparison to serum neutrophil gelatinase-associated lipocalin. *Intensive Care Med.* **43**, 764–773.

[86] Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJGM, Lip GYH (2010) A Novel User-Friendly Score (HAS-BLED) To Assess 1-Year Risk of Major Bleeding in Patients With Atrial Fibrillation: The Euro Heart Survey. *Chest* **138**, 1093–1100.

[87] Gage BF, Yan Y, Milligan PE, Waterman AD, Culverhouse R, Rich MW, Radford MJ (2006) Clinical classification schemes for predicting hemorrhage: Results from the National Registry of Atrial Fibrillation (NRAF). *Am. Heart J.* **151**, 713–719.

[88] Le Gall JR, Lemeshow S, Saulnier F A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* **270**, 2957–63.

[89] Knaus WA, Draper EA, Wagner DP, Zimmerman JE (1985) APACHE II: a severity of disease classification system. *Crit. Care Med.* **13**, 818–29.

[90] Capuzzo M, Valpondi V, Sgarbi A, Bortolazzi S, Pavoni V, Gilli G, Candini G, Gritti G, Alvisi R (2000) Validation of severity scoring systems SAPS II and APACHE II in a single-center population. *Intensive Care Med.* **26**, 1779–85.

[91] Alves GC, Silva Júnior GB da, Lima RSA, Sobral JB, Mota RMS, Abreu KLS de, Rocha NA, Nogueira CB, Daher EDF (2010) Fatores de risco para óbito em pacientes idosos gravemente enfermos. *Rev. Bras. Ter. Intensiva* **22**, 138–143.

[92] Li G, Cook DJ, Thabane L, Friedrich JO, Crozier TM, Muscedere J, Granton J, Mehta S, Reynolds SC, Lopes RD, Lauzier F, Freitag AP, Levine MAH, PROTECT Investigators for the Canadian Critical Care Trials Group, and the Australian and New Zealand Intensive Care Society Clinical Trials Group the PI for the CCCT, Group and the A and

NZICSCT (2016) Risk factors for mortality in patients admitted to intensive care units with pneumonia. *Respir. Res.* **17**, 80.

[93] Rivera-Fernández R, Navarrete-Navarro P, Fernández-Mondejar E, Rodriguez-Elvira M, Guerrero-López F, Vázquez-Mata G (2006) Six-year mortality and quality of life in critically ill patients with chronic obstructive pulmonary disease. *Crit. Care Med.* **34**, 2317–2324.

Appendix 1: Ethics Committee Approval (1/3)

Comitato Etico Regionale delle Marche
c/o Azienda Ospedaliero-Universitaria Ospedali Riuniti
Via Conca, 71 - 60126 Torrette di Ancona
Presidente Prof. Paolo Pelaia

Seduta del 21-06-2018
Prot. 2018 168

Il Comitato Etico Regionale delle Marche (CERM), istituito con determina n. 402 del 9/5/2017 in esecuzione della DGRM n. 244 del 20/3/2017, è organizzato ed opera secondo quanto indicato dal D.M. 8 febbraio 2013 e s.m.i., e per le sue decisioni ed attività relative alle sperimentazioni cliniche fa riferimento, alle norme di "Good Clinical Practice" (ICH-GCP) nella versione più recente.
Il Comitato Etico nella seduta del **21-06-2018** ha preso in considerazione il protocollo della sperimentazione dal titolo *"AFICILL (Atrial Fibrillation in the Critically Ill)"* - codice protocollo: AFICILL - (sponsor/promotore:Dott. Lorenzo Falsetti) proposto dal Dott. Lorenzo Falsetti, SOD PRONTO SOCCORSO E MEDICINA D'URGENZA, A.O.U. Ospedali Riuniti Ancona.

Dopo aver esaminato:

Titolo	Ver	Data Ver	Nota	File
Modulo 10. Dichiarazione di conformità al Decreto Ministero della Salute 17.12.2004	1.0	24-04-2018		Modulo.n.10.pdf
Modulo 11. Modello di dichiarazione pubblica sul conflitto di interessi dello sperimentatore	1.0	24-04-2018		Modulo.n.11.pdf
Curriculum dello sperimentatore	1.0	02-05-2018		CV Lorenzo Falsetti.pdf
Modulo 7. Domanda di autorizzazione aziendale e richiesta di parere per sperimentazioni cliniche non interventistiche - no profit	1.0	24-04-2018		Modulo.n.7_bis.pdf
Lettera di intenti	1.0	24-05-2018		Lettera Intenti CERM.pdf
Modulo 8. Dichiarazione sulla natura osservazionale centro coordinatore	1.0	24-05-2018		Modulo n.8_b.pdf
Elenco dei centri partecipanti (solo per sperimentazioni multicentriche)	1.0	24-04-2018		Elenco Centri Partecipanti.docx
Informativa e manifestazione del consenso al trattamento dei dati personali	1.0	24-04-2018		Informativa su Trattamento Dati Personali.docx

Segreteria Tecnico Scientifica
Tel. 071 596 3135-3667
comitato.etico@ospedaliriuniti.marche.it

COMITATO ETICO REGIONALE
DELLE MARCHE

Via Conca,

Appendix 1: Ethics Committee Approval (2/3)

Comitato Etico Regionale delle Marche
c/o Azienda Ospedaliero-Universitaria Ospedali Riuniti
Via Conca, 71 - 60126 Torrette di Ancona
Presidente Prof. Paolo Pelaia

Scheda clinica per la raccolta dati (Case Report Form)	1.0	24-04-2018		Scheda Clinica Raccolta Dati_PDF.pdf
Modulo per il consenso informato	1.0	24-04-2018		Modulo per Consenso Informato.docx
Elenco in word della documentazione sottomessa	1.0	24-04-2018		Elenco Documenti Presentati.docx
Protocollo sperimentale e relativi allegati	2.0	11-06-2018	Protocollo aggiornato secondo le richieste del CERM 31/5/2018	AFICILL Protocollo Aggiornato 2.0.docx

Valutata l'idoneità della struttura ove sarà effettata la ricerca e l'idoneità dello sperimentatore all'effettuazione dello specifico studio, il Comitato esprime **parere favorevole** alla sperimentazione in esame con la raccomandazione di utilizzare le definizioni "mortalità" o "probabilità di decesso" invece del termine "prevalenza di mortalità".
Ogni cambiamento del protocollo di studio dovrà essere trasmesso al Comitato Etico.
Si richiede che questo Comitato Etico venga informato dell'inizio della sperimentazione e della sua conclusione o eventuale interruzione. Inoltre dovrà essere informato di ogni successivo emendamento al protocollo e degli eventi avversi, seri o inattesi insorti nel corso dello studio, che potrebbero influire sulla sicurezza dei soggetti o sul proseguimento dello studio.

IL PRESIDENTE
Prof. Paolo Pelaia

COMITATO ETICO REGIONALE
DELLE MARCHE (CERM)
c/o Azienda Ospedaliero-Universitaria
Ospedali Riuniti
Via Conca, 71 - 60126 Torrette - ANCONA

Segreteria Tecnico Scientifica
Tel. 071 596 3135-3667
comitato.etico@ospedaliriuniti.marche.it

Appendix 1: Ethics Committee Approval (3/3)

Elenco presenze della Riunione del Comitato Etico Regionale delle Marche (C.E.R.M.) del 21 giugno 2018

NOMINATIVO COMPONENTE	CARICA	ENTE DI RIFERIMENTO	RUOLO	PRESENZA SI/NO
PROF. PAOLO PELAIA	PRESIDENTE CERM	AOU OSP.RIUNTI	CLINICO	SI
PROF. SALVATORE AMOROSO		AOU OSP.RIUNTI	FARMACOLOGO	NO
DOTT. VINCENZO BERDINI		ESTERNO	MMGT	NO
DOTT. STEFANO BIANCHI	In relazione a studi svolti presso Marche Nord	A.O. MARCHE NORD	DIRETTORE SANITARIO (SOSTITUTO PERMANENTE)	SI
DOTT. GIUSEPPE BRAICO		ESTERNO	PEDIATRA LIBERA SCELTA	SI
DOTT. MARINO BRUNORI		A.O. MARCHE NORD	CLINICO	NO
PROF. FLAVIA CARLE		ESTERNO (UNIVPM)	BIOSTATISTICO	SI
DOTT.SSA VALENTINA COLA		AOU OSP.RIUNITI	ESPERTO DISPOSITIVI MEDICI	SI
ING. GIANCARLO CONTI	In relazione all'area medico-chirurgia con dispositivo medico	A.O. MARCHE NORD	INGEGNERE CLINICO	NO
PROF. MARCELLO M. D'ERRICO	PRESIDENTE VICARIO E RESPONSABILE STS	AOU OSP.RIUNITI	CLINICO	SI
DOTT.SSA EMMA ESPINOSA		A.O. MARCHE NORD	CLINICO	SI
DOTT. DOMENICO GABRIELLI		ASUR	CLINICO	NO
DOTT. PIERO GALIENI		ASUR	CLINICO	SI
DOTT. MICHELE GENTILI		ASUR	FARMACISTA SSR	NO
DOTT. MICHELE GIUA		ESTERNO	RAPPRESENTANTE VOLONTARIATO O DELL'ASSOCIAZIONISMO DI TUTELA DEI PAZIENTI	SI
DOTT. COSTANTINO GOBBI		ESTERNO	PEDIATRA LIBERA SCELTA	SI
DOTT. VINCENZO LARICCIA		ESTERNO (UNIVPM)	BIOTECNOLOGO	SI
DOTT.SSA STEFANIA MAGGI	In relazione a studi con nuove procedure tecniche, diagnostiche e terapeutiche, invasive e semi invasive	A.O. OSPEDALI RIUNITI	CLINICO	NO
DOTT. MASSIMILIANO MARINELLI		ESTERNO	BIOETICISTA	SI
DOTT. ANDREA MARINOZZI		A.O. OSPEDALI RIUNITI	FARMACISTA SSR	SI
DOTT. VINCENZO MASSETTI		ESTERNO	RAPPRESENTANTE VOLONTARIATO O DELL'ASSOCIAZIONISMO DI TUTELA DEI PAZIENTI	SI
DOTT. RODOLFO MATTIOLI		A.O. MARCHE NORD	CLINICO	SI
PROF SSA LAURA MAZZANTI	In relazione a studi di prodotti alimentari	ESTERNO (UNIVPM)	ESPERTO IN NUTRIZIONE	SI
DOTT. FRANCESCO PELLEGRINI		ESTERNO	CLINICO	SI
DOTT. FRANCESCO PAOLO PERRI		ASUR MARCHE	PEDIATRA DI LIBERA SCELTA	SI
DOTT. GIANLUCA SERAFINI	In relazione a studi svolti presso A.O. Ospedali Riuniti	A.O. OSPEDALI RIUNITI	DIRETTORE SANITARIO	NO
DOTT. PAOLO SIGNORE		ESTERNO	MMGT	SI
DOTT.SSA ROSA RITA SILVA	VICE PRESIDENTE	ASUR MARCHE	CLINICO	SI
DOTT.SSA ELISABETTA SIMONETTI		A.O. OSPEDALI RIUNITI	RAPPRESENTANTE AREE SANITARIE	SI
DOTT.SSA NADIA STORTI	In relazione a studi svolti presso ASUR Marche	ASUR MARCHE	DIRETTORE SANITARIO	NO
PROF. ADRIANO TAGLIABRACCI		A.O. OSPEDALI RIUNITI	MEDICO LEGALE	NO
DOTT. MARCELLO TAVIO		A.O. OSPEDALI RIUNITI	CLINICO	SI
DOTT.SSA GIADA TORTORA	In relazione studi di genetica	ESTERNO	ESPERTO IN GENETICA	SI

Componenti del CERM presenti che non hanno partecipato alla votazione: Dott. Stefano Bianchi;