1875-0362/20

Send Orders for Reprints to reprints@benthamscience.net

The Open Bioinformatics Journal

Content list available at: https://openbioinformaticsjournal.com

SYSTEMATIC REVIEW

Machine Learning Algorithms in Cardiology Domain: A Systematic Review

Aleksei Dudchenko[1,2], Matthias Ganzinger[1] and Georgy Kopanitsa[2,*]

[1] *Institute of Medical Biometry and Informatics, Heidelberg University, Heidelberg, Germany*

[2] *Center for Cognitive Technologies, ITMO University, Saint-Petersburg, Russia*

Abstract:

Background:

It could be seen in the previous decades that Machine Learning (ML) has a huge variety of possible implementations in medicine and can be of great use. Nevertheless, cardiovascular diseases cause about a third of the total global deaths.

Does ML work in the cardiology domain and what is the current progress in this regard? To answer this question, we present a systematic review aiming at 1) identifying studies where machine learning algorithms were applied in the domain of cardiology; 2) providing an overview based on the existing literature about the state-of-the-art ML algorithms applied in cardiology.

Methods:

For organizing this review, we adopted the PRISMA statement. We used PubMed as the search engine and identified the search keywords as "Machine Learning", "Data Mining", "Cardiology", and "Cardiovascular" in combinations. Scientific articles and conference papers published between 2013-2017 reporting about implementations of ML algorithms in the domain of cardiology have been included in this review.

Results:

In total, 27 relevant papers were included. We examined four aspects: the aims of ML systems, the methods, datasets, and evaluation metrics. The major part of the paper was aimed at predicting the risk of mortality. A promising branch of Machine Learning, the 'Reinforcement Learning', was also never proposed in the observed papers. Tree-based ensembles are common and show good results, whereas deep neural networks are poorly represented. Most papers (20 of 27) have used datasets that are hardly available for other researchers, *e.g*. unpublished local registries. We also identified 28 different metrics for model evaluation. This variety of metrics makes it difficult to compare the results of different researches.

Conclusion:

We suppose that this systematic review will be helpful for researchers developing medical machine learning systems and for cardiology in particular.

Keywords: Cardiology, Systematic review, Machine learning, Neural networks, Natural Language Processing (NLP), PubMed.

Article History	Received: October 30, 2019	Revised: February 06, 2020	Accepted: February 11, 2020

1. INTRODUCTION

1.1. Background

Cardiovascular diseases (CVD) are a group of disorders of the heart and blood vessels. Data from the World Health Organization shows that CVDs are a leading cause of deaths worldwide for both sexes and all ages. In particular, CVDs caused 17.3 million deaths in 2013. This is 45% of all non-communicable disease deaths and 31.5% of all global deaths. More deaths worldwide were caused by CVDs than all communicable, maternal, neonatal, and nutritional disorders combined, which is twice more than those caused by cancer [1].

These facts illustrate the importance of dealing with CVDs. Artificial intelligence and clinical decision support can help doctors provide better and more personalized treatment to their patients. A lot of efforts have been applied during the previous years to implement clinical decision support systems. A big class of clinical decision support systems is based on machine learning. It has been shown in the previous decades that machine learning, as well as other branches of artificial

[*] Address correspondence to this author at the Center for Cognitive Technologies, ITMO University, Saint-Petersburg, Russia;
E-mail: georgy.kopanitsa@gmail.com

intelligence (AI), has a broad variety of possible implementations in medicine and can be very helpful. The first and currently used definition of ML was proposed by A. Samuel [2]: "ML is a field of study that gives computers the ability to learn without being explicitly programmed". An ML program can learn from medical data that has been collected by physicians and devices for years to make predictions, prognosis, or diagnosis.

There are many particular applications of AI and ML algorithms as tools to support decision making for different medical tasks. For instance, artificial intelligence classifiers have been used in urology diagnosis [3], in oncology and breast cancer diagnosis [4 - 6], in the diagnosis of hypoglycemic episodes [7], in skin cancer classification and diagnosis [8, 9], as well as for medical image analysis [10].

In our research, we aimed to identify and analyze the current applications of ML algorithms that are employed in the cardiology domain and presented in recent scientific papers.

1.2. Related Works

There are previous works devoted to the analysis of different aspects of artificial intelligence systems in medicine. We discuss some of them below:

Considered the main characteristics of predictive clinical data mining and focused on two specific aspects: the methods able to deal with temporal data and the efforts performed to build data mining models based on the results of molecular medicine.

A systematic review by Palaniappan *et al*. [11] examined the processing of sensor data, signal processing, classification, and statistical methods to analyze lung sounds reported in previous research.

Review of the Literature by Triantafyllidis *et al*. [12] observed applications of machine learning in real-life digital health interventions, aiming to improve the understanding of researchers, clinicians, engineers, and policymakers in developing robust and impactful data-driven interventions in the health care domain. The finding of the review is the fact that health interventions engaging machine learning algorithms in real-life studies can be useful and effective. The authors also reported about the necessity to conduct further studies in intervention settings following evaluation principles and demonstrating the potential of machine learning in clinical practice.

The survey by Wallert *et al*. [13] described biomedical information systems for decision support, their application protocols and methodologies, and also suggested the future challenges and directions.

Melillo *et al*., in their review [14 - 20], underlined that clusters of computer technologies are used for pain management by processing clinical data for the development of clinical decision support systems (CDSS). The clusters are rule-based algorithms, artificial neural networks, nonstandard set theory, and statistical learning algorithms. The authors detected methodologies for content processing such as terminologies, questionnaires, and scores.

Machine learning in lung sound analysis was examined by Kalidas *et al*. [19]. The authors highlighted specific lung sounds/disorders, the number of subjects, the signal processing and classification methods, and the outcome of lung sounds analysis using machine learning methods based on previous research. This review also contains recommendations for further improvements.

The survey by Eerikainen *et al*. [20] presents existing clinical decision support systems, recapitulates actual data on the application and impact of clinical decision support systems in practice, and recommendations for using these systems outside the research.

Applying natural language processing (NLP) techniques in healthcare was considered in the paper by Rajagopalan *et al*. [21]. This review provides the concept of NLP, the applications of NLP, and the challenges of NLP systems in healthcare.

The review "Using data mining techniques in heart disease diagnosis and treatment" [22] indicates shortcomings in the research on heart disease diagnosis and suggests a model to systematically resolve those shortcomings.

Despite the great interest in the topic, a number of issues have not been considered in detail. The existing reviews do not reflect the aspects people developing their AI systems might be interested in. For instance, what ML methods were applied and how efficient they were; in what way the efficiency of the algorithms and systems is usually measured and which metrics are used if so; what kind of data is used for such systems and what can be a source of the data. An overview of these questions can be relevant for researchers and developers who are going to develop an ML system for a particular medical field. We have discussed some of these issues as the objectives of our study.

There has been an increasing interest in the application of artificial intelligence since 2014 in the industry [17]. In this regard, we have considered the papers published since 2013, for a systematic review.

1.3. Objectives

This review aims at 1) identifying studies where machine learning algorithms were applied in the cardiology domain; 2) providing an overview based on the identified literature of the state-of-the-art ML algorithms applied in cardiology.

None of these aims have been reported in detail in previous works. It makes this systematic review a significant contribution to developing the field by helping and guiding researchers on what can be useful in their work and what can be a further way in their research and development.

2. RESEARCH METHODS

For organizing this review, we have employed the PRISMA statement. PRISMA is a set of items for reporting systematic reviews and meta-analyses that are focused on reporting reviews and evaluating randomized trials but can also be used as a basis for reporting systematic reviews [14, 15]. For the review, we have adopted the PRISMA statement and have identified the following items: review questions, information sources, search strategy, and selection criteria.

2.1. Review Questions

We analyzed the studies in terms of the following four questions.

Aims of the system. What is the system focused on and what is the output of the system?

We identify papers that report on the systems engaging ML methods in the domain of cardiology. We want to find out what tasks are the systems aimed at and what kind of output they produce.

Methods and algorithms. What algorithms were applied in the system?

There are plenty of ML algorithms that can be used to predict, classify, or estimate medical data. Different algorithms could be more suitable and efficient or less dependent on specific data or a particular task. We aimed at identifying what algorithms were implemented in research projects that applied ML techniques in cardiology.

Data sources. What dataset is used? How big is it? How many features/parameters does it have?

Any machine learning system needs a relevant dataset to be trained and validated. One of the biggest issues in developing a machine learning system is to get data for training and evaluation. We explored the datasets that were used in the observed researches, how many features and samples they had, and whether other researchers could access the dataset.

Algorithm evaluation. What metrics were used to evaluate the system?

Every system needs to be evaluated and there are many different metrics to evaluate them. We examined observed papers to identify and analyze the metrics that were engaged to evaluate the system.

2.2. Bibliographic Search Process

We identified the search keywords *machine learning*, *data mining, cardiology,*and *cardiovascular*. The keywords were combined in the search statement as machine learning OR data mining AND cardiology OR cardiovascular. PubMed has been employed as the search engine.

2.3. Selection Criteria

Table **1** provides the inclusion and exclusion criteria that were applied to select papers. We included scientific articles and conference papers in English published between 2013 and 2017 devoted to the application of ML methods in the field of cardiology. In addition to the criteria listed in Table **1**, we did not consider the papers where we could not clearly identify information regarding our review questions.

3. RESULTS

Bibliographic search identified 372 papers (Fig. **1**). Screening by titles revealed that 218 works were not related to the topic. Further screening by abstracts excluded 33 works not related to the topic and 54 papers that did not cover the implementation of any ML method. Full-text examination of 67 works led to excluding 8 more works as those were not related to the topic, 11 for not containing any ML method implementation, and 21 for not being scientific articles (review, dissertation/thesis, book chapter, comparative study, *etc*). Finally, 27 scientific articles or conference papers written in English and reporting about the implementation of an ML method or algorithm in cardiology were included in this review.

Table 1. Inclusion criteria.

Facets	Inclusion Criteria	Exclusion Criteria
Published	between 2013-2017	until 2013
Study Type	scientific article or proceedings paper	not a scientific article (review, dissertation/thesis, book chapter, comparative study, *etc.*)
Research content	devoted to the application of ML methods in cardiology	not related to the topic; does not contain a description of any certain ML method
Language	English	Not in English

3.1. Aims of the Systems and the Outcomes

Based on the declared aims of the papers, we divided the papers into four groups and several papers were left out of that classification (Table **2**). The major part of the articles (12 of 27) was aimed at risk prediction or mortality prediction. For instance, Lezcano–Valverde *et al*. [16] reported about the development of a mortality prediction model for rheumatoid arthritis patients based on demographic and clinical-related variables collected during the first two years after disease diagnosis; Verma *et al*. [13] presented a model to identify and confirm coronary artery disease cases by using clinical data that can be easily collected at hospitals.

Table 2. Objective classification of systems.

Aim	Paper	Number
Risk or mortality prediction	[16, 18, 19, 27, 33 - 35]	12
Diagnosis of CVD	[13, 36 - 42]	8
Classification of cardiac alarm as true or false	[20, 21]	2
ECG signal and heart sound classification	[23, 43]	2
Others	[24 - 26]	3

Eight works, as an objective, considered a system for the diagnosis of cardiovascular disease development. For example, Wallert *et al*. [18] reported construction algorithms predicting two-year survival. Another example from this category is the study of Melillo *et al*. [19] where the aim was to develop predictive models for risk classification for hypertensive patients.

The next category contains two works [20, 21] aimed at classifying a cardiac alarm as true or false and reducing false alarms in intensive care units. The last group is represented by two papers [22, 23] related to the classification of ECG signals or heart sounds.

Three works were left out of the classification. Xiong *et al*. [24] proposed a system for determining the physiological manifestation of coronary stenosis from CTA images. Sengupta *et al*. [25] provided a pilot study to aid standardized assessments and support the quality of interpretations of cardiac imaging. The study of Seyednasrollah *et al*. [26] used childhood clinical factors and genetic risk factors for predicting adulthood obesity. This relates to cardiology because the authors emphasize that obesity is one of the risk factors for cardiovascular disease and early prediction of obesity is essential for CVD prevention.

All papers in our study are related to solving the problems of classification, which is a type of task, where a computer program is asked to specify, which k category a certain input belongs to. To solve this task, the learning algorithm usually produces a function f. When y=f(x), the model assigns an input described by vector x to a category identified by numerical code y [44]. Such systems take an input of a set of samples, each sample belonging to a class. The output for a new sample is a class that the sample has the highest probability of belonging to.

We grouped the papers according to the number of classes that they deal with. Most of the works (21 of 27) classify data into two classes, the others deal with 3, 5, or 6 (Table **3**). Below, we provide some examples for every classification group.

Table 3. Number of classes in observed works.

Task	Papers	Number
Classification, n=2 (binary classification)	[13, 18, 19, 20, 21, 23 - 26, 28, 29 - 34, 36, 38, 39, 41, 42]	21
Classification n=3	[27, 37, 40]	3
Classification n=5	[16, 43]	2
Classification n=6	[35]	1

Arabasadi *et al*. [36] reported about a system that predicts whether the patient has CAD or not. Wallert *et al*. presented [18] an algorithm that differentiates survivors and non-survivors in the two years after their first myocardial infarction. Ruiz-Fernández *et al*. [27] implemented a system for classifying a risk related to congenital heart disease surgery among three types: low complexity, medium complexity, and high complexity. Li *et al*. [43] provided a five-level ECG signal quality classification algorithm instead of the commonly used two-level (clean or noisy) classification. In the paper by Ambale-Venkatesh *et al*. [35], the system predicts six cardiovascular outcomes in comparison with standard cardiovascular risk scores. Table **4** shows the aim of the system and the number of classes used for classification.

3.2. ML Methods and Algorithms

The basic concept of ML is that machines use data to create a program or to learn a target function F that best maps input variable X to output variable Y. In contrast to traditional programming, in ML we do not give the computer a function or a program to get output according to it. Instead, we give the computer examples of inputs and desired outputs to create a program to get the right output for unobserved inputs (Fig. **2**).

There are three groups of Machine Learning algorithms: supervised learning, unsupervised learning, and reinforcement learning. In our work, we observed methods that had been applied in the selected papers. In some papers, ML methods were applied not only for the main task of the work such as diagnosis or predictions but also for selecting attributes or data preprocessing. The work is focused on the ML methods that were used in order to reach the main research task. In other words, here we report about classification algorithms and methods or the prediction model itself.

Table 4. System objective classification and the number of classes.

Aim/N of classes	2 classes	3 classes	5 classes	6 classes
Diagnosis of CAD	[13, 36, 38, 39, 41, 42]	[37, 40]	-	-
Risk or mortality prediction	[18, 19, 28 - 34]	[27]	[16]	[35]
Classification of cardiac alarm as true or false	[20, 21]	-	-	-
ECG signal and heart sound classification	[23]	-	[43]	-
Other	[24 - 26]	-	-	-

Fifteen out of 27 works adopted more than one method. For example [29], adopted six supervised classification ML algorithms and compared their predictive performance. We consider all the indicated methods. Table **5** provides a group of methods and works that applied these methods. Supplement 2 also provides methods and papers but in a paper-oriented way.

Table 5. ML algorithms applied in observed works.

Algorithm		Papers	N of papers	
Trees and Boosting	RF Random Forest	[16, 18, 19, 21, 23, 24, 29, 30, 34, 35, 38, 42]	12	28
	DT Decision trees	[13 - 19, 27, 28, 37, 38, 40]	8	
	Gradient boosting DT	[26, 30, 34]	3	
	AdaBoost	[19, 24, 29, 38]	4	
	LogitBoost	[32]	1	
ANN	MLP Multi-layer perceptron	[13, 19, 27, 30, 33, 36, 37]	7	12
	Other	[27, 38, 41, 42]	5	
	LogitBoost	[32]	1	
LR Logistic Regression		[13, 18, 29 - 31, 34, 39]	7	7
SVM Support Vector Machines		[18 - 20, 37, 41 - 43]	7	7
Others		[17, 27, 35, 37, 38, 41]	6	6
Naïve Bayes (NB)		[19, 24, 29, 38, 39]	5	5
k-NN k-Nearest Neighbors		[38, 41]	2	2

3.3. Decision Trees

Tree structures are used for classification. Nodes represent features from instances, and branches represent values. Leaf nodes represent decisions or classes. Based on the feature values of instances, the decision trees classify the instances [45].

There is a decision tree generating algorithms such as ID3 [46], its extension C4.5 [47], and C5/See5. Some improvements done in C4.5 are handling training data with missing attribute values, attributes with differing costs, and continuous attributes. C5 is an improvement of C4.5. C5 is more efficient in terms of speed, memory usage, size of a decision tree, and it includes boosting [48].

3.4. Boosting and Ensembles

The main idea of ensemble methods is to combine "weak" classifiers (*e.g*. SVM or decision trees) to create a final "strong" one [49]. Predictions from a single classifier can be weighted to get the final prediction (boosting), or the final prediction can be obtained as the average or major value (bagging). The difference between ensemble algorithms is the way of getting the final prediction and classifiers.

Ensembles of decision trees are very common and known. These include rotation forests with alternating decision tree as an underlying classifier [28, 29], RF, and GBDT (Table **5**). AdaBoost (Adaptive Boosting) is a popular boosting classification algorithm and it was the first algorithm that could adapt to weak learners. It was applied in Shouval's work [29] to predict mortality after myocardial infarction and was reported in Melillo's paper [19] to predict cardiovascular events. Both papers applied six different ML methods (including AdaBoost) and compared the results. Both LogitBoost and AdaBoost are based on an additive logistic regression. However, AdaBoost minimizes the exponential loss and LogitBoost minimizes the logistic loss. Motwani *et al*. [32] implemented LogitBoost to predict mortality in patients with coronary artery disease.

In contrast to the previous works, the two following works did not use DT as underlying models. Narula *et al*. [42] presented an ensemble model with three different algorithms (support vector machines, random forests, and artificial neural networks) and the final prediction was obtained by majority voting. Lo *et al*. [38] also reported about an ensemble voting mechanism where multiple classifiers were combined to obtain better prediction performance.

3.5. Random Forest or Random Decision Forests

A random forest classifier is an ensemble learning method for classification and regression that combines a collection of decision trees [50, 51]. To build a random forest classifier, a training dataset is divided into subsets. Every subset builds an independent decision tree. For every tree, different training examples are used. In other words, subsets must not overlap. To classify a new object, results from all trees are compared and the class that appears more often is considered as an answer. The important advantage of this method is that more trees will not overfit a model.

3.6. Artificial Neural Network and MLP

Artificial neural networks are the models taking their inspiration from the human brain. The core element of the model is a neuron that has connectors called synapses. Neurons are connected to each other and the model can be represented with Graph theory. A multilayer perceptron (MLP) is a simple architecture that is able to classify objects. MLP has three types of neurons that form layers: input, output, and hidden. All neurons from one layer connect with all the neurons of the next layer. Every connection has a weight that is to be adjusted during the training process. Beside MLP, plenty of other neural network architectures were applied in medical image processing, data extracting from medical records, and other areas.

3.7. Logistic Regression (LR)

Logistic Regression is a classification algorithm uses a logistic function. The output of logistic regression is a probability that the given input belongs to a certain class [52 - 54].

3.8. Support Vector Machine (SVM)

SVM is one of the most robust and accurate methods among all ML algorithms [45]. It requires a small sample set and generates patterns from that. It is insensitive to the number of dimensions. In the case of a linearly separable dataset, a classification function is a separating hyperplane f(x) that passes through the middle of the two classes to separate them [45]. When the function is determined, data instance x_n can be classified by simple testing. If $f(x_n) > 0$, then x_n belongs to the positive class.

3.9. Naive Bayes Classifier

Naive Bayes is a set of algorithms based on Bayes' theorem of the assumption of the probability of independence among predictors. One of the main advantages of these algorithms is that it is easy and fast to apply for both binary and multi-class classifications. The algorithm works especially well if predictors are actually independent, but it is almost impossible for real-life data [55 - 57].

3.10. K-NN

The k-nearest neighbors algorithm searches for the k-nearest training instances and classifies the new instance into the most frequent class of these k instances.

3.11. Datasets and Data Sources

Every system needs a dataset for training and validation. We divided all datasets used in the observed studies into two groups: publicly available datasets and datasets that could not be easily accessed. Examples of the latter are datasets that are collected in medical institutions or obtained from a register. The first group comprises seven studies and eight databases (Table **6**), but some datasets are used in more than one study and some researches use more than one dataset. Li *et al*. [43] used the PhysioNet/CinC Challenge 2011 database to develop a classifier and engaged real ECGs from the MIT-BIH arrhythmia database (MITDB) to evaluate the classification performance. Lo *et al*. [38] obtained an integrated dataset collected from 4 datasets provided by the UC Irvine Machine Learning Repository. They included the Hungarian dataset, the Switzerland dataset, the Cleveland dataset, and the Long Beach VA dataset. The new dataset contains 822 cases diagnosed either with or without CAD. The dataset is also available from the UC Irvine Machine Learning Repository as Heart Disease Data Set [58].

The second group includes 17 papers and datasets (Table **7**). For the works not indicated in Tables **6** and **7**, the sources of the data could not be clearly identified.

Table 6. Open available datasets.

Title	# of samples	# of attributes	Paper
PhysioNet/CinC Challenge 2011 database	2658	ECG	[43]
MIT-BIH arrhythmia database (MITDB)	47	ECG	[43]
Z-Alizadeh Sani dataset	303	54	[36]
Cleveland Heart Disease data set	303	76	[38]
MESA study (Multi-Ethnic Study of Atherosclerosis)	6 814	735	[59]
PhysioNet/CinC Challenge 2015 dataset.	1 250	ECG	[21, 60]
Physionet/CinC Challenge 2016 dataset	3126	heart sound	[23]
Heart Disease Data Set	822	76	[38]

3.12. UC Irvine Machine Learning Repository

The UCI Machine Learning Repository collects datasets that can be used for empirical analysis of machine learning algorithms. The website of the repository says: "it has been cited over 1000 times, making it one of the top 100 most cited "papers" in all of computer science" [58]. The repository contains a total of 399 datasets classified by domains, task type, data type, *etc*. The Life Science category includes 91 datasets related to biology and medicine. In our research, three papers [13, 36, 38] used three datasets from this repository: Z-Alizadeh Sani dataset, the Cleveland Heart Disease dataset, and the Heart Disease Data Set. We will consider them below.

3.13. Z-Alizadeh Sani Dataset

Z-Alizadeh Sani dataset contains records of 303 patients, each of them having 54 features. The features are arranged into four groups: demographic, symptom and examination, ECG, and laboratory and echo features [61]. Each patient can belong to one of the two possible categories: CAD or Normal. Patients are categorized as CAD if their coronary artery diameter narrowing is greater or equal to 50%, otherwise, a patient is categorized as Normal [61, 62]. The dataset was employed in a study [36] and provided by the UCI Machine Learning Repository; an updated version of this dataset is available there as well.

3.14. Cleveland Heart Disease Dataset

The dataset was published in 1988 and contains 76 attributes and 303 instances. The proposed task for the dataset is to predict the presence of heart disease for the patient. The target field is an integer-valued from 0 to 4 [58]. Fifty-four percent of samples represent patients without heart disease and 46% with heart disease. The dataset webpage also says that researchers usually use a subset of 14 of the 76 presented attributes. The subset includes age, sex, chest pain type, resting blood pressure, serum cholesterol in mg/dl, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression, slope of the peak exercise ST segment, number of major vessels and diagnosis of heart disease (the predictable attribute). The dataset is provided by the UCI Machine Learning Repository. The dataset was used as the main data source in one study [38] and adopted as a benchmark dataset in another [13].

Table 7. Unpublished datasets.

Name	# of samples	# of attributes	Paper
Cardiovascular Foundation of Colombia	2 432	87	[27]
Japanese health check-up data	61 313	11	[34]
Northwestern Medical Faculty Foundation (NMFF)	7 463	980	[28]
Department of Cardiology, Indira Gandhi Medical College, Shimla, India	335	26	[13]
Acute Coronary Syndrome Israeli Survey (ACSIS) registry [71]	13 422 included 2 782	54	[29]
Clinical Practice Research Datalink	383 592, included 378 256	30	[30]
Bio-signal Research Center of the Korea Research Institute of Standards and Science	214	20	[37]
SWEDEHEART, the national quality Register for Information and Knowledge about Swedish Heart Intensive Care Admissions (RIKS-HIA)	51 943	>100	[18]
Cardiology service of Hospital Clínic in Barcelona (Spain)	1 390	2100	[41]
The clinical trial of patients after non-ST-elevation acute coronary syndrome (NSTEACS)			[31]
Korean Health and Genome Epidemiology study database (KHGES)	12 789	41	[39]
COronary CT Angiography EvaluatioN For Clinical Outcomes: An InteRnational Multicenter (CONFIRM) registry	10 030	69	[32]
Cardiovascular Risk in YFS (Young Finns Study)	2 262	97	[26]
Centre of Hypertension of the University Hospital Federico II	139	33	[19]
Autonomic nervous system (ANS) unit of the cardiology department of Avicenne hospital	263	11	[40]
Hospital Clínico San Carlos RA cohort (HCSC-RAC) and Hospital Universitario de La Princesa Early Arthritis Register Longitudinal (PEARL)	1 741	12	[16]
Coronary Care Unit of Clinical Hospital Center Bezanijska Kosa, Belgrade, Serbia	1705	11	[33]

3.15. Heart Disease Data Set

The dataset is comprised of the Hungarian dataset (with 294 participants) provided by the Hungarian Institute of Cardiology, the Swiss dataset (with 123 participants) provided by Switzerland University Hospital, the dataset from Cleveland (with 303 participants) provided by the Cleveland Clinic Foundation, and the Long Beach VA dataset (with 200 participants) provided by the VA Medical Center, Long Beach, California, USA. Data from these four resources were combined into a new dataset, and incomplete entries were removed. The dataset contains 822 cases including 453 patients diagnosed with CAD and 369 cases without CAD symptoms, 642 men and 180 women aged from 28 to 77 years.

3.16. PhysioNet/Computing in Cardiology (CinC) Challenge

The PhysioNet web resource for complex physiological signals [63, 64] provides access to a collection of recorded physiological signals. PhysioNet jointly with the Computing in Cardiology conference [65] hosts a series of challenges, inviting participants to tackle clinically interesting problems. Four papers [20, 21, 23, 43] using the datasets provided as CinC challenges data sources were included in our review.

3.17. PhysioNet/CinC Challenge 2011 Database

The CinC challenge 2011 was titled “Improving the quality of ECGs collected using mobile phones”. The dataset includes standard 12-lead ECG recordings (leads I, II, II, aVR, aVL,aVF, V1, V2, V3, V4, V5, and V6) with full diagnostic bandwidth (0.05 through 100 Hz). The leads were recorded simultaneously for a minimum of 10 seconds; each lead was sampled at 500 Hz with 16-bit resolution. ECGs collected for the challenge were reviewed independently by a group of annotators who examined each ECG and assigned it a signal quality letter grade from A (excellent) to F (unacceptable). The average grade was calculated in each case, and each record was assigned to one of three groups: acceptable, indeterminate, unacceptable. The collection of 1500 twelve-lead ECGs, each being 10 seconds long, are available and split into training and test sets.

3.18. Physionet/CinC Challenge 2015 Database

The CinC Challenge 2015 was titled Reducing False Arrhythmia Alarms in the ICU. The dataset provides ECG, ABP (arterial blood pressure), PPG (photoplethysmogram) and respiratory data from intensive care with arrhythmia alarms for five life-threatening arrhythmia types: asystole, extreme bradycardia, extreme tachycardia, ventricular tachycardia, and ventricular flutter or fibrillation. The data consists of 750 records for the training set and 500 records for the unrevealed test set. Both in the training and test set, half of the records are 5 min long and the other half contains an additional 30 s after the alarm. In every record, the alarm occurs at 5 min from the beginning of the record.

3.19. PhysioNet/CinC Challenge 2016 database

The CinC Challenge 2016 was aimed at the development of heart sounds classification algorithms. The sound recordings were collected in either a clinical or non-clinical (such as in-home visits) environment, from both healthy subjects and pathological patients. The Challenge training set consists of a total of 3,126 heart sound recordings lasting from 5 seconds to over 120 seconds.

3.20. MIT-BIH Arrhythmia Database (MITDB)

MIT-BIH Arrhythmia Database [66, 67] was completed and its distribution started in 1980. This is also provided by the PhysioNet web resource. The database contains ECG recordings collected at Boston's Beth Israel Hospital. The MIT-BIH Arrhythmia Database contains 48 fragments of half-hour ECG recordings obtained from 47 subjects studied by the BIH Arrhythmia Laboratory between 1975 and 1979.

3.21. MESA (Multi-Ethnic Study of Atherosclerosis)

The Multi-Ethnic Study of Atherosclerosis (MESA) is a study of the characteristics of subclinical cardiovascular disease (disease detected non-invasively before it produces clinical signs and symptoms) and the risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease [68, 69]. The Multi-Ethnic Study of Atherosclerosis started in 2000 and included 6814 asymptomatic participants aged 45-84. The data includes Traditional Risk Factors, Demographics, Atherosclerotic markers, Magnetic Resonance Imaging (MRI) markers, Lab Biomarkers, etc. There are 735 features in total.

3.22. Unpublished Datasets / Not Publicly Accessible Datasets

Eighteen papers used datasets not available in public repositories. Table **7** provides those datasets, the number of features, and the number of attributes for each of them. The biggest dataset in terms of the number of samples is the dataset presented in the Clinical Practice Research Datalink. CPRD website [70] says that CPRD collects de-identified patient data from a network of GP practices across the UK. Primary care data in combination with other health-related data constitute a representative UK population health dataset. The data include over 35 million patients. However, the study reported a subset of 383 592 records, and 378 256 of them met the criteria of the work and were included. All the remaining datasets contain less than 62 000 samples.

The biggest amount of features comprises 2100 attributes [41]. However, 22 of 27 (81%) works used datasets with less than 100 features.

Fig. (**3**) is a plot that shows the destitution of the datasets with the number of samples as the X-axis and the number of features as the Y-axis. Most datasets (18 of 27) are located in a square limited by 14 000 samples and 100 features.

3.23. Evaluation of Algorithms

The papers reviewed employed different metrics to evaluate their results. The full list of metrics comprises 26 items. Table **8** provides the most commonly used metrics observed in the papers reviewed. A comprehensive table with all metrics is provided in Supplement 1.

Table 8. Evaluation metrics.

Metric	Papers	N
Sensitivity / Recall / TPR	[16, 18, 19, 21, 24, 26, 28, 30, 33, 34, 36 - 39, 41, 42]	18
Specificity / TNR	[16, 18, 19, 21, 24, 26, 28, 30, 32 - 34, 36 - 39, 41, 42]	17
AUC ROC	[18, 19, 24 - 26, 28 - 32, 33 - 36, 39, 42]	16
Accuracy	[13, 18, 19, 23, 24, 27, 28, 33, 36, 37, 38, 41, 43]	13
Precision / PPV	[18, 21, 24, 28, 30, 37]	6
F-score / F1 / F-measure	[21, 28, 33, 37 - 39]	6

Table 9. Number of metrics per study.

N of metrics	N of papers	Papers
1	5	[23, 27, 29, 31, 40]
2	4	[16, 17, 20, 35]
3	5	[25, 26, 32, 41, 42]
4	2	[19, 43]
5	6	[21, 30, 34, 36, 37, 39]
6	2	[24, 33]
7	2	[28, 38]
8	1	[18]

Not only the metrics, but also the number of metrics each paper used was reviewed in our study (Table **9**). Fourteen works engaged less than four different metrics and eleven papers considered five or more different evaluation metrics. Below, we give an overview of the frequent metrics we identified and how they were calculated.

The most basic approach to evaluating classification results is a confusion matrix (Table **10**). To build such a matrix, every classified example must be labeled as one of the following types: True Positive (TP) examples are classified as positive and are actually positive (right classification); True Negative (TN) examples are classified as negative and are actually negative (right classification); False Positive (FP) examples are classified as positive but are actually negative (type I error); False Negative (FN) examples are classified as negative but are actually positive (type II error). Most of the considered examples are based on the confutation matrix.

Accuracy is the most common and easy to understand evaluation metric. Accuracy was used to measure the classification performance in 13 of 27 works (66%) included in this review. It is the ratio of all correct predictions to the total amount of all predicted samples. In many cases, accuracy is not a very helpful metric.

$$\mathrm{Accuracy} = \frac{\mathrm{True}}{\mathrm{Total}} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \quad \textbf{(1)}$$

To get a more indicated and helpful evaluations for classifications, Precision and Recall algorithms are used.

Precision is the proportion of predicted positive outcomes that are really correct positive results to all positive predicted samples. Precision gives an answer to the question: of all the samples we classified as true, how many are actually true? This metric was applied in six considered papers.

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\text{predicted positives}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad \textbf{(2)}$$

The Recallis the ratio of obtained relevant instances (true positive outcomes) and the total amount of relevant samples. It shows how many of all actual positive examples were classified correctly. The higher the Recall, the fewer the positive examples missed in the classification.

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\text{actual positives}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad \textbf{(3)}$$

Both metrics are connected to each other, a higher level of Precision may be obtained by decreasing recall and vice versa. Since that, separating using neither Precision nor recall is a good evaluator of a classification algorithm. In order to combine both metrics into one, the F-Score is used.

Table 10. Confusion matrix.

–		Actual	
		Positive	**Negative**
Predictive	Positive	TP	FP (Type I error)
	Negative	FN (Type II error)	TN

F-Score (F-measure or F_1 score) is the weighted harmonic mean of precision and recall. This metric demonstrates how many cases the model predicts correctly, and how many true instances the model does not miss. F-score appears in six papers considered.

$$\mathrm{FScore} = 2 * \frac{\mathrm{precision} * \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} \quad \textbf{(4)}$$

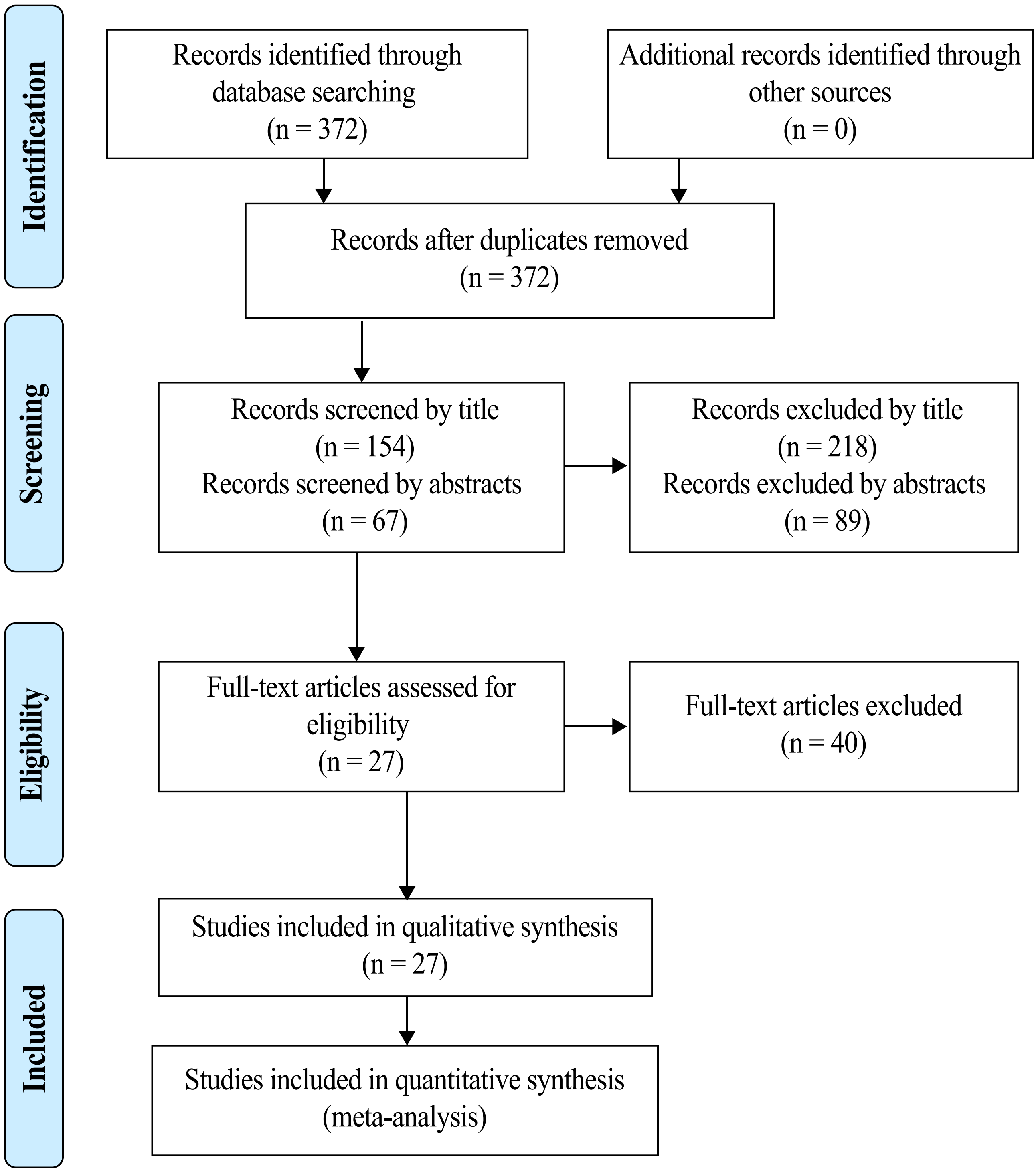

Fig. (1). PRISMA flow diagram.

Another informative and very common metric is Area Under the Receiver Operating Characteristic Curve (AUC ROC) [72]. The curve is a graph showing the performance of the classification model and it plots two parameters: True Positive Rate (TPR) and False Positive Rate (FPR). It summarizes the trade-off between TPR and FPR using all different classification thresholds. The big advantage of that metric is that AUC evaluates models independently from the threshold. AUC ROC is a common metric among the considered papers and it was used in 16 works.

$$\text{TPR} = \text{Sensitivity} = \text{Recall} = \frac{\text{TP}}{\text{actual positives}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \textbf{(5)}$$

As a matter of fact, TPR completely equals to recall and sensitivity and shows the proportion of positive examples that are correctly classified to all actual positive examples. In its turn, FPR is the proportion of actual negative examples that are mistakenly classified as positive (FP), to all actual negative ones. The higher the FPR, the more negative examples are classified wrong.

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\text{actual negative}} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \quad \textbf{(6)}$$

$$Specificity = TNR = 1 - FPR = \frac{TN}{actual\ negative} = \frac{TN}{TN + FP} \quad \textbf{(7)}$$

In some sources, another metric is called Specificity or True negative rate (TNR). TNR measures the proportion of actual negative examples that are correctly classified as negative. This is the opposite metric to the FPR and is often used coupled with sensitivity or TPR. 17 papers cited this metric.

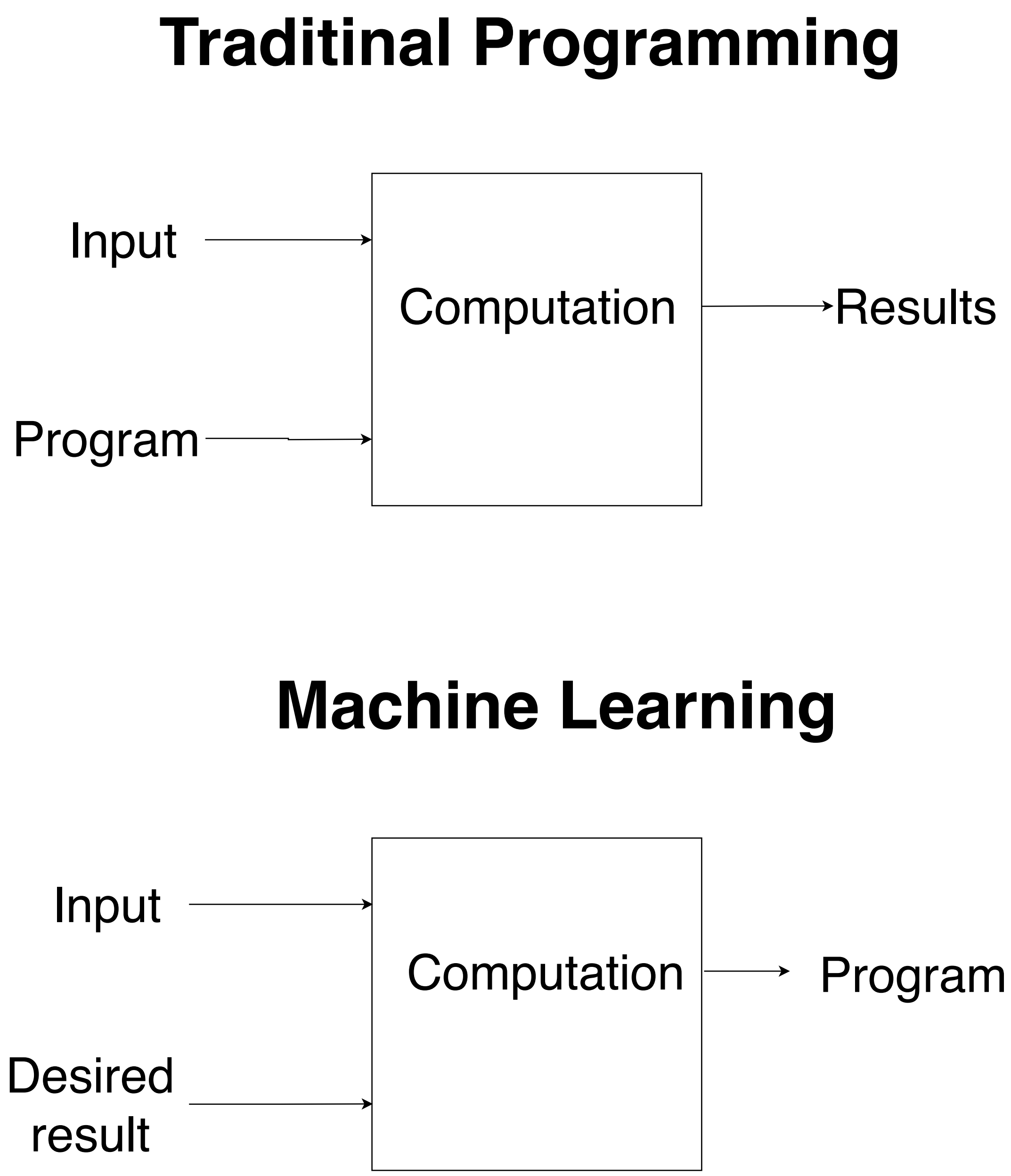

Fig. (2). Machine Learning concept.

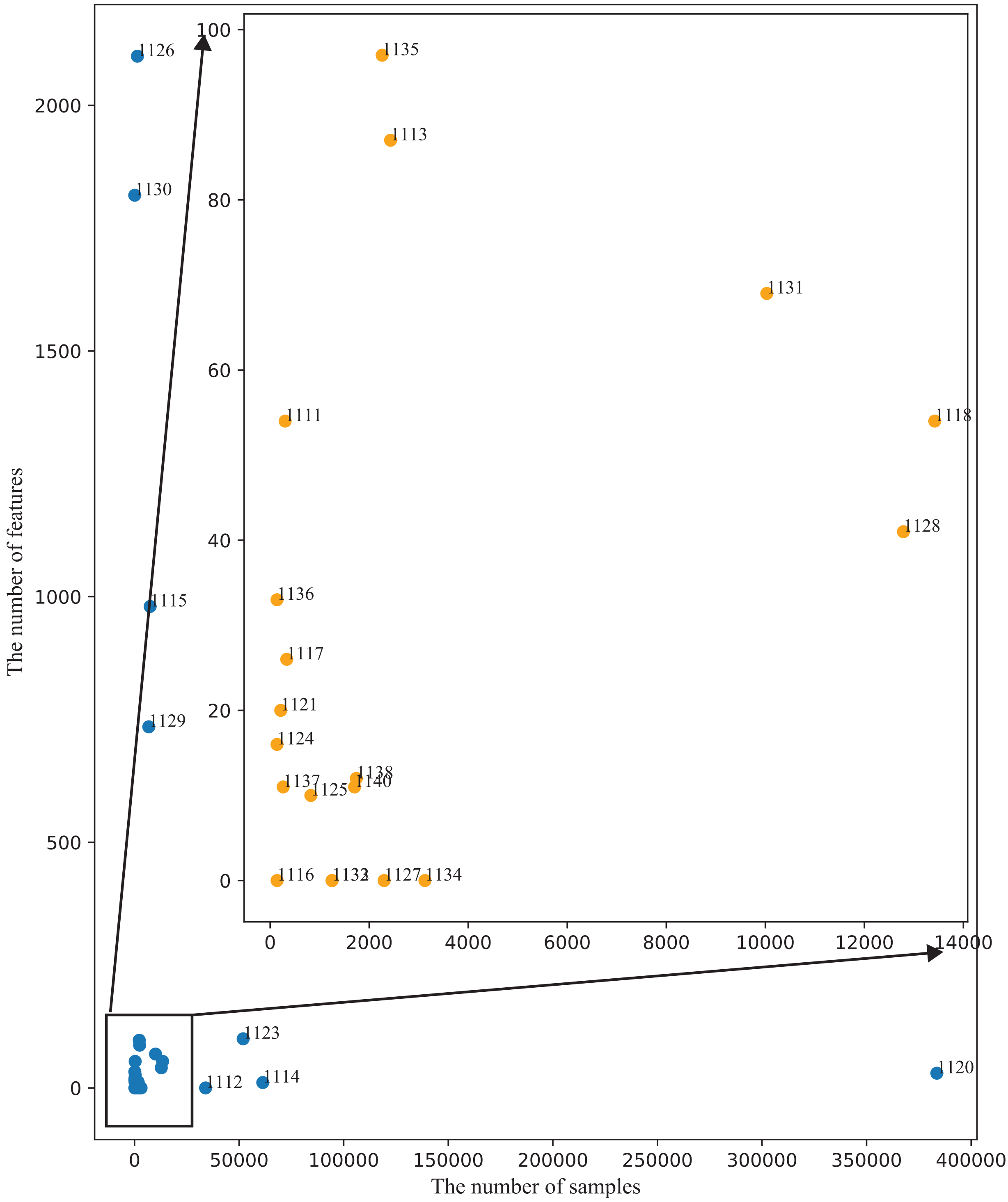

Fig. (3). The number of samples and features in the datasets.

4. DISCUSSION

We systematically reviewed 27 papers describing machine learning algorithms applied for the decision support in cardiology. The diversity and maturity machine learning methods presented in these papers allow making conclusions on what the state of the art is and what future directions of the ML in cardiology can be. In this section, we discuss findings of the review, including aims and outcomes, ML methods, datasets, and evaluation metrics.

4.1. Aims and Outcomes

All observed papers deal with classification. Nonetheless, ML methods can handle regression and clusterization tasks as well, but we did not reveal such works. Many of the studies deal with binary classification only and predict mortality or diagnose CAD in a binary way. We assume that there is room for improvement here, and aims for ML-based systems can be given in a more sophisticated manner.

4.2. Methods

Our study revealed more than ten different ML methods implemented in the observed studies. This is a good variety, but all methods are related to classification, not regression. Moreover, all methods belong only to supervised learning, and none of the reports are about unsupervised learning or clustering.

None of the papers observed in our review reported the implementation of the RL approach. Extra search for RL implementations in cardiology on PubMed produced no paper that could be included. Kipp *et al*. [73] also report the fact that "Application of reinforcement learning to health care and cardiology thus far has been scarce".

We associate this with the complexity of RL algorithms implementation and the lack of suitable data. The data should be not only sufficient, but it needs to be presented in the appropriate form. To get such data representation is not a trivial task. Nonetheless, the application of RL in medicine and cardiology, in particular, is very promising and needs a closer look.

Some of the observed papers implement several methods independently and compare results, which we consider a good practice since there is no way to be sure which method has a better performance in every particular task. Several works implemented ensembles. Tree-based ensembles are the most common methods in our research, but all the methods are well-known and were used in many different tasks. However, there is a lack of more sophisticated and innovative methods such as, for example, XGBoost [74], an ensemble method that has empirically proven to be a highly effective approach by gaining the best results in numerous machine learning competitions [75]. Deep neural networks are also poorly represented.

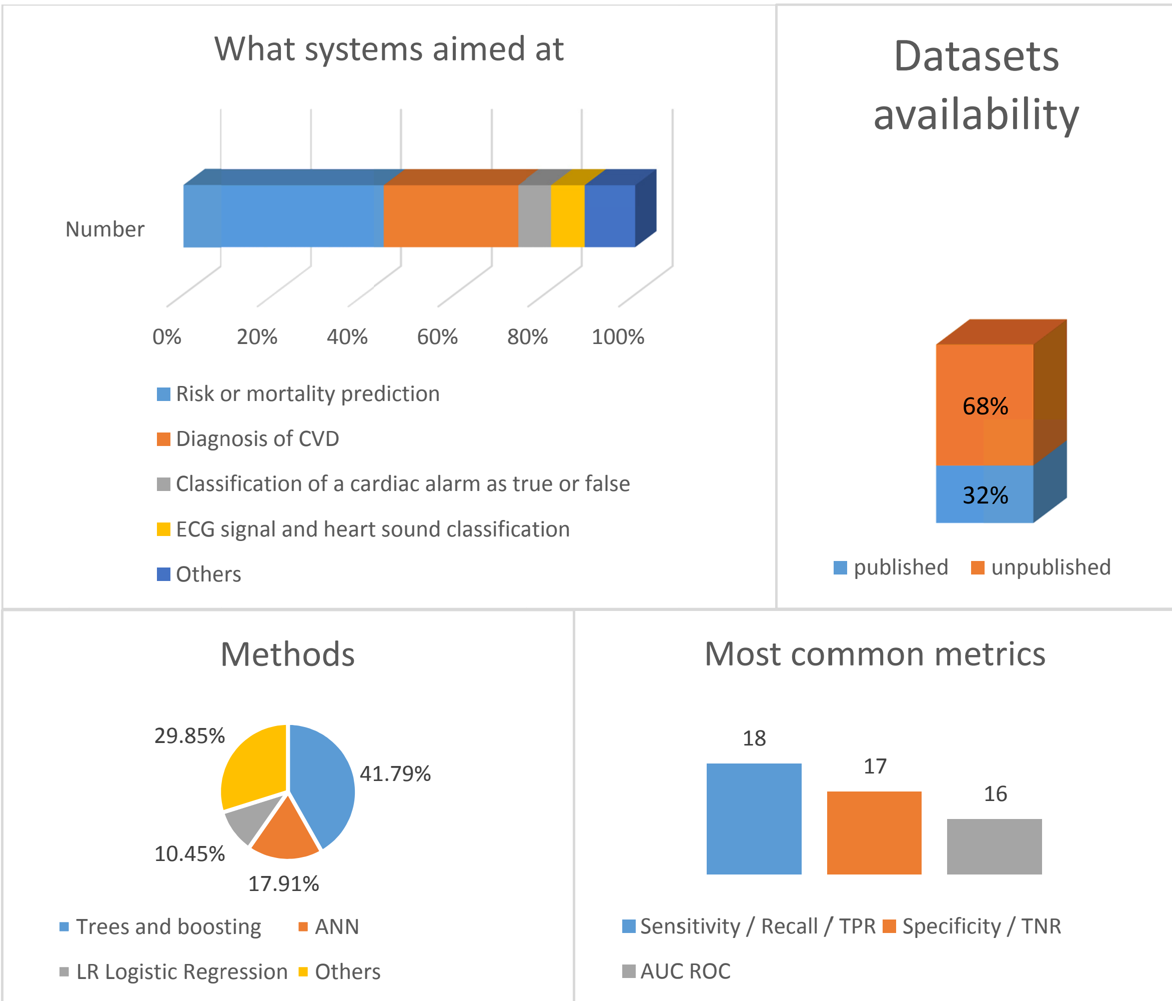

Fig. (4). Aggregated quantitative results of the study.

A broad variety of different network architectures was not covered by the papers. State-of-the-art algorithms remained outside the reviewed works. The implementation of advanced ML algorithms in the field of cardiology is still an under-developed niche.

4.3. Datasets

Repositories containing the datasets report about hundreds of studies referring to them. Nevertheless, only 7 of 27 works (26%) in our review employed publicly available datasets. That might be explained as a commitment to solving a particular real-life task with particular real data collected in the environment that the solution is supposed to work in later. At the same time, the use of data that is hardly available for other researchers leads to the lack of reproducibility and comparability of the results.

When developing a system, it is good to have data of the same structure and from the same source as the data that will be used in the system. In contrast, for the works given in order to evaluate applications of a certain method to the specific task as well as for works aimed to show improvement in existing algorithms or presenting new algorithms, it is meaningful to employ well-known and accessible datasets. This gives the opportunity to compare results and see real advantages or disadvantages of the proposed algorithm. We also suggest that publication of de-identified data along with the research papers will have a positive impact on future developments.

4.4. Evaluation

We identified 28 different metrics for a classification task. This variety of metrics makes it difficult to compare and understand research results. The metrics show incomparable aspects of algorithm efficiency. This way, selecting a metric is an important task that should be done at the beginning of developing an ML system. It is necessary to decide which is more important: to classify some healthy patients as ill and start treatment they do not need to misclassify healthy, but omit some ill patients and do not give them the treatment they need. The decision depends on many factors and should be made for every particular task.

The second issue here is the difference in the names of metrics that are calculated in the same way but come from different domains. Thus, recall and sensitivity are the same function, but the first term is mostly used in information retrieval and the second one is more typical for statistics and medical tests. Moreover, this function sometimes might be called as TRR (true positive rate) or 1 – FNR (false-negative rate). That is true not only for recall/sensitivity but also for some other metrics. It makes interpreting and comparing results more complicated.

The third issue of algorithm evaluation is a decrease in model performance due to the difference between training and real data. Li *et al*. [43] evaluated an algorithm on real unseen data and discovered a lower performance score. This is a common problem for machine learning algorithm evaluation. First of all, it means that we cannot directly compare the efficiency of two algorithms or systems if they were evaluated on different datasets.

4.5. Future Directions

Based on the obtained results of our review, we suppose that, in the nearest future, systems employing deep neural networks will actively spread in the domain along with a growing variety of tasks and applications of ML systems in cardiology. We also expect works reporting on another very promising branch of ML: Reinforcement Learning [76].

In addition, we assume that creating a pool of openly accessible datasets related to cardiological issues will open new opportunities for researchers and result in a big positive impact on the field.

CONCLUSION

In this systematic review, we identified studies where machine learning algorithms were applied in the domain of cardiology and examined four aspects: aims, ML methods, datasets, and evaluation metrics. We showed that a broad variety of methods are applied in cardiology, but all methods belong to a group of supervised learning classification methods; more often than not, researchers use unpublished, hardly available data; different studies aimed at similar tasks and engaging similar methods could not easily compare to each other (Fig. **4**).

We hope that this systematic review will be a helpful tool for researchers who are developing machine learning based systems in medicine and particularly in cardiology.

LIMITATIONS

The paper covers studies published between 2013-2017. Only one literature database, PUBMED, was used in the research. This paper evaluated the included papers on the basis of only the four questions mentioned, however, additional aspects, such as data preprocessing, feature selection, dimensionality reduction, content, and structure of input data should also be analyzed in future works. Electrocardiogram (ECG) signal processing was not considered in this review, however, there can also be some valuable results in the field of machine learning. We could not give a quantitative estimate for the algorithms due to the heterogeneity of the metrics used in different studies.

LIST OF ABBREVIATIONS

ABP	=	Arterial Blood Pressure
AI	=	Artificial Intelligence
ANN	=	Artificial Neural Networks
ANS	=	Autonomic Nervous System
AUC ROC	=	Area Under the Receiver Operating Characteristic Curve
CDSS	=	Clinical Decision Support Systems
CVD	=	Cardiovascular Diseases
ECG	=	Electrocardiogram
FN	=	False Negative
FNR	=	False Negative Rate
FP	=	False Positive

FPR	=	False Positive Rate
GBDT	=	Gradient Boosted Decision Trees
K-NN	=	K Nearest Neighbor
LR	=	Logistic Regression
ML	=	Machine Learning
MLP	=	Multi-layer Perceptron
MRI	=	Magnetic Resonance Imaging
NB	=	Naïve Bayes
NLP	=	Natural Language Processing
PPG	=	Photoplethysmogram
RF	=	Random Forest
SVM	=	Support Vector Machines
TN	=	True Negative
TNR	=	True Negative Rate
TP	=	True Positive
TPR	=	True Positive Rate

AUTHORS' CONTRIBUTIONS

AD was responsible for data collection, analysis and manuscript drafting. MG was responsible for the manuscript writing. GP was responsible for the methodological part of the study.

CONSENT FOR PUBLICATION

Not applicable.

STANDARD OF REPORTING

PRISMA Guideline and methodology were followed.

FUNDING

This work was financially supported by the Government of the Russian Federation through the ITMO fellowship and professorship program. This work was supported by a Russian Fund for Basic research 18-37-20002.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publishers web site along with the published article.

REFERENCES

[1] World Health Organization Cardiovascular diseases (CVDs) [Internet]. WHO. World Health Organization. 2016.

[2] Samuel AL. Some Studies in Machine Learning Using the Game of Checkers. IBM J Res Develop 1959; 3(3): 210-29. [http://dx.doi.org/10.1147/rd.33.0210]

[3] Krafczyk S, Tietze S, Swoboda W, Valkovič P, Brandt T. Artificial neural network: A new diagnostic posturographic tool for disorders of stance. Clin Neurophysiol [Internet] Elsevier 2006; 117(8): 1692-8. [http://dx.doi.org/10.1016/j.clinph.2006.04.022]

[4] Mishra BK, Singh SK, Bhala S. Breast cancer diagnosis using back-propagation algorithm. Proc Int Conf Work Emerg Trends Technol - ICWET '11 [Internet] New York. 470. [http://dx.doi.org/10.1145/1980022.1980123]

[5] Dietzel M, Baltzer PAT, Dietzel A, *et al.* Artificial Neural Networks for differential diagnosis of breast lesions in MR-Mammography: a systematic approach addressing the influence of network architecture on diagnostic performance using a large clinical database. Eur J Radiol 2012; 81(7): 1508-13. [http://dx.doi.org/10.1016/j.ejrad.2011.03.024] [PMID: 21459533]

[6] Liu M, Dong X. The application of improved BP neural network in the diagnosis of breast tumors. Int Conf Syst Informatics IEEE. 2012; 2012; pp. 1239-42. [http://dx.doi.org/10.1109/ICSAI.2012.6223260]

[7] Chan KY, Ling SH, Dillon TS, Nguyen HT. Diagnosis of hypoglycemic episodes using a neural network based rule discovery system. Expert Syst Appl Pergamon 2011; 38(8): 9799-808. [http://dx.doi.org/10.1016/j.eswa.2011.02.020]

[8] Esteva A, Kuprel B, Novoa RA, *et al.* Dermatologist-level classification of skin cancer with deep neural networks. Nature Nature Publishing Group 2017; 542(7639): 115-8. [http://dx.doi.org/10.1038/nature21056]

[9] Haenssle HA, Fink C, Schneiderbauer R, *et al.* Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol 2018; 29(8): 1836-42. [http://dx.doi.org/10.1093/annonc/mdy166] [PMID: 29846502]

[10] Litjens G, Kooi T, Bejnordi BE, *et al.* A survey on deep learning in medical image analysis. Med Image Anal 2017; 42: 60-88. [http://dx.doi.org/10.1016/j.media.2017.07.005] [PMID: 28778026]

[11] Palaniappan R, Sundaraj K, Ahamed N, Arjunan A, Sundaraj S. Computer-based Respiratory Sound Analysis: A Systematic Review. IETE Tech Rev 2013; 30(3): 248. [http://dx.doi.org/10.4103/0256-4602.113524]

[12] Triantafyllidis AK, Tsanas A. Applications of machine learning in real-life digital health interventions: Review of the literature. J Med Internet Res 2019; 21(4)e12286 [http://dx.doi.org/10.2196/12286] [PMID: 30950797]

[13] Verma L, Srivastava S, Negi PCC. A Hybrid Data Mining Model to Predict Coronary Artery Disease Cases Using Non-Invasive Clinical Data. J Med Syst 2016; 40(7): 178. [http://dx.doi.org/10.1007/s10916-016-0536-z] [PMID: 27286983]

[14] Moher D, Shamseer L, Clarke M, *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015; 4(1): 1. [http://dx.doi.org/10.1186/2046-4053-4-1] [PMID: 25554246]

[15] Moher D, Liberati A, Tetzlaff J, Altman DG, Group TP. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009; 6(7)e1000097 [http://dx.doi.org/10.1371/journal.pmed.1000097] [PMID: 19621072]

[16] Lezcano-Valverde JM, Salazar F, León L, *et al.* Development and validation of a multivariate predictive model for rheumatoid arthritis mortality using a machine learning approach. Sci Rep 2017; 7(1): 10189. [http://dx.doi.org/10.1038/s41598-017-10558-w] [PMID: 28860558]

[17] The AI Industry Series. Top Healthcare AI Trends To Watch

[18] Wallert J, Tomasoni M, Madison G, Held C. Predicting two-year survival versus non-survival after first myocardial infarction using machine learning and Swedish national register data. BMC Med Inform Decis Mak 2017; 17(1): 99. [http://dx.doi.org/10.1186/s12911-017-0500-y] [PMID: 28679442]

[19] Melillo P, Izzo R, Orrico A, *et al.* Automatic prediction of cardiovascular and cerebrovascular events using heart rate variability analysis. PLoS One 2015; 10(3)e0118504 [http://dx.doi.org/10.1371/journal.pone.0118504] [PMID: 25793605]

[20] Kalidas V, Tamil LS. Cardiac arrhythmia classification using multi-modal signal analysis. Physiol Meas 2016; 37(8): 1253-72. [http://dx.doi.org/10.1088/0967-3334/37/8/1253] [PMID: 27454417]

[21] Eerikäinen LM, Vanschoren J, Rooijakkers MJ, Vullings R, Aarts RM. Reduction of false arrhythmia alarms using signal selection and machine learning. Physiol Meas 2016; 37(8): 1204-16. [http://dx.doi.org/10.1088/0967-3334/37/8/1204] [PMID: 27454128]

[22] Li Q, Rajagopalan C, Clifford GDGD. A machine learning approach to multi-level ECG signal quality classification. Comput Methods Programs Biomed 2014; 117(3): 435-47. [http://dx.doi.org/10.1016/j.cmpb.2014.09.002] [PMID: 25306242]

[23] Kay E, Agarwal A. DropConnected neural networks trained on time-frequency and inter-beat features for classifying heart sounds. Physiol Meas 2017; 38(8): 1645-57. [http://dx.doi.org/10.1088/1361-6579/aa6a3d] [PMID: 28758641]

[24] Xiong G, Kola D, Heo R, Elmore K, Cho I, Min JK. Myocardial perfusion analysis in cardiac computed tomography angiographic images at rest. Med Image Anal NIH Public Access 2015; 24(1): 77-89. [http://dx.doi.org/10.1016/j.media.2015.05.010]

[25] Sengupta PP, Huang Y-M, Bansal M, *et al.* Cognitive Machine-Learning Algorithm for Cardiac Imaging: A Pilot Study for Differentiating Constrictive Pericarditis From Restrictive Cardiomyopathy. Circ Cardiovasc Imaging 2016; 9(6)e004330 [http://dx.doi.org/10.1161/CIRCIMAGING.115.004330] [PMID: 27266599]

[26] Seyednasrollah F, Mäkelä J, Pitkänen N, *et al.* Prediction of Adulthood Obesity Using Genetic and Childhood Clinical Risk Factors in the Cardiovascular Risk in Young Finns Study. Circ Cardiovasc Genet 2017; 10(3)e001554 [http://dx.doi.org/10.1161/CIRCGENETICS.116.001554] [PMID: 28620069]

[27] Ruiz-Fernández D, Monsalve Torra A, Soriano-Payá A, Marín-Alonso O, Triana Palencia E. Aid decision algorithms to estimate the risk in congenital heart surgery. Comput Methods Programs Biomed 2016; 126: 118-27. [http://dx.doi.org/10.1016/j.cmpb.2015.12.021] [PMID: 26774238]

[28] Mathias JS, Agrawal A, Feinglass J, Cooper AJ, Baker DW, Choudhary A. Development of a 5 year life expectancy index in older adults using predictive mining of electronic health record data. J Am Med Inform Assoc 2013; 20(e1): e118-24. [http://dx.doi.org/10.1136/amiajnl-2012-001360] [PMID: 23538722]

[29] Shouval R, Hadanny A, Shlomo N, *et al.* Machine learning for prediction of 30-day mortality after ST elevation myocardial infraction: An Acute Coronary Syndrome Israeli Survey data mining study. Int J Cardiol 2017; 246: 7-13. [http://dx.doi.org/10.1016/j.ijcard.2017.05.067] [PMID: 28867023]

[30] Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS One 2017; 12(4)e0174944 [http://dx.doi.org/10.1371/journal.pone.0174944] [PMID: 28376093]

[31] Liu Y, Scirica BM, Stultz CM, Guttag JV. Beatquency domain and machine learning improve prediction of cardiovascular death after acute coronary syndrome. Sci Rep 2016; 6(1): 34540. [http://dx.doi.org/10.1038/srep34540] [PMID: 27708350]

[32] Motwani M, Dey D, Berman DS, *et al.* Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. Eur Heart J 2017; 38(7): 500-7. [PMID: 27252451]

[33] Gligorijević T, Ševarac Z, Milovanović B, *et al.* Follow-up and risk assessment in patients with myocardial infarction using artificial neural networks. Complexity Hindawi 2017; pp. 1-8.

[34] Ichikawa D, Saito T, Ujita W, Oyama H. How can machine-learning methods assist in virtual screening for hyperuricemia? A healthcare machine-learning approach. J Biomed Inform 2016; 64: 20-4. [http://dx.doi.org/10.1016/j.jbi.2016.09.012] [PMID: 27658886]

[35] Ambale-Venkatesh B, Yang X, Wu CO, *et al.* Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis. Circ Res 2017; 121(9): 1092-101. [http://dx.doi.org/10.1161/CIRCRESAHA.117.311312] [PMID: 28794054]

[36] Arabasadi Z, Alizadehsani R, Roshanzamir M, Moosaei H, Yarifard AA. Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. Comput Methods Programs Biomed 2017; 141: 19-26. [http://dx.doi.org/10.1016/j.cmpb.2017.01.004] [PMID: 28241964]

[37] Kim H, Ishag MIM, Piao M, Kwon T, Ryu KH. A data mining approach for cardiovascular disease diagnosis using heart rate variability and images of carotid arteries. Symmetry (Basel) 2016; 8(6) [http://dx.doi.org/10.3390/sym8060047]

[38] Lo Y-T, Fujita H, Pai T-W, *et al.* Prediction of coronary artery disease based on ensemble learning approaches and co-expressed observations. J Mech Med Biol Nature Publishing Group 2016; 16(1): 178. [http://dx.doi.org/10.1142/S0219519416400108]

[39] Lee BJ, Kim JY. A comparison of the predictive power of anthropometric indices for hypertension and hypotension risk.PLoS One Public Library of Science. Zhu Z. 2014; 9: p. (1)e84897.. [PMID: PMID:24465449] [http://dx.doi.org/10.1371/journal.pone.0084897]

[40] Idri A, Kadi I. Evaluating a decision making system for cardiovascular dysautonomias diagnosis. Springerplus 2016; 5(1): 81. [http://dx.doi.org/10.1186/s40064-016-1730-7] [PMID: 26844028]

[41] Ribas Ripoll VJ, Wojdel A, Romero E, Ramos P, Brugada J. ECG assessment based on neural networks with pretraining. Appl Soft Comput J Elsevier 2016; 49: 399-406. [http://dx.doi.org/10.1016/j.asoc.2016.08.013]

[42] Narula S, Shameer K, Salem Omar AM, Dudley JT, Sengupta PP. Machine-Learning Algorithms to Automate Morphological and Functional Assessments in 2D Echocardiography. J Am Coll Cardiol 2016; 68(21): 2287-95. [http://dx.doi.org/10.1016/j.jacc.2016.08.062] [PMID: 27884247]

[43] Li Q, Rajagopalan C, Clifford GD. A machine learning approach to multi-level ECG signal quality classification. Comput Methods Programs Biomed 2014; 117(3): 435-47. [http://dx.doi.org/10.1016/j.cmpb.2014.09.002] [PMID: 25306242]

[44] Goodfellow I, Bengio Y, Courville A. Deep learning.

[45] Wu X, Kumar V, Ross Quinlan J, *et al.* Top 10 algorithms in data mining. Knowl Inf Syst Springer-Verlag 2008; 14(1): 1-37. [http://dx.doi.org/10.1007/s10115-007-0114-2]

[46] Quinlan JR. Induction of Decision Trees 1986; 1: 81-106. http://hunch.net/~coms-4771/quinlan.pdf [http://dx.doi.org/10.1007/BF00116251]

[47] Quinlan JR, John R, Ross J. programs for machine learning Morgan Kaufmann Publishers 1993. Available from: https://dl.acm.org/citation.cfm?id=152181ISBN:1558602380

[48] Kuhn M. Central Iowa R Users Group “Predictive Modeling” 2013. Available from: http://link.springer.com/content/pdf/10.1007/978-1-4614-6849-3.pdf

[49] Zhou Z-H. (Computer scientist) Ensemble methods : foundations and algorithms. CRC Press 2012.http://cds.cern.ch/record/1487876ISBN:9781439830055 [http://dx.doi.org/10.1201/b12207]

[50] Breiman L. Random Forests. Random Forests. Mach Learn Kluwer Academic Publishers 2001; 45: pp. (1)5-32. [http://dx.doi.org/10.1023/A:1010933404324]

[51] Hastie T, Tibshirani R, Friedman JH. Jerome H The elements of statistical learning : data mining, inference, and prediction. 2009.

[52] Park H-A. An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain J Korean Acad Nurs 2013; 43(2): 154. [http://dx.doi.org/10.4040/jkan.2013.43.2.154]

[53] Sperandei S. Understanding logistic regression analysis. Biochem medica Croatian Society for Medical Biochemistry and Laboratory Medicine 2014; 24(1): 12-8. [PMID: 24627710] [http://dx.doi.org/10.11613/BM.2014.003]

[54] Peng C, Lee KL, Ingersoll GM. An Introduction to Logistic Regression Analysis and Reporting 2003. Available from: https://www.semanticscholar.org/paper/An-Introduction-to-Logistic-Regression-Analysis-and-Peng-Lee/889c94e7440b1d2ad2cc7ff4be2b72c1dafe6347

[55] Murphy KP. Naive Bayes classifiers Available from: https://datajobsboard.com/wp-content/uploads/2017/01/Naive-Bayes-Kevin-Murphy.pdf

[56] Rish I. An empirical study of the naive bayes classifier 2001.

[57] Lewis DD. Naive (Bayes) at forty: The independence assumption in information retrieval. Berlin, Heidelberg: Springer 1998; pp. 4-15.

[58] Machine Learning Repository UCI. Learning Repository Available from: http://archive.ics.uci.edu/ml/

[59] Ambale-Venkatesh B, Yang X, *et al.* Cardiovascular Event Prediction by Machine LearningNovelty and Significance. Circ Res 2017; 121(9): 1092-101. [http://dx.doi.org/10.1161/CIRCRESAHA.117.311312] [PMID: 28794054]

[60] Russell SJ. Stuart J, Norvig P, Davis E. Artificial intelligence : A modern approach . 3rd ed. Upper Saddle River NJ Prentice Hall 2010. Available from: http://www.worldcat.org/title/artificial-intelligence-a-modern-approach/oclc/359890490ISBN:9780136042594

[61] Alizadehsani R, Zangooei MH, Hosseini MJ, *et al.* Coronary artery disease detection using computational intelligence methods. Knowledge-Based Syst Elsevier 2016; 109: 187-97. [http://dx.doi.org/10.1016/j.knosys.2016.07.004]

[62] Alizadehsani R, Habibi J, Hosseini MJMJMJ, *et al.* A data mining

approach for diagnosis of coronary artery disease. Comput Methods Programs Biomed 2013; 111(1): 52-61.
[http://dx.doi.org/10.1016/j.cmpb.2013.03.004] [PMID: 23537611]
[63] PhysioNet. Available from: https://physionet.org/
[64] Costa M, Moody GB, Henry I, Goldberger AL. PhysioNet: an NIH research resource for complex signals. J Electrocardiol 2003; 36(Suppl.): 139-44.
[http://dx.doi.org/10.1016/j.jelectrocard.2003.09.038] [PMID: 14716615]
[65] Cin C. Available from: http://www.cinc.org/
[66] Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. IEEE Eng Med Biol Mag 2001; 20(3): 45-50.
[http://dx.doi.org/10.1109/51.932724] [PMID: 11446209]
[67] MIT-BIH Arrhythmia Database Directory. Available from: https://www.physionet.org/physiobank/database/html/mitdbdir/mitdbdir.htm
[68] Bild DE, Bluemke DA, Burke GL, *et al.* Multi-Ethnic Study of Atherosclerosis: Objectives and design. Am J Epidemiol 2002; 156(9): 871-81.
[http://dx.doi.org/10.1093/aje/kwf113] [PMID: 12397006]
[69] MESA - Multi-Ethnic Study of Atherosclerosis. Available from: https://www.mesa-nhlbi.org/
[70] Clinical Practice Research Datalink | CPRD . Available from: https://cprd.com/home
[71] Kornowski R. The ACSIS Registry and primary angioplasty following coronary bypass surgery. Catheter Cardiovasc Interv 2011; 78(4): 537-9.
[http://dx.doi.org/10.1002/ccd.23345] [PMID: 21953750]
[72] Fawcett T. An introduction to ROC analysis 2005.
[http://dx.doi.org/10.1016/j.patrec.2005.10.010]
[73] Johnson KW, Torres Soto J, Glicksberg BS, *et al.* Artificial Intelligence in Cardiology. J Am Coll Cardiol 2018; 71(23): 2668-79.
[http://dx.doi.org/10.1016/j.jacc.2018.03.521] [PMID: 29880128]
[74] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min - KDD '16 New York. New York, USA:. ACM Press 2016; pp. 785-94.
[http://dx.doi.org/10.1145/2939672.2939785]
[75] Nielsen D. Tree Boosting With XG Boost Why Does XGBoost Win "Every" Machine Learning Competition? . Available from: https://brage.bibsys.no/xmlui/bitstream/handle/11250/2433761/16128_FULLTEXT.pdf
[76] Sutton R, Barto A. Reinforcement Learning: An Introduction 2018. Available from: https://books.google.de/books?hl=en&lr=&id=sWV0DwAAQBAJ&oi=fnd&pg=PR7&ots=1Zbar1hnYj&sig=RVefEU0hZ8L3WN5nmH8qsat8FMA