Computer Science Review 39 (2021) 100334

Contents lists available at ScienceDirect

Computer Science Review

journal homepage: www.elsevier.com/locate/cosrev

Review article

Prediction of Atrial Fibrillation using artificial intelligence on Electrocardiograms: A systematic review

Igor Matias [a,b,*], Nuno Garcia [a,b], Sandeep Pirbhulal [a,c], Virginie Felizardo [a,b], Nuno Pombo [a,b], Henriques Zacarias [a,b], Miguel Sousa [a], Eftim Zdravevski [d,e]

[a] Universidade da Beira Interior, Covilhã, Portugal
[b] Instituto de Telecomunicações, Covilhã, Portugal
[c] Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Norway
[d] Faculty of Computer Science and Engineering, Saints Cyril and Methodius University in Skopje, Skopje, North Macedonia
[e] COPELABS, Universidade Lusófona de Humanidades e Tecnologias, Lisboa, Portugal

A R T I C L E I N F O

Article history:
Received 16 July 2020
Received in revised form 14 November 2020
Accepted 27 November 2020
Available online xxxx

Keywords:
ECG waveform
Electrocardiogram
Artificial Intelligence
Prediction algorithms
Atrial Fibrillation

A B S T R A C T

Atrial Fibrillation (AF) is a type of arrhythmia characterized by irregular heartbeats, with four types, two of which are complicated to diagnose using standard techniques such as Electrocardiogram (ECG). However, and because smart wearables are increasingly a piece of commodity equipment, there are several ways of detecting and predicting AF episodes using only an ECG exam, allowing physicians easier diagnosis. By searching several databases, this study presents a review of the articles published in the last ten years, focusing on those who reported studies using Artificial Intelligence (AI) for prediction of AF. The results show that only twelve studies were selected for this systematic review, where three of them applied deep learning techniques (25%), six of them used machine learning methods (50%) and three others focused on applying general artificial intelligence models (25%). To conclude, this study revealed that the prediction of AF is yet an under-developed field in the context of AI, and deep learning techniques are increasing the accuracy, but these are not as frequently applied as it would be expected. Also, more than half of the selected studies were published since 2016, corroborating that this topic is very recent and has a high potential for additional research.

Contents

* Correspondence to: Universidade da Beira Interior, Departamento de Informática, Rua Marquês d'Ávila e Bolama, 6201-001, Covilhã, Portugal.
E-mail address: igor.matias@ubi.pt (I. Matias).

https://doi.org/10.1016/j.cosrev.2020.100334
1574-0137/

1. Introduction

Atrial Fibrillation (AF) is a type of arrhythmia, which is characterized by irregular heartbeats, that can lead to blood clots, heart failure, stroke, and other heart-related complications including death, and is commonly underdiagnosed [1,2]. It can assume four different types, Paroxysmal AF, Persistent AF, Long-standing Persistent AF, and Permanent AF. According to [3], only the severest Long-standing Persistent and Permanent can be easily detected with an ECG exam, the other types being harder to identify due to the irregularity of its symptoms.

This irregularity of symptoms makes it very hard to detect both less severe types of AF, due to the high probability of these patients not showing any symptoms during an ECG. To avoid this low efficiency on the detection of AF, predictive models were developed, allowing the diagnosis of a patient's AF state only based in a short ECG signal, avoiding extra-long and intrusive devices and methodologies.

The detection of AF is commonly performed by analysing the signal collected from an ECG, a non-invasive and painless exam with quick results, typically outputting several charts resulting from a 12 lead collection setup [4].

Nowadays, portable devices such as smartwatches, smart fitness bands, or portable medical signal collectors have a crucial role in evolving the way we diagnose several health disorders before they step into a high-risk medical field. Due to its ease on the recording, for example, ECG and pulse signals [5–8], by being always with the patient itself, they are able to collect data from several moments of the day, within different activity and emotional states. Some of these devices, despite using a smaller number of ECG leads, sometimes 3 or 2, have been proved to be as efficient as Hospital grade ECG equipment, as tested in [9].

However, in the last years, there were developed several new methods to detect and to predict the existence of the different types of AF. These new approaches all require powerful algorithms combined with innovative sensors, applying several different types of AI.

There is a multitude of benefits from integrating AI into healthcare, including automation tasks and analysing big patient's datasets to deliver better healthcare faster, and at a lower cost [10]. The usage of AI into healthcare, and consequently AF detection and prediction, does allow the analysis of bigger datasets, with the faster result, easing the workload of healthcare professionals, possibilitating automated and real-time diagnosis, anytime and anywhere.

With AI applications in such area, it is possible to diagnose AF conditions ahead of time, as well as to predict the AF episodes onset, allowing to better prepare, or even revert, an AF episode, this way preventing many possible severe health conditions.

This paper presents a systematic literature review on ECG-based models for AF Prediction using AI techniques covering the last ten years. At the time of this review, we did not find any report that covers this topic. Therefore, the selected studies reviewed here present the most recent work in applying computational methods in the analysis and evaluation of biomedical signals for the prediction of AF, according to the carried search as described in the next sections.

The main contributions of this article are:
(1) to present a discussion on how the prediction of AF have been and is currently addressed;
(2) to indicate what databases, features, pre-processing and predictive algorithms have been and are presently used in these systems;
(3) a benchmark to conclude which model from the studied articles performs better.

The remainder of this paper is organized as follows: Section 2 presents a description of the method that was designed for eligibility selection and extraction of information. Section 3 includes the results of the search by displaying the selected studies and their features in summary tables. The discussion and the answer to the research questions are presented in Section 4. Finally, Section 5 of this review contains the highlights and limitations of this study.

2. Methods

This systematic literature review was conducted informed by recommendations from the Cochrane Handbook for Systematic Reviews of Interventions, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [11–13], and based on the guidelines from [14].

This section explains in detail the methodology used for conducting this review.

2.1. Search strategy

The databases Web of Science,[1] Scopus,[2] ACM Digital Library,[3] IEEE Xplore,[4] PubMed,[5] and Science Direct[6] were used to search for relevant peer-reviewed publications from January 1, 2009, 00:00 to December 13, 2019, 04:22, Lisbon time.

The first two used databases are interdisciplinary databases. ACM Digital Library is, according to [15], the number one database related to academic databases for computer science and IEEE Xplore was chosen due to its high number of articles from the field of computer science. Finally, PubMed was used due to its content regarding research in biomedicine and Science Direct because of its high number of articles from thousands of books and journals.

We searched titles and abstracts using the keywords presented below. The list of references from the selected articles was manually screened for the inclusion of additional relevant articles.

The keywords used in all the databases were:

("machine learning" OR "artificial intelligence") AND ("ECG" OR electrocardio*) AND ("Atrial Fibrillation" OR "AF" OR "arrhythmia") AND ("prediction" OR "prognosis" OR "foresee").

[1] S. C. Collection, "Web of Science [v.5.14] - Web of Science Core Collection 引用レポート," pp. 8–9.
[2] "Scopus - Document search | Signed in." [Online]. Available: https://www.scopus.com/search/form.uri?display=basic. [Accessed: 15-Jan-2020].
[3] C. The et al. "ACM Digital Library," 1985.
[4] "IEEE Xplore Digital Library." [Online]. Available: https://ieeexplore.ieee.org/Xplore/home.jsp. [Accessed: 15-Jan-2020].
[5] "Home - PubMed - NCBI." [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/. [Accessed: 15-Jan-2020].
[6] "ScienceDirect.com | Science, health and medical journals, full-text articles and books." [Online]. Available: https://www.sciencedirect.com/. [Accessed: 15-Jan-2020].

Table 1
Inclusion and exclusion criteria used in the review, as in [17].

Type	Inclusion	Exclusion
Date	All	None
Exposure of interest	All	None
Geographic location of study	All	None
Language	English	Any other language
Participants	With no recent surgical procedures or drugs effects during the ECG collection	With any recent surgical procedure or ingestion or drugs effects during the ECG collection
Peer review	Journal and Conference	All others
Reported outcomes	At least one: accuracy, sensitivity, specificity, confusion matrix	All others that did not report any metric
Setting	All	None
Study design	All	None
Type of publication	Journal and Conference	All others

2.2. Study selection

We screened the titles and abstracts of all identified publications for eligibility, using the web application Rayyan QCRI [16].

The inclusion criteria were broadly defined to increase the sensitivity of the search. The aim was to identify the articles that applied any AI method on ECG signals for prediction of AF in patients with no previous clinical conditionings. Studies that applied techniques to detect the presence of an actual AF episode in a patient were not considered, as this review's main goal is focused on the prediction of AF rather than detection, that is, the prediction of AF onset before it actually occurs, instead of detecting the start of the episode.

Additional inclusion/exclusion criteria are summarized in Table 1.

According to the inclusion and exclusion criteria presented in Table 1, all the articles not excluded after its analysis had full texts reviewed for eligibility.

2.3. Extraction of study characteristics

The extraction of information from the selected publications was based on the pre-defined categories, to collect the relevant data and to assess, analyse the model characteristics and its experimental setup:

- Study Information: defines the study citation and year of publication;
- Inputs: assess the inputs used to develop the algorithm, including dataset used, amount and age of the individuals from where the dataset was collected;
- Signal treatment: defines the usage of the ECG signals received as input, namely the features extracted from it, the duration of the signal used for training, and the tools used for the process;
- Methods: defines the methods/algorithms applied to the pre-processing of the ECG signal, the prediction of AF and evaluation of the model, as well as the number of iterations, and the data separation into training and testing;
- Performances: defines the evaluation metrics used to assess the predictions.

2.4. Research questions

The research questions of this review were:

- (RQ1) How is the prediction problem assessed?
- (RQ2) What databases and features are used?
- (RQ3) What pre-processing algorithms are used?
- (RQ4) What predictive algorithms are used?
- (RQ5) Which model does provide the best performance?

The (RQ1) motivation was to identify the trends and possible opportunities for research topic focus.

The motivation for (RQ2) and (RQ3) was to identify new advances on features and databases and pre-processing techniques used for prediction of AF, respectively.

The motivation for (RQ4) was to identify the new predictive algorithms used to predict AF using ECG data on recent studies.

Finally, for (RQ5) motivation, it was intended to identify the models that can more accurately predict AF episodes, this way identifying trends and possible opportunities for the use of research methods.

3. Results

At the beginning of the search, it yielded 375 unique records, after the removal of duplicates.

After the review of the title and abstract and following the inclusion and exclusion criteria presented in Table 1, 293 records were excluded; 82 full-text publications were assessed for eligibility and after full-text review, of which 72 records were excluded.

The excluded records can be described as follows. Sixty-four studies reported research related to AF, but there was no prediction of AF during its execution. Two studies could not be fully read because the authors of this systematic literature review were not able to obtain the full articles. Two articles did not present the evaluation metrics included in the Inclusion Criteria of this search presented in Table 1. Two studies were focused on reviewing state-of-the-art related to AF identification. One study had a publication date before 2009, and another one did the work with ECG collected from patients with surgical proceedings (prophylactic ICD-implantation).

From the remaining 10 records, reference tracking was performed, and two studies were added, totalizing 12 studies to be included for the data extraction and the qualitative synthesis stage. The flow diagram of the identification and inclusion of articles is shown in Fig. 1.

3.1. Eligibility of the studies

Despite all the selected studies that met the inclusion and exclusion criteria, it is useful to clarify the selection of some studies.

The study [18] presents an algorithm for short term prediction of Persistent AF, but the ECG data used was collected from sheep instead of human individuals. Despite this, we considered that this article is eligible, not so much because of the nature of the ECG signal, but mostly because of the described methodologies and algorithmic approaches the paper describes.

In the studies [19,20] and [21] the prediction of AF was only performed between pre and post AF moments, not allowing for cases with no AF prediction. However, they were included because of the insight the papers report to this research.

Finally, in the study [21] the reported measurements with the single fold method matched neither the tables nor the text of the paper. It was decided to include this last study, but only to consider the best measurements for the 10-fold method, that has valid reporting of values in the tables and the study's text.

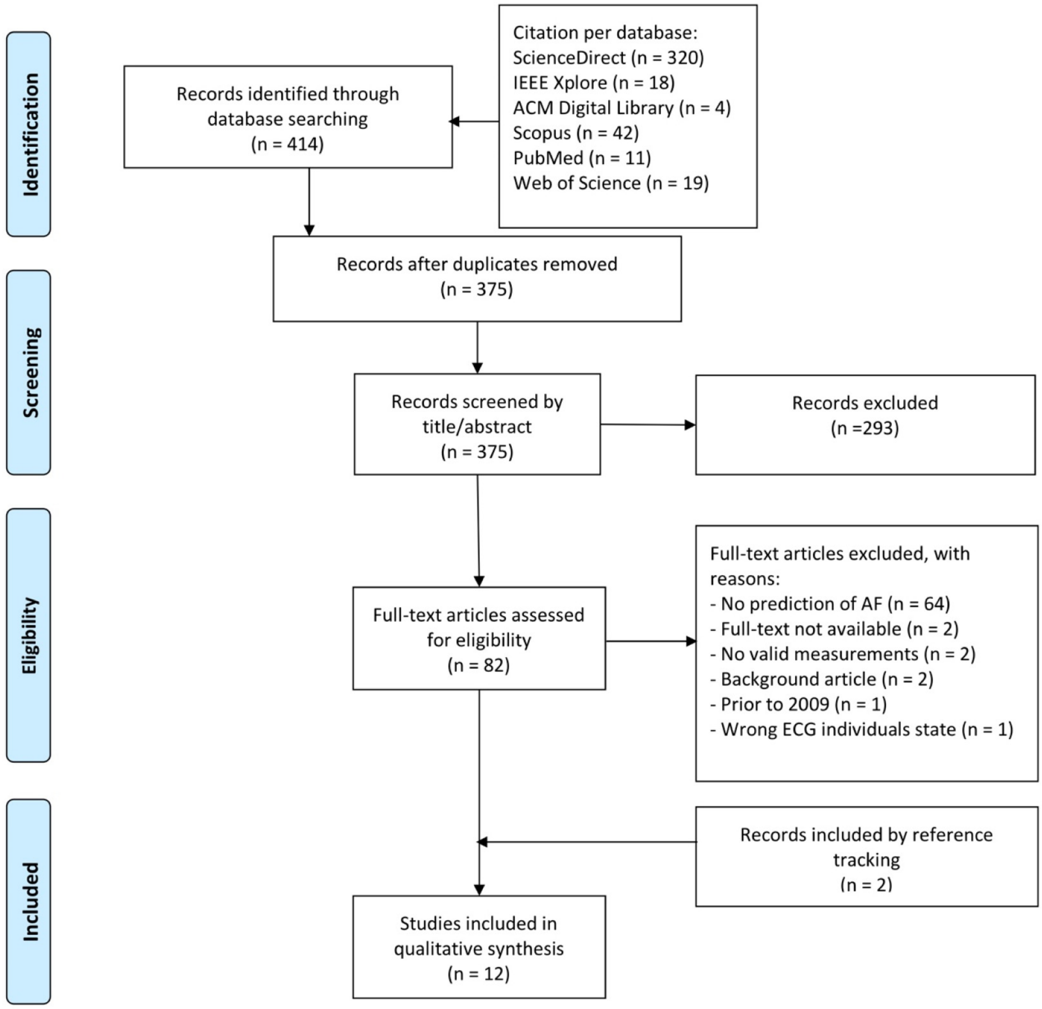


Fig. 1. Flow diagram of identification and inclusion papers.

Table 2
Number of publications by journal or conference type.

	Description	Number of studies	Portion of total
Journal or Conference type	Medicine Journal	1	8%
	Bioinformatics Journal	4	33%
	Computer Science Journal	1	8%
	IEEE Conference or sponsored by IEEE	3	25%

3.2. Source of evidence

From the resulting twelve studies to be included, half of them were published after the start of the year of 2018, as Fig. 2 shows. Table 2 presents the classification of the selected studies by its type of publication, and by its publication place's main area of focus.

3.3. Study participants and design

Seven studies (58.33%) were based on databases with small samples of individuals (less than 100), two studies (16.67%) with samples between 100 and 25 000 individuals, one study (8.33%) with a sample of around 126 000 individuals, and two studies (16.67%) did not report the sample size.

Three studies (25.00%) used a personal database of ECG records, one (8.33%) used a UCI Repository Warehouse's [28] database, one was based on the Mayo Clinic [32] ECG Laboratory's database, one used the Medical Information Mart for Intensive Care III database [34], five articles (41.67%) have done the research using the Atrial Fibrillation Prediction Database [22], and one study (8.33%) used a China Kadoorie Biobank's [26] database. Both [34] and [22] datasets are available at the Physionet Repository. Only [22,34] and [26] are publicly available.

3.4. Prediction methods

The selected articles used several different methods of AI for prediction of AF:

- Five articles [18,20,21,25,30] applied Support Vector Machine;
- Two articles [23,33] used statistical AI methods;
- Two articles [24,31] used Convolutional Neural Network;
- One study [27] applied its study using Arrhythmia Fuzzy Hybrid Classifier;
- Another study [29] used Markov Chain;
- Finally, the last of all articles [19] used the method Mixture of Experts for the prediction of AF.

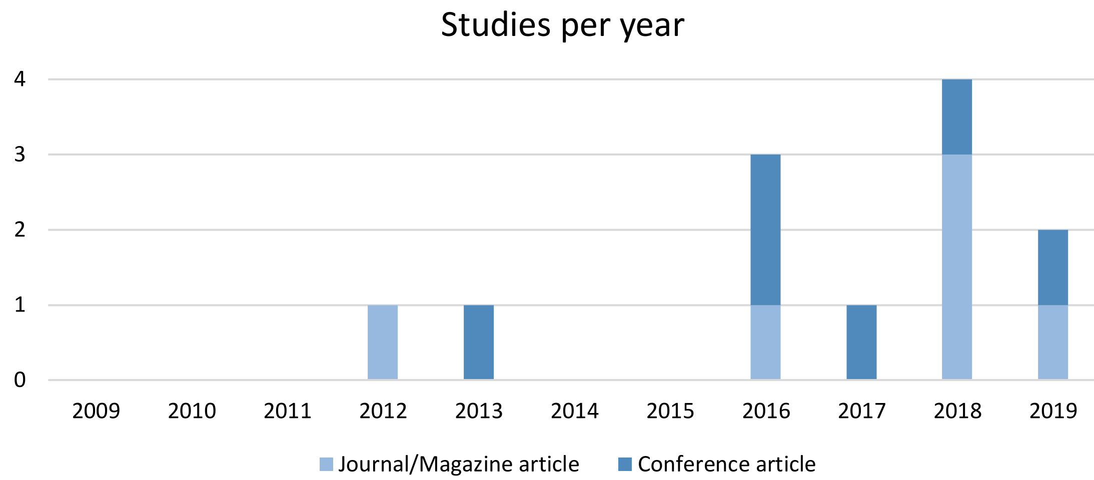


Fig. 2. Number of studies from 2009 to 2019.

Table 3
Input information collected from studies. NR = Not Reported/Not Applicable.

Year	Study	Dataset used	Number of participants	Age of participants
2012	Mohebbi et al. [20]	Atrial Fibrillation Prediction Database [22]	NR	NR
2013	Costin et al. [23]	Atrial Fibrillation Prediction Database [22]	75	NR
2016	Kim et al. [24]	Own collected dataset	1	NR
2016	Shen et al. [25]	China Kadoorie Biobank [26]	24369	NR
2016	Boon et al. [21]	Atrial Fibrillation Prediction Database [22]	53	NR
2017	ElMoaqet et al. [18]	Own collected dataset	33	NR
2018	Rajalakshmi et al. [27]	UCI Repository Warehouse [28]	NR	NR
2018	Li et al. [29]	Own collected dataset	5	NR
2018	Boon et al. [30]	Atrial Fibrillation Prediction Database [22]	53	NR
2018	Ebrahimzadeh et al. [19]	Atrial Fibrillation Prediction Database [22]	53	NR
2019	Attia et al. [31]	Mayo Clinic ECG Laboratory [32]	126526	>18, average 60.3
2019	Mohamed et al. [33]	Medical Information Mart for Intensive Care III database [34]	246	NR

3.5. Data collected from selected studies

During the quality synthesis process, it is essential to get as much information from the selected studies as possible. However, despite all the articles having some extra data, some of it was not comparable, the reason why they are not mentioned in the next collected data tables.

Table 3 shows the dataset used in each one of the selected studies, including the number of individuals, where the data came from, and its age, if provided.

From all the twelve selected studies, not all of them apply AI methods that do not need feature selection and extraction from the source ECG signal. Table 4 presents the number of frequency-domain, time-domain, space-domain, and non-linear features extracted from each one of the studies, as well as the signal duration used as input to the AI model/method and the tools used at the collecting and pre-processing phase of the studies.

Regarding the signal duration used in each one of the selected studies, it is a noticeable difference between the minimum and maximum among all. The majority used a signal of 300 s (three studies), followed by 30 and 10 s (two studies each). Some other articles reported usage of signals with 120, 180, 0 and 3600 s length (one study each).

Table 5 has information about the different methods used in each one of the selected studies. It is divided into the methods used for pre-processing the input data for the prediction phase and the performance evaluation. The table also includes the number of iterations used on the training as well as the data split between training and testing subsets.

The identified models/algorithms in all selected studies were compared with the reported accuracy. Some of them did not report the sensitivity and specificity, neither the F-Score nor the Area Under the Curve. Table 6 contains information about the achievements of each study.

As Table 4 indicated, almost all of the selected studies performed feature extraction. This extraction was not performed by those that only implemented a deep learning method. Going deeper into the analysis and comparison of the selected studies, Table 7 presents all the features selected and extracted by each one of them. Despite not all the studies used the same dataset, the results of each one can prove that its methods can predict AF, which is the goal they all have in common.

Table 4
Signal treatment information collected from studies. NR = Not Reported/Not Applicable.

Year	Study	Features extracted from ECG signal	Signal duration (seconds)	Tools used
2012	Mohebbi et al. [20]	4 frequency-domain 6 time-domain 4 non-linear	300	NR
2013	Costin et al. [23]	1 frequency-domain 1 time-domain	300	Pan-Tompkins algorithm [35], MATLAB 2008 [36]
2016	Kim et al. [24]	NR	30	Caffe deep learning framework [37]
2016	Shen et al. [25]	1 time-domain 1 space-domain	10	NR
2016	Boon et al. [21]	8 frequency-domain 1 time-domain	1800	NR
2017	ElMoaqet et al. [18]	1 frequency-domain 5 time-domain 3 non-linear	30	MATLAB [36], LibSVM toolbox [38]
2018	Rajalakshmi et al. [27]	5 time-domain	NR	Excel[a], MATLAB 2015 [36], Rapid Miner[b]
2018	Li et al. [29]	NR	120	NR
2018	Boon et al. [30]	3 frequency-domain 2 time-domain 2 non-linear	900	C++ [39], LibSVM library [38]
2018	Ebrahimzadeh et al. [19]	4 frequency-domain 5 time-domain 8 non-linear 11 time-frequency	300	NR
2019	Attia et al. [31]	NR	10	GE-Marquette ECG machine[c], MUSE system[d], Keras[e], TensorFlow [40], Python[f], R[g]
2019	Mohamed et al. [33]	5 time-domain	3600	NR

[a]Microsoft Portugal, "Microsoft Excel," 2019. [Online]. Available: https://products.office.com/pt-pt/excel?legRedir=true&CorrelationId=3e4e9d3a-7d82-42a5-977c-fa3f430fa6ce&rtc=1. [Accessed: 29-Jan-2020].
[b]RapidMiner, "Lightning Fast Data Science Platform for Teams | RapidMiner©," RapidMiner, 2019. [Online]. Available: https://rapidminer.com/. [Accessed: 29-Jan-2020].
[c]"MAC 2000 - Resting ECGs - Diagnostic Cardiology - Categories | GE Healthcare." [Online]. Available: https://www.gehealthcare.com/products/mac-2000. [Accessed: 29-Jan-2020].
[d]"MUSE v9 | GE Healthcare." [Online]. Available: https://www.gehealthcare.com/products/diagnostic-ecg/cardio-data-management/muse-v9. [Accessed: 29-Jan-2020].
[e]"Home - Keras Documentation." [Online]. Available: https://keras.io/. [Accessed: 29-Jan-2020].
[f]Python Software Foundation, "Welcome to Python.org," 2001, 2019. [Online]. Available: https://www.python.org/. [Accessed: 29-Jan-2020].
[g]The R Foundation, "R: The R Project for Statistical Computing," 2018. [Online]. Available: https://www.r-project.org/. [Accessed: 29-Jan-2020].

Table 8 shows the horizon of the prediction made by every one of the selected studies, this is, in how much time can the resultant models predict AF episodes.

4. Discussion

This systematic literature review aims to identify, assess and analyse the recent state-of-the-art of ECG-based models for AF Prediction using Artificial Intelligence techniques. The following paragraphs discuss the previously defined research questions.

4.1. How is the prediction problem addressed? (RQ1)

From the selected articles, most of them only address the problem of predicting AF, that is, their main focus is to predict AF and no other types of arrhythmia or heart pathologies.

All the selected articles performed classification prediction, that is, all classified the prediction with discrete labels.

From all the twelve selected studies, only one performed a risk-based approach on the prediction problem, which means the majority did a time series prediction of AF. Regarding the number of classes used for the prediction process, only two articles reported a study using a multi-class approach, all the remaining used binary (between "pre-AF" and "not pre-AF" events).

Despite not all the studies reported the event horizon of the prediction, two of them used a 30 min horizon, and the remaining used 14 days, 60 min, 5 min, 2 min and under a 0-min horizon (immediately before the AF event).

Eight out of twelve of the selected studies performed prediction of AF with input signals shorter or equal to 300 s (five minutes long), being the most used length of signal by the studies.

When looking at the datasets used by the selected articles, the three most accurate models are from three of the five studies that used the dataset [22], thus identifying this as a good option for further work on assessing the problem.

4.2. What databases and features are used? (RQ2)

Despite some of the selected studies do not perform ECG signal features extraction, when performing it, the selected features directly impact the model's capability of predicting AF existence with higher accuracy.

Table 7 indicates the different features selected by the articles considered in this systematic literature review.

Table 5
Methods applied by the selected studies. NR = Not Reported/Not Applicable.

Year	Study	Pre-processing method(s)	Prediction method(s)	Evaluation method(s)	Number of iterations	Data split (training/testing) %
2012	Mohebbi et al. [20]	Noise removal, QRS detection	Support Vector Machine	NR	NR	47/53
2013	Costin et al. [23]	Noise removal	HRV analysis and Morphologic Variability of QRS complexes	NR	NR	50/50
2016	Kim et al. [24]	NR	Convolutional Neural Network with ON/OFF ReLU	NR	30 000	90/10
2016	Shen et al. [25]	NR	Support Vector Machine	5-fold Cross-Validation	NR	NR
2016	Boon et al. [21]	Hamilton and Tompkins algorithm, McNames algorithm	Support Vector Machine	10-fold Cross-Validation	10	90.6/9.4
2017	ElMoaqet et al. [18]	Noise removal	Weighted Support Vector Machine	10-fold Cross-Validation	100	75/25
2018	Rajalakshmi et al. [27]	Normalisation, Missing values removal	Novel Arrhythmia Fuzzy Hybrid Classifier Algorithm	NR	NR	NR
2018	Li et al. [29]	Noise removal, QRS detection	Markov Chain	NR	NR	NR
2018	Boon et al. [30]	McNames algorithm	Support Vector Machine	10-fold Cross-Validation	5	90.6/9.4
2018	Ebrahimzadeh et al. [19]	Noise removal, QRS detection	Mixture of Experts	10-fold Cross-Validation	NR	47/53
2019	Attia et al. [31]	NR	Convolutional Neural Network	NR	NR	70/20
2019	Mohamed et al. [33]	NR	Belief Functions Theory	NR	30	67/33

Table 6
Evaluation of the selected studies. NR = Not Reported/Not Applicable.

Year	Study	Accuracy	Sensitivity	Specificity	F-Score	Area Under Curve
2012	Mohebbi et al. [20]	96.64%	96.30%	93.10%	NR	NR
2013	Costin et al. [23]	90.00%	89.44%	89.29%	NR	89.40%
2016	Kim et al. [24]	83.58%	NR	NR	NR	NR
2016	Shen et al. [25]	75.60%	NR	NR	NR	83.00%
2016	Boon et al. [21]	80.20%	81.10%	79.30%	NR	NR
2017	ElMoaqet et al. [18]	84.90%	66.70%	97.00%	NR	93.50%
2018	Rajalakshmi et al. [27]	82.80%	0.40%	0.43%	1.21%	NR
2018	Li et al. [29]	82.00%	86.00%	80.00%	74.51%	90.88%
2018	Boon et al. [30]	87.70%	86.80%	88.70%	NR	NR
2018	Ebrahimzadeh et al. [19]	98.21%	100.00%	96.55%	NR	NR
2019	Attia et al. [31]	83.30%	82.30%	83.40%	45.40%	90.00%
2019	Mohamed et al. [33]	70.49%	77.07%	63.90%	NR	NR

The most used features are Standard Deviation of RR Intervals, Low-frequency band power, Mean of RR Intervals and Standard Deviation, being used by, at least, 3 different selected articles.

Most of the approaches are based solely on ECG signals, but one study combined ECG signal's data with heart morphology data. Almost half of the selected articles used the Atrial Fibrillation Prediction Database [22], a quarter of them used an own collected dataset of ECG signals and others used a UCI Repository Warehouse [28] dataset, the Mayo Clinic [32] ECG Laboratory's database, the Medical Information Mart for Intensive Care III database [34], and a China Kadoorie Biobank's [26] database.

According to the article [27], the UCI Repository Warehouse dataset consists of 452 instances with 279 attributes, where the ECG reports are in image format.

The Mayo Clinic ECG Laboratory's database used by [31], included "all patients aged 18 years or older with at least one digital, normal sinus rhythm, standard 10-second, 12-lead ECG acquired in the supine position" between 1993 and 2017. The signals were acquired at a sampling rate of 500 Hz using a GE-Marquette ECG Machine[7] and stored using the MUSE data management system.[8] All the records were "over-read by a physician-supervised, trained technician, with corrections made to the diagnostic labels as needed".

As used by [33], the Medical Information Mart for Intensive Care III database was collected from 2001 to 2012. It contains information from over 40 thousand patients, about Heart Rate, Arterial Blood Pressure, and Respiration. This database also contains "charts at a higher frequency like ECG and continuous blood pressure from Intensive Care Units patients". For the study, only the patients who have developed AF during their recordings are considered.

[7] "MAC 2000 - Resting ECGs - Diagnostic Cardiology - Categories | GE Healthcare." [Online]. Available: https://www.gehealthcare.com/products/mac-2000. [Accessed: 29-Jan-2020].

[8] "MUSE v9 | GE Healthcare." [Online]. Available: https://www.gehealthcare.com/products/diagnostic-ecg/cardio-data-management/muse-v9. [Accessed: 29-Jan-2020].

Table 7
Features extracted from input on each one of the selected articles.

Domain	Features	Studies
Frequency	Low-frequency band power (LF)	[19,20,30]
	High-frequency band power (HF)	[19,20]
	LF/HF ratio	[19,23]
	Low-frequency component of Fast Fourier Transforms (FFT-LF)	[21]
	High-frequency component of Fast Fourier Transforms (FFT-HF)	[21]
	LL-H1	[21,30]
	LL-H2	[21]
	HH-H3	[21]
	ROI-H1	[21]
	ROI-H2	[21]
	ROI-H3	[21]
	QRS segment duration	[27]
	P-R waves interval	[27]
	Q-T waves interval	[27]
	T wave interval	[27]
	P wave interval	[27]
	Weighted centre of the bispectrum (ROI-WCOB)	[30]
	Very Low-Frequency band power (VLF)	[19]
Time	Standard Deviation of Average of all NN interval for all 5-min periods of the entire recording (SDANN)	[23]
	ST level	[25]
	Standard Deviation of RR intervals (SDRR)	[18,19,21,33]
	Mean of RR intervals	[18,19,33]
	Skewness of RR intervals	[18,33]
	Kurtosis of RR intervals	[18,33]
	Number of adjacent RR intervals differing by more than 50 ms (NN50)	[30]
	Sum of NN50 divided by the total number of all RR intervals (PNN50)	[19,30]
	Square root of the mean of the squares of differences between adjacent RR intervals (RMSSD)	[19]
	Standard deviation of differences between adjacent RR intervals (SDSD)	[19]
	Smoothed Pseudo Winger Ville distribution (SPWVD)	[19]
Space	Amplitude of P wave	[25]
	Amplitude of Q wave	[25]
	Amplitude of R wave	[25]
	Amplitude of S wave	[25]
	Amplitude of T wave	[25]
Nonlinear	Standard Deviation 1 (SD1)	[19,20]
	Standard Deviation 2 (SD2)	[19,20,30]
	SD1/SD2 ratio	[19,20]
	Sample Entropy	[20,30]
	Approximate Entropy	[18]

The studies [19–21,23,30] used the Atrial Fibrillation Prediction Database, which "consists of excerpts of two-channel long-term ECG (Holter) recordings and is divided into a learning set and a test set of equal size. The database includes the digitized ECG signals (sampled at 128 Hz per signal, with 12-bit resolution) and a set of unaudited, automatically-generated QRS annotations", as in [22]. The records were collected from 48 individuals, although the selected articles always refer to 53 or 75 participants, as in Table 3.

Table 8
Prediction horizon on each one of the selected articles. NR = Not Reported.

Year	Study	Prediction horizon
2012	Mohebbi et al. [20]	NR
2013	Costin et al. [23]	30 min
2016	Kim et al. [24]	NR
2016	Shen et al. [25]	NR
2016	Boon et al. [21]	30 min
2017	ElMoaqet et al. [18]	14 days
2018	Rajalakshmi et al. [27]	NR
2018	Li et al. [29]	2 min
2018	Boon et al. [30]	NR
2018	Ebrahimzadeh et al. [19]	5 min
2019	Attia et al. [31]	NR
2019	Mohamed et al. [33]	60 min

Last, the study [25] was based on a database from the China Kadoorie Biobank, which is a cohort study of over 520 000 adults from 10 different areas from China, collected from 2004 to 2008 using questionnaires and anthropometric and physiological measurements as well as blood samples of every participant. For the study, the 12-lead ECG data of 10 s duration at 500 Hz were used, which were collected from 24 369 participants using a Mortara ELIx50 device during 2013 and 2014, as well as the blood pressure data (systolic and diastolic).

4.3. What pre-processing algorithms are used? (RQ3)

The pre-processing methods used in all the twelve selected studies are presented in Table 5.

Although not all the articles indicate the pre-processing methods applied, due to some of them were elaborated applying prediction methods that do not need any pre-processing of the signal, the most used pre-processing technique is Noise Removal (5 studies), followed by QRS Detection (4 studies) and Correction of Signal (2 studies). Both Normalisation and Missing Value Removal methods were applied by one study each.

4.4. What predictive algorithms are used? (RQ4)

The most used prediction method/algorithm is Support Vector Machine [18,20,21,25,30], followed by Convolutional Neural Network [24,31].

Some other selected studies applied either statistical AI methods (HRV analysis and Morphologic Variability of QRS complexes, Belief Functions Theory), or Arrhythmia Fuzzy Hybrid Classifier, Markov Chain, or, at last, Mixture of Experts.

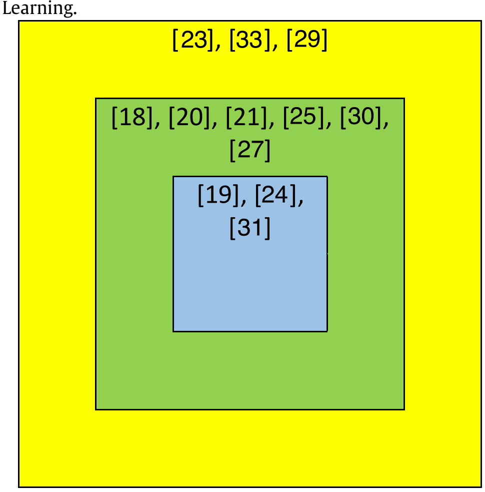
Dividing the predictive algorithms into three classes, this is, Deep Learning, Machine Learning, and Artificial Intelligence, we can identify the type of prediction approach executed by each one of the selected studies, as presented in Table 9.

4.5. Which model does provide the best performance? (RQ5)

To address a comparative evaluation of the models used by the selected studies, the authors of this systematic literature review cluster the discussion in terms of:

1. Studies using the same datasets;
2. Studies applying the same prediction method or algorithm;
3. Studies based on the same input signal duration;
4. Studies within the same class of Artificial Intelligence applied method (according to Table 9);
5. All the studies.

Table 9
Class of Artificial Intelligence methods applied by the selected studies. Yellow/Outer = Artificial Intelligence, Green/Middle = Machine Learning, Blue/Inner = Deep Learning.



1. From all the selected studies, only five of them used the same dataset, leaving all the remaining ones working with a dataset that only they used.

Thus, and comparing all the studies that used the Atrial Fibrillation Prediction Database [22], three of them achieved accuracies above or equal to 90% by applying (ordered by accuracy level decreasing) Mixture of Experts, Support Vector Machine, and Statistical AI methods [19,20,23]. The two worst performing studies both used Support Vector Machine [21,30], thus not being possible to indicate what was the best method to apply.

2. Regarding studies applying Support Vector Machine, those who perform the best both used as features LF, SD2, and Sample Entropy [20,30].

When looking at the studies that applied Convolutional Neural Networks as a prediction method, both acquired very similar accuracy rates [24,31].

3. Relatively to the studies based on the input of signals with 300 s length, the authors of this systematic literature review highlight the article that used as method Mixture of Experts [19], also linking the two best performances with the usage of the features LF, HF, SD1, SD2, SD1/SD2 and Sample Entropy [19,20].

In the studies using signals of 30 s [18,24], both performed around 85% of accuracy, but the second achieved higher performance, having a higher amount of individuals from whom the data was collected, as well as applying Support Vector Machine instead of Convolutional Neural Network as the first.

From the two articles that report work done with signals with 10 s length [25,31], the second performed better than the first, and applied Convolutional Neural Network method instead of Support Vector Machine, as well as having a higher number of individuals from whom the data was collected (approximately 5,25 times).

4. From the two studies that applied Deep Learning methods [24,31], both acquired very similar accuracy rates.

Looking into the articles working with Machine Learning methods (excluding those who apply Deep Learning techniques) [18–21,25,27,30], the two that outperformed all the others, achieving accuracies above 95%, used the Atrial Fibrillation Prediction Database, worked with signals of 300 s long and with frequency-domain, time-domain and non-linear features extracted from the input ECG signals.

5. The results of the studies revealed that the increase in the length of the period of ECG signal sent for prediction does not necessarily increase the accuracy of the model created. The best prediction accuracies were obtained in the studies [19] (98.21%), [20] (96.64%) and [23] (90.00%), in which there were used signal parts of 300 s. Contrasting, the worst accuracies achieved by the models from the selected articles were obtained in the studies [33] (70.49%), [25] (75.60%) and [21] (80.20%), with signal durations of 3600, 10 and 1800 s respectively. These data can indicate that signals too short (10 s only) or too long (1800 s or above) are not the best approach to the problem being assessed in this systematic literature review.

At last, from the results from the three studies that applied Artificial Intelligence methods [23,29,33], the authors highlight the achieved accuracy of the first study, which worked with Atrial Fibrillation Prediction Database, having signals with 300 s long instead of 120 (second study) or 3600 (third study), performing better among the three.

At last, the authors highlight the achieved accuracy of the study [23], that worked with Atrial Fibrillation Prediction Database, with ECG signals 300 s long instead of 120 (as used on the study [33]), or 3600 (on the study [29]). All these three studies used Artificial Intelligence methods.

5. Conclusion

The present systematic literature review presents and summarizes the current data-based work on predicting Atrial Fibrillation (AF) using Electrocardiogram (ECG) data as input and Artificial Intelligence (AI) methods. Twelve studies were analysed, and the main findings are summarized as follows:

- (RQ1) Despite not existing a current high number of articles published based on studies focused on prediction of AF using AI and ECG signals, most of the existing ones assess the problem by predicting only AF cases, not spending time in the prediction of other cardiovascular issues at the same time, thus being the major number of studies a binary prediction system. The higher part of the existing studies worked with ECG signals 300 s long, that is, five minutes. Although some studies tried increasing the length of the period of ECG signal used as input for the prediction models, it does not necessarily increase the accuracy of the obtained final model;
- (RQ2) From all the studies selected for this systematic literature review, the most accurate models were achieved using the Atrial Fibrillation Prediction Database for training. This database was also the most used, by almost half of all the selected articles. The most used features are Standard Deviation of RR Intervals, Low-Frequency band power, Mean of RR Intervals and Standard Deviation, all collected from the ECG signal inputted. This can indicate the higher importance of the RR intervals in the ECG exams, for AF prediction purposes.
- (RQ3) Among all the selected articles, there were applied many pre-processing techniques, from which the Noise Removal was the most used, followed by the QRS complex detection, to allow the collection of the most used features related to the RR intervals and peaks;
- (RQ4) The trend in predictive methods based on Machine Learning techniques is increasing. From all the selected studies, the two most used methods were Support Vector Machine and Convolutional Neural Network, indicating the Machine Learning techniques as a trend in this field. However, the authors of this systematic literature review noticed that the usage of deep learning techniques is yet not highly accurate when comparing to simpler Support Vector Machine methods, allowing bigger inconsistency of results and higher difficulty of getting the desired results from the analysed data;

- (RQ5) Generally, the models based on Machine Learning methods achieved higher accuracy rates. The higher accuracy was obtained by applying a Mixture of Experts method, followed by a Support Vector Machine implementation. The selected features that conducted to higher accuracy were LF, SD2, and Sample Entropy. Also, the usage of ECG signals 300 s long as input for the method's training led to a high rate of prediction accuracy. The database that conducted all the three most accurate models achieved was the Atrial Fibrillation Prediction Database.

As shown by Fig. 2, between 2009 and 2019 (the time window of reference for this systematic literature review), more than 80% of the total published studies were performed after 2016, 50% belonging to the last two years (2018 and 2019).

The amount of work on the prediction of AF episodes is rapidly increasing and showing promising results. Although deep learning methods have already shown outstanding results on the prediction of several areas, namely healthcare, but were not yet applied to many studies, this is, focusing on the prediction of AF using ECG signals. The best results tend to be achieved using Machine Learning and Deep Learning techniques, namely Support Vector Machine and Mixture of Experts.

At last, some limitations of this systematic literature review should be mentioned.

First, this systematic literature review only concerned research in papers written in English. Second, the research for articles returned few articles, even with a cross-reference of the selected studies. Third, this review excluded all the studies that included data collected from patients with recent surgical proceedings or with known cardiovascular conditions that could infer the results of an ECG exam. Finally, the selected studies had to contain evaluation measurements such as accuracy, sensitivity, specificity, or the confusion matrix, excluding any article without any of these evaluations.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is funded by FCT/MCTES through national funds and when applicable co-funded EU funds under the project UIDB/EEA/50008/2020 (Este trabalho é financiado pela FCT/MCTES através de fundos nacionais e quando aplicável cofinanciado por fundos comunitários no âmbito do projeto UIDB/EEA/50008/2020).

This article/publication is based on work from COST Action IC1303 - AAPELE - Architectures, Algorithms and Protocols for Enhanced Living Environments and COST Action CA16226 - SHELD-ON - Indoor living space improvement: Smart Habitat for the Elderly, supported by COST (European Cooperation in Science and Technology). More information in www.cost.eu.

References

[1] D.J. Gladstone, et al., Atrial fibrillation in patients with cryptogenic stroke, New. Engl. J. Med. 370 (26) (2014) 2467–2477, http://dx.doi.org/10.1056/NEJMoa1311376.

[2] T. Sanna, et al., Cryptogenic stroke and underlying atrial fibrillation, New. Engl. J. Med. 370 (26) (2014) 2478–2486, http://dx.doi.org/10.1056/NEJMoa1313600.

[3] Mayo Clinic, Atrial fibrillation - Symptoms and causes - Mayo Clinic, 2019, [Online]. Available: https://www.mayoclinic.org/diseases-conditions/atrial-fibrillation/symptoms-causes/syc-20350624. (Accessed 29 January 2020).

[4] Mayo Clinic Staff, Electrocardiogram (ECG or EKG) - Mayo Clinic, Mayo Clinic, 2018, [Online]. Available: https://www.mayoclinic.org/tests-procedures/ekg/about/pac-20384983. (Accessed 29 January 2020).

[5] I. Matias, N. Pombo, N. Garcia, Towards a fully automated bracelet for health emergency solution, in: IoTBDS 2018 - Proceedings of the 3rd International Conference on Internet of Things, Big Data and Security, Vol. 2018-March, 2018, pp. 307–314, http://dx.doi.org/10.5220/0006800803070314.

[6] N. Pombo, P. Araújo, J. Viana, M.D. Da Costa, Evaluation of a ubiquitous and interoperable computerised system for remote monitoring of ambulatory post-operative pain: A randomised controlled trial, Technol. Health Care 22 (1) (2014) 63–75, http://dx.doi.org/10.3233/THC-130774.

[7] N. Pombo, N.M. Garcia, UbiSleep: An ubiquitous sensor system for sleep monitoring, in: International Conference on Wireless and Mobile Computing, Networking and Communications, 2016, http://dx.doi.org/10.1109/WiMOB.2016.7763192.

[8] N. Pombo, P. Araújo, J. Viana, B. Junior, R. Serrano, Contribution of web services to improve pain diaries experience, in: Lecture Notes in Engineering and Computer Science, Vol. 2195, 2012, pp. 589–592.

[9] H.P. Da Silva, J. Guerreiro, A. Lourenço, A. Fred, R. Martins, BITalino: A novel hardware framework for physiological computing, in: PhyCS 2014 - Proceedings of the International Conference on Physiological Computing Systems, 2014, pp. 246–253, http://dx.doi.org/10.5220/0004727802460253.

[10] T. Lysaght, H.Y. Lim, V. Xafis, K.Y. Ngiam, AI-assisted decision-making in healthcare: The application of an ethics framework for big data in health and research, Asian Bioethics Rev. 11 (3) (2019) 299–314, http://dx.doi.org/10.1007/s41649-019-00096-0.

[11] S. Higgins, P.T. Julian, Green, Cochrane handbook for systematic reviews of interventions, 2019.

[12] A. Liberati, et al., The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration, PLoS Med. 6 (7) (2009) e1000100, http://dx.doi.org/10.1371/journal.pmed.1000100.

[13] D. Moher, et al., Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement, Ann. Intern. Med. 151 (4) (2009) 264–269, http://dx.doi.org/10.7326/0003-4819-151-4-200908180-00135, American College of Physicians.

[14] F. Weidt, R. Silva, Systematic Literature Review in Computer Science-A Practical Guide, Vol. 1, No. 0, Relatórios Técnicos do DCC/UFJF, 2016, pp. 1–7, http://dx.doi.org/10.1027/1016-9040.11.3.244.

[15] The best research databases for computer science [Update 2019] - Paperpile, 2020, [Online]. Available: https://paperpile.com/g/research-databases-computer-science/. (Accessed 12 April 2020).

[16] M. Ouzzani, H. Hammady, Z. Fedorowicz, A. Elmagarmid, Rayyan-a web and mobile app for systematic reviews, Syst. Rev. 5 (1) (2016) 210, http://dx.doi.org/10.1186/s13643-016-0384-4.

[17] 5: Inclusion and exclusion criteria - Systematic Reviews - Library Guides at University of Melbourne, 2020, [Online]. Available: https://unimelb.libguides.com/c.php?g=492361&p=3368110. (Accessed 15 January 2020).

[18] H. ElMoaqet, Z. Almuwaqat, M. Ryalat, N. Almtireen, A new algorithm for short term prediction of persistent atrial fibrillation, in: 2017 IEEE Jordan Conf. Appl. Electr. Eng. Comput. Technol., 2017, pp. 1–6.

[19] E. Ebrahimzadeh, et al., Prediction of paroxysmal Atrial Fibrillation: A machine learning based approach using combined feature vector and mixture of expert classification on HRV signal, Comput. Methods Programs Biomed. 165 (2018) 53–67.

[20] M. Mohebbi, H. Ghassemian, Prediction of paroxysmal atrial fibrillation based on non-linear analysis and spectrum and bispectrum features of the heart rate variability signal, Comput. Methods Programs Biomed. 105 (1) (2012) 40–49.

[21] K.H. Boon, M. Khalil-Hani, M.B. Malarvili, C.W. Sia, Paroxysmal atrial fibrillation prediction method with shorter HRV sequences, Comput. Methods Programs Biomed. 134 (2016) 187–196.

[22] G.B. Moody, A.L. Goldberger, S. McClennen, S.P. Swiryn, Predicting the onset of paroxysmal atrial fibrillation: The computers in cardiology challenge 2001, Comput. Cardiol. (2001) 113–116, http://dx.doi.org/10.1109/CIC.2001.977604.

[23] H. Costin, C. Rotariu, A. Păsărică, ru, Atrial fibrillation onset prediction using variability of ECG signals, in: 2013 8th Int. Symp. Adv. Top. Electr. Eng., 2013, pp. 1–4.

[24] J. Kim, O. Sangjun, Y. Kim, M. Lee, Convolutional neural network with biologically inspired retinal structure, in: 7th Annu. Int. Conf. Biol. Inspired Cogn. Archit., Vol. 88, BICA 2016, held July 16 to July 19, 2016 New York City, NY, USA, 2016, pp. 145–154.

[25] Y. Shen, Y. Yang, S. Parish, Z. Chen, R. Clarke, D.A. Clifton, Risk prediction for cardiovascular disease using ECG data in the China kadoorie biobank, in: 2016 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 2016, pp. 2419–2422.

[26] R. Questionnaire, China kadoorie biobank, 2013, [Online]. Available: https://www.ckbiobank.org/site/. (Accessed 18 January 2020).

[27] S. Rajalakshmi, K.V. Madhav, A collaborative prediction of presence of Arrhythmia in human heart with electrocardiogram data using machine learning algorithms with analytics, J. Comput. Sci. 15 (2) (2019) 278–287.
[28] A. Frank, A. Asuncion, {UCI} machine learning repository, 2010.
[29] Z. Li, H. Derksen, J. Gryak, H. Ghanbari, P. Gunaratne, K. Najarian, A novel atrial fibrillation prediction algorithm applicable to recordings from portable devices, in: 2018 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 2018, pp. 4034–4037.
[30] K.H. Boon, M. Khalil-Hani, M.B. Malarvili, Paroxysmal atrial fibrillation prediction based on HRV analysis and non-dominated sorting genetic algorithm III, Comput. Methods Programs Biomed. 153 (2018) 171–184.
[31] Z.I. Attia, et al., An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction, Lancet 394 (10201) (2019) 861–867.
[32] Mayo Clinic, Mayo Clinic Mayo Clinic, 2019.
[33] M. Mohamed, M. Farah, A. Fahed, Evidence-based prediction of Atrial Fibrillation using physiological signals, in: 2019 3rd Int. Conf. Bio-engineering Smart Technol., 2019, pp. 1–4.
[34] M. Laboratory For Computational Physiology, The MIMIC III clinical database, 2015, http://dx.doi.org/10.13026/C2XW26.
[35] J. Pan, W.J. Tompkins, A real-time QRS detection algorithm, IEEE Trans. Biomed. Eng. BME-32 (3) (1985) 230–236, http://dx.doi.org/10.1109/TBME.1985.325532.
[36] MATLAB, MATLAB documentation, MATLAB, 2012, p. R2012b, http://dx.doi.org/10.1201/9781420034950.
[37] Y. Jia, et al., Caffe: convolutional architecture for fast feature embedding, in: MM 2014 - Proceedings of the 2014 ACM Conference on Multimedia, 2014, pp. 675–678, http://dx.doi.org/10.1145/2647868.2654889.
[38] C.C. Chang, C.J. Lin, LIBSVM: A Library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (3) (2011) http://dx.doi.org/10.1145/1961189.1961199.
[39] International Organization for Standardization, International standard ISO/IEC 14882:2014(E) – programming language C++, 2014, p. 1358.
[40] Martin Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, et al., Tensorflow: A system for large-scale machine learning, in: 12th $USENIX$ Symposium on Operating Systems Design and Implementation, $OSDI$ 16, 2016, pp. 265–283.