# A Personal Health Agent for Decision Support in Arrhythmia Diagnosis

Tezira Wanyana[1,3], Mbithe Nzomo[1,3(✉)], C. Sue Price[2,3], and Deshendran Moodley[1,3]

[1] University of Cape Town (UCT), Cape Town, South Africa
{twanyana,mnzomo,deshen}@cs.uct.ac.za
[2] University of KwaZulu-Natal (UKZN), Durban, South Africa
pricec@ukzn.ac.za
[3] Centre for Artificial Intelligence Research (CAIR), Cape Town, South Africa

**Abstract.** We propose an architecture for a personal health agent (PHA) that combines machine learning and a Bayesian network (BN) for detecting and diagnosing heart disease, specifically arrhythmia. Machine learning (ML) is used for classifying a patient's ECG signal. Four ML models, i.e. gradient boosting, random forest, multilayer perceptron and support vector machine, are compared and evaluated using a dataset of 5,340 records containing 12-lead ECG signals created from the Chapman-Shaoxing database. Among the four models, the gradient boosting model produces the best accuracy of 82.88% when classifying an ECG signal as either atrial fibrillation, other arrhythmia, or no arrhythmia. The detected pattern is integrated into a BN that captures expert knowledge about the causes of arrhythmia. The BN structure and parameters are informed by expert knowledge from the literature and evaluated using Pitchforth and Mengersen's framework. The agent uses a decision support module to guide the diagnosis process. It suggests what questions to ask to increase certainty of the presence of arrhythmia, and it suggests what arrhythmia causes to follow up. This is achieved using sensitivity analysis and diagnostic Bayesian reasoning respectively. The architecture is evaluated using application use cases.

**Keywords:** ECG · Arrhythmia · Machine learning · Bayesian networks · Agent architecture

## 1 Introduction

With the current surge in popularity of wearable devices such as smart watches and bands, many people are becoming increasingly motivated to monitor their health digitally [38]. Such devices contain sensors that can monitor physical activity by providing the heart rate, step count and sleep patterns among others. Some wearable devices, such as Withings Move and Apple Watch, have the ability to monitor the heart rhythm and to provide ECG readings [44]. Data collected by such devices should be presented back to a user in an understandable format that motivates their actions towards improved health [38].


L. A. Maciaszek et al. (Eds.): ICT4AWE 2021/2022, CCIS 1856, pp. 385–407, 2023.
https://doi.org/10.1007/978-3-031-37496-8_20

This study explores an architecture for an intelligent personal health agent (PHA) that incorporates both machine learning (ML) and knowledge representation techniques for pattern detection, situation analysis and decision support. The goal of the agent is to retrieve data from a wearable device and assist an individual to understand their heart rhythm. The PHA focuses on electrocardiogram (ECG) readings to determine if arrhythmia is present or not and if present, what its possible causes might be. Arrhythmia is a cardiac condition characterised by heart rhythm irregularities which, if left unattended, may lead to stroke [52]. The most common arrhythmia is atrial fibrillation (AF) [26]. The prevalence of AF has increased by 33% in the past two decades, and this prevalence is expected to increase over the next 30 years. Though this study focuses on arrhythmia, the PHA can be extended to cover additional health conditions.

In previous work [50], we proposed an initial version of the PHA which combined ML and a Bayesian network (BN) to interpret and explain the occurrence of AF in a patient in terms of its risk factors. The architecture consisted of four modules: two exogenous modules, i.e. the AI service and the Domain Expert, which were external to the agent, and two endogenous modules, i.e. the Perception and Deliberation modules. The AI service used an ML model to classify an ECG signal as either *P-wave present* or *P-wave absent*. An absent P-wave is the hallmark characteristic of AF in an ECG signal. In the deliberation module, a BN was used to represent the causal relations between different risk factors that influenced the presence of AF. The probabilities of the ML classification were used as the likelihood values for the evidence captured by the P-wave node. When the detected ECG feature was entered into the P-wave node, and the probabilities were propagated, the probabilities of the states of the AF node changed so that the state that corresponds to the detected ECG feature, "AF present", had a higher probability. The PHA then identified the most probable risk factors of the patient's condition from the BN.

In this paper, we describe an extension and refinement of the ML module, the BN, and the overall agent architecture. In the initial PHA, the user would have had to understand how to use BNs in order to interact with the PHA. The architecture now includes a new decision support module, which allows the agent to offer guidance to the clinician on what questions to ask the patient or what risk factors to follow up. Using the decision support module, the clinician can now easily interact with the BN and the agent. The ML module and the BN previously only detected a single form of arrhythmia, AF, in the ECG signal. Individuals may have other types of arrhythmias beyond AF. The problem is now formulated as multiclass classification of the ECG signal to detect either AF, other arrhythmia, or no arrhythmia (none). The inclusion of the third class (other arrhythmia) increases the generalisability and usability of the ML model, allowing it to detect the presence of not only AF but also other types of arrhythmia and understand its causes. Additionally, we evaluated our architecture on a new dataset. In the original study, a combination of the MIT-BIH Arrhythmia and MIT-BIH AF databases were used to train the ML model. The dataset only represented two rhythm types (AF or none) with 24 ECG records. In this study we use the Chapman-Shaoxing database which contains 5,340 ECG records and 11 types of rhythms.

The deliberation module, which involves obtaining the most probable risk factors of an individual's condition, is carried out using the BN. The scope of the BN has

accordingly been extended to cater for three possibilities: AF, other arrhythmia or no arrhythmia. A node (i.e. `ML Prediction: Arrhythmia`) is introduced to form the interface between the ML and the BN. The conditional probability tables (CPTs) of this node are populated using recall values of the applied ML algorithm which is the algorithm with the best accuracy (gradient boosting in this work). In addition, we factor the effect of COVID-19 on cardiac arrhythmia into the design of our BN.

The contribution of this paper is threefold: firstly, an improved PHA agent architecture with comprehensive internal (endogenous) modules including an explicit decision support module; secondly, a multi-class ML model which identifies whether a patient has AF, some other type of arrhythmia, or none, based on an ECG input; and thirdly, an improved prototype arrhythmia BN for interpretation and explanation of the ECG result of a particular patient in terms of arrhythmia risk factors.

The rest of the paper is organised as follows: Sect. 2 discusses related work. Section 3 presents an improved version of the PHA architecture, discussing the essence and relevance of each of the modules. In Sect. 4, details about the dataset used, the model building and the results of the ML experiments are presented. Section 5 discusses in detail how the arrhythmia BN was built and validated. In Sect. 6, we describe the decision support module, and evaluate the PHA with use cases. A detailed discussion is presented in Sect. 7 and we conclude and provide the limitations and future work in Sect. 8.

## 2 Related Work

### 2.1 Electrocardiogram Classification

The electrocardiogram (ECG) remains the gold standard in arrhythmia detection [16]. A typical ECG signal consists of P, T and U waves, as well as the QRS complex [51]. ECG is measured using electrodes placed on the skin. A particular arrangement of electrodes gives rise to a lead, with the simplest lead being a pair of electrodes [14]. The most commonly used lead system is the 12-lead ECG. It is derived from 10 electrodes placed on the legs, arms and chest that form 12 leads, namely Leads I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5 and V6 [23]. The first six leads are referred to as limb leads, and are derived from electrodes placed on the arms and legs. The latter six leads are referred to as precordial leads and are derived from electrodes placed on the chest [23].

The 12-lead ECG is considered the benchmark because it captures a more complete picture of the heart compared to reduced lead systems. Each lead provides a different angle of the electrical activity in the heart; therefore, with a 12-lead ECG, the same electrical event can be viewed from 12 different angles [20]. Wearable devices for measuring ECG often rely on fewer electrodes and, subsequently, fewer leads. Although more leads currently result in better arrhythmia detection [44], reduced-lead ECGs generated from wearable devices are starting to be viable for detecting arrhythmia and are increasingly showing comparable performance to standard 12-lead ECGs [18,42].

### 2.2 ECG Classification Using Machine Learning

The classification of ECG data is part of the situation detection module in the PHA. A situation refers to “an external semantic interpretation of sensor data” [54] in an

application domain. Example situations from an ECG pattern would be “AF” or “other arrhythmia”. Situations are used in the context of the state of monitored features in a physical environment [1]. Situation detection techniques can be categorised as either specification-based or learning-based [54]. Specification-based techniques such as ontologies and evidence theory rely on expert knowledge to model situations and then reason on them with input sensor data. On the other hand, learning-based techniques such as ML uncover patterns or correlations in the data.

ML for arrhythmia classification from ECG signals has gained traction in recent years. Traditional ML algorithms such as tree-based methods and linear models have been widely used and shown to produce good results in ECG analysis, while being simple and computationally inexpensive to train [36]. Popular ML algorithms used in ECG classification include support vector machine (SVM) and ensemble decision trees using techniques such as bagging (e.g. random forest) and boosting (e.g. gradient boosting and adaptive boosting).

### 2.3 Bayesian Networks

Many systems developed for ECG analysis apply ML techniques [2,27]. These systems have a number of limitations. They are not able to deal efficiently with uncertainty; they are considered black boxes and are hard for domain experts to understand; and they demand large datasets [11]. Also, while these techniques have registered great success in heartbeat recognition, beat segmentation and ECG classification, they have had little success in decision support. BNs have the ability to deal with the uncertainty that is embedded in reasoning in cardiology as well as medical reasoning at large; they are understandable to non-technical users [5] and play a large role in decision support. They have been proposed to support the screening, diagnosis, selection of treatment, prognosis and multimorbidity modelling of different medical conditions e.g. cancer and heart disease [11,21].

BNs describe causal relationships between variables using directed acyclic graphs (DAGs). The variables are represented by nodes. For discrete BNs, each node or variable has a number of exhaustive and non-overlapping states [21]. The interaction between the nodes is specified using conditional probability tables (CPTs) in each node [21], which gives the probability of a certain state occurring, given the state of a parent node. The structure of the BN can be developed by hand or from data. The parameters in the CPTs can be generated from domain experts, data and/or literature [21,40,47], making BNs a flexible modelling tool [15]. The CPT values of input nodes (i.e. those without parents) are usually obtained from a distribution of how the states naturally occur [21]. On compiling the BN, if the user is certain of some information, this information is entered into the BN as evidence [21], and the probability of that state becomes 100%. As a result, the probabilities of each of the nodes in the network are updated using Bayes’ rule (Eq. 1 in the Appendix).

**Validation of BNs.** The structure of a BN is usually validated by experts. The parameters of the BN can be validated by data, if it is available [21]. If not, the framework proposed by Pitchforth and Mengersen [40] can be used to evaluate an expert-elicited BN. This framework addresses seven different types of validity: nomological-,

face-, content-, concurrent-, convergent-, discriminant- and predictive validity. One of the tests for assessing predictive validity is sensitivity analysis. This measures how sensitive the network is to changes in input (evidence) and parameter (CPT) values [21]. Measures for analysing sensitivity to evidence, like entropy and mutual information (MI), can be used to determine the degree to which adding evidence about one variable will reduce the uncertainty in our belief of the value of a target variable [21]. Entropy tells us the current uncertainty we have in our belief of the value of some target variable, while MI gives us an indication of the degree to which we can reduce this uncertainty by adding evidence at another variable (see Eqs. 2 and 3 in the Appendix). To gain the largest reduction in uncertainty in the target variable, we should add evidence to the variable with the highest MI value. Thus the MI values can be used to determine a priority ranking in terms of which evidence to gather next. Using this ranking, the PHA can determine which questions should be asked to increase certainty in the target variable.

**BNs in Cardiology.** BNs were explored as a tool for premature ventricular contraction beat classification based on the ECG features of the sinus rhythm and the shape of the beat waveform [9–11]. Domain knowledge on risk factors that have causal influence on the arrhythmia are not incorporated in the BN.

BNs have been used by a number of authors to model the risk of coronary heart disease (CHD), e.g. [13,15,35,37], cited in Korb & Nicholson [21]. These BNs model the factors that lead to CHD, such as age [13,35,37], sex [13,35], smoking [13,15,35,37], obesity [13,15,37], alcohol [15], diabetes [37] and hypertension [13,15,37]. The structures of these BNs varied; some BNs modelled some risk factors as influencing others, while some BNs modelled the risk factors as being independent of each other. Only one of these BNs included a node for an ECG and for rapid heartbeats, which are child nodes of the heart disease node [15]. To our knowledge, no BNs in the literature focus on arrhythmia.

### 2.4 Agent Architecture

The knowledge discovery and evolution (KDE) agent architecture [48] was developed recently for agents that analyse patterns from sensor data in physical systems, and detect, interpret and explain patterns found in this data. The architecture combines "bottom-up" ML techniques for pattern detection with "top-down" knowledge representation and reasoning (KRR) techniques such as BNs for interpreting and explaining these patterns [48,49]. Although these techniques are rarely combined in agents, there is now a convergence on the fact that purely data driven or knowledge driven systems alone are not sufficient for AI systems and that these systems could benefit from incorporating both techniques [17]. Such systems are known as hybrid systems. In this work, we leverage the strengths of both techniques for the different tasks of the PHA.

## 3 PHA Architecture

The architecture for the PHA is shown in Fig. 1. The architecture design draws from the KDE architecture [48]. It consists of four modules: the pattern detection module, per-

ception module, deliberation module and the decision support module, that contribute to the various steps of the agent's overall operation. Each module is discussed in the following subsections.

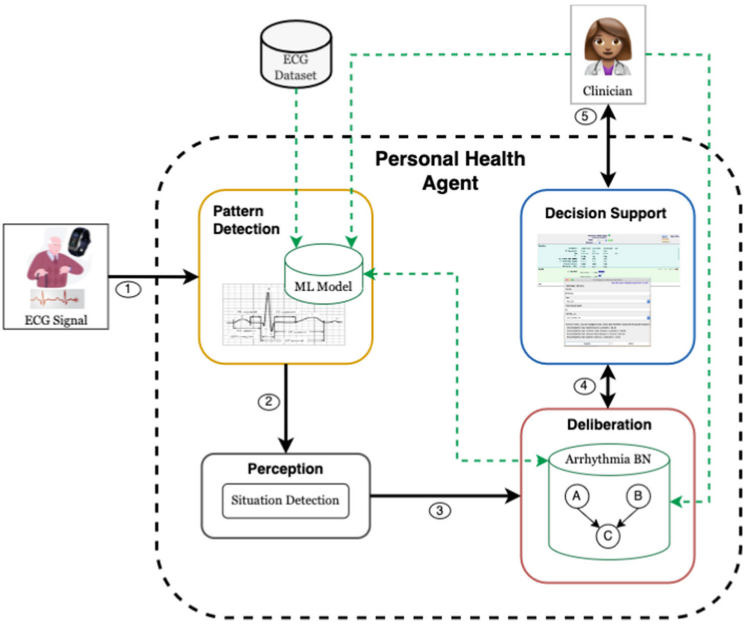


**Fig. 1.** Agent architecture for a personal health agent (PHA).

The clinician is the domain expert, who plays a key role in different model development activities. These activities are represented by the green dashed arrows. The clinician 1) curates the ECG dataset with labels to train and update the ML model; 2) develops the structure and conditional probability tables of the BN, including how it integrates with the ML model; and 3) continuously updates and refines the models as s/he gains new knowledge. For example, s/he may add new data and retrain the ML model and/or revise the BN.

### 3.1 Pattern Detection

The first step of the PHA's operational loop is pattern detection. The incoming ECG signal from a wearable device is passed to the pattern detection module (arrow 1). The module uses a pre-trained ML model to classify the ECG signal. The ML model is trained prior to the agent's execution. Details are discussed in Sect. 4. The clinician is responsible for the curation of the ECG dataset.

### 3.2 Perception

The perception module provides a bridge between the pattern detection module and the deliberation module. It consists of a situation detection module which uses rules, of the form $if$ $<pattern>$ $then$ $<situation>$, to map patterns detected from the ML model to situations in the BN. For the current version of the PHA agent, the patterns (AF, other arrhythmia and none) can be added directly to the BN as evidence, so these rules are not used.

### 3.3 Deliberation

The deliberation module focuses on how the PHA arrives at its suggestion for following up a given situation. The main component of the deliberation module is the BN. BNs have been used in other agent architectures in their deliberation processes, e.g. [12,41]. The PHA agent has knowledge that is captured using the BN presented in detail in Sect. 5. The objective of the PHA is to suggest the best way to follow up the observed situation; it does not provide treatment options. For instance, the PHA suggests what further information to find about an individual's lifestyle or what risk factors to check for. On this note, the cost of having false positives is much lower that the cost of having false negatives. The information encoded in the BN can be drawn from expert knowledge or from literature; it can also be designed from data.

Using the BN, prior to deliberation, situation analysis is conducted. The situation obtained in the perception module is mapped onto a corresponding state in the observable node (i.e. `ML Prediction: Arrhythmia`) in the BN by entering the matching state as evidence. The agent then determines whether a given situation should be followed up or not. A situation where AF or other arrhythmia is suspected to be present needs to be followed up. A normal ECG which indicates that no arrhythmia is present requires no further attention.

In the deliberation module, the agent considers specific evidence of the example form *ML Prediction: Arrhythmia == Atrial fibrillation*. The selected state corresponds to the situation detected during perception. During deliberation, the agent requires context. The agent starts by incorporating context that consists of information about the demographic risk factors, specifically age and sex. The lifestyle risk factors (alcohol abuse, smoking and obesity) also form part of the context, if information about them is available. After all evidence and context has been added to the BN, the belief values of all the disease risk factors (e.g. hypertension) that have causal influence on the arrhythmia node are extracted.

### 3.4 Decision Support

The decision support module provides an interactive tool which serves as the key interface between the agent and the clinician. The tool allows the clinician to add information about a patient iteratively and to determine the effect of this information on the presence of arrhythmia in the patient. Based on sensitivity analysis of the BN, the agent informs the clinician on the most important information that is required to reduce the uncertainty of the patient having arrhythmia. As more information about the patient is provided, the

agent obtains a clearer picture of the patient's situation and is able to determine the risk the patient having arrhythmia or not. In addition, the agent also displays the likelihood of other disease risk factors that the clinician may need to follow up.

## 4 Machine Learning for Arrhythmia Detection

In this section, we present details about the model building phase of the pattern detection module, which was conducted prior to the agent execution. In this phase, we use a dataset of 5,340 records containing 12-lead ECG signals to train ML models. The full source code for the data preprocessing, hyperparameter tuning and model building is publicly available on GitHub[1].

### 4.1 Dataset

The Chapman-Shaoxing database [55], a large publicly available ECG database, was used to create the dataset. The database was developed as a result of a collaboration between Chapman University, California and Shaoxing Hospital Zhejiang University School of Medicine, China. It contains 10,646 12-lead ECG records, each from a unique patient, with a duration of 10 s and a frequency 500 Hz. The database contains 11 rhythm categories, with each record belonging to only one: sinus bradycardia, sinus rhythm, atrial fibrillation, sinus tachycardia, atrial flutter, sinus irregularity, supraventricular tachycardia, atrial tachycardia, atrioventricular node reentrant tachycardia, atrioventricular reentrant tachycardia, and sinus atrium to atrial wandering rhythm. The sinus rhythm category denotes a normal rhythm with no arrhythmia, while the sinus irregularity category denotes an unspecified irregular rhythm.

Each signal in the database was filtered using a bandpass Butterworth filter to remove noise. Two demographic features, age and sex, were obtained for each record. Additionally, we computed statistical features from the ECG signals, as proposed in the Physionet/Computing in Cardiology 2020 and 2021 challenges [3,43]. Seven statistical features, i.e. mean, median, standard deviation, variance, skewness, and kurtosis, were calculated for both the R-R intervals and the R-peak values for each of the 12 leads, resulting in 144 statistical features. The R-R intervals and R-peak values were computed using Sznajder and Łukowska's QRS detector [46] based on the Pan-Tomkins algorithm. The root mean square for each lead was also calculated, resulting in another 12 features. Together with age and sex, this resulted in a total of 158 features. The statistical features were then normalised using Scikit-learn's `StandardScaler`, which removes the mean and scales the data to unit variance. The mathematical formula for this is $z = \frac{(x-u)}{s}$, where $z$ is the standard score of a sample $x$, $u$ is the mean of the samples, and $s$ is the standard deviation of the samples.

There were 1,780 AF records in the database. To ensure a balanced dataset, 1,780 records from each of the other two classes were randomly selected. To determine the effect of the number of classes on the ML model performance, we used a subset of the dataset with two classes, AF and None. Table 1 shows the details of the multiclass and binary datasets, including summary statistics for age and sex. The impact of the number of classes is discussed in Sect. 4.3.

[1] https://github.com/mbithenzomo/personal-health-agent-ml.

**Table 1.** Dataset details.

| Dataset | Classes | Records per class | Total records | Sex ratio | Age range | Age percentiles ($25^{th}$, $50^{th}$, $75^{th}$) |
|---|---|---|---|---|---|---|
| Multiclass | AF | 1780 | 5340 | M: 2,842 F: 2,498 | Min: 4 Max: 98 | 51, 64, 75 |
| | Other | 1780 | | | | |
| | None | 1780 | | | | |
| Binary | AF | 1780 | 3560 | M: 1,826 F: 1,734 | Min: 5 Max: 98 | 53, 67, 77 |
| | None | 1780 | | | | |

### 4.2 Model Development

Four ML classification algorithms were implemented using Scikit-learn [39]: support vector machine (SVM), random forest, gradient boosted decision trees, and multilayer perceptron (MLP). The first three were selected due to their wide use in the literature for ECG classification, while the MLP was selected as it was the best performing model in our previous work. Hyperparameter tuning for the four algorithms was done using Scikit-learn's `GridSearchCV`, which performs an exhaustive search over specified parameter values and returns the best combination of parameters. The options and selected hyperparameters for each algorithm are shown in the Appendix.

As we did in our previous study [50], we used 10-fold cross-validation to evaluate the performance of the models. Cross-validation has been shown to be effective in accurately assessing the generalisation performance of ML models [19]. A stratified approach was chosen to maintain the distribution of each class in each fold.

### 4.3 Model Evaluation and Selection

The ML models were evaluated using several metrics, averaged across the test folds during the cross-validation process: overall accuracy, overall confusion matrix, and the precision, recall, and F1-score for each class. Accuracy refers to the percentage of correct predictions for the test data, and is obtained by dividing the number of correct classifications by the total number of both correct and incorrect classifications. The confusion matrix shows the number of correctly classified examples and incorrectly classified examples, with the number of correctly classified examples forming the main diagonal of the matrix.

The precision for a particular class is calculated by dividing the number of correctly classified examples for that class by the total number of examples that were classified as belonging to that class, whether correctly or incorrectly. On the other hand, the recall for a particular class is the number of correctly classified examples for that class, divided by the actual number of examples for that class. The F1 score is a computation of the harmonic mean of the precision and recall. The formulae for these metrics are shown in the Appendix.

Including the demographic features generally resulted in an increase in accuracy for all the models. The impact of data normalisation was most significant in the SVM

and MLP models, which performed much better with normalised data. In contrast, the gradient boosting and random forest models were not significantly affected by normalisation. With regards to the number of output classes, the binary classification accuracy was significantly better than the multiclass accuracy. This can be attributed to the fact that many arrhythmias have similar irregularities in the ECG, such as in the R-R interval [7].

Table 2 and Table 3 show the classification results for the multiclass and binary datasets respectively, including the precision, recall and F1-score by class as well as the overall average confusion matrices and accuracy scores. Overall, the best performing model was the gradient boosting model, which achieved an average accuracy of 82.88% for the multiclass dataset and 93.85% for the binary dataset. The 93.85% accuracy achieved for the binary classification is an improvement on the performance of the best performing model in our previous work, which was an MLP that achieved an accuracy of 89.61%.

**Table 2.** Multiclass classification results (AF, Other and None).

| Algorithm | Classes | Precision | Recall | F1-Score | Confusion Matrix | Accuracy |
|---|---|---|---|---|---|---|
| **Gradient Boosting** | AF | 86.24% | 89.10% | 87.65% | $\begin{pmatrix} \underline{1586} & 113 & 81 \\ 146 & \underline{1364} & 270 \\ 107 & 197 & \underline{1476} \end{pmatrix}$ | **82.88%** |
| | Other | 81.48% | 76.63% | 78.98% | | |
| | None | 80.79% | 82.92% | 81.84% | | |
| Random Forest | AF | 80.76% | 87.02% | 83.78% | $\begin{pmatrix} \underline{1549} & 141 & 90 \\ 192 & \underline{1228} & 360 \\ 177 & 170 & \underline{1433} \end{pmatrix}$ | 78.84 % |
| | Other | 79.79% | 68.99% | 74.00% | | |
| | None | 76.10% | 80.51% | 78.24 % | | |
| MLP | AF | 77.04% | 79.38% | 78.20% | $\begin{pmatrix} \underline{1413} & 198 & 169 \\ 218 & \underline{1255} & 307 \\ 203 & 297 & \underline{1280} \end{pmatrix}$ | 73.93% |
| | Other | 71.71% | 70.51% | 71.10% | | |
| | None | 72.89% | 71.91% | 72.40% | | |
| SVM | AF | 73.82% | 81.74% | 77.58% | $\begin{pmatrix} \underline{1455} & 181 & 144 \\ 312 & \underline{1075} & 393 \\ 204 & 279 & \underline{1297} \end{pmatrix}$ | 71.67% |
| | Other | 70.03% | 60.39% | 64.86% | | |
| | None | 70.72% | 72.87% | 71.78% | | |

**Table 3.** Binary classification results (AF and None).

| Algorithm | Classes | Precision | Recall | F1-Score | Confusion Matrix | Accuracy |
|---|---|---|---|---|---|---|
| **Gradient Boosting** | AF | 92.77% | 95.11% | 93.93% | $\begin{pmatrix} \underline{1693} & 87 \\ 132 & \underline{1648} \end{pmatrix}$ | **93.85%** |
| | None | 94.99% | 92.58% | 93.77% | | |
| Random Forest | AF | 88.88% | 92.53% | 90.67% | $\begin{pmatrix} \underline{1647} & 133 \\ 206 & \underline{1574} \end{pmatrix}$ | 90.48% |
| | None | 92.21% | 88.43% | 90.28% | | |
| MLP | AF | 88.10% | 87.75% | 87.93% | $\begin{pmatrix} \underline{1562} & 218 \\ 211 & \underline{1569} \end{pmatrix}$ | 87.95% |
| | None | 87.80% | 88.15% | 87.97% | | |
| SVM | AF | 86.09% | 89.04% | 87.54% | $\begin{pmatrix} \underline{1585} & 195 \\ 256 & \underline{1524} \end{pmatrix}$ | 87.33% |
| | None | 88.66% | 85.62% | 87.11% | | |

## 5 Arrhythmia Bayesian Network

In this section, we describe the development and validation of the BN used in the deliberation module.

### 5.1 Developing the Bayesian Network

We designed a prototype arrhythmia BN using the Netica GUI[2]. The arrhythmia BN prototype is illustrated in Fig. 2. The aim of the BN is to show factors which could have causal influence on arrhythmia. The structure of the BN was informed by medical idioms for BNs [24]: nodes with prefix "RF" denote risk factors, and the condition (Arrhythmia) is denoted with prefix "C". Other medical idioms [24,33] used in developing the BN were the definitional/synthesis idiom (`Lifestyle risk factors` summarises the alcohol abuse, smoking and obesity risk factors); and the cause-consequence idiom (`Age` and `Lifestyle risk factors` cause the four traditional risk factors of hypertension, ischemic and valvular heart disease and diabetes). CPT values for the nodes were based on literature, including our previous work [50]. Using literature means that models can be created when domain experts are not available [21]. This knowledge has the advantage of having been peer reviewed, compared to eliciting the knowledge directly from domain experts [21].

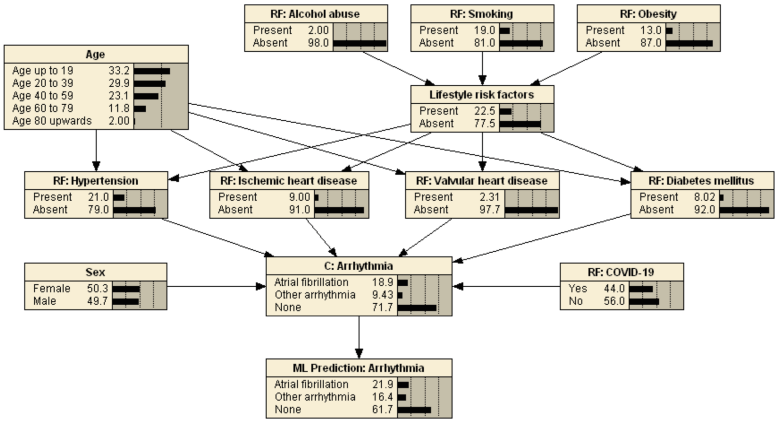


**Fig. 2.** A prototype BN for explaining the causes of arrhythmia with no evidence added.

There are three lifestyle risk factors in the BN: alcohol abuse, smoking and obesity. Other traditional risk factors are hypertension, ischemic heart disease, valvular heart disease, diabetes mellitus and whether the person had contracted COVID-19 or not. COVID-19 was added to the AF BN [50], since patients who have had COVID-19 developed different types of arrhythmia [29], the most common being AF [8]. Age and sex are demographic factors. Prior probabilities of the age node were obtained from global population percentages [4].

The ML model is applied in the BN in the following way: the chosen ML algorithm is the one with the best accuracy, i.e. gradient boosting (see Table 3). Of special interest is the `ML Prediction: Arrhythmia` node, which is the node in which the ML prediction is entered as evidence. The CPT values of the `ML Prediction: Arrhythmia` node contain recall values based on the confusion matrix given in Table 2 for gradient boosting as shown in the example instance below.

[2] https://www.norsys.com/download.html.

P(ML Prediction: Arrhythmia == Atrial fibrillation | C:Arrhythmia == Atrial fibrillation) = recall (Atrial fibrillation)

These represent the proportion of correctly vs incorrectly classified samples identified by the ML model. These CPT values are defined as the probability of the ML algorithm detecting a given state given that it is indeed the patient's condition. The CPT of the ML Prediction: Arrhythmia is shown in Fig. 3. Based on the output of the situation detection module, evidence is entered into one of the states of the ML Prediction: Arrhythmia node, depending on whether the ML algorithm identified AF, other arrhythmia or none.

| C: Arrhythmia | Atrial fibrillation | Other arrhythmia | None |
|---|---|---|---|
| Atrial fibrillation | 89.1 | 6.35 | 4.55 |
| Other arrhythmia | 8.2 | 76.63 | 15.17 |
| None | 6.01 | 11.07 | 82.92 |

**Fig. 3.** The CPT for the ML Prediction: Arrhythmia node.

The BN reasons diagnostically (against the flow of the arrows) to infer the most likely risk factors of the patient, given the evidence in the ML Prediction: Arrhythmia node. These give explanations of the occurrence of arrhythmia in the patient in terms of the risk factors. If more information is known about the patient, such as their demographic or lifestyle factors, these can also be entered into the BN as evidence.

### 5.2 Validating the Bayesian Network

Pitchforth and Mengerson's criteria for evaluating expert-elicited BNs [40] were used by the authors to evaluate the BN. The BN falls within the cardiology domain, with an emphasis on arrhythmia. This confirms nomological validity. The structure, node discretisation and parameters in the BN are what would be expected. This confirms face validity.

In the prototype arrhythmia BN, the main risk factors for AF match those mentioned in the literature (e.g. [6,25,34,53]). The risk factors for arrhythmia are similar to those of AF [28,45], and the most prevalent arrhythmia risk factors found in the literature are modelled in the BN. The states of the nodes cover the range of values for each node, with no gaps. The CPTs of input nodes are based on prevalence values from literature. These evaluations confirm content validity. However, it should be noted that in certain populations, some risk factors are more prevalent than others. This would also affect the CPT values of the BN.

Concurrent validity is determined by comparing how the BN and other theoretically similar BNs act. Apart from the AF BN [50], we do not have access to other working BNs in the cardiology domain. The nodes and discretisation of the prototype arrhythmia BN are the same as those of the AF BN, apart from the C: Arrhythmia, COVID-19, ML Prediction: Arrhythmia nodes which are new. In addition, the discretisation of the Age node changed from four states to five states: an additional state was added for ages of under 20, since it is possible to experience other types of

arrythmia during these ages. The CPTs of the arrhythmia BN are the same as the AF BN, apart from those of the `Age` nodes and the four traditional risk factors. Both networks show a similar prior distribution (see Fig. 2). The probability in the AF state increases in both BNs when evidence for the disease risk factors is added, which is as expected. The arrhythmia BN also follows the medical idioms of Kyrimi et al. [24], as outlined in Sect. 5.1.

By comparing the BN to others in the literature, convergent validity could be established. Apart from the AF BN [50], no BNs explaining the causes of arrhythmia were found. The structure of the arrhythmia BN mirrors sub-networks of other cardiology-related BNs, with minor exceptions. For example, in the Busselton BN which modelled the risk of CHD, risk factors such as being overweight, drinking alcohol, smoking, diabetes, age and sex were identified [35]. In this BN, all factors led to the node which predicted a risk of a CHD event in the following 10 years. In another CHD BN [37], age, smoking and obesity led to diabetes, which is similar to the arrhythmia BN. In Ghosh and Valtorta [15], obesity, smoking and alcohol leads to hypertension, and heart disease leads to rapid heartbeats; however, they model hypertension as leading to heart disease, whereas in our BN, these are modelled as two independent risk factors for arrhythmia. A BN modelling cardiovascular risk [13] also identified the factors of weight, smoking, sex and age. Weight could influence hypertension, as in our BN. In their BN though, sex influenced age, and age influenced the smoking habits and weight of the patient. It should be noted, however, that the structure of this cardiovascular risk BN was generated by averaging the structures of 500 different networks whose structure was learned from data. The face validity of the arrhythmia BN was judged to be fair.

Predictive validity is evaluated by assessing the behaviour of the BN when it is executed, the sensitivity of the BN to findings or to parameters, and how the BN behaves for extreme conditions. Evidence was entered into the `C: Arrhythmia` node's "Atrial fibrillation" state, modelling that a person has AF (see Fig. 4). The inferred values of

**Table 4.** Comparison of the four traditional risk factor values, given AF, to Nguyen et al. [34].

| Node | BN value (Fig. 4) | Min prevalence value in [34] | Max prevalence value in [34] |
|---|---|---|---|
| `RF: Hypertension` | 32.8% | 10.3% | 71.9% |
| `RF: Ischemic heart disease` | 16.9% | 6.4% | 47% |
| `RF: Valvular heart disease` | 6.02% | 5.6% | 66% |
| `RF: Diabetes mellitus` | 11.8% | 3.3% | 33% |

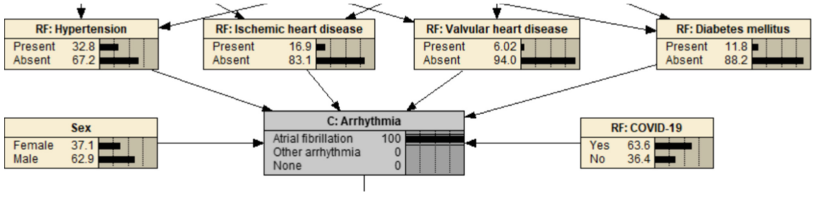


**Fig. 4.** Extract of the prototype BN showing arrhythmia risk factors, given that AF is present.

these four nodes fall within the ranges outlined in Nguyen et al.'s systematic literature review of the prevalence of AF [34] in developing contexts (see Table 4).

Extreme conditions such as the sex and age of the person were used to evaluate the BN. The BN's results for AF matches the statement of Naccarelli et al. [32] that AF increases with age, and that more men have AF than women, at any age. Adding evidence for lifestyle factors to the BN increased the probability of AF and of other arrhythmia.

The sensitivity of the BN to findings was examined (see Table 5). This table shows what evidence should be sought to increase certainty about the values of the `C: Arrhythmia` node. Whether the patient has had COVID or not gives the most added certainty, followed by hypertension, the patient's sex, ischemic heart disease, the patient's age, valvular heart disease, diabetes and other lifestyle risk factors.

This ranking reflects the main modifiable risk factors of the most prevalent arrhythmia, AF [22,34]. However, the risk factors may have a larger or smaller effect on arrhythmia, depending on the population being modelled.

Based on the validation, the BN suitably represents factors causing arrhythmia. However, this prototype BN should be tested and evaluated further before it is deployed for real world use.

**Table 5.** Sensitivity of `C: Arrhythmia` to findings at other nodes.

| Node | Mutual Information | Percent | Variance of Beliefs |
|---|---|---|---|
| `C: Arrhythmia` | 1.11949 | 100 | 0.2590261 |
| `ML Prediction: Arrhythmia` | 0.54069 | 48.3 | 0.1175418 |
| `RF: COVID-19` | 0.04423 | 3.95 | 0.0098642 |
| `RF: Hypertension` | 0.02255 | 2.01 | 0.0049399 |
| `Sex` | 0.02009 | 1.79 | 0.0045437 |
| `RF: Ischemic heart disease` | 0.01959 | 1.75 | 0.0040420 |
| `Age` | 0.01798 | 1.61 | 0.0039373 |
| `RF: Valvular Heart Disease` | 0.01517 | 1.35 | 0.0025202 |
| `RF: Diabetes mellitus` | 0.00525 | 0.469 | 0.0011528 |
| `Lifestyle risk factors` | 0.00029 | 0.0258 | 0.0000659 |
| `RF: Smoking` | 0.00006 | 0.00552 | 0.0000141 |
| `RF: Obesity` | 0.00004 | 0.00368 | 0.0000094 |
| `RF: Alcohol Abuse` | 0.00001 | 0.000544 | 0.0000014 |

## 6 Evaluation of the PHA

### 6.1 Decision Support Module

The Decision Support module provides the key interface between the clinician and the agent. The agent is able to provide two types of decision support, i.e. predictive support and diagnostic support.

**Predictive Support.** The agent assists the clinician to find out the most relevant information that might increase certainty about the presence of arrhythmia in a patient. This is achieved using the sensitivity analysis shown in Table 5. For example, assuming that the ML prediction is already provided: the next most important information is whether the patient has had COVID, and then whether there is hypertension present. At each stage of the process, the system recommends the best questions to ask based on the previous answers given and it updates its conclusions. A screenshot of the dashboard obtained using AutoNetica[3] is shown in Fig. 5. It provides an ordered set of the most important questions, based on the MI scores, that would increase the certainty in the `C: Arrhythmia` node. In this way the clinician is guided through a set of questions. As the answers are provided, the posterior probabilities (at the bottom of the screen) of the different arrhythmia states are updated.

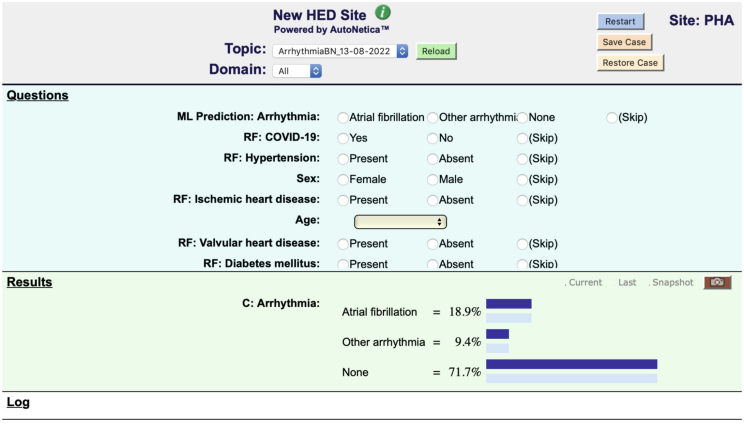


**Fig. 5.** A display of the of the most relevant questions to ask (using Autonetica).

**Diagnostic Support.** The agent also provides a diagnostic support interface to the clinician. Assuming that the clinician believes that a patient has Arrhythmia, then the agent is able to advise the clinician of the most likely underlying diseases that may have caused the arrhythmia in this patient. The diagnostic support interface is shown in Fig. 6. The clinician supplies demographic information and lifestyle risk factors for the patient, if this is available. Based on the supplied information, the PHA displays the probabilities of the patient having traditional risk factors that should be followed up (see bottom of Fig. 6). At the top of this interface is a link to the predictive decision support described above. The interface is implemented in NeticaJ[4], the Java version of the Netica API.

## 6.2 PHA Evaluation

To evaluate the PHA, we apply case-based evaluation [21] on a set of application use cases. Cases are generated to test different situations i.e., a situation where atrial fib-

[3] https://www.norsys.com/WebHelp/NETICA/X_AutoNetica.htm.
[4] https://www.norsys.com/netica-j.html#download.

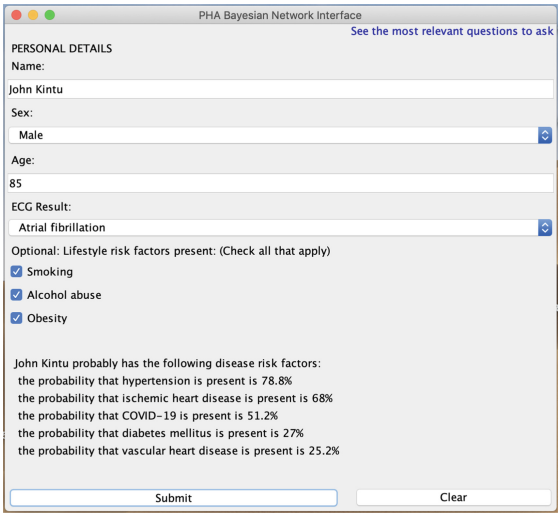


**Fig. 6.** A display of the traditional risk factors and their probabilities for hypothetical patient "John Kintu": an 85 year old male who is obese, smokes, abuses alcohol and has an ECG which shows the presence of AF.

rillation is present and a situation where some other arrhythmia is present. The context variables are changed to extreme values in the different cases to ensure reasonable network performance across a variety of possible cases. For example, the extreme upper case would be an obese male individual above 80 years who smokes and abuses alcohol.

The first use case is for the hypothetical patient, "John Kintu", an 85 year old male who is obese, smokes, abuses alcohol and has presented with an ECG that shows the presence of AF. Figure 7 shows the BN with the same information about the hypothetical patient, "John Kintu" as shown in Fig. 6 entered as evidence. The agent suggests that for John's demographic factors and his lifestyle risk factors (obesity, smoking and alcohol abuse), he would have 78.8% chance of having hypertension and 68% chance

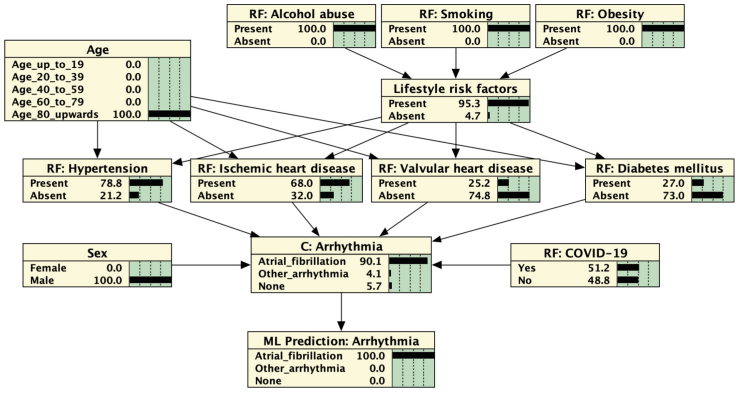


**Fig. 7.** BN for hypothetical patient "John Kintu" with evidence entered, generated in Java using NeticaJ.

of ischemic heart disease. He would therefore have to confirm using medical tests if these conditions are indeed present and manage them.

For the second use case, let us consider another hypothetical individual who is a 25 year old female with two lifestyle risk factors, alcohol abuse and smoking. Assuming that from her ECG, the predicted situation is "other arrhythmia", this is captured as evidence in the `ML prediction: Arrhythmia` node. The patient's age and sex as well as lifestyle risk factors are also captured. As shown in Fig. 8, the chances that indeed, the patient has "other arrhythmia" is 33.7%. This patient has a 16.4% probability of having hypertension and 12% probability of having ischemic heart disease. The low probabilities of the traditional risk factors can be attributed to the fact that the patient is young.

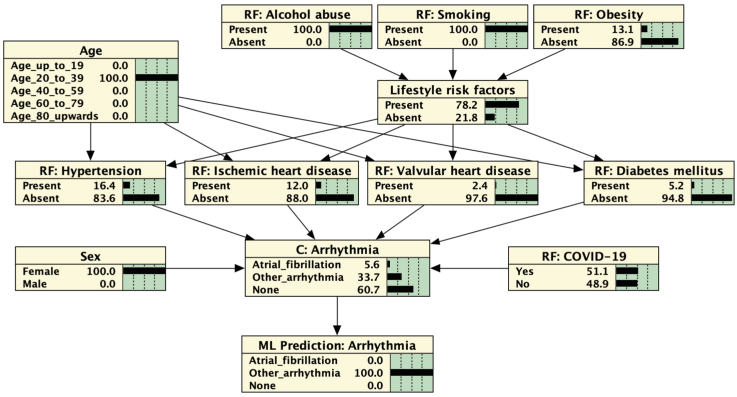


**Fig. 8.** The BN with evidence for a hypothetical 25 year old female with two lifestyle risk factors: alcohol abuse and smoking. The situation detected is that she has "other arrhythmia".

It is important to note that the agent does not suggest with full certainty that someone has AF, other arrhythmia or none. The moment the agent predicts that the patient has AF or other arrhythmia, the clinician should consider confirming the condition using a medical test.

## 7 Discussion and Conclusions

In this paper, we have described an architecture for a personal health agent (PHA) to support situation detection, situation analysis and decision support for diagnosing arrhythmia. The PHA is based on an agent architecture that incorporates an ML model for detecting irregular patterns and situations in a patient's ECG signal, a BN for capturing expert knowledge about the causes of arrhythmia, and a decision support module which guides the diagnosis process. The architecture combines an ML model, a BN and a decision support module to detect the presence of arrhythmia in a patient and to determine its likely causes. The ML model takes in an ECG signal and detects whether

arrhythmia is present in the patient. The BN, which captures expert knowledge from the literature contains an ML prediction node. This is used to integrate the ML model predictions into the inference process. Once evidence is entered in the ML prediction node, together with the demographic, lifestyle and disease risk factors of the patient, the probability of the patient having arrhythmia is determined using Bayesian inferencing.

We introduced a decision support module that provides specific support for clinical decision making. The module supports both predictive and diagnostic reasoning. It uses sensitivity analysis and predictive reasoning to identify the most relevant questions for the clinician to ask the patient to determine whether arrhythmia is present or not. It uses diagnostic reasoning to determine the possible causes of a patient's arrhythmia given the information that is currently available for the patient. The decision support module provides interactive user interfaces which guides the clinician through the diagnosis process without requiring knowledge about using and interpreting BNs. This is an improvement on the previous PHA architecture [50] which did not have a decision support module.

The other key extension was the augmentation of the agent to cater for other arrhythmia conditions besides AF. In our previous study [50], the focus was on distinguishing between AF and normal rhythms from an ECG signal. This was done by training an ML model to classify an ECG pattern based on the absence of a P-wave. The best performing model was an MLP which achieved an accuracy of 89.61% on a combination of the MIT-BIH Arrhythmia and MIT-BIH AF databases. In this study, we include a third class for other types of arrhythmia and use a different dataset. The best performing model is now a gradient boosting classifier, which achieves an accuracy of 82.88% when distinguishing between the three classes on the Chapman-Shaoxing database. The drop in accuracy is not surprising given the similarities in certain ECG characteristics in different arrhythmias, such as the R-R interval [7]. When evaluating the gradient boosting classifier on the binary problem (AF or normal) the model achieves an accuracy of 93.85%, which is an improvement on our previous work.

Like the ML model, the BN was extended to accommodate AF, other arrhythmia and none. In our previous BN, the `Age` node only catered for patients 20 years old and over. AF is typically not experienced in younger people, but other types of arrhythmia can be. An additional state in the `Age` node was added to cater for under 20 year olds. As a result, the CPT values of the traditional risk factors were updated. In the current BN and PHA, there is a stronger interface between the ML model and the BN since we added a node (the `ML Prediction: Arrhythmia` node) for which the CPT is obtained from the recall values of the ML classifier. This allows the BN to account for the false positives of the ML classifier, such that a prediction of arrhythmia by the ML classifier does not provide a definitive diagnosis of the presence of arrhythmia.

A key aspect of this architecture is the integration of ML and BNs to support situation detection, situation analysis and decision-making. Some architectures that incorporate BNs for decision support have been proposed previously [12,30,31,41]. Similar to these architectures, our architecture incorporates a BN in deliberation and decision support. However, these architectures did not incorporate ML into the BN inference process and did not provide an interactive decision making module.

## 8 Limitations and Future Work

The PHA has some limitations. Firstly, the ML model was trained using a 12-lead dataset. This dataset was selected due to its large size, expert annotation, inclusion of demographic information, and that it incorporated various types of arrhythmia. To our knowledge, there is no equivalent publicly available dataset for wearables. As part of our ongoing work, we will evaluate our architecture on ECG datasets from wearable devices as these become available. Additionally, the feature extraction in this study focused on the R-R interval and R-peak of the ECG signal. The use of additional features, such ventricular rate, atrial rate and characteristics of the QRS complex may result in better differentiation between AF and other arrhythmias. The need for manual feature extraction can be eliminated by deep learning, which will be explored in future work.

In the current arrhythmia BN, we modelled only the key risk factors for arrhythmia. The BN can be extended to incorporate other risk factors associated with arrhythmia. While the BN was evaluated using Pitchforth and Mengersen's framework based on expert findings from the literature, we did not consult any experts to verify the BN. We acknowledge that this should be done before it can be deployed for real world use. Furthermore, the CPT values for the BN may differ, and should be customised for the population where it will be used.

**Acknowledgements.** This work was financially supported by the Hasso Plattner Institute for Digital Engineering through the HPI Research School at UCT.

## Appendix

### Bayes' Rule

$$Pr(A|B) = \frac{Pr(B|A)Pr(A)}{Pr(B)} \tag{1}$$

where Pr(A|B) is the posterior probability of A given B; Pr(B|A) is the posterior probability of B given A; and Pr(A) and Pr(B) are prior probabilities of A and B respectively.

### Entropy and Mutual Information

$$ENT(X) = -\sum P(x)logP(x) \tag{2}$$

$$MI(X|Y) = ENT(X) - ENT(X|Y) \tag{3}$$

### Hyperparameter Selection Table

| Algorithm | Hyperparameter | Options | Selected Options |
|---|---|---|---|
| Gradient Boosting | n_estimators | 300, 500, 800 | 800 |
| | criterion | friedman_mse, squared_error, mse | friedman_mse |
| | loss | log_loss, exponential | log_loss |
| | max_depth | 1, 3, 10 | 3 |
| Random Forest | n_estimators | 300, 500, 800 | 800 |
| | criterion | gini, entropy, log_loss | entropy |
| | max_depth | 50, 100, None | None |
| | max_features | sqrt, log2, None | sqrt |
| SVM | C | 0.5, 1, 1.5 | 1.5 |
| | kernel | poly, rbf, sigmoid | rbf |
| | gamma | scale, auto | auto |
| | decision_function_shape | ovo, ovr | ovo |
| MLP | hidden_layer_sizes | N/A | (158, 100, 50) |
| | activation | identity, logistic, tanh, relu | tanh |
| | batch_size | auto, 64, 100 | auto |
| | solver | lbfgs, sgd, adam | adam |
| | learning_rate | constant, invscaling, adaptive | adaptive |
| | max_iter | 200, 500, 1000, 2000 | 500 |

### ML Metrics: Accuracy, Precision, Recall, F1-Score

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{7}$$

## References

1. Adeleke, J.A., Moodley, D., Rens, G., Adewumi, A.O.: Integrating statistical machine learning in a semantic sensor web for proactive monitoring and control. Sensors **17**(4), 807 (2017). https://doi.org/10.3390/s17040807
2. Al Rahhal, M.M., Bazi, Y., Al Zuair, M., Othman, E., BenJdira, B.: Convolutional neural networks for electrocardiogram classification. J. Med. Biol. Eng. **38**(6), 1014–1025 (2018). https://doi.org/10.1007/s40846-018-0389-7
3. Alday, E.A.P., et al.: Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. Physiol. Measur. **41**(12), 124003 (2020). https://doi.org/10.1088/1361-6579/abc960

4. Ang, C.: Visualizing the world's population by age group (2021). https://www.visualcapitalist.com/the-worlds-population-2020-by-age/. Accessed 12 Aug 2022
5. Chen, S.H., Pollino, C.A.: Good practice in Bayesian network modelling. Environ. Model. Softw. **37**, 134–145 (2012). https://doi.org/10.1016/j.envsoft.2012.03.012
6. Chung, M.K., et al.: Lifestyle and risk factor modification for reduction of atrial fibrillation: a scientific statement from the American Heart Association. Circulation **141**(16), e750–e772 (2020). https://doi.org/10.1161/CIR.0000000000000748
7. Clifford, G.D., et al.: AF classification from a short single lead ECG recording: the PhysioNet/Computing in Cardiology Challenge 2017. In: 2017 Computing in Cardiology (CinC), pp. 1–4. IEEE (2017). https://doi.org/10.22489/CinC.2017.065-469
8. Coromilas, E.J., et al.: Worldwide survey of COVID-19-associated arrhythmias. Circ.: Arrhythmia Electrophysiol. **14**(3), e009458 (2021). https://doi.org/10.1161/CIRCEP.120.009458
9. De Oliveira, L.S., Andreão, R.V., Sarcinelli-Filho, M.: Detection of premature ventricular beats in ECG records using Bayesian networks involving the P-wave and fusion of results. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, pp. 1131–1134. IEEE (2010). https://doi.org/10.1109/IEMBS.2010.5627116
10. De Oliveira, L.S., Andreão, R.V., Sarcinelli-Filho, M.: Premature ventricular beat classification using a dynamic Bayesian network. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4984–4987. IEEE (2011). https://doi.org/10.1109/IEMBS.2011.6091235
11. De Oliveira, L.S.C., Andreão, R.V., Sarcinelli-Filho, M.: The use of Bayesian networks for heart beat classification. In: Hussain, A., Aleksander, I., Smith, L., Barros, A., Chrisley, R., Cutsuridis, V. (eds.) Brain Inspired Cognitive Systems 2008. AEMB, vol. 657, pp. 217–231. Springer, Cham (2010). https://doi.org/10.1007/978-0-387-79100-5_12
12. Drake, R., Moodley, D.: INVEST: ontology driven Bayesian Networks for investment decision making on the JSE. In: Second Southern African Conference for Artificial Intelligence Research, pp. 252–273 (2022)
13. Fuster-Parra, P., Tauler, P., Bennasar-Veny, M., Ligeza, A., Lopez-Gonzalez, A., Aguiló, A.: Bayesian network modeling: a case study of an epidemiologic system analysis of cardiovascular risk. Comput. Methods Programs Biomed. **126**, 128–142 (2016). https://doi.org/10.1016/j.cmpb.2015.12.010
14. Geselowitz, D.B.: On the theory of the electrocardiogram. Proc. IEEE **77**(6), 857–876 (1989). https://doi.org/10.1109/5.29327
15. Ghosh, J.K., Valtorta, M.: Building a Bayesian network model of heart disease. In: 38th Annual on Southeast Regional Conference (ACM-SE 38), pp. 239–240 (2000). https://doi.org/10.1145/1127716.1127770
16. Hagiwara, Y., et al.: Computer-aided diagnosis of atrial fibrillation based on ECG signals: a review. Inf. Sci. **467**, 99–114 (2018). https://doi.org/10.1016/j.ins.2018.07.063
17. Harmelen, F.: Preface: the 3rd AI wave is coming, and it needs a theory. In: Neuro-Symbolic Artificial Intelligence: The State of the Art, pp. V–VII. IOS Press BV (2022). https://doi.org/10.3233/FAIA210347-fm
18. Hartikainen, S., et al.: Effectiveness of the chest strap electrocardiogram to detect atrial fibrillation. Am. J. Cardiol. **123**(10), 1643–1648 (2019). https://doi.org/10.1016/j.amjcard.2019.02.028
19. Hastie, T., Tibshirani, R., Friedman, J.: Model assessment and selection. In: Hastie, T., Tibshirani, R., Friedman, J. (eds.) The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edn., pp. 219–259. Springer, New York (2009). https://doi.org/10.1007/978-0-387-84858-7_7
20. Klabunde, R.E.: Cardiac electrophysiology: normal and ischemic ionic currents and the ECG. Adv. Physiol. Educ. **41**(1), 29–37 (2017). https://doi.org/10.1152/advan.00105.2016

21. Korb, K.B., Nicholson, A.E.: Bayesian Artificial Intelligence. CRC Press (2011)
22. Kornej, J., Börschel, C.S., Benjamin, E.J., Schnabel, R.B.: Epidemiology of atrial fibrillation in the 21st century: novel methods and new insights. Circ. Res. **127**(1), 4–20 (2020). https://doi.org/10.1001/jama.2019.18058
23. Kusumoto, F.M.: ECG Interpretation: From Pathophysiology to Clinical Application. Springer, Boston (2009). https://doi.org/10.1007/978-0-387-88880-4
24. Kyrimi, E., Neves, M.R., McLachlan, S., Neil, M., Marsh, W., Fenton, N.: Medical idioms for clinical Bayesian network development. J. Biomed. Inform. **108**, 103495 (2020). https://doi.org/10.1016/j.jbi.2020.103495
25. Lau, D.H., Nattel, S., Kalman, J.M., Sanders, P.: Modifiable risk factors and atrial fibrillation. Circulation **136**(6), 583–596 (2017). https://doi.org/10.1161/CIRCULATIONAHA.116.023163
26. Lippi, G., Sanchis-Gomar, F., Cervellin, G.: Global epidemiology of atrial fibrillation: an increasing epidemic and public health challenge. Int. J. Stroke **16**(2), 217–221 (2021). https://doi.org/10.1177/1747493019897870
27. Lu, J., et al.: Efficient hardware architecture of convolutional neural network for ECG classification in wearable healthcare device. IEEE Trans. Circuits Syst. I Regul. Pap. **68**(7), 2976–2985 (2021). https://doi.org/10.1109/TCSI.2021.3072622
28. Mayo Clinic: Heart arrhythmia (2022). https://www.mayoclinic.org/diseases-conditions/heart-arrhythmia/symptoms-causes/syc-20350668. Accessed 13 July 2022
29. Mohammad, M., Emin, M., Bhutta, A., Gul, E.H., Voorhees, E., Afzal, M.R.: Cardiac arrhythmias associated with COVID-19 infection: state of the art review. Expert Rev. Cardiovasc. Ther. **19**(10), 881–889 (2021). https://doi.org/10.1080/14779072.2021.1997589
30. Moodley, D., Simonis, I.: A new architecture for the sensor web: the SWAP framework. In: ISWC 2006: 5th International Semantic Web Conference (2006)
31. Moodley, D., Simonis, I., Tapamo, J.R.: An architecture for managing knowledge and system dynamism in the worldwide sensor web. Int. J. Semant. Web Inf. Syst. (IJSWIS) **8**(1), 64–88 (2012). https://doi.org/10.4018/jswis.2012010104
32. Naccarelli, G.V., Varker, H., Lin, J., Schulman, K.L.: Increasing prevalence of atrial fibrillation and flutter in the United States. Am. J. Cardiol. **104**(11), 1534–1539 (2009). https://doi.org/10.1016/j.amjcard.2009.07.022
33. Neil, M., Fenton, N., Nielson, L.: Building large-scale Bayesian networks. Knowl. Eng. Rev. **15**(3), 257–284 (2000). https://doi.org/10.1017/S0269888900003039
34. Nguyen, T.N., Hilmer, S.N., Cumming, R.G.: Review of epidemiology and management of atrial fibrillation in developing countries. Int. J. Cardiol. **167**(6), 2412–2420 (2013). https://doi.org/10.1016/j.ijcard.2013.01.184
35. Nicholson, A.E., Twardy, C.R., Korb, K.B., Hope, L.R.: Decision support for clinical cardiovascular risk assessment. In: Pourret, O., Naim, P., Marcot, B. (eds.) Bayesian Networks: A Practical Guide to Applications, pp. 33–52. Wiley, Chichester (2008). https://doi.org/10.1002/9780470994559.ch3
36. Olier, I., Ortega-Martorell, S., Pieroni, M., Lip, G.Y.: How machine learning is impacting research in atrial fibrillation: implications for risk prediction and future management. Cardiovasc. Res. **117**(7), 1700–1717 (2021). https://doi.org/10.1093/cvr/cvab169
37. Orphanou, K., Stassopoulou, A., Keravnou, E.: DBN-extended: a dynamic Bayesian network model extended with temporal abstractions for coronary heart disease prognosis. IEEE J. Biomed. Health Inform. **20**(3), 944–952 (2015). https://doi.org/10.1109/JBHI.2015.2420534
38. Patel, M.S., Asch, D.A., Volpp, K.G.: Wearable devices as facilitators, not drivers, of health behavior change. JAMA **313**(5), 459–460 (2015). https://doi.org/10.1001/jama.2014.14781
39. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al.: Scikitlearn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)

40. Pitchforth, J., Mengersen, K.: A proposed validation framework for expert elicited Bayesian networks. Expert Syst. Appl. **40**(1), 162–167 (2013). https://doi.org/10.1016/j.eswa.2012.07.026
41. Price, C.S., Moodley, D., Pillay, A.W., Rens, G.B.: An adaptive probabilistic agent architecture for modelling sugarcane growers' decision-making. S. Afr. Comput. J. **34**(1), 152–191 (2022). https://doi.org/10.18489/sacj.v34i1.857
42. Rajakariar, K., Koshy, A.N., Sajeev, J.K., Nair, S., Roberts, L., Teh, A.W.: Accuracy of a smartwatch based single-lead electrocardiogram device in detection of atrial fibrillation. Heart **106**(9), 665–670 (2020). https://doi.org/10.1136/heartjnl-2019-316004
43. Reyna, M.A., Sadr, N., Alday, E.A.P., Gu, A., Shah, A.J., Robichaux, C.: Will two do? Varying dimensions in electrocardiography: The PhysioNet/Computing in Cardiology Challenge 2021. In: 2021 Computing in Cardiology (CinC), pp. 1–4. IEEE (2021). https://doi.org/10.23919/CinC53138.2021.9662687
44. Scholten, J., et al.: Six-lead device superior to single-lead smartwatch ECG in atrial fibrillation detection. Am. Heart J. **253**, 53–58 (2022). https://doi.org/10.1016/j.ahj.2022.06.010
45. Scripps: Top 10 things you should know about heart rhythm (2015). https://www.scripps.org/sparkle-assets/documents/heart_rhythm_facts.pdf. Accessed 09 July 2022
46. Sznajder, M., Łukowska, M.: Python online and offline ECG QRS detector based on the Pan-Tomkins algorithm (2017). https://zenodo.org/record/591747. Accessed 22 July 2022
47. Uusitalo, L.: Advantages and challenges of Bayesian networks in environmental modelling. Ecol. Model. **203**(3–4), 312–318 (2007). https://doi.org/10.1016/j.ecolmodel.2006.11.033
48. Wanyana, T., Moodley, D.: An agent architecture for knowledge discovery and evolution. In: Edelkamp, S., Möller, R., Rueckert, E. (eds.) KI 2021. LNCS (LNAI), vol. 12873, pp. 241–256. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87626-5_18
49. Wanyana, T., Moodley, D., Meyer, T.: An ontology for supporting knowledge discovery and evolution. In: Gerber, A. (ed.) Southern African Conference for Artificial Intelligence Research (SACAIR), pp. 206–221 (2020)
50. Wanyana, T., Nzomo, M., Price, C.S., Moodley, D.: Combining machine learning and Bayesian networks for ECG interpretation and explanation. In: 8th International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE), pp. 81–92. INSTICC, SciTePress (2022). https://doi.org/10.5220/0011046100003188
51. Wasilewski, J., Polonski, L.: An introduction to ECG interpretation. In: Gacek, A., Pedrycz, W. (eds.) ECG Signal Processing, Classification and Interpretation: A Comprehensive Framework of Computational Intelligence, pp. 1–20. Springer, Cham (2012). https://doi.org/10.1007/978-0-85729-868-3_1
52. Weimann, K., Conrad, T.O.F.: Transfer learning for ECG classification. Sci. Rep. **11**(1), 1–12 (2021). https://doi.org/10.1038/s41598-021-84374-8
53. Williams, B.A., Chamberlain, A.M., Blankenship, J.C., Hylek, E.M., Voyce, S.: Trends in atrial fibrillation incidence rates within an integrated health care delivery system, 2006 to 2018. JAMA Netw. Open **3**(8), e2014874–e2014874 (2020). https://doi.org/10.1001/jamanetworkopen.2020.14874
54. Ye, J., Dobson, S., McKeever, S.: Situation identification techniques in pervasive computing: a review. Pervasive Mob. Comput. **8**(1), 36–66 (2012). https://doi.org/10.1016/j.pmcj.2011.01.004
55. Zheng, J., Zhang, J., Danioko, S., Yao, H., Guo, H., Rakovski, C.: A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. Sci. Data **7**(1), 1–8 (2020). https://doi.org/10.1038/s41597-020-0386-x