Interpretation of SVM Using Data Mining Technique to Extract Syllogistic Rules

Sanjay Sekar Samuel, Nik Abdullah, Anil Raj

Interpretation of SVM using Data Mining Technique to Extract Syllogistic Rules

Exploring the Notion of Explainable AI in diagnosing CAD

Sanjay Sekar Samuel[1]

Nik Nailah Binti Abdullah[2]

Anil Raj[3]

[1] Monash University, Kuala Lumpur, Malaysia,
mail@sanjaysekarsamuel.com
[2] Monash University, Kuala Lumpur, Malaysia,
nik.nailah@monash.edu
[3] IHMC, Florida. USA,
araj@ihmc.us

Abstract. Artificial Intelligence (AI) systems that can provide clear explanations of their behaviors have been suggested in many studies as a critical feature for human users to develop reliance and trust when using such systems. Medical Experts (ME) in particular while using an AI assistant system must understand how the system generates disease diagnoses before making patient care decisions based on the AI's output. In this paper, we report our work in progress and preliminary findings toward the development of a human-centered explainable AI (XAI) specifically for the diagnosis of Coronary Artery Disease (CAD). We applied syllogistic inference rules based on CAD Clinical Practice Guidelines (CPGs) to interpret the data mining results using a Support Vector Machine (i.e., SVM) classification technique— which forms an early model for a knowledge base (KB). The SVM's inference rules are then explained through a voice system to the MEs. Based on our initial findings, we discovered that MEs trusted the system's diagnoses when the XAI described the chain of reasoning behind the diagnosis process in a more interpretable form—suggesting an enhanced level of trust. Using syllogistic rules alone, however, to interpret the classification of the SVM algorithm lacked sufficient contextual information— which required augmentation with more descriptive explanations provided by a medical expert.

Keywords: Explainable AI, Coronary Artery Disease, Support Vector Machine, Data mining, Medical Expert, Artificial Intelligence, Human-centered.

1 Introduction and Motivation

Explainable artificial intelligence (XAI) or Interpretable AI seeks to make expert systems more transparent by ensuring the understandability of the design's complexities and operations of the AI, thereby promoting trust to those who use the systems [1]. Explanation capabilities are not just desirable, but also crucial for the success of an expert system [2]. These explanations will enable users to understand the contents and limitations of the system's knowledge base (KB) and its chain of reasoning process across that KB [3]. Machine learning (ML), a branch of AI which includes pattern identification from input streams and predict the outputs [4] can be classified as supervised, unsupervised and semi-supervised learning techniques. A Support Vector Machine (SVM), one of the well-known supervised ML approaches, can use data mining techniques to classify the data into categorical class labels while supporting visual representations of how the ML algorithm categorized the processed data [5]. Currently, SVM methods are being used to improve diagnosis of diseases like Coronary Artery Disease (CAD) in clinical settings [6, 7]. They have also demonstrated high performance in solving classification problems in bioinformatics [8, 9]. Though analogous ML algorithms like neural networks (NNs) have also shown remarkable performance while classifying outcomes in CAD, they lack direct support for visualization capabilities like SVM algorithm. Decision trees on the other hand have also demonstrated excellent performance in visualizing their tree branching. However, subject matter experts or medical experts (ME) in particular, have found the visual explanation approach given by the SVM algorithm's graphs more accordant compared to the complex neural network and decision tree graphs. SVM also come handy in representing cases that are similar through the visualization of shared patient characteristics [10]. Moreover, SVM also allows the use of syllogisms, a form of deductive reasoning where one infers a conclusion 'C' through a rule-based examination of two other premises 'A & B' which can be given as A+ B = C [11]. Resultantly, since SVM algorithms comes right after neural networks in the interpretability and accuracy graph, this swayed us to choose SVM as the ML model for our initial work to build the components of an XAI model to accommodate the ME's request.

In the domain of healthcare, ML solutions must provide an explanation of the AI's rationale for the resultant classifications or predictions to enable Medical Experts (ME) to understand it themselves and explain the information to their patients [12]. The provided explanation develops trust in the ML by both the MEs and patients. Moreover, when MEs were asked to rank in the order of importance the top fifteen computer-based consultation systems that develops trust in them, they ranked the "ability to explain their diagnostic and treatment decisions to ME users" as the most essential [13]. At our initial stage of work, we define XAI as an AI model that can provide an accurate and contextualized explanation that can establish the aspect of trust with MEs. To explore the notion of XAI, we use SVM that used inference data mining techniques to extract

syllogistic rules from unseen data. The syllogistic inference rules were then validated with MEs and accepted Clinical Practice Guidelines (CPGs) from the American College of Cardiology [14]. CPGs or commonly known as Medical guidelines are universally accepted documents that are accepted by all the members of the medical community. CPGs have guided decisions and criteria regarding diagnosis, management and treatment in specific areas of health care [14]. Because of their universal standardization and accepted diagnostic rules that have been used and updated for many years, these CPG rules were additionally used to validate the MEs inference of the syllogisms extracted from the data mining results of SVM's classification through graphs. The model then used text-to-speech to explain the MEs which factors in the KB inference rules contributed to the CAD classification. The synthetic voice method was used to evaluate goodness and satisfaction of the MEs mental models of the XAI. Once the goodness and satisfaction are evaluated by the MEs, we can then calculate the precision and performance score of the complete XAI model.

With that, we organize this chapter as follows: First, we will discuss the problem statement and related work briefly. Then we introduce our research methodology. Next, we elaborate on our experiments using data mining techniques and syllogistic inference rules to interpret SVM's classification. Lastly, we conclude with a discussion of the results and future work.

2 Problem Statement

CAD manifests as narrowing of the lumen of the coronary arteries of the heart— which can impair or block blood flow to the cardiac tissue. CAD has risen to be one of the greatest scourges of the human population since the industrial revolution [15]. The multifactorial causes of CAD (elevated cholesterol, smoking, diabetes, aging, etc.,) can make early diagnosis challenging without using invasive coronary catheterization, or costly computed tomography (CT), angiography [16]. Applying XAI in the context of early CAD diagnosis could provide recommendations for the CPG-recommended tests based on model outputs for each patient. Deployment of an XAI model that can address these challenges could result in significant cost savings in developing countries such as Malaysia and India, where the government heavily subsidizes patient care [17]. In light of these challenges, our research study seeks to explore the components required to develop an XAI model that can assist MEs in the diagnosis of CAD by focusing on outlier cases.

Thus, at the initial stage, we focus on two objectives:

1. Interpreting SVM-based classification, and the modeling of syllogistic rules in a KB.
2. Communicating the interpretation to MEs in a manner that develops trust.

Therefore, the preliminary aim of our project is to answer the following research question—how do we represent the classification provided by a ML algorithm to help MEs understand how the XAI arrived at a specific diagnostic output?

3 Related Work

Prior implementations have shown the potential of data mining techniques on SVM-based CAD diagnostic systems [18, 19, 20, 21]. Though other ML methods, like decision tree and neural network models, can demonstrate significant performance in classifying heart disease patients, SVM-approaches are compatible with large and complex data (e.g., "omic" data) and can process data with higher accuracy than other supervised learning algorithmic approaches [22]. Omic data in this context refers to genomics, metabolomics, proteomics, and data from standardized electronic health records (EHRs) or precision medicine platforms. In light of the ML algorithms mentioned earlier, NN-based AI assistants have also shown effective performance in the diagnoses of CAD and other chronic diseases in many medical studies [23]. However, compared to the hidden layers in NNs, SVM algorithm provide a direct pathway to visualize the underlying AI model feature characteristics through graphs while reaching similar or higher levels of accuracy [24]. Additionally, we see that decision tree ML algorithms also support design tools useful for explaining and interpreting predictions derived by the ML algorithms [25]. Representing algorithmic predictions through human-machine[1] explanatory dialogue systems that employ contrastive explanations and exemplar-based explanations can provide transparency and improve trust in XAI [26]. In this context, the XAI uses contrastive rather than direct explanations to illustrate the cause of an event concerning some other event. This is because, with contrastive explanation, the AI has ability to explain itself by identifying the chain of reasoning it went through to reach the diagnostic outcome [27]. Such explanations are a major requirement when a ME evaluates a patient.

When users accept an explanation as good and relatively complete, they can develop a representative mental model, which in turn, fosters appropriate trust towards the AI system [28]. Moreover, these explanations are key contributor to trust because enhancing the explanatory power of an AI systems can result in systems that are easier to use with improved decision-making and problem-solving performance [29]. In our proposed work, we aim to explore the components needed for an XAI model by using SVM and a data mining technique through which we extract syllogistic rules that support transparency to the MEs for CAD diagnosis. We will use voice communication as a tool to explain the interpretation of the classification with contextual information to enhance the level trust while using the XAI. In the next section, we will elaborate more on the methodology involved in discovering and building our early components of an XAI.

[1] Human-machine system is a system in which the functions of a human user and a machine are integrated.

4 Research Methodology

For this study, we undertook a combination of constructive and a human-centered methodology which was iteratively conducted in two stages. The "constructive" methodology here refers to a new concept, theory or model being developed in a research [30], which will then be evaluated by the target user population. And a "human-centered" approach involves an iterative development process where a target user population (MEs in our case) were involved in the loop. These target user population play an important role in this research to make improvements iteratively and identify inconspicuous components required to construct an XAI model [31, 32]. Human-centered approach is crucial for this research as these data sets may contain missing data, unwanted data, dirty data noisy data and most importantly data that has missing contextual information [33, 34]. Thereby, with the integration of experienced MEs in the loop, we can ensure an enhanced level of knowledge discovery process pipeline. This is because, these MEs will have the tacit knowledge required to fill the missing links in these data sets which is critical for an XAI to make its diagnosis for a given condition.

During the first stage of our research we used a constructive approach by using real-world datasets specifically to identify the components needed to construct an early XAI model for the diagnosis of CAD. In the second stage, we applied a human-centered approach, using questionnaires to get qualitative feedback from MEs regarding any improvements needed for the model throughout the iterative stages. The following sections will detail our early model construction and research results.

5 A Syllogistic Approach for Interpreting SVM's Classifications

In this section, we will present the steps involved in designing our early XAI model and its different components, starting with an analysis of the data used to develop the model.

5.1 Data Selection

The data selected for this project was obtained from the University of California-Irvine (UCI) Cleveland heart disease database and the Framingham Heart Study Repository [35, 36]. The UCI repository has 304 records (patients) with 22 attributes (physiological condition). The Framingham heart study database includes 4241 records (patients) with 17 attributes (physiological condition). The attributes/features used for our experiment from the UCI data set are Max heart rate, ST-depression, Number of blood vessels highlighted by fluoroscopy, typical angina, exercise-induced angina, cholesterol level, age, and non-anginal pain. Whereas the attributes/features used for our experiment from the Framingham data set are: SysBP, DiaBP, age, sex, cholesterol level, cigsPerday, diabetes, prevalent hypertension, and glucose level. CPGs and MEs inputs were used to validate the selected attributes, which represent the top ten results given by the uni-

variate feature selection method. Univariate feature selection method is used here because of its unique ability to examine each feature independently by using Pearson correlation to identify the strength of each feature with the target variable [37].

5.2 Analyzing and Interpreting the Information from SVM by using Syllogistic Rules

When analyzing the data as a whole using the SVM algorithm, the interpretation behind the chain of reasoning used by the algorithm tend to get complicated and hard to process because of the vast number of attributes involved in the prediction. However, breaking down the data and analyzing the visualization of the predictions one at a time reduces complexity and simplifies interpretation. To accomplish this, we used classification techniques to interpret information from the SVM algorithm output. Classification technique is a technique where the algorithm parts the data into their respective dependencies depending upon their internal relations. Classification is also one of the most important techniques for data mining [10]. In order to easily extract syllogistic rules, the attributes from the selected data (as detailed from the previous section) were individually classified by the SVM algorithm in the form of visual graphs (see Fig 2 for a detailed excerpt). Both the datasets have a target column that indicates if a patient has heart disease or not (1 for yes and 0 for no). A single attribute selected from the dataset will then be correlated with the target to identify the correlation coefficient as seen in (Fig 1). Correlation coefficient here refers to the numerical measurement of the statistical relationship between two variables [38]. With this correlation coefficient number illustrated on a graph, we were able to identify and extract syllogistic rules from the hyperplane's division.

To delineate this in reference to Fig 1, selected attributes from the dataset (blood pressure, cholesterol, etc.,) placed in variable "X" were individually evaluated to find its correlation coefficient with the target variable "Y" (CAD +/-). The inferred outcomes from the correlation coefficient graph are then used to develop the rules manually in the form of syllogisms.

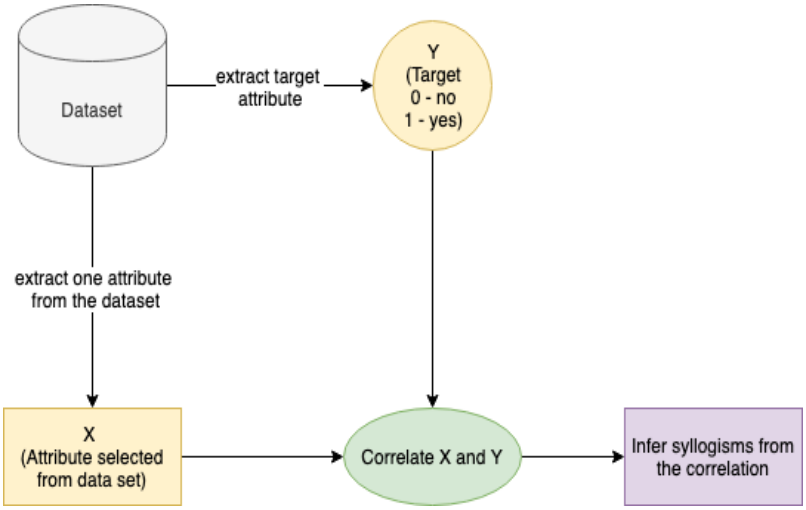

Fig. 1. Overview of the SVM process workflow

5.3 Modelling of Syllogistic Inference Rules

In this section, we illustrate the process of building syllogistic rules through the interpretation of the individual features from the SVM's classification output.

First, we define the following:

$$X = (X_{f1}, X_{f2}, X_{f3} \ldots . X_{fn}) \text{ Set of all features available in the dataset} \quad (1)$$

$$Y = \text{Target} \quad (2)$$

First, to simplify the complexity of the SVM algorithm and make it more interpretable, we extract multiple subsets from the main feature set X ($X_{f1}, X_{f2}, X_{f3} \ldots . X_{fn}$) and find the correlation coefficient of them individually on the target Y. Next, we visualize each of these coefficient values on a graph, which will correspond to the visual reasoning task that is decomposed into a small set of primitives that facilitate the interpretability of the model. Thus, by applying this principle, we know that each subset (X_{f1}, X_{f2}) extracted and visualized from the main feature set X on the target Y will represent a portion of the interpretable classification from the original feature set X. Thereby, with all these subsets combined ([(X_1, X_2), (X_2, X_3)...] on the target Y will attain the complete transparent output of the set X. Thus, extracting a syllogism A_n from each of these subsets and combining these syllogisms together will correspond to have complete interpretation of the classification outcome given by the SVM algorithm.

This can be formulated as follows:

$$\sum (\boldsymbol{X_{f1}}, \boldsymbol{X_{f2}}, \boldsymbol{X_{f3}} \ldots . \boldsymbol{X_{fn}}) \rightarrow \mathrm{Y} \tag{3}$$

$$(\boldsymbol{X_{f1}} + \boldsymbol{X_{f2}}) \rightarrow \mathrm{Y} \Rightarrow \boldsymbol{A_n}(\boldsymbol{X_{f1}} + \boldsymbol{X_{f2}}) \Rightarrow \mathrm{Z} \tag{4}$$

Table 1 (below) defines the variables and symbolic representations used in the syllogism.

Table 1. Variables Explanation

Variable	Representation
X_{f1}	Feature 1
X_{f2}	Feature 2
Y	Target
$A_n(X_{f1} + X_{f2})$	The rule generated with two selected features on the target variable
Z	Interpretable output from the selected features

An Excerpt of Extracting Syllogistic rules from SVM.

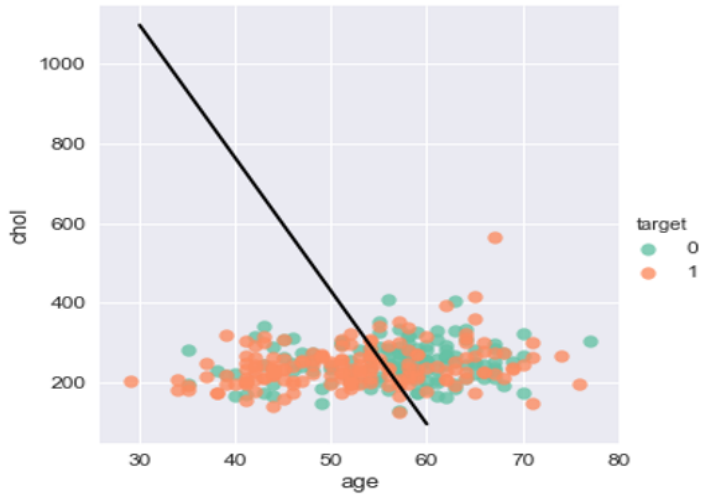

Fig. 2. SVM graphical classification with hyperplane on Cholesterol VS Age

To illustrate the above-mentioned method graphically, we use the example of cholesterol vs age in Fig 2 as the two features extracted from the main feature set. Using the data mining approach on Age vs. Cholesterol as the first result iteration, we observe the effect each selected feature Xage and Xchol (X_{f1}, X_{f2}) has on the target Y. From the hyperplane's division of the plotted points, we could easily interpret that patients below the age of 60 with a cholesterol level of over 200 are more prone to have CAD.

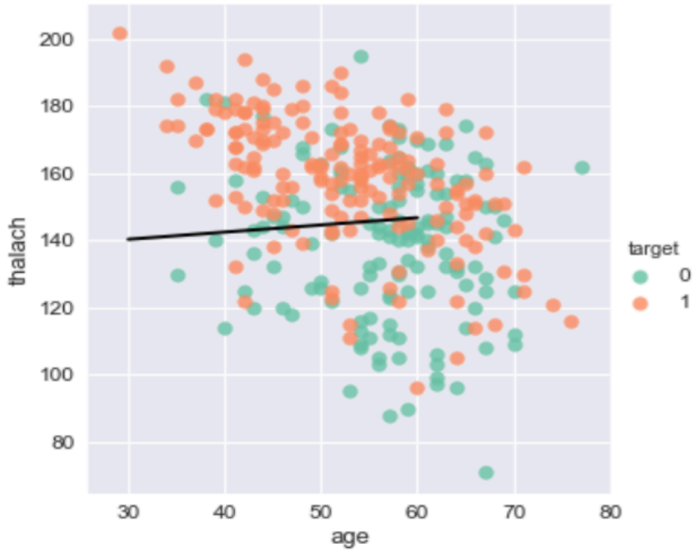

Fig. 3. SVM graphical classification with hyperplane on Maximum heart rate VS Age

Let's take another example with maximum heart rate vs age in Fig 3. From this hyperplane's division we noticed that there is high coagulation of data points in between 140-200 beats per minute. A maximum heart rate a person can have is 220 subtracted from the patient's age. This formula signifies that as the patient gets older, their maximum heart rate decreases. Additionally, we can observe from the graph that as the patient gets older, the maximum heart rate also goes down for a CAD condition to persist. This inference correlates with the formula and shows that as the patient's heart gets weaker, the probability of having CAD increase.

Additionally, these interpretations were also validated with the MEs and also with CPGs in the loop to ensure that there are no conflicts between the interpretations provided by the ME. This is because we as humans tend to have different opinions based on our tacit knowledge. Likewise, MEs may also have different treatment regimens for different symptoms. Thereby, having these universally accepted CPGs resolves such conflicts. Similar approach was followed for all the extracted subsets until a satisfactory number of syllogistic rules as prescribed by the MEs were interpreted.

5.4 Updating the CAD Knowledge Base with the Extracted Rules from SVM

The syllogistic rules extracted from the SVM algorithm's output are then used to build an early KB. In our work, we do not apply a complete KB with every representation of events. Instead, the KB here stores background information extracted from the SVM algorithm to provide an interpretation of its classification. This information is then used by XAI to generate inferences about a specific CAD condition in rule-based form.

The diagnosis of CAD is highly complex due to the additive and synergistic effects of the multifactorial underlying disease variables. Therefore, constructing a complete KB with all the possible inference rules is not tractable. We were, however, able to make incremental improvements to the KB by integrating interactions with MEs in the iterative development cycle and merging their interpretations with the SVM's output. With this approach, we were able to develop a XAI which was robust enough to diagnose CAD patients in a more precise manner than the original SVM's classifications. More on how the XAI was able to perform better than SVM will be explained in the testing and results section.

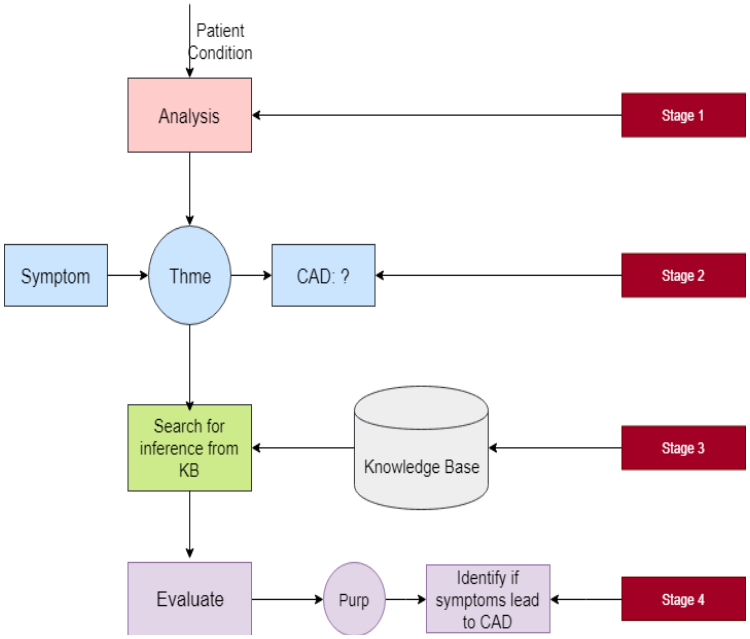

Fig. 4. Workflow for the development of an early KB

Fig. 4. adapted from the workflow of Sowa [11] gives an illustrative representation of the stages involved on how an XAI uses its KB to infer diagnosis for a particular patient condition. In stage 1, the anonymized patient's data is fed to the algorithm for analysis.

In stage 2, the algorithm identifies if the anonymized patient condition has symptoms of CAD by going through the syllogisms stored in the KB. Here the symbol “Thme” refers to the symptoms that match the theme of CAD. And finally, in stage 3, the XAI evaluates the inference from the KB to identify if the given patient condition may lead to CAD.

5.5 Validation by MEs in the loop for Early Experiment Output

In this section, we describe the iterative process of integrating MEs in-the-loop when building the components of our XAI. A total of five MEs from the Cardiovascular department of well acclaimed medical institutions were involved in this stage. Additionally, we also involved two MEs from different medical department to ensure that the XAI’s outcome can be understood and trusted by them who have less experience in this particular field.

First, we use SVM algorithm’s data mining technique to classify the data in the database in the form of graphs to determine if a given patient has CAD or not for a given physiological condition. Second, we use syllogistic approach to interpret rules from the SVM’s classification in graphs by analyzing how the hyperplane divides the data. Third, MEs were involved to evaluate the syllogistic rules interpreted from the SVM’s hyperplane division to identify any missing contextual information that may be vital for diagnosis based on their tacit knowledge. Contextual information here refers to the hidden information within the data that are usually excluded by the algorithm. Some examples of contextual information for a patient are medication record, innate physiological conditions and demographic data. These syllogistic rules were also then validated with CPGs because when it comes to human evaluators, there will always be biases in their evaluations. Hence, by the use of a universally accepted gold standard CPGs, we were able to nullify such dissentions. Lastly, these validated syllogistic rules with the necessary contextual information are then transformed into explainable outcomes by using a voice system that was implemented with "pytsx3" [39] to explain the diagnosis in a natural dialogue manner. These explanations were then validated with MEs in iterative stages to measure the level of goodness and satisfactoriness of each of these explanations. Fig 5 shows a pictorial representation of the workflow consisting of the different components developed for the XAI.

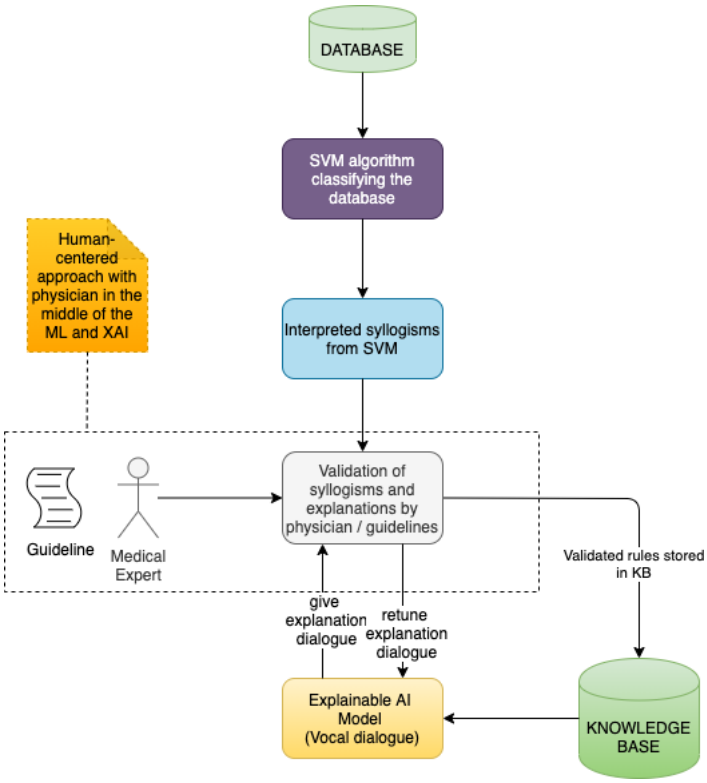

Fig. 5. XAI model components and workflow with ME in loop

6 Preliminary Results from Testing

In this section, we describe the results of applying our process experimentally to the UCI and Framingham datasets. We partitioned the repositories into training sets of (80%) and reserved 20% for testing and provide a representative excerpt of the syllogistic rules interpreted as our preliminary results. The symbol ∃ here represents the existence of a specific type of risk.

Blood pressure:
(SystolicBloodPressure > 140) ∧ (DiastolicBloodPressure > 90) →
(∃:x(HighBloodPressure))

Cholesterol:
(ch_LDL_lev(x) > 160) ∧ (age(x) > 21) ∧ (ch_HDL_lev(x) < 40 →
(∃:x(cholesterol_risk_high))

Blood sugar:
(blood_sugar(x) > 4) ∧ (blood_sugar(x) <= 5.4) → ¬(∃:x(DiabeticRisk))

Chest pain risk:
(anginal_cp) ∨ (atypical_anginal_cp) ∨ (non_anginal_cp) → (∃:x(chest_pain_risk)

Physiological triggering factors (KB inference Engine)
Rule 1: if there exists a high cholesterol risk, and diabetic risk, and high blood pressure risk, then CAD is present:
(∃:x(cholesterol_risk_high)) ∧ (∃:x(DiabeticRisk)) ∧ (∃:x(HighBloodPressure)) →
(∃:x(CoronaryArteryDisease))

Rule 2: if the patient's age is between 30 and 33 years old and diastolic blood pressure falls between: 60 and 110 mmHg, then CAD is present:
((age(x) > 30 ∧ age(x) < 33) ∧ (DiastolicBloodPressure(x) > 60 ∧ DiastolicBloodPressure(x) < 110)) →
(∃:x(CoronaryArteryDisease))

Rule 3: if there exists diabetic risk and a body-mass index (BMI) risk, then CAD is present:
(∃:x(DiabeticRisk)) ∧ (∃:x(BMI_risk)) → (∃:x(CoronaryArteryDisease))

The syllogistic rules generated above represent only a small portion of the knowledge background for the early KB. When a specific patient condition is given to the model, the XAI analyzes the condition based on stored syllogistic information in the KB to determine if a CAD condition can be warranted. We incrementally improved the explainability of this reasoning process during the study by verbalizing the rules via the voice system for MEs to understand and validate the syllogisms. Given below is an excerpt of the incremental development of the dialogues given by the XAI model.

For the explanation dialogues used in Phase I testing (13 April 2019), the XAI model only reported the risk factors in the rules without the incorporation of any contextual information:

A

"Male, and has cholesterol level risk, blood pressure risk, smoking risk. Prone to have CAD"

B

"Patient has normal glucose level. Prone to have CAD"

Explanation dialogues used in Phase II testing (7 May 2019), utilized a more natural language dialogue to represent the XAI model's output with the incorporation of the missing links from the contextual data:

A

"Since the patient is Male, and males have a higher chance of attaining CAD, and patient's test results show that they have high cholesterol, high blood pressure, and high diabetic risk. This infers that the patient has a high probability risk for CAD"

B

"The patient seems to be on glucose medication which normalizes blood glucose level. This infers why patient may have high potential risk for CAD though they have a normal glucose level"

From the above two iterative excerpts, we can infer how the inclusion of ME's contextual information within the explanation has drastically improved the level of reasoning given by the XAI. We will see more on how we measured the level of trust and understandability of the XAI's explanations in the next section.

6.1 Using Questioners to evaluate the Precision and Performance

To measure the trust and understandability of the XAI, we performed qualitative analysis with 5 MEs to evaluate the precision and performance of the XAI model using a human-machine work system analysis approach. The MEs used checklist-based questionnaires to evaluate the goodness and satisfactoriness of the explanations generated by the XAI model [28]. The evaluators must have extensive experience in the specific disease process supported by the XAI. With their expert understanding of CAD and general patient care we were able to evaluate and validate the XAI model's explanations.

The testing of performance and precision of the XAI model were conducted in two phases using Robert Hoffman's chart [28] as described below. The score attained in each phase represents the precision, performance and trustworthiness of the XAI model. Phase I represents preliminary stage testing where the MEs qualitatively assessed the dialogues given by the XAI model. While, the Phase II testing incorporated Phase I MEs feedback and recommendations to improve the model's explanation

Phase I.

Explanatory goodness and satisfaction test

For each of the statements listed below, on a scale of 1-5 (where 1 being very bad and 5 being very good), please indicate how well the AI was able to explain its diagnosis to you by circling one of the numbers from 1-5 below.

No.	Statement					
1.	I understand the explanation of how the AI works.	1	2	3	4	5
2.	The explanation given by the AI is satisfying.	1	2	3	4	5
3.	The explaining given by the AI has sufficient details.	1	2	3	4	5
4.	The explanation lets me judge when I can trust and not trust the AI.	1	2	3	4	5
5.	The explanation says how accurate the AI is.	1	2	3	4	5
6.	The explanation will aid me in diagnosis.	1	2	3	4	5

Fig. 6. Phase I. Iteration 1. XAI model testing results

An excerpt of comment given by ME for testing Phase I, iteration 1:

"*The explanation could be made more versatile. Will be more helpful if it could help me in identifying the special cases in CAD*"

Precision & performance score I.1:

$$\frac{\text{Attained Score from (Fig 5)}}{\text{Total score from (Fig 5)}} = \frac{5+4+4+3+5+5}{5+5+5+5+5+5} = \frac{26}{30} = 86\%$$

An excerpt of comment given by ME for testing Phase I, iteration 2:

"Needs more explanation and contextual information about patient is necessary to be convincing and trustworthy. Naming the factors that caused the patient to acquire CAD is not enough. With better explanation, it can perform better than scorecards"

Precision & performance score I.2: 83% (calculated using the same formula shown in Phase I.1)

Phase II.

In Phase II of the development, we modified the dialogues and added in more contextual information about the patient's conditions like medication intake, daily routine etc. as per the feedback given by the ME. With the incorporation of these information into the model, there was a drastic improvement in the diagnostic performance of the model and enabled phase 2 to score better on the precision and performance test.

No.	Statement					
1.	I understand the explanation of how the AI works.	1	2	3	4	5
2.	The explanation given by the AI is satisfying.	1	2	3	4	5
3.	The explaining given by the AI has sufficient details.	1	2	3	4	5
4.	The explanation lets me judge when I can trust and not trust the AI.	1	2	3	4	5
5.	The explanation says how accurate the AI is.	1	2	3	4	5
6.	The explanation will aid me in diagnosis.	1	2	3	4	5

Fig. 7. Phase II XAI model's precision and performance testing result

An excerpt from a comment given by a ME for testing Phase II:

> *"Explanations have been improved to be more trustworthy and precise. Additionally, we were able to get more insights through the explanations. Yet, fails to identify the anomaly cases in CAD."*

Precision & performance score II by ME 1: 93%
Precision & performance score II by ME 2: 95%
Precision & performance score II by ME 3: 89%
Precision & performance score II by ME 4: 96%
Precision & performance score II by ME 5: 92%
(calculated using the same formula shown in Phase I.1)

7 Results

Through the two iterative testing phases, we observed that the MEs rated Phase I XAI performance and precision lower than the Phase II version. This is likely due to the incompleteness and lack of narrative of the explanations given by the Phase I model. The Phase I model lacked the type of information the MEs felt necessary for understanding and trusting the XAI model. The second iterative phase, however, through the iterative involvement of MEs in every stage returned a much more comprehensive model explanation. This model included the qualitative and contextual information presented in a natural dialogue form which the MEs felt necessary for a more trustworthy and understandable interaction with the model's outputs. As seen from the score given by five MEs, the Phase II explanations were rated as more satisfactory by the MEs as it enabled them to develop complete and trustworthy mental models about the XAI's diagnosis compared to the simple explanations given by the Phase I model. The sub-

jective effectiveness of the XAI model for MEs is based on these ratings which demonstrates the degree to which a person could gain understanding from the explanations [40].

Additionally, an interesting finding we observed was that the final phase of our early XAI model scored 96% (according to the precision and performance chart given by Robert Hoffman, IHMC) from the MEs which is considerably better than the 86% scored by the SVM algorithm standing alone with a training and testing cohort of 80% and 20% from its respective records. Since the XAI was built through the interpretation of the individual feature sets on SVM algorithm as illustrated in section 5.2, it should score approximately the same accuracy rating as the SVM algorithm. However, with the incorporation of the human-centered approach and integrating contextual information (e.g., treatment regimens, demographic data, pharmaceutical history, etc.), the XAI was able to perform considerably better than the original SVM algorithm.

8 Conclusion and Future Work

Our preliminary work in exploring the notion of Explainable AI by the use of SVM and data mining techniques using syllogistic rules revealed two main findings. Firstly, we showed the importance of employing a human-centered approach iteratively when developing and interpreting the SVM output. By including MEs in the process, we found that data lacked the contextual, demographic and treatment information from the patients' records. The significance of contextual information can be illustrated by a finding whereby some of the interpretations from the SVM algorithm's graph gave contradicting results when validated against the CPGs. For example, blood glucose results from the Framingham dataset showed that patients with normal blood glucose are more prone to have CAD. However, the MEs inferred that blood glucose level is one of the most important factors contributing to CAD. In this case, the MEs surmised that most of these CAD patients were compliant with their diabetes treatment, and, thus, they maintained target therapeutic levels (i.e., normal range blood glucose) despite being diagnosed with CAD. Therefore, this proposed method will rely on MEs to discover the hidden rules and patterns within the data (i.e., specific types of CAD). While ML algorithms exhibit a hard to interpret behavior, the XAI provides an alternative pathway (through data visualization and natural language dialogue) to iteratively decode the inherent complexity of how risk factor variables interactions affect patient health, disease progression and outcomes. Although intelligent assistants often apply KBs, we employed a novel technique to improve the KB quality and precision using syllogistic rules to interpret SVM classification coupled with integral, iterative human-in-the-loop feedback during the development process.

Secondly, we observed that the XAI model's diagnostic capacity improved with more contextual information. With this, the MEs found the diagnosis performed by the XAI model to be more reliable than other conventional ML algorithms (Neural Net-

works, Decision Trees etc.) they have worked with before. Additionally, it also improves MEs assessment of the XAI model by enabling them to have a better understanding and trust with the model's chain of reasoning for a particular output (diagnoses). Moreover, the MEs trust in a the XAI model improved by employing a voice-based natural language dialogue system rather than a text-based output. Perhaps this may be due to the voice-based dialogue system providing the explanations of the XAI's decision-making process when proposing a diagnosis. The trust and confidence of the AI assistant is increased by providing more specific details and contextual information with explanations that employ natural language dialogue. Hence with this, the XAI was able to accomplish something a ML algorithm could not do on its own.

In the near future, we plan to explore the following two potential directions. First, we will investigate how to apply multiagent system techniques to automate the acquisition and integration of the multivariate contextual information needed to develop fully automated XAI systems. With this automation, the need for human technical support to integrate newly discovered contextual information would not be necessary anymore. Second, we would investigate the use of this automated XAI system to identify various forms of cardiovascular anomalies, which are generally deemed to be difficult for MEs to diagnose. Apart from that, these anomalies are also often misdiagnosed by ML algorithms due to the lack of requisite contextual information, which again strains the crucial need for the automation of XAI system.

With this proposed XAI model, we conclude this study with hopes to inspire and convince other researches to join and invest their experience and expertise in this emerging research field and transform the field of health care for multitudes around the world.

References

1. Core, M., Lane, C., Lent, M.V., Gomboc, D., Solomon, S., Rosenberg, M.: Building explainable artificial intelligence systems. In Proceedings of the 18th conference on Innovative applications of artificial intelligence - Volume 2 (IAAI'06). AAAI Press, 1766–1773 (2006).
2. Moore, J., Swartout, W.: Explanation in Expert Systems-A Survey. University of Southern California. Information Sciences Institute Tech Report ISI/RR-88-228, (1988).
3. Buchanan, B., Shortliffe, E.: Rule based expert systems. The MYCIN Experiments of the Stanford Heuristic Programming Project (1984).
4. Bishop, C.: Pattern recognition and machine learning. Springer Publication, (2006).
5. Liu, B., Hsu, W., Ma, Y.: Integrating classification and association rule mining, 98, 80-86. KDD Publication, (1998).
6. Maglogiannis, I., Loukis, E., Zafiropoulos, E., Stasis, A.: Support vectors machine-based identification of heart valve diseases using heart sounds. Computer methods and programs in biomedicine 95(1), 47-61, (2009).
7. Thurston, R., Matthews, K., Hernandez, J., Torre, F.D.L.: Improving the performance of physiologic hot flash measures with support vector machines. Psychophysiology, 46(2), 285-292, (2009).
8. Adrienne, C., Hongshik, A., Bhawna, H., Bruce, K., Everson L.A., Alan, B., Michael, L., Atul, K.: A decision support system to facilitate management of patients with acute gastrointestinal bleeding. Artificial intelligence in medicine 42(3), 247-259, (2008).
9. Rice, S., Nenadic, G., Stapley, B.: Mining protein function from text using term-based support vector machines. BMC bioinformatics, 6(1), S22, (2005).
10. Lamy, J.B, Sekar, B., Guezennec, G., Bouaud, J., Séroussi, B.: Explainable artificial intelligence for breast cancer: A visual case-based reasoning approach. Artificial intelligence in medicine 94, 42-53, (2019).
11. Sowa, J.: Knowledge representation: logical, philosophical, and computational foundations, PWS Publishing Company, (2000).
12. London, A.J.: Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability. Hastings Center Report, 49(1), 15-21, (2019).
13. Teach, R., Shortliffe, E.: An analysis of physician attitudes regarding computer-based clinical consultation systems. Computers and Biomedical Research 14(6), 542-558. DOI: 10.1016/0010-4809(81)90012-4, (1981).
14. American College of Cardiology Homepage, https://www.acc.org/guidelines#doctype=Guidelines, last accessed 2019/28/12.
15. Cohn, P.: Diagnosis and therapy of coronary artery disease: Springer Science & Business Media (2012).
16. Shavelle, D. Almanac 2015: coronary artery disease. Heart, 102(7), 492-499, (2016).
17. Abdullah, N., Clancey, W., Raj, A., Zain, A., Khalid, K.F., Ooi, A.: Application of a double loop learning approach for healthcare systems design in an emerging market. IEEE/ACM International Workshop on Soft-ware Engineering in Healthcare Systems (SEHS) 10-13. IEEE, (2018).
18. Babaoğlu, I., Fındık, O., Bayrak, M.: Effects of principle component analysis on assessment of coronary artery diseases using support vector machine. Expert Systems with Applications 37(3), 2182-2185, (2010).
19. Hongzong S., Tao W., Xiaojun Y., Huanxiang L., Zhide H., Mancang L. and BoTao, F.: Support vector machines classification for discriminating coronary heart disease pa-tients from non-coronary heart disease. West Indian Medical Journal 56(5), 451-457, (2007).

20. Xing, Y., Wang, J., Zhao, Z.: Combination data mining methods with new medical data to predicting outcome of coronary heart disease. International Conference on Convergence Information Technology, 868-872. ICCIT, (2007).
21. Zhu, Y., Wu, Z., Fang, Y.: Study on application of SVM in prediction of coronary heart disease. Journal of bio-medical engineering 30(6). 1180-1185, (2013).
22. Krittanawong, C., Zhang, H.J., Wang, Z., Aydar, M., Kitai, T.: Artificial intelligence in precision cardiovascular medicine. Journal of the American College of Cardiology 69(21), 2657-2664, (2017).
23. Çolak, M.C., Çolak, C., Kocatürk, H., Sagiroglu, S., Barutçu, I.: Predicting coronary artery disease using different artificial neural network models. AKD, 8(4), 249, (2008).
24. Guidi, G., Pettenati, M.C., Melillo, P., Iadanza, E.: A machine learning system to improve heart failure patient assistance. IEEE journal of biomedical and health informatics, 18(6), 1750-1756. IEEE, (2014).
25. Sokol, K., Flach, P.: Conversational Explanations of Machine Learning Predictions Through Class-contrastive Counterfactual Statements, 5785-5786 (2018). IJCAI Publication.
26. Miller, T.: Explanation in artificial intelligence: Insights from the social sciences. arXiv: 1706.07269 (2018).
27. Mittelstadt, B., Russell, C., & Wachter, S.: Explaining explanations in AI. In Proceedings of the conference on fairness, accountability, and transparency. ACM Library, (2019).
28. Hoffman, R., Mueller, S., Klein, G., Litman, J.: Metrics for explainable AI: Challenges and prospects. arXiv:1812.04608 (2018).
29. Nakatsu, R.: Explanatory power of intelligent systems. In Intelligent Decision-making Support Systems, 123-143. Springer (2006).
30. Lehtiranta, L., Junnonen, J.M., Kärnä, S., Pekuri, L.: The constructive research approach: Problem solving for com-plex projects. Designs, methods and practices for research of project management 95-106 (2015).
31. Trentesaux, D., Millot, P.: A human-centered design to break the myth of the "magic human" in intelligent manufacturing systems. *In Service orientation in holonic and multi-agent manufacturing*, 103-113. Springer (2016).
32. Kass, R., Finin, T.: The need for user models in generating expert system explanations: International Journal of Expert Systems 1(4) (1988).
33. Clark, P., & Niblett, T.: Induction in Noisy Domains. (pp. 11-30). EWSL, (1987).
34. Holzinger, A. Interactive machine learning for health informatics: when do we need the human-in-the-loop?. Brain Informatics, 3(2), 119-131, (2016).
35. Heart Disease UCI. Kaggle, https://www.kaggle.com/ronitf/heart-disease-uci, last accessed 2019/8/12.
36. Framingham Heart study dataset. Kaggle, https://www.kaggle.com/amanajmera1/framingham-heart-study-dataset, last accessed 2019/8/12.
37. Scikit learn Homepage, https://scikit-learn.org/stable/auto_examples/feature_selection/plot_feature_selection.html, last accessed 2019/28/12.
38. Benesty, J., Chen, J., Huang, Y., & Cohen, I.: Pearson correlation coefficient. In Noise reduction in speech processing (pp. 1-4). Springer, Berlin, Heidelberg, (2009).
39. PyPI Homepage, https://pypi.org/project/pyttsx3/, last accessed 2019/28/12.
40. Keil, F.: Explanation and understanding. Annu. Rev. Psychol., 57, 227-254 (2006).