Keywords

The accurate prediction of heart disease is essential to treat the patient efficiently.

Feature fusion can provide rich healthcare data for heart disease diagnosis.

Existing systems cannot handle high-dimensional datasets related to heart disease.

Traditional methods are failed to extract valuable features for heart treatment.

A smart system is proposed to enhance the accuracy of heart disease diagnosis.
A Smart Healthcare Monitoring System for Heart Disease Prediction Based On Ensemble Deep Learning and Feature Fusion

Farman Ali¹, Shaker El-Sappagh², S.M. Riazul Islam³, Daehan Kwak⁴, Amjad Ali⁵*, Muhammad Imran⁶, Kyung-Sup Kwak⁷

¹Department of Software, Sejong University, Seoul, South Korea
²Department of Information and Communication Engineering, Inha University, Incheon, South Korea
³Department of Computer Science and Engineering, Sejong University, Seoul, South Korea
⁴Department of Computer Science, Kean University, Union, USA
⁵Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Lahore, Pakistan
⁶College of Applied Computer Science, King Saud University, Riyadh, Saudi Arabia

E-mail: farmankanju@gmail.com; farmankanju@sejong.ac.kr, shaker_elsapagh@yahoo.com, riaz@sejong.ac.kr, dkwak@kean.edu, dr.m.imran@ieee.org

*Corresponding author: K.S. Kwak (kskwak@inha.ac.kr); Amjad Ali (amjad.ali@cuilahore.edu.pk)

Abstract: The accurate prediction of heart disease is essential to efficiently treating cardiac patients before a heart attack occurs. This goal can be achieved using an optimal machine learning model with rich healthcare data on heart diseases. Various systems based on machine learning have been presented recently to predict and diagnose heart disease. However, these systems cannot handle high-dimensional datasets due to the lack of a smart framework that can use different sources of data for heart disease prediction. In addition, the existing systems utilize conventional techniques to select features from a dataset and compute a general weight for them based on their significance. These methods have also failed to enhance the performance of heart disease diagnosis. In this paper, a smart healthcare system is proposed for heart disease prediction using ensemble deep learning and feature fusion approaches. First, the feature fusion method combines the extracted features from both sensor data and electronic medical records to generate valuable healthcare data. Second, the information gain technique eliminates irrelevant and redundant features, and selects the important ones, which decreases the computational burden and enhances the system performance. In addition, the conditional probability approach computes a specific feature weight for each class, which further improves system performance. Finally, the ensemble deep learning model is trained for heart disease prediction. The proposed system is evaluated with heart disease data and compared with traditional classifiers based on feature fusion, feature selection, and weighting techniques. The proposed system obtains accuracy of 98.5%, which is higher than existing systems. This result shows that our system is more effective for the prediction of heart disease, in comparison to other state-of-the-art methods.

Keywords: Feature extraction, feature fusion, heart disease prediction, deep learning, ontology

1. Introduction

The automatic prediction of heart disease is one of the most essential and difficult health problems in the real world. Heart disease affects the functionality of blood vessels, and causes coronary artery infections that weaken the body of the patient, especially adults and old people. The World Health Organization (WHO) has determined that more than 18 million deaths occur every year in the world due to cardiovascular diseases [1]. Furthermore, the United States of America spends $1 billion on treatments for heart disease per day [19]. The main cause of death in America is heart diseases such as stroke, heart attack, and hypertension. Therefore, early prediction of heart disease is very important to effectively treat cardiac patients before a heart attack or stroke can occur.

Cardiovascular diseases can be identified by conducting medical tests and using wearable sensors. However, extracting valuable risk factors for heart disease from electronic medical tests is difficult as physicians try to quickly and accurately diagnose patients. These electronic medical records (EMRs) are unstructured and increasing in size constantly due to daily medical tests. Currently, wearable sensors are also utilized to continuously monitor the patient’s body internally and externally to detect heart disease. However, wearable sensor data for heart disease prediction are corrupted by signal artifacts such as missing values and noise, which decreases system performance and generates inaccurate results. In the first place, utilizing both wearable sensors and EMRs together is a significant and difficult task when monitoring cardiac patients. In the second place, extracting relevant and meaningful features from the data is a challenging task for heart disease prediction. Therefore, there is a need for an intelligent system that can automatically fuse the extracted information from both sensor data and EMRs, and that can analyze the extracted data to identify the hidden symptoms of heart problems and predict heart disease before a heart attack occurs.
Currently, several systems have been proposed to predict and diagnose cardiovascular disease by utilizing data mining techniques and hybrid models, as discussed in the related work section below. A data mining technique extracts risk factors from unstructured textual data [23,36]. In addition, a hybrid model is the integration of two different methods that work better together than any one individual method. These hybrid models basically contain two main phases. In the first phase, feature selection or a feature weighting approach is applied to select a feature’s subset or identify the feature weight. In the second phase, the feature’s subset or weight is utilized as input for classifiers to predict heart disease [1,33,40,42,48]. However, a collected dataset on heart disease contains relevant features along with many redundant and irrelevant features. Both redundant and irrelevant features create confusion and noise over the definition of a target class. The handling of these features is not only time-consuming but also affects the accuracy of classification. Furthermore, the existing systems for heart disease diagnoses are based on general feature-weighting methods. These methods allocate the same weight to each feature for all classes. However, they utilize uncertain combination operations, which may set the feature importance for differentiation. They increase the mean square error (MSE) and decrease the accuracy of the prediction model due to a lack of theoretical supports. Therefore, it is necessary to remove useless features and assign a specific weight to the features before applying classification models.

In this paper, a novel, smart healthcare monitoring system is proposed for heart disease prediction using ensemble deep learning and feature fusion. First, the data of the heart patient are efficiently collected using two different methods: wearable sensors and electronic medical tests. Second, valuable Framingham risk factors (FRFs) are extracted from the EMRs. The feature fusion approach is then used to combine FRFs and sensor data and generate rich healthcare data on heart disease. Next, information gain (IG) and conditional probability are employed to reduce the set of features and compute a specific weight for heart disease features. Finally, an ensemble deep learning classifier is trained to predict heart disease in patients. The main contributions of this paper are as follows.

- For heart disease prediction, a novel, information framework is proposed to process both sensor data and medical records using ensemble deep learning and a feature fusion technique.
- An FRF extraction module is developed to detect and extract low-dimensional risk factors of heart disease from unstructured EMRs. In addition, feature-level fusion is conducted to combine both sensor data and extracted FRFs for heart disease prediction.
- The IG approach is proposed for feature selection that reduces noise by eliminating irrelevant features, decreasing the complexity and dimensionality of the dataset. In addition, a conditional probability approach is utilized that computes a specific feature weight for each class in order to increase the prediction accuracy.
- The boosting algorithm called 
  LogitBoost is utilized as a meta-learning classifier that reduces bias and variance, and boosts the deep learning model to achieve high accuracy. Furthermore, an ontology based on Semantic Web Rule Language (SWRL) rules is developed that automatically recommends a dietary plan or activities for the heart patient.
- The proposed system enhances the performance and obtains a higher accuracy of 98.5% in heart disease prediction, compared to other state-of-the-art methods.

The rest of this paper is structured as follows. Section 2 presents a brief review of heart disease diagnostic systems, along with feature extraction and feature identification. Section 3 introduces the proposed system framework, and discusses data preprocessing, ensemble deep learning, and the ontology. Section 4 presents the experimental results. Finally, Section 5 offers conclusions about the proposed work and directions for future study.

2. Related work

Wearable sensors and EMRs play a key role in a healthcare monitoring system for heart disease patients. However, feature extraction from sensor data and EMRs, and then fusing them in order to transform them into structured data, is a challenging task. In addition, feature selection from structured data and then assigning valuable weights to them is another challenge for machine learning (ML)-based systems. Therefore, this section first looks at wearable sensor–based heart disease diagnostic systems, and then focuses on information extraction from textual data, and then feature fusion. This section also presents a brief review of the identification of feature importance in healthcare data in the domain of heart disease prediction.

2.1 Wearable sensor–based heart disease diagnostic systems

In recent years, various systems have been proposed using wearable sensors to improve the process of heart disease prediction [34,35]. Al-Makhadmeh and Tolba presented a wearable medical devices–based system that collects details about cardiac patients before and after a heart attack [3]. They transmitted the collected data to the healthcare system, and then, utilized feature extraction techniques and a deep learning model for valuable feature extraction and correct classification. However, this system has limitations in its efficient feature extraction and feature weighting approach. In addition, they used 23 attributes for training,
which increases system complexity and dimensionality. A system called HealthFog was presented to automatically treat heart patients using Internet of Things (IoT) devices and a deep learning model [53]. The main aim of this system is to competently manage the heart patient data coming from the IoT devices. Furthermore, another novel framework for a decision support system was presented for disease detection in patients [57]. This system integrates collected medical data and data from wearable medical devices. In addition, a multi-tier structure approach along with deep learning is utilized for disease diagnoses. A hybrid recommender system based on IoT devices was presented to diagnose heart disease [20]. That system recommends a dietary plan and physical activities according to the cardiac patient’s condition. However, this recommender system is based on simple rules that need semantic information in order to extract more sensitive information about the patient for correct recommendations. An IoT-based disease prediction system was presented that uses a fuzzy neural classifier [29]. The system provides a novel framework of a mobile healthcare monitoring system for serious disease diagnosis. Furthermore, a three-tier framework with an ML model is presented to store and process the data from a heart patient using wearable devices [28]. The first tier collects physiological data from the sensors; the second tier stores the healthcare data in the cloud, and the third tier predicts heart disease using logistic regression. However, this system conducts experiments using seven features, which is not enough to correctly identify the health condition of heart patients. In addition, data preprocessing steps are missing, and therefore, it is difficult to understand the data preparation for ML classifiers.

2.2 Information extraction and fusion

The extraction of valuable features from medical textual data and the fusion of sensor data is another key problem for a diagnostic system of heart diseases. The techniques of feature extraction retrieve meaningful information from healthcare big data. The procedure of data fusion merges different sources of data for relevant and valuable data generation. Recently, various models have been proposed to extract features from healthcare textual data [16,18,49] and fuse sensor data with the other data [13,39,44]. A real-time system based on ML classifiers was presented to predict heart disease [1]. This system employs two different approaches to feature extraction (univariate and relief) that extract an important feature’s healthcare dataset. A novel and robust structure was presented to mine EMRs for heart disease diagnoses. This system uses a word embedding model for feature identification and long short-term memory (LSTM) for heart failure prediction. Furthermore, another system used the results of echocardiography to predict the mortality rate of cardiac patients in the hospital [31]. This system utilizes text mining techniques to extract features, and then applies a deep learning model for mortality rate predictions. A novel text mining approach is presented to extract information from EMRs [24]. The authors of this system utilize a rules-based engine that automates information extraction and decision-making tasks. However, many medical records contain unstructured data, and it is difficult to handle these data with a rules-based engine. Furthermore, a heart disease risk score is identified using unstructured EMRs of patients [23]. In this system, the authors use text mining techniques to extract heart-critical factors from unstructured EMRs and calculate a heart disease risk score in diabetes patients.

Wearable sensors are utilized to detection the emotions of users for music recommendations [12]. In this system, the authors utilized feature-level fusion that extracts features from each sensor independently, and then combines the extracted features for emotion detection. Furthermore, in yet another system, the authors utilized both environmental and wearable body sensors for emotion detection [27]. They applied all three levels of fusion (data, feature, and decision) in order to investigate the impact of the environment on human health. A recurrent convolutional neural network (RCNN)-based system was presented to predict disease risk [19]. This system extracts features from the patient’s structured and unstructured data, and fuses them using a deep belief network to improve the accuracy of the RCNN.

2.3 Feature identification models

In the literature, there have been various studies about feature selection and feature significance identification in the domain of disease diagnosis in general or heart disease prediction specifically. A framework of multi-sensor data fusion was proposed using fog computing with kernel random forest for the classification of heart diseases [39]. In this system, a correlation technique is utilized to select the most relevant features, and then, it applies both data-level and feature-level fusion to combine the selected features. This system achieved accuracy of 98% based on eight selected features from activity sensors. However, these features cannot cover all the risk factors of heart patients. Fuzzy logic along with a genetic algorithm was proposed to deal with the uncertainty of healthcare data [43]. This system uses wavelet transformation for feature selection that reduced the computational burden owing to the small number of selected features. A hybrid system based on random forest and a linear model was introduced to enhance the performance of heart disease diagnosis [37]. This system applied various feature selection and classification methods, and achieved accuracy of 88.7%. A recommendation system based on fast Fourier transformation was proposed for heart disease prediction [58]. This system selects features in a time series, and makes a bond of features for an ML model in order to effectively detect disease with an appropriate recommendation. Similarly, a feature reduction system based on
chaos firefly and rough sets was presented for cardiac disease classification [33]. This approach handles the uncertainty and high-dimensionality of datasets, and achieved an accuracy of 88% via type-2 fuzzy logic.

In the field of healthcare data analysis, various feature weighting approaches can be utilized to improve the performance of classifiers [21,22,25,56]. A diagnostic system based on correlation coefficient and weighted least squares techniques was presented to identify heart failure in patients. These approaches select critical features that help the heart disease prediction system to improve accuracy by using fuzzy logic and genetic algorithms [45]. A decision support system based on a fuzzy analytic hierarchy process (AHP) and an artificial neural network (ANN) was presented to predict heart diseases in patients [48]. This system first utilizes fuzzy AHP to identify the significance of features, and assigns global weights to the features. The weighted features are then employed to train the ANN for heart disease prediction. However, the general feature-weighting technique may not identify the correct feature significance for all classes. In addition, it may set the feature importance for differentiation by using an uncertain combination operation, which decreases the accuracy of prediction models and increases the error rate.

![Fig. 1. The structure of the smart healthcare monitoring system for heart disease prediction.](image)

3. **Smart healthcare monitoring system for heart disease prediction**

This section discusses the structure of the proposed smart healthcare monitoring system (SHMS) in detail. First, the general structure of the SHMS is described. The structure is divided into different layers in order to thoroughly describe the information groundwork of each stage in the proposed system. Finally, the structure of the ensemble deep learning model and ontology are presented, which are employed by the SHMS to predict heart disease in patients and recommend dietary plans and activities.

The framework of the proposed SHMS is in Fig. 1. The SHMS has two main data sources. The first source, presented on the left, is the wireless body sensor network (WBSN). The other data source is EMRs. The system uses a WBSN based on medical sensors to collect internal and external physiological data, such as an electrocardiogram (ECG), an electroencephalogram (EEG), an electromyogram (EMG), the heart rate, blood pressure (BP), position, activities, respiration rate, and blood sugar, oxygen saturation, and cholesterol levels of the patient for daily health monitoring (Task 1 in Fig. 1). EMRs provide patient observation reports, medical history, smoking history, diabetes history, and detailed clinical examinations (lab reports) (Task 2 in Fig. 1). After sensing the data from a heart patient, the proposed system transfers the data to the associated gateway devices. Various devices can be used as a gateway to collect and forward the sensed data for further processing. In this system, the physiological data are transmitted via Bluetooth and WiFi devices (Task 3 in Fig. 1). Both the sensed data and the EMRs are securely stored in...
a database considered healthcare big data (Task 4 in Fig. 1). The purpose of the SHMS is to predict disease risk in patients based on the collected data. Therefore, we utilize health condition prediction and a disease diagnosis engine to predict heart disease based on the collected data, both structured and unstructured, (Task 5 in Fig. 1). This engine consists of four main steps: 1) data fusion, 2) preprocessing, 3) ensemble deep learning–based heart disease prediction, and 4) ontology-based recommendations. In the first step, the extracted features from both structured and unstructured data are fused using the proposed fusion scheme. The next step is where the data are preprocessed using data mining techniques. This step includes data filtering, normalization, valuable feature selection using data mining techniques, and feature weighting using conditional probability. In the third step, the preprocessed data are passed to a deep learning classifier trained on a heart disease dataset for the final prediction of heart diseases. In the fourth step, an ontology is used to recommend a dietary plan or activities per the patient’s health condition (Task 6 in Fig. 1). As shown in Fig. 2, the workflow of the proposed SHMS uses four sequential layers: data collection, data fusion and feature extraction, data preprocessing, and disease prediction and recommendation. These layers are briefly discussed in the following subsections.

3.1 Data collection layer

The proposed SHMS considers two different types of data for heart disease prediction: patient physiological data and EMRs, as shown in Fig. 2. The patient physiological data are collected with the help of wearable sensors. Two types of sensor are utilized to collect the physiological data: medical and activity sensors. Medical sensors include a respiration rate sensor, an oxygen saturation sensor, a blood pressure sensor, a cholesterol level sensor, a glucose level sensor, a temperature sensor, an EMG sensor,
an EEG sensor, and an ECG sensor. These are connected to the patient’s body to collect physiological data without interruption. In addition, a wearable watch is utilized to record physical activities and the heart rate. The patient’s physical activities provide valuable information for the prediction of heart disease. This model uses network devices to transfer physiological data to a personal database for the analysis of heart conditions. In the data collection layer, identities are assigned to all sensor data (e.g., S is EEG sensor data), as shown in Fig. 2. The system considers these identities as features during processing. In addition, the sensor data are presented in columns, along with identities and numerical values, for further processing. Moreover, unstructured EMRs of the patient are collected to identify risk factors for heart disease prediction. EMRs include lab reports, a medical history, questions and observations, allergies and medications, and personal statistics. The EMRs can be analyzed to extract FRFs that can help provide valuable information for disease prediction. The FRFs include age, gender, the presence of diabetes, smoking history, cholesterol levels, BP, heart rate, ejection fraction (EF), body mass index (BMI), level of obesity (if present), and diet. However, the volume of EMR data is usually large, and each record comprises data with distributed variables of high dimensionality. Therefore, we use a text mining approach to effectively extract FRFs from unstructured textual data.

![Fig. 3. Extraction of Framingham risk factors from an unstructured EMR.](image)

### 3.2 Feature extraction and data fusion layer

In this section, we extensively discuss the procedure of feature extraction from unstructured data, and the data’s transformation into a structured format. First, we introduce the extraction of FRFs. Then, we present the detailed procedure of the data fusion scheme for heart disease prediction.

#### 3.2.1 Framingham risk factor extraction

We extract information from unstructured EMRs through a separate module called the FRF extraction module, as shown in Fig. 3. The unstructured EMR is assigned to the FRF extraction module to identify and extract risk factors related to heart disease. This module is divided into two submodules: structured FRF extraction, and categorical FRF extraction. In structured FRF extraction, the system extracts factors where values are already available in a structured form. For example, the system extracts age, height, heart rate, EF, gender, BP, and BMI along with values from structured fields. In categorical FRF extraction, the system extracts risk factors where values appear in different classes (e.g., diabetes history, smoking history, and family history of coronary artery disease [CAD]). These factors are extracted using text mining techniques and a rules-based engine. The text mining technique comprises three main steps. In the first step, morphological and lemmatization algorithms are applied to all unstructured data to identify the lemma of each word [10]. The next step is tokenization, which separates complex, unstructured text into small chunks [9]. In the third step, N-gram approaches are used to extract risk factors. Usually, the value of a risk factor appears in two or three adjacent words. Therefore, bigrams and trigrams are utilized to extract two and three adjacent factor words, respectively. Rules are generated to identify abbreviations and particular features that indicate gender, age, cholesterol levels, and ejection fraction. For example, EMRs may contain a phrase with 48y; the value 48 is extracted as the age of the patient based on a set of abbreviations. Categorical FRF extraction obtains information about the patient’s diabetes history, family CAD history, and smoking history based on a set of rules. A separate bag of words model is built based on diabetes, CAD, and smoking terms to identify the abovementioned histories. This module filters the records with the help of the rules-based engine to remove records.
that do not contain information about heart disease, diabetes, and/or smoking. In addition, it removes risk factors without a numerical value. The result of this module is FRFs with values or categories in a structured format.

3.2.2 Feature fusion layer

In this section, we discuss the fusion of sensor data with extracted FRFs from textual data, as shown in Fig. 4. Fusion is the procedure of merging different sources of data to generate more valuable and relevant data for classification [12,27]. There are three levels of fusion: data level, feature level, and decision level [15,52,55]. Data-level fusion combines different data from heterogeneous sources that match each other. Data fusion can be categorized as feature level and decision level. At the feature level, features are retrieved from different datasets individually and then merged to create the best set of features for prediction. At the decision level, the decisions of various processes are considered in order to enhance the accuracy of the system. Usually, data-level fusion contains a large amount of redundant data, and therefore, it is not desirable. In contrast, the feature level contains sufficient information to identify the disease risk. In the proposed system, both data-level and feature-level fusion are performed. The workflow for data fusion is shown in Fig. 4. First, the patient’s physiological data are collected using sensors, and FRFs are extracted from EMRs, as discussed in sections 3.1 and 3.1.1, respectively. Then, sensor data are fused with other extracted FRFs, such as age, gender, diabetes history, and smoking history. Finally, the sensor data and extracted features are converted into comma separated value (CSV) files for easy parsing. In this way, the system finds the best combination of features related to cardiac disease. The main goal of the proposed system is to use sensor data with relevant and lower-dimensional features extracted from EMRs for heart disease prediction. However, the sensor data may have missing values, and the extracted features may comprise irrelevant information, which decreases the prediction accuracy and increases feature dimensionality. In addition, these also increase the memory needs and complexity for classification. Therefore, data preprocessing is applied before the actual processing, which improves the quality of the data, and saves time and memory.

![Feature fusion-based structured data for heart disease prediction](image)

3.3 Data preprocessing

Data preprocessing is the most essential step before applying ML algorithms. It is not possible to utilize real-world data directly in the prediction task, as it tends to be noisy, incomplete, and inconsistent. Therefore, a preprocessing step is applied to represent the data effectively for heart disease prediction. Data preprocessing includes missing-data filtering, normalization, feature selection, and feature weighting.
3.3.1 Missing-data filtering

Data collected using wearable sensors, and data extracted from EMRs contain useless and incorrect information. Wearable-sensor data for heart disease prediction are damaged by signal artifacts, such as missing values and noise, which decreases the prediction accuracy or generates an inaccurate result. In addition, data extracted from EMRs are presumed to be missing when they do not contain at least a single value. In the extracted data, information may be missed due to the failure of text mining techniques to recognize an FRF value, or FRF values may not have been recorded. We filter the data by using a well-known filtering approach called Kalman filtering [17]. This filter cleans the data by removing noise, duplicate records, and inconsistencies. In addition, we also utilize two unsupervised filters in the data filtering stage: RemoveUseless and ReplaceMissing Values [45]. The first filter removes useless attributes with a maximum variance of 90%. The second filter replaces all missing values in the structured dataset with mean and median values from existing data by using the following equation:

\[
\overline{X} = \frac{1}{N} \sum_{i=1}^{N} X_i
\]

where \( \overline{X} \) represents features such that \( X = \{ \text{“age”}, \text{“chol”}, \text{“sex”}, \text{“heart rate”}, \ldots \text{“CAD history”}, \text{“smoking history”} \} \), a category level such that \( \mathbb{P}(0, 1, 2) \), the pattern number, the \( \mathbb{P} \) pattern of feature \( X \) under the \( \mathbb{P} \) category, and the mean of feature \( X \) under category \( \mathbb{P} \) respectively. In this work, \( \overline{X} \) replaces the missing values of feature \( X \) within category \( \mathbb{P} \). In addition, we also utilized extracted FRF values to overcome the limitations of wearable sensor–based generated data; for example, we replace a missing value with a current FRF attribute value in the dataset.

3.3.2 Normalization

The heart disease dataset \( \mathcal{X} \) contains a number of features, and every feature includes different numerical values, which increases the difficulties during the computation process. Therefore, a normalization technique is used to normalize dataset \( \mathcal{X} \) in the range between zero and 1, as well as to decrease the numerical complexity during the computational process of heart disease prediction. Various methods can be utilized for data normalization. In the proposed system, the well-known min-max normalization method is used [33,41]. This method plots a numerical value, \( \mathcal{X} \), of original dataset \( \mathcal{X} \) into \( \mathcal{X} \) within the interval \([0, 1]\) by using the following equation:

\[
\mathcal{X} = \frac{\mathcal{X} - \min \mathcal{X}}{\max \mathcal{X} - \min \mathcal{X}} 
\]

Here, \( \mathcal{X} \), \( \mathcal{X} \), \( \mathcal{X} \), and \( \mathcal{X} \) are the normalized data value, the original data value, the minimum data value, and the maximum data value, respectively, in the entire dataset, while \( \mathcal{X} \) and \( \mathcal{X} \) indicate the range of the converted dataset. We use \( \mathcal{X} \) and \( \mathcal{X} \) to select features with a maximum value. Using this method, all the features’ values lie within the interval \([0, 1]\).

3.3.3 Information gain-based feature selection

Patient records normally consist of many irrelevant features that reduce the accuracy of a prediction. However, the extraction of meaningful information from medical records, reducing noise by eliminating irrelevant features, and accurate prediction of heart disease with a limited number of features, are all challenging tasks. Before applying any prediction model, it is essential to remove noisy data, select useful features that help achieve accurate results, and reduce the complexity and dimensionality of the dataset. Therefore, feature selection is an important step that improves the clarity of the data and decreases the training time of ensemble deep learning models. Different methods of feature selection are used in healthcare datasets, such as sequential forward selection, weighted least squares, rough sets, and univariate feature selection [1,20,39,45,47]. We utilize the information gain method, which affects the prediction results by eliminating noisy features. There are 27 attributes in structured dataset \( \mathcal{X} \) for heart disease prediction. Only a few of them are useful for classifying the disease into one of the given categories. The system can learn about specific problems based on the importance of features in the dataset. In the proposed system, by using IG, features are selected that measure importance according to the classification task. The proposed system uses entropy to measure system uncertainty. It finds the difference between prior entropy and post entropy of two given distinct variables, A and B, as shown in the following equations:

\[
\Delta H(A,B) = H(A) - H(A|B) 
\]
where $A$ and $B$ are discrete random variables, and the prior entropy of feature $A$ can be computed using equation 4:

$$ H(A) = - \sum_{b \in B} P(b) \sum_{a \in A} P(a | b) \log P(a | b) $$

where $P(b)$ denotes the prior probability for $b$. After being given post entropy $B$, the conditional entropy of $A$ can be calculated using equation 5:

$$ H(A | B) = - \sum_{b \in B} P(b) \sum_{a \in A} P(a | b) \log P(a | b) $$

The IG can be measured by putting equations 4 and 5 into equation 3, as shown in equation 6:

$$ IG(A, B) = H(A) - H(A | B) $$

The proposed system estimates the importance of each feature to the task of heart disease prediction using equation 6. This system deletes the least important features after measuring the IG for each feature. It reduces features by eliminating one feature at a time until the decline in performance ends.

### 3.3.4 Conditional probability–based feature weighting

Feature weighting is a method of assigning a weight to each feature based on its significance. It is different from the feature selection method, which completely removes redundant and irrelevant features from the training dataset. Various methods have been utilized for feature weighting, such as the analytic hierarchy process and rules-based weighting [25,42,48]. These methods assign the same weight to each feature for all classes, which is called the general weight of the feature [21,22,56]. In general feature weighting, the feature significance is first determined for each class. The maximization and summation functions are then used to identify the overall feature significance for all classes. However, there are some limitations in the existing approaches. One of the major issues is that they are based on an uncertain combination operation, which may set the feature importance for differentiation. In addition, these existing approaches increase the MSE and decrease the accuracy of the prediction model due to a lack of theoretical supports. In this paper, the probabilistic approach is utilized to achieve a specific feature weight for each class. It is a more refined approach than general weighting. It is important that the feature weight should be specific to the class for the prediction task. This improves the performance of the prediction model by learning the different significance of the feature for each class. The matrix representations of general and specific feature weighting are shown in Table 1 and Table 2, respectively. In Table 1, each feature shares the same weight for all classes, whereas in Table 2, each feature shares a different weight for all classes. Let $P_j$, $W_j$, $b_j$, $\tilde{b}_j$ be $J$-feature variables. $b_j$ is the instance signified by feature vector $\mathbf{d}_j$, $W_j$ be the weight of $P_j$, and $\tilde{b}_j$ represents the value of $P_j$. The specific feature weight for instance $b_j$ can be computing using the following equation:

$$ b_j \cdot W_j \cdot \tilde{b}_j $$

where $b_j$ and $\tilde{b}_j$ denote the class variable value and the particular weight of feature value $\tilde{b}_j$ for class $c$, respectively. The value of $b_j \cdot W_j$ is related to feature value $\tilde{b}_j$, where a different weight is assigned to each feature value. The range of $b_j \cdot W_j$ is from zero to 1, which determines the importance of feature value $\tilde{b}_j$ for the heart disease prediction task. This approach is useful for converting the values of the heart disease features into a more manageable form. The main aim of obtaining these weights for feature values is to use them as initial weights with a deep learning model to obtain better prediction results.

**Table 1: Representation of general weights for each feature.**

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Age</th>
<th>Gen</th>
<th>Chol</th>
<th>Smoke</th>
<th>CAD history</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_1$</td>
<td>$W_1$</td>
<td>$W_2$</td>
<td>$W_3$</td>
<td>$W_4$</td>
<td>$W_n$</td>
</tr>
<tr>
<td>$P_2$</td>
<td>$W_1$</td>
<td>$W_2$</td>
<td>$W_3$</td>
<td>$W_4$</td>
<td>$W_n$</td>
</tr>
<tr>
<td>$P_3$</td>
<td>$W_1$</td>
<td>$W_2$</td>
<td>$W_3$</td>
<td>$W_4$</td>
<td>$W_n$</td>
</tr>
<tr>
<td>$P_4$</td>
<td>$W_1$</td>
<td>$W_2$</td>
<td>$W_3$</td>
<td>$W_4$</td>
<td>$W_n$</td>
</tr>
</tbody>
</table>

**Table 2: Representation of specific feature weights for all instances.**
<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Age</th>
<th>Gen</th>
<th>Chol</th>
<th>Smoke</th>
<th>...</th>
<th>CAD history</th>
</tr>
</thead>
<tbody>
<tr>
<td>P₁</td>
<td>W₁₁</td>
<td>W₂₁</td>
<td>W₃₁</td>
<td>W₄₁</td>
<td>...</td>
<td>Wₙ₁</td>
</tr>
<tr>
<td>P₂</td>
<td>W₁₂</td>
<td>W₂₂</td>
<td>W₃₂</td>
<td>W₄₂</td>
<td>...</td>
<td>Wₙ₂</td>
</tr>
<tr>
<td>.</td>
<td>W₁₃</td>
<td>W₂₃</td>
<td>W₃₃</td>
<td>W₄₃</td>
<td>...</td>
<td>Wₙ₃</td>
</tr>
<tr>
<td>Pₙ</td>
<td>W₁₄</td>
<td>W₂₄</td>
<td>W₃₄</td>
<td>W₄₄</td>
<td>...</td>
<td>Wₙ₄</td>
</tr>
</tbody>
</table>

3.4 Disease prediction and recommendation layer

In this section, we first present the ensemble deep learning model for heart disease prediction, and we then look at the ontology-based recommendation system.

3.4.1 Ensemble deep learning model for heart disease prediction

The fourth layer of the proposed SHMS contains an ensemble deep learning model where the framework is conceptualized as presented in Fig. 5. This model is a feed-forward network that utilizes back propagation techniques and gradient algorithms for binary classification of heart diseases. We first use different numbers of features from the Cleveland dataset to train the model, and we then utilized it to predict the results of real-time input data as presented in Fig. 5. We utilized the boosting algorithm called LogitBoost as a meta-learning classifier, which boosts the deep learning model to achieve high accuracy [26]. Based on our experiments and on the literature review, the boosting algorithm is more appropriate for handling noisy data, as compared to the AdaBoost algorithm. It is designed for reducing bias and variance to enhance classification performance.

![Fig. 5. Deep learning model for heart disease prediction.](image)

The ensemble deep learning model is composed of five layers: the input layer, three hidden layers, and the output layer. The input layer comprises 16 nodes according to the number of features in the dataset. It is essential to evaluate the hidden layers of the neural model by using different numbers of nodes. In the proposed work, we consider a fully connected hidden layer with 20 nodes, which produces consistent performance. The attributes are assigned to a deep neural network model using the input layer. These attributes are then passed to the three hidden layers by multiplying their corresponding weights with values. The weighted sum is calculated, and bias is added in order to process the input data in the hidden layer nodes, as shown in equation 8:

\[ \sum_{\lambda} \theta_{\lambda} \sum_{j} W_{j} \cdot b_{j1} + \beta_{j} \]

where \( \theta_{\lambda}, b_{j1}, \) and \( \beta_{j} \) represent the input data, the weight between nodes, and the bias, respectively. The ensemble deep learning model often considers many dead neurons using rectified linear unit (ReLU) activation function, which affect the output. This factor is emulated using leaky ReLU activation function. The leaky ReLU has a small negative slope of 0.01 when \( x < 0 \), which produces a leak and increases the range of ReLU. Therefore, \( \sum_{\lambda} \theta_{\lambda} \sum_{j} W_{j} \cdot b_{j1} + \beta_{j} \) is converted by using a leaky ReLU as expressed in equation 9.
The converted data are then brought to the output nodes to predict heart disease. The size of the output layer is two nodes, which represents the results of binary classification (the presence of heart disease or the absence of heart disease). The training process of deep learning began with a set of initial weights to each inputs value. Subsequently, the error between actual and predicted output is minimized using back propagation algorithm. While training the ensemble learning model, the weights are updated using equation 10.

\[
\frac{\partial E}{\partial w_j} = \sum_{i=1}^{T} f_i \sigma(\sum_{j=1}^{T} w_{ij} x_j + b_j) - y_i
\]

where \(\frac{\partial E}{\partial w_j}\) represents the learning rate, called positive constant, and \(E\) denotes the error which is defined by equation 11.

\[
\frac{\partial E}{\partial w_j} = \sum_{i=1}^{T} f_i (\sigma(\sum_{j=1}^{T} w_{ij} x_j + b_j) - y_i) x_j
\]

where \(p, o, T, Y\) denote the number of samples, the number of outputs, target output, and actual output, respectively. All the weights are updated based on the training error and the predicted values of the output layer are re-evaluated. The process is repeated until the network achieved the minimal error between the actual and predicted output. In this work, the learning rate with a value of 0.03 and the Adam optimizer is used for training the proposed deep learning model. The pseudo code of the ensemble deep learning model is presented in Algorithm 1.

**Algorithm 1.** Ensemble deep learning-based heart disease prediction.

| Input: Number of features \(Y\) for heart disease prediction |
| Output: The presence of heart disease or the absence of heart disease |

**Begin**

1. For number of training iterations do
2. Calculate the weighted sum and add bias in each hidden layer node by \(I^T \sigma(\sum_{j=1}^{T} w_{ij} x_j + b_j) \eta_j\)
3. \(\text{Compute} \ \frac{\partial E}{\partial w_j} = \sum_{i=1}^{T} f_i \sigma(\sum_{j=1}^{T} w_{ij} x_j + b_j) - y_i\)
4. \(\eta_j = \frac{\partial E}{\partial w_j} \sum_{i=1}^{T} f_i \sigma(\sum_{j=1}^{T} w_{ij} x_j + b_j) - y_i \sum_{j=1}^{T} w_{ij} x_j\)
5. \(\text{Choose} \ \alpha \ \text{and update} \ \frac{\partial E}{\partial w_j}\)
6. Repeat until error become minimal between \(T\) and \(Y\)

**End for**

7. Apply leaky ReLU activation function
8. Predict heart disease.
9. **Use LogitBoost as a meta-learning classifier**

### 3.4.2 Ontology-based recommendation

An ontology is a collection of classes (features) and the relationships among them. The ontology uses vocabulary to share knowledge from a specific domain among different modules. The definitions and construction procedures of the ontology are extensively discussed in our previously published papers [4,6–8]. The proposed ontology was developed using several stages in the Protégé Web Ontology Language (OWL). First, classes are constructed. Second, the data properties and object properties are implemented and utilized to determine relationships among classes. Finally, the inference rules are built for recommendations by using Semantic Web Rule Language rules. We use the proposed ontology for three main reasons. First, most of the heart disease prediction systems make physicians recheck medical reports for any kinds of advice for the patient. Therefore, a novel recommendation system is needed that can automatically suggest a dietary plan or activities for the patient, either at the hospital or at remote locations. Second, the recommendation module requires domain knowledge that can be used as expert knowledge for advising pre-diagnostic and diagnostic heart patients. Third, the symptoms of heart disease are similar to other diseases, which may confuse the physician, and therefore, a rules-based system is needed to eliminate confusion during recommendations. Fig. 6 presents the proposed ontology that is used for making recommendations for the heart patient.
An ontology-based recommendation model is the most essential part of the proposed SHMS. The developed ontology consists of facts and rules. The facts include extracted features from sensor data and EMRs, treatments, symptoms, and laboratory test results, as shown in Fig. 6. The rules include SWRL rules, which is expert knowledge that provides a recommendation for the patient according to the predicted health condition results. These rules are collected from online papers and cardiologists for all types of heart disease [2, 5, 20]. In this paper, we have designed 56 rules that accurately recommend activities and dietary plans for cardiac patients. Some of the SWRL rules are thoroughly discussed in our previously published paper [5]. However, the following rule is described in order to understand the use of all rules:

\[
\text{If } X \text{ is a cardiac patient and has BMI, HR, and Exercise results of high, normal, and No, respectively, then the recommendation is physical activities. This rule represents a recommendation that the patient increase physical activity such that calories consumed are burned.}
\]

The deep learning model first predicts heart disease using patient data. After prediction, the system identifies the patient’s gender. The main reason behind this is that a recommendation for a male cardiac patient is different than one for a female. In addition, the system finds the patient’s age and then identifies the group in which the age belongs (young, adult, old). The model then recommends a dietary plan or activities based on the patient’s gender, age, and predicted results (e.g., stop smoking, increase physical activities, control weight, and decrease meat consumption). In addition, this module calls rescue units and emergency services if the values of the extracted features are too high, and the predicted result is negative.

![Fig. 6. Features, Bayes axioms, and data and object properties of the proposed ontology.](image)

4. Experiments

The data collected using sensors and EMRs were discussed in subsections 3.1 and 3.2.1, respectively. The fusion of collected data was described in subsection 3.2.2. The data preparation for the use of the deep learning model was illustrated in subsection 3.3. In this section, the performance of the proposed system is evaluated, and the results are discussed.
4.1 Dataset

The proposed model was tested with two different heart disease datasets: Cleveland and Hungarian. These datasets are taken from the University of California, Irvine (UCI) online ML and data mining repository [14,30,43,46]. These datasets are considered to identify patients with heart disease, which is done with a number: 1 (present) or 0 (absent). The original Cleveland dataset consists of 303 cases with 76 features. In our study, only 14 and 16 features were considered in order to find the patient’s health condition. The Hungarian dataset comprises 294 cases with 14 features. Both datasets have cases with missing values, which are handled by the proposed filtering approach as discussed in subsection 3.3.1. We combined these datasets to make a single dataset for a better test of performance from the proposed model. The combined dataset contains 597 cases with 14 features. It should be noted that the Hungarian and the combined datasets use 14 features, and the Cleveland dataset was individually utilized but with only 16 features. A deep learning model cannot be used for a nominal dataset. Therefore, we converted nominal data into numeric data before using the ensemble deep learning model. The feature descriptions are shown in Table 3.

### Table 3: Feature information of heart disease from the Cleveland and Hungarian datasets.

<table>
<thead>
<tr>
<th>Label</th>
<th>Feature name</th>
<th>Description</th>
<th>Range</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_1$</td>
<td>AGE</td>
<td>Age of a patient in completed years.</td>
<td>29–79</td>
<td>Numeric</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>SEX</td>
<td>Gender of the patient (female is denoted by 0; male by 1).</td>
<td>0, 1</td>
<td>Nominal</td>
</tr>
<tr>
<td>$\rho_3$</td>
<td>CPT</td>
<td>Chest pain types in four categories: 1) typical angina, 2) atypical angina, 3) non-angina pain, and 4) asymptomatic.</td>
<td>1, 2, 3, 4</td>
<td>Nominal</td>
</tr>
<tr>
<td>$\rho_4$</td>
<td>RBP</td>
<td>The level of resting blood pressure in millimeters of mercury (mm Hg) when admitted to hospital.</td>
<td>94–200</td>
<td>Numeric</td>
</tr>
<tr>
<td>$\rho_5$</td>
<td>CHL</td>
<td>Level of cholesterol in milligrams per deciliter (mg/dl).</td>
<td>126–564</td>
<td>Numeric</td>
</tr>
<tr>
<td>$\rho_6$</td>
<td>YEARS</td>
<td>Number of years as a smoker.</td>
<td>5–45</td>
<td>Numeric</td>
</tr>
<tr>
<td>$\rho_7$</td>
<td>FBS</td>
<td>Level of fasting blood sugar (FBS is considered true if FBS is greater than 120 mg/dl, and false if it is below that).</td>
<td>0, 1</td>
<td>Nominal</td>
</tr>
<tr>
<td>$\rho_8$</td>
<td>FHIST</td>
<td>Family history of CAD (1=true, 0=false).</td>
<td>0, 1</td>
<td>Nominal</td>
</tr>
<tr>
<td>$\rho_9$</td>
<td>RER</td>
<td>The electrocardiogram reading at rest. It is presented in one of three categories: normal is denoted by the value 0, ST-T wave; abnormality (ST segment &gt; 0.05 mV) is represented by the value 1; and left ventricular (LV) hypertrophy by Estes’ criteria, denoted by the value 2.</td>
<td>0, 1, 2</td>
<td>Nominal</td>
</tr>
<tr>
<td>$\rho_{10}$</td>
<td>MHR</td>
<td>The maximum heart rate.</td>
<td>71–202</td>
<td>Numeric</td>
</tr>
<tr>
<td>$\rho_{11}$</td>
<td>EIA</td>
<td>Exercise-induced angina (0 represents No, 1 represents Yes).</td>
<td>0, 1</td>
<td>Nominal</td>
</tr>
<tr>
<td>$\rho_{12}$</td>
<td>OPK</td>
<td>Old Peak-ST depression induced by exercise in comparison with rest.</td>
<td>1, 2, 3</td>
<td>Numeric</td>
</tr>
<tr>
<td>$\rho_{13}$</td>
<td>SLP</td>
<td>ST segment measured in terms of slope of the peak exercise, which is represented as one of three categories (1 = up slope, 2 = flat, 3 = down slope).</td>
<td>1, 2, 3</td>
<td>Nominal</td>
</tr>
<tr>
<td>$\rho_{14}$</td>
<td>CA</td>
<td>Number of major vessels (0-3) colored by fluoroscopy.</td>
<td>0, 1, 2, 3</td>
<td>Numeric</td>
</tr>
<tr>
<td>$\rho_{15}$</td>
<td>THA</td>
<td>Period of exercise test in minutes. It demonstrates the heart status with one of three values (3 = normal, 6 = fixed defect, and 7 = reversible defect).</td>
<td>3, 6, 7</td>
<td>Nominal</td>
</tr>
<tr>
<td>$\rho_{16}$</td>
<td>NUM</td>
<td>Diagnosis of heart disease illustrated with one of five values: 0 (&lt;50% diameter narrowing) represents the absence of heart disease, and 1 to 4 (&gt;50% diameter narrowing) indicating the presence of heart disease to different degrees.</td>
<td>0, (1, T1, T2, T3, T4)</td>
<td>Nominal</td>
</tr>
</tbody>
</table>

### 4.2 Performance evaluation

This section presents the various experiments carried out to show how well the proposed SHMS can predict heart disease. The data were collected from the patient’s body using sensors and from the EMR. The collected data were fused in order to represent them in a structured form. The preprocessing module analyzed the structured data for further processing. Furthermore, the Cleveland and Hungarian datasets were then utilized in order to train the model for cardiac disease prediction. To evaluate the proposed deep learning model, we compared it with other classifiers: the support vector machine (SVM), logistic regression, multi-layer perceptron (MLP), random forest, the decision tree, and naive Bayes. The proposed model and other classifiers were utilized before and after feature selection, and with general feature weight and specific feature weight, and performance was compared. In addition, the proposed model and the other classifiers were tested based on various sizes of feature set and feature fusion. The results of the proposed model were also compared with the state-of-the-art methods. This system was developed using the Protégé OWL tool and the Waikato Environment for Knowledge Analysis (WEKA) with Java. The datasets were divided randomly into 70% and 30%, respectively, for training and testing the abovementioned models.
4.3 Evaluation metrics

Different performance metrics were utilized to determine the efficiency of the abovementioned ML models, as shown in Table 4. Accuracy represents the overall prediction ability of the proposed deep learning model. True positive (TP) and true negative (TN) measure the ability of classifier models to predict the absence and presence of patient heart disease. False positive (FP) and false negative (FN) identify the number of false predictions generated by the models. Precision and recall measure the success and sensitivity of the heart disease classification model, respectively. Function measure (FM) is utilized to determine the prediction performance. Root mean square error (RMSE) and mean absolute error (MAE) measure the difference and absolute variation between the actual values and predicted values, respectively. In the formulas for RMSE and MAE, N, \( t \) and \( \hat{y} \) denote the total number of observations, the actual value, and the predicted value, respectively.

Table 4: Performance metrics.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t )</td>
<td>( \sum_{i=1}^{N} y_i )</td>
</tr>
<tr>
<td>( \hat{y} )</td>
<td>( \sum_{i=1}^{N} \hat{y}_i )</td>
</tr>
<tr>
<td>( b )</td>
<td>( \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 )</td>
</tr>
<tr>
<td>( \hat{y} )</td>
<td>( \sum_{i=1}^{N} (\hat{y}_i - t)^2 )</td>
</tr>
</tbody>
</table>

4.4 Results

This section presents the results of the experiments defined in subsection 4.2. First, the effect of the fusion of sensor data and EMRs was evaluated based on the accuracy rate of the classifiers. We collected various features from both sensor data and EMRs, and fused them to generate more valuable and relevant data for classification. Thereafter, combinations of different heart disease features were tested in order to identify the best feature set for prediction.

4.4.1 Test results based on sensor data, EMRs, and feature fusion

Fig. 7. Classifier accuracy for fusion of sensor data and EMRs.