Tech Science Press

# Fusion-Based Machine Learning Architecture for Heart Disease Prediction

**Muhammad Waqas Nadeem[1,2], Hock Guan Goh[1,*], Muhammad Adnan Khan[3], Muzammil Hussain[4], Muhammad Faheem Mushtaq[5] and Vasaki a/p Ponnusamy[1]**

[1]Faculty of Information and Communication Technology (FICT), Universiti Tunku Abdul Rahman (UTAR), Kampar, Perak, 31900, Malaysia
[2]Department of Computer Science, Lahore Garrison University, Lahore, 54000, Pakistan
[3]Department of Computer Science, Faculty of Computing, Riphah International University, Lahore Campus, Lahore, 54000, Pakistan
[4]Department of Computer Science, School of Systems and Technology, University of Management and Technology, Lahore, 54000, Pakistan
[5]Department of Information Technology, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, 64200, Pakistan
*Corresponding Author: Hock Guan Goh. Email: gohhg@utar.edu.my
Received: 05 October 2020; Accepted: 21 November 2020

**Abstract:** The contemporary evolution in healthcare technologies plays a considerable and significant role to improve medical services and save human lives. Heart disease or cardiovascular disease is the most fatal and complex disease which it is hardly to be detected through our naked eyes, as numerous people have been suffering from this disease globally. Heart attacks occur when the ranges of vital signs such as blood pressure, pulse rate, and body temperature exceed their normal values. The efficient diagnosis of heart diseases could play a substantial role in the field of cardiology, while diagnostic time could be reduced. It has been a key challenge for researchers and medical experts to diagnose heart diseases accurately and timely. Therefore, machine learning-based techniques are used for the diagnosis with higher accuracy, using datasets compiled from former medical patients' reports. In recent years, numerous studies have been presented in the literature propose machine learning techniques for diagnosing heart diseases. However, the existing techniques have some limitations in terms of their accuracy. In this paper, a novel Support Vector Machine (SVM) based architecture for heart disease prediction, empowered with a fuzzy based decision level fusion, is presented. The SVM-based architecture has improved the accuracy significantly as compared to existing solutions, where 96.23% accuracy has been achieved.

**Keywords:** Heart disease; machine learning; support vector machine; fuzzy logic; fusion; cardiovascular

## 1 Introduction

Heart disease (HD) is a serious health issue around the world and numerous peoples are affected by this disease [1]. The most common symptoms of HD are physical body weakness, breath shortness, and swollen feet [2]. In recent years, many researchers present various machine learning methods and techniques for early prediction of heart disease but the existing diagnostic techniques for heart disease are not efficient and effective due to several reasons such as execution time and accuracy of the machine learning models [3]. Due to the unavailability of a medical expert and modern technology, the diagnosis and treatment of heart disease are difficult to be carried out appropriately [4]. The life of numerous people can be saved by using effective and accurate diagnostic technologies [5]. According to the European Society of Cardiology, there are 3.6 million people have diagnosed as HD patients annually around the world [6,7]. Most of the people in the United States (US) are affected by HD [8]. Approximately 50% of heart patients that are suffering from heart disease can survive within 1–2 years, and 3% of the financial healthcare budget is used for the management of heart disease [9]. Traditionally, the physician use concerning symptoms, patient medical history, and physical examination reports for the diagnosis of heart disease. The results obtained from these methods are not effective and accurate for the identification of HD patient. Moreover, these methods are computationally difficult and expensive [10].

The development of machine learning-based noninvasive diagnostic systems is needed for effective diagnosis of HD [11–16]. Machine learning-based expert decision systems and applications of Artificial Fuzzy Logic (AFL) efficiently diagnose the heart disease patient that results in the decreases in death ratio [17,18]. The Cleveland heart disease data set used by several researchers [19–24] for the prediction of HD. The proper data are required by the predictive machine learning models for their training and testing. The use of balanced data set for the training and testing improves the performance machine learning model. Furthermore, the use of proper and related features from the data set improves the predictive capabilities of the model. Hence, features selection and data balancing are key parameters to improve the performance of the model. In the literature, several machine learning-based diagnostic methods and techniques such as Neuro Fuzzy, Artificial Neural Network (ANN), Support Vector Machine (SVM), Decision Tree (DT), Naïve Bayes (NB) etc. have been proposed by researchers, but these techniques have some limitations that include lack of large training data, inconsistency accuracy, proper data balancing, and so on. Furthermore, these techniques do not effectively diagnose heart disease. The data standardization at the data processing layer also improves the predictive capabilities of the machine learning models. Since more, some other preprocessing techniques that include Min–Max Scalar, removal of missing features from the dataset, and standard scalar improve the performance of the model [20]. Several features selection techniques such as Principle Component Analysis (PCA), Local Learning-Based Features Selection (LLBFS), Greedy Algorithm (GA) etc. are used for the selection of important parameters. Furthermore, several optimization techniques that include Bacterial Foraging Optimization (BFO), Ant Colony Optimization (ACO) and so on also used for the optimization of features before training of machine learning models [25].

Furthermore, in recent years, several machine learning algorithms such as ANN, SVM, K-Nearst Neighbour (KNN) etc. are used in the Internet of Things (IoT) based systems for prediction and classification [26]. The unsupervised machine learning algorithms are used to label the data which is collected by the different IoT devices. The data which is labeled by the machine learning algorithms gives more accurate results as compared to manual labeling.

Hence more, Neural Networks based tools achieved state-of-the-art performance for the prediction of brain and heart diseases. In recent years, Carotid Artery Stenting (CAS) treatment

is commonly used in the field of medicine. The CAS methods give an overview of the Major Adverse Cardiovascular Events (MACE) of the HD patients at an early stage. The ANN produces more accurate results as compared to the simple CAS method [27]. The proposed ANN-based methods do not only combine posterior probabilities but also produce vales from multiple predecessor techniques. The ANN-based methods achieved much better results as compared to existing methods [28].

In this paper, supervised machine learning architecture empowered with fuzzy-based decision level fusion medical expert system is proposed for the prediction of heart disease. The proposed architecture consists of two phases: supervised machine learning phase and fuzzy-based decision level fusion phase. The main objective of this proposed architecture is to improve the accuracy of machine learning-based solution for the diagnosis of heart disease. Furthermore, in recent years, many studies restrict the use of feature selection methods for the model. Therefore, the proposed model working on the mechanism of parallel computation that allows us to use all the features without any restriction of feature selection method at the pre-processing layer. The experiment results show that the proposed architecture has effective results in terms of accuracy for the diagnosis of heart disease as compared to existing machine learning methods.

The rest of the paper is organized as follows: Section 2 describes the related work. Section 3 presents the materials and methods for the diagnosis of heart disease. Section 4 discusses the simulation results of the proposed architecture. Section 5 concludes the study.

## 2 Related Work

In the literature, numerous machine learning-based medical expert systems were designed by the researchers for the diagnosis of heart disease. This paper gives an overview of some existing machine learning-based diagnosis systems for heart disease and highlights the importance of the proposed work. ANN-based diagnostics models give the highest prediction accuracy in the domain of healthcare [27]. Similarly, Big Data and Optimal Artificial Neural Network (OANN) based model is presented in [29] that achieved the 90.91% prediction accuracy. Kaggle and UCI laboratory heart disease data sets have been used by the researchers to discover the patterns using different machine learning algorithms such as DT, ANN, NB, and SVM. The hybrids methods give more accuracy as compared with a single machine learning algorithm [30].

Furthermore, numerous machine learning-based noninvasive medical support systems such as ANN, SVM, DT, NB, KNN, Logistic Regression (LR), Fuzzy logic (FL), Adaboost (AB) are developed by the researchers in the recent years for the diagnosis of heart disease [18,31]. The use of machine learning-based medical expert systems for the diagnosis of heart disease gradually increases which decreases the death ratio of heart patients [32]. Several machine learning-based medical expert system for the diagnosis of HD has been reported in numerous scientific research studies.

Support Vector Machine and Principal Component Analysis (SVM-PCA) based system is present in [33] which achieved 88.24% classification accuracy. Another SVM based model is presented in [34] to predict the risk of heart disease and achieved 89.9% accuracy. In [35], ANN and Neuro Fuzzy based predictive model for heart disease that obtained 87.04% accuracy was presented. Olaniyi et al. [36] presented a three-phase ANN model to diagnose heart disease and achieved 88.89% classification accuracy and the proposed system easily deployed in health care information systems. Another Ensemble-based ANN predictive model is also presented in [29] which used statistical analysis technique for the diagnosis of heart disease and obtained 89.01%

accuracy, 95.91% specificity, and 80.09% sensitivity. Furthermore, ANN and Fuzzy Analytical Hierarchical Processing (F-AHP) based integrated decision support medical system is presented in [37] that achieved 91.10% classification accuracy.

## 3 Materials and Method

The section briefly describes the research method and materials of the paper.

### 3.1 Dataset

Two different heart disease datasets are used in this paper to train the supervised machine learning algorithm. The first "heart disease dataset 2019," which is used by various researchers [13] in recent years for the diagnosis of heart disease. The "heart disease dataset 2019" is also publically available on the online Kaggle repository. The heart disease dataset has 1025 number of samples, 13 features, and some missing values. The target output label has two classes that represent the patient is normal or heart patient. The second "cardiovascular disease dataset 2019" is also used in this paper. The cardiovascular disease dataset 2019 is also available on the online Kaggle repository. The cardiovascular disease dataset 2019 has 70,000 number of patient samples, 11 unique features, and some missing values. The detailed description of these datasets is given in Tabs. 1 and 2.

### 3.2 Experimental Design Setup

The supervised prediction experiment has been conducted to evaluate the performance of the proposed architecture. First, we evaluate the performance of the Support vector machine (SVM) on two different data sets. The K-fold cross-validation method is applied to split the data. To access the performance of the architecture several performance evaluation metrics are computed. All the computation experiment has been performed in Python 3.7 environment using several machine learning libraries on an Intel® Core™ i3-3217U CPU@1.80 GHz PC.

### 3.3 Proposed System Model

The proposed supervised machine learning architecture empowered with fuzzy-based decision level fusion is presented in Fig. 1. The data set which is generated by the Internet of Medical Things (IoMT) enabled devices are used for the training of machine learning algorithms. The proposed architecture consists of two phases: the supervised machine learning phase and the fuzzy-based decision level fusion phase. The supervised machine learning phase has three distinct layers that include the pre-processing layer, application layer, and performance layer. The pre-processing layer receives raw data and maybe the raw data has some missing values and noise. At the pre-processing layer, different methods such as mean, mode, and average are applied for the prediction of missing values and remove the noise using normalization. Furthermore, the application layer receives the processed data and the processed data is used to train the supervised machine learning technique named SVM. The same mechanism is executed in parallel inside the proposed architecture.

#### 3.3.1 Preprocessing

In the proposed architecture, the preprocessing step includes handling missing values, moving average, and normalization are describe as follows:

At the first step, the null and missing values are filled in the data set, because they can lead towards the wrong prediction of any machine learning model. In the proposed architecture, mean

method is selected to impute the missing or null values because the mean method is beneficial as it impute continuous data without introducing outliers. The mean method is formulated as:

$$Q(x) = \begin{cases} mean(x), & if \ x = null/missed \\ x, & otherwise \end{cases}$$

where $x$ represents the instances of feature vectors that lies in n-dimensional space, $x \in R$.

**Table 1:** Description of heart disease dataset 2019 with feature information

| Sr # | Feature name | Feature code | Description | Type | Value range (min–max) |
|---|---|---|---|---|---|
| 1 | Age | Age | Age in years | Numeric | 29<age>77 |
| 2 | Sex | Sex | 1 = Male | Nominal | 1 |
|   |   |   | 0 = Female |   | 0 |
| 3 | Chest pain type | Cp | 1 = atypical angina | Nominal | 1 |
|   |   |   | 2 = typical angina |   | 2 |
|   |   |   | 3 = asymptomatic |   | 3 |
|   |   |   | 4 = nonanginal pain |   | 4 |
| 4 | Resting blood pressure | Trestbps | In mm Hg on admission to the hospital | Numeric | 94–200 |
| 5 | Serum cholesterol | Chol | In mg/dl | Numeric | 126–564 |
| 6 | Fasting blood sugar | Fbs | Fasting blood sugar &gt; 120 mg/dl) (1 = true; 0 = false) | Nominal | 1 0 |
| 7 | Resting electrocardiographic results | Rest ECG | 0 = normal | Nominal | 0 |
|   |   |   | 1 = having ST-T |   | 1 |
|   |   |   | 2 = hypertrophy |   | 2 |
| 8 | Maximum heart rate achieved | Thalach | Not mention | Numeric | 77–202 |
| 9 | Exercise-induced angina | Exang | 1 = yes | Nominal | 1 |
|   |   |   | 0 = no |   | 0 |
| 10 | Old peak = ST depression induced by exercise relative to rest | oldpeak | Not mention | Numeric | 0–6.2 |
| 11 | Slope of the peak exercise ST segment | slope | 1 = up sloping | Nominal | 1 |
|   |   |   | 2 = flat |   | 2 |
|   |   |   | 3 = down sloping |   | 3 |
| 12 | Number of major vessels (0–3) colored by fluoroscopy | Ca | Not mention | Nominal | 1 2 3 |
| 13 | Thallium scan | Thal | 3 = normal | Nominal | 3 |
|   |   |   | 6 = fixed defect |   | 6 |
|   |   |   | 7 = reversible defect |   | 7 |

**Table 2:** Description of cardiovascular disease dataset 2019 with feature information

| Sr # | Feature name | Feature code | Description | Type | Value range (min–max) |
|---|---|---|---|---|---|
| 1 | Age | Age | In days | Numeric | 10798<day>23713 |
| 2 | Gender | Gender | 1 = women | Nominal | 1 |
|   |   |   | 2 = men |   | 2 |
| 3 | Height | Height | In cm | Numeric | 55–250 |
| 4 | Weight | Weight | In Kg | Numeric | 10–200 |
| 5 | Systolic blood pressure | ap_hi | Not mention | Numeric | −150–16020 |
| 6 | Diastolic blood pressure | ap_lo | Not mention | Numeric | −70–11000 |
| 7 | Cholesterol | Cholesterol | 1 = normal | Nominal | 1 |
|   |   |   | 2 = above normal |   | 2 |
|   |   |   | 3 = well above normal |   | 3 |
| 8 | Glucose | Gluc | 1 = normal | Nominal | 1 |
|   |   |   | 2 = above normal |   | 2 |
|   |   |   | 3 = well above normal |   | 3 |
| 9 | Smoke | Smoke | Whether patient | Nominal | 1 |
|   |   |   | 1 = smoke |   | 0 |
|   |   |   | 0 = not smoke |   |   |
| 10 | Alcohol intake | Alco | whether patient take | Nominal | 1 |
|   |   |   | 1 = take alcohol |   | 0 |
|   |   |   | 0 = not take alcohol |   |   |
| 11 | Physical activity | Active | whether physical active | Nominal | 1 |
|   |   |   | 1 = active |   | 0 |
|   |   |   | 0 = not active |   |   |

In Moving average (MA), to reduce the noise from the data set, a series of averages is computed of different subsets using full data set. Arithmetic mean of given set of values is taken to calculate the moving average. The moving average is formulated as:

$$MA = \frac{x1 + x2 + x3 + \ldots + xn}{N}$$

where $x1, x2, x3, \ldots, xn$ represents instances of the feature vector and $N$ represents total number of attributes.

In normalization, standardization or Z-score normalization techniques is used to rescale the values of attributes. The standardization method normalize the distribution of data with zero mean and also reduce the skewness of the data distribution. The standardization is formulated as:

$$R(x) = \frac{x - \bar{x}}{\sigma}$$

where $x$ is the instances of feature vectors with n-dimensional space, $x \in R^n$. $\bar{x}$ and $\sigma$ represent mean and standard deviation of attributes respectively.
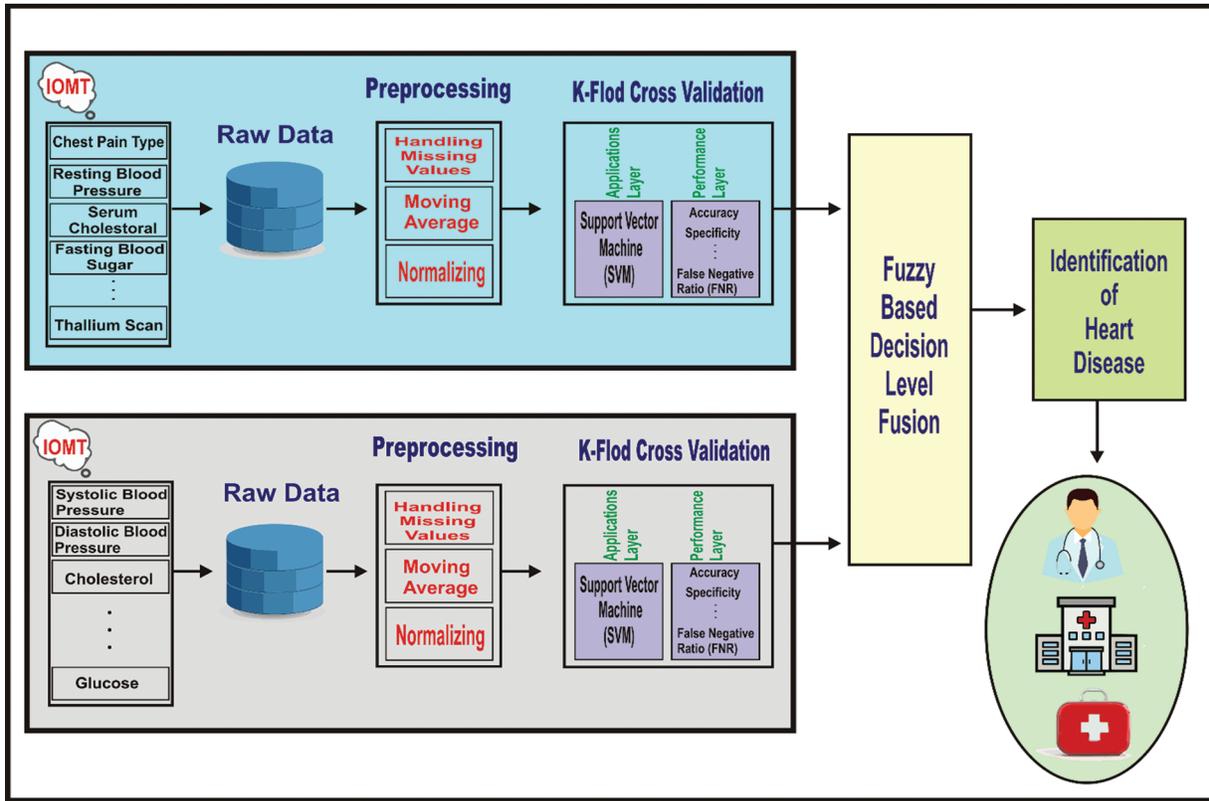


**Figure 1:** The systematic diagram of proposed supervised machine learning architecture empowered with fuzzy based decision level fusion

### 3.3.2 K-Fold Cross Validation

The K-fold cross validation method is widely used by the researchers for the selection of machine learning model and estimation of classifiers error [28]. In the proposed architecture, 5-fold cross-validation is used to split the data set for the training and testing of SVM. The fold-1 is used to train and fine-tune of hyper-parameters in inner loop where grid search algorithm is employed [29]. In outer loop (k times), the performance of the model is evaluated using test data. Since more, the data sets which are used for the training and testing of proposed architecture has imbalanced negative and positive samples. The stratified KCV is used to preserve the ratio of each class. The final performance of the model is evaluated by using the following formula:

$$M = \frac{1}{K} \times \sum_{n=1}^{K} P_n \pm \sqrt{\sum_{n=1}^{K} \frac{\left(P_n - \bar{P}\right)^2}{K-1}}$$ where M denotes the final performance metric for the classifiers

and $Pn \in R$, n = 1, 2, 3, ..., K represents the performance metric for each fold.

### 3.3.3 Support Vector Machine

SVM algorithm is used for regression and classification. In SVM based models, the data points are categorized into groups, represent on the space and the points which have similar properties falls in same group. In linear SVM, the p-dimensional vector is considered for the given data and separated by maximum of $p - 1$ planes that are known as hyper planes. These planes are used to separate the data space among different data groups to regression and classification. The mathematical representation of SVM is formulated as:

The equation of the line is described as:

$$a_1 = a_2 x + b \tag{1}$$

In Eq. (1) 'x' is the slope of the line and 'b' is intersect, so

$$a_1 - a_2 x + b = 0$$

Let $\bar{a} = (a_1, a_2)^T$, and $\bar{z} = (x, -1)$ so the above equation can be written as

$$\bar{z} \cdot \bar{a} + b = 0 \tag{2}$$

The Eq. (2) is derived from 2-dimensional vectors. The above equation is also applicable for any number of dimensions. The Eq. (2) is also called the hyper lane equation.

Vector direction $\bar{a} = (a_1, a_2)^T$ is written in the form of $\bar{z}$ and defined as:

$$\bar{z} = \frac{a_1}{\|a\|} + \frac{a_2}{\|a\|} \tag{3}$$

where

$$\|a\| = \sqrt{a_{1+}^2 a_{2+}^2 a_{3+}^2 \ldots a_n^2}$$

As we know that

$$\cos(\theta_1) = \frac{a_1}{\|a\|} \quad \text{and} \quad \cos(\theta_2) = \frac{a_2}{\|a\|}$$

So, Eq. (3) can also be written as

$$\bar{z} = (\cos(\theta_1), \cos(\theta_2))$$

$$\bar{z} \cdot \bar{a} = \|\bar{z}\| \|a\| \cos(\theta)$$

$$\theta = \theta_1 - \theta_2$$

$$\cos(\theta) = \cos(\theta_1 - \theta_2) = \cos(\theta_1)\cos(\theta_2) + \sin(\theta_1)\sin(\theta_2)$$

$$= \frac{\bar{z}_1}{\|\bar{z}\|} \frac{a_1}{\|a\|} + \frac{\bar{z}_2}{\|\bar{z}\|} \frac{a_2}{\|a\|}$$

$$= \frac{\bar{z}_1 a_1 + \bar{z}_2 a_2}{\|\bar{z}\| \|a\|}$$

$$\bar{z} \cdot a = \|\bar{z}\| \|a\| \left[ \frac{\bar{z}_1 a_1 + \bar{z}_2 a_2}{\|\bar{z}\| \|a\|} \right]$$

$$\bar{z} \cdot \bar{a} = \sum_{i=1}^{n} \bar{z}_i a_i$$

For n-dimensional vectors, the dot product of the above equation is computed as:

Let, $f = y(\bar{z} \cdot a + b)$

If sign $(f) > 0$ mean the classification is correct and sign $(f) < 0$ mean the classification is incorrect. If D is given dataset, then f is computed on a training dataset

$$f_i = y_i(\bar{z} \cdot a + b) \tag{4}$$

We also compute the functional margin (F) of a dataset as:

$$F = \min_{i=1 \ldots m} f_i$$

Through the comparison of hyperplanes, the hyperplane which has the largest F will be selected. Where F is known as the geometric mean of the dataset. We need to find the optimal values of $\bar{z}$ and b for the selection of optimal hyperplane.

The Lagrangian function is

$$L(\bar{z}, b, \alpha) = \frac{1}{2}\bar{z} \cdot \bar{z} - \sum_{i=1}^{m} \alpha_i [y: (\bar{z} \cdot a + b) - 1]$$

$$\nabla_{\bar{z}} L(\bar{z}, b, \alpha) = \bar{z} - \sum_{i=1}^{m} \alpha_i y_i a_i = 0 \tag{5}$$

$$\nabla_b L(\bar{z}, b, \alpha) = -\sum_{i=1}^{m} \alpha_i y_i = 0 \tag{6}$$

By using Eqs. (5) and (6) we get

$$\bar{z} = \sum_{i=1}^{m} \alpha_i y_i a_i \quad \text{and} \quad \sum_{i=1}^{m} \alpha_i y_i = 0 \tag{7}$$

After the substitution of Lagrangian function L we get

$$\bar{z}(\alpha, b) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j a_i a_j$$

Thus,

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j a_i a_j \tag{8}$$

Subject to

$$\alpha_i \geq 0, \quad i = 1 \ldots m, \quad \sum_{i=1}^{m} \alpha_i y_i = 0$$

Due to the inequality of constraints, the Lagrangian function can be extended to Karush-kuhn-tucker (KKT) conditions. Eq. (9) describes the complementary conditions of KKT.

$$\alpha_i \left[ y_i \left( \bar{z}_i.a^* + b \right) - 1 \right] = 0 \tag{9}$$

where $a^*$ is the optimal point and $\alpha$ is a positive value. For other points, the value of $\alpha$ is $\approx 0$

So,

$$y_i \left( \left( \bar{z}_i.a^* + b \right) - 1 \right) = 0. \tag{10}$$

Eq. (10) describe the support vectors which are closest points to the hyperplane.

$$\bar{z} - \sum_{i=1}^{m} \alpha_i y_i a_i = 0$$

$$\bar{z} = \sum_{i=1}^{m} \alpha_i y_i a_i \tag{11}$$

The value of b is computed as

$$y_i \left( \left( \bar{z}_i.a^* + b \right) - 1 \right) = 0 \tag{12}$$

Multiply by y on both sides of Eq. (12)

$$y_i^2 \left( \left( \bar{z}_i \cdot a^* + b \right) - y_i \right) = 0, \quad \text{where } y_i^2 = 1$$

$$\left( \left( \bar{z}_i \cdot a^* + b \right) - y_i \right) = 0$$

$$b = y_i - \bar{z}_i \cdot a^* \tag{13}$$

Then

$$b = \frac{1}{S} \sum_{i=1}^{s} (y_i - \bar{z} \cdot a) \tag{14}$$

In Eq. (14) 's' represents the number of support vectors. These support vectors make the hyperplanes and then hyperplanes are used for prediction.

The hypothesis function is described as:

$$h (\bar{z}_i) = \begin{bmatrix} +1 & \text{if } \bar{z} \cdot a + b \geq 0 \\ -1 & \text{if } \bar{z} \cdot a + b < 0 \end{bmatrix} \tag{15}$$

If the point is above the hyperplane then it will be classified as +1 class mean the HD found and if the point is below the hyperplane then it is classified as −1 class mean the HD does not found.

### 3.3.4 Fuzzy Based Fusion

After the training of the SVM, different evaluation parameters such as accuracy, sensitivity, specificity etc. are used to evaluate the performance of the proposed architecture. Once the process of performance evaluation is completed for both SVM individually then a fuzzy-based decision level fusion process is applied to integrate the performance of both SVM for the final decision as:

$$\mu_{FHD}(fh) = \mu_{SVM1 \cap SVM2}(s1, s2)$$
$$\mu_{FHD}(fh) = \min[\mu_{SVM1}(s1), \mu_{SVM2}(s2)] \tag{16}$$

In Eq. (16) the FHD denotes the fusion-based heart disease prediction. The t-norm function for fuzzy-based fusion is defined as:

$$t: [0, 1] \times [0, 1] \to [0, 1]$$

After the implementation of the t-norm function fuzzy-based fusion implication is applied as:

$$\mu Q_6(s1, s2) = \min[\mu_{FP1}(s1), \mu_{FP2}(s2)]$$

The Eq. (17) shows the relationship between SVM1 and SVM2:

$$Q_6 = \bigcup_{e=1}^{6} Ru^e \tag{17}$$

Eq. (18) integrates the performance of SVM1 and SVM2 in crisp form is as follow:

$$\mu_{\varphi}(L) = \max_{1 \leq i \leq 6} \left[ sup_{I \in (s1,s2)} \left( \prod_{k=1}^{6} \left( \mu_{s1_k, s2_k}(s1, s2) \right) \right) \right] \tag{18}$$

Defuzzifier is a very important component of an expert system. It is the process of mapping the fuzzy sets to the crisp output. The center of gravity defuzzifier is used to get the final fused decision of the proposed architecture. The center of gravity defuzzifier specifies the $\Omega^*_{COG}$ as the center of the area covered by the membership function of $\varphi$ for fuzzy-based decision level fusion, that is,

$$\Omega^*_{COG} = \frac{\int_v \varphi \mu_{\varphi}(\varphi) \, d\varphi}{\int_v \mu_{\varphi}(\varphi) \, d\varphi}$$

## 4 Results and Discussion

The proposed supervised machine learning architecture empowered with fuzzy-based decision level fusion has been applied to two different datasets. The proposed architecture working on the mechanism of parallel computing. Furthermore, the k-fold cross-validation method is used to split

the dataset into different folds for the training and testing of the proposed architecture. Different evaluation metrics are used to access the performance of the architecture which are as follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Miss\ rate = \frac{FP + FN}{TP + TN + FP + FN} \times 100\%$$

$$Sensitivity/recall = \frac{TP}{TP + FN} \times 100\%$$

$$Specificity = \frac{TN}{TN + FP} \times 100\%$$

$$Percision = \frac{TP}{TP + FP} \times 100\%$$

$$False\ positive\ ratio = 1 - specificity$$

$$False\ negative\ ratio = 1 - sensitivity$$

The proposed architecture predicts the output as positive (+1) and negative (−1). The +1 indicates the presence of heart disease and −1 indicates that no symptoms of heart disease found in the patient. The performance of the proposed supervised machine learning architecture empowered with fuzzy-based decision level fusion using different statistical metrics are shown in Tab. 3. In Tab. 3 it is clearly shown that the proposed architecture achieved effective results during fold-5 cross-validation. The architecture achieved 96.23%, 95.64%, 94.36%, 97.01%, 3.7%, 4.36%, and 2.99% in terms of accuracy, specificity, precision, sensitivity, miss rate, false positive ratio, and false negative ratio respectively.

**Table 3:** The performance of the proposed architecture

| K-fold | Accuracy (%) | Specificity (%) | Precision (%) | Sensitivity (%) | Miss rate (%) | False positive ratio (FPR) (%) | False negative ratio (FNR) (%) |
|--------|----------|-------------|-----------|-------------|-----------|-----------------------|-----------------------|
| 2-fold | 84.41 | 83.06 | 76.62 | 86.52 | 15.58 | 16.94 | 13.48 |
| 3-fold | 94.70 | 94.10 | 92.32 | 95.75 | 5.19 | 5.9 | 4.25 |
| 4-fold | 94.80 | 95.85 | 94.70 | 95.52 | 4.29 | 4.15 | 4.48 |
| 5-fold | 96.23 | 95.64 | 94.36 | 97.01 | 3.7 | 4.36 | 2.99 |

The comparison of proposed supervised machine learning architecture empowered with fuzzy-based decision level fusion with existing methods is described in Tab. 4. Different machine learning methods and architectures for the diagnosis of heart disease which is presented by the researcher in recent years such as Multilayer Perceptron combine with SVM (MLP+SVM), Hybrid machine learning-based diagnostic system, ANN combine with FL (ANN + FL), Hybrid Random Forest with a Linear Model (HRFLM), ANN combine with Fuzzy Analytical Hierarchy Process (AHP) (ANN + Fuzzy AHP) etc. are studied for the comparative analysis of proposed architecture. The accuracy performance metric is used to compare the performance of proposed architecture with

existing methods in the field of heart disease. It is observed that the proposed architecture gives better results in terms of accuracy as compared to the other existing methods.

**Table 4:** Comparative analysis of proposed architecture with existing methods

| Study | Method | Year of proposed | Evaluation in term of accuracy (%) |
|---|---|---|---|
| [33] | Support vector machine + principle component analysis (SVM + PCA) | 2018 | 88.24 |
| [28] | Hybrid machine learning-based diagnostic system | 2019 | 88.47 |
| [35] | Artificial neural network + Neuro Fuzzy logic (ANN + NF) | 2014 | 87.04 |
| [34] | Support vector machine based heart disease risk prediction model | 2020 | 89.9 |
| [29] | Big data + optimal artificial neural network (OANN)-based diagnostic system | 2020 | 90.91 |
| [36] | ANN-based three-phase method | 2015 | 88.89 |
| [37] | Artificial neural network + Fuzzy analytical hierarchy process (AHP) (ANN + Fuzzy AHP) | 2017 | 91.1 |
| [38] | HD detection system based on a set of relief-rough | 2017 | 92.32 |
| [25] | Fast conditional mutual information (FCMIM) + support vector machine (FCMIM + SVM) | 2020 | 92.37 |
| [39] | Cloud computing and machine learning algorithm support vector machine (SVM) | 2020 | 93.33 |
| **The Proposed architecture** | **Fussion based machine learning** | **2020** | **96.23** |

## 5 Conclusion

The early diagnosis of heart abnormalities and information related to heart condition from raw health care data is very important which could help to save human lives in the long term. In recent years, machine learning methods and techniques have achieved effective performance to process raw data and give a novel and new discernment toward heart disease. The prediction of heart disease is an important and a challenging task in the field of medical. However, the mortality rate of heart disease can be significantly controlled if heart disease is diagnosed at an early stage and adopt preventative measures. Furthermore, different machine learning methods and techniques for the diagnosis of heart disease are presented in recent years. The existing machine learning methods have some limitations in terms of accuracy. Therefore, the proposed supervised machine learning architecture empowered with fuzzy-based decision level fusion has achieved 96.23% accuracy which is much better than the existing methods. The proposed work can be extended by using different machine learning algorithms such as Artificial Neural Network, Decision Tree, Random Forest etc. along with SVM.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

[1]   A. L. Bui, T. B. Horwich and G. C. Fonarow, "Epidemiology and risk profile of heart failure," *Nature Reviews Cardiology*, vol. 8, no. 1, pp. 30–41, 2011.

[2]   M. Durairaj and N. Ramasamy, "A comparison of the perceptive approaches for preprocessing the data set for predicting fertility success rate," *International Journal of Control theory and Applications*, vol. 9, no. 27, pp. 1–7, 2016.

[3]   L. A. Allen, L. W. Stevenson, K. L. Grady, N. E. Goldstein, D. D. Matlock *et al.,* "Decision making in advanced heart failure: A scientific statement from the American Heart Association," *Circulation*, vol. 125, no. 15, pp. 1928–1952, 2012.

[4]   S. Ghwanmeh, A. Mohammad and A. A. Ibrahim, "Innovative artificial neural networks-based decision support system for heart diseases diagnosis," *Journal of Intelligent Learning Systems and Applications*, vol. 5, pp. 1–6, 2013.

[5]   Q. K. A. Shayea, "Artificial neural networks in medical diagnosis," *International Journal of Computer Science*, vol. 8, no. 2, pp. 150–154, 2011.

[6]   A. J. S. Coats, "Ageing, demographics, and heart failure," *European Heart Journal Supplements*, vol. 21, no. Supplement_L, pp. L4–L7, 2019.

[7]   I. Spoletini and M. Lainscak, "Epidemiology and prognosis of heart failure," *International Cardiovascular Forum Journal*, vol. 10, pp. 1–6, 2017.

[8]   P. A. Heidenreich, J. G. Trogdon, O. A. Khavjou, J. Butler, K. Dracup *et al.,* "Forecasting the future of cardiovascular disease in the United States: A policy statement from the American Heart Association," *Circulation*, vol. 123, no. 8, pp. 933–944, 2011.

[9]   A. U. Haq, J. P. Li, M. H. Memon, S. Nazir and R. Sun, "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms," *Mobile Information Systems*, vol. 2018, no. 8, pp. 1–21, 2018.

[10]  A. Tsanas, M. A. Little, P. E. M. Sharry and L. O. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity," *Journal of the Royal Society Interface*, vol. 8, no. 59, pp. 842–855, 2011.

[11]  A. H. Gonsalves, F. Thabtah, R. M. A. Mohammad and G. Singh, "Prediction of coronary heart disease using machine learning: An experimental analysis," in *Int. Conf. on Deep Learning Technologies*, Xiamen, China, pp. 51–56, 2019.

[12]  P. Sharma, K. Choudhary, K. Gupta, R. Chawla, D. Gupta *et al.,* "Artificial plant optimization algorithm to detect heart rate & presence of heart disease using machine learning," *Artificial Intelligent Medicine*, vol. 102, pp. 101752–101765, 2020.

[13]  Y. Khan, U. Qamar, N. Yousaf and A. Khan, "Machine learning techniques for heart disease datasets: A survey," in *Int. Conf. on Machine Learning and Computing*, Zhuhai, China, pp. 27–35, 2019.

[14]  G. P. Diller, A. Kempny, S. V. B. Narayan, M. Henrichs, M. Brida *et al.,* "Machine learning algorithms estimating prognosis and guiding therapy in adult congenital heart disease: Data from a single tertiary centre including 10019 patients," *European Heart Journal*, vol. 40, no. 13, pp. 1069–1077, 2019.

[15]  N. S. C. Reddy, S. S. Nee, L. Z. Min and C. X. Ying, "Classification and feature selection approaches by machine learning techniques: Heart disease prediction," *International Journal of Innovative Computing*, vol. 9, no. 1, pp. 39–46, 2019.

[16] Y. Meng, W. Speier, C. Shufelt, S. Joung, J. E. V. Eyk *et al.,* "A machine learning approach to classifying self-reported health status in a cohort of patients with heart disease using activity tracker data," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 3, pp. 878–884, 2019.

[17] S. I. Ansarullah and P. Kumar, "A systematic literature review on cardiovascular disorder identification using knowledge mining and machine learning method," *International Journal of Recent Technology and Engineering*, vol. 7, no. 65, pp. 1009–1015, 2019.

[18] S. Nazir, S. Shahzad, S. Mahfooz and M. Nazir, "Fuzzy logic based decision support system for component security evaluation," *International Arab Journal of Information Technology*, vol. 15, no. 2, pp. 224–231, 2018.

[19] J. Nahar, T. Imam, K. S. Tickle and Y. P. P. Chen, "Computational intelligence for heart disease diagnosis: A medical knowledge driven approach," *Expert Systems with Applications*, vol. 40, no. 1, pp. 96–104, 2013.

[20] J. Nahar, T. Imam, K. S. Tickle and Y. P. P. Chen, "Association rule mining to detect factors which contribute to heart disease in males and females," *Expert Systems with Applications*, vol. 40, no. 4, pp. 1086–1093, 2013.

[21] C. B. Gokulnath and S. P. Shantharajah, "An optimized feature selection based on genetic approach and support vector machine for heart disease," *Cluster Computing*, vol. 22, no. S6, pp. 14777–14787, 2019.

[22] M. S. Amin, Y. K. Chiam and K. D. Varathan, "Identification of significant features and data mining techniques in predicting heart disease," *Telematics and Informatics*, vol. 36, pp. 82–93, 2019.

[23] H. Ahmed, E. M. G. Younis, A. Hendawi and A. A. Ali, "Heart disease identification from patients' social posts, machine learning solution on spark," *Future Generation Computor Systems*, vol. 111, pp. 714–722, 2020.

[24] R. E. Bialy, M. A. Salamay, O. H. Karam and M. E. Khalifa, "Feature analysis of coronary artery heart disease data sets," *Procedia Computer Science*, vol. 65, pp. 459–468, 2015.

[25] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan *et al.,* "Heart disease identification method using machine learning classification in e-healthcare," *IEEE Access*, vol. 8, pp. 107562–107582, 2020.

[26] Y. Meidan, M. Bohadana, A. Shabtai, J. D. Guarnizo, M. Ochoa *et al.,* "Profiliot: A machine learning approach for IoT device identification based on network traffic analysis," in *Proc. of the Symp. on Applied Computing*, Marrakech, Morocco, pp. 506–509, 2017.

[27] L. Baccour, "Amended fused TOPSIS-VIKOR for classification (ATOVIC) applied to some uci data sets," *Expert Systems with Applications*, vol. 99, pp. 115–125, 2018.

[28] S. Mohan, C. Thirumalai and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019.

[29] R. T. Selvi and I. Muthulakshmi, "An optimal artificial neural network based big data application for heart disease diagnosis and classification model," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, pp. 1–11, 2020.

[30] C. A. Cheng and H. W. Chiu, "An artificial neural network model for the evaluation of carotid artery stenting prognosis using a national-wide database," in *Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, Seogwipo, South Korea, pp. 2566–2569, 2017.

[31] S. Nazir, S. Shahzad and L. S. Riza, "Birthmark-based software classification using rough sets," *Arabian Journal for Science and Engineering*, vol. 42, no. 2, pp. 859–871, 2017.

[32] A. Methaila, P. Kansal, H. Arya and P. Kumar, "Early heart disease prediction using data mining techniques," *Computer Science and Information Technology Journal*, vol. 5, pp. 53–59, 2014.

[33] C. Yang, B. An and S. Yin, "Heart-disease diagnosis via support vector machine-based approaches," in *IEEE Int. Conf. on Systems, Man, and Cybernetics*, Miyazaki, Japan, pp. 3153–3158, 2018.

[34] H. Y. Lu, "Applying propensity score and support vector machine to construct a predictive model for heart disease," in *4th Int. Conf. on Medical and Health Informatics*, Kamakura, Japan, pp. 18–21, 2020.

[35] M. A. M. Abushariah, A. A. M. Alqudah, O. Y. Adwan and R. M. M. Yousef, "Automatic heart disease diagnosis system based on artificial neural network and adaptive neuro-fuzzy inference systems approaches," *Journal of Software Engineering and Applications*, vol. 7, no. 12, pp. 1055–1064, 2014.

[36] E. O. Olaniyi, O. K. Oyedotun and K. Adnan, "Heart diseases diagnosis using neural networks arbitration," *International Journal of Intelligent Systems and Applications*, vol. 7, no. 12, pp. 72–78, 2015.

[37] O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang and G. Li, "An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction," *Expert Systems with Applications*, vol. 68, pp. 163–172, 2017.

[38] X. Liu, X. Wang, Q. Su, M. Zhang, Y. Zhu *et al.,* "A hybrid classification system for heart disease diagnosis based on the rfrs method," *Computational and Mathematical Methods in Medicine*, vol. 2017, pp. 1–10, 2017.

[39] M. A. Khan, S. Abbas, A. Atta, A. Ditta, H. Alquhayz *et al.,* "Intelligent cloud based heart disease prediction system empowered with supervised machine learning," *Computers, Materials & Continua*, vol. 65, no. 1, pp. 139–151, 2020.