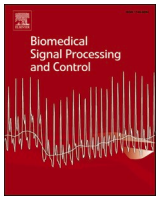
Biomedical Signal Processing and Control 86 (2023) 105129

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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



D-t-SNE: Predicting heart disease based on hyper parameter tuned MLP

Sonam Palden Barfungpa [a,b], Leena Samantaray [c,*], Hiren Kumar Deva Sarma [d], Rutuparna Panda [e,*], Ajith Abraham [f]

[a] Biju Patnaik University of Technology, Odisha, India
[b] Advanced Technical Training Centre, Bardang, Sikkim, India
[c] Ajay Binay Institute of Technology, Cuttack, Odisha, India
[d] Department of Information Technology, Gauhati University, Guwahati, Assam, India
[e] Dept. of Electronics & Telecommunication Engg., VSS University of Technology, Odisha, India
[f] Machine Intelligence Research Labs, Scientific Network for Innovation and Research Excellence, WA 98071-2259, USA

ARTICLE INFO

Keywords:
Data mining
Heart disease prediction
Distributed-t-stochastic neighborhood embedding
Hyper parameter tuned MLP

ABSTRACT

Heart disease has recently become a major cause of high mortality rates. Concurrently, data mining (DM) has also attracted increasing attention in the healthcare field. Identifying this disease in the starting stage helps to minimize treatment costs, thereby saving people's lives. Although several classification models have been applied in recent years, they are deficient in their prediction accuracy. Hence, this research intends to apply DM methods for heart disease prediction by concentrating on maximum accuracy. The proposed scheme is evaluated for the performances in terms of various performance metrics using HD datasets (Statlog + Hungary + Cleveland + Switzerland + long beach VA datasets). Deep Convolutional Neural Network (CNN) models have been proposed to extract relevant features owing to their capability for automatic and effective learning. Subsequently, the fusion was performed. Following this, D-t-SNE (Distributed-t-Stochastic Neighborhood Embedding) is introduced to reduce dimensionality reduction to solve over fitting issues and remove redundant data to improve the classifier performance for predicting heart disease. Furthermore, efficient classification is undertaken by the introduced hyper-parameter-tuned MLP (H-MLP), as it has the ability to solve classification issues. Finally, the proposed work was assessed through comparison with traditional techniques with respect to accuracy, precision, sensitivity, Matthew's correlation coefficient (MCC), F1-score, specificity, and negative predictive value (NPV). The outcomes showed the superior prediction of this system compared to conventional research.

1. Introduction

According to reports by the World Health Organization (WHO), heart disease is the major cause of mortality worldwide taking nearly 17.9 million lives every year. Prediction of this disease is important in its early stages affording appropriate treatments in a timely manner and minimizing the death rate. In recent years, DM techniques have been used to solve several challenges in managing and examining specific data in medical centers [1,2,31]. Several traditional studies have applied various DM methods to predict this type of disease with enhanced accuracy. In addition, the common issues of healthcare centers have been that all experts do not possess equal skill and knowledge for treating patients. They apply their own decision-making which might provide poor outcomes leading to the death of patients. To solve this issue, predicting the presence of this disease through DM methods plays a crucial role in its diagnosis. Hence, a review was undertaken to explore the few ML algorithms utilized to predict heart diseases namely, support vector machine (SVM), artificial neural network (ANN), K-nearest neighbor (K-NN), naïve Bayes (NB), and decision tree (DT). The main benefit of the survey is that it enhances traditional research in making better decisions through the use of feature selection techniques and various algorithms. Appropriate selection of features plays a major role in improving classification accuracy. In addition, dimensionality reduction provides support for improving prediction accuracy [3]. Applying classification techniques to disease datasets provides better results through the development of an intelligent, adaptive, and automated system for predicting heart diseases [432].

Traditional research has attempted various approaches to diagnose heart disease. Accordingly, a hybrid random forest with linear model (HRFLM) has been endorsed to find significant features through the

* Corresponding authors.
E-mail addresses: leena.samantaray@abit.edu.in (L. Samantaray), rpanda_etc@vssut.ac.in (R. Panda).

https://doi.org/10.1016/j.bspc.2023.105129
Received 24 January 2023; Received in revised form 3 May 2023; Accepted 8 June 2023
Available online 16 June 2023

employment of ML methods, leading to enhanced prediction accuracy. Various feature combinations and renowned classification methods have been presented with an accuracy of 88.7% accuracy [5]. To enhance the model performance, it is vital to select correct and significant feature combinations. Thus, significant features were identified and DM methods were used to improve the prediction rate. Seven classification methods, DT, NB, K-NN, SVM, NN, LR (logistic regression), and vote (NB + LR), were applied. The empirical outcomes revealed that the prediction of heart disease through voting accomplished 87.4% accuracy [6,7]. Three DM classification algorithms, RF (Random Forest), NB, and DT were addressed and utilized for developing a prediction system to analyze heart disease probability. This study helps to find an effective classification method suitable for achieving high accuracy. The RF algorithm showed better performance, with an 81% precision rate [8].

Although conventional research aimed to effectively perform heart disease prediction, it lacked relevant feature extraction and dimensionality reduction, which eventually impacted the accuracy rate. Thus, an effective technique using DM methods to enhance the detection rate is vital for diagnosing this disease. Hence, this research introduces distributed t-stochastic neighbor embedding (D-t-SNE) to reduce dimensionality as it is capable of preserving global and local structures. It also helps to easily comprehend high-dimensional data and project them into a low-dimensional space, which makes it valuable when handling CNN networks. The present work performs feature extraction through two deep CNNs, as it is computationally efficient and possesses several layers. It can also learn several parameters required to solve the challenges that enable the efficient retrieval of relevant features. Furthermore, the hyper parameters are also vital as they manage the model's overall behavior. The primary aim of this process is to discover an optimal hyper parameter combination that reduces the predefined loss function to obtain better outcomes. In addition, the adaptive learning ability of MLP and the suitability of this algorithm for solving classification issues have made this study propose a hyper-parameter-tuned MLP (H-MLP). Owing to these merits, the proposed methods are expected to yield better predictions which are confirmed through the results.

The major contributions of this study are listed below:

- Feature extraction and feature-level fusion were performed through introduced deep CNN models to select only significant and relevant features.
- Accomplished dimensionality reduction by the proposed D-t-SNE (Distributed-t-Stochastic Neighbour Embedding) for enhancing the classifier performance.
- Classified the absence and presence of the disease through the introduced hyper-parameter-tuned multilayer perceptron (H-MLP) so as to attain better accuracy.
- Evaluated the effectiveness of the proposed system in prediction of heart disease through comparative analysis with regard to accuracy, precision, sensitivity, F1-score, Matthew's correlation coefficient (MCC), specificity, and negative predictive value (NPV).

The paper is organized as follows. Section 1 explores WHO reports on heart disease prediction, various approaches used by the existing system, the main problems faced by conventional works and the methods proposed to solve these issues. Section 2 reviews the existing research to highlight the methods used and the obtained results. This section presents the problems encountered during this analysis. Section 3 comprehensively describes the main ideas of the proposed system with a flow chart, architecture, algorithm, and pseudo code. The results are discussed in section 4. Finally, the overall research is concluded in Section 5, along with the future scope of this work.

2. Related work

Various DM methods have been used by different traditional investigators for heart disease prediction. These methods were then analyzed and explored. The significant and general issues identified in this analysis are also presented in this section.

2.1. Feature extraction for predicting heart disease

An approach has been endorsed for cardiac arrhythmia detection in ECG signals by focusing mainly on feature extraction and classification. Accuracy, specificity, computational cost and sensitivity were the evaluation metrics employed for the analysis. Higher-order statistics (HOS), structural co-occurrence matrix (SCM), Goertzel, and Fourier have been the methods used for feature extraction. In addition, SVM, multilayer perceptron (MLP), optimum path forest (OPF) and Bayesian methods have been used as classifiers. These methodologies were tested and compared with six conventional feature-extraction techniques. Accuracy has been found to be 94.3% [9]. Medical test results have also been used as inputs for extracting features with minimum dimensions to afford better heart disease diagnosis. Suggested system extracts high influencing new projection features through Probabilistic Principal Component Analysis (PPCA). Feature subclasses with minimized dimensions have been afforded to radial basis function (RBF)-based SVM for classification that achieved 82.18% accuracy for the Cleveland dataset, 85.82% for the Hungarian dataset and 91.30% accuracy for the Switzerland dataset [10,11]. To further enhance the system, neural network-feature correlation analysis (NN-FCA) was used in two phases. The initial stage is feature selection, which makes features agree with the significance of predicting the risk of heart disease. The subsequent stage is FCA, in which learning occurs about the prevalence of correlations among feature relations and data of individual NN predictor outcomes are found. The recommended system has been found to be better than the Framingham Risk Score (FRS) with respect to risk prediction of this disease [12]. Similarly, different approaches have been used by conventional works; accordingly, the feature selection technique called incremental feature selection algorithm (IFSA) has been endorsed as an integrated intelligent conditional random field (ICRF) concept, ICRF-LCFS (ICRF-Linear Correlation-coefficient based Feature Selection) and T-CNN (traditional CNN with temporal features). The recommended system for predicting the disease has been assessed to achieve better accuracy with a minimum false alarm rate (FAR). T-CNN has also been valuable for improving the level of performance with respect to prediction accuracy above 93% [13]. Thus, feature selection among datasets has been the main factor that affects prediction accuracy. The MCC was also considered in this study. Modified particle swarm optimization (PSO) was employed to select suitable attributes. An enhanced fuzzy ANN was used for the prediction. The maximum prediction rate was 88.82% for male patients and 88.05% for female patients [14].

2.2. Classification for predicting heart disease

Hybrid classification systems have been suggested that rely on the relief and rough set (RFRS) for diagnosing heart disease. The jackknife cross-validation method has been accomplished that showed a 92.59% accuracy [15]. Similarly, an effective decision-making system in medicine is vital. Thus, twin SVM has been recommended for predicting the absence or presence of heart disease. This method discovers dual non-parallel and hyper-planes, for which each one has been identical to the initial class and is far from the subsequent class as probable. Experiments were conducted on a real-time dataset and an evaluation was performed that exposed an 86.75% accuracy rate [16,17]. To further improve accuracy, an enhanced deep learning-assisted convolutional neural network (EDCNN) has been endorsed for supporting and enhancing heart disease prognostication. The performance of the system was assessed based on the overall and reduced features. It has been found that feature reduction affects classifier efficiency with respect to accuracy and processing time [18]. To enhance the system, a DM algorithm has been proposed. Many prediction studies have utilized complex

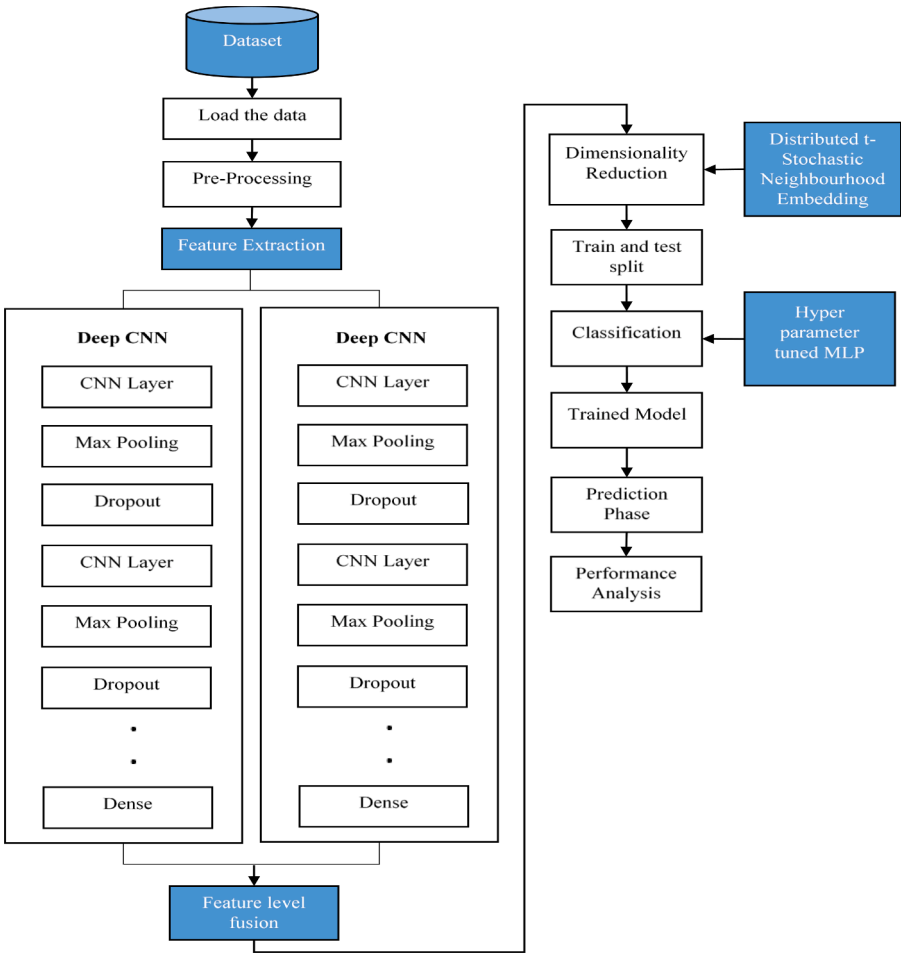


Fig. 1. Proposed flow for predicting heart disease.

patient data, such as biomarkers, pathological measurements and biomedical images. Method similar to language model has been demonstrated to predict high risk prognostication from patient's history through deep RNNs (Recurrent Neural Networks). Mentioned system utilized multiple RNNS to learn from the prediction of patient's code sequences for identifying diseases of high risk [19]. Finally, the empirical outcomes revealed that the endorsed technique could achieve better outcomes. In contrast, gated recurrent units (GRUs) have been suggested to predict heart failure. Compared to renowned techniques such as LR, MLP, K-NN, and SVM, the GRU models exhibited better performance. Through analysis, the significance of the medical record sequence has been discussed [20]. Different methods have shown varied classification outcomes. Correspondingly, ANN showed 97% accuracy, CART showed 87.6% accuracy, NN exhibited 97.4% accuracy, SVM exhibited 95.6% accuracy, LR explored 72% accuracy, and multi-criteria oriented decision analysis and genetic algorithms showed 91% accuracy [21]. Although better outcomes were attained by each method, the NN showed a high accuracy rate of 97.4%. However, the accuracy must be enhanced further to predict the disease correctly.

2.3. Other methods for predicting heart disease

Ensemble classification has been analyzed and utilized to enhance the accuracy of weak algorithms through the incorporation of several classifiers. Experiments were conducted using data and comparative analysis to determine the degree to which this method could be employed to enhance the accuracy in predicting heart disease. The recommended model, namely majority-vote with NB, RF, MLP and BN (Bayes Net) showed 85.48% accuracy [22]. To remove irrelevant features, the X^2-statistical model has been endorsed and a deep neural network (DNN) has been searched through an exhaustive search method. The efficiency of this model was determined through a comparative analysis with the traditional DNN and ANN models. Accuracy has been found to be 93.33% [23]. Through these outcomes, it was revealed that identifying important features and efficient DM methods could enhance the prediction rate. Thus, nine classifiers were applied: DT, LR (logistic regression), Adaboost (adaptive boosting), SGD (stochastic gradient descent), RF, GBM (gradient boosting), SVM, G-NB (Gaussian-Naïve Bayes) and ETC (extra tree classifier). The imbalance class issue was addressed through the synthetic minority oversampling technique (SMOTE). Moreover, ML models were trained on maximum-ranked features using RF.

The outcomes were compared with those of traditional ML algorithms using a full feature set. Empirical outcomes showed that ETC performed better than other models, with 92.62% accuracy in predicting the survival rate of patients with heart disease [24,25]. In addition, a hybrid OFBAT-RBFL (oppositional firefly with BAT and rule-based fuzzy logic) has been designed to diagnose heart disease. Initially, relevant features were chosen from the dataset through locality preserving projection (LPP), which assists the diagnosis system in developing classification models using fuzzy logic. Fuzzy rules were then created from the data sample. Among the overall rules, the significant and associated rules were selected using the OFBAT algorithm. Subsequently, a fuzzy system was designed with the assistance of framed membership functions and fuzzy rules, for which classification could be undertaken with the designed fuzzy system. Finally, the experiment was conducted using publicly accessible UCI datasets from Hungarian, Switzerland, and Cleveland. The results showed that the proposed system performed

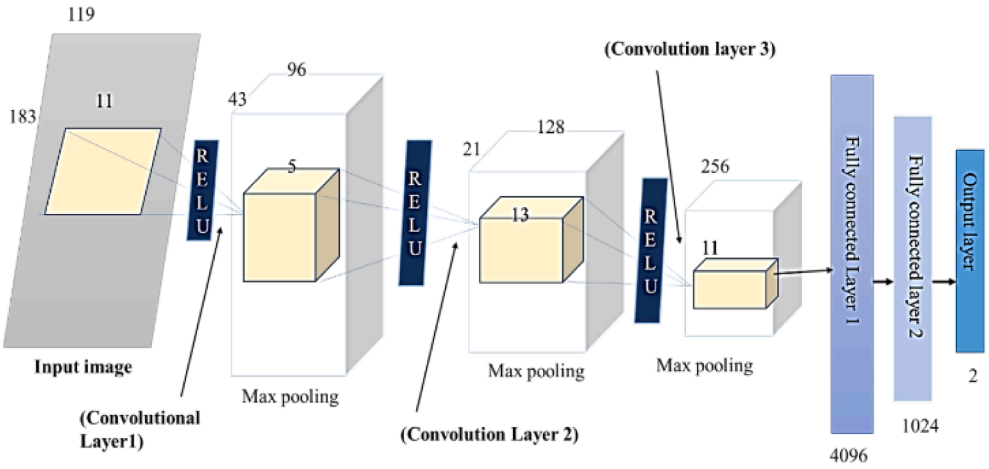


Fig. 2. Architecture of Deep CNN.

better than a conventional system with 78% accuracy [26,27].

2.4. Problem identification

Various problems identified through the analysis of above existing works are listed below.

- Traditional studies have attempted to attain better accuracy in predicting heart disease by using various methods. Accordingly, the OFBAT-RBFL method explored 78% accuracy [26], and PSO with an improved fuzzy ANN showed 88.82% for male patients and 88.05% for female patients [14]. In addition, feature extraction by Parallel Probabilistic PCA (PPPCA) explored 82.18%-Cleveland dataset, 85.82%-Hungarian and 91.30% for Switzerland dataset [10]. Moreover, RFRS showed 92.59% accuracy [15], and a novel feature extraction method to classify cardiac arrhythmia explored 94.3% accuracy [9].
- Effective algorithms have to be planned for selecting significant features for discovery, which is highly suitable for classification [16].
- A feature fusion technique was applied [7]. However, the performance of this process must be improved using DM methodologies.

3. Proposed methodology

The study mainly concentrated on predicting heart disease as it has been the reason for many deaths. The selection of relevant and significant features, effective dimensionality reduction and enhanced accuracy have been major challenges in traditional methods. Hence, this study aims to solve these problems by relying on DM methods that encompass ML and DL for feature extraction and classification. The overall flow of the proposed system is shown in Fig. 1. Several processes are involved in disease prediction. First, the Cleveland dataset is loaded. After this preprocessing is performed which utilizes data cleaning to eliminate unrequired data, thereby enabling the user to attain a dataset possessing useful information? Then, feature extraction and fusion are performed by the introduced deep CNN models that assist in finding a compact and informative set of features to enhance the reliability and efficiency of the classifier. Subsequently, dimensionality reduction is performed using the proposed D-t-SNE. This process minimizes the data storage space because a reduction in the dimensions is undertaken. This also assists in minimizing the computation or training time and helps to visualize the data. Then, it was fed into the train and test splits. Subsequently, the classification process is accomplished through the proposed H-MLP, which supports the differentiation of the presence or absence of the disease. Finally, the prediction is achieved through a trained model. The efficiency of the system was evaluated by performing performance analysis.

Table1
Feature Extraction through Deep CNN.

Extracted Features	Deep CNN 1	Deep CNN 2	Total	Dimensionality Reduced
	100	500	600	10

3.1. Feature extraction and fusion: Deep convolutional neural network

Generally, a CNN is a type of feed-forward artificial neural network (FFANN), which is biologically stimulated through visual cortex organization. These are extensively employable in various areas such as image and video recognition, Recommender System (RS) and Natural Language Processing (NLP). A CNN encompasses two main parts: convolutional and max-pooling layers. Moreover, the convolutional layer provides a feature map as the output through computation of the dot product consisting of the local region of input feature map and filter. The nonlinear function then estimates the complex functions by squashing the output of the NN. In addition, the pooling layer accomplishes downsampling for the feature map by calculating the maximum or average value on the sub-region. Fully connected (FC)layers follow stacked and pooling layers and Softmax is the last FC layer that calculates the scores for each class. Overall, the deep CNN and the CNN were identical. However, a deep CNN comprises of several layers. The CNN consists of basic parts for feature extraction and classification. In this study, a deep CNN was used to extract features. Owing to the numerous layers and computational efficiency of CNN, it learns various parameters than are needed to solve the challenges that eventually enhance efficiency. The deep CNN architecture is shown in Fig. 2 consisting of three convolutional and FC layers.

For the first convolutional-layer agrresing toe structure of CNN, images possessing 183 × 119 pixels as the input are given into a convolutional-layer having 96-filters exploring 11 × 11 pixels as the size with pixels (2 × 2) as the stride. Then, 96 feature maps were fed to the max-pooling layer to obtain a robust CNN structure to translate the image. Therefore, the initial layer output comprised 96 feature maps with a size of 43 × 27 pixels. The subsequent layer was positioned following the starting layer to fine-tune it, which had 128 filters with a size of 5 × 5 × 96. Subsequently, additional max-pooling occurs. By starting two layers, 128 feature maps were obtained with a size of 21 × 13 pixels. The first two layers are used to extract low-level features from the image. To extract high-level features, three additional convolution layers were utilized, as shown in Fig. 2. The third layer has 256 filters with a size of 3 × 3 × 128 that are later fed to 3FC layers encompassing

neurons with orders 4096, 1024, and 2. Following this, features are utilized at the 2nd FC layer to extract the features of the image. Hence, the feature vector can be extracted, comprising 1024 components for each image.

Thus, the deep CNN 1 model extracts 100 features, whereas deep CNN 2 extracts 500 features, as shown in Table1. In total, 600 features were extracted. When all of these features are fed into the classification stage, more time is consumed which degrades the performance of the classifier. For this purpose, dimensionality reduction was performed, where all of these features were reduced to 10 based on their significance and relevance.

3.2. Dimensionality Reduction: Distributed-t-Stochastic Neighborhood Embedding

This research proposes D-t-SNE for reducing nonlinear and high-dimensional data. It accomplishes dimension minimization through projection of a Gaussian distribution corresponding to a high-dimensional spatial neighborhood to a low-dimensional D-t-SNE. This algorithm can efficiently capture a significant portion of a high-dimensional local data structure. It also explored the global structure at various scales. To preserve the similarity among high-dimensional data and map it into the low-dimensional space, this algorithm converts the distances among actual data points into Gaussian joint probabilities (GJP) through computation of pairwise similarity among these data points. The stepwise process involved in this process is discussed below:

Considering a high-dimensional dataset $A = \{f_1, f_2, \cdots\cdots f_n\} R^D$, the conditional probability $S_{b|a}$ of data-point f_b to data-point f_a is given by (1).

$$S_{b|a} = \frac{exp(-||f_a - f_b||^2/2\sigma_a^2)}{\sum_{k\neq a} exp(-||f_a - f_k||^2/2\sigma_a^2)} \tag{1}$$

In Eq. (1), σ_a represents the Gaussian variance, which is centered on the data point f_a. When $S_{a|a} = 0$, the joint probability S_{ab} in high-dimensional space is given by (2).

$$S_{a,b} = \frac{S_{b|a} + S_{a|b}}{2n} \tag{2}$$

A low-dimensional dataset is given by $B = \{b_1, b_2, \cdots, b_n\} R^D$. Similar to Equation2, when σ_a of conditional probabilities $q_{j|i}$ is $\frac{1}{\sqrt{2}}$, then joint probability $Q_{i,j}$ in low-dimensional space is given by (3).

$$Q_{i,j} = \frac{\left(1 + ||B_a - B_b||^2\right)^{-1}}{\sum_{k\neq l}\left(1 + ||B_k - B_l^2\right)^{-1}} \tag{3}$$

To make low-dimensional space possess a similar joint-probability distribution along with high-dimensional data, the introduced D-t-SNE intends to discover B which reduces the mismatch between P and Q. Similarity evaluation among P and Q could be computed through Kullback Leibler divergences among low-dimensional distributions and high-dimensional data. Moreover, the loss function Q is given by (4).

$$Q(B_1, B_2, \cdots\cdots B_n) = \sum_a KL(P_a \| Q_a) = \sum_a \sum_b P_{ab} log \frac{P_{ab}}{Q_{ab}} \tag{4}$$

The objective function Q is reduced by the gradient descent. Thus, the gradient of D-t-SNE is given by (5).

$$\frac{\partial Q}{\partial B_a} = 4\sum_b (P_{ab} - Q_{ab})(B_a - B_b)\left(1 + ||B_a - B_b||^2\right)^{-1} \tag{5}$$

From Eq. (5), the update rule below is derived and is shown in (6).

$$Y^{(e)} = Y^{(e-1)} + n(\partial C)/\partial Y + \alpha(t)B^{(e-1)} - B^{(e-2)}) \tag{6}$$

The overall algorithm for dimensionality reduction using D-t-SNE is presented in Algorithm I.

Algorithm I: Distributed-t-Stochastic Neighbourhood Embedding

Input the dataset: $A = \{f_1, f_2, \cdots\cdots f_n\} \in R^D$, perplexity (perp), number-of-iterations (T), learning-rate (n) and momentum (a(t))
Begin
Step 1: Calculate high dimensional similarities having perplexity (perp) through equation1.
Step 2: Construct the matrix P through equation2.
Step 3: Initialise $B^{(o)}$ from $S\left(0, 10^{-4}\right)$
Step 4: for t = 1 to T do
Step 5: Calculate low dimensional similarities $Q_{a,b} = \frac{\left(1 + ||B_a - B_b||^2\right)^{-1}}{\sum_{k\neq l}\left(1 + ||B_a - B_b||^2\right)^{-1}}$ by equation3.
Step 6: Construct the matrix Q
Step 7: Calculate gradient $\frac{\partial C}{\partial B}$ by equation5.
Step 8: Set $B^e = B^{e-1} + n\frac{\partial C}{\partial B} + \propto(t)\left(B^{(e-1)} - B^{(e-2)}\right)$
Step 9: End
Step 10: Output: Low dimensional dataset $B^{(T)} \in R^d$

Initially, the dataset was used as the input. The high-dimensional similarities that possess perplexity are then given in Step 1. This matrix was constructed based on Eq. (2). Subsequently, Steps 3 and 4 were performed. Following this, low-dimensional similarities were computed based on Equation3. Then, the matrix Q is constructed. Following this, the gradient was computed based on Equation5. Finally, steps 8 and 9 were performed to obtain a dataset with a low dimension.

3.3. Classification: Hyper parameter tuned MLP

MLP is a type of feed forward NN (FFNN), which transfers information in one-way through NNs and its respective neurons are systemized in various parallel layers. An initial layer exists in the input layer in the various parallel layers. The final layer is termed the output layer, whereas the intermediate layers indicate hidden layers. When FFNNs are hidden later, this is called MLP. Moreover, the hypothesis space indicates four-dimensional spaces encompassing several weights (i.e., the weight vector group). The delta rule (gradient descent) is selected as the training rule for finding the weight vector $\vec{w}$ which fits best with the training instances and the search approach within the hypothesis space is for discovering $\vec{w}$ which has the ability to reduce the training error (E) for all the training instances. In accordance with the general definition of the training error, the hypothesis is computed as per (7).

$$E(\vec{w}) = \frac{1}{2}\sum_{d\in D}(e_d - o_d)^2 \tag{7}$$

In Eq. (7), D indicates the training instance set, e_d represents the targeted result for a particular training instance, o_d and d represent the network results for the training instance. This was then customized into (8).

$$E(\vec{w}) = \frac{1}{2}\sum_{d\in D}[(e_r - o_d) + (e_d - o_d)]^2 \tag{8}$$

In Eq. (8), e_d represents the maximum values for columns j and row i for afforded cells i, j in the similarity matrix. The overall algorithm of the hyper-parameter-tuned MLP is shown in Algorithm II.

Algorithm II: Hyper parameter tuned MLP

Input : The original similarity matix M, between two ontologies/schema and a set of training samples
Output: Learned weight vector $\underset{w}{\rightarrow}$
Step 1: Intialisation of $\underset{w}{\rightarrow}$: $T_a \leftarrow 0.25$;
Step 2: For a←1 to a predefined iteration number do
Step 3: Save training instances to a temporary variable;
$\Delta T_a \leftarrow 0$;

(*continued on next page*)

(*continued*)

Algorithm II: Hyper parameter tuned MLP
Step 4: While Training instances are not empty do d ← Get Current Training Instances (); r←Obtain Row Number In Matrix(d); c←Obtain Column Number In Matrix(d) o_d←Calculate Network Output(d); e_r←FindMaximum similarity In Column(c) $\Delta T_a \leftarrow \Delta T_a + n[(e_r - o_d) + (e_c - o_d)]s_{id}$ **Step 5:**Remove Current Training Instances (); End **Step 6:**$T_a \leftarrow T_a + \Delta T_a$; **Step 7:**Restore training instances from temporary variables End **Step 8:**Output updated ($\vec{w}$)

Initially, the similarity matrix and training sample sets were used as the inputs. Then, initialization followed by Step 2 was performed. Subsequently, training instances were saved as temporary variables. If the training instances are non-empty, Step 4 is undertaken. Current training instances were removed. Finally, the training instances were restored from the temporary variables to obtain the learned weight vector. In addition, various parameters were considered by H-MLP and are shown in table-2.

From Table 2, the number of inputs is 10; thus, the number of selected parameters is 10. Then, the number of hidden layers was [10,20,30]. In this case, 30 is selected as the optimal parameter. Similarly, the number of hidden units in the initial hidden layer ([50, 70, 100]) and ([50, 80, 100]) was 100. Subsequently, the best parameter of the hidden units in the second hidden layer ([100,150,200]) and ([100, 150, 200]) is 150. The number of outputs was considered to be one. The selected parameters of the connection weight ([20, 30, 50]) and ([40, 60, 80]) were 30.

4. Results and discussion

The results obtained from the execution of the proposed system are discussed in this section. The considered dataset, performance metrics, experimental results, and comparative analysis are also presented. The analysis confirms the efficiency of the proposed system compared with the conventional system in predicting the presence (affected) or absence (unaffected) of heart disease.

4.1. Dataset description

The research uses the HD datasets (Hungary + Cleveland + Switzerland + VA long beach datasets) to predict the presence or absence of heart disease that encompasses 11 features. To date, this database has been used by various ML investigators. This HD dataset has been presented by integrating various datasets. These datasets are already available individually and are not combined before. When theses datasets are integrated, the resultant is the largest HD dataset which is accessible for research purposes. The HD dataset (Cleveland + Switzerland + Hungary + VA long beach datasets) comprised of records of patients from the US, Hungary and Switzerland. It encompasses 11 different major features.

Some features of the HD dataset are enlisted in Table 3.

Table 2
Parameters considered by H-MLP.

Network name	Incremental back propagation	Batch back propagation	
Network topology			
Network type	feed-forward fully connected network	feed-forward fully connected network	Selected Parameters
No. of inputs	10	10	10
No. of hidden layers	[10,20,30]	[10,20,30]	[10,20,30]
Hidden units in the 1st hidden layer	[50,70,100]	[50,80,100]	100
Hidden units in the 2nd hidden layer	[100,150,200]	[100,150,200]	150
No. of outputs	1	1	1
Connection weight	[20,30,50]	[40,60,80]	30

Table 3
Features of Dataset.

Features	Depiction
Gender/ sex	Female is represented as 0Male is represented as 1
trestbps	RBP: Resting Blood Pressure on the admittance in hospital
age	Age: in years
cp	Type of the chest painTypical angina: value 0Atypical angina: value 1Non angina: value 2Asymptomatic: value 3
ST slope	Peak-slope employing ST-segmentUp-slopping: value 0Flat: value 1Down-slopping: value 2
Rest ECG	Outcomes of the resting electrocardiographsNormal: value 0Possessing an ST-T wave abnormality (elevation of ST or the depression rate > 0.05 mV): value 1Finding possible or defined LVH (Left Ventricular Hypertrophy) through estes' criteria: value 2
chol	Serum cholesterol in mg/dl
fbs	Blood-sugar > 120 mg/dlTrue: 1False: 0
oldpeak	ST depression prompted relying on rest
thalach	HRR: High Heart Rate accomplished
exang	Utilise prompted anginaYes: 1No: 0

4.2. Performance metrics

The introduced work was assessed by a comparative analysis with respect to accuracy, sensitivity, Matthew's correlation coefficient (MCC) precision, F1-score, specificity, and negative predictive value (NPV) to prove the effectiveness of the proposed system for heart disease prediction.

Accuracy

It is defined as the proportion of samples that are correctly classified into the overall samples and is indicated by (9).

$$Accuracy = \frac{TN + TP}{TP + TN + FP + FN} \tag{9}$$

Sensitivity

It is defined as the proportion of positive samples that are correctly classified to the overall positive instances and is indicated by (10).

$$Sensitivity = \frac{TP}{TP + FN} \tag{10}$$

5. Precision

It is defined as the computation of the count of accurate positive detects to the overall count of positive prediction, and is given by (11).

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

Specificity

It is defined as the proportion of samples that are correctly classified as negative examples to the overall negative samples, and is indicated by (12).

$$Specificity = \frac{TN}{TN + FP} \tag{12}$$

F1-Score

It is also known as the F-measure which is defined as the harmonic mean of the precision and recall. It was used to determine measurement

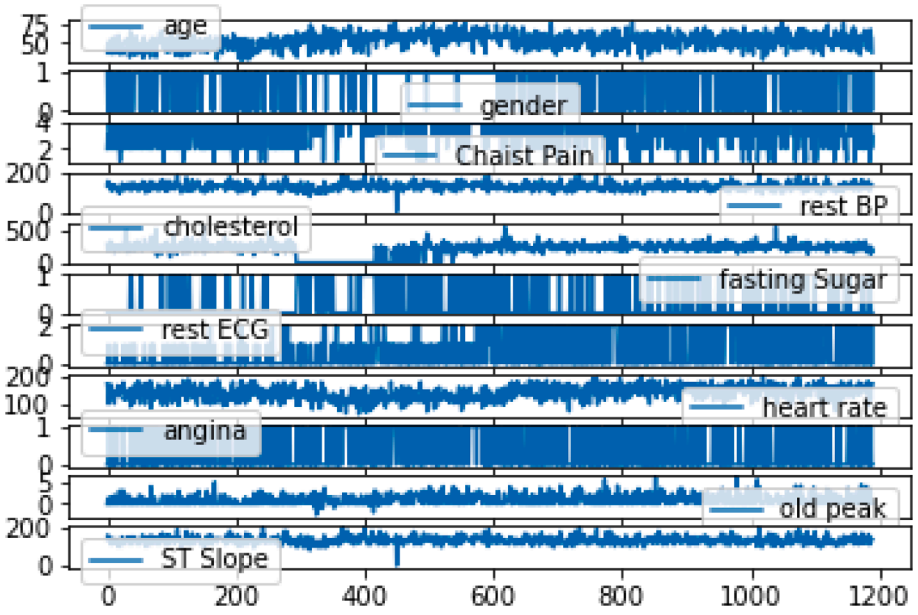


Fig. 3. Individual Feature Representation.

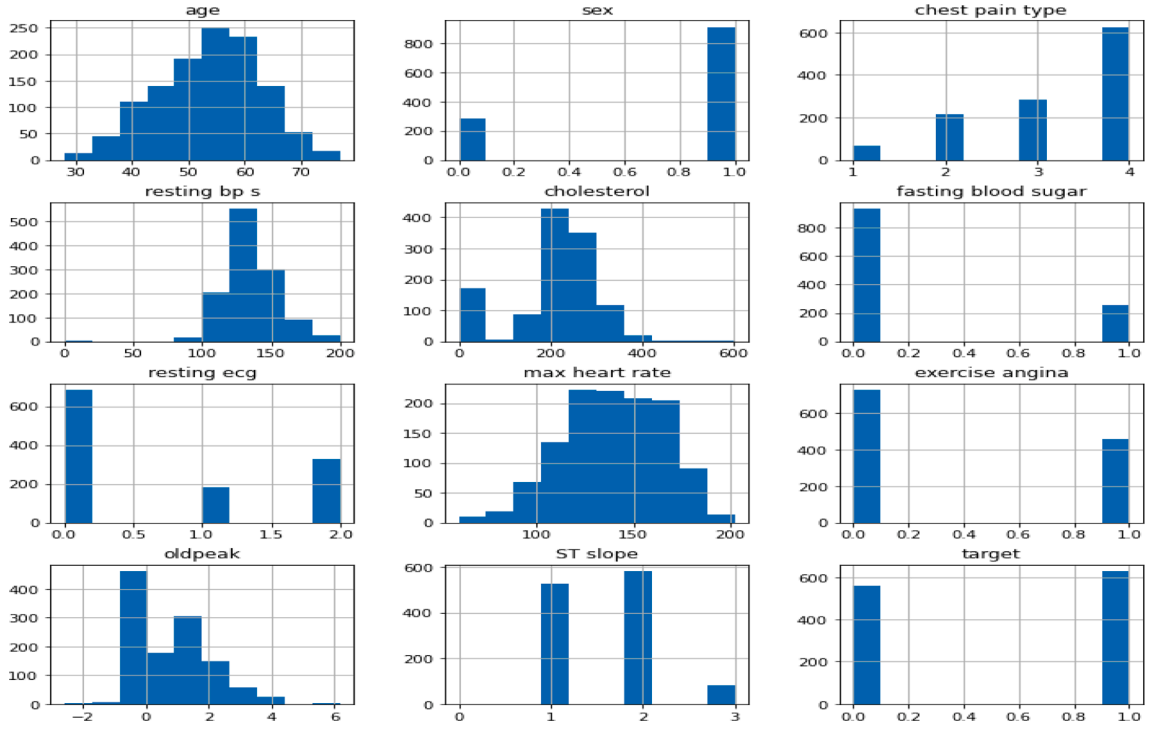


Fig. 4. Bar-plot for representing different dataset features.

efficacy. F1-score ranges between 0 and 1, where the worst value represents 0 and the best value is 1. This is given by (13).

$$F1 - Score = \frac{2(TP)}{2TP + FN + FP} \tag{13}$$

MCC (Matthews correlation Coefficient)

The correlation coefficient between the perceived and detected binary classifications ranges between 1 and −1, where −1 indicates the worst performance and + 1 represents the best performance. This is given by (14).

$$MCC = \frac{(TP*TN) - (FP*FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{14}$$

NPV (Negative predictive Value)

It is the proportion of accurately classified positive samples to the overall identified negative samples and is given by (15).

$$NPV = \frac{TN}{TN + FN} \tag{15}$$

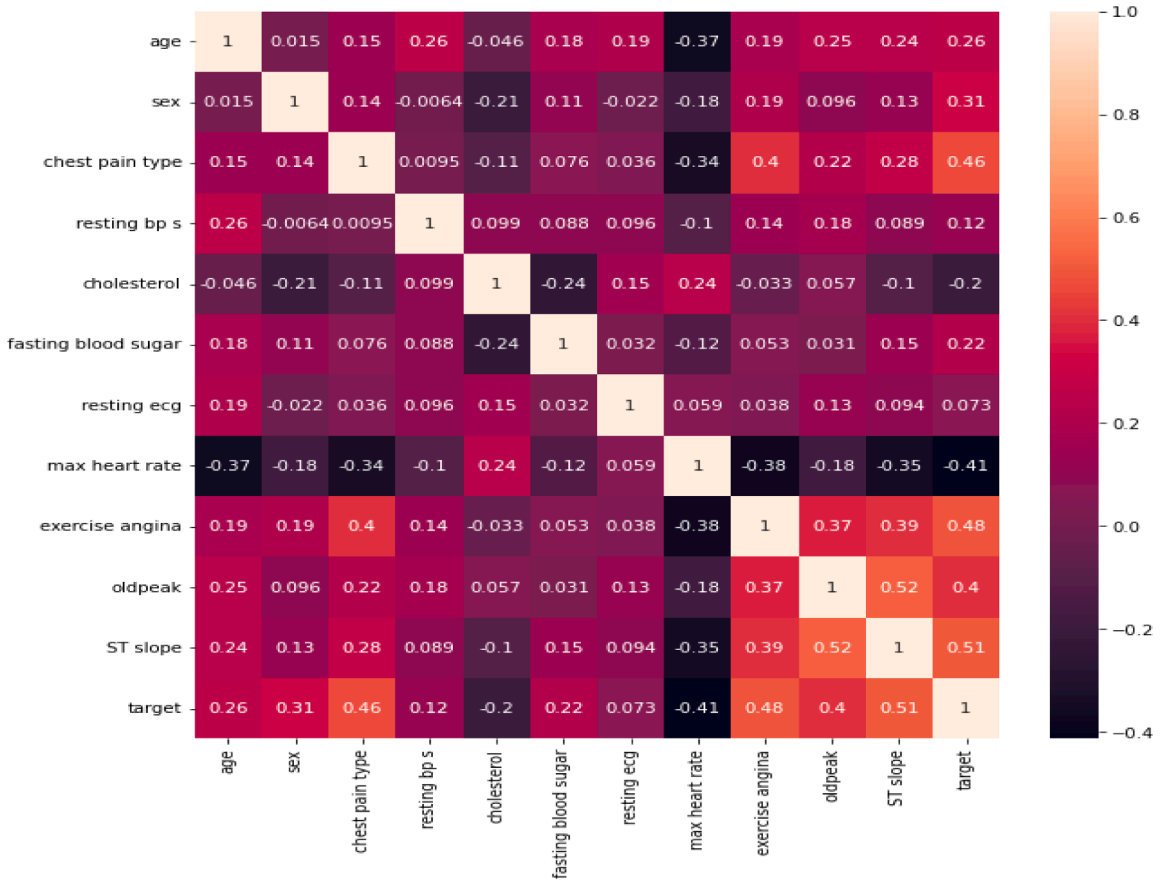


Fig. 5. Correlation plot.

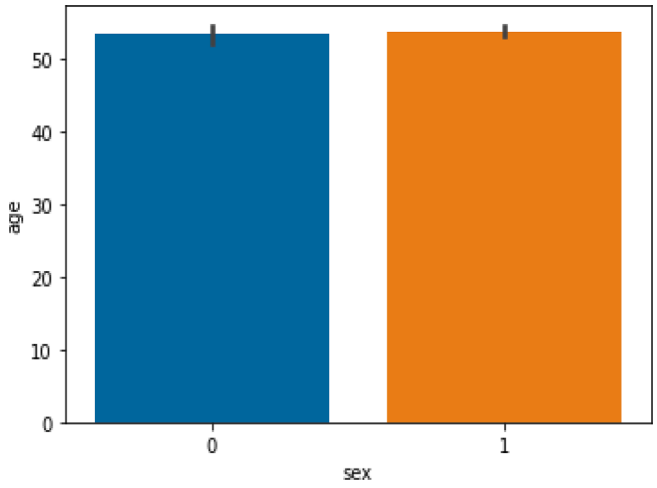


Fig. 6. Sex vs age Count plot.

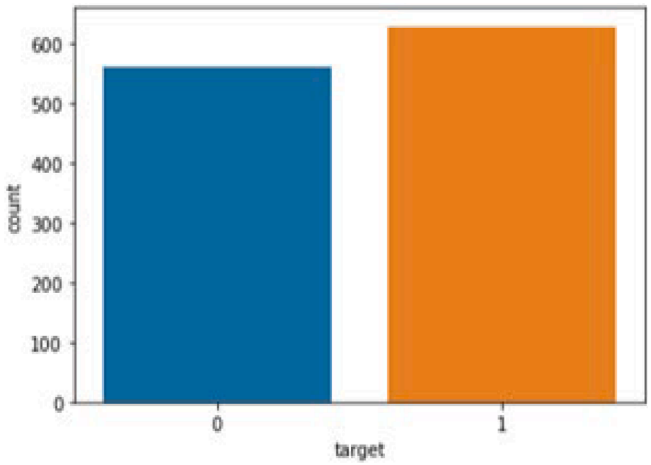


Fig. 7. Target vs Count plot.

5.1. Experimental results

The results obtained after implementation of the proposed system is discussed in this section.

5.1.1. Representation of each feature of the dataset

The heartbeat signals obtained after implementation for each feature set (age, gender, cp, rest BP, cholesterol, fasting sugar, rest ECG, heart rate, angina, old peak, and ST slope) of the HD dataset are shown in Fig. 3.

In addition, a bar plot for individual features is presented in Fig. 4 to provide a comprehensive view of the empirical outcomes.

5.1.2. Correlation matrix

In general, the correlation matrix indicates a table that explores correlation and is best utilized for variables that explore linear relationships among one another. The best data fit is visually indicated by the scatter plot shown in Fig. 5. This matrix comprises rows and columns that explore the variables (features) in the dataset.

5.1.3. Plots for varied features

The number of diseases was plotted by considering sex with respect

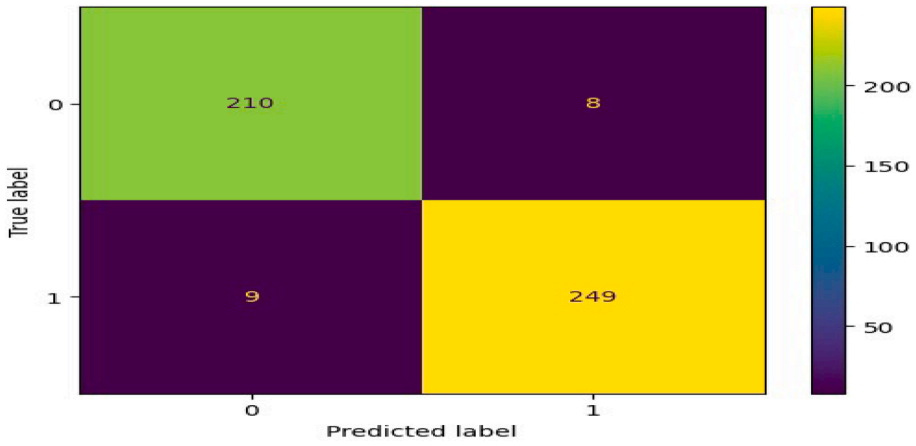


Fig. 8. Confusion Matrix.

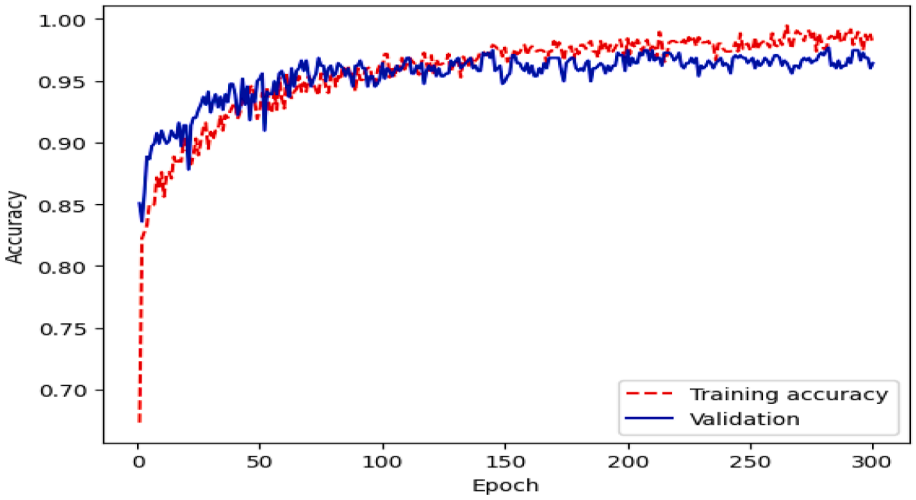


Fig. 9. Performance analysis with respect to model accuracy.

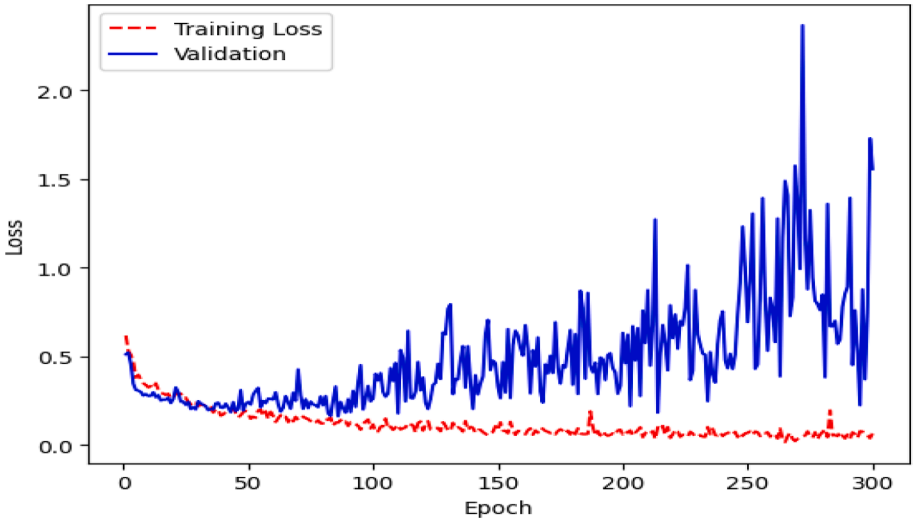


Fig. 10. Performance analysis with respect to model loss.

to age, as shown in Fig. 6. Both sexes were found to be equally affected by the disease. In addition, the number of patients with and without diseases were computed and are plotted in Fig. 7. From Fig. 7, it can be seen that the absence of disease is above 500, and the presence of the disease is nearly 600.

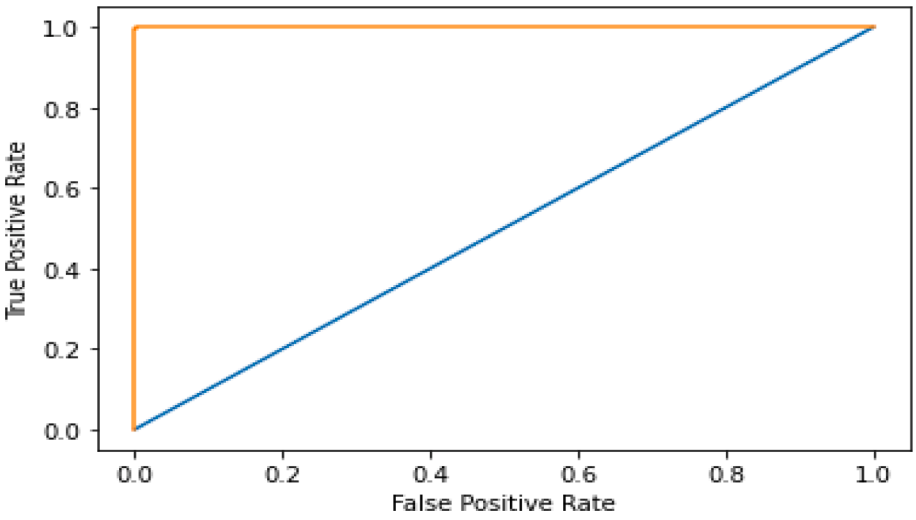


Fig. 11. Performance analysis with respect to ROC.

Table 4
Analysis of existing [28] and proposed system in terms of accuracy.

Methods	Accuracy (%)
SVM	91.18
Logistic Regression	86.13
Decision Tree	90.34
Naïve Bayes	85.71
Random Forest	94.96
KNN	88.66
Proposed	96.68

5.1.4. Confusion matrix

The confusion matrix indicates a table that is often utilized to describe the classifier's performance on the test dataset for which true values are known. Accordingly, confusion matrix for the proposed classifier is plotted and the obtained results are shown in Fig. 8.

Fig. 8 shows that the proposed model correctly predicted 210 normal cases and 249 attack cases. In contrast, it misclassified one case of a normal patient as an attack. In addition, 6 attack patients were misclassified as normal. Since the correctly classified rate was higher than the misclassified outcomes, the proposed system was found to be efficient in predicting the disease.

5.1.5. Performance analysis

The performance of the proposed system was analyzed with respect to the model accuracy, loss, and receiver operating characteristic (ROC). The analytical outcomes of the model accuracy are shown in Fig. 9, and the results of the model loss are shown in Fig. 10.

From the analytical results, it was found that the accuracy of the trained model was enhanced for each epoch (as shown in Fig. 9). However, the loss of the trained model was found to be minimal for individual epochs (as shown in Fig. 10). The high accuracy and low loss of the introduced system indicate the efficiency of the introduced system. The analytical outcomes in terms of ROC are shown in Fig. 11.

Hence, the analysis of the performance of the proposed system with respect to accuracy, loss and ROC confirms its effectiveness. The proposed H-MLP could solve classification issues, therefore, it affords better outcomes.

5.2. Comparative analysis

The proposed system is compared with conventional systems to evaluate the efficacy of the introduced system compared to traditional systems. SVM, LR, DT, NB, RF, and KNN are the existing studies considered for analysis. The results are shown in Table 4.

From the results, it was found that various systems had different accuracies. Accordingly SVM (91.18%), LR (86.13%), DT (90.34%), NB (85.71%), RF (94.96%), and KNN (88.66%), whereas the proposed system showed 96.68% accuracy rate. The results are presented in Table 4.

From the results, it is evident that the proposed system has a higher accuracy rate (96.68%) than that for conventional systems for predicting heart disease. In addition, various other parameters, such as sensitivity, specificity, precision, F1-score, and MCC were considered for the analysis. The results are presented in Table 5. This is shown graphically in Fig. 12.

The results showed that the accuracy of the proposed system was 96.68%, sensitivity rate was 96.79%, specificity was 96.51%, precision rate was 97.27%, F1-score was 96.89%, and MCC exposure was 93.24%. The outcomes of the conventional methods are lesser than those of the proposed system. For instance, the KNN exhibited 88.66% accuracy (Table 5). However, it is lesser than that of the introduced system that explores the efficacy of the proposed method. Another comparative analysis was conducted by considering different existing methods. The majority vote with NB, BN (Bayes Net), RF and MLP, linear SVM + linear

Table 5
Analysis of existing [28] and proposed system in terms of various performance metrics.

Comparative Analysis							
Performance measurement parameters	SVM	LR	DT	NB	RF	KNN	**Proposed**
Accuracy (%)	91.18	86.13	90.34	85.71	94.96	88.66	**96.68**
Sensitivity (%)	86.92	84.11	91.59	83.18	91.59	83.18	**96.79**
Specificity (%)	94.66	87.79	89.31	87.79	97.71	93.13	**96.51**
Precision (%)	89.86	87.12	92.86	86.47	93.43	87.14	**97.27**
F1-score	92.19	87.45	91.05	87.12	95.52	90.04	**96.89**
MCC	82.21	71.96	80.63	71.09	89.88	77.13	**93.24**

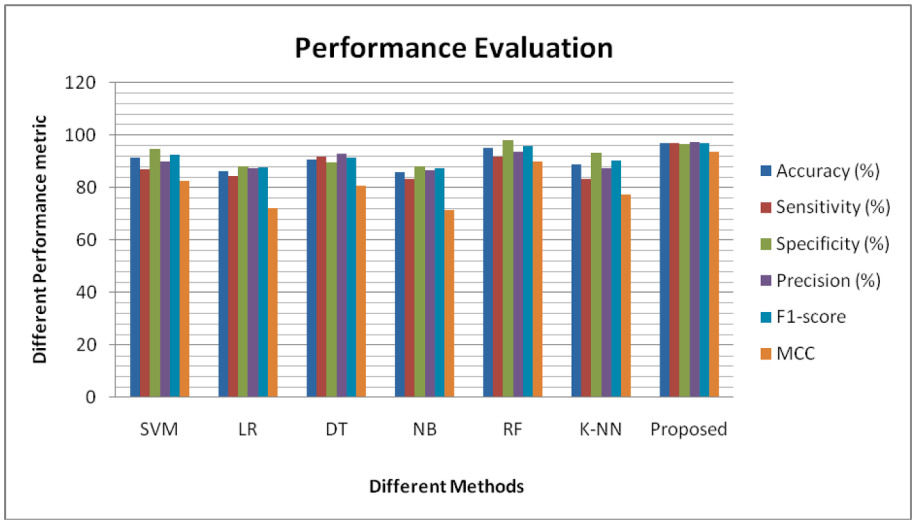


Fig. 12. Comparative analysis with respect to performance metrics [28].

Table 6
Analysis with respect to performance metrics [29].

Method	Accuracy (%)	Precision (%)	Sensitivity (%)	F1 Score
Majority vote with NB, BN, RF, and MLP	85.48	NA	NA	NA
Linear SVM + Linear & RBF SVM	92.22	NA	82.92	NA
HRFLM	88.4	90.1	92.8	90
NB and AES	89.77	NA	NA	NA
ANN and Fuzzy_AHP	91	NA	NA	NA
Randomized Decision Tree Ensemble	93	96	91	93
Proposed	**96.68**	**97.27**	**96.79**	**96.89**
*** NA: Not Available**				

Table 7
Analysis with respect to performance metrics [30].

Method	Accuracy (in %)	Sensitivity (in %)	Specificity (in %)
Rules based classifier	86.7	NA	NA
PPCA	82.18	75	90.57
PSO with SVM	84.36	NA	NA
Relief + LR	89	77	98
mRMR + NB	84	77	90
LASSO + SVM	88	75	96
Proposed	**96.68**	**96.79**	**96.51**
*** NA: Not Available**			

and RBF (radial basis function) SVM, HRFLM, NB, AES (Advanced Encryption Standard), ANN and Fuzzy AHP (Analytical Hierarchy Process), and randomized decision tree ensemble have been traditionally considered for analysis. Table 6 presents the results.

The results revealed varied outcomes have been explored. Accordingly, the majority vote with NB, BN (Bayes Net), RF, and MLP (85.48%), linear SVM + linear and radial basis function (RBF) SVM (92.22%), HRFLM (88.4%), NB, AES (Advanced Encryption Standard) (89.77%), ANN and fuzzy AHP (analytical hierarchy process)(91%), and randomized decision tree ensemble (93%). In contrast, the proposed system showed a high accuracy rate of 96.68% than that for conventional systems. In addition, analysis is carried out by considering traditional methods such as the rule-based classifier, PSO with SVM, PPCA, Relief + LR, mRMR (minimal redundancy and maximal relevance) + NB, LASSO (shrinkage and selection operator) + SVM, HRFLM, factor analysis of mixed data (FAMD) + RF, and L1 linear SVM + L2 linear SVM. Sensitivity, accuracy and specificity were considered as the performance metrics. The outcomes are presented in Table 7.

6. Conclusion

The main intention of this study was to predict heart disease using DM techniques. In this study, DL and ML methods were utilized in various processes. Introduced deep CNN models carried out feature extraction by eliminating irrelevant data through effective learning, D-t-SNE minimized feature dimensionality to enhance the classifier performanceand the proposed H-MLP classified the absence (unaffected) and presence (affected) of the disease. The performance was confirmed by comparison with three traditional studies (Tables 5, 6, & 7) in terms of the significant metrics. The results showed that H-MLP performed classification with a better accuracy than conventional methodologies at 96.68% for the HD dataset. The maximum accuracy attained by this system is highly appropriate for diagnosing heart diseases. The limitation of this work is that the proposed method has not been provisioned for real time applications. In the future, several combinations of DM methods can be employed to further improve the prediction rate for diagnosing heart disease. For example, other deep learning techniques like Deep Neural Networks (DNN), Long Short Term Memory (LSTM) etc for feature extraction and Kernel PCA (Principal Component Analysis) for dimensionality reduction may be explored.

CRediT authorship contribution statement

SonamPalden Barfungpa: Methodology. **Leena Samantaray:** Conceptualization, Methodology. **Hiren Kumar Deva Sarma:** . **Rutuparna Panda:** . **Ajith Abraham:** Supervision.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

References

[1] M. Diwakar, A. Tripathi, K. Joshi, M. Memoria, P. Singh, Latest trends on heart disease prediction using machine learning and image fusion, Materials Today: Proceedings 37 (2021) 3213–3218.

[2] C. Beyene, P. Kamat, Survey on prediction and analysis the occurrence of heart disease using data mining techniques, International Journal of Pure and Applied Mathematics 118 (2018) 165–174.

[3] P. Sharma, K. Choudhary, K. Gupta, R. Chawla, D. Gupta, A. Sharma, Artificial plant optimization algorithm to detect heart rate & presence of heart disease using machine learning, Artificial intelligence in medicine 102 (2020), 101752.

[4] D. Jain, V. Singh, Feature selection and classification systems for chronic disease prediction: A review, Egyptian Informatics Journal 19 (2018) 179–189.

[5] S. Mohan, C. Thirumalai, G. Srivastava, Effective heart disease prediction using hybrid machine learning techniques, IEEE access 7 (2019) 81542–81554.

[6] M.S. Amin, Y.K. Chiam, K.D. Varathan, Identification of significant features and data mining techniques in predicting heart disease, Telematics and Informatics 36 (2019) 82–93.

[7] F. Ali, S. El-Sappagh, S.R. Islam, D. Kwak, A. Ali, M. Imran, et al., A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion, Information Fusion 63 (2020) 208–222.

[8] H. David and S. A. Belcy, heart disease prediction using data mining techniques, *ICTACT Journal on Soft Computing,* vol. 9, 2018.

[9] L.B. Marinho, N. de MM Nascimento, J. W. M. Souza, M. V. Gurgel, P. P. Reboucas Filho, and V. H. C. de Albuquerque,, A novel electrocardiogram feature extraction approach for cardiac arrhythmia classification, Future Generation Computer Systems 97 (2019) 564–577.

[10] S.M.S. Shah, S. Batool, I. Khan, M.U. Ashraf, S.H. Abbas, S.A. Hussain, Feature extraction through parallel probabilistic principal component analysis for heart disease diagnosis, Physica A: Statistical Mechanics and its Applications 482 (2017) 796–807.

[11] D.C. Yadav, S. Pal, Prediction of heart disease using feature selection and random forest ensemble method, International Journal of Pharmaceutical Research 12 (2020) 56–66.

[12] J. K. Kim and S. Kang, Neural network-based coronary heart disease risk prediction using feature correlation analysis, *Journal of healthcare engineering,* vol. 2017, 2017.

[13] S. Sandhiya, U. Palani, An effective disease prediction system using incremental feature selection and temporal convolutional neural network, Journal of Ambient Intelligence and Humanized Computing 11 (2020) 5547–5560.

[14] B. Narasimhan, A. Malathi, Altered particle swarm optimization based attribute selection strategy with improved fuzzy Artificial Neural Network classifier for coronary artery heart disease risk prediction, Int J. Adv. Res. Ideas Innov. Technol 5 (2019) 1196–1203.

[15] X. Liu, X. Wang, Q. Su, M. Zhang, Y. Zhu, Q. Wang, et al., A hybrid classification system for heart disease diagnosis based on the RFRS method, *Computational and mathematical methods in medicine,* vol. 2017, 2017.

[16] Y. Brik, M. Djerioui, and B. Attallah, An efficient Prediction System for Heart Disease based on Twin Support Vector Machine, *International Journal of Computing and Digital System,* 2021.

[17] Y. Khourdifi, M. Bahaj, Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization, International Journal of Intelligent Engineering and Systems 12 (2019) 242–252.

[18] Y. Pan, M. Fu, B. Cheng, X. Tao, J. Guo, Enhanced Deep Learning Assisted Convolutional Neural Network for Heart Disease Prediction on the Internet of Medical Things Platform, IEEE Access 8 (2020) 189503–189512.

[19] N. S. R. Pillai, K. K. Bee, and J. Kiruthika, "Prediction of heart disease using rnn algorithm," *International Research Journal of Engineering and Technology,* vol. 5, 2019.

[20] E. Choi, A. Schuetz, W.F. Stewart, J. Sun, Using recurrent neural network models for early detection of heart failure onset, Journal of the American Medical Informatics Association 24 (2017) 361–370.

[21] S. Safdar, S. Zafar, N. Zafar, N.F. Khan, Machine learning based decision support systems (DSS) for heart disease diagnosis: a review, Artificial Intelligence Review 50 (2018) 597–623.

[22] C.B.C. Latha, S.C. Jeeva, Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques, Informatics in Medicine Unlocked 16 (2019), 100203.

[23] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, J.A. Khan, An automated diagnostic system for heart disease prediction based on ${\chi^{2}} $ statistical model and optimally configured deep neural network, IEEE Access 7 (2019) 34938–34945.

[24] A. Ishaq, S. Sadiq, M. Umer, S. Ullah, S. Mirjalili, V. Rupapara, et al., Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques, IEEE Access 9 (2021) 39707–39716.

[25] K. Mathan, P.M. Kumar, P. Panchatcharam, G. Manogaran, R. Varadharajan, A novel Gini index decision tree data mining method with neural network classifiers for prediction of heart disease, Design automation for embedded systems 22 (2018) 225–242.

[26] G.T. Reddy, N. Khare, An efficient system for heart disease prediction using hybrid OFBAT with rule-based fuzzy logic model, Journal of Circuits, Systems and Computers 26 (2017) 1750061.

[27] H. Sharma, M. Rizvi, Prediction of heart disease using machine learning algorithms: A survey, International Journal on Recent and Innovation Trends in Computing and Communication 5 (2017) 99–104.

[28] S.I. Ayon, M.M. Islam, M.R. Hossain, Coronary artery heart disease prediction: a comparative study of computational intelligence techniques, IETE Journal of Research (2020) 1–20.

[29] I.D. Mienye, Y. Sun, Z. Wang, An improved ensemble learning approach for the prediction of heart disease risk, Informatics in Medicine Unlocked 20 (2020), 100402.

[30] A. Gupta, R. Kumar, H.S. Arora, B. Raman, MIFH: A machine intelligence framework for heart disease diagnosis, IEEE Access 8 (2019) 14659–14674.

[31] V. Veera Anusuya, V. Gomathi, An Efficient Technique for Disease Prediction by Using Enhanced Machine Learning Algorithms for Categorical Medical Dataset, Information Technology and Control 50 (1) (2021) 102–122, https://doi.org/10.5755/j01. itc.50.1.25349.

[32] Ogundokun, R.O., Misra, S., Sadiku, P.O., Gupta, H., Damasevicius, R., Maskeliunas, R. (2022), "Computational Intelligence Approaches for Heart Disease Detection", In: Singh, P.K., Singh, Y., Chhabra, J.K., Illés, Z., Verma, C. (eds) Recent Innovations in Computing. Lecture Notes in Electrical Engineering, vol 855. Springer, Singapore. https://doi.org/10.1007/978-981-16-8892-8_29.