Biomedical Signal Processing and Control 68 (2021) 102820

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Heart disease detection using deep learning methods from imbalanced ECG samples

Adyasha Rath [a], Debahuti Mishra [a], Ganapati Panda [b], Suresh Chandra Satapathy [c,*]

[a] Department of Computer Science and Engineering, Siksha O Anusandhan (Deemed to be) University, Bhubaneswar, 751030, Odisha, India
[b] Department of Electronics and Tele Communication, C. V. Raman Global University, Bhubaneswar, 752054, Odisha, India
[c] Department of Computer Science and Engineering, KIIT Deemed to be University, Bhubaneswar, 751024, Odisha, India

ARTICLE INFO

Keywords:
HD detection
DL model
GAN
GAN-LSTM model
ECG samples-based HD detection
Imbalanced ECG sample

ABSTRACT

Heart disease (HD) is a fatal disease which takes the lives of maximum people compared to other diseases across the world. Early and accurate detection of the disease will help to save many valuable lives. The HD can be detected from medical tests, Electrocardiogram (ECG) signal, heart sounds, Computed Tomography (CT) Images etc. Out of all types of detection of HD from ECG signals plays a vital role. In this paper, the ECG samples of the subjects have been considered as the required inputs to the HD detection model. In recent past, many useful articles have been reported for classification of HD using different machine learning (ML) and deep learning (DL) models. It is observed that with imbalanced HD data the detection accuracy is lower. With an objective to achieve better detection of HD, suitable DL and ML models have been identified in this paper and the required classification models have been developed and tested. The Generative Adversarial Network (GAN) model is chosen with an objective to deal with imbalanced data by generating and using additional fake data for detection purpose. Further, an ensemble model using long short-term memory (LSTM) and GAN is developed in this paper which demonstrates higher performance compared to individual DL model used in this paper. The simulation results using standard MIT-BIH show that the proposed GAN-LSTM model provides the highest accuracy, F1-score and area under curve (AUC) of 0.992, 0.987 and 0.984 respectively compared to other models. Similarly, for PTB-ECG dataset the GAN-LSTM model outperforms all other models with accuracy, F1-score and AUC of 0.994, 0.993 and 0.995 respectively. It is observed that out of the five models investigated, the GAN model performs the best whereas the detection potentiality of the NB model is the lowest. Further research work can be carried out by choosing all other different ensemble models and using other different datasets and the performance can be similarly obtained and compared. The proposed best detection methodology can also be applied to other diseases and healthcare problems.

1. Introduction

CVD is a common non-communicable disease which is the vital cause of mortality around the globe. The various types of CVDs are: HF, heart attacks, hemorrhagic stroke, ischemic stroke and different types of problems associated with arrhythmia and heart valves. It is reported [31,36] by WHO that about 17.7 million people of the globe die each year from the CVDs. This accounts for about 31 % of total deaths and 75 % of this pertains to the developing countries. It is predicted that by 2030 the death due to CVD will touch 23 million per year. Cardiac arrhythmia is the most common type of CVDs and it can be properly detected from the electrocardiogram records. The ECG is an important

Abbreviations: AUC, area under curve; AUROC, area under ROC; Bi-LSTM, bi-layer long short-term memory; CNN, convolutional neural network; CT, computed tomography; CVD, cardiovascular disease; DL, deep learning; ECG, electrocardiogram; FGCNet, fusion of GCN and CNN network; GAN, generative adversarial network; GCN, graph convolution network; HD, heart disease; HF, heart failure; ICG, impedance cardiography; KNN, K-nearest neighbor; LSTM, long short-term memory; ML, machine learning; MCG, magnetocardiography; MRI, magnetic resonance imaging; MIT-BIH, Massachusetts Institute of Technology-Beth Israel hospital; MLP, multi-layer perceptron; NB, Naïve Bayes; PET, positron emission tomography; RBF, radial basis function; ROC, receiver operating characteristics; ReLu, rectified linear unit; RNN, recurrent neural network; SVM, support vector machine; STFT, short time Fourier transform; WHO, World Health Organization.

* Corresponding author.
E-mail addresses: adyasharath1996@gmail.com (A. Rath), debahutimishra@soa.ac.in (D. Mishra), ganapati.panda@gmail.com (G. Panda), suresh.satapathyfcs@kiit.ac.in (S.C. Satapathy).

https://doi.org/10.1016/j.bspc.2021.102820
Received 1 April 2021; Received in revised form 5 May 2021; Accepted 23 May 2021
Available online 31 May 2021

medical tool which records the characteristics of the cardiac transmission, recovery and excitatory for automatic detection of CVD. It is vital to detect the irregular heart rhythms present in the ECG signals. For interpretation of the ECG recording manual examination is essential which is tedious and time consuming. Therefore, different ML and DL methods have been employed for automatic identification of cardiac arrhythmias. The conventional ML algorithms mostly require signal processing techniques related to feature extraction, feature selection, feature reduction and finally classification. The major drawbacks of such techniques are to identify and use the appropriate features from the ECG signals. In recent past, DL methods have played significant role in applications which required prediction and classification tasks. These types of models get rid of the problem of feature extractions and selection. To have a clear picture on the latest status on HD detection using DL and ML methods through literature review is required. This review would assist to identify the research gap and to provide research motivation and objectives of research.

2. Literature review

This Section deals with the review of literature related to detection of HDs to have a better understanding and to know the current trend. In total fourty two numbers of standard articles published during in the journals during 2008–2021 have been chosen for this purpose. These papers have employed different techniques primarily based on ML and DL. In addition, two recently published review papers [1,2] on HD detection are also included. Out of the total number of papers reviewed, the distribution of number of papers which employ ML and DL techniques are twenty-one and fourteen respectively. It is further observed that the ML based articles have employed, the SVM, NB, MLP, RBF and KNN techniques for detection purpose. In case of DL based papers, the LSTM, CNN, GAN have been chosen by authors for HD detection. Considering the types of input into HD detection model, it is learnt that of the reviewed [3,4,6,8–10,12,14] and [21] have employed test data of the subjects, twenty-six number of papers [5,7,11,13,15–20,22–37] have used different types of signals of human body. The distribution of papers using types of input data is presented in Table 1. A through study on the collected papers reveal that various types of signal from human body have been used as inputs to different classification models for detection of HDs. It is found that twenty-six out of fourty two number of papers have used signals such as ECG, heart murmur and CT, ICG and MCG. The details of the distribution of the reviewed papers using types of signal are shown in Table 2. A study on the collected articles also shows that the standard source of data is used in the development of classification model. It is presented in Table 3. The DL methods have been used in articles [22–32] and [35,23–37]. In [24] atrial fibrillation has been automatically detected using LSTM network. It is reported that for 20 subjects using ten-fold cross validation and for three subjects using blind fold validation, accuracies of 98.5 % and 99.77 % are achieved respectively. The bidirectional LSTM based DL method is used in [25] for classification of ECG signal. In the paper wavelet transform is introduced to obtain the sub bands of the main signal and then used as inputs to the deep network. A recognition performance of 99.39 % has been reported by the proposed method. In [26,27,31] and [32] the ECG signal has been identified using combined CNN and LSTM models. The CNN is used in the first few layers for extraction of features and then the LSTM classification model has been developed using the extracted features by the CNN. The combined approach is reported to provide a higher detection accuracy.

Table 1
Distribution of papers using types of input data for HD detection.

Types of data	Details of paper numbers
Test	[3,4,6,8,9,10,12,14,21]
Signal (ECG, ICG, MCG, CT Images)	[5,7,11,13,15,16,17,18,19,20,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37]
Review papers	[1,2]

Table 2
Distribution of papers using types of signals for HD detection.

Types of Signals	Reference Numbers
Electrocardiogram (ECG)	[5,7,18,24,25,26,27,28,29,31,32,33,34,35,36,37]
Impedancecardiography (ICG)	[17,19]
Magnetocardiography (MCG)	[16]
Heart sound	[11,13,20,30]
Computed Tomography (CT Images)	[15,22,23]

Table 3
Standard sources of data for HD detected obtained from the literature review.

Sources of Data	References
EST data	[12]
UCI	[9,10,14,21]
SPECTF, SPECT	[15,16]
StatLog (UCI)	[6,8]
PhysioNet	[7,25,26,27]
MIT-BIH	[4,24,25,27,28,29,31,35,36,37]
Fuwai, PTB-ECG	[31,33,35]
MITDB	[34]
MITHSDB	[30]

In [28], four different types of features are first extracted from the ECG signal and then fed to the LSTM model for detection purpose. The simulation-based experiment yields satisfactory precision, accuracy and F1-score. In [29], features based bi-LSTM model is developed for classification of arrhythmia from ECG signals. Segmentation of heart sound using duration LSTM network is reported in [30]. It is shown that the proposed model provides promising performance. In another interesting paper [22] reduction of noise in CT has been implemented using GAN. Imbalanced CT data has been used in the simulation study and the use of GAN approach has demonstrated satisfactory reduction of noise component. A combined hybrid model is suggested in [23] which provided improved performance and noise reduction present in CT images. In [35–37] the CNN based CVD classification is reported to generate modified ECG signal whereas in [36] a STFT is used for preprocessing the ECG before it is fed to the CNN model. In [37], an improved CNN model is suggested which provides an arrhythmia detection accuracy of 99.06 %. The classification of fruit category is a challenging task. This problem has been addressed [38] by using CNN based approach. It is reported that the proposed method has achieved an accuracy of classification of 94.94 % which is higher than other approaches. In a recent work the authors have presented [39] a complete overview of multimodal data fusion of neuro images. The fusion of CT, MRI, PET, optical imaging as well as radionuclide imaging has led to comprehensive information on pathology. It is reported that the image fusion-based approach has shown improved clinical diagnosis. The authors have developed a CNN based diagnosis system for secondary pulmonary tuberculosis [40]. The paper has employed stochastic pooling and optimization of hyper parameters associated within the CNN. It is demonstrated that the proposed system has exhibited higher sensitivity, specificity and accuracy compared to other four methods. In an interesting paper, the authors have developed FGCNet [41] by combining features from GCN and CNN. The multiple-way data augmentation and rank based average pooling have been chosen in the CNN structure. This approach diagnoses chest CT images. It is claimed that the FGCNet can assist to rapidly detect Covid 19 using chest CT images. The authors in [42] have proposed CCSHNet model for COVID 19 detection which incorporates deep fusion with discriminant correlation analysis and transfer learning. Exhaustive comparison of various performance results, reveals that the CCSHNet is a potential candidate for detection of lung infectious diseases including COVID 19.

It is observed from the literature review that many papers using conventional ML based detection of arrhythmia in ECG signal have been reported. But the number of reported papers on HD detection using ECG signal using DL methods is comparatively less. Some critical observations from the review of literature are presented. In [6] an ensemble model for HD prediction using standard image datasets is reported. The authors have employed the principle of majority votes scheme using two types of DT, SVM, NB and memory-based learner. Therefore, this paper has not used any DL model nor they have used ECG samples for HD prediction. In another article [9] the authors have used a hybrid but not ensemble model and for training and testing purpose the authors have taken Cleveland test dataset. In [23], the authors have predicted the HD risk using combination of GAN-LSTM models. The GAN is used to segment the calcified areas of CT-scan images. In addition, the LSTM model is used for HD risk classification. The ECG signals of subjects have been classified into five types using PhysioNet and MIT-BIH arrhythmia datasets [27]. They have chosen ten-fold cross validation approach for training and testing.

The authors in [32] have presented a CNN and LSTM models for classifying cardiomyopathy with anemia using MIT-BIH arrhythmia datasets. The CNN extracts features and then the LSTM network identifies the inputs into two classes. In another recent work [36], the STFT based Spectrogram of the ECG signals have been obtained and these are then used as inputs to CNN for classifying the ECG signal. The heartbeat classification using MIT-BIH arrhythmia database has been reported [37] by employing improved CNN method. The critical review of these related works provides the research gap, motivation and objectives of proposed investigation.

3. Problem formulation, contribution and organization of the paper

This Section deals with research gap, motivation of research, research objective, contributions and organization of the paper.

3.1. Research gap

The articles reported in the literature on HD detection from the ECG samples mostly have employed conventional ML techniques. Very few articles have used DL approach for features extraction and classification of the HD from the ECG samples of subjects particularly using imbalanced data. In general, it is observed that the DL based approach performs better than the ML classifiers as the features extracted from DL techniques are more suitable. Further, if the available input data belong to imbalanced output classes, then very few methods classify correctly. Mostly the ensemble model offers better detection performance compared to its component models. Some work on HD detection using hybrid model has been reported [23,27,32]. But the reported articles on ensemble classification models are few. These research gaps associated with HD classification need to be addressed for achieving better and reliable detection performance.

3.2. Motivation of research

The research gap outlined in Section 3.1 to suggest that other different DL models which have not been used for detection of the arrhythmia in ECG signal can be employed and performance efficacy can be compared. The MLP, GAN and LSTM based DL methods can be used for this purpose. The three DL methods are chosen because these offer consistent and better performance in many applications such as classification and forecasting. There are many methods for classifying imbalanced data. But these are associated with overfitting problems. However, the GAN model solves data imbalance problem by generating virtual data which are similar to the available data. Hence GAN model has been chosen for classification task. It serves two purposes: generation of virtual data and then detection of arrhythmia using the total dataset. In LSTM model, unlike the RNN, the vanishing and exploding gradient problems are not present. Further, the ensemble model using two better performing DL methods can also be developed for arrhythmia detection from the ECG signal as such model offers better performance. The performance of these DL models can be evaluated using standard ECG datasets and the performance obtained can be compared with those obtained by simulating other ML and ensemble models.

3.3. Objectives of research

The objectives of the present paper have thus formulated considering the research gap and the motivation of research identified in the Sub Sections 3.1 and 3.2.

i. To develop MLP and GAN arrhythmia detection models using imbalanced ECG signal samples of different subjects obtained from the MIT-BIH arrhythmia dataset.
ii. To develop an ensemble model using the GAN and LSTM models using imbalanced ECG signals for detection of HD.
iii. To obtain different performance measures of the proposed ensemble model and compare the same with individual three DL models and two conventional ML models such as the NB and the SVM.

3.4. Contributions

The detailed analysis of the various performance measures obtained in the simulation study using two standard datasets, leads to the following contribution of the paper.

i. Out of the two ML models the SVM yields higher performance measures compared to the NB model.
ii. From among the three DL models, MLP, LSTM and GAN, the GAN offers the best HD classification performance in terms of all three performance measures.
iii. It is further noticed that, in general the proposed three DL models have higher detection potentiality compared to two ML models studied.
iv. By taking into account three performance measures achieved and using two standard datasets, it is observed that the out of six models studied, the ensemble version of GAN-LSTM model offers the best HD detection performance of 0.992, 0.987 and 0.984 for dataset (MIT-BIH) and 0.994, 0.993 and 0.995 for dataset (PTB-ECG) in terms of accuracy, F1-score and AUC of ROC respectively.
v. The findings are consistent with both imbalanced datasets as inputs which demonstrate that the detection reliability of the proposed methods.

3.5. Organization of the paper

The organization of the paper proceeds as follows:

Sections 1 presents a brief Introduction to the area of investigation. The Literature survey and the analysis of review are presented in details in Section 2. The research gap, motivation and objectives of research, contribution and the organization of the paper are outlined in Section 3. Section 4 presents the data sources and details of the data related to ECG signals of subjects used in these articles. The details of MLP, LSTM and the GAN are presented in Section 5. It also deals with the proposed GAN and LSTM based ensemble model for detection of arrhythmia. In Section 6, brief outline of NB and SVM models which have been used for comparison of results are dealt. The simulation-based experiment is carried out using the proposed DL and ensemble models as well as two conventional ML models in Section 7. The imbalanced data presented in Section 4 has been used as inputs to these models and various performance measures have been evaluated. The discussion and interpretation of results obtained from the simulation study are presented in Section 8. Finally, in Section 9, the conclusion of the paper highlighting the major contribution, the limitations of the paper as well as future scope of work are discussed.

4. Details of two Imbalanced data sources used in Simulation Study

In this paper, the ECG signals of various subjects have been obtained from MIT-BIH arrhythmia and PTB-ECG databases [35–37]. The details of the data are presented in Table 4. The ECG samples of each subject have been obtained by averaging 12 recordings of each subjects. The MIT-BIH data under A of this table is used for all six models studied. In total, the samples of 360 subjects with 146 HD patients and remaining 214 normal subjects have been taken from the MIT-BIH dataset. Out of 360 subjects, 220 are males and 140 are females. As may be seen from this table the number of datasets of normal and HD are imbalanced and have been used to develop the ML, DL and ensemble models of this paper. For GAN and GAN-LSTM models, the details of data used for training and testing phases are listed under B of Table 4. The generator of GAN produces 140 fake data (120 normal + 20 HD). Thus, for GAN and GAN-LSTM ensemble models 500 datasets (334 normal and 166 HD) have been employed in the model development. The details of PTB-ECG dataset are also provided in Table 4. There is total 400 datasets, out of which 54 are normal and remaining 346 are of HD patients. In this case also 200 normal fake data and 100 HD fake data are generated before classification.

5. HD detection model

5.1. DL methods

In the present study, MLP, LSTM, GAN and GAN-LSTM ensemble models have been developed to detect the presence of arrhythmia in ECG signal of subjects. The basic briefly principle, the block diagram and the training and testing phases of each of the DL models are presented in sequel.

5.1.1. Multi-layer Perceptron (MLP)

The block diagram of a MLP classification model is shown in Fig. 1. The proposed MLP structure comprises of five layers which has been chosen based on achieving better performance during simulation study. The samples of ECG signal of each subjects are fed to the MLP classifier. The five hidden layers contain 256, 128, 64, 32 and 16 number of artificial neurons respectively. The ReLu activation function is used in each artificial neuron. In third and fifth hidden layers 50 % dropout of neurons is chosen to avoid overfitting problem. Finally, the last layer contains two output neurons to identify whether the input samples belong to a HD category or not. It may be noted that ECG samples of imbalanced number of subjects are used during training and testing phases. From the MIT-BIH arrhythmia datasets the ECG samples of a total 288 subjects 116 belongs to HD category and 172 subjects are normal. In [3–5] the authors have used MLP but classification task has been performed using balanced dataset.

5.1.2. Long short-term memory (LSTM)

The block diagram of LSTM model is shown in Fig. 2. It comprises of three stages of 256, 128 and 64 basic modules. After each LSTM block, 50 % dropouts of basic modules are made to reduce overfitting problem, In the last classification layer, the softmax activation function is used. The number of LSTM modules in each stage and percentage of dropout have been chosen based on achieving better classification performance reported in the literature. It may be noted that most of LSTM based HD detection-based articles deal with balanced data during training or testing phases. Before training the LSTM network the available data have been splitted into training and testing sets. Randomly 80 % of the dataset have been employed for training purpose and the remaining 20 % for testing operation. At a time 1024 ECG samples of a subject have been used as inputs to 256 LSTM models. Then following the block diagram of Fig. 2 the computation continues until the estimated output is obtained. The desired class is compared with the estimated output and the resultant error is used to train the internal parameters of the LSTM model. The process of training is continued by applying the training samples sequentially. The training process is continued until the loss function attains a possible minimum value. After satisfactory training phase the performance of the model is tested with the remaining 20 % datasets.

5.1.3. Generative adversarial network (GAN)

The GAN based model is suitable for classification if the number of available data is less or imbalanced. This DL model has the potentiality to generate fictious data using its generator. Subsequently both fake and original data are used for classification purpose. It follows latent space interpolation principle. The block diagram of a GAN model is shown in Fig. 3. The role of the generator (Fig. 3(a)) is to produce fake data and it tries to fool the discriminator. But the discriminator (Fig. 3(b)) is trained to distinguish between the actual and fake data. After successful training the discriminator it fails to separate the actual and artificial data. Subsequently, the discriminator is not used and the generator produces new realistic dataset. The challenging task of the GAN model is to obtain its satisfactory training. A healthy competition takes place between the generator and discriminator for simultaneous learning. In this case, the loss function of the model is computed from the output of the

Table 4
Details of Original MIT-BIH and PTB-ECG including Fake datasets used in different ML and DL models.

Datasets	A		B			Total
	Types	Number of sets	Fake data produced by generator of GAN	Number of training phase data	Number of testing phase data	
1. MIT-BIH	Normal	214	120	292 (172 + 120) (original) (fake data)	42 (Original)	334
	HD	146	20	136 (116 + 20) (Original) (fake data)	30 (Original)	166
	Total	360	140	428	72	500
2. PTB-ECG	Normal	54	200	43 + 200 (Original + fake)	11 (Original)	254
	HD	346	100	277 + 100 (Original + fake)	69 (Original)	446
	Total	400	300	620	80	700

Note
A- Datasets in A used in all DL and ML models except GAN and GAN-LSTM.
B- Datasets in B used in GAN and GAN-LSTM.

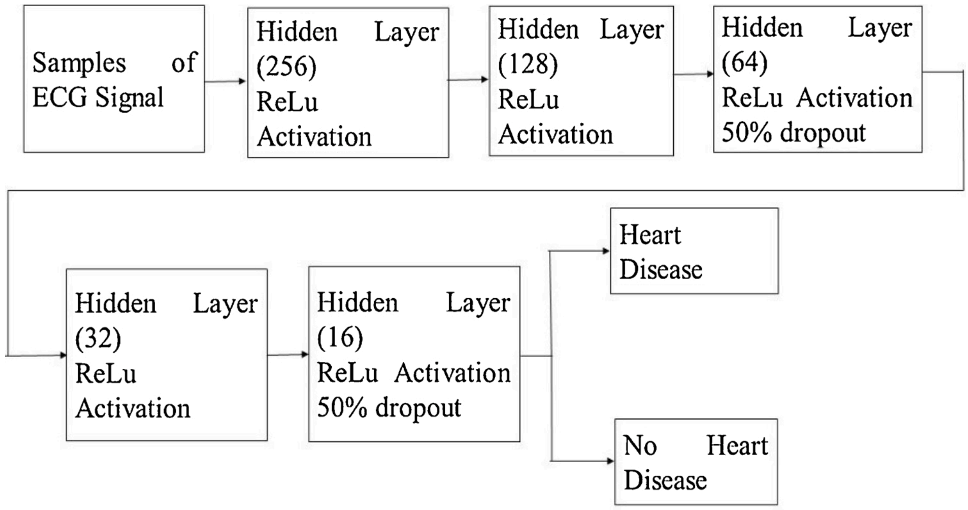


Fig. 1. Block Diagram of MLP Model.

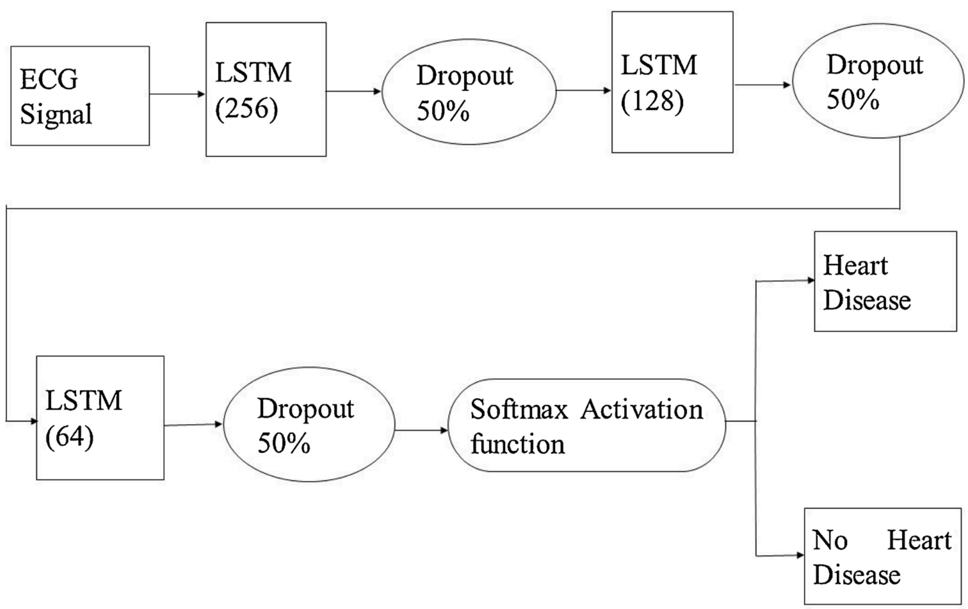


Fig. 2. Block diagram of LSTM Model.

discriminator. Therefore, the discriminator updates its parameters faster. On the other hand, parameters update of the generator are slower and sometimes fails to converge. Another drawback of this model is that the generator produces similar outputs for different random inputs. In Fig. 3 the generator is like a counter feiter whereas the discriminator acts like a policeman. During training phase, the discriminator parameters are temporarily not changed. The generator uses the gradients of the cost function to update the parameters and improves the ability to synthesize the signal. In essence, the GAN consists of two networks competing with each other and also cooperating at the same time. When the training of GAN is successful the discriminator considers the synthesize signals as real. At this stage the discriminator is discarded. Subsequently the generator produces meaningful patterns from random inputs.

In GAN model, the generator captures the joint probability between the data and the corresponding labels. But the discriminator captures the conditional probability between the input and label sets. In essence the generator models how the data is placed throughout the space. But the discriminator attempts to draw boundaries in the data space. The better detection performance of the GAN model is demonstrated using two imbalance datasets (MIT-BIH, PTB-ECG).

5.2. *Ensemble model using DL methods GAN-LSTM*

Fig. 4 shows the GAN-LSTM ensemble learning model which comprises of two independent deep learning models. In the present investigation the two deep learning models chosen are GAN and LSTM. In this case also imbalanced data is used as input. The final prediction is considered as concatenation of two individual predictions. These two models chosen are on the basis of yielding higher classification accuracy for each of the two classes. In the present case, the LSTM performs better for the majority output class whereas the GAN model shows improved performance for minority output class. The model averaging method is employed to train the two DL models. Finally, the predicted class of the ensemble model is achieved concatenation of the two predicted outputs.

6. Brief outlines of ML models

To have a performance comparison of HD detection with, NB and

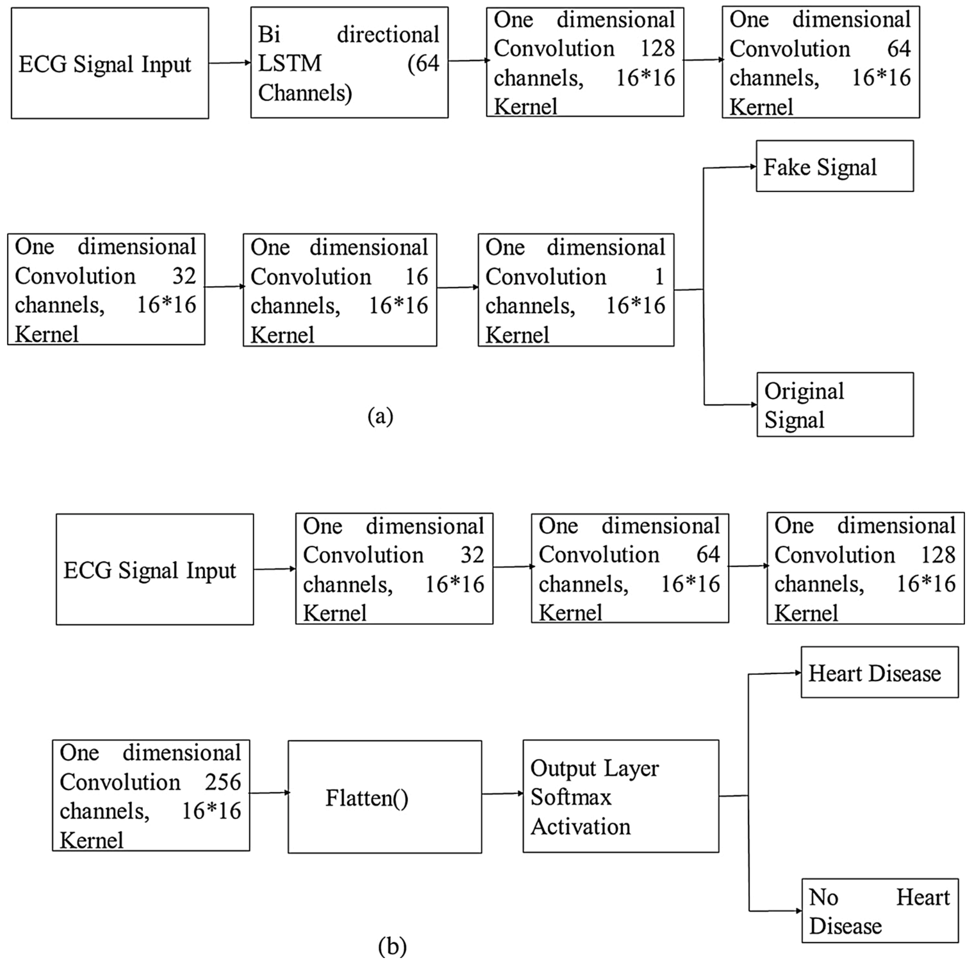


Fig. 3. Block diagram of GAN Model (a) Generator (b) Discriminator.

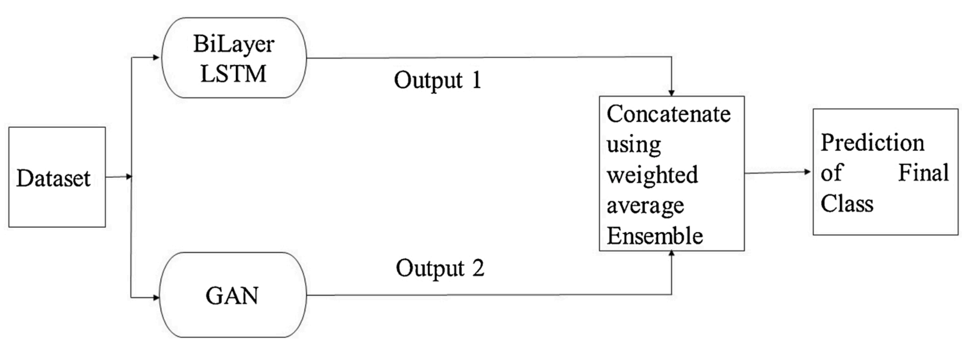


Fig. 4. Block diagram of GAN-LSTM Ensemble Model.

SVM are chosen as these two models show consistent and satisfactory past performance using different types of datasets. In this Section the basic methodology of these two ML methods is dealt.

6.1. Naïve Bayes (NB)

The concept behind NB classification algorithm is based on posterior probability of a data instance n_i in a class cs_j of the model. The posterior probability $P(n_i|cs_j)$ refers to the probability that n_i belongs to class cs_j. The posterior probability is computed by taking the product of all the probabilities of different attributes of the data. It is computed as

$$P(n_i|cs_j) = \prod_{m=1}^{M} P(x_{im}|cs_j)$$

Where M = number of attributes and x_{im} is mth attribute of ith instance.

It may be noted that the posterior probability of each class is separately computed. The highest of these probabilities refers to the label of the instance under consideration. The detailed flowchart based on this principle is shown in Fig. 5.

6.2. Support vector machine (SVM)

The SVM is a dedicated classification algorithm which is essentially

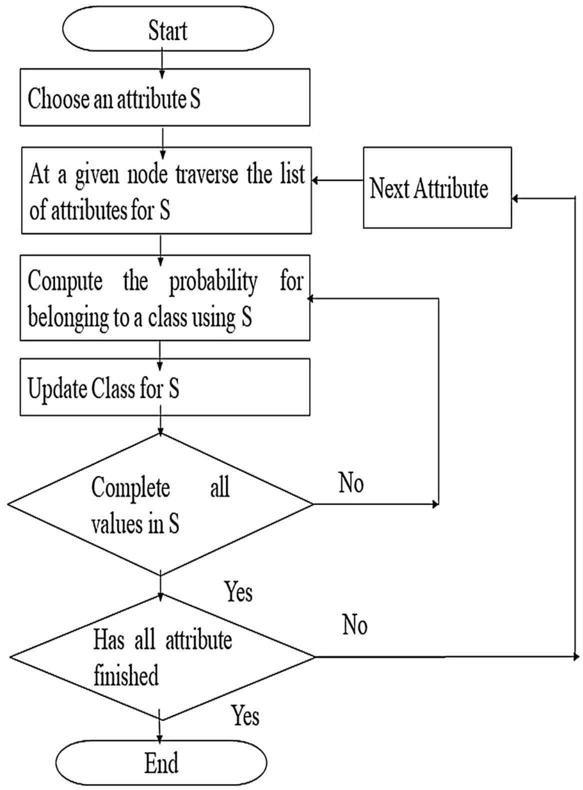


Fig. 5. Flowchart of the NB Model.

classifies the input datasets into two groups. The support vector represents the separating line between the two classes which maximizes the margin. The kernel SVM is suitable for nonlinear data as more features can be added to fit a hyperplane. In the present case four kernels: linear, sigmoid, RBF and polynomial are used in the SVM model and the best of the four is selected for HD detection. The block diagram of the SVM model is shown in Fig. 6. To build the SVM model the step followed are:

i. Split the dataset into training and test data
ii. Simulate the model using Support Vector based principle
iii. Implement the Kernel functions
iv. Predict the class by applying the test data to the developed model
v. Find various performance measures during the testing phase.

7. Simulation based experimental study

In this Section, the block diagrams of MLP, LSTM and GAN shown in Figs. 1–3 have been simulated. The training and testing phases are carried out using MIT-BIH arrhythmia and PTB-ECG datasets. The details of the datasets are provided in Table 4. The randomly chosen 80 % of the data presented under column A are used for training all ML and DL models. The remaining 20 % of the datasets are used for testing purpose. For GAN and GAN-LSTM ensemble models, the datasets listed under column B of Table 4 have been employed as inputs during simulation study. In case of GAN and GAN-LSTM models 500 datasets have been used out of which 428 datasets are used for training phase and remaining 72 have been used for validation purpose. Out of 500 datasets, the GAN and GAN-LSTM models have employed. 360 original data of MIT-BIH and 140 numbers of fake data. In case of PTB-ECG, 620 data (243 normal +377 HD) are used for training purpose. The remaining 80 datasets (11 original normal + 69 original HD) are employed in testing phase. During the process of simulation, the variation of accuracy of classification with number of epochs has been obtained both during training and validation phases. These comparative plots for both the datasets have been shown in Figs. 7–10 for MLP, LSTM, NB and SVM models respectively. Similarly, for the GAN model the variation of loss value with number of iterations during training phase of generator and discriminator has been obtained from simulation-based experiment. These have been compared and displayed in Fig. 11(a) and (b) for MIT-BIH and PTB-ECG datasets respectively. In the ensemble model the variation of generator loss with number of epochs both during training and validation phases are also obtained from the experiment. These have been compared in Fig. 12(a) and (b) and the similar comparative loss plots for discriminator of the ensemble model during training and validation states have been shown in Fig. 13(a) and (b). The ROC plots exhibiting the relation between true positive rate and false negative rate for all models have also been obtained through simulation n study. The corresponding comparative plots have been presented in Fig. 14(a) and (b). The observation and the analysis of all these plots have been dealt in the next section.

8. Discussion and interpretation of results

In this Section, the analysis of various plots and results presented in tables obtained from simulation-based experiments dealt in the previous section is made. The observations of Figs. 7 and 8 on model accuracy of MLP and LSTM indicates that the MLP model exhibits slightly higher training accuracy during training phase compared to the corresponding validation accuracy obtained after training is completed. However, from

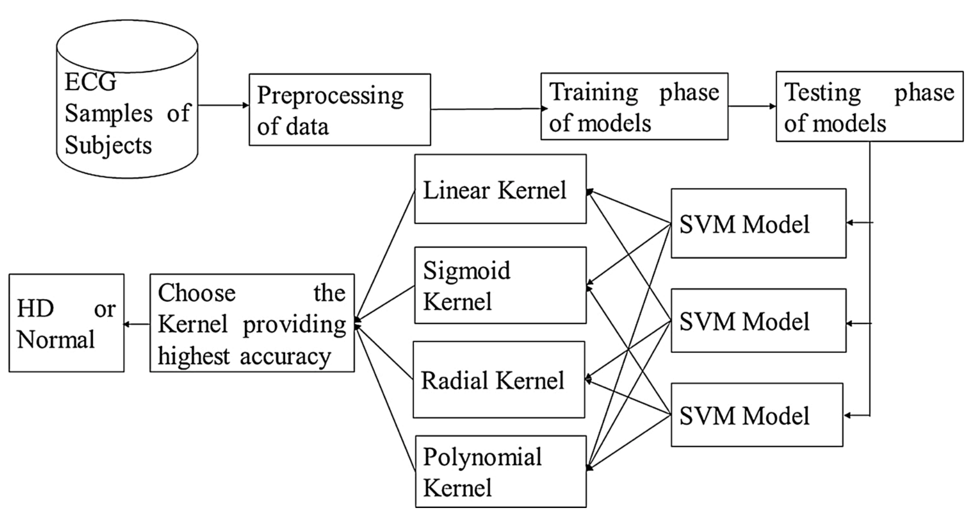
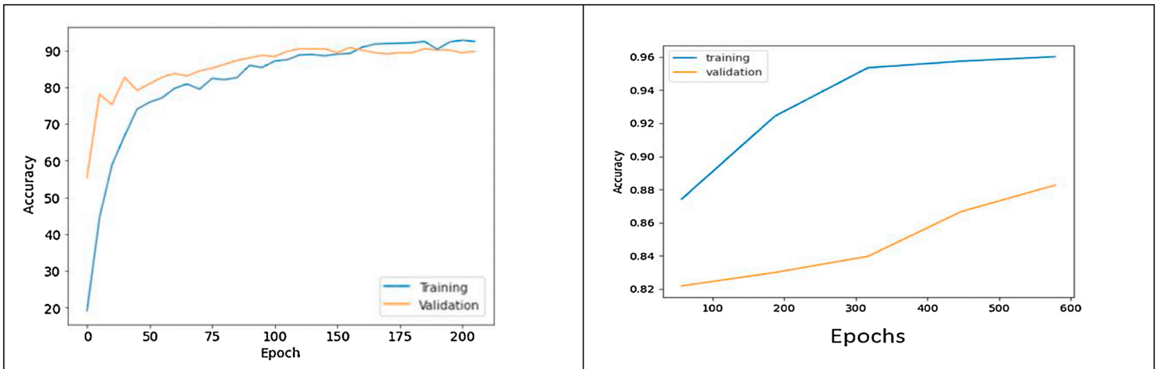
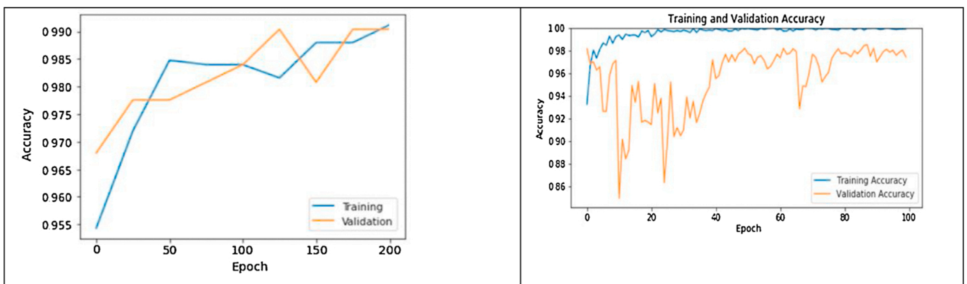


Fig. 6. Block diagram of the SVM Model.



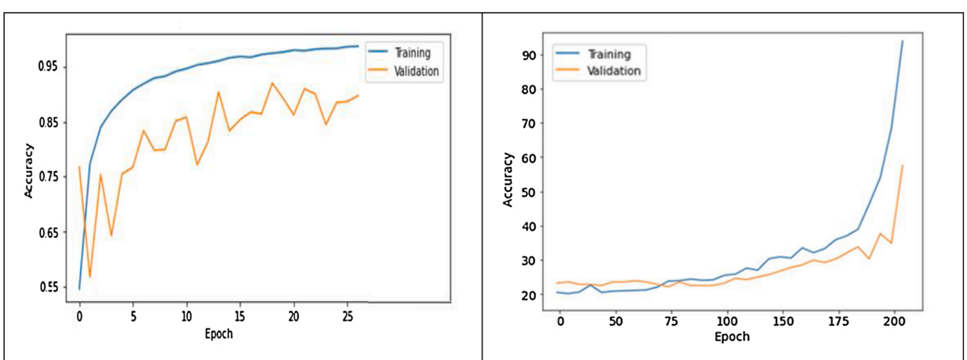
(a) (b)

Fig. 7. Comparison of plots of variation of accuracy with number of epochs during training and validation phase of MLP Model (a) MIT-BIH (b) PTB-ECG.



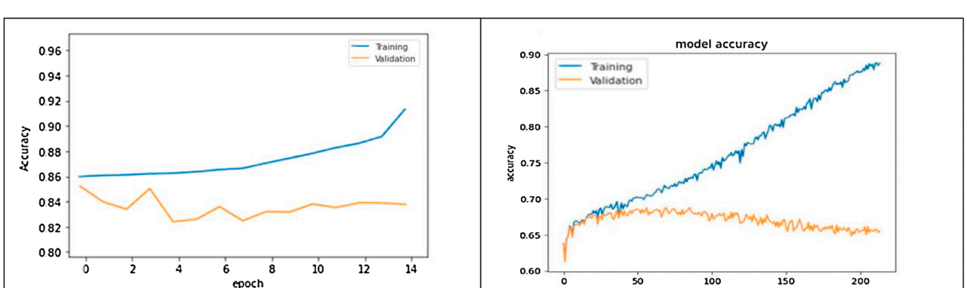
(a) (b)

Fig. 8. Comparison of plots of variation of accuracy with number of epochs during training and validation phase of LSTM Model (a) MIT-BIH (b) PTB-ECG.



(a) (b)

Fig. 9. Comparison of plots of variation of accuracy with number of epochs during training and validation phase of NB Model (a) MIT-BIH (b) PTB-ECG.



(a) (b)

Fig. 10. Comparison of plots of variation of accuracy with number of epochs during training and validation phase of SVM Model (a) MIT-BIH (b) PTB-ECG.

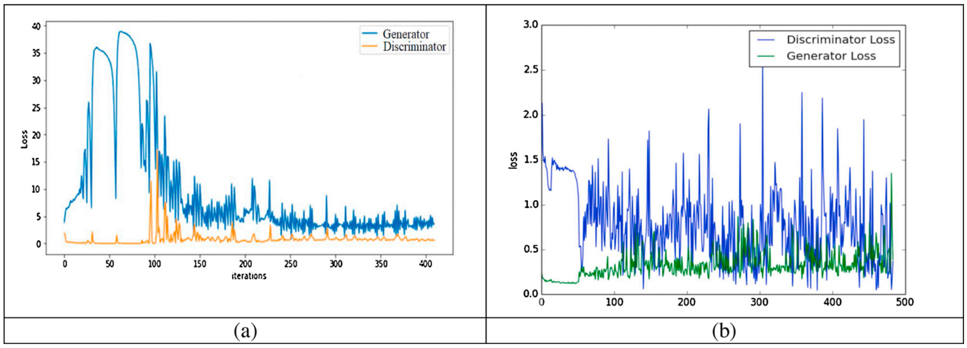


Fig. 11. Comparison of loss with variation of number of iterations during training phase of GAN Model (a) MIT-BIH (b) PTB-ECG.

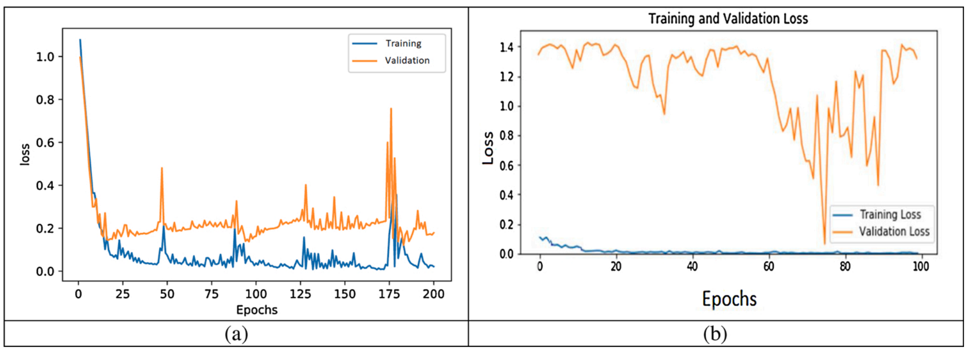


Fig. 12. Comparison of variation of loss of Generator of GAN during training and validation phases (a) MIT-BIH (b) PTB-ECG.

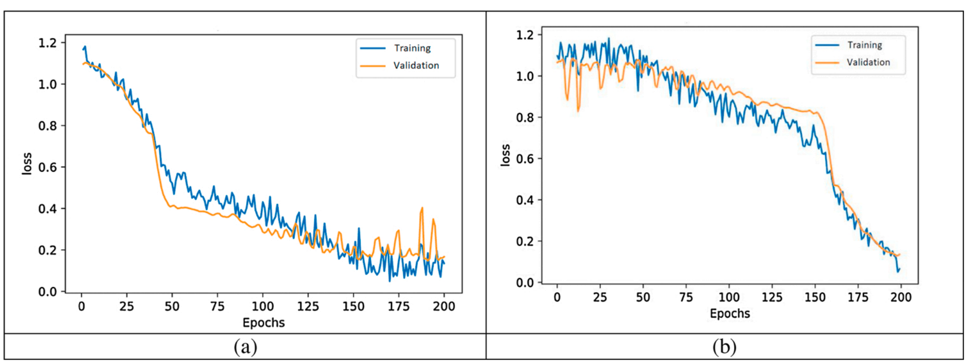


Fig. 13. Comparison of variation of loss of Discriminator of GAN during training and validation phases (a) MIT-BIH (b) PTB-ECG.

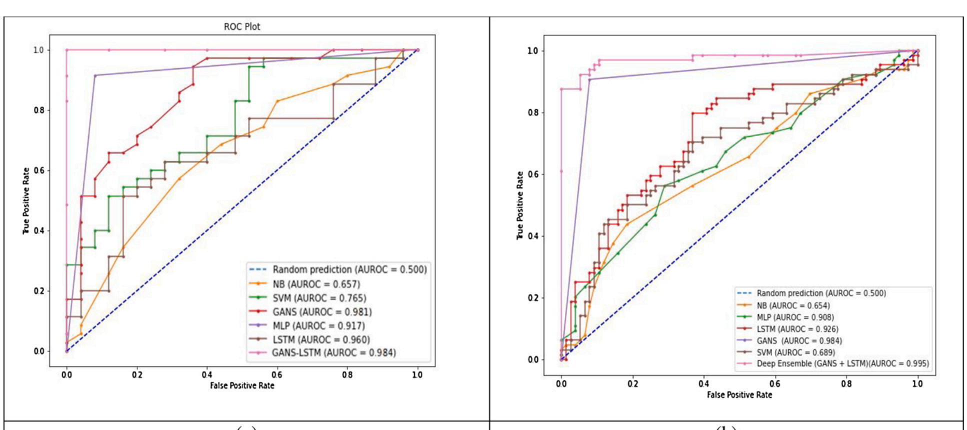


Fig. 14. Comparison of ROC Plots (a) MIT-BIH (b) PTB-ECG.

Fig. 8, it is noticed that the average accuracy value of LSTM model remains almost the same both during training and validation phases for MIT-BIH datasets. However, in case of PTB-ECG dataset, the training accuracy is higher than the corresponding validation accuracy. Thus, it is observed that LSTM offers more stable performance compared to the MLP model for MIT-BIH datasets. The accuracy plots shown in Figs. 9 and 10 reveal that in both NB and SVM models, the accuracy during training phase is higher than the validation accuracy for both the datasets. This observation is similar to other ML classification models. In case of GAN model, it is noticed from Fig. 11 that the loss value of the generator is higher than the loss value of the discriminator. In the initial period of training the difference in loss value of the two models of generator and discriminator is higher. As the training progresses the loss value decreases and almost remains constant. This observation is quite natural in almost all types of datasets. Similarly, for the ensemble model the variation of loss with epochs for both generator and discriminator are shown in Figs. 12 and 13. It is observed from these two plots that in case of generator the validation loss is higher than the training loss. However, in case of discriminator (Fig. 13(a)) the training loss shows higher value compared to the validation loss for MIT-BIH datasets. But the training loss of PTB-ECG dataset is higher during initial epochs and then subsequently remains almost same. From the simulation study, the ROC plots of all the six models have been obtained and compared in Fig. 14(a) and (b) for MIT-BIH and PTB-ECG datasets respectively. The area under ROC plots of these models have been obtained and are listed in the two Fig. 14(a) and (b). The observation of these AUROC shows that the proposed ensemble model provides the highest value 0.984 (MIT-BIH) and 0.995 (PTB-ECG) compared to all other five models. It is noticed that GAN offers the next best value whereas the NB model provides the least AUROC value of 0.657 and 0.654 for MIT-BIH and PTB-ECG datasets respectively. The complete performance measures in terms of the accuracy, F1-score, AUC of all the two ML models, three DL models and one ensemble model have been listed in Tables 5 and 6 for MIT-BIH and PTB-ECG datasets. Based on the three performance measures the ranking of the models has been done and are shown in the respective tables. The analysis of these performance measures clearly exhibits that the GAN-LSTM ensemble model gets the first rank whereas the NB model attains the sixth rank. This observation is true for both the datasets which demonstrate that the performance of all the proposed models is consistent irrespective of the datasets used.

Table 5
Comparison of various performance measures of HD detection using imbalanced datasets (MIT-BIH dataset).

Performance Measures	SVM	NB	MLP	LSTM	GAN	GAN-LSTM
Accuracy	0.846	0.821	0.915	0.966	0.978	0.992
(Value/Rank)	(V)	(VI)	(IV)	(III)	(II)	(I)
F1-Score	0.813	0.817	0.902	0.937	0.975	0.987
(Value/Rank)	(VI)	(V)	(IV)	(III)	(II)	(I)
AUC	0.765	0.657	0.917	0.960	0.981	0.984
(Value/Rank)	(V)	(VI)	(IV)	(III)	(II)	(I)
Overall Winner	GAN-LSTM					

Table 6
Comparison of various performance measures of HD detection using imbalanced datasets (PTB-ECG dataset).

Performance Measures	SVM	NB	MLP	LSTM	GAN	GAN-LSTM
Accuracy	0.664	0.634	0.888	0.973	0.987	0.994
(Value/Rank)	(V)	(VI)	(IV)	(III)	(II)	(I)
F1-Score	0.726	0.721	0.912	0.965	0.976	0.993
(Value/Rank)	(V)	(VI)	(IV)	(III)	(II)	(I)
AUC	0.689	0.658	0.908	0.926	0.984	0.995
(Value/Rank)	(V)	(VI)	(IV)	(III)	(II)	(I)
Overall Winner	GAN-LSTM					

9. Conclusion

This paper has addressed the problem of HD classification using ECG samples of normal subjects and HD patients as input information. The imbalanced MIT-BIH arrhythmia dataset as well as the PTB-ECG have been chosen to assess the HD detection performance of subjects. Six classification models including two ML, three DL and one ensemble models have been chosen for this purpose. The GAN model has been used to generate fake samples with an objective to perform better under imbalanced category of input data. Analysis of various performance measures show that the GAN-LSTM model outperforms the individual GAN and LSTM model in terms of all three performance measures evaluated in the paper. This is true for the two standard datasets used in simulation study.

Based on the analysis of three performance measures during validation phase, the ranks of the GAN-LSTM, GAN, LSTM and MLP are I, II, III and IV respectively. Further all the three DL models demonstrate higher HD detection potentiality compared to the two ML models. The proposed work can be extended for other ensemble DL models as well as using other standard databases. Further, the proposed methods can also be applied for classification of HD using datasets relating to ICG, MCG, CT and heart sound signals of patients. Different types of ensembling of models have been suggested in literature for providing better classification accuracy. A detailed investigation on the detection performance obtained from all the ensemble models can be carried out and the model which would provide consistent and the best performance can be identified and recommended for use in practice. The performance ranking of the proposed models is based on two datasets. The ranking consistency of these models can be verified using other different standard ECG datasets. The performance of the proposed detection methods can also be applied to COVID 19, Pneumonia and other fatal diseases.

CRediT authorship contribution statement

Adyasha Rath, Debahuti Mishra, Ganapati Panda and Suresh Chandra Satapathy have equally contributed to the work in the paper.

Declaration of Competing Interest

All authors declare that we have no conflict of interest and authors have not used any image/data for our work which needs authorization.

References

[1] M. Nabih-Ali, E.S.A. El-Dahshan, A.S. Yahia, A review of intelligent systems for heart sound signal analysis, J. Med. Eng. Technol. 41 (7) (2017) 553–563.
[2] A.K. Dwivedi, S.A. Imtiaz, E. Rodriguez-Villegas, Algorithms for automatic analysis and classification of heart sounds–a systematic review, IEEE Access 7 (2018) 8316–8345.
[3] I. Kurt, M. Ture, A.T. Kurum, Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease, Expert Syst. Appl. 34 (1) (2008) 366–374.
[4] J.J. Oresko, Z. Jin, J. Cheng, S. Huang, Y. Sun, H. Duschl, A.C. Cheng, A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing, IEEE Trans. Inf. Technol. Biomed. 14 (3) (2010) 734–740.
[5] J. Wang, C. Liu, L. Li, W. Li, L. Yao, H. Li, H. Zhang, A stacking-based model for non-invasive detection of coronary heart disease, IEEE Access 8 (2020) 37124–37133.
[6] S. Bashir, U. Qamar, F.H. Khan, M.Y. Javed, MV5: a clinical decision support framework for heart disease prediction using majority vote-based classifier ensemble, Arab. J. Sci. Eng. 39 (11) (2014) 7771–7783.
[7] M. Yasin, T. Tekeste, H. Saleh, B. Mohammad, O. Sinanoglu, M. Ismail, Ultra-low power, secure IoT platform for predicting cardiovascular diseases, IEEE Trans. Circuits Syst. I Regul. Pap. 64 (9) (2017) 2624–2637.
[8] A.K. Dwivedi, Performance evaluation of different machine learning techniques for prediction of heart disease, Neural Comput. Appl. 29 (10) (2018) 685–693.
[9] S. Mohan, C. Thirumalai, G. Srivastava, Effective heart disease prediction using hybrid machine learning techniques, IEEE Access 7 (2019) 81542–81554.

[10] J.P. Li, A.U. Haq, S.U. Din, J. Khan, A. Khan, A. Saboor, Heart disease identification method using machine learning classification in E-healthcare, IEEE Access 8 (2020) 107562–107582.
[11] A. Sengur, An expert system based on linear discriminant analysis and adaptive neuro-fuzzy inference system to diagnosis heart valve diseases, Expert Syst. Appl. 35 (1–2) (2008) 214–222.
[12] İ. Babaoglu, O. Findik, E. Ülker, A comparison of feature selection models utilizing binary particle swarm optimization and genetic algorithm in determining coronary artery disease using support vector machine, Expert Syst. Appl. 37 (4) (2010) 3177–3183.
[13] C. Kwak, O.W. Kwon, Cardiac disorder classification by heart sound signals using murmur likelihood and hidden Markov model state likelihood, IET Signal Process. 6 (4) (2012) 326–334.
[14] J. Nahar, T. Imam, K.S. Tickle, Y.P.P. Chen, Computational intelligence for heart disease diagnosis: a medical knowledge driven approach, Expert Syst. Appl. 40 (1) (2013) 96–104.
[15] S. Shilaskar, A. Ghatol, Feature selection for medical diagnosis: evaluation for cardiovascular diseases, Expert Syst. Appl. 40 (10) (2013) 4146–4153.
[16] R. Tao, S. Zhang, X. Huang, M. Tao, J. Ma, S. Ma, C. Shen, Magnetocardiography-based ischemic heart disease detection and localization using machine learning methods, IEEE Trans. Biomed. Eng. 66 (6) (2018) 1658–1667.
[17] S. Chabchoub, S. Mansouri, R.B. Salah, Detection of valvular heart diseases using impedance cardiography ICG, Biocybern. Biomed. Eng. 38 (2) (2018) 251–261.
[18] R.L. Devi, V. Kalaivani, Machine learning and IoT-based cardiac arrhythmia diagnosis using statistical and dynamic features of ECG, J. Supercomput. 1- (2019) 12.
[19] I.B. Salah, R. De la Rosa, K. Ouni, R.B. Salah, Automatic diagnosis of valvular heart diseases by impedance cardiography signal processing, Biomed. Signal Process. Control 57 (2020), 101758.
[20] A. Saeidi, F. Almasganj, Cardiac valves disorder classification based on active valves appearance periodic sequences tree of murmurs, Biomed. Signal Process. Control 57 (2020), 101775.
[21] G. Magesh, P. Swarnalatha, Optimal feature selection through a cluster-based DT learning (CDTL) in heart disease prediction, Evol. Intell. (2020) 1–11.
[22] J.M. Wolterink, T. Leiner, M.A. Viergever, I. Išgum, Generative adversarial networks for noise reduction in low-dose CT, IEEE Trans. Med. Imaging 36 (12) (2017) 2536–2545.
[23] Z.Z. Zou, K. Xie, Y.F. Zhao, J. Wan, L. Lan, C. Wen, Intelligent assessment of percutaneous coronary intervention based on GAN and LSTM models, IEEE Access 8 (2020) 90640–90651.
[24] O. Faust, A. Shenfield, M. Kareem, T.R. San, H. Fujita, U.R. Acharya, Automated detection of atrial fibrillation using long short-term memory network with RR interval signals, Comput. Biol. Med. 102 (2018) 327–335.
[25] Ö. Yildirim, A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification, Comput. Biol. Med. 96 (2018) 189–202.
[26] J.H. Tan, Y. Hagiwara, W. Pang, I. Lim, S.L. Oh, M. Adam, et al., Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals, Comput. Biol. Med. 94 (2018) 19–26.
[27] S.L. Oh, E.Y. Ng, R. San Tan, U.R. Acharya, Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats, Comput. Biol. Med. 102 (2018) 278–287.
[28] S.K. Pandey, R.R. Janghel, Automatic arrhythmia recognition from electrocardiogram signals using different feature methods with long short-term memory network model, Signal Image Video Process. 14 (6) (2020) 1255–1263.
[29] B. Ganguly, A. Ghosal, A. Das, D. Das, D. Chatterjee, D. Rakshit, Automated detection and classification of arrhythmia from ECG signals using feature-induced long short-term memory network, IEEE Sens. Lett. 4 (8) (2020) 1–4.
[30] Y. Chen, J. Lv, Y. Sun, B. Jia, Heart sound segmentation via duration long– short term memory neural network, Appl. Soft Comput. 95 (2020), 106540.
[31] M. Thirugnanam, M.S. Pasupuleti, Cardiomyopathy-induced arrhythmia classification and pre-fall alert generation using Convolutional Neural Network and Long Short-Term Memory model, Evol. Intell. (2020) 1–11.
[32] K. Yamamoto, R. Hiromatsu, T. Ohtsuki, ECG signal reconstruction via doppler sensor by hybrid deep learning model with CNN and LSTM, IEEE Access 8 (2020) 130551–130560.
[33] M. Deng, C. Wang, M. Tang, T. Zheng, Extracting cardiac dynamics within ECG signal for human identification and cardiovascular diseases classification, Neural Netw. 100 (2018) 70–83.
[34] J. Chen, A. Valehi, A. Razi, Smart heart monitoring: early prediction of heart problems through predictive analysis of ECG signals, IEEE Access 7 (2019) 120831–120839.
[35] N.I. Hasan, A. Bhattacharjee, Deep learning approach to cardiovascular disease classification employing modified ECG signal from empirical mode decomposition, Biomed. Signal Process. Control 52 (2019) 128–140.
[36] J. Huang, B. Chen, B. Yao, W. He, ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network, IEEE Access 7 (2019) 92871–92880.
[37] H. Wang, H. Shi, X. Chen, L. Zhao, Y. Huang, C. Liu, An improved convolutional neural network based approach for automated heartbeat classification, J. Med. Syst. 44 (2) (2020) 35.
[38] Y.D. Zhang, Z. Dong, X. Chen, W. Jia, S. Du, K. Muhammad, S.H. Wang, Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation, Multimed. Tools Appl. 78 (3) (2019) 3613–3632.
[39] Y.D. Zhang, Z. Dong, S.H. Wang, X. Yu, X. Yao, Q. Zhou, et al., Advances in multimodal data fusion in neuroimaging: overview, challenges, and novel orientation, Inf. Fusion 64 (2020) 149–187.
[40] Y.D. Zhang, D.R. Nayak, X. Zhang, S.H. Wang, Diagnosis of secondary pulmonary tuberculosis by an eight-layer improved convolutional neural network with stochastic pooling and hyperparameter optimization, J. Ambient Intell. Humaniz. Comput. (2020) 1–18.
[41] S.H. Wang, V.V. Govindaraj, J.M. Górriz, X. Zhang, Y.D. Zhang, Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network, Inf. Fusion 67 (2021) 208–229.
[42] S.H. Wang, D.R. Nayak, D.S. Guttery, X. Zhang, Y.D. Zhang, COVID-19 classification by CCSHNet with deep fusion using transfer learning and discriminant correlation analysis, Inf. Fusion 68 (2021) 131–148.