Feature Augmentation and Convolutional Neural Networks for Accurate Prediction of Heart Disease

T.G. Ramnadh babu[1], Kolipaka Bhavana[2], , Gude chaitanya[2], , Madagunda Sravani[2], , Rafi Shaik[3], and , Dr. Sireesha Moturi[4]

[1] Professor, Dept Of Computer Science and Engineering
Narasaraopeta Engineering College(Autonomous),
Narasaraopet 522601, Palnadu District,Andhra Pradesh,India.
baburamnadh@gmail.com

[2] Dept of Computer Science and Engineering
Narasaraopeta Engineering College(Autonomous),
Narasaraopet 522601, Palnadu District,Andhra Pradesh,India.
bhavanakolipaka2004@gmail.com, chaitanyagude434@gmail.com,
madagundasravani@gmail.com

[3] Asst.Professor, Dept Of Computer Science and Engineering
Narasaraopeta Engineering College(Autonomous),
Narasaraopet 522601, Palnadu District,Andhra Pradesh,India.
shaikrafinrt@gmail.com

[4] Assoc.Professor, Dept Of Computer Science and Engineering
Narasaraopeta Engineering College(Autonomous),
Narasaraopet 522601, Palnadu District,Andhra Pradesh,India.
sireeshamoturi@gmail.com

Abstract. Heart diseases are now becoming the cause of major deaths in developing countries. Prediction of heart disease is crucial for risk evaluation of any patient. This paper presents a new method to enhance the representation features and classifies the accuracy for the prediction of heart disease by embedding Convolutional Neural Networks into the Sparse Autoencoder. In this paper, with a dataset of 918 patients' records containing 11 clinical variables, our method sidesteps the pitfall problems that lie in traditional classifiers by elaborating new constructive features using the feature augmentation techniques. Experimental results of this design show that it is possible to reach an accuracy up to 93.478%, over 4.98% compared to traditional designs, such as MLP and RF. Determining the size of the latent space with 100 features greatly improved the output from the model. Our findings are that there is increased prediction accuracy with the incorporation of deep learning techniques likely to have therapeutic benefit for a clinical decision regarding a patient's predisposition to heart disease. It is expected that this present research report will show that advanced feature extraction methods are highly important for diagnostic improvement and, in the final analysis, in care.

Keywords: Heart Disease Prediction · Deep Learning · Feature Augmentation · Sparse Autoencoder (SAE) · Multilayer Perceptron (MLP)

1 Introduction

Cardiovascular diseases are still one of the highest rates of morbidity and mortality in the world, claiming several million lives annually. According to estimations from the WHO, CVDs account for about 32% of all deaths that occur around the world; that calls for urgent implementation of effective strategies for risk assessment and early intervention [1]. By nature, heart disease is a multifactorial disease moderated by an interaction between genetic, lifestyle, and environmental factors, which makes the identification of at-risk individuals more complex [2]. Conventional methods are based on a relatively small set of clinical indicators that would be quite incapable of capturing any complex interactions between Identify applicable funding agency here. If none, delete this. various risk factors [3]. In this backdrop, there has been a growing interest in research in recent years in machine learning and artificial intelligence approaches to help with predicting diseases [4]. Among all the variants, deep learning is a popular variant due to its inherent hierarchal feature learning capability from raw data; especially, it is well-suited to high-dimensional data [5]. In analyzing such complex patterns in the heart patients' data, hardly any traditional statistical method can decipher, but deep learning models are capable of doing that [6]. This work highlights a new approach with the use of SAE, having complementary strengths combined with CNNs for better representation of features and improved classification performance in the prediction of heart disease [7]. Consequently, SAE will provide effective unsupervised augmentation of features to bring out latent meaningful features from the initial clinical data. We map a dataset with the use of SAE, having complementary strengths combined 918 patient records characterized by 11 clinical features to an augmented feature set that enhances the ability of the CNN to detect critical patterns correlated to cardiovascular risk [8]. Our experimental results showed that the proposed architecture could indeed provide a high level of accuracy, 88.454%, much higher than traditional classifier architecture through MLP and Random Forest by about 88.135%. With an improvement of more than 4.4%, it marked a high utility value with the approach of using deep learning techniques for medical diagnoses [9]. Furthermore, at 200 features, the best latent space size maximizes the predictive performance of the model [10], [11]. This research work is not only going to achieve high accuracies but also underpins how methodologies in deep learning might change the ways of clinical practice by giving healthcare professionals strong tools for risk assessment [12]. Improved predictive models would then translate into better patient outcomes and efficient delivery of health through early identification for timely intervention [13]. This goes beyond the mere improvement of accuracy: it also opens a vista for deep learning methodologies to affect clinical practice with such strong risk assessment weaponry for health professionals [14]. Models likethis, that go one step further in refining early detection and intervention, may substantially impact improvement in patient outcomes and effective delivery in health care delivery [15]. It underlines, in other words, the strong impact of deep learning on effective improvements to predictive modeling for heart disease and some of the most important inventive feature extraction techniques for attacking the complexity involved in this very important topic of cardiovascular risk assessment.

2 Related Work

Before the start of the mission, one carried out a literature survey over a wide range of research papers regarding this area, in order to recognize which dataset was used and how the models were skilled. The case study gives insight into transport ahead in the following assignment:

According to various researchers like Mana Saleh Al Reshan et al., a robust heart disease prediction system is one that involves the usage of CNN and LSTM networks based on HDNN. It managed an accuracy of 98.86% on a few datasets of heart diseases outperforming other techniques. It has also discussed the effectiveness of a classifier, namely, an extra tree classifier for feature selections. Future work on achieving better model adaptability includes looking at [1] deep ensemble learning techniques. The paper which is proposed by S. Ghorashi et al. discusses how regression analysis is applied to get the symptoms for CVDs. Here, SLR and MLR have been used on clinical datasets in order to study critical symptoms such as chest pain and fatigue. SPSS software analyzes data so that improved diagnoses can be attained with predictive modeling. This, in essence, confirms the importance that symptom overlap brings to the fore [2] regarding cardiovascular conditions. The paper which is proposed by Abdulwahab Ali Almazroi et al. also presents a review of the already existing machine learning frameworks and points out their shortcomings. A new scheme is proposed, which would be based on data imputation and Locality Preserving Projection for feature selection. This study further calls for optimizations to increase [3] the clinical performance of this model. Related work should revise the existing literature about machine learning techniques applied to ischemic disease prediction, summarizing models such as Support Vector Machines and Random Forest. It also should revise the datasets, including but not limited to UCI heart disease and Kaggle cardiovascular disease, in order to describe the gaps in feature representation at high risk. Comparisons with [4] state-of-the-art methods will contextualize the contributions of the proposed framework.

The paper which is proposed by Sumit Sharma et al. focuses on the working of the deep learning methodologies regarding the prediction of heart diseases through Talos hyperparameter optimization. This study has been done on the Heart Disease UCI dataset after comparing a set of machine learning algorithms that comprises K-NN, SVM, Naive Bayes, Random Forest, and logistic regression. While the best among those was the proposed deep learning model [5] with Talos optimization, which showed an accuracy of 90.78 %. Conclusion: The study identified that heart disease prediction for a medical application can effectively be improved by using deep learning models. This work which is proposed by Ali M. A. Barhoom et al is concerned with the increased prevalence of heart diseases and the need to predict them early for better results. The paper, therefore, proposes a model comprising several machine and deep learning algorithms that estimate the probabilities of a patient having a heart disease given a set of medical features. This research, based on the analysis [6] of 18 attributes in the dataset derived from Kaggle, tries to enhance the ability of health professionals to identify people at risk with great efficiency. The authors Sadia Arooj et al. have suggested a CNN-based classifier, which incorporates the multihead self-attention mechanism for heart disease prediction. The model was trained on Google Colaboratory; at the end of 100 epochs, it reached 91.71 % with respect to validation accuracy, while precision was 88.88 %, recall was 82.75%, [7] and the F1 score was 85.70 %. It had outperformed many other existing techniques and showed the major contributions of deep learning in medical diagnostics of heart disease. A paper which is proposed by Syed Nawaz Pasha et al. conducted by Syed Nawaz Pasha et al. reviews several machine learning algorithms-SVM, KNN, DT, and ANN-over the outcome of heart disease using the Kaggle dataset. This concludes that the ANN model performs best among the other models, with an accuracy as high as 85.24 %. The current study is designed in a way to indicate that [8] the only possible effective diagnosis and treatment of cardiovascular diseases are those done with the most possible accurate prediction models, exhibiting deep learning potential in medical diagnostics. The paper which is proposed by Chintan M. Bhatt reviews related work in heart disease prediction using machine learning.

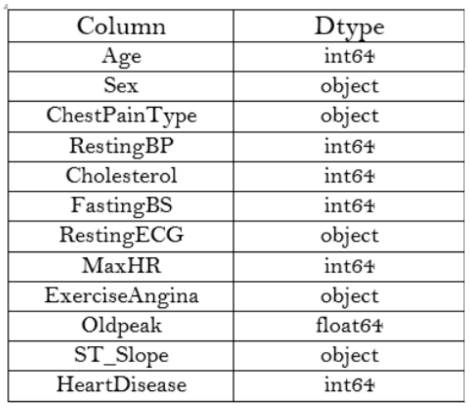
This paper insists that it is necessary to compare the clustering algorithms, including k-modes with other clustering algorithms like k-means; also, powerful classifiers such as random forest and XGBoost have shown very good accuracy in previous studies. Moreover, [9] the authors also outline that model generalization to unseen datasets is necessary, if robust predictions in a clinical setting are to be assured Work related to the paper which is proposed by Sivakannan Subramani1 et al. presented the trends and advancement in machine learning applications toward the prediction of cardiovascular diseases. Most of these studies have indeed shown that random forests and support vector machines are performing far better, as they [10] have been able to model nonlinear associations with much success. Furthermore, combining heterogeneous data sources has also been envisioned as a promising approach toward developing improved predictive accuracy.

3 Proposed Work

Select the most appropriate features and proceed with relevant preprocessing steps on data: handling missing values, removing duplicates, scaling features, and encoding categorical variables which is shown in Fig.3.1. Next step would involve feature transformations like scaling or applying dimensionality reduction. Finally, perform k-fold cross-validation to really ensure the correctness of the model in terms of training it by dividing your dataset into several folds.

3.1 Dataset

The dataset contains 11 clinical features that give health information about the patients, which is shown in the Fig 3.1.



Column	Dtype
Age	int64
Sex	object
ChestPainType	object
RestingBP	int64
Cholesterol	int64
FastingBS	int64
RestingECG	object
MaxHR	int64
ExerciseAngina	object
Oldpeak	float64
ST_Slope	object
HeartDisease	int64

Fig.3.1.1: Dataset

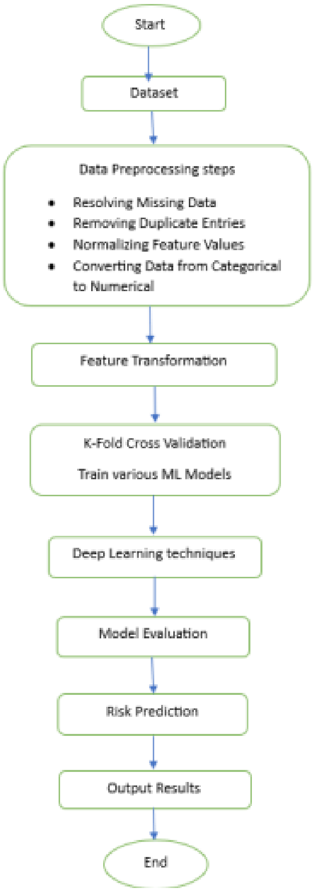


Fig.3.1.2: End-To-End Risk Prediction Pipeline.

The dataset has one binary output class, which is a heart disease diagnosis for the patient:

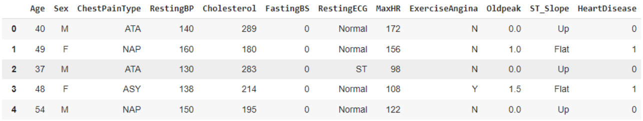
• 1: Indicates that a patient has heart disease.

• 0: Indicates the patient is healthy-that is,the patient does not have heart disease.

• Total number of samples in the dataset: 918.

3.2 Preprocessing Techniques

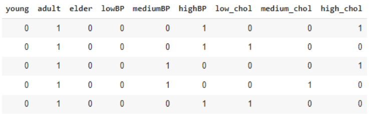
Before applying preprocessing techniques on the given dataset it is appeared in the below format which is shown in the following fig 3.2.1.



	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
0	40	M	ATA	140	289	0	Normal	172	N	0.0	Up	0
1	49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	1
2	37	M	ATA	130	283	0	ST	98	N	0.0	Up	0
3	48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
4	54	M	NAP	150	195	0	Normal	122	N	0.0	Up	0

Fig.3.2: Raw Dataset

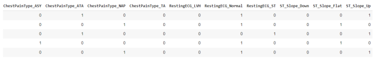
Feature Engineering: Similarly, after the feature engineering on 'Age', 'Resting Blood Pressure ', and 'Cholesterol',each of the columns went to three numerical columns as appears below.



young	adult	elder	lowBP	mediumBP	highBP	low_chol	medium_chol	high_chol
0	1	0	0	0	1	0	0	1
0	1	0	0	0	1	1	0	0
0	1	0	0	1	0	0	0	1
0	1	0	0	1	0	0	1	0
0	1	0	0	0	1	1	0	0

Fig.3.2.1: After Feature Engineering

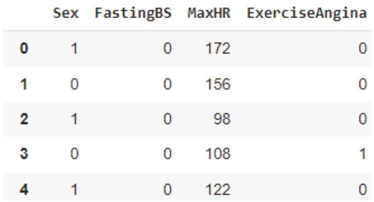
One Hot Encoding: Categorical Features ' ChestPainType', 'RestingECG', and 'ST-Slope'. One-Hot Encoding: This is a method used to transform categorical values into binary matrices, using 1 for presence and 0 for absence, with columns for each category, so that ordinal relationships are not brought in place, which goes with the flow of machine learning.



ChestPainType_ASY	ChestPainType_ATA	ChestPainType_NAP	ChestPainType_TA	RestingECG_LVH	RestingECG_Normal	RestingECG_ST	ST_Slope_Down	ST_Slope_Flat	ST_Slope_Up
0	1	0	0	0	1	0	0	0	1
0	0	1	0	0	1	0	0	1	0
0	1	0	0	0	0	1	0	0	1
1	0	0	0	0	1	0	0	1	0
0	0	1	0	0	1	0	0	0	1

Fig.3.2.2: One-Hot Encoding

Label Encoding: Preprocessed Features: Sex, ExerciseAngina This way of handling encoding refers to assigning different integers to different categories; this works for ordinal features. For non-ordinal features, one-hot encoding is favored so as not to mislead the algorithm with incorrect clues.



	Sex	FastingBS	MaxHR	ExerciseAngina
0	1	0	172	0
1	0	0	156	0
2	1	0	98	0
3	0	0	108	1
4	1	0	122	0

Fig.3.2.3: Label Encoding

Outlier Detection and Outlier Handling: Outlier Detection: Once the computation of the Z-score method is done, data would come up with the magnitude of the Z-score. That gives how many standard deviations the point is far away from the mean. Generally, a Z-score $>$ 3 can be taken as outliers.

Outlier Treatment: It replaces the outliers with their corresponding column medians; hence, the outliers are less influential over the model's performance. The shape of the data is not severely compromised as the outliers are replaced by the median value of that column.

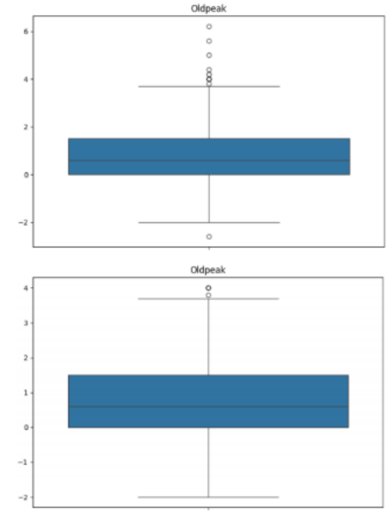


Fig.3.2.4: Before and After Removal of Outliers.

3.3 Algorithms Description

The algorithms will range from most classic machine learning, such as k-Nearest Neighbors, Random Forest, and Decision Trees, to advanced ensemble methods such as XGBoost and AdaBoost. Later, we present some deep learning techniques that will be explored in more detail: Multilayer Perceptron (MLP), whose architecture allows learning non-linear relationships present in the data through neural networks.

1) K-Nearest Neighbors : The k-NN classifier is an instance-based, non-parametric learning algorithm that classifies samples based upon the majority class of the input samples among its 'k' nearest neighbours in feature space. It measures the distance between the training sample and the given input sample, basically using the Euclidean distance, to find the closest neighbors.

$$\mathbf{x}_i = (x_{1i}, x_{2i}, \ldots, x_{pi}) \in X \tag{1}$$

2) Random Forest: Random Forest is the ensemble learning technique that at the time of training, creates a lot of decision trees and provides the mode of predictions of trees in case of classification problems. It creates diverse trees by reducing overfitting via bagging or bootstrap aggregating.
3) Decision Tree: It is in the form of a tree, where every leaf node carries the result, every branch presents the decision rule, and every internal node provides

one feature. This recursively splits the data according to the value of features with an aim to get to a model in predicting the target variable.

$$Gini = 1 - \sum_{i=1}^{c} p_i^2 \tag{2}$$

4) XGBoost: An efficient gradient boosting system using decision trees as base learners. It employs regularization to prevent overfitting; the trees are built greedily in a serial manner, correcting for errors of all previously built trees. 5) ADABoost: ADABoost stands for Adaptive Boosting. It is among one of the ensemble learning techniques that constructs a single strong classifier out of various weak classifiers, most of the time in the form of decision trees. It modifies the weight of such examples that have been misclassified previously by earlier classifiers so that later classifiers can pay more attention to those challenging cases. 6) Multilayer Perceptron: MLP is feed-forward ANN that typically consists of three layers of neurons: input layer, one or more hidden layers, and output layer. Typically it can learn any approximation of the underlying pattern of the data with just one hidden layer that contains a reasonable number of neurons. MLP is trained by backpropagation.

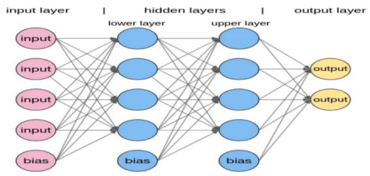


Fig.3.3: MLP Architecture

7) Naive Bayes-Gaussian: Gaussian NB is a Bayesian classifier which assumes that features are participating in Gaussian distribution by working on Bayes' theorem. It calculates posterior probabilities of each class given an input feature and then picks the class with the higher probability.

3.4 Our Proposal Method

1) Convolutional Neural Networks (CNNs): Convolutional Neural Networks represent a special deep learning model;

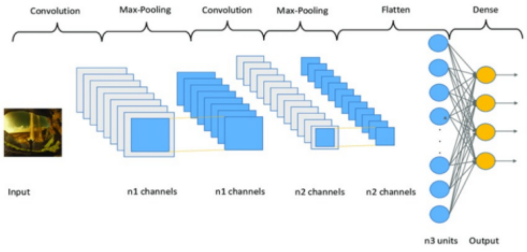


Fig.3.4: CNN Architecture

they are designed for grid-like data, for instance, images. Moreover, CNNs are translation invariant; that is, they are able to recognize patterns irrespective of their locations in input data, which is pretty useful for both image and time-series data.

3.5 Analysis of Graph

More specifically, this graph is depicting the performance of the various models with respect to the prediction of heart disease risk. Various models are included along the x-axis and their corresponding accuracies are recorded on the yaxis: RandomForestClassifier-88.1%. The best accuracy obtained using MLPClassifier was 88.454%.

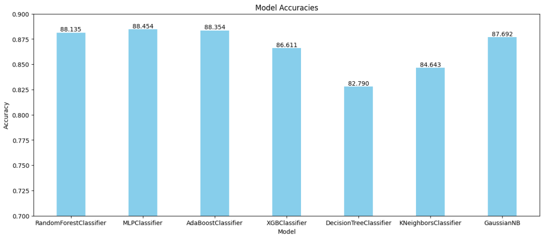


Fig.3.5: Different model Accurcies

The XGBClassifier had an accuracy of 86.6%. DecisionTreeClassifier reached an accuracy of 83.8%. While simple, decision trees can overfit and hence limit their performance. The KNeighborsClassifier had an accuracy of 84.6%. GaussianNB had an accuracy of 87.7%. As expected, the highest accuracy is achieved by MLPClassifier with an accuracy of 88.6% among these models tested.
1) MLP with SAE: This graph represents the accuracy of an MLP model with different latent sizes. This following represents the graph for the MLP:

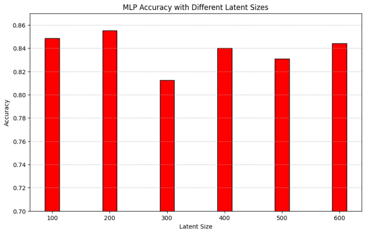


Fig.3.5.1: MLP Accuracy with different latent sizes

This usually means the number of neurons or units within a hidden layer in an MLP model. Shown on the x-axis is the value of that ranges from 100 to 600. The y-axis is the model's accuracy, ranging from 0.80 to 0. The model reaches its peak at an approximate 0.84 accuracy for a latent size of 100. The accuracy increases once more to about 0.84 when this increases to a latent size of 400. The latent sizes of 100 and 400 are where the model performs the best; in some cases, when the latent size goes beyond 100, the model decreases performance instead of increasing accuracy. The model performs best when the size of the latent varies a bit in the range 0.82-0.84.

2) CNN with SAE: The following graph represents the built CNN model by various latent sizes that represent accuracy. Here, on the x-axis, latent size represents a range from 100 to 600 in a step of 100, and on the y-axis, it is the accuracy of the CNN model. However, it only goes from 0 to 1. Latent size 100 Accuracy: 93.478%, Latent size 200 Accuracy: 90.761%, Latent size 300 Accuracy: 90.217%, Latent size 400 Accuracy: 92.391%, Latent size 50CNN with SAE: By that latent size can be seen by the x direction, that varies in stage of 100 from 100 to 600, and the y-axis represents the accuracy of the CNN model.0 Accuracy: 92.935%, Latent size 600 Accuracy : 91.848%. The accuracy is highest with the latent size of 100 and decreases from the best as it moves from Latent Size 100 to 200. After the first drop for Latent Sizes 200 and 300, accuracies rise again at Latent Sizes 400 and 500.

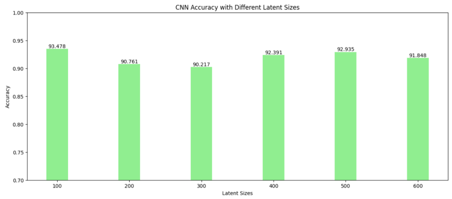
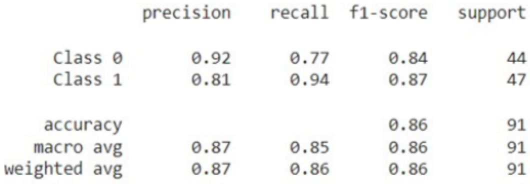


Fig.3.5.2: CNN accuracy with different Latent sizes

4 Results and Discussion

4.1 Classification report

Precision of Class 0 is 92% of the instances in Class 0. It captures only 77% of those instances, though. Class 1, which has a precision of 0.81, identifies instances equal to 81%. The general accuracy for the model is 0.86—it predicts 86% of all instances. It supports both classes with rough consistencies; it has close support values in Class 0 and Class 1.



```
              precision    recall  f1-score   support

     Class 0       0.92      0.77      0.84        44
     Class 1       0.81      0.94      0.87        47

    accuracy                           0.86        91
   macro avg       0.87      0.85      0.86        91
weighted avg       0.87      0.86      0.86        91
```

Fig.4.1: Classification Report

Precision: True positive rate of actual cases of interest among the total positive predictions. Precision = TP/(TP+FP) Recall: Number of true positives predicted to the total actual positives. Recall = TP/(TP+FN) F1-score: The harmonic average of precision and recall. F1=2 F1-score = Precision.Recall/(Precision+Recall) Support: The number of occurrences of the class in the particular data set.

4.2 Confusion Matrix

This is where this binary classification problem really shows the best of what the CNN could offer: it achieves 0.94 recall for class 1; on the other hand, it missed 23% of class 0. The model also includes committing 10 false positives in confusion between the classes. RESULT: Fewer Class 0 false positives for an overall better accuracy measurement.

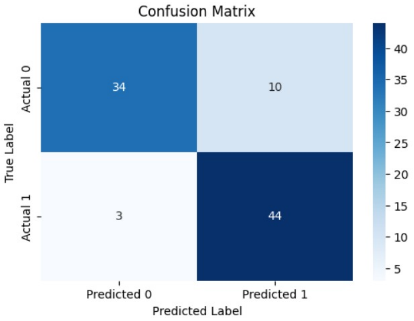
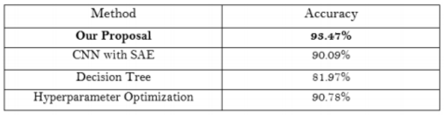


Fig. 4.2.1: Confusion matrix for CNN



Method	Accuracy
Our Proposal	**93.47%**
CNN with SAE	90.09%
Decision Tree	81.97%
Hyperparameter Optimization	90.78%

Fig. 4.2.2: Comparision with Results obtained different models

The table as shown above illustrates the accuracy rates of models used in heart disease prediction. CNN with SAE at 90.09% is trained and hyperparameter optimization was 90.78%. CNNs were used to optimize the performance in high-dimensional data, where traditional, old models of Decision Trees were at 81.97%. Therefore, the best performances were found to be 93.47%. Also, there's a misspelling in the title, and it should be rephrased into "Results Comparison from Various Models."

4.3 ROC Curve

It is a plot of the true positive rate against the false positive rate at different threshold settings of a binary classification model. The area under the ROC derived is 0.96, which is good balancing between sensitivity and specificity; hence it will be robust for classification tasks. Nevertheless, such balance may be further optimized by the adjustment of the decision threshold.

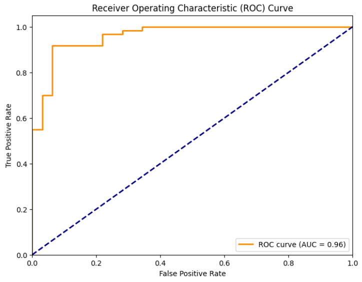


Fig.4.3: ROC Curve

5 CONCLUSION

This paper presents a deep learning framework for predicting heart disease using a data set of 918 samples with 11 clinical features. Feature augmentation was done using SAE which resulted in a more information-bearing feature besides reconstructing the data into a 2D array suitable for training CNN. The joint model SAE-CNN improved feature extraction as a result of backpropagation since CNN learns spatial dependencies. It outperformed MLP and showed an improvement of 0.6% accuracy, making it more effective for spatial data processing. MLP contributed to it, but CNN had a stronger impact on enhancing feature extraction for SAE. It reached the best performance for feature extraction when using 200 neurons in the latent space. Extra neuron addition then offered decreasing returns. This model achieved an accuracy of 93.478%, giving an improvement of 4.4% over classic classifiers such as MLP and RF. It, of course, outperformed all state-of-the-art methods based on the stacked approach with the advantage of better computational efficiency. The outcome is bound to influence patient results and generate clinical approach in practice that could improve diagnosis earlier and more efficiently.

6 DATASET AVAILABILITY

The dataset available in the link https://www.kaggle.com/ fedesoriano/heart-failure-prediction.

References

1. Mana Saleh Al Reshan, Samina Amin, Muhammad Ali Zeb, Adel Sulaiman, Hani Alshahrani, and Asadullah Shaikh, ”A study on A Robust Heart Disease Prediction System Using Hybrid Deep Neural Networks,” IEEE, 2023. https://ieeexplore.ieee.org/document/10302290
2. Sireesha Moturi, S.N. Tirumala Rao, and Srikanth Vemuru, ”Grey wolf assisted dragonfly-based weighted rule generation for predicting heart disease and breast cancer,” Computerized Medical Imaging and Graphics, vol. 91, 2021, pp. 101936. https://doi.org/10.1016/j.compmedimag.2021.101936.
3. Sara Ghorashi, Khunsa Rehman, Anam Riaz, Hend Khalid Alkahtani, Ahmed H. Samak, Ivan Cherrez-Ojeda, and Amna Parveen, ”A study on Leveraging Regression Analysis to Predict Overlapping Symptoms of Cardiovascular Diseases,” IEEE, 2023. https://ieeexplore.ieee.org/document/10151859.
4. Abdulwahab Ali Almazroi, Eman A. Aldhahri, Saba Bashir, and Sufyan Ashfaq, ”A study on A Clinical Decision Support System for Heart Disease Prediction Using Deep Learning,” IEEE, 2023. https://ieeexplore.ieee.org/document/10148957.
5. Sireesha Moturi, Srikanth Vemuru, and S.N. Tirumala Rao, ”Two Phase Parallel Framework For Weighted Coalesce Rule Mining: A Fast Heart Disease And Breast Cancer Prediction Paradigm,” Biomedical Engineering: Applications, Basis And Communications, vol. 34, no. 3, 2022. https://doi.org/10.4015/S1016237222500107.

6. Ghulam Muhammad, Saad Naveed, Lubna Nadeem, Tariq Mahmood, Amjad R. Khan, Yasar Amin, and Saeed Ali Omer Bahaj, ”A study on Enhancing Prognosis Accuracy for Ischemic Cardiovascular Disease Using K Nearest Neighbor Algorithm: A Robust Approach,” IEEE, 2023. https://ieeexplore.ieee.org/document/10239171
7. M. Sireesha, Srikanth Vemuru, and S.N. Tirumala Rao, ”Classification Model for Prediction Of Heart Disease Using Correlation Coefficient Technique,” International Journal of Advanced Trends in Computer Science and Engineering, vol. 9, no. 2, Mar.-Apr. 2020, pp. 2116–2123. https://www.researchgate.net/publication/341210689 Classification Model for Prediction of Heart Disease using Correlation Coefficient Technique.
8. Sumit Sharma and Mahesh Parmar, ”A study on Heart Diseases Prediction using Deep Learning Neural Network Model,” IEEE, 2020. https://ieeexplore.ieee.org/abstract/document/9112443.
9. Ali M. A. Barhoom, Abdelbaset Almasri, Bassem S. Abu-Nasser, and Samy S. Abu-Naser, ”A study on Prediction of Heart Disease Using a Collection of Machine and Deep Learning Algorithms,” PhilArchive, 2022. https://philarchive.org/archive/BARPOH-4.r-Prediction-of-Heart-Disease-using-Correlation-Coefficient-Technique.
10. Sireesha Moturi, S.N. Tirumala Rao, and Srikanth Vemuru, ”Predictive Analysis of Imbalanced Cardiovascular Disease Using SMOTE,” International Journal of Advanced Science and Technology, vol. 29, no. 5, 2020, pp. 6301–6311. http://sersc.org/journals/index.php/IJAST/article/view/15633.
11. Sadia Arooj, Saifur Rehman, Azhar Imran, Abdullah Almuhaimeed, A. Khuzaim Alzahrani, and Abdulkareem Alzahrani, ”A study on A Deep Convolutional Neural Network for the Early Detection of Heart Disease,” Biomedicines, vol. 10, no. 11, 2022. https://doi.org/10.3390/biomedicines10112796
12. Syed Nawaz Pasha et al., ”A study on Cardiovascular disease prediction using deep learning techniques,” IOP, 2020. https://iopscience.iop.org/article/10.1088/1757-899X/981/2/022006.
13. Chintan M. Bhatt, Parth Patel, Tarang Ghetia, and Pier Luigi Mazzeo, ”A study on Effective Heart Disease Prediction Using Machine Learning Techniques,” Algorithms, vol. 16, no. 2, 2023. https://www.mdpi.com/1999-4893/16/2/88.
14. M. Sireesha, S.N. Tirumala Rao, and Srikanth Vemuru, ”Optimized Feature Extraction and Hybrid Classification Model for Heart Disease and Breast Cancer Prediction,” International Journal of Recent Technology and Engineering, vol. 7, no. 6, Mar. 2019, pp. 1754–1772. https://www.ijrte.org/wp-content/uploads/papers/v7i6/F2343037619.pdf.
15. Sivakannan Subramani, Neeraj Varshney, M. Vijay Anand, Manzoore Elahi M. Soudagar, Lamya Ahmed Al-keridis, Tarun Kumar Upadhyay, Nawaf Alshammari, Mohd Saeed, Kumaran Subramanian, and Krishnan Anbarasu, ”A study on Cardiovascular diseases prediction by machine learning incorporation with deep learning,” Frontiers in Medicine, 2023. https://www.frontiersin.org/articles/10.3389/fmed.2023.1150933/full.