International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Healthcare Predictive Analytics based on Machine Learning Techniques for Identifying Cardiovascular Risks Screening

Jahnavi Anilkumar Kachhia*

Independent Researcher, California State University Fullerton, USA

Received 01 Dec 2023, Accepted 24 Dec 2023, Available online 26 Dec 2023, **Vol.13, No.6 (Nov/Dec 2023)**

Abstract

Cardiovascular diseases (CVDs) can be considered a severe concern to the universal health that affects the mortality rates. Clinical decision-making and early diagnosis can be challenged with the help of intelligent systems, which are based on data-driven models. The suggested solution would utilize the DenseNet-based deep learning model in making accurate predictions of heart disease using UCI Cleveland dataset. The data undergoes pre-processing with systematic data treatment of missing values, removal of duplicates, feature selection via correlation and standardization. In terms of predicting cardiac disease, the proposed model performs exceptionally well, with an accuracy of 99.65, precision of 98.38, recall of 98.77, and F1-score of 97.96. To provide comparative analysis, conventional machine learning classifiers such as KNN, MLP and Logistic Regression. The effectiveness model as demonstrated by the comprehensive performance gains indicate a high level of capturing of complex clinical patterns, which is effective in screening early cardiovascular risks in healthcare predictive analytics and is able to maintain model interpretability, hence making it a promising solution in early and automated heart disease detection in clinical environment.

Keywords: *Cardiovascular prediction, Heart disease UCI Cleveland dataset, Machine learning, DenseNet, Deep learning artificial intelligence.*

1. Introduction

Healthcare predictive analytics has now become an essential part of any contemporary healthcare system, providing an opportunity to support decision-making in disease prevention and management with the use of data [1]. Specifically, the high prevalence and mortality rate of cardiovascular diseases in all parts of the world has made the early detection of cardiovascular risks to be a priority [2][3]. Legacy screening and diagnostic systems tend to be based on constrained clinical variables and manual evaluation, possibly leading to delayed or inaccurate risk identification [4][5]. Thus, the increasing demand is the development of smart analytical systems that can be used to screen early cardiovascular risks using healthcare data.

Cardiovascular disease (CVD), the leading cause of death globally, has become a major public health issue [6]. This has resulted in high socioeconomic costs for patients, their families and the governments of these nations. Prediction algorithms that employ risk stratification can be used to identify patients who are at high risk for CVD [7][8]. Then, specific interventions for this population, such as statin usage and dietary modifications, can help lower that risk and support primary CVD prevention.

The combination of ML, as well as DL methods, to be incorporated in a single healthcare predictive analytics system, is a scalable and robust approach to early cardiovascular risk detection [9]. These smart systems not only improve clinical decision support but also provide individualised health care, better patient care outcomes, and improvement of data-driven cardiovascular disease management. In order to overcome these shortcomings, the growing popularity of deep learning methods has been based on their capacity to extract discriminative features and learn non-linear relationships in healthcare data automatically [10][11]. DL models learn non-linearity interactions about hierarchical features and auto-learn about healthcare data, which results in higher prediction performance and stability. Machine learning models, K-nearest neighbour, support vector machines and multilayer perceptron's have all been extensively studied in cardiovascular risk prediction [12]. Although these models have demonstrated encouraging outcomes, their applicability is usually limited by manual feature selection, data imbalance, and inability to scale to different populations.

*Corresponding author's ORCID ID: 0000-0000-0000-0000
DOI: https://doi.org/10.14741/ijcet/v.13.6.17

Motivation and Contributions of this Study

Early and precise detection of heart diseases because manual analysis of clinical data is time-consuming and can be affected by human error. This study employs state-of-the-art DL algorithms, such as Dense Net, to automate and optimize the diagnostic procedure of healthcare providers that assist in detecting high-risk patients at an early stage and improve the clinical outcomes. This research also has the goal of classifying heart disease by utilizing automated analysis of the clinical features of the heart disease UCI Cleveland dataset using a deep learning model. The main contribution of cardiovascular prediction is as follows:

The dataset utilized in this study is the well-known UCI Cleveland Heart Disease dataset that comprises a wide variety of clinical and demographic characteristics applicable to the diagnosis of heart diseases.

To guarantee the quality of data and model preparedness, a robust pre-processing pipeline is enforced with the standardization of missing values, duplicated records, feature selection to decrease the dimensions and increase the learning efficiency.

The proposed Dense Net model combines with the attention-based selection of features and deep learning to make predictions that are both accurate and interpretable.

To offer an in-depth overview of the model's performance, a number of measures are used, including accuracy, precision, recall, F1-score, confusion matrix, and ROC curve.

Justification and Novelty

The novelty of this work lies in the integration and incorporation of highly developed data preprocessing, feature selection, and DL based on the DenseNet model in predicting heart diseases automatically. The methodology assures good input data and learning with the application of approaches like missing value imputation, duplicate record handle, standardization, and feature selection. A combination of attention-based feature selection and deep learning allows the use of DenseNet architecture to effectively extract complex patterns in clinical characteristics, and this additionally allows the model to generate accurate and understandable predictions in detecting early signs of a disease.

Structure of the Paper

The study is structured as follows: Section II reviews the existing literature on cardiovascular disease prediction. Section III presents the methodology, including data preprocessing, model design, and implementation. Section IV provides the results and performance analysis of the proposed model. Finally, Section V presents the conclusion for future work.

Literature Review

This section reviews existing studies on machine learning–based healthcare predictive analytics for early cardiovascular risk identification. Table I summarizes the key methods, datasets, and findings from the reviewed literature:

Liu (2025) developed a cardiovascular disease risk prediction model utilizing Long Short-Term Memory Network (LSTM), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN), and evaluated their comparative performance. Analysis of EHR data from a major tertiary hospital revealed that the LSTM model achieved the highest precision, recall, and F1 score, with precision at 87.3%, recall at 84.7%, and F1 score at 89.2% [13].

Alampally et al. (2025) contrast the predictive strength of XGBoost against heart disease prediction and other prediction ML techniques, including Logistic Regression, Support Vector Machine (SVM), CatBoost, Random Forest (RF), Naïve Bayes, Decision Tree and K-Nearest Neighbour (KNN). The Kaggle Heart Disease dataset was tested, cleaned, and features were normalized, and parameters were tuned to ensure accurate results. Experiments show that XGBoost worked best, has a 95.45% accuracy rate, 4.55% error rate, 97.85% AUC, 94.93% recall, 96.15 precision, and 95.54% F1-score [14].

López-Saynes et al. (2024) employed various ML algorithms, like RF, DT, MLP, CNN, and SVM, to improve the accuracy of risk assessment. Utilizing a dataset consisting of 12 features, including the target variable, sourced from IEEE, aimed to surpass the 73% accuracy reported in state-of-the-art. To achieve this, a Variational Autoencoder-Generative Adversarial Network (VAE-GAN) was applied to generate additional synthetic data, which underwent rigorous validation procedures to ensure coherence and reliability. Results demonstrate a significant improvement in precision and reliability in cardiovascular risk assessment, with an accuracy of 90.50% achieved thus far [15].

Doiphode, Bora and Navghare (2024) employ four distinct ML algorithms, SVM, DT, RF, and KNN are used to predict the risk of cardiovascular diseases. The dataset used in this exploration consists of 14 essential features and has been sourced from Kaggle, a prominent platform for data science competitions. The primary idea is to estimate the predictive performance of these algorithms and determine which algorithm gives us maximum accuracy. Among the algorithms examined, the KNN classifier emerges as the top performer, attaining 84.48 accuracy. This result shows how effectively the KNN algorithm manages the complexity of defending against CVD attacks [16].

Prakash, Mahesh and Gouthaman (2023) The LR technique, a ML model with good accuracy and interpretability, was used to create a cardiac risk assessment system. Information from a variety of patients was included in the dataset utilized in this investigation. The LR model was trained using 13 characteristics in total, including blood pressure, cholesterol, age, and gender. With an accuracy of 86.89, the findings showed that the LR method was highly successful in predicting the risk of CVD [17].

Baghdadi et al. (2023) accessing the extensive health data on CVD that is already accessible in hospital databases has a great deal of promise for reliable, efficient, and effective ML algorithms, particularly those designed for the automated recognition of significant attributes and the prompt diagnosis of cardiovascular illness, early prediction and intervention for CVD, and lowering the concomitant financial strain on patients and the healthcare system. The suggested Catboost model produces an average accuracy of 90.94% and an F1-score of around 92.3% [18].

Pal et al. (2022) Cardiovascular disease (CVD) causes malfunction in the heart and blood arteries and frequently results in mortality or physical paralysis. Two trustworthy ML methods, K-nearest neighbour (K-NN) and multi-layer perceptron (MLP), have been used to diagnose CVD using publicly accessible Data from the University of California, Irvine repository. Experimental-based results demonstrate that a higher accuracy of 86.41% are obtained using the MLP model. Therefore, the proposed MLP model was suggested for the automated detection of CVD [19]

Table 1 Literature Review Summary on Machine Learning Techniques for Identifying Cardiovascular Disease

Author & Year	Methods/Models Used	Dataset Used	Key Findings/Results	Limitations / Future Work
Liu (2025)	LSTM, CNN, RNN	EHR data from tertiary hospital	LSTM achieved best performance with precision 87.3%, recall 84.7%, and F1-score 89.2%.	Limited to single-hospital EHR data; future work can include multi-hospital datasets and real-time clinical deployment.
Alampally et al. (2025)	XGBoost, LR, SVM, CatBoost, RF, NB, DT, KNN	Kaggle Heart Disease dataset	XGBoost achieved highest accuracy (95.45%) and AUC (97.85%).	Dataset size limited; future work can use real-time clinical data and deep learning models for improved generalization.
López-Saynes et al. (2024)	RF, DT, MLP, CNN, SVM with VAE-GAN	IEEE dataset (12 features)	Achieved 90.50% accuracy using synthetic data augmentation.	Synthetic data may introduce bias; future work should validate on real-world clinical datasets and larger feature sets.
Doiphode et al. (2024)	KNN, RF, DT, SVM	Kaggle dataset (14 features)	KNN achieved highest accuracy of 84.48%.	Limited feature diversity and dataset size; future work includes hybrid/ensemble models and larger datasets.
Prakash et al. (2023)	Logistic Regression	Clinical dataset (13 features)	Logistic Regression achieved 86.89% accuracy with good interpretability.	Limited predictive complexity; future work can integrate deep learning and feature selection techniques.
Baghdadi et al. (2023)	CatBoost	Hospital health records	Achieved 90.94% accuracy and 92.3% F1-score.	Needs validation across diverse populations; future work can explore ensemble and explainable AI models.
Pal et al. (2022)	MLP, KNN	UCI repository dataset	MLP achieved 86.41% accuracy and recommended for CVD detection.	Limited dataset size; future work includes large-scale datasets and hybrid deep learning approaches.

Research gap: Significant progress has been made in employing machine learning (ML) and deep learning (DL) techniques for predicting cardiovascular disease (CVD), there are some research gaps to be addressed. The vast majority of the available literature is based on either classical ML methods or single-application DL methods, which are both highly sensitive to manual feature selection and extremely small datasets, and thus cannot be generalized across a variety of populations. Moreover, little focus has been put on the incorporation of deep learning structures, including Dense Net, to organized clinical data, and interpretability and scalability of models in clinical practice. These gaps put a strong emphasis on the necessity of an effective, end-to-end predictive framework structure that integrates successful pre-processing, automated learning of features and thorough model comparison to improve reliable and early cardiovascular risk screening.

Methodology

The proposed methodology for the approach to cardiovascular risk identification through the use of ML techniques is shown in Figure 1.

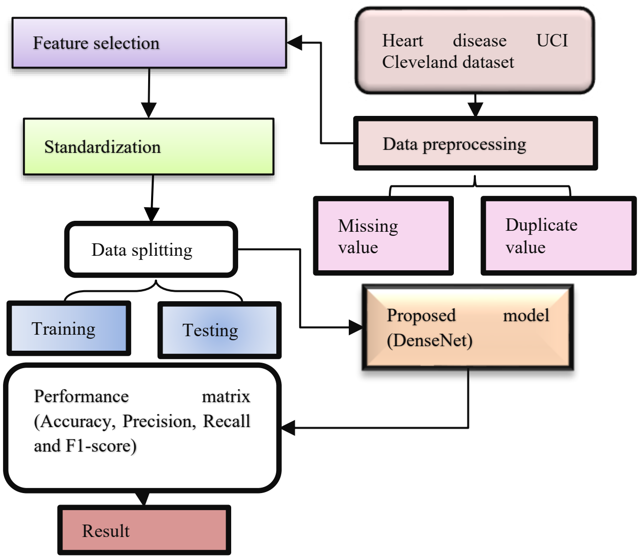

Fig.1 Flowchart for Cardiovascular Disease Using Machine Learning Model

First, the input dataset used is the Heart Disease UCI Cleveland. The data is then subjected to a data preprocessing step that involves missing values and duplicate values. After preprocessing, the most pertinent features are selected to construct the models. The data set is then separated into sets for testing and training. The proposed DenseNet model is developed using the training data. The trained models were evaluated using the standard performance metrics of F1-score, recall, accuracy, and precision on the test data set.

Data Collection

The Heart Disease Cleveland dataset, a popular resource for constructing heart disease prediction models, is accessible via the UCI ML Repository. Age, sex, number of main arteries (ca), maximum heart rate (thalach), ST depression (old peak), slope, and kind of chest pain (cp) are among the 13 clinical and demographic features it includes. Despite its small size, the dataset is often used to train and assess ML algorithms for predicting cardiovascular risk.

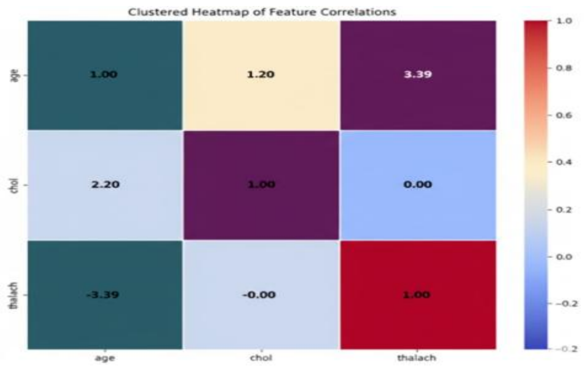

Fig.2 Heatmap of Feature Correlation

The heatmap demonstrates the relationship among the most important features, age, cholesterol (Chol), and maximum heart rate (thalach) in Figure 2.

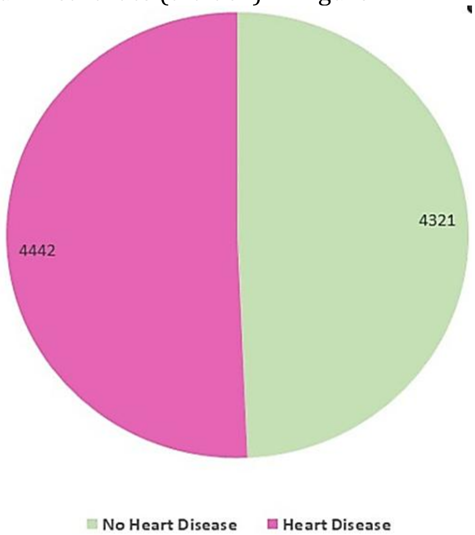

Fig.3 Pie Chart for Class Distribution

There is a moderate negative relationship between thalach and age, whereas there is a weak positive relationship between age and Chol. The fact that Chol and thalach have a low correlation (-0.00) suggests that multicollinearity is minimal, which means that each of the features gives independent contribution to prediction of heart disease.

The pie chart, as seen in Figure 3, shows that class 0 denotes no heart illness (4321 cases), and class 1 denotes heart disease (4442 instances). ML and ensemble learning models perform poorly if for the issue statement, the dataset is not balanced. In some circumstances, sampling strategies can be employed to produce a dataset that is balanced.

Data Preprocessing

Preprocessing data is the procedure that transforms unprocessed data into a clean and useful format before subjecting the data to ML models or data analysis. It includes multiple methods of missing value and duplicate record missing, transforming, and organizing the data to enhance the models' accuracy and performance. The primary phases of data preparation include:

Missing value: Missing values across all features were confirmed. The records that had large gaps were removed, and the dataset was complete and consistent because the ones with employing the proper mean or mode to impute small gaps.

Duplicate value: Examine the columns to identify any duplication. These unnecessary records were eliminated during the model's training process to avoid bias, overfitting, or data leakage.

Outlier Identification and Removal: An outlier is a data point or group of data points that differs from the values found in the remainder of the dataset [20]. In this scenario, a datapoint or data points in a dataset are those that don't fit into the larger distribution of data values.

Feature Selection

In the ML process, feature selection is crucial since the dataset may contain a large number of unrelated aspects that compromise the algorithms' accuracy. Feature selection enhances algorithm performance by minimizing such disconnected features. To determine which features were most crucial based on their relevance, it employed a variety of feature ranking strategies.

Standardization

Data can be standardized by giving values that represent the difference between standard deviations from the mean value and scaling the risk variables. By ensuring that the data is of a similar size, standardization helps avoid certain characteristics from taking center stage in the learning process [21]. The dataset's separation into training and testing sets enables model evaluation and validation of data that hasn't been made public. It is derived in Equation (1).

$$Standardization\ of\ X = \frac{X - mean\ of\ X}{standard\ deviation\ of\ X} \quad (1)$$

To improve machine learning classifier performance, it rescales the risk factor value with a mean (μ) of 0 and a standard deviation (σ) of 1.

Data Splitting

The resampled dataset is divided into a training 80% and a testing 20% sets using the train-test split technique. To guarantee that every group is fairly represented, data categorisation separates the training and testing datasets.

Proposed Dense Net model for Cardiovascular prediction

Dense Net allows each layer to access a wide range of features from earlier levels by concatenating the feature maps of all preceding layers. This method enhances overall performance by enabling the network to learn more intricate patterns. Because of the close relationships [22]. Each layer's neurons' output values undergo a nonlinear change by the activation function. In completely connected layers, ReLU can be used to improve the model's functionality [23]. Accordingly, Dense Net's structure consists of three layers: Dense Block layers, transition layers, basic convolution and pooling layers, an average pooling layer before the classification layer, and another Dense Block layer.

The Dense Net-based model's detection accuracy for cardiomegaly was assessed as follows Equation (2):

$$y = \sigma(w_x + b) \quad (2)$$

y stands for the possibility of falling into the cardiomegaly class or the probability of predicting cardiomegaly. W, x, and b stand for the weight vector, input vector, and bias, respectively. The nonlinear function known as the sigmoid function was by $\sigma(x)$, is defined by Equation (3):

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (3)$$

The input x is defined as Equation (4):

$$x = W_{Dense}.\, GAP.\, b_{Dense} \quad (4)$$

The weight vector in the last dense layer is represented by $WDense$, while the output of the global average pooling layer, which collects the characteristics that the Dense Net backbone has acquired, is represented by GAP. The bias of the dense layer is $bDense$. The outputs of all previous layers are used by each of the many convolutional layers that make up a Dense Block layer.

Performance Metrics

The experimental model's confusion matrix has been computed to determine whether the predicted class is possible. It is possible to assess the efficacy and efficiency of the implemented classifiers using performance metrics [24]. The following formulae for the specified performance measures have been used in the computations. Typically, the confusion matrix includes the four terms described below:

True Positive (TP) refers to circumstances in which actual positives are accurately anticipated to be positive.

False Positive (FP) refers to circumstances in which actuals are mispredicted as positives.

True Negative (TN) refers to circumstances in which a negative outcome is accurately foreseen.

False Negative (FN) refers to circumstances where accurate predictions of actual positives are misinterpreted as negatives.

Accuracy

A classifier's accuracy is the number of times it makes an accurate prediction, as given in Equation (5):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (5)$$

Precision

Precision is a measure of how frequently the favourable forecast is actual. It can be measured as Equation (6):

$$Precision = \frac{TP}{TP+FR} \times 100 \quad (6)$$

Recall

The number of times a class's positive value is accurately predicted is known as recall. The following formula is used to measure it Equation (7):

$$Recall = \frac{TP}{TP+FN} \times 100 \quad (7)$$

F1 Score

To compute the F-measure, divide the total of accuracy and recall by twice the precision times recall. The F-Measure equation is provided by Equation (8):

$$F1 - \text{score} = \frac{2 \times recall \times precision}{recall + precision} \quad (8)$$

ROC Curve

To assess the effectiveness of classification algorithms, the ROC curve is utilized [25]. The graph compares, at different threshold values, the FPR and the TPR, often known as recall.

Result Analysis and Discussion

In this section, the findings of the research on heart disease UCI Cleveland datasets that used DL to predict heart illness, the performance of the model, and the Python programming language, Jupyter Notebook, Google Colab, and necessary computational platform with 32 GB RAM and 8 GB VRAM NVIDIA RTX 3070, together with Python packages such as scikit-learn, keras, pandas, numPy, seaborn, tensorflow, and matplotlib were utilized to manage the computational needs of the suggested DenseNet architecture comparison analysis. The following sections present the outcomes of suggested techniques in anticipating heart illness using the DenseNet model.

Table 2 DenseNet model Performance on Heart Disease UCI Cleveland dataset

Measure	DenseNet
Accuracy	99.65
Precision	98.38
Recall	98.77
F1-score	97.96

Table II: Assessment of the suggested model's performance using the Heart Disease UCI Cleveland dataset. How well the DenseNet model performed on the Heart Disease UCI Cleveland dataset was astounding. In particular, it obtained an F1-score of 97.96, recall of 98.77, accuracy of 99.65, and precision of 99.38. The above findings demonstrate that the DenseNet structure was able to extract the inherent patterns in the heart disease data.

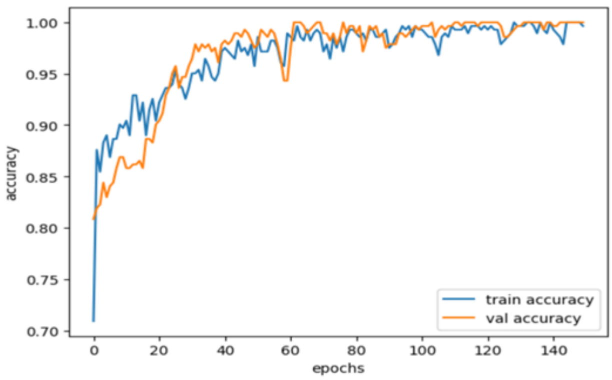

Fig.4 Accuracy graph of DenseNet model

Figure 4 illustrates how accurate the training and validation, starting as low as 75% and rising past 99% in the initial 50 epochs. The overlapping curves in great model generalization without overfitting during the training process, proves DenseNet worthy of the prediction of heart diseases with reliability.

In Figure 5 indicates that both models are effectively trained as both training and validation losses are steadily falling between 0.5 and nearly zero after 150 epochs, which confirms DenseNet's effectiveness in accurately forecasting heart illness.

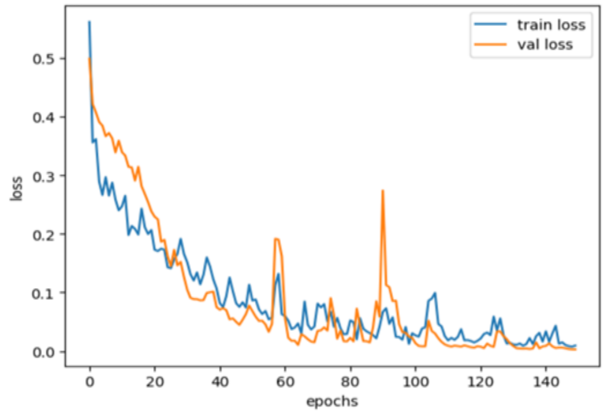

Fig.5 Loss graph of DenseNet model

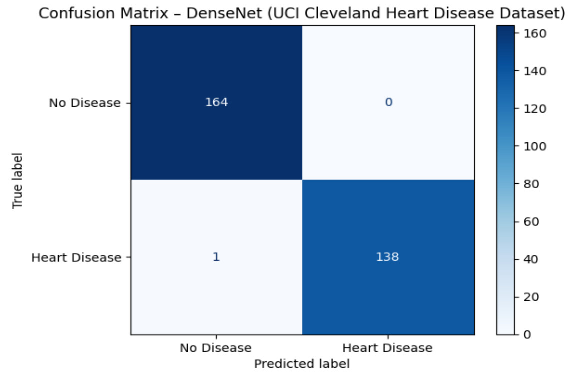

Fig.6 Confusion matrix of DenseNet model

The DenseNet model's confusion matrix utilizing the UCI Cleveland Heart Disease data demonstrates how well the model works in classification, figure 6 indicates that 138 instances of heart disease and 164 cases of no heart disease were accurately identified. There was only 1 case of misclassification of heart disease as no disease, and no false positives.

Table 3 Comparison between DenseNet and Existing models for cardiovascular disease prediction

Measure	DenseNet	KNN[26]	MLP[27]	Logistic regression[28]
Accuracy	99.65	86.45	84.25	82.9
Precision	98.38	87.53	82.08	89.6
Recall	98.77	86.21	89.43	90.2
F1-score	97.96	86.22	85.60	91.1

The model of Dense Net that is suggested displays remarkable results in terms of predicting heart diseases with an accuracy of 99.65%. The model has the ability to predict with high accuracy since it takes advantage of high levels of connectivity and extensive feature reuse in its design and therefore, captures the non-linear relationships between clinical and demographic variables well.

Conclusion and future work

Predicting heart diseases refers to the application of data-driven models to identify those who are susceptible to cardiovascular illnesses in order to get an early diagnosis, intervene promptly, and achieve better patient outcomes. In this work, the predictive framework of heart disease is developed using the DenseNet deep learning structure and is expected to improve cardiovascular risk screening. The data is preprocessed through an extensive data cleaning, normalization, and pipeline for feature pre-processing to improve the model's resilience and forecasting precision. The suggested DenseNet model's efficacy is assessed and contrasted with the traditional ML models systematically in order to prove its suitability as an effective tool to be used in healthcare context where the knowledge of the model behaviour is critical to make informed clinical judgment This quality makes the proposed DenseNet model a powerful tool to be used as a reliable model in detecting early cardiovascular risks.

The further research will be dedicated to validation of the model on increasingly extensive and diverse datasets to enhance the level of generalizability, and the implementation of the suggested method into clinical decision support systems to apply the proposed approach in practice and to positively trust AI-based healthcare solutions.

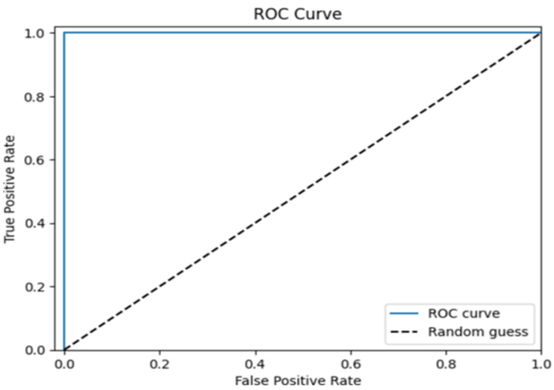

Fig.7 ROC curve of DenseNet model

The Dense Net's ROC curve is shown in Figure 7, which demonstrates how a good performance in terms of classification. A low number of FP and a high positive. The model is close to perfect discrimination of classes in heart disease prediction basing on electronic medical records.

Comparative Discussion

The performance evaluation indicates that the Dense Net network has been performing much better when compared to the conventional machine learning-based classifiers on all measures of evaluation in Table III. Dense Net demonstrated the highest accuracy of 99.65 and high precision (98.38), recall (98.77), and F1-score (97.96), which is great heart disease predictive accuracy. Comparatively the KNN model achieved an accuracy whereas the MLP model achieved 84.25% accuracy, 82.08% precision, 89.43% recall and F1-score of 85.60. Logistic regression least accurate with the accuracy of 82.9. All in all, these findings prove the high effectiveness of Dense Net in predicting cardiovascular disease risk, as opposed to traditional models.

References

[1] K. K. Mohammed, "A Survey on Digital Health Care Data Analysis Techniques for Developing Machine Learning Models," Int. J. Sci. Eng. Technol., vol. 13, no. 5, October, pp. 1–6, 2025, doi:10.5281/zenodo.17339813.

[2] M. Badawy, N. Ramadan, and H. A. Hefny, "Healthcare predictive analytics using machine learning and deep learning techniques: a survey," J. Electr. Syst. Inf. Technol., vol. 10, no. 1, p. 40, 2023, doi: 10.1186/s43067-023-00108-y.

[3] H. Alharthi, "Healthcare predictive analytics: An overview with a focus on Saudi Arabia," J. Infect. Public Health, vol. 11, no. 6, pp. 749–756, Nov. 2018, doi: 10.1016/j.jiph.2018.02.005.

[4] S. Yadav, M. R. M. Veeramanickam, R. Kumar, Shubneet, A. R. Yadav, and N. S. Talwandi, "Federated Learning for Privacy-Preserving Medical Image Analysis Across Global Healthcare Networks," Springer Nat., pp. 388–398, 2025.

[5] S. Murumkar, "AI and Machine learning in Healthcare," IJERT, vol. 14, no. 12, December, p. 6, 2025.

[6] R. Khademi, "Deep learning models for predicting heart disease risk using the UCI database: methods, performance, and clinical context," Am. J. Cardiovasc. Dis., vol. 15, no. 6, pp. 374–386, 2025, doi: 10.62347/CEBK2916.

[7] S. Subramani et al., "Cardiovascular diseases prediction by machine learning incorporation with deep learning," Front. Med., vol. 10, p. 1150933, 2023, doi: 10.3389/fmed.2023.1150933.

[8] S. N. M. Ali and N. M. Ahmed, "Comparing some Machine Learning Models for Cardiovascular Disease," J. Pioneer. Med. Sci., vol. 14, no. 04, pp. 60–66, May 2025, doi: 10.47310/jpms2025140408.

[9] S. J. Curry et al., "Screening for Cardiovascular Disease Risk With Electrocardiography," JAMA, vol. 319, no. 22, p. 2308, jun. 2018, doi: 10.1001/jama.2018.6848.

[10] T. Liu, A. Krentz, L. Lu, and V. Curcin, "Machine learning based prediction models for cardiovascular disease risk using electronic health records data: systematic review and meta-analysis," Eur. Hear. J. - Digit. Heal., vol. 6, no. 1, pp. 7–22, 2025, doi: 10.1093/ehjdh/ztae080.

[11] C. Tayal, "Data Quality Assessment and Cleaning Framework for Healthcare Databases Using Python," Int. J. Artif. Intell. Data Sci. Mach. Learn., vol. 3, no. 4, pp. 107–112, 2022, doi: 10.63282/3050-9262.IJAIDSML-V3I4P112.

[12] Y. Macha, "Cloud-Based CRM for Healthcare Data Management: A Review of HIPAA Compliance and Validation Rules," TIJER - Int. Res. J., vol. 10, no. 11, pp. 118–123, 2023.

[13] C. Liu, "Research on Risk Prediction of Cardiovascular Disease Based on Machine Learning," in 2025 4th International Symposium on Computer Applications and Information Technology (ISCAIT), IEEE, Mar. 2025, pp. 382–385. doi: 10.1109/ISCAIT64916.2025.11010381.

[14] S. P. Alampally, K. K. R. C, S. S. M, J. Puttanapura, P. Boddupally, and K. M. J, "Optimizing Cardiovascular Disease Prediction Using Machine Learning Models on Heart Disease Dataset," in 2025 7th International Conference on Information Systems and Computer Networks (ISCON), IEEE, Sep. 2025, pp. 1–6. doi: 10.1109/ ISCON65210. 2025.11341639.

[15] J. L. López-Saynes et al., "Analysis of Physiological Parameters for Assessing the Risk Level of Cardiovascular

Diseases Using Machine Learning Algorithms," in 2024 10th International Conference on Control, Decision and Information Technologies (CoDIT), IEEE, Jul. 2024, pp. 2763–2768. doi: 10.1109/CoDIT62066.2024.10708546.
[16] A. Doiphode, K. Bora, and P. R. Navghare, "Comparative Analysis of Machine Learning Algorithms for Cardiovascular Disease Risk Prediction," in 2024 1st International Conference on Trends in Engineering Systems and Technologies (ICTEST), IEEE, Apr. 2024, pp. 1–6. doi: 10.1109/ICTEST60614.2024.10576161.
[17] N. Prakash, M. Mahesh, and P. Gouthaman, "Cardiovascular Disease Risk Assessment using Machine Learning," in 6th International Conference on Inventive Computation Technologies, ICICT 2023 - Proceedings, 2023. doi: 10.1109/ICICT57646.2023.10133957.
[18] N. A. Baghdadi, S. M. Farghaly Abdelaliem, A. Malki, I. Gad, A. Ewis, and E. Atlam, "Advanced machine learning techniques for cardiovascular disease early detection and diagnosis," J. Big Data, vol. 10, no. 1, p. 144, 2023, doi: 10.1186/s40537-023-00817-1.
[19] M. Pal, S. Parija, G. Panda, K. Dhama, and R. K. Mohapatra, "Risk prediction of cardiovascular disease using machine learning classifiers," Open Med., vol. 17, no. 1, pp. 1100–1113, Jun. 2022, doi: 10.1515/med-2022-0508.
[20] S. Dalal et al., "Application of Machine Learning for Cardiovascular Disease Risk Prediction," Comput. Intell. Neurosci., vol. 2023, no. 1, 2023, doi: 10.1155/2023/9418666.
[21] A. Saboor, M. Usman, S. Ali, A. Samad, M. F. Abrar, and N. Ullah, "A Method for Improving Prediction of Human Heart Disease Using Machine Learning Algorithms," Mob. Inf. Syst., no. 15, pp. 1–9, Mar. 2022, doi: 10.1155/2022/1410169.
[22] D. V. G. Krishnan, D. M. V. V. Saradhi, D. S. S. Kumar, G. Dhanalakshmi, P. Pushpa, and D. V. Vijayaraja, "Hybrid Optimisation-based Feature Selection with DenseNet Model for Heart Disease Prediction," Int. J. Electr. Electron. Res., vol. 11, no. 2, pp. 253–261, Apr. 2023, doi: 10.37391/ijeer.110203.
[23] W. Hadi, T. Jaware, T. Khalifa, F. Aburub, N. Ali, and R. Saini, "Enhancing Cardiovascular Disease Detection Through Exploratory Predictive Modelling Using DenseNet-Based Deep Learning," Computers, vol. 14, no. 8, p. 330, Aug. 2025, doi: 10.3390/computers14080330.
[24] S. Arooj, S. ur Rehman, A. Imran, A. Almuhaimeed, A. K. Alzahrani, and A. Alzahrani, "A Deep Convolutional Neural Network for the Early Detection of Heart Disease," Biomedicines, vol. 10, no. 11, p. 2796, Nov. 2022, doi: 10.3390/biomedicines10112796.
[25] N. Chandrasekhar and S. Peddakrishna, "Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization," Processes, vol. 11, no. 4, p. 1210, Apr. 2023, doi: 10.3390/pr11041210.
[26] A. Alqahtani, S. Alsubai, M. Sha, L. Vilcekova, and T. Javed, "Cardiovascular Disease Detection using Ensemble Learning," Comput. Intell. Neurosci., vol. 2022, pp. 1–9, Aug. 2022, doi: 10.1155/2022/5267498.
[27] A. Tiwari, A. Chugh, and A. Sharma, "Ensemble framework for cardiovascular disease prediction," Comput. Biol. Med., vol. 146, 2022, doi: 10.1016/j.compbiomed.2022.105624.
[28] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," IEEE Access, vol. 7, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.