Detection of Cardiovascular Diseases in ECG Images

DR K VENKATA SUBBAIAH1, T. Geetha Yasaswini2, N. Vyshnavi3, P. Sumathi4, V. Srinivasulu5

#1Professor in Department of CSE, PBR Visvodaya Institute of Technology and Science, Kavali.

#2#3#4#5B.Tech with Specialization of Computer Science and Engineering- Artificial Intelligence in PBR Visvodaya Institute of Technology and Science, Kavali.

Abstract_ Cardiovascular diseases encompass various conditions affecting the heart and blood vessels, including myocardial infarction, commonly known as a heart attack. This occurs when blood flow to a part of the heart is obstructed, leading to damage or death of heart muscle tissue. Symptoms of myocardial infarction may include chest pain, shortness of breath, fatigue, and dizziness. Early detection and management of cardiovascular diseases are crucial for preventing complications and improving patient outcomes.

With the use of deep learning and machine learning models, this research attempts to digitize paper ECG records for automated diagnosis. The ECG report is divided into 13 Leads, which are then turned into signals and scaled to binary pictures using thresholding, scaling, and smoothing. Principal Component Analysis and other dimension reduction techniques are used after feature extraction. The method combines a voting classifier with SVM, KNN, ANN, Logistic Regression, and another pathway that uses Convolutional Neural Network (CNN) to extract patterns from ECG images. Metrics like accuracy, precision, recall, and F1-score are used to compare the outcomes from the two pathways. About 90.68% accuracy is attained by the technique. Finally, using ECG reports, the model determines if the patient has a myocardial infarction, an abnormal pulse, or a normal state of health.

Keywords – Voting Classifier, Convolution Neural Network

1. INTRODUCTION

Heart disease is the second most common cause of mortality in low-income countries and the top cause of death in high-income countries, according to the World Health Organization. For the past 20 years, it has continued to be the world's greatest cause of death. Cardiovascular disease (CVD) is a major global health concern that

impacts millions of people globally. It includes a range of disorders that impact the heart and blood arteries, such as peripheral artery disease, heart attacks, and strokes. High blood pressure, smoking, physical inactivity,
poor diets, obesity, and physical inactivity are risk factors that lead to the high prevalence of CVD in a variety of populations. Significant effects of CVD include disability, a lower quality of life, and an early death.

Reports from electrocardiograms (ECGs) are a vital source of information on the electrical activity of the heart, which helps with both diagnosis and condition monitoring. The heart's rhythm, pace, and any anomalies in its electrical conduction channels are all visible on an ECG recording.

The patient's heart rate, rhythm, existence of any arrhythmias, and any symptoms of ischemia or myocardial infarction are usually included in the ECG report. Expertise is needed to interpret ECG results since even minute variations in the waveform can reveal serious heart disease. Technological and machine learning developments have made it possible to automatically analyze ECG reports, which has sped up and improved the identification of heart problems. Furthermore, by digitizing paper ECG records, patient information may be shared, stored, and retrieved more easily, which streamlines healthcare delivery and enhances patient outcomes.

CATEGORY	DESCRIPTION
Normal	Heartbeat within parameters, Regular cardiac rhythm, no abnormalities detected
Abnormal	Chest pain, shortness of breath, fatigue, dizziness
Myocardial Infraction	It is commonly known as heart attack, occurs when blood flow to a part of the heart is blocked for an extended period leading to damage or death of heart muscle tissue.
History of Myocardial Infraction	History of myocardial infraction, signifies prior heart muscle damage due to the restricted blood flow.

Table 1.1 Different Conditions of ECG Reports

The above Table1.1, describes the different conditions of ECG reports.

While traditional diagnostic methods rely on clinical assessments and invasive procedures, recent advancements in medical imaging technology and machine learning, particularly through electrocardiography (ECG), offer non-invasive approaches for disease prediction and diagnosis. Leveraging CNN for analyzing ECG images has shown promise, allowing for the extraction of patterns and automation of diagnosis through digitization of ECG records. This approach aims to expedite the detection and analysis of CVD, ultimately improving patient care and management strategies.

2. LITERATURE SURVEY

i) Park and Kim (2015) presented a model tailored to the Korean population for predicting CVD by leveraging machine learning and deep learning methods. Their model achieved high accuracy rates, surpassing traditional machine learning approaches, indicating the effectiveness of deep learning in predicting CVD without heavy reliance on feature engineering.

ii) Choi et al. (2017) investigated the use of recurrent neural network (RNN) models to predict HF onset based on temporal relations among events in electronic health records (EHRs). The study demonstrated that RNN models, particularly those using gated recurrent units (GRUs), outperformed conventional methods in detecting incident HF, showcasing the potential of deep learning models to improve early detection of cardiovascular conditions.

iii) Poplin et al. (2018) explored the application of deep learning to predict cardiovascular risk factors from retinal fundus photographs. Their study showed that deep learning models trained on retinal images could accurately predict various risk factors, such as age, gender, smoking status, and blood pressure, highlighting the capability of deep learning to extract valuable information from medical images for cardiovascular risk assessment.

iv) Stewart and Sun (2016) developed Doctor AI, a predictive model based on recurrent neural networks, for anticipating clinical events using EHR data. Doctor AI demonstrated effectiveness in performing differential diagnosis and showed generalizability across different healthcare institutions, indicating its potential for improving clinical decision-making and patient care.

3. PROPOSED SYSTEM

We provide an approach that creates a 1-D signal from ECG recordings using deep learning and machine learning algorithms. On our online platform, users input an ECG image and then go through a number of image processing processes. These include of separating the important parts of the ECG signal, converting photos to gray scale, and eliminating extraneous information. For analysis, the captured signal is subsequently converted into a straightforward 1-D representation. After feeding the normalized 1D signal into our machine learning and deep learning models that have already been trained, the data is analyzed.

These models collaborate by using a voting classifier technique, which improves the analysis's robustness and accuracy. Upon completion of the analysis, the model provides the user with the results based on its findings. We present the findings in an understandable manner to enable people to better care for themselves by being aware of their cardiac condition.

For the purpose of classifying our ECG images, we have employed four categories. Normal, Myocardial infarction, Abnormal, Myocardial infarction history.

4. MODULES DESCRIPTION

4.1 Data Collection and Pre-processing

ECG lead data is stored in a structured manner (CSV file), which makes operations related to analysis and machine learning easier. Filenames are sorted, CSVs are concatenated, and target labels are numerically encoded during preprocessing. With a total of roughly 99.8%, Principal Component Analysis (PCA) provides insights into the variance explained by each component. To help with feature selection, the descending order of variance explained is emphasized. To improve ECG analysis and classification, the resulting `result_df} contains PCA-transformed data with 100 components and target labels.

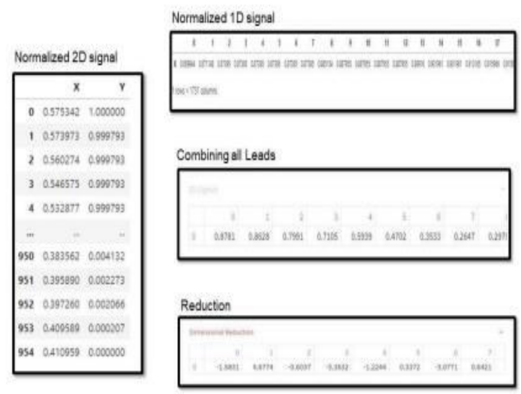

Fig 4.1 Feature Transformation

This method simplifies data processing and improves the effectiveness of later machine learning algorithms.

4.2 Model Configuration

Several machine learning algorithms, each with a unique set of hyperparameters, make up the model:

- Support Vector Machine (SVM) with gamma = 0.01 and C = 100 parameters.

• An artificial neural network (ANN) with 50 neurons in the hidden layer.

• One nearest neighbor when using k-Nearest Neighbors (KNN).

• 300 decision trees in a Random Forest (RF).

4.3 Models Training

Split your dataset into training, validation, and testing sets. Train the CNN model using the training data. Implement hyperparameter tuning to optimize model performance. Monitor the validation accuracy and loss during training to prevent overfitting.

***Soft Voting Classifier*:** By averaging the probability distributions of each class's predictions from various base classifiers, the soft voting classifier integrates the predictions. Unlike hard voting, it chooses the class with the highest average probability as the final prediction, producing a smoother decision boundary. By utilizing the advantages of each classifier and minimizing its disadvantages, an ensemble approach enhances predictive performance. Soft voting improves overall generalization performance when base classifiers show distinct sorts of errors or display varying levels of competence.

Convolution Neural Network (CNN): For identifying cardiovascular anomalies, CNNs are essential in the automated feature extraction and classification from raw ECG images. Convolutional filters and fully connected layers are two examples of parameters that CNNs optimize across ten epochs, using labeled datasets to train in order to reduce loss and improve accuracy. Approximately 93.01% of the test dataset's cases were correctly classified, according to the test accuracy stated. The hierarchical architecture of CNNs enables them to learn complex features at different abstraction levels, which helps them identify subtle patterns that are important for cardiovascular diagnosis.

4.4 Models Evaluation

The following Table 4.1, are some of the hyperparameters for the ensemble model that have been optimized for the dataset.

Support Vector Machine	**'SVM__C': 100, 'SVM__gamma': 0.01**
Artificial Neural Network	**'ann__hidden_layer_sizes': (50,)**
K –Nearest Neighbours	**'knn__n_ neighbors': 1**
Random Forest	**'rf__n_estimators': 300**

Table 4.1 Best Hyperparameters

Utilizing a combination of deep learning architectures and conventional ML techniques, the ensemble model achieves an impressive accuracy rate of 90.68%, showcasing the effectiveness of optimized hyperparameters in enhancing predictive performance.

A CNN gathers patterns from ECG pictures, but a Voting Classifier aggregates predictions from different methods. The accuracy of CNN model is approximately 87% Both are trained and assessed using labeled data, with precision and accuracy being compared. The final prediction, which determines the heart problems from ECG readings, is guided by the both soft voting classifier and the DL models (CNN) results.

5. RESULTS AND DISCUSSION

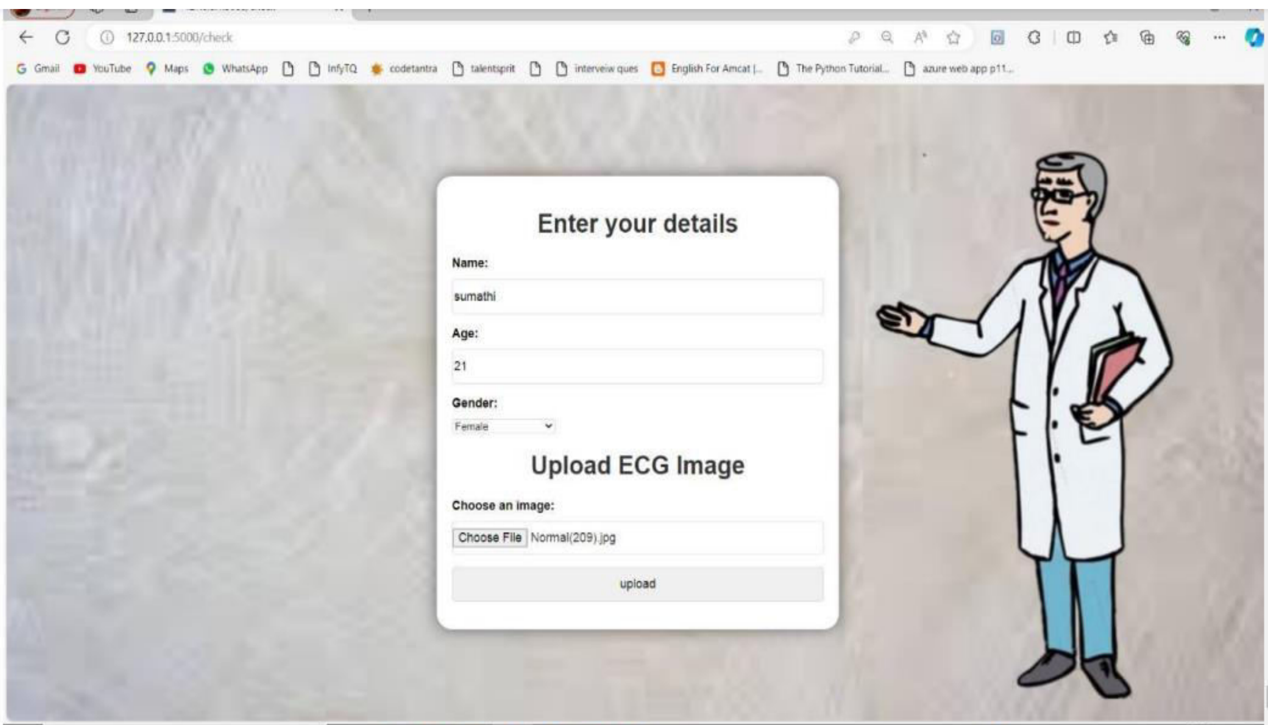

Fig 5.1 ECG Uploading Page

Based on the heath condition of the user, the corresponding display format will appear. Display format like,

- if condition-normal: Screen will appear in green color
- if condition-abnormal: Screen will appear in red color

- if condition- Myocardial Infarction: Screen will appear in orange color
- if condition- History of Myocardial Infarction: screen will appear in blue color

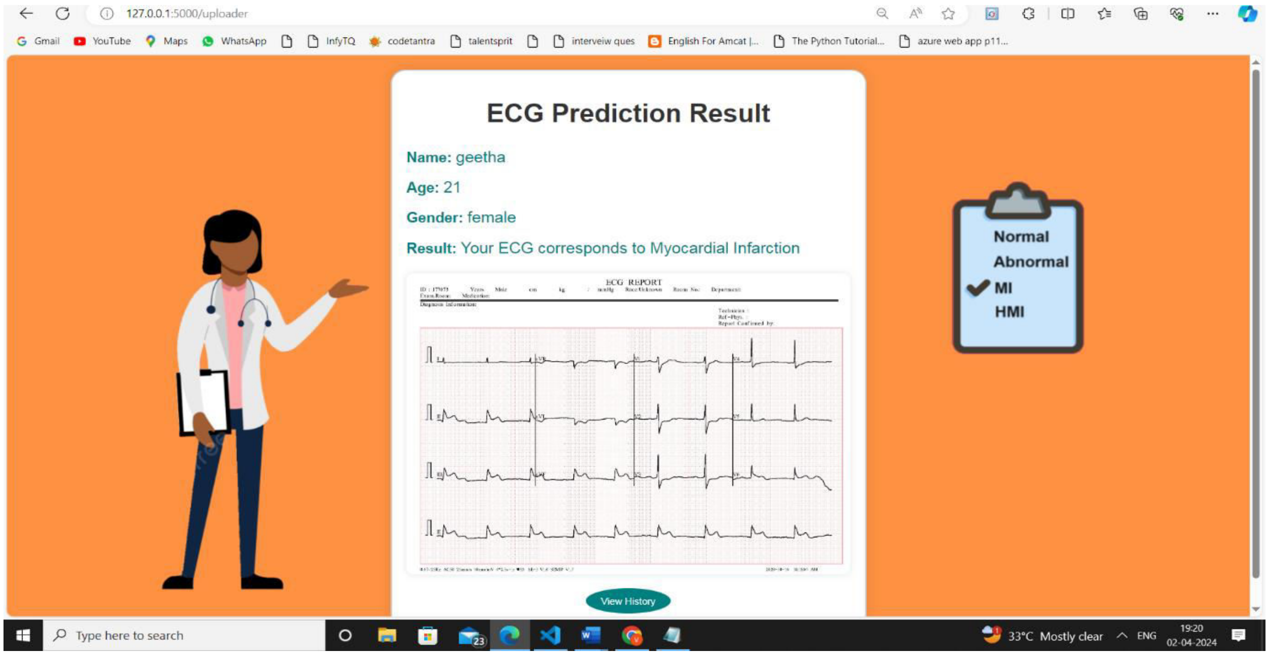

Fig 5.2: Health Condition: Myocardial Infarction (Orange)

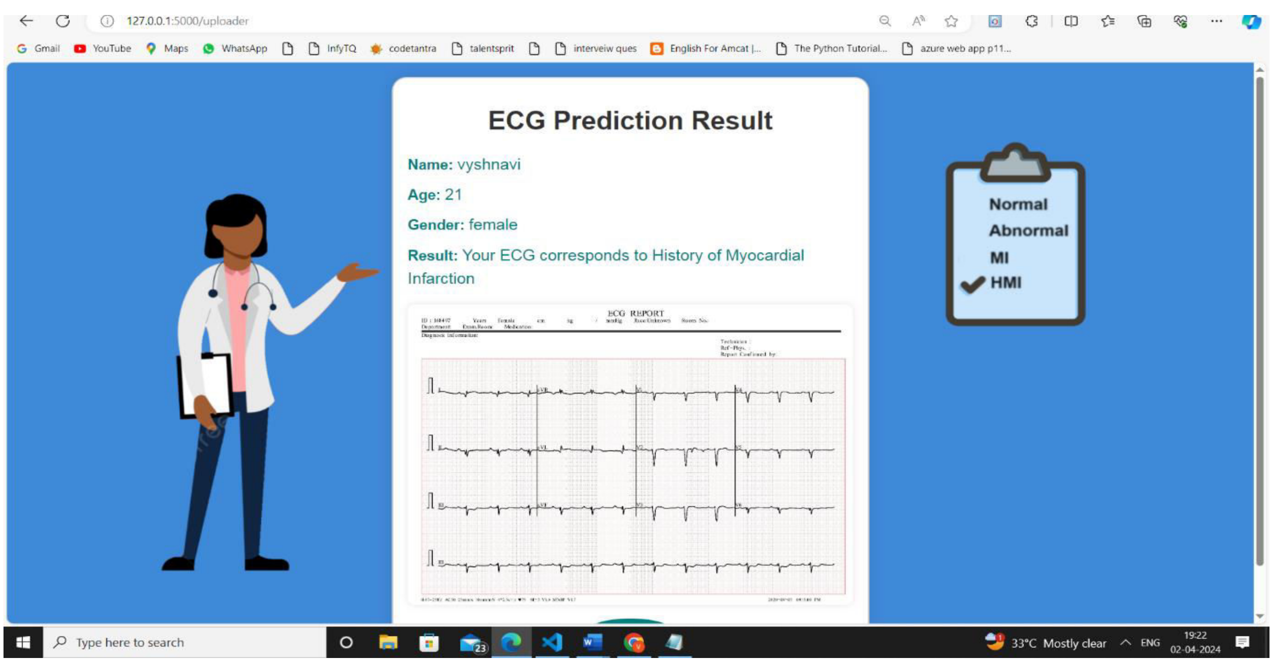

Fig 5.3: Health Condition: History of Myocardial Infarction (Blue)

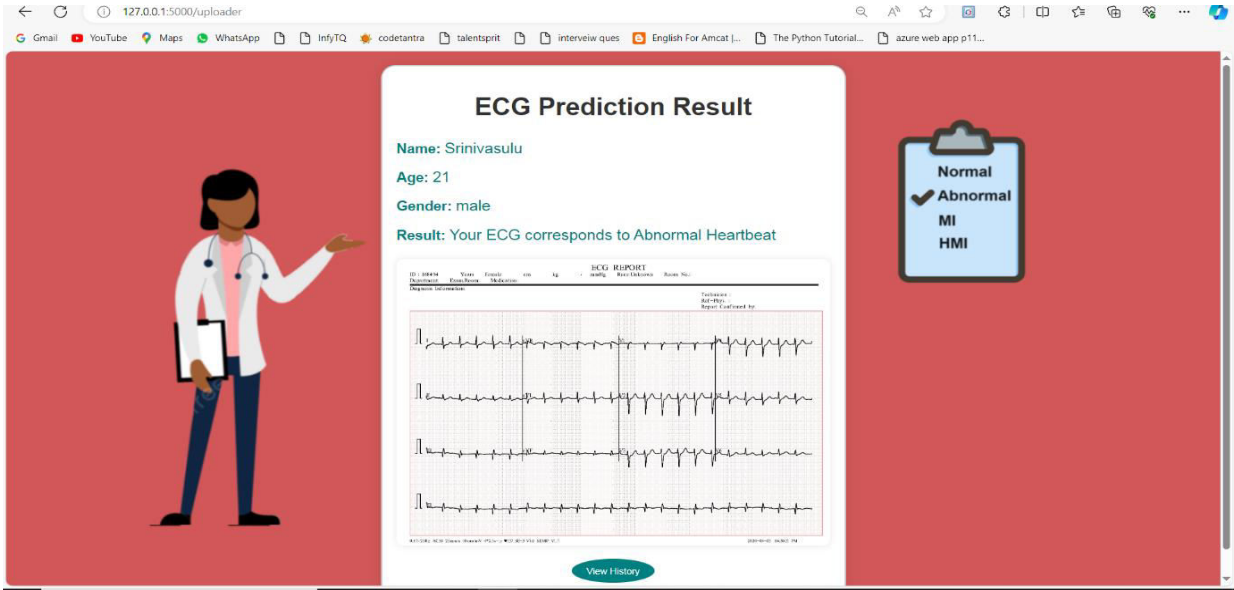

Fig 5.4: Health Condition: Abnormal (Red)

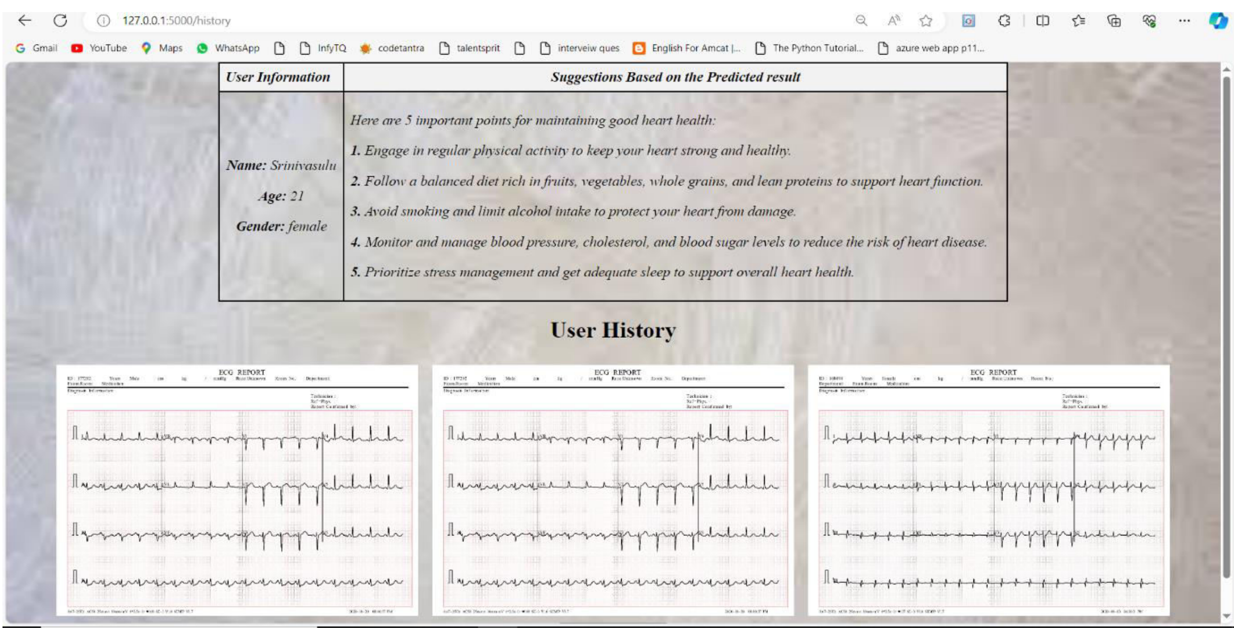

Fig 5.5: History Page and Logout

6. CONCLUSION

By creating an automated system with ML and DL, this study addressed issues with manual ECG interpretation. Raw data was prepared for analysis by means of preprocessing procedures and feature extraction. With independent analysis of ECG data, the Voting Classifier and CNN obtained accuracies of almost 90.68% and 87.68%, respectively. A final prediction was produced by outcome comparison. The model helps physicians diagnose and treat patients by identifying a variety of cardiovascular disorders. Its clinical utility can be improved with additional refining and validation on larger datasets. The project is significant because it has the potential to simplify ECG interpretation and provide medical practitioners with a useful tool. It's a start in the right direction toward bettering patient outcomes via early illness detection. To ensure robustness and generalizability, future work will require improving the model and evaluating it on a variety of datasets.

7. FUTURE SCOPE

Automated cardiovascular disease diagnosis will be further improved using a variety of approaches, including ensemble learning for better performance, model design and hyperparameter tuning, and data augmentation for model generalization. Real-time processing facilitates timely diagnosis, and the integration of extra clinical data improves interpretability and context. Validation and improvement are ensured by working with domain experts.

Trying out more complex CNN architectures, such as Transformer-based models and ResNet, can improve accuracy. Predictive powers are enhanced by adding lifestyle variables, medical history, and demographic data. Relevance is guaranteed by putting a feedback loop in place for ongoing model improvement. Creating intuitive mobile applications encourages proactive health management on the part of the user. Through increased clinical relevance, efficacy, and efficiency of automated diagnostic systems, patient outcomes and treatment quality will eventually be improved.

8. REFERENCES

1. G.-M. Park and Y.-H. Kim, "Model for predicting cardiovascular disease: Insights from a Korean cardiovascular risk model," Pulse, vol. 3, no. 2, pp. 153–157, 2015.

2. Choi, E., Schuetz, A., Stewart, W. F., & Sun, J. (2017). Using recurrent neural network models for early detection of heart failure onset. Journal of the American Medical Informatics Association, 24(2), 361-370.

3. Poplin.R, Varadarajan.A.V, Blumer.K, Liu.Y, McConnell.M.V, Corrado.G.S., ... & Topol, E. J. (2018). Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning.

Nature Biomedical Engineering, 2(3), 158-164.

4. Stewart, W. F., & Sun, J. (2016). Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. Journal of Machine Learning Research, 56, 301-318.

DR K VENKATA SUBBAIAH received his Ph.D from Rayalaseema University, Kurnool in 2017 in Digital Image processing. He has total 26 years of teaching experience and his area of interests are Machine learning, Deep learning and Image processing. Currently he is working as a professor in Dept of CSE, PBR VITS, KAVALI.

T. Geetha Yasaswini, B.Tech with Specialization of Computer Science and Engineering- Artificial Intelligence in PBR Visvodaya Institute of Technology and Science, Kavali.

N. Vyshnavi, B.Tech with Specialization of Computer Science and Engineering-Artificial Intelligence in PBR Visvodaya Institute of Technology and Science, Kavali.

P. Sumathi, B.Tech with Specialization of Computer Science and Engineering-Artificial Intelligence in PBR Visvodaya Institute of Technology and Science, Kavali.

V. Srinivasulu, B.Tech with Specialization of Computer Science and Engineering-Artificial Intelligence in PBR Visvodaya Institute of Technology and Science, Kavali.