AI-BASED DISEASE DETECTION USING MACHINE LEARNING AND CONVOLUTIONAL NEURAL NETWORKS

Dr. K. Soumya[1], Gorla Sukanya[2], Jakkula Madhurima[3], Jeeri Alekhya Reddy[4], Kallakuri Vyshnavi[5], Dr. S. Pallam Setti[6]

[1] Assistant Professor, Dept. of Computer Science and Systems Engineering, Andhra University College of Engineering for Women, Andhra Pradesh, India
[2-5]B.Tech, Final year Student, Andhra University College of Engineering for Women, Andhra Pradesh, India
[6]Founder & CEO, Dr Pallam Setti Center for Research & Technology, A-hub, Andhra University, Visakhapatnam, Andhra Pradesh, India

---***---

Abstract - *Artificial Intelligence serves as a vital healthcare component by detecting diseases with excellent precision. Biomedical researchers utilize Convolutional Neural Networks (CNN) with machine learning techniques to extract image features and sequential data through this AI-Based Disease Detection system which leads to improved model performance. The system requires different types of data input such as images and patient records and time-series measurements for chronic kidney disease and Parkinson's disease diagnosis. This hybrid approach which uses various neural networks ensures robust feature learning, temporal pattern recognition and optimization making it suitable for real world diseases and clinical applications. Experimental results demonstrate reduced false positives, improved accuracy, and better generalization across datasets. The AI system enables swift early diagnosis of diseases, feature selection and enhances medical care while facilitating better deliverance of health results to patients with applications of deep learning.*

Key Words: AI in healthcare, disease detection, CNN, chronic kidney disease, Parkinson's, medical diagnostics, machine learning, early diagnosis, feature extraction, deep learning.

1.INTRODUCTION

In past years, use of Artificial Intelligence (AI) in healthcare has completely changed how illnesses are identified and treated. For better treatment, an early and precise diagnosis is essential, but traditional diagnostic techniques are costly, time-consuming procedures, and depend on availability of medical professionals or assistance with specialised medical knowledge.

To address these challenges, our project, **"AI-Based Disease Detection System"** aims to develop a comprehensive, user-friendly web application that facilitates the early detection of Chronic kidney disease and Parkinson's disease. The application allows users to input relevant medical information which is then processed by specialized AI models to provide accurate predictions. This disease detection system aims to serve as an effective decision-support tool for individuals seeking initial medical evaluations.

The core technology behind this system uses deep learning algorithms, particularly Convolutional Neural Networks (CNNs) and various Machine Learning(ML) algorithms. CNNs are employed for image-based diseases. On the other hand, Machine Learning algorithms are utilized for analyzing sequential medical data, such as patient health records, time-series data which are critical for conditions like chronic kidney disease, and Parkinson's disease. The user-centric design ensures that individuals with minimal technical knowledge can benefit from AI-driven diagnostics, making healthcare more reachable.

Despite the advancements in AI and machine learning for disease detection, most existing systems are limited to identifying a single disease. This lack of versatility forces individuals to rely on multiple tools for different conditions, leading to increased complexity and inefficiency in the diagnostic process. Additionally, many AI-based diagnostic applications are not user-friendly, making them difficult to access for individuals without technical expertise. So, this project aims to address these challenges by developing an AI-based disease detection system using Machine Learning with CNNs through a user-friendly web application.

Most existing AI-based diagnostic systems are designed to detect only one specific disease, requiring separate tools for different conditions. This fragmentation increases the complexity for healthcare professionals and patients, making the diagnostic process inefficient. Existing techniques often struggle to integrate different types of medical data (e.g., images, clinical text, lab results) effectively. This limits the ability to provide a comprehensive diagnosis based on multiple data sources. Many AI-powered healthcare tools are designed for medical professionals who have technical knowledge, making them inaccessible to the general public, especially in rural areas.

The system incorporates real-time predictive analytics powered by advanced machine learning algorithms, enabling it to not only detect diseases but also predict potential health risks based on historical and current patient data. It provides healthcare professionals with actionable decision support by highlighting critical insights. The proposed system is built with a modular architecture, allowing it to scale and adapt to include additional diseases or data types as new medical research and datasets become available. This ensures long-term relevance and flexibility, enabling the platform to evolve with advancements in machine learning techniques and healthcare needs, such as integrating emerging diagnostic markers or supporting new imaging technologies.

2. REVIEW OF LITERATURE

The survey explores how deep learning transforms medical image analysis through applications in segmentation, classification, and disease detection. It highlights the ability of convolutional neural networks to process complex visual data, offering insights into its widespread adoption in clinical settings. The work underscores the technology's potential to enhance diagnostic precision and efficiency [1]. Deep convolutional adversarial networks are employed to synthesize high-quality medical images, addressing the scarcity of labelled data for training diagnostic models. These synthetic images mimic real patient scans, enabling robust testing and validation of AI systems. This innovation holds promise for advancing research in medical imaging and diagnostics [2]. Machine learning facilitates disease prediction by integrating diverse patient data, including imaging, lab results, and clinical histories. This approach allows for the early identification of multiple conditions, enhancing preventative care strategies. It demonstrates the power of data-driven insights in improving healthcare outcomes [3]. A hybrid CNN-RNN model combines convolutional and recurrent neural networks to predict multiple diseases by analyzing spatial and temporal patient data. This method excels in capturing both image-based anomalies and time-series trends, offering a comprehensive diagnostic tool. Its success highlights the synergy of advanced architectures in medical AI [4]. Machine learning identifies Parkinson's disease by analyzing motor symptoms and physiological data, such as tremors and gait patterns. This approach offers a non-invasive, data-driven method to support early diagnosis and monitoring. It highlights the versatility of AI in tackling neurological disorders [5]. Efficient detection of Chronic Kidney Disease (CKD) relies on selecting the most relevant predictors while maintaining high accuracy. Feature selection techniques play a crucial role in minimizing computational complexity without compromising performance. Various machine learning models, including Support Vector Machines (SVM), Decision Trees, Random Forest, and k-Nearest Neighbors (k-NN), have been explored for CKD classification, optimizing accuracy and handling imbalanced datasets effectively [6]. Deep learning further enhances healthcare analytics by processing time-series data, such as patient vitals and laboratory results, for improved disease prediction. Its ability to model temporal dependencies makes it invaluable for critical care monitoring, forecasting patient deterioration, and supporting timely medical interventions [7]. A hybrid deep learning model integrates imaging, genomic, and clinical data for disease diagnosis, improving predictive accuracy. This comprehensive approach allows clinicians to address complex cases involving overlapping conditions. It represents a step towards personalized medicine through AI [8]. Transfer learning accelerates medical image classification by adapting pre-trained models to specific diagnostic tasks, minimizing training time. This technique leverages knowledge from large datasets, making it efficient for specialized medical applications. It broadens the accessibility of AI in clinical practice [9].

3.METHODOLOGY

The AI-Based Disease Detection System functions as an advanced framework which discovers diseases through analysis of different medical data types like medical images and patient details and laboratory test output. The system makes medical diagnosis more exact through deep learning Convolutional Neural Networks (CNNs) for image analysis which helps medical professionals detect diseases early. This innovative system possesses built-in scalability features with modular architecture that adapts effectively to diverse medical areas. This system focuses on disease identification from X-rays and CT scans and blood test reports when completing accurate disease predictions quickly. The system performs disease predictions at once which enables doctors to get complete diagnoses in one session without any hassle. The system processes massive amounts of medical information that consists of scans, reports, images and user-supplied symptoms which leads to quick computing with little wait time. The system brings together a user-friendly healthcare professional dashboard which improves both accessibility and ease of use. The system deployment of this work is shown in figure1.

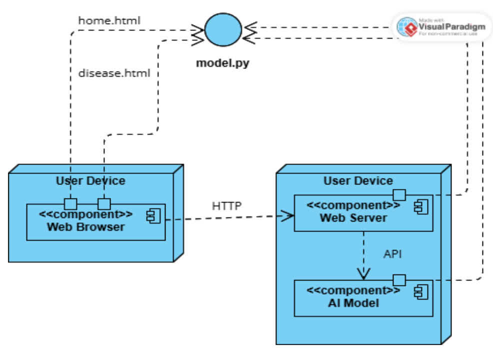

Fig1: Deployment diagram

The system workflow starts by processing collected data and medical information from numerous sources such as medical images along with patient information in EHRs and laboratory test results including blood sugar data, kidney. Data collectors normalize and format the data to establish compatibility with systems using Artificial Intelligence analysis. The process of data preparation involves multiple pre-processing techniques like cleaning, normalization, standardization and augmentation to improve prediction accuracy while using AI models. Data cleaning operates to get rid of errors and fix data gaps and normalize distribution patterns during processing. The combination of image enhancement procedures and the extraction process results in more generalized models while CNNs analyze images and RNNs analyze sequential data to locate vital medical indicators.

The system utilizes medical data analysis through Random Forest and Generative Adversarial Networks (GANs) to process tabulated lab results and blood testing data. The RNN for time-series data is shown in figure2.

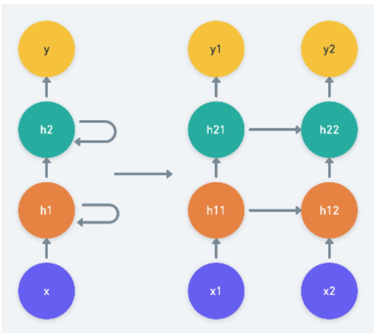

Fig2: Recurrent Neural Networks for time-series data

Model robustness gets enhanced through GANs since they create synthetic medical data which addresses class imbalances in datasets. A multi-label classification system enables simultaneous disease predictions because it overrides the single-disease diagnosis limitation. This produces good results for challenging healthcare situations. The interface was developed with an intention to let medical professionals access information easily and efficiently. Flask operates as the system's backend due to its Python web framework architecture which enables seamless connections of machine learning models with frontend application components thus creating real-time predictions during user interactions.

The Federated Learning patients will benefit from model training systems which operate without the need to transfer their personal information to central server locations thus delivering better privacy and security measures. Personalized disease prediction utilizes AI to generate healthcare and advises by processing individual patient medical information. Patients will benefit from continuous health monitoring through the implementation of IoT and wearable devices. This gives real-time data about their chronic disease conditions. As advancements in federated learning, cloud computing, and explainable AI continue to evolve, the AI-Based Disease Detection System will further solidify its role as a transformative solution in modern healthcare, revolutionizing disease diagnosis and improving patient outcomes.

4. IMPLEMENTATION

CNN stands for Convolutional Neural Network. Deep CNN is an advancement of the Convolutional Neural Network, which helps in processing the images fed to the system. It has been proven effective to use Deep CNN as it considerably uses lesser number of artificial neurons to process the image than the other neural networks. The implementation of disease detection is shown in figure3.

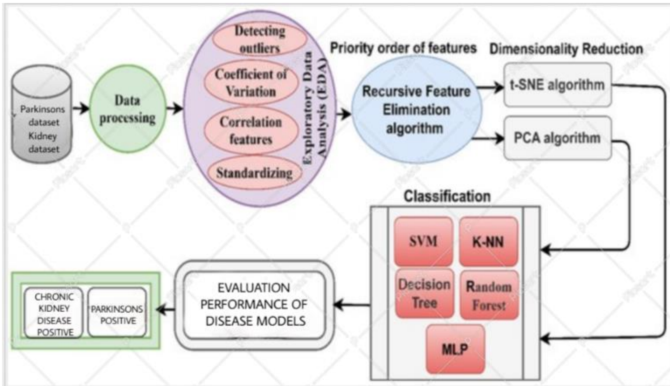

Fig3: Implementation of Disease detection

The distribution of collected data for this work is given in figure 4.

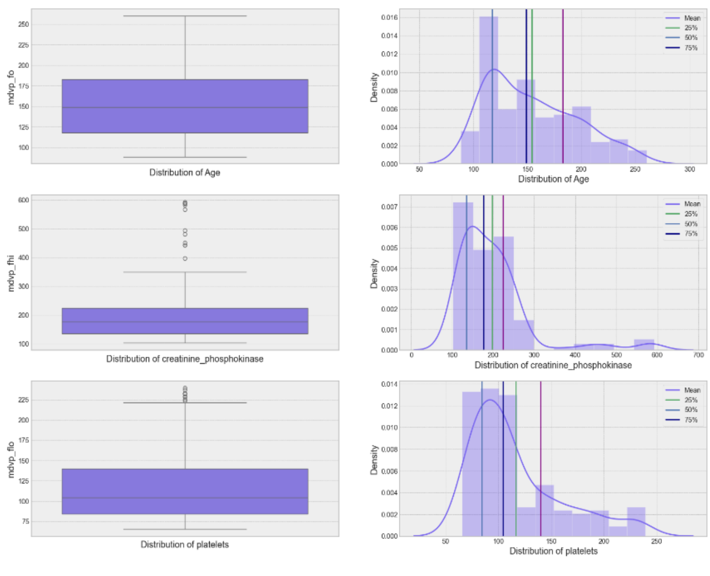

Fig4: Distribution of collected data

The architecture of a convolutional network typically consists of four types of layers: convolution, pooling, activation, and fully connected. The Deep CNN works by training the images. The processing is further carried out step by step and detects the disease associated with the image.

- **Convolutional Layer:** Applies a convolution filter to the image to detect features of the image. The following are the steps that take place in the Convolutional layer of the Deep CNN.
- **ReLU Activation Layer:** The convolution maps are passed through a nonlinear activation layer, such as Rectified Linear Unit (ReLU), which replaces negative numbers of the filtered images with zeros. It helps in extracting meaningful patterns from the given data to the system like the medical images or the patient records.
- **Pooling Layer:** The pooling layers gradually reduce the size of the image by removing unnecessary pixels, thus by keeping only the important information.
- **Fully Connected Layer:** Fully connected layers receive an input vector containing the flattened pixels of the image or the image with the unnecessary pixels removed, which have been filtered, corrected and reduced by convolution and pooling layers.

The Deep CNN architecture for Parkinsons detection is shown in figure5.

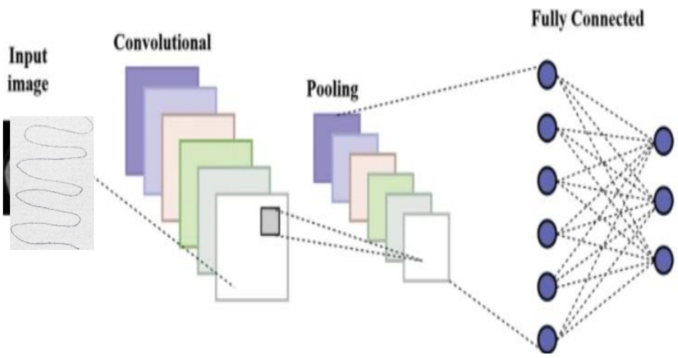

Fig5: Deep CNN architecture for Parkinsons detection Parkinsons

The Random Forest Classifier is defined as an ensemble of multiple decision trees, where each tree is trained on a random subset of the data. The final prediction is obtained by aggregating the predictions from all individual trees, typically through majority voting in classification tasks or averaging in regression tasks. The Random Forest architecture of the work is shown in figure6.

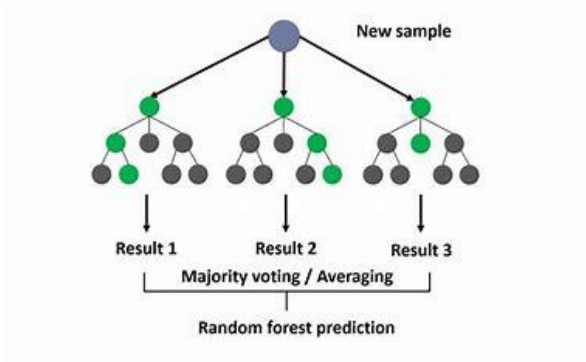

Fig6: Random forest architecture

XGBoost (Extreme Gradient Boosting) is a powerful machine learning algorithm used mainly for classification and regression tasks. It is based on the **gradient boosting** technique, which builds multiple decision trees sequentially, where each tree corrects the errors of the previous one. This iterative learning process helps to improve accuracy and reduce mistakes. XGBoost is widely used because it is fast, efficient, and handles missing data well, making it an excellent choice for medical diagnosis. The XGB of this work in shown in figure7.

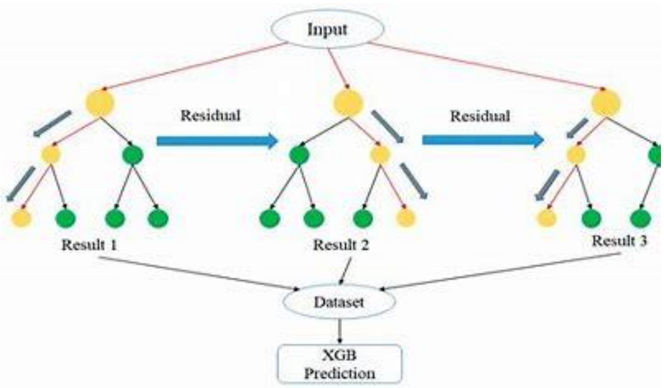

Fig7: XGB architecture

5. RESULTS AND ANALYSIS

The best accuracy of the model for the diseases is achieved by running tests with the inputs over many models like Random Forest, Decision trees, XG Boost, KNN etc. Only one among them gave the best accuracy.

We implemented and compared two CNN-based deep learning approaches for analyzing Parkinsons hand drawings:

- Baseline CNN: A simple convolutional neural network model trained on input features to recognize patterns.
- CNN with Data Augmentation: Used synthetic data generation techniques like random

transformations and feature variations to enhance generalization.

We employed a mixed approach for accurate detection of Parkinsons where voice signals were normalized into time-series data in order to detect tremors and lower frequency values. Along with this time-series data, a CNN model is integrated to detect the onset of Parkinsons by analysing hand drawings of spiral and waves.

From the results, we found that for chronic kidney disease, **RANDOM FOREST CLASSIFIER** gave the better results with the accuracy of **0.991667**. The overall test results for chronic kidney disease using different algorithms are tabulated in table1. The ROC Curve and correlation matrix for Chronic Kidney Disease is given in figures 8&9.

Table -1: Accuracy Scores for Kidney disease

Algorithms used for prediction of Chronic Kidney Disease	
Model	**Score**
Random Forest Classifier	**0.991667**
Gradient Boosting	0.975000
XGBoost	0.966667
Decision Tree Classifier	0.941667
Logistic Regression	0.908333

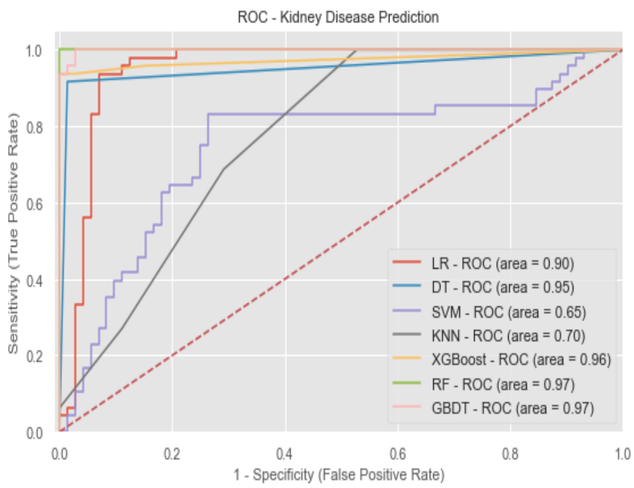

Fig8: ROC Curve for Chronic Kidney Disease

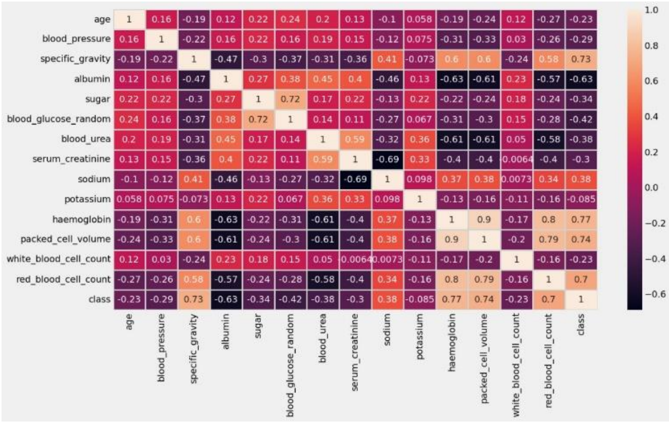

Fig9: ROC Curve for Chronic Kidney Disease

From the results, it is observed that For Parkinson's disease, **XGBoost** gave the better results with the accuracy of **0.983051**. The overall test results for Parkinson's disease using different algorithms are tabulated in table2. The ROC curve and correlation matrix is also shown in figure10&11.

Table -2: Accuracy Scores for Kidney disease

Algorithms used for prediction of Parkinson's Disease	
Model	**Score**
XGBoost	**0.983051**
Random Forest Classifier	0.949153
Bagging Classifier	0.932203
Logistic Regression	0.915254
Support Vector Classifier (SVC)	0.898305

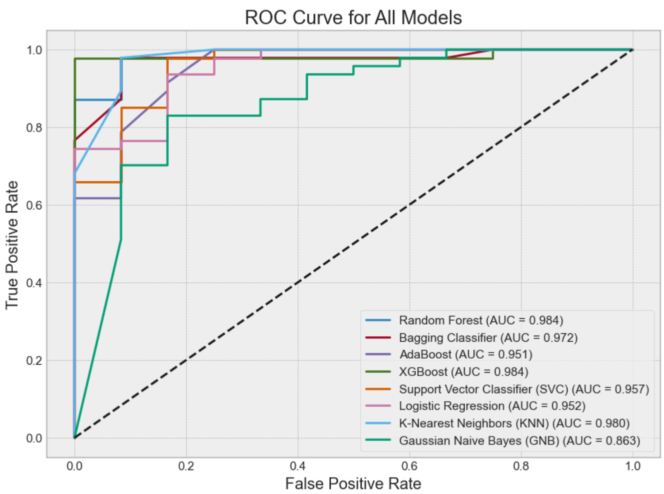

Fig10: ROC Curve for Parkinsons Disease

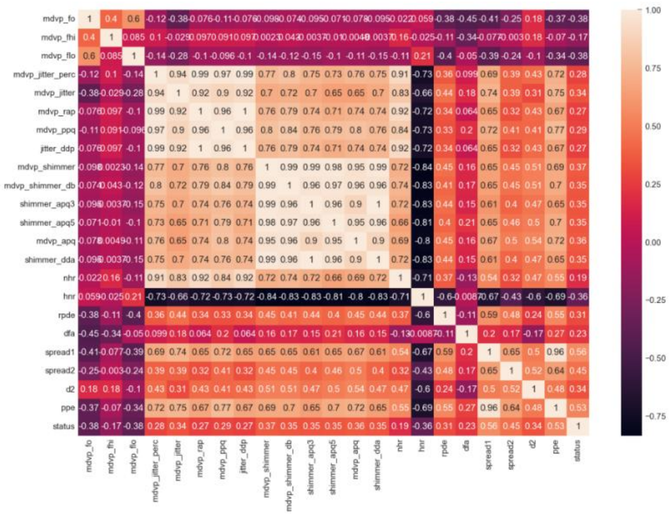

Fig11: Correlation matrix of Parkinsons time-series data

6.CONCLUSION & FUTURE SCOPE OF WORK

The AI-Based Disease Detection System brings vital improvements to medical diagnostics through CNNs. This work unites kidney and Parkinson's diseases under a single integrated platform. This system provides a structured diagnosis framework which unites various medical tests into one process thus simplifying assessment procedures while both shortening evaluation times and enhancing evaluation precision and operational efficiency. The system resolves problems in current diagnostic techniques by employing several extensive datasets combined with advanced deep learning algorithms which leads to trustworthy diagnostic outcomes even when handling multichannel data while addressing user accessibility issues. The web interface system provides user-friendly design for non-technical users and extends healthcare services into areas with minimal healthcare infrastructure.

The system holds promise for improvement through development of a mobile application accessible on Android and iOS platforms which would enable users to run health assessment tests at any time in locations with restricted healthcare services.

The progress in healthcare technology leads to a decrease in healthcare inequalities through increased detection of early diseases together with AI-based medical insights that enhance health outcomes throughout the community.

The future scope of work aims to provide AI-based disease detection using machine learning and convolutional neural networks for other different disease under a single platform.

8.REFERENCES

[1] D. Shen, G. Wu, and H.-I. Suk, "A survey on deep learning in medical image analysis," *Science*, vol. 294, Dec. 2001, pp. 2127-2130.

[2] D. Nie, R. Trullo, C. Petitjean, S. Ruan, and D. Shen, "Medical image synthesis with deep convolutional adversarial networks," *J. Name Stand. Abbrev.*, in press.

[3] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Multi-disease prediction using machine learning," *J. Name Stand. Abbrev.*, in press.

[4] P. Rajpurkar et al., "Title of paper if known," unpublished.

[5] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Machine learning-based Parkinson's disease detection," *J. Name Stand. Abbrev.*, in press.

[6] M. Almasoud and T. E. Ward, "Chronic kidney disease classification using machine learning," *J. Name Stand. Abbrev.*, in press.

[7] Z. Che, S. Purushotham, R. Khemani, and Y. Liu, "Deep learning for time-series disease prediction," *J. Name Stand. Abbrev.*, in press.

[8] S. Wang, Y. Zhang, Z. Dong, and S. Du, "Hybrid deep learning for multi-disease diagnosis," *J. Name Stand. Abbrev.*, in press.

[9] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfer learning in medical image analysis," *J. Name Stand. Abbrev.*, in press.

9.BIOGRAPHIES

Gorla Sukanya, Student, Andhra University College of Engineering for Women

Jakkula Madhurima, Student, Andhra University College of Engineering for Women

Jeeri Alekhya Reddy, Student, Andhra University College of Engineering for Women

Kallakuri Vyshnavi, Student, Andhra University College of Engineering for Women