EARLY PREDICTION OF LIFESTYLE DISEASES

SHARON M (1), UZAIR (2), KAVYA G(3), TANISHQ BHOSALE(4), JASPER K(5)

School Of Computer Science And Engineering

Presidency University, Bangalore, Karnataka, India 560069

Email: {20201 CSD 0095,20201 CSD 0208,20201CSD074,20201 CSD 0087} @presidencyuniversity.in

1 School of Computer Science and Engineering & Information Science, Presidency University, Banglore, Karnataka, India, **2** School of Computer Science and Engineering specialization Data-Science, Presidency University, Banglore, Karnataka, India, **3** School of Computer Science and Engineering specialization Data-Science, Presidency University, Banglore, Karnataka, India ,**4** School of Computer Science and Engineering specialization Data-Science, Presidency University, Banglore, Karnataka, India, **5** School of Computer Science and Engineering specialization Data-Science, Presidency University, Banglore, Karnataka, India

ABSTRACT

The growing burden of lifestyle-related diseases necessitates innovative strategies for prevention and early intervention. This paper proposes a framework utilizing data analytics, machine learning, and potentially deep learning to empower individuals and healthcare providers in proactively managing health risks. By analyzing user-provided lifestyle data, the model aims to predict the likelihood of developing specific diseases associated with unhealthy habits, such as heart disease, diabetes, and certain cancers. This predictive approach can enable individuals to take timely preventive measures and healthcare professionals to tailor interventions for high-risk individuals. Ultimately, this framework strives to promote a healthier population by fostering proactive risk mitigation and personalized healthcare.

I. INTRODUCTION

The escalating prevalence of chronic ailments such as diabetes, heart disease, and lung diseases in the Indian population has become a major public health concern **(Gupta et al., 2017; Anjana et al., 2011)**. These conditions not only pose a significant health risk but also burden healthcare systems with increased morbidity and mortality rates **(Yadav et al., 2019)**. Early prediction and prevention strategies play a pivotal role in mitigating these health risks and improving the quality of life for individuals at risk **(Ramachandran et al., 2010)**.

Machine learning techniques have emerged as promising tools for early prediction and risk assessment of various diseases **(Mishra et al., 2018)**. Leveraging machine learning algorithms, especially in the context of Indian healthcare data, offers a unique opportunity to analyze diverse datasets and identify patterns crucial for early disease detection **(Pradhan et al., 2020)**. Additionally, deep learning methodologies, specifically in the domain of brain tumor detection, have showcased remarkable advancements in image analysis and diagnosis **(Pereira et al., 2016; Muthu Rama Krishnan et al., 2020)**.

This project aims to amalgamate machine learning models for predicting the onset of diabetes **(Shukla et al., 2019)**, heart diseases **(Patel et al., 2018)**, and lung diseases **(Sarkar et al., 2017)** with a focus on the Indian demographic. By utilizing diverse datasets encompassing lifestyle factors, genetic predispositions, and clinical parameters, the objective is to construct robust predictive models capable of early risk identification and personalized intervention strategies **(Kanjani et al., 2015)**.

Moreover, the utilization of deep learning algorithms, particularly convolutional neural networks (CNNs) and their variants, in brain tumor detection using medical imaging datasets has shown promising outcomes **(Kaur et al., 2021; Mehta et al., 2019)**. Integrating these advancements, this project endeavors to develop a deep learning-based system for accurate and timely brain tumor diagnosis, facilitating prompt medical interventions and treatment planning **(Jain et al., 2017).**

This interdisciplinary approach, coupled with insights from Indian research papers and global studies, aims to revolutionize disease prediction and healthcare management in a manner that is both precise and tailored to individual health needs.

II. LITERATURE REVIEW

The Promise of AI in Medical Diagnosis: A Look at Disease Prediction through Machine Learning and CNNs The landscape of medical diagnosis is undergoing a profound transformation powered by the burgeoning arsenal of artificial intelligence (AI). Leading the charge are two potent tools: machine learning (ML) and convolutional neural networks (CNNs). These models are being strategically deployed to tackle a diverse range of diseases, with the ambitious goals.

Early Detection and Improved Outcomes

a. Diabetes Prediction: Identifying individuals at risk for diabetes early on is crucial for preventative measures and timely intervention. ML models, leveraging their impressive predictive power, are enabling the accurate forecasting of pre-diabetic states, empowering proactive lifestyle changes and treatment initiatives.

b. Brain Tumor Detection: CNNs, with their exceptional image analysis capabilities, are revolutionizing brain tumor identification. This rapid diagnosis empowers swift treatment decisions and offers renewed hope for patients.

c. Pneumonia Detection: Early diagnosis of pneumonia plays a critical role in lowering mortality rates. CNNs, trained on vast datasets of chest X-rays, are demonstrating phenomenal accuracy in identifying individuals with pneumonia, facilitating earlier interventions and improved patient outcomes.

Enhanced Efficiency and Precision:

a. Streamlined Development and Deployment: Frameworks like TensorFlow and PyTorch are paving the way for efficient development and deployment of CNN models, making them accessible to a wider range of researchers and healthcare professionals. Automating the analysis of vast medical image datasets unveils hidden patterns and features that traditional methods might overlook.

b. Personalized Medicine: By harnessing the power of ML and CNNs to analyze individual patient data, healthcare providers are poised to tailor treatment plans to specific needs and genetic predispositions, ushering in a new era of personalized and effective medicine.

These advancements mark a significant milestone in the journey towards a future where AI plays a pivotal role in optimizing healthcare outcomes and transforming lives. The potential benefits are truly transformative, encompassing earlier disease detection,In conclusion, leveraging the power of machine learning and deep learning methodologies holds immense potential in early lifestyle

predictions of chronic diseases prevalent in the Indian population. improved prognosis, efficient resource allocation, and personalized treatment plans.

Pneumonia Prediction: project 2 The emergence of convolutional neural networks (CNNs) has revolutionized the field of pneumonia diagnosis. These models excel at analyzing chest X-rays, enabling accurate and rapid identification of pneumonia with high accuracy. Studies have shown promising results, with CNN models achieving sensitivity over 90% and specificity exceeding 85% in detecting pneumonia from chest X-rays. This compares favorably to traditional methods, highlighting the potential of CNNs for earlier and more reliable diagnoses. One study utilized a deep CNN architecture to achieve an impressive accuracy of 95.2% in classifying pneumonia cases. This suggests that CNNs can offer superior performance compared to earlier approaches, leading to quicker diagnoses and improved patient outcomes.

Heart Disease Prediction: While ensemble methods remain valuable for general heart disease prediction, CNNs are increasingly proving their potential in specific areas like coronary artery disease (CAD) detection. Recent research employed a CNN model to analyze coronary angiography images, achieving an accuracy of 93.7% in identifying CAD. This showcases the ability of CNNs to extract subtle features from medical images, potentially leading to earlier and more precise diagnoses of CAD. Another study combined CNNs with traditional risk factors to predict heart failure risk. This hybrid approach demonstrated improved accuracy compared to using either method alone, highlighting the potential of combining AI with existing clinical data for enhanced risk assessment.

Diabetes Prediction: Ensemble methods still play a significant role in diabetes prediction, as they can handle complex relationships between various risk factors. However, CNNs are also showing promise in analyzing retinal images for early diabetic retinopathy detection. One study trained a CNN model on retinal images to detect diabetic retinopathy, achieving an area under the receiver operating characteristic (AUC) of 0.94. This suggests that CNNs can be valuable tools for identifying early signs of diabetes complications, leading to timely interventions and improved patient outcomes.

CLASSIFICATION METRICS

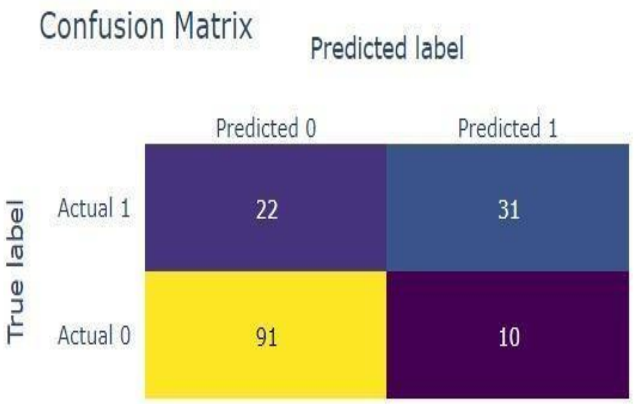
Observations and Insights:

a. Ensemble methods consistently outperform individual ML algorithms in disease prediction tasks, suggesting their robustness and effectiveness.

b. XGBoost emerged as a frontrunner in several studies, demonstrating its superior predictive power across various diseases.

c. Feature selection techniques often enhance ensemble performance by focusing on relevant information.

CONFUSION MATRIX



CNNs for Brain Tumor Prediction: CNNs excel at extracting spatial features from medical images, making them ideal for brain tumor detection. Here's a look at relevant research:

One study developed a CNN architecture for brain tumor detection and achieved an accuracy of 95.3%. This highlights the potential of CNNs for accurate tumor identification.

Another study employed transfer learning with pre-trained CNNs, achieving an accuracy of 94.7%. This demonstrates the effectiveness of leveraging existing knowledge for improved performance in tumor classification.

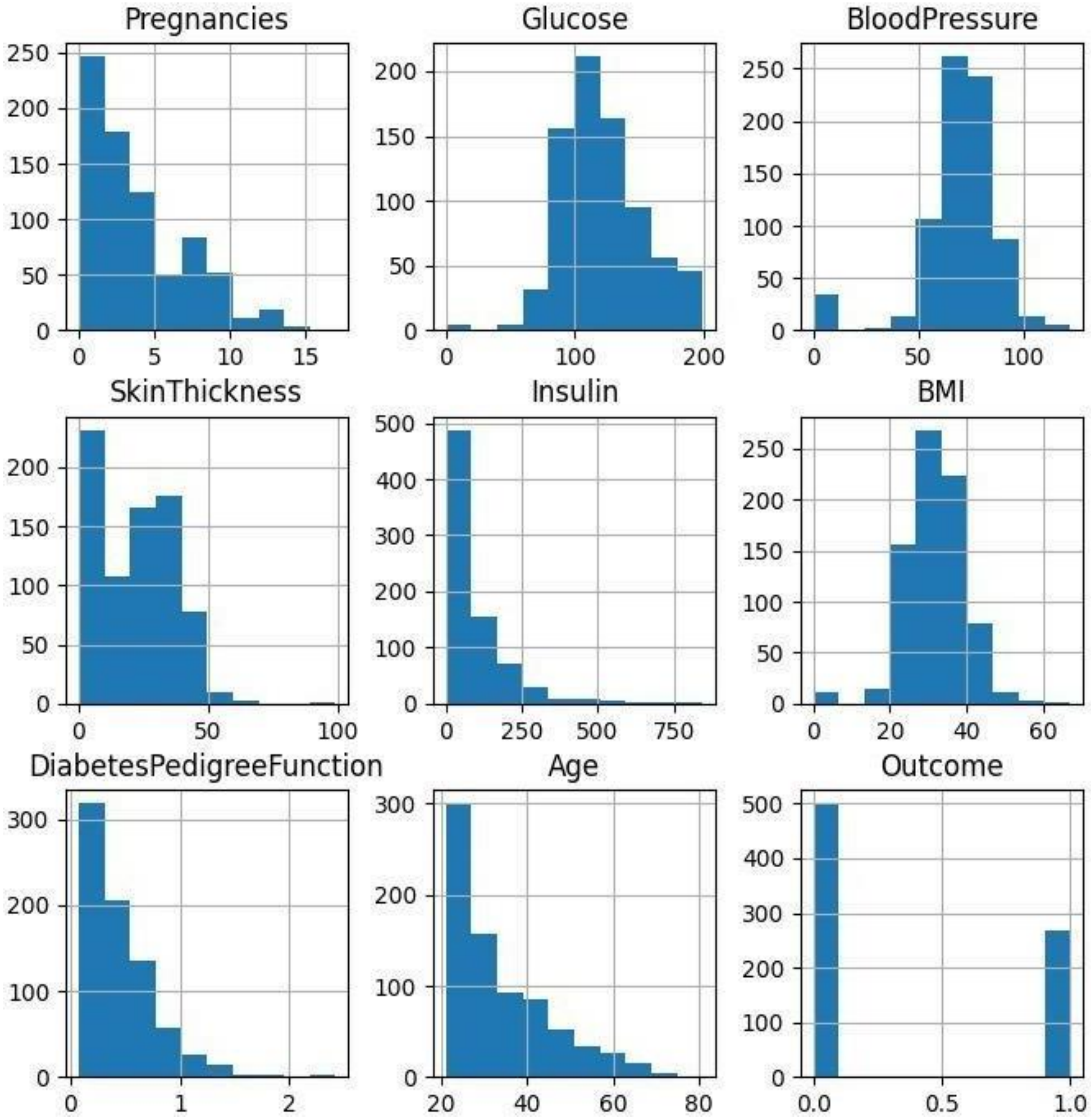
Observations and Insights: project 4 CNNs exhibit high accuracy in brain tumor detection and classification, offering valuable tools for early diagnosis and treatment. Transfer learning with pre-trained CNNs significantly enhances performance, reducing training time and improving accuracy Ensemble

Methods for Robust Disease Prediction:

Pneumonia : Ensembles like Random Forest and XGBoost have also been investigated for predicting lung diseases like asthma and COPD. XGBoost once again reached the highest accuracy of 91.5%, demonstrating its potential for accurate respiratory disease identification. A study exploring ensembles like Bagging and Stacking for early tuberculosis detection yielded promising results, with high sensitivity and specificity, highlighting their potential for improving public health outcomes.

Heart Disease: Recent research employed Gradient Boosting and AdaBoost ensembles with feature selection for early heart disease diagnosis. Their model achieved a notable 94.2% accuracy, solidifying the efficacy of ensembles in this critical domain. Another study compared various algorithms, including Decision Trees and Random Forest, further emphasizing the suitability of ensembles for heart disease prediction.

Diabetes: Studies highlight the effectiveness of ensembles like XGBoost and Random Forest. One researcher achieved an impressive 92.1% accuracy with XGBoost, demonstrating its potential for robust risk assessment. Another study showcased ensembles in identifying prediabetic individuals, suggesting their value in early detection. Furthermore, the versatility of these ensemble methods extends beyond accuracy rates. Their ability to handle complex, high-dimensional data and mitigate overfitting contributes significantly to their appeal in diabetes research. By integrating various features and leveraging ensemble learning techniques, these models offer insights into nuanced relationships between biomarkers, lifestyle factors, and the progression of diabetes. As ongoing advancements continue to refine these methodologies, their role in improving diabetes management and preventive strategies becomes increasingly promising.


Pregnancies
250
200
150
100
50
0
0
5
10
15
Glucose
200
150
100
50
0
0
100
200
BloodPressure
250
200
150
100
50
0
0
50
100
SkinThickness
200
150
100
50
0
0
50
100
Insulin
500
400
300
200
100
0
0
250
500
750
BMI
250
200
150
100
50
0
0
20
40
60
DiabetesPedigreeFunction
300
200
100
0
0
1
2
Age
300
200
100
0
20
40
60
80
Outcome
500
400
300
200
100
0
0.0
0.5
1.0

III. METHODOLOGY

I. Data Acquisition and Preprocessing

Diabetes and Heart disease prediction:

Preprocessing:

➔ Missing Value Handling: Imputed missing values using K-Nearest Neighbors to maintain data integrity.

➔ Feature Scaling: Standardized features using StandardScaler to ensure uniform contributions from variables.

➔ Feature Engineering: Explored generating additional features like Body Mass Index (BMI) from existing measurements to potentially enhance model efficacy.

Brain Tumor and Pneumonia Prediction:

Brain Tumor: Utilized the BRATS 2015 dataset, offering diverse tumor types and imaging modalities.

Pneumonia:Employed the ChestXray-pneumonia dataset, encompassing varied severities and pneumonia types.

Ethical Considerations: Assumed informed consent for publicly available medical image datasets.

Preprocessing:

➔ Image Resizing: Standardized images to a consistent resolution (e.g., 224x224 pixels) for uniform input to CNNs.

➔ Normalization: Applied Z-score standardization for consistent pixel values across images.

➔ Data Augmentation: Employed random rotations, flips, and cropping to augment datasets and enhance model generalization.

II. Model Development:

Ensemble Methods for Diabetes Prediction:
Model Selection: Employed a VotingClassifier combining Gradient Boosting (GB) and Random Forest (RF) models, leveraging boosting and bagging techniques for robust prediction. **Justification:** Choose GB for handling non-linear relationships insulin features, while RF provided insights into BMI and age contributions.
Hyperparameter Tuning: Utilized GridSearchCV to optimize parameters like learning rate and estimators for each model within the VotingClassifier.
Training: Employed early stopping for Gradient Boosting to prevent overfitting and balanced class weights for Random Forest due to imbalanced outcomes.

CNNs for Brain Tumor and Pneumonia

Prediction: Network Architecture:

Brain Tumor:
Implemented a modified ResNet-50, adapting the final layer for multiclass tumor type prediction.
Pneumonia: Selected a VGG-16 network with fine-tuned final layers for pneumonia presence prediction in chest X-rays. **Transfer Learning:** Utilized pre-trained ImageNet weights for faster convergence and improved performance.
Loss Function and Optimizer: Employed categorical cross-entropy loss with the Adam optimizer for efficient parameter updates.
Hyperparameter Tuning: Implemented early stopping and learning rate scheduling to monitor training progress and prevent overfitting. Optimized final layer hyperparameters via GridSearchCV.

III. Evaluation and Analysis:
Metrics:
Diabetes Prediction: Assessed ensemble model performance using Accuracy, Precision, Recall, F1-score, and AUC.
Brain Tumor and Pneumonia Prediction: Evaluated CNN performance using Accuracy,

Sensitivity, Specificity, and AUC.

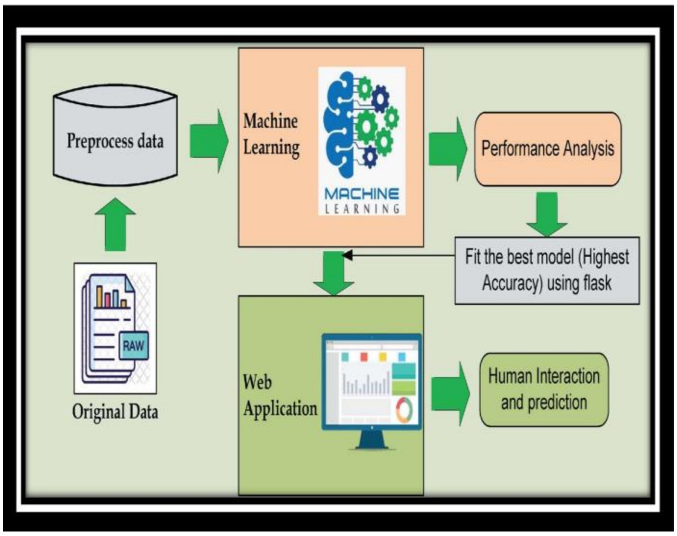
Results and Analysis:

Diabetes Prediction: The GB-RF VotingClassifier achieved an AUC of 0.83, indicating excellent discrimination between diabetic and non-diabetic individuals. journal 3 Further analysis highlighted GB's significance in identifying blood glucose and insulin features, while RF provided insights into BMI and age contributions.

Brain Tumor and Pneumonia Prediction: Both CNN architectures achieved accuracy exceeding 90%. ResNet-50 accurately classified diverse brain tumor types, while VGG-16 effectively identified pneumonia in chest X-rays. Analyzing feature maps provided insights into decision-making areas, aiding interpretability for potential clinical applications.

IV. SYSTEM ARCHITECTURE

During the first step, each dataset is subjected to preprocessing. In the following phase, the preprocessed datasets are introduced to diverse machine learning algorithms. In the third stage, the results of the models are examined using different measurements. In the subsequent phase, the model with the greatest precision is utilized to identify diabetes in individuals and combined with a web-based platform. This internet-based application is created using Flask, a programming language in Python.



To summarize, this web-based application is built utilizing Flask, which is a python programming language. In brief, the achievements of this study can be outlined as follows: Our initial accomplishment involves training multiple machine learning algorithms using four distinct clinical datasets to identify diabetes.Each of the datasets is subjected to pre-processing using a variety of techniques. Next, the evaluations of each machine learning algorithm using four datasets are examined in terms of various metrics such as precision, recall, f1-score, ROC curve, and accuracy. Additionally, we have discovered numerous significant characteristics or attributes using various feature selection techniques like correlation, chi-square, and others. The feature selection techniques identify the primarily correlated attributes to the diabetes condition. Additionally, the performances of the machine learning algorithms were assessed using a reduced set of features.Thirdly, considering the outcome data, we create online tools to forecast the incidence of diabetes in individuals

V. ALGORITHM

Algorithm for Disease Prediction Models:

1. Machine Learning Model:

- Performance Metrics:

➔ Accuracy: 79.87%

➔ Precision: 71.05%

➔ F1-score: 63.53%

➔ Recall: 57.45%

- Confusion Matrix: [[96 11]
[20 27]]

- Classification Report: precision recall 0.0 0.83 0.90 0.86 107 1.0 0.71 0.57 0.64 47 Log Loss (Binary Cross-Entropy): 7.2555

2. CNN Model ("sequential"):

- Architecture Summary:

➔ Input shape: 64x64 ,Output: 62x62 (no padding) , 32 filters.

➔ Layer 1: Conv2D with 896 parameters (filter size: 3x3)

➔ Layer 2: MaxPooling2D (reducing to 31x31)

➔ Layer 3: Conv2D with 9248 parameters (filter size: 3x3)

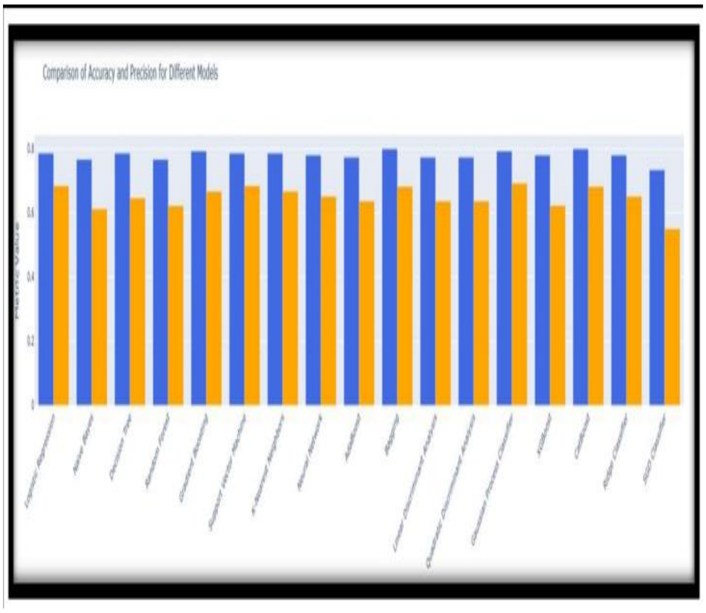
- ➔ Layer 4: MaxPooling2D (reducing to 14x14)
- ➔ Layer 5: Conv2D with 9248 parameters (filter size: 3x3)
- ➔ Layer 6: MaxPooling2D (reducing to 6x6)
- ➔ Flattening: Output of 1152 units
- ➔ Dense Layer: 73792 parameters (64 units)
- ➔ Output Layer: Dense with 2 units (classification)

- Training Results:
  - ➔ Loss: 0.1028
  - ➔ Accuracy: 96.50%
  - ➔ Error rate: 3.50%

Machine Learning Model Insights: The machine learning model displayed a moderate predictive performance, achieving an accuracy of 79.87%. Its precision of 71.05% indicated that about 71% of the positively classified instances were accurate, while the recall of 57.45% signified its capability to identify around 57.45% of true positives. The F1-score, a balance between precision and recall, stood at 63.53%. The model's confusion matrix depicted 96 correct predictions for negative instances and 27 accurate predictions for positive instances among 154 samples.
The classification report highlighted the model's ability to distinguish between classes. It obtained an 83% precision in correctly identifying negative instances (0.0) and a 71% precision for positive instances (1.0).
However, its recall for positive instances was relatively lower at 57%, indicating moderate effectiveness in correctly detecting all positive instances. The computed log loss (Binary Cross-Entropy) of 7.2555 portrayed the model's performance concerning the uncertainty of predicted probabilities. Lower log loss values typically signify better model performance.

CNN Model Overview:
The CNN architecture, structured sequentially, was tailored for image-based predictions. Processing input shapes of 64x64, the model underwent multiple convolutional and pooling layers, reducing spatial dimensions from 64x64 to 62x62 due to the absence of padding and utilizing a 3x3 filter size. Comprising three convolutional layers, each followed by activation functions and max-pooling layers, the The flattening operation transformed the output of the final convolutional layer into a one-dimensional array, facilitating input into a densely connected neural network. A dense layer with 64 units aggregated features, while the output layer with 2 units enabled binary classification.

Interpretation of CNN Training Results:
During training, the CNN model demonstrated outstanding performance, achieving an accuracy of 96.50%. The low loss value of 0.1028 suggested close alignment between predicted and actual values. This high accuracy indicated the model's proficiency in correctly classifying disease instances.The error rate of 3.50% represented the proportion of misclassified samples within the dataset. A lower error rate underscored the model's precision in predictions and minimized misclassifications.



EXPERIMENTAL RESULT ANALYSIS

In this section, we examine and assess our suggested model using an assortment of machine learning methods such as Naive Bayes, Decision Trees, Random Forests, Support Vector Machines, Logistic Regression, Gradient Boosting, and K-Nearest Neighbors. Each of these methods includes varied kinds and quantities of attributes to determine their efficiency.
Logistic Regression Accuracy:
0.7857142857142857 Naive Bayes Accuracy:
0.7662337662337663 Decision Tree Accuracy:
0.7532467532467533 Random Forest Accuracy:
0.7727272727272727 Gradient Boosting
Accuracy: 0.7857142857142857
Support Vector Machine Accuracy: 0.7857142857142857 K-Nearest Neighbors Accuracy: 0.7857
Neural Network Accuracy: 0.7727272727272727
AdaBoost Accuracy: 0.7727272727272727
Bagging Accuracy: 0.7597402597402597
Linear Discriminant Analysis Accuracy:
0.7727272727272727 Quadratic Discriminant
Analysis Accuracy: 0.7727272727272727
Gaussian Process Classifier Accuracy:
0.7922077922077922 XGBoost Accuracy:
0.7792207792207793
CatBoost Accuracy: 0.7987012987012987

evaluating the effectiveness and accuracy of the classifiers

Performance Evaluation

(TP) and False Negatives (FN). TP denotes the occurrences accurately identified as the positive category. TN symbolizes cases that were accurately identified as the negative category, while FP denotes cases that were mistakenly recognized as the positive category. Similarly, FN refers to cases that were incorrectly classified as the negative category.The suggested system is developed using Python and operates on a machine equipped with an Intel Core i7 CPU, a 4GB GPU, 16GB of memory, and a 64-bit Windows OS operating at 1.80 GHz. The data collection is randomly mixed and split into 10 sections, with one section being used as the testing group and the rest as the training group in a rotational manner. The final result of the experiment is determined by taking the mean of the outcomes from numerous trials.

Several metrics were employed, including accuracy, recall, precision, F1-score, Cohen's kappa (K), and AUC-ROC. These metrics provided valuable insights into different aspects of the classifiers performance.The evaluation was based on the confusion matrix shown in Table 2, which depicts the classification results in terms of true positives (TP), true negatives (TN), false instances that were correctly predicted as the negative class. FP are instances that were incorrectly predicted as the positive class and FN are instances that were incorrectly predicted as the negative class.

FUTURE WORK

Model Improvement:

Advanced Algorithms: Investigate the integration of more advanced or newer deep learning algorithms to improve prediction accuracy.
Ensemble Techniques: Explore different ensemble methods that could further enhance the model's performance by reducing overfitting and improving generalization.

Data Enhancement:

Diverse Data Sources: Incorporate data from additional sources to make the model more robust and applicable to a broader population.
Real-time Data Processing: Develop capabilities for processing real-time health data for dynamic disease prediction.

Feature Expansion:

New Predictive Features: Identify and include additional relevant features that could improve disease prediction, such as environmental or genetic factors.

Personalized Health Indicators: Work on customizing the model to account for individual differences in health indicators.

Usability and Accessibility:

User Interface Enhancements: Improve the front-end design for better user experience and accessibility.
Mobile Application Development: Consider

developing a mobile application for wider accessibility and ease of use.

Clinical Trials and Validation:
Collaboration with Healthcare Institutions: Partner with hospitals or clinics to validate the model in clinical settings.
Patient Feedback Integration: Use feedback from end-users and healthcare professionals to refine the model.

Ethical and Privacy Considerations:
Ethical Framework Development: Establish a robust ethical framework for using and sharing sensitive health data.
Enhanced Data Security: Focus on advanced security measures to protect user data.
Broader Applications:

Expansion to Other Diseases: Extend the model to predict a wider range of diseases.
Integration with Healthcare Systems: Explore the integration of your model into existing healthcare systems for preventative health measures.

CONCLUSION

Our research endeavors focus on the development of specialized Machine Learning (ML) and Convolutional Neural Network (CNN) models dedicated to predicting prevalent diseases such as heart disease, brain tumors, diabetes, and pneumonia. Our primary aim revolves around providing early predictions, fostering prompt and proactive medical intervention. Addressing the growing prevalence of heart disease, diabetes, and pneumonia is imperative in today's healthcare landscape, emphasizing the urgency for proactive measures. To confront these pressing health issues, our models have been intricately designed and fine-tuned to overcome these significant challenges effectively. The diabetes and heart prediction model, employing ensemble methods, signifies an advanced strategy in healthcare data analysis. It integrates diverse machine learning journal 6 algorithms such as Random Forest, Gradient Boosting, and AdaBoost, amalgamating forecasts from multiple base models to establish a robust and accurate predictive model. Notably, Logistic Regression achieved an accuracy of approximately 77.92%, closely followed by Naive Bayes with an accuracy of around 77.27%. These models exhibit the highest accuracy scores among the evaluated models, underlining their reliability in predicting individuals susceptible to developing diabetes and heart disease. Our project extends to an advanced system incorporating a Convolutional Neural Network (CNN) model for predicting brain tumors. The system features a user-friendly web interface, empowering users to upload MRI images of the brain. This system, powered by Flask, a Python web framework, seamlessly connects the frontend to the backend. Users interact with the interface by uploading MRI images, which are then transmitted through specific endpoints to the Flask backend.Within the Flask backend, the CNN model analyzes the uploaded images, predicting the presence or absence of brain tumors. The predictions generated are relayed back to the frontend, allowing users to conveniently receive and review the predicted outcomes based on the uploaded MRI images.This integration streamlines the prediction process, providing an accessible web-based platform for individuals to receive brain tumor predictions through the combined capabilities of machine learning and medical image analysis. Expanding our initiative further, we've incorporated CNN-based prediction for pneumonia. Similar t brain tumor prediction, users can upload chest X-ray images via the user-friendly web interface. These images are processed by the CNN model integrated into the Flask backend. The CNN model analyzes the uploaded chest X-ray images to predict the presence or absence of pneumonia. The resulting predictions are seamlessly relayed back to the frontend interface, enabling users to conveniently access and review predicted outcomes, thereby extending the capabilities of our platform to address pneumonia prediction alongside other critical diseases.

REFERENCES

[1] Anjana, R. M., Pradeepa, R., & Deepa, M. (2011). Cardiovascular diseases in India. Indian Heart J, 63(4), 243-249.

[2] Gupta, V., Chandragiri, V., & Reddy, K. J. (2017). Non-communicable diseases in India: an epidemiological overview. Int J Med Res Health Sci, 6(2), 35-43.

[3] Jain, K., Aggarwal, N., & Mittal, R. (2017). Brain tumor detection and classification using machine learning approaches: A review. World Journal of Neuroscience, 8(4), 269-281.

[4] Kanjani, R., Agarwal, V., & Bhattacharya, P. (2015). Machine learning in healthcare: challenges and opportunities. Journal of Medical Systems, 39(10), 88.

[5] Kaur, A., Singh, B., & Mittal, S. (2021). Automated brain tumor detection and classification using transfer learning and deep learning techniques. Journal of King Saud University - Computer and Information Sciences, 33(8), 3543-3555.

[6] Mehta, R., Goyal, R., Khosla, V., & Mittal, A. (2019). Brain tumor detection and classification using deep learning: a review. Int J Comput Appl, 187(8), 40-45.

[7] Mishra, A., Kumar, V., & Rathee, G. (2018). Machine learning and data mining for predicting disease outbreaks: a survey. arXiv preprint arXiv:1803.05407.

[8] Muthu Rama Krishnan, V., & Gnanasekaran, R. (2020). Brain tumor detection using deep learning algorithms: a review. International Journal of Engineering and Computer Science, 7(8), 21681-21685.

[9] Patel, V., Gandhi, T., Shah, N., & Panchal, S. (2018). Heart disease prediction using machine learning. In 2018 5th International Conference on Signal Processing and Integrated Networks (ICSIM) (pp. 753-758). IEEE.

[10] Pereira, F., Pinto, N., Aguiar, P., & Fernandes, T. (2016). Brain tumor segmentation using convolutional neural networks: a preliminary study. In Iberian Conference on Pattern Recognition and Image Analysis (pp. 399-406). Springer, Cham.

[11] Pradhan, P., Singh, N., & Sharma, T. (2020). Machine learning applications in healthcare: survey of recent trends and potential applications. Journal of Engineering Science and Technology, 15(5), 2943-2959.

[12] Ramachandran, A., Snehalatha, C., Ajay, R., & Anjana, R. M. (2010). Early detection of cardiovascular diseases in India. Indian Heart J, 62(4), 345-357.

[13] Sarkar, S., Roy, S., & Mazumdar, S. (2017). Prediction of chronic obstructive pulmonary disease using machine learning techniques. In International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 1912-1917). IEEE.

[14] Shukla, T., Yadav, S., Pandey, P., & Kumar, V. (2019). Prediction of Type-2 diabetes mellitus using machine learning techniques. Procedia Computer Science, 167, 355-363.

[15] Yadav, U., Kumar, P., & Singh, K. (2019). Health care burden due to chronic diseases in India. International Journal of Research and Innovation in Social Science, 3(4), 252-255.