GMR
The Original

MULTI DISEASE DETECTION USING DEEP NEURAL NETWORK

Shradha Prakash Patil[1], Smita S. Sangewar[2], Pratiksha M. Jamdade[3]

[1]P.G. Student, Department of CSE, DKTE Society's Textile and Engineering Institute, Ichalkaranji, Maharashtra, 416115, India.
[1]Assistant Professor, CSE Department, Ashoakrao Mane Group of Institutions, Vathar, Maharashtra 416112, India.
[2]Assistant Professor & P.G. Coordinator, Department of CSE, DKTE Society's Textile and Engineering Institute, Ichalkaranji, Maharashtra 416115, India.
[3]P.G. Student, Department of CSE, DKTE Society's Textile and Engineering Institute, Ichalkaranji, Maharashtra 416115, India.
[3]Assistant Professor, AIDS Department, Padmabhooshan Vasantraodada Patil, Institute of Technology (PVPIT), Budhgaon, Sangli, 416416, India .

ABSTRACT
This study introduces a comprehensive intelligent medical prediction framework, which combines Structured clinical data and deep convo disease-specific machine learning models The aim of this study is to create lutional neural networks to detect pneumonia in chest X-rays. The framework keeps data diversity and strengthens the predictability in multiple medical fields including chronic kidney disease, heart disease, diabetes, and so on. Disease, liver disorders, Parkinson's, diabetes and pneumonia. The experimental assessment showed a good predictive ability for most of the disease type's categories. With the classification of chronic kidney disease, there was a near-perfect discrimination of which the area under the curve was 0.96. Good ROC-AUC values for all classifiers (1.00), indicating high separability of clinical biomarkers. Good results also obtained with ensemble and neural network techniques in the detection of Parkinson's disease. Nonlinear models showed the successfulness which gave a ROC-AUC of more than 0.97. Acoustic feature learning. Ensemble fusion was shown to aid in prediction of heart disease, and to have superior classification/stability and calibration performance compared to tree-based methods while retaining these benefits models. The neural mod works very well on biochemical interactions which are complex, namely detecting liver disorderelling and ensemble aggregation. The prediction of diabetes was still limitedly complicated due to the overlapping characteristics feature in the metabolism and that led to moderate performance in the classification, indicating the need for more complex feature representations. Medical image analysis was achieved using a deep learning model based on ResNet50V2. Chest x-ray diagnosis of pneumonia. The model shows high recall, f1-score and correctness (above 95%) and VGG19 is performing well, although not as well. Minimal overfitting, and convergence. A confusion matrix analysis also was performed and showed excellent.

KEYWORDS: Ensemble Learning, clinical decision support system, multimodal healthcare, disease prediction, chronic kidney disease, heart disease, liver disease, Parkinson's disease, diabetes prediction, pneumonia detection, Chest X-ray classification, convolutional neural network (CNN)

1. INTRODUCTION

Artificial Intelligence (AI) and Machine Learning (ML) have had a massive impact in the present times in the healthcare sector with intelligent and data-driven solutions. Medical decision-making. There is a lot of information available for healthcare organizations. Using any electronic health records, lab results, wearable devices, diagnostic imaging, etc. There are monitoring systems as well as clinical monitoring systems. This type of highly dimensional and heterogeneous data is analyzed. Manually processing medical data is a time-consuming and error-prone process. Thus, there is growing demand for automatic diagnostic systems to identify the meaningful patterns from medical data and provide timely and correct diagnosis for the prediction of the disease. The traditional medical diagnosis is primarily on the skill of doctors and the disease. Specific testing procedures. While the above mentioned strategies continue to be important, they can be challenging to implement scalability, diagnostic consistency and early detection of chronic illnesses. The computational diagnostic systems are primarily employed to predict/explore existing systems. A single disease (diabetes, heart disease, kidney disease, liver disease etc.) disorders. However, in real clinical life, patients may be faced with several medical conditions simultaneously or several medical conditions are called multiple co-existing medical conditions. The several diseases may have co-occurrence in the same host, which makes diagnosis more complicated due to symptoms that are present in the host with both diseases, there are potential overlaps between biomarkers, and physiological abnormalities. This makes it challenging for health-care workers to identify all diseases properly and efficiently. Machine Learning techniques can provide an answer to these issues by learning. Understand and simulate complex nonlinear relationships with clinical data from the past. The supervised learning algorithms that have proven to have good disease prediction skills include Decision Trees, Random Forests, Support Vector Machines, Logistic Regression, Gradient Boosting methods, and Artificial Neural Networks. Besides, the approached ensemble learning methods can be used to increase the predictive stability and generalization of the output of a number of classifiers. Multiple specialized models. Deep learning algorithms, specifically Convolutional Neurons Networks (CNNs) have also picked up tremendous successes in medical

image analysis, the image processing includes detection of pneumonia and other thoracic diseases based on chest X-ray images. A common framework for the structured clinical data and medical imaging that are collected. Along with the fact that it uses ligent framework, it's a required step towards a well-integrated healthcare prediction systems. They are applicable in hospitals, telemedicine system and public applications. Facilitate early diagnosis, risk assessment and personalized treatment planning through mentoring health organizations. Also, intelligent disease prediction systems can decrease health care. Addresses costs, reduces diagnostic delays and enhances access to health services, particularly in low resource areas. Hence, the establishment of a solid multi-country dataset on the city level is crucial. The research on multiple-disease prediction model based on the modal approach has become significant. In today's health care information technology.

Key Contributions

Put forward an integrated AI health prediction system that could predict multiple diseases from clinical and medical imaging information. Improved accuracy in predicting disease by creating special Machine Learning models for diabetes, heart disease, kidney disease, liver disease and Parkinson's disease. Adopted an ensemble learning approach to get more reliable, robust and overall system performance by combining the predictions of different models. Designed a model for the detection of pneumonia based on deep learning using chest x-ray images that has highly accurate and reliable prediction results. Developed a multimodal scalable clinical decision support system with structured medical information and image based analysis, to provide intelligent healthcare diagnosis.

2. RELATED WORK

Various researchers have explored Machine Learning (ML) and Deep Learning (DL) techniques to forecast disease and for diagnosis. Traditional Machine Learning models such as Random Forest, Support Vector Machine (SVM), Logistic Regression and XGBoost have been successfully applied to traditional structured clinical datasets to predict disease conditions like diabetes, heart disease, kidney disease and liver disorders. The derived techniques achieved acceptable levels of prediction error and emphasized the need to have automated healthcare analytics. The combination of multiple classifiers to obtain more stable and robust predictions is another ensemble learning methods that have been the subject of recent studies. Ensemble based systems were demonstrated to be more accurate, more recall and more generalizing than individual systems in a heterogenous medical data set and imbalanced classes. Deep Learning (DL) models, specifically Convolutional Neural Networks (CNNs) have exhibited outstanding results in disease detection using radiological images in medical imaging applications. The detection of pneumonia has been successfully carried out using deep learning architectures such as ResNet, VGG and DenseNet, achieving a high classification accuracy and a low false negative rate, in chest X-ray images. In the context of multimodal healthcare systems, structured clinical information and medical imaging have also been explored as a means to enable exhaustive disease diagnosis. But, a lot of the current systems only make predictions for a single disease or utilise only one kind of health care information. This kind of constraint requires the need to find a common intelligent system to treat various diseases and disparate data in the same clinical decision support system.

3. RESEARCH GAP AND PROBLEM STATEMENT

3.1 Research Gap

Most current prediction systems for healthcare are only able to predict one disease at a time and are unable to deal with multiple diseases in one framework. Medical imaging information and structured clinical information have been used in most studies, and a few studies have integrated them into an intelligent diagnostic system. Healthcare data is often complex and diverse, with a need to capture complex and nonlinear relationships. Healthcare data is complex and diverse and sometimes requires capturing complex and nonlinear relationships, which traditional Machine Learning models may not be able to handle. On the other hand, the Deep Learning models can achieve a high accuracy in medical imaging tasks, but the resources required are relatively high, and the models cannot be used for clinical interpretation. The available literature also has many limitations on handling diverse data sets, missing data, class imbalance and the distribution of features over disease classes. Most of those systems also only consider the accuracy of prediction without considering robustness, reliability of the calibration, and the system's performance in other medical fields. Therefore, a multimodal healthcare framework with scalable and efficient performance is needed that uses various machine learning, deep learning and ensemble learning techniques to provide accurate, reliable and robust prediction of multiple diseases, without the need to develop multiple clinical decision support systems. Moreover, a number of the currently available hybrid systems which integrate encryption and steganography cannot find a good balance between the strength of security, carrying capacity and image quality and efficiency.

3.2 Problem Statement

Chronic and lifestyle-related diseases, such as diabetes, heart disease, chronic kidney disease, liver diseases, Parkinson's disease and pneumonia are commonplace and are a major concern for healthcare systems today. Conventional diagnostics usually target a single disease, and the process is lengthy, complicated and not very efficient for each patient who suffers from multiple diseases. Current prediction models are mainly built separately for distinct diseases, and lack a common framework that can be used to process and predict heterogeneous clinical and imaging data. In addition, the existing diagnostic systems are not reliable and scalable because of issues of class imbalance, missing values, feature distribution and lack of model generalizability. Thus, an intelligent and unified multi-disease prediction framework is needed which combines the combination of Machine Learning, Deep Learning and ensemble learning techniques to offer accurate,

robust, and reliable clinical decision assistance that leverages structured medical information and imaging (radiographic) data.

4. PROPOSED METHODOLOGY

The proposed system adopts a single system for the case of healthcare prediction, while the techniques of machine learning (ML), deep learning (DL) and ensemble learning (EL) are incorporated for multi-disease diagnosis. The methodology starts with the gathering of publicly available medical data of the ailments including diabetes, cardiovascular ailment, chronic kidney ailment, liver disorders, Parkinson's sickness and pneumonia. Multimodal analysis of healthcare is helped by structured clinical data and image data of chest X-rays. In order to improve the quality of data, the missing values, noisy records and duplicate data are taken care of during the pre-processing stage. The application of feature engineering techniques like normalization, encoding, feature selection, and class balancing aims to boost predictive accuracy and minimize the dataset's bias. Processed data is then split into training and testing data sets to develop the model. In the case of structured clinical data, Machine Learning algorithms such as Random Forest, XGBoost, Logistic Regression, Support Vector Machine and Artificial Neural Networks are put in place to forecast diseases. For the pneumonia detection task, the Deep Learning models use Convolutional Neural Networks (CNNs) to analyze the chest X-ray images and extract important visual attributes. Multiple classifiers' outputs are combined with ensemble learning, and the outputs from multiple classifiers are fused based on the decision fusion technique, in order to improve the accuracy and robustness of prediction. The models obtained are evaluated with performance parameters such as accuracy, precision, recall, f1 score, roc-auc score and confusion matrix analysis. Last but not least, the top-performing models are combined into a single clinical decision support system with the ability to make accurate and reliable multi-disease predictions for intelligent healthcare systems.

5. Mathematical Model and Algorithm

Let the complete healthcare dataset be represented as:

$$D=\{(xi,yi)\}i=1N \quad (1)$$

Where:

- xi=[x1,x2,x3,...,xn]x_i = [x_1, x_2, x_3, ..., x_n]xi=[x1,x2,x3,...,xn] represents the feature vector containing patient medical attributes.
- yiy_iyi represents the disease class label.
- NNN represents the total number of samples.

1. Data Preprocessing

The normalized feature value is calculated using Min-Max Normalization:

$$x'=xmax−xminx−xmin \quad (2)$$

Where:

- xxx = original feature value
- xminx_{min}xmin = minimum feature value
- xmaxx_{max}xmax = maximum feature value

2. Machine Learning Prediction Model

For disease prediction, classifiers learn a mapping function:

$$f(x)=y \quad (3)$$

Where:

- f(x)f(x)f(x) = prediction function
- xxx = input medical features
- yyy = predicted disease output

3. Ensemble Prediction

The final ensemble prediction is obtained using majority voting:

$$Y=mode(y1,y2,y3,...,yn) \quad (4)$$

Where:

- y1,y2,...,yny_1, y_2, ..., y_ny1,y2,...,yn are outputs from individual classifiers.
- YYY is the final predicted disease.

4. CNN-Based Pneumonia Detection

For chest X-ray image classification:

$$O = CNN(I) \quad (5)$$

Where:

- III = input chest X-ray image
- CNNCNNCNN = convolutional neural network model
- OOO = predicted output class

6.Dataset

The proposed research is based on the use of multiple benchmark healthcare datasets from publicly available repositories like UCI machine learning repository and Kaggle. The datasets contain both structured clinical records and medical imaging data, allowing the establishment of a multi-modal intelligent healthcare prediction framework. Such datasets can be applied to numerous research projects in healthcare analytics and are appropriate for assessing Machine Learning and Deep Learning models for disease diagnosis applications.

Dataset Source:

In this study, the datasets are retrieved from trusted and public data repositories like UCI Machine Learning Repository and Kaggle. These repositories offer benchmark medical data which is commonly used in research of Machine Learning and Deep Learning in the medical field for prediction and diagnosis of diseases.

Dataset Domain:

Data sets are from the medical diagnosis and the healthcare industry. They include information on clinical, physiological, biochemical, voice and/or radiographic data on diseases such as diabetes, heart disease, chronic kidney disease, liver diseases, Parkinson's disease and pneumonia. The datasets are used to make intelligent decisions in clinical settings and automatic prediction of disease.

Dataset Composition:

The dataset is comprised of two categories of medical datasets: Imaging dataset and Structured tabular medical dataset. The structured dataset consists of patient clinical data, laboratory test results, and biomedical measurements, while the imaging dataset is chest X-ray radiographs for detecting pneumonia. The multimodal dataset structure enables to build a single healthcare prediction model on heterogeneous healthcare data.

Chronic Kidney Disease (CKD) Data Set

The clinical and laboratory information for the evaluation of kidney function is included in the CKD dataset. Common features are blood pressure, serum creatinine, blood urea, Hb, Albumin, Na, K and glucose. The information should be divided into two categories: chronic disease of the kidney, or normal kidney. The data has high biomarker separability, which enables highly accurate predictive modelling.

Heart Disease Dataset

The heart disease dataset contains information about risk factors for heart disease collected from a medical exam of heart disease patients. Age, cholesterol level, resting blood pressure, kind of chest pain, fasting blood sugar, ECG measures, maximum heart rate and exercise-induced angina. The dataset is employed for the prediction of cardiovascular disease risk and evaluation of effectiveness of ensemble learning technologies in clinical decision support.

Liver Disease Dataset

The biochemical test parameters that are associated with liver dysfunction are included in the liver disorder dataset. These include total bilirubin, direct bilirubin, alkaline phosphatase, alanine aminotransferase (ALT), aspartate aminotransferase (AST), albumin and total protein. Liver Disease Classification: The dataset is utilized for classification of liver disease, while Machine Learning models analyze complex biochemical interactions.

Parkinson’s disease Dataset

The Parkinson's data set is a set of bio-medical voice measurements collected from a set of Parkinson's patients. They include jitter, shimmer, harmonic-to-noise ratio (HNR) and recurrence period density entropy (RPDE) measures and pitch perturbation. The data set is used in the detection of Parkinson's disease with the use of nonlinear acoustic feature analysis and neural learning methods.

Diabetes Dataset

Metabolic and physiological health parameters were included in the diabetes dataset for diabetes prediction. Main variables are glucose concentration, body mass index (BMI), insulin level, blood pressure, skin thickness, age and number of pregnancies. The data have also similar metabolic characteristics, making the prediction of the disease somewhat difficult and a good test of the model's generalization abilities across the data.

Pneumonia Chest X-ray Dataset

Pneumonia data set contains chest radiographic images which are divided into pneumonia and normal classes. The data set is used for deep learning medical image classification using Convolutional Neural Networks (CNNs) using ResNet50V2 and VGG19 architecture. The data set can be used to automatically detect pneumonia in the radiographic images with high diagnostic accuracy and low false negative rate

Dataset Characteristics

- Structured tabular data and radiographic image data combined.
- Works with binary disease classification problems.

• Has a mix of medical information from different health areas.
• Needs to be preprocessed, including normalization, feature encoding, missing value treatment, and class balancing.
For training, validation and evaluation of Machine Learning and Deep Learning models.

Table 1. Detailed Dataset Overview

Parameter	Details
Dataset Source	UCI Machine Learning Repository and Kaggle
Total Datasets Used	6
Diseases Covered	Chronic Kidney Disease, Heart Disease, Liver Disease, Parkinson's Disease, Diabetes, Pneumonia
Dataset Type	Structured Clinical Data and Medical Imaging Data
Data Format	CSV files and Chest X-ray Images
Classification Type	Binary Classification
Input Features	Clinical biomarkers, physiological parameters, voice measurements, radiographic images
Preprocessing Techniques	Missing value handling, normalization, encoding, feature selection, class balancing
Machine Learning Models	Random Forest, XGBoost, Logistic Regression, SVM, ANN
Deep Learning Models	CNN, ResNet50V2, VGG19
Evaluation Metrics	Accuracy, Precision, Recall, F1-score, ROC-AUC
Application Area	Intelligent Healthcare and Clinical Decision Support System

7. Experimental Model and Metrics of Evaluation.

7.1 Experimental Setup

The proposed system has several Machine learning model and Deep learning model for Disease prediction and medical image analysis. For structured clinical datasets, classification algorithms such as Random Forest, XGBoost, Logistic Regression, Support Vector Machine (SVM) and Artificial Neural Networks (ANN) are used with the task of disease-specific classification. Pre-processed healthcare data sets that include clinical and biochemical attributes of chronic kidney disease, heart diseases, liver disorders, Parkinson's disease and diabetes are used in training these models. Deep Learning models using Convolutional Neural Networks (CNNs) are used for detecting pneumonia by interpreting chest X-ray images. For feature extraction and image classification ResNet50V2 and VGG19 architectures are followed. Moreover, the predictions of each of the classifiers are combined with other methods of ensemble learning to improve the prediction accuracy, robustness and generalization on the heterogeneous healthcare datasets.

7.2 Evaluation Details

A thorough exploratory data analysis (EDA) of the multi-disease data sets is explored to assess data quality, relationships between features, and relationships between features and disease outcome. Correlation heatmaps are used to identify clinically relevant feature clusters and linear relationships to assist in model selection and ensemble approaches.

1.1 Chronic Kidney Disease (CKD) Dataset

The CKD dataset contains 400 patient records, each having 26 clinical attributes, including numerical and categorical. High rates of missing data were found (22% for potassium, 21.75% for sodium, 13% for hemoglobin, and urinary markers), due to the discrepancies found in clinical practice, requiring powerful imputation methods. The class distribution shows moderate imbalance between the number of CKD positive and non-CKD cases (62.5% and 37.5% respectively).

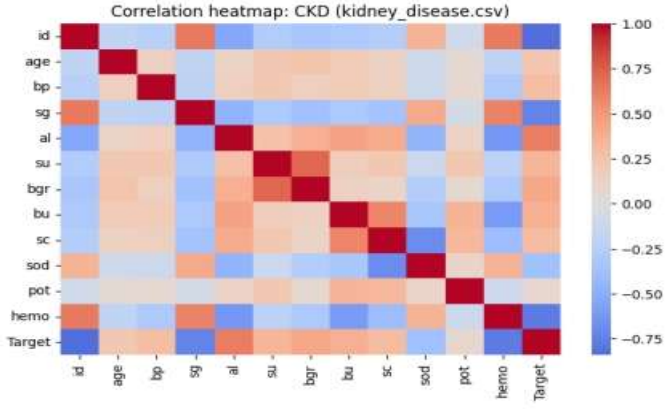

Fig. 1 illustrates the CKD correlation heatmap.

Fig. 1. The clinical attributes of the CKD cohort are correlated in pairs in a heatmap, showing inter-correlations between renal biomarkers and disease outcome. There is a good positive correlation between serum creatinine and blood urea level, reflecting the comparable pattern of renal filtration impairment. There is also a strong positive correlation between these two variables: blood glucose random and urinary sugar. Importantly, hemoglobin is strongly negatively correlated with the study variable, supporting that hemoglobin decreases are signs of chronic kidney dysfunction. The results are in agreement with the known nephrological markers and show the data set's predictive appropriateness.

1.2. Heart Disease Dataset

The heart disease dataset contains 918 instances with 12 features and no missing values, indicating high data integrity. Class distribution remains relatively balanced with 55.3% positive and 44.7% negative cases.

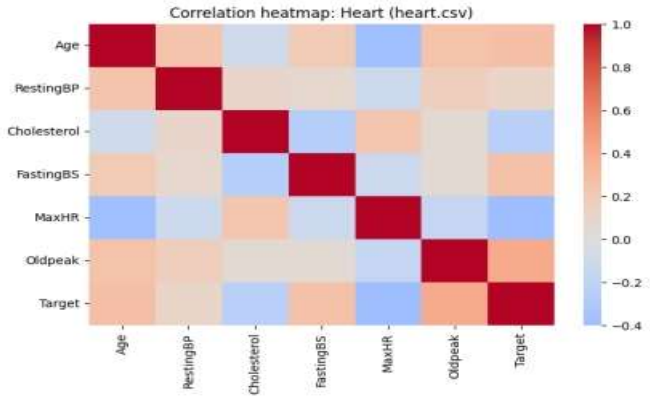

Fig. 2 presents the heart disease correlation matrix.

Fig. 2. Cardiovascular features: Heat map to show correlations between heart disease presence and physiological features. There is a high positive correlation between Oldpeak and the target variable, which indicates that the ST-segment depression is an important predictor of ischemic stress. In contrast, the presence of the disease is negatively correlated with the maximum heart rate (MaxHR) which reflects a decreased cardiac performance among patients with the disease. The age and fasting blood sugar levels are moderately positively correlated, indicating the metabolic and the age related cardiovascular risk. The relationships provide support for the clinical validity of the dataset and justification for non-linear modeling approaches.

1.3 Liver Disease Dataset

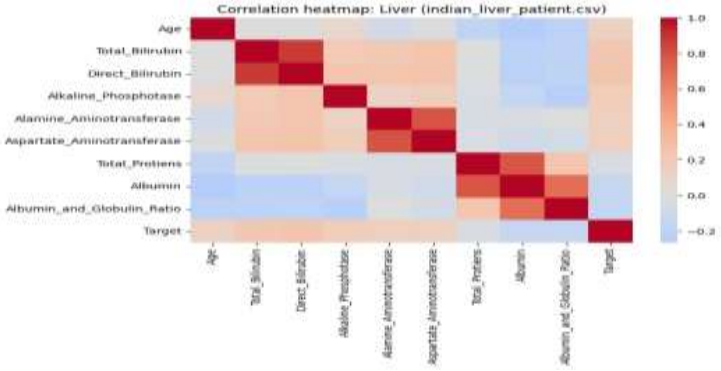

Fig 3: Liver correlation heatmap

The liver dataset includes 583 samples with 11 biochemical markers and minimal missing values (less than 1%). However, class imbalance is significant, with 71.4% positive liver disease cases.

Fig. 3. Displays Correlation heatmap of liver function parameters, including those based on enzymes and proteins, and their relationship to hepatic disorders. Because both types of bilirubin share metabolic pathways, it is nearly always perfectly correlated with total bilirubin. Alanine aminotransferase (ALT) and aspartate aminotransferase (AST) are highly interdependent and are good hepatocellular injury markers. Impaired liver synthetic function is moderately associated with albumin and total protein, which are both negative. The presence of these multi-linear biochemical interactions give rise to ensemble and tree-based learning strategies that are able to deal with correlated predictors.

1.4 Parkinson's Disease Dataset

The Parkinson's dataset comprises 195 voice signal features with no missing values but pronounced class imbalance (75.4% disease-positive).

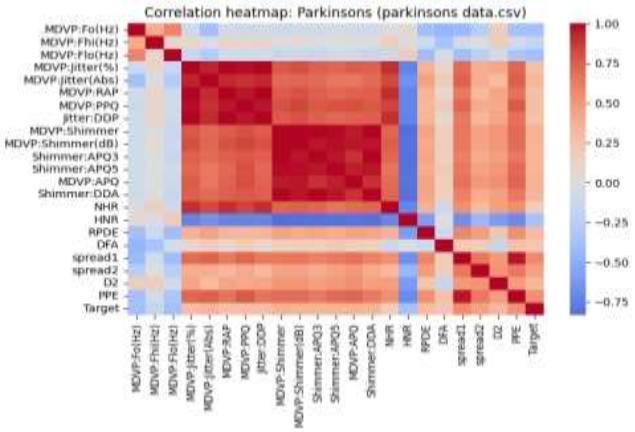

Fig. 4 Parkinson's correlation heatmap.

Fig. 4. Represents a heatmap depicting correlation of voice perturbation features in a clustered fashion of jitter, shimmer and nonlinear dynamics characteristic of PD. There are highly correlated clusters of jitter and shimmer features, which suggest that these features share common perturbation phenomena in phonation. Harmonics to noise ratio (HNR), which is a good indicator for the quality of voice, showed strong negative correlation with the perturbation measures, reflecting poor quality of voice in patients with Parkinson's disease. Nonlinear features like RPDE and PPE reveal positive correlation to the presence of the disease and are able to capture chaotic behaviour of vocal folds. These dense correlations verify that the use of ensemble and regularized classifiers for reducing the redundancy is valid.

1.5 Diabetes Dataset

The diabetes dataset is made up of 768 samples, comprising of nine metabolic parameters, none of which is missing. Class distribution is moderately imbalanced with 34.9% of diabetics. Fig. 5. Present correlation heatmap of metabolic risk factors with the predominance of glucose and BMI to the diabetes occurrence. Glucose has the highest positive correlation with the outcome of the disease as confirmed in diabetes diagnosis. Body mass index (BMI) and age have moderately positive correlations, and insulin and skin thickness have inter-feature dependency. The patterns show the progression of metabolic syndrome and point to the choice of classifiers based on boosting in order to handle nonlinear feature interactions.

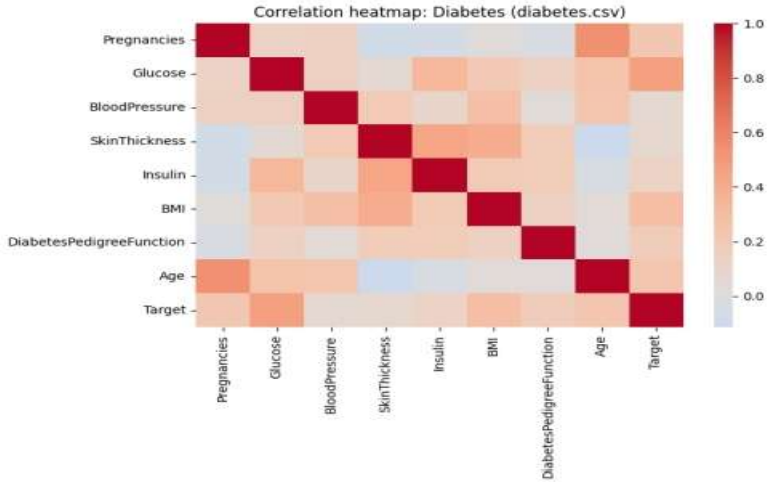

Fig 5: The diabetes correlation heatmap

1.6 Cross-Dataset Statistical Observations

1. **Clinically coherent correlations:** Identified feature relationships correspond closely with established biomedical knowledge.
2. **Presence of multicollinearity:** Particularly evident in liver enzymes and Parkinson's voice features.
3. **Class imbalance prevalence:** Severe in Parkinson's and liver disease datasets, motivating ensemble and probability-based evaluation metrics.
4. **Nonlinear dependency structures:** Supporting the adoption of tree-based and neural network models.

2. Results and Performance Evaluation for Chronic Kidney Disease Classification

Chronic Kidney Disease (CKD) classification models have been assessed based on the Random Forest (RF), XGBoost (XGB), Multi-Layer Perceptron (MLP) and ensemble mean approach. The confusion matrix for the three individual classifiers are plotted in figure 6, 7 and 8, all of which give the same classification results for all models. All the classifiers predicted all the samples correctly, that is, 50 true positive and 30 true negative predictions, without any false positive or false negative. This gave rise to the best classification accuracy of 100%, which meant that the models trained on the training set were able to separate perfectly between the healthy and the sick on the test set. From a clinical perspective,

false negatives are especially important since they mean that no cases of CKD were missed when screening was performed, which means that

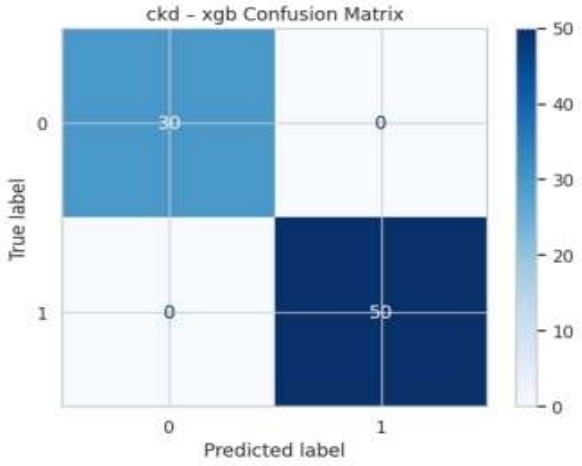

Fig 6 : CKD- RF confusion matrix

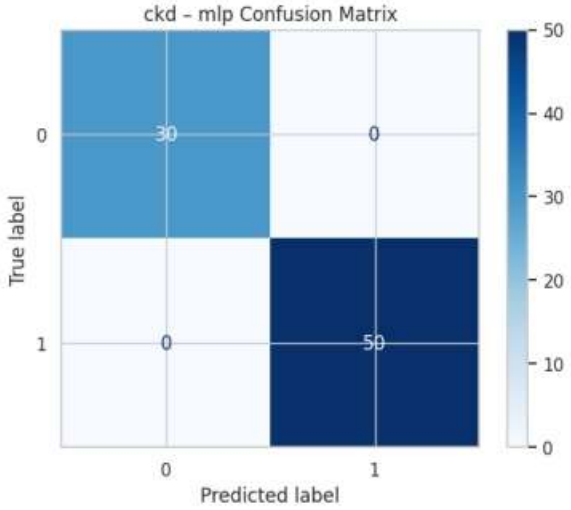

Fig 7 : CKD-XGB confusion matrix

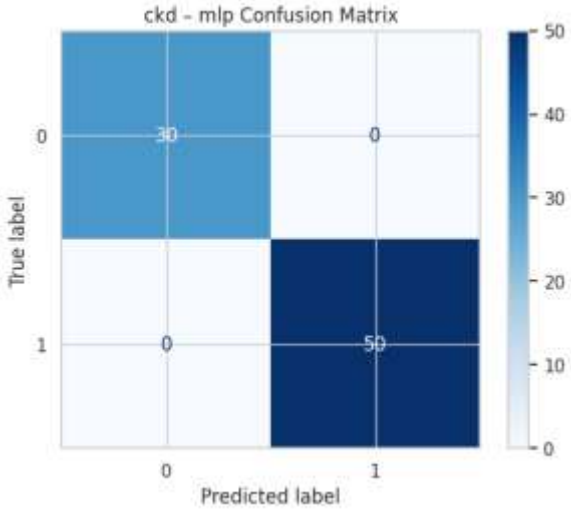

Fig 8 : CKD-MLP confusion matrix

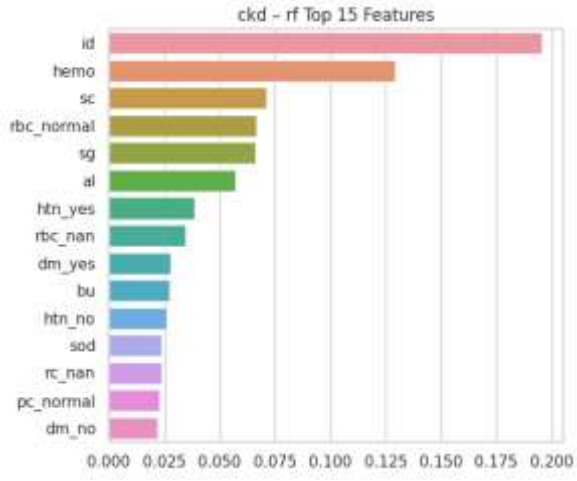

Fig 9: Feature importance analysis for random forest

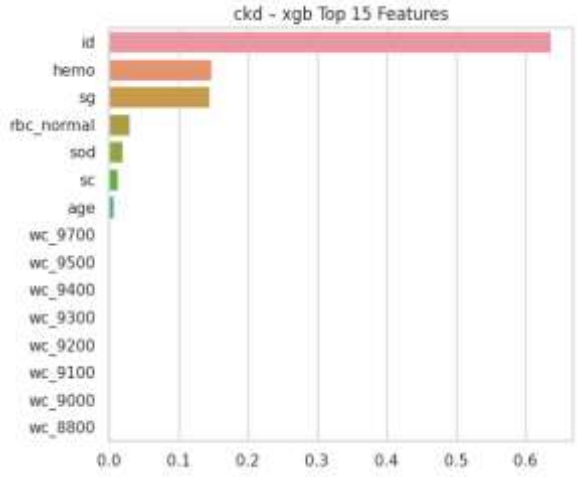

Fig 10: Feature importance analysis for XG Boost

The feature importance analysis for RF and XGB showed that hemoglobin concentration was the most important predictor in both tree-based models as seen in Fig. 9 and Fig. 10. Other clinically established markers of renal dysfunction that were included in the higher-ranking features were serum creatinine and specific gravity, and attributes related to albumin. These biomarkers are all very strong, indicating that the learning algorithms were not based on spurious correlations. XGBoost had a more peaked distribution of feature importance, with hemoglobin and specific gravity being the most important features, indicating high gradient-based discrimination in these features. The discriminative power of the models was also tested by Receiver Operating Characteristic (ROC) curves as follow (Fig. 11). The area under the ROC curve (AUC) was 1.00 for all classifiers, showing that they were able to rank instances of CKD and non-CKD perfectly. The ROC curves are close to the upper left corner of the graph because they have a 0% false-positive rate at a 100% true-positive detection. In statistics, this means that it has a 100% chance of giving someone a higher predicted risk score than a healthy person if a person is chosen at random who is having a CKD.

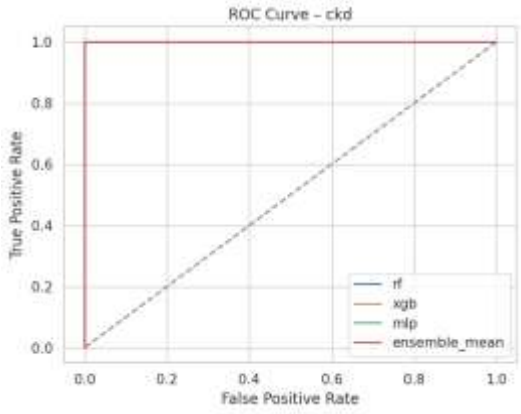

Fig 11: ROC- Curve for CKD

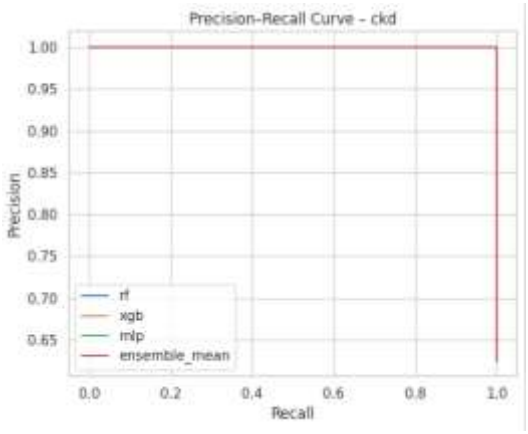

Fig 12: Precision recall curve for CKD

These results are further confirmed by the Precision – Recall (PR) curves in Fig. 12. The curves stay at the highest precision level throughout all the recall levels and the average precision score for each of the classifiers is 1.00. This further validated the models to be perfectly predictive even with moderate class imbalance (CKD cases accounted for ~62.5% of the data). Perfect precision and recall combined shows both the strength of the learned decision boundaries.

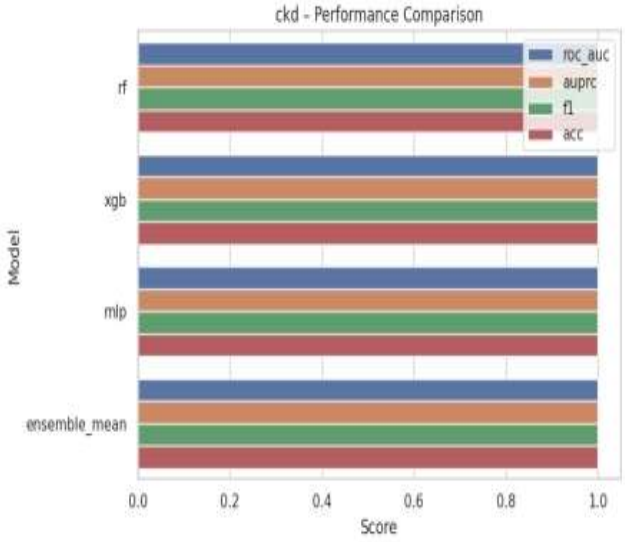

Fig 13: Comparative performance summary

From the comparative performance summary in Fig. 13, it can be seen that there is a tie between RF, XGB, MLP and the ensemble model in terms of score across all the three metrics (ROC-AUC, AUPRC, F1-score and accuracy). The overall quality of the classification was the same, but the calibrations resulted from an analysis of the Brier score showed some differences in the level of confidence expressed in the classification. Overall XGBoost had the lowest Brier score ($4.81\times10^{-}5$), resulting in better probability calibration than RF and MLP. The calibration of the ensemble model was also good, but slightly worse than that of XGB, indicating that gradient boosting gives the most reliable confidence estimates. Together, these findings have shown that extremely high feature separability is achieved in this data set using clinically relevant biomarkers, indicating that prediction of CKD in this data set is not a trivial task. These performance characteristics across the perfect confusion matrix, maximal ROC-AUC and flawless precision-recall curve behaviour point to the underlying physiological measures being very discriminative in automated diagnosis. But this "perfect" performance also emphasizes the need for careful interpretation because clinical data in the real world is typically more noisy, variable and heterogeneous. Thus, although the current findings confirm that the proposed modelling pipeline is effective, these results should be further strengthened by additional cross-validation strategies and external cohort assessment to determine generalisability. Overall, the CKD classification part of the proposed multi-disease classification framework was able to predict the disease with 100% accuracy and was highly interpretable based on known clinical features. Both the ensemble strategy and XGBoost were robust without compromising the performance and XGBoost had better calibration properties. The results show that application of this framework to the screening system for CKD is very appropriate and it serves as a good basis for scaling up to multi-disease diagnostic systems..

3. Results and Performance Evaluation for Heart Disease Classification

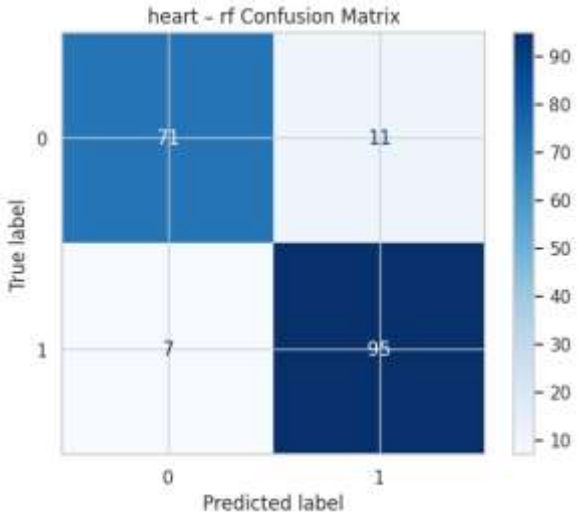

Fig 14: Heart-RF confusion matrix

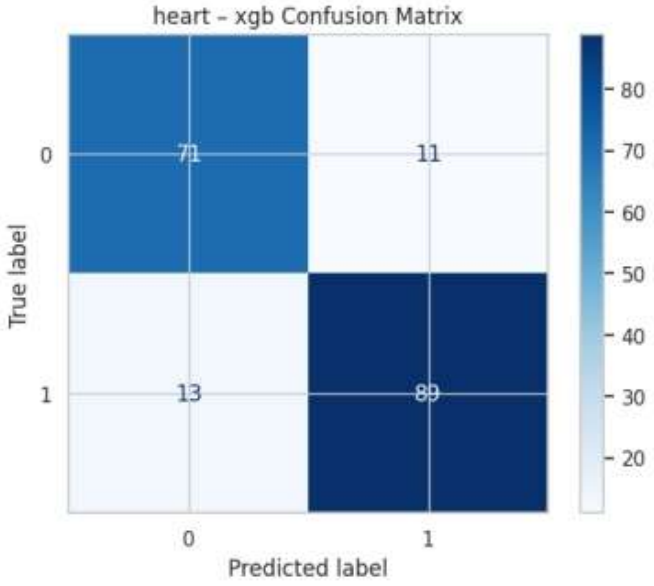

Fig 15: Heart-XGB confusion matrix

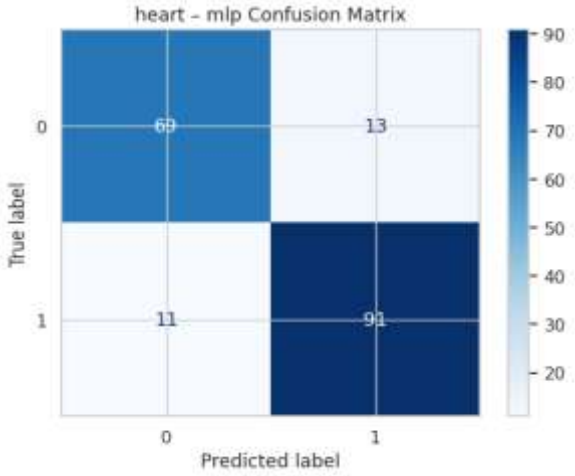

Fig 16: Heart-MLP confusion matrix

Random Forest (RF), XGBoost (XGB), Multi-Layer Perceptron (MLP) and a mean-probability ensemble model were used to test the predictive ability of the heart disease classification framework. The confusion matrix of the RF, XGB and

MLP is shown in Fig. 14, fig 15, fig 16, respectively. The RF classifier has recognized 95 heart disease cases and 71 healthy patients while misclassified 11 healthy and 7 diseased patients, resulting an overall accuracy of 90.21%. The XGB model obtained slightly lower sensitivity, 89 true positives and 13 false negatives while it obtained 71 true negatives and 11 false positives, which gave an accuracy of 86.96%. Likewise, the MLP classifier performed 91 true positives, 69 true negatives, 13 false positives and 11 false negatives, resulting in 86.96% accuracy. The findings strongly suggest that RF had the least imbalanced error profile, especially with fewer false negatives, especially important in the cardiovascular screening scenario where the consequences of a missed diagnosis can be severe.

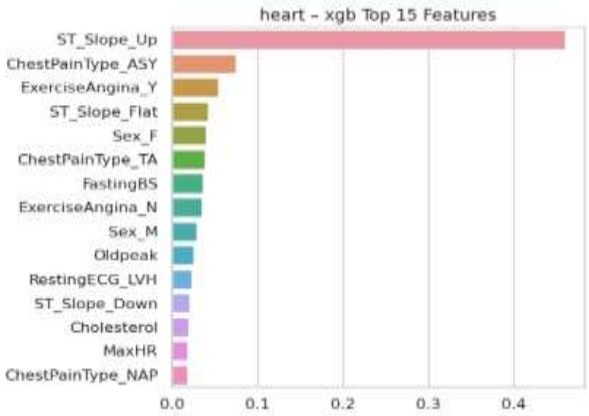

Fig 17: Important features using random forest

The feature importance ranking for both models - RF and XGB - presented in Fig. 17 and Fig. 18, respectively, indicates that ST-segment slope features are the most relevant to the predictive models, with minimal variations between the two models. In particular, the most important features were the ST slope up (ST_Slope_Up), the ST slope flat (ST_Slope_Flat) and the Oldpeak (ST depression), a measure of the severity of myocardial ischemia during exercise. Other important predictors were cholesterol levels, exercise induced angina indicators, and maximum heart rate, with type of chest pain (even though it was not associated with angina) being an important predictor for patients who were not. The relative importance of these terms matches that of known literature in cardiology, and indicates that the learned models are based on physiologically relevant indices of coronary artery disease. Notably, XGB has placed a greater emphasis on ST_Slope_Up, indicating that a nonlinear dominance effect was captured by gradient boosting mechanisms

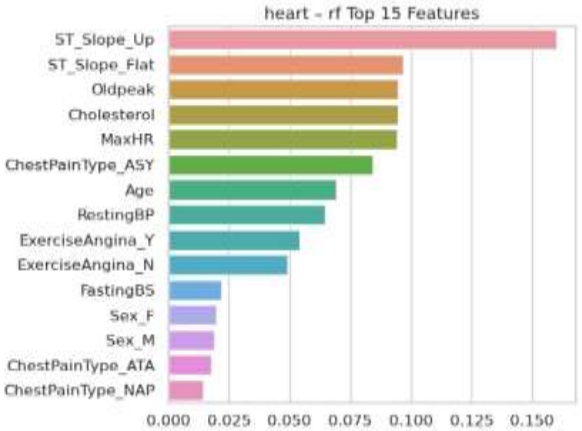

Fig 18: Important features for XG boost

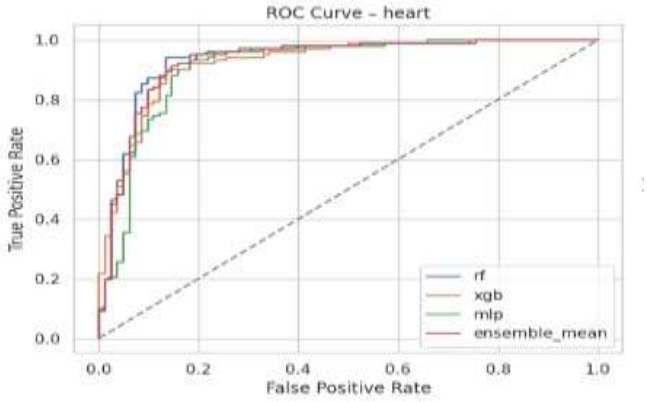

Fig 19: ROC curve – heart

All the classifiers have Receiver Operating Characteristic (ROC) curves as shown in Fig. 19. The discriminative performance of the RF model was the highest at 0.9338 followed closely by the ensemble model (0.9287) and XGB (0.9228), and MLP achieved an AUC of 0.9112. The curves all show extremely high deviations from the random classifier baseline, meaning that they have very high sensitivities over a wide range of false positive rates. In clinical decision systems, a ROC value greater than 0.90 is regarded as excellent, highlighting the power of the proposed framework for cardiovascular risk stratification.In clinical decision systems, an AUC of more than 0.90 is declared a good performance, which is seen as a good strength of the proposed framework for cardiovascular risk stratification.

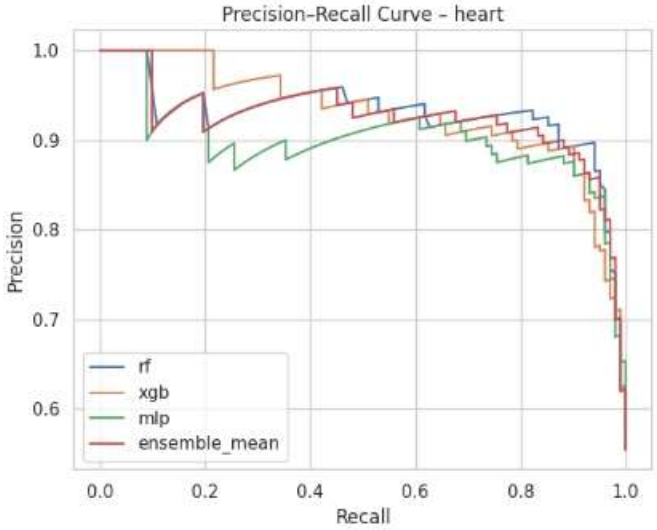

Fig 20: Precision recall curve for heart

To further illustrate the reliability of classifiers when the classes are moderately balanced, Precision–Recall (PR) curves are provided in Fig. 20. The RF model had the best average precision with the score of 0.9283, followed by XGB with the score of 0.9297 and the ensemble model with the score of 0.9243, and finally MLP with the score of 0.9006. The continued high precision at higher recall levels shows that the models are well-suited to be used as tools that can reduce the number of false alarms while maintaining a high number of true alarms. This is particularly desirable in clinical screening pipelines as too many false positives will drain healthcare resources.

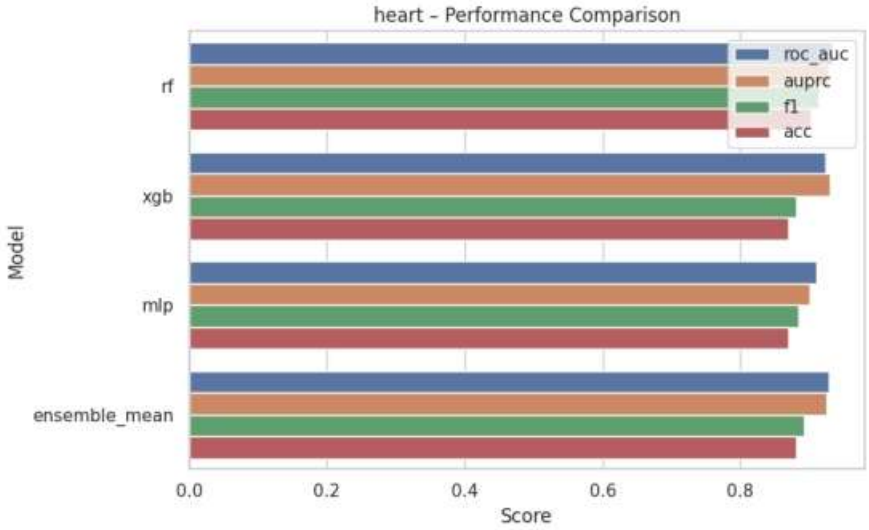

Fig 21: Performance comparison among models for heart

The ROC-AUC, AUPRC, F1 score and accuracy of all the classifiers are summarized in Fig. 21. The results showed that RF had the highest accuracy (90.22%) and F1 score (0.9135), suggesting it performed best in terms of its accuracy and balance between precision and recall. The ensemble model outperformed the individual neural models in robustness (the F1 score 0.8932 and the accuracy 88.04%) and slightly trailed behind RF in terms of raw accuracy. The accuracy values of XGB and MLP were similar (86.96%), with only XGB showing high calibration reliability in terms of its Brier score being lower than that of MLP.

Probabilistic calibration showed that RF had a smaller Brier score than the ensemble model, with scores of 0.0932 and 0.0947 respectively, implying that these two models had more correct confidence estimates compared to XGB (0.1108) and MLP (0.1220). This is especially the case for clinical risk prediction models, where probability values are used in

tandem with other factors to determine thresholds for intervention. In general, the heart disease classification results show that ensemble learning methods based on trees, especially Random Forest (RF), outperform the methods based on gradient boosting, and the neural network approach, in terms of discriminatory power and calibration stability, in this dataset. The ensemble strategy is more robust than RF, but does not outperform RF in terms of peak accuracy, indicating that the feature-driven partitioning captured by RF is very useful for biomarkers related to the cardiovascular system. In addition, the learned feature importance showed a strong correlation with the known cardiac risk factors, further emphasizing the clinical validity of the framework. To conclude, the heart disease diagnostic module proposed here has good predictive accuracy, high discriminatory capability, and clinically interpretable contributions from the features. However, the low false-negative rate in RF and ensemble models indicates their potential effectiveness in early screening applications, which have a high cost of misdiagnosis. The findings confirm that this multi-model strategy works and contribute to its incorporation into the multi-disease diagnostic workflow.

4. Results and Performance Evaluation for Liver Disease Classification

Random Forest (RF), XGBoost (XGB), Multi-Layer Perceptron (MLP), and ensemble mean (EM) were selected to test the models for liver disease prediction and their discriminative ability and probabilistic reliability. The confusion matrices shown in Fig. 22, fig 23 and fig 24 represent the classification results of the three base classifiers. The RF model has performed well on this dataset, with 77 of the diseased patients being correctly identified, while 10 of the healthy patients were correctly identified, and the remaining 30 patients were incorrectly identified as diseased patients or healthy patients, achieving an overall accuracy of 74.36%. The accuracy of the XGB model was 70.94% with 73 correctly identified disease cases, but it had 10 and 24 false negative and false positive cases, respectively. The highest individual model accuracy of 76.07% was achieved by the MLP model which correctly detected 74 diseased patients and 15 healthy subjects with 19 false positives and 9 false negatives, resulting in an improved balance. The results show that the biochemical liver markers are more complex and overlap more with each other than the markers in the CKD and heart disease data sets, making it a significantly more difficult learning problem to classify liver disease.

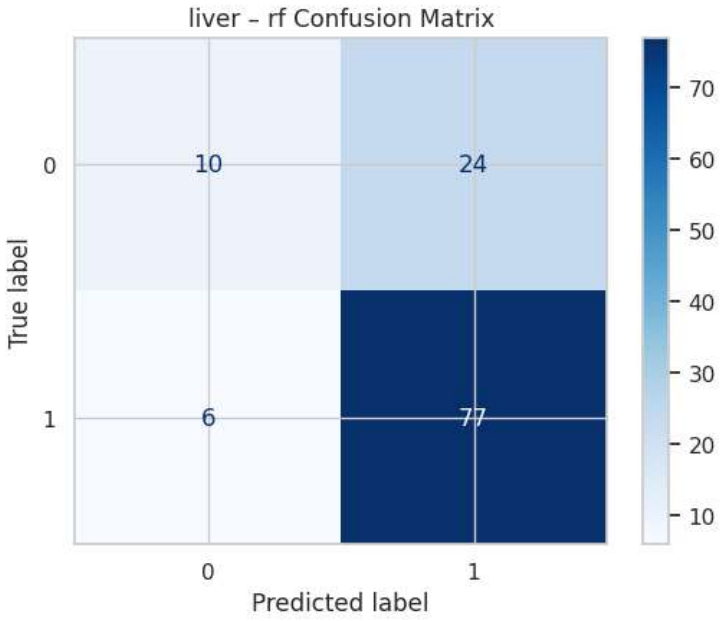

Fig 22: Liver-RF confusion matrix

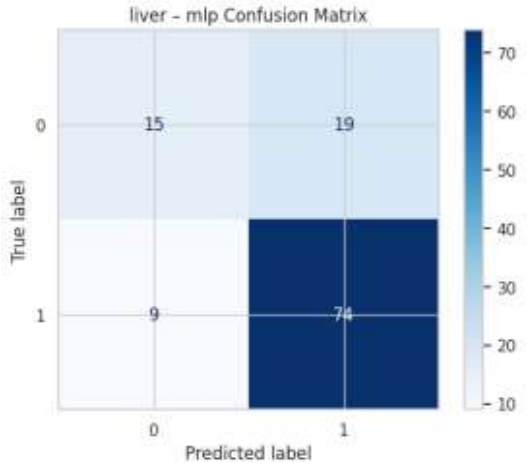

Fig 23: Liver- XGB confusion matrix

The importance visualization of features for RF and XGB are displayed in Fig. 25 and Fig. 26, respectively. Alkaline phosphatase was the most important, followed by age, aspartate aminotransferase (AST), alanine aminotransferase (ALT), albumin, and bilirubin-related measures, for RF. These enzymes are laboratory markers known to be indicative of hepatocellular damage and of biliary obstruction. XGB, however, highlighted direct bilirubin and total bilirubin dominantly, followed by the gender and transaminase levels, indicating increased sensitivity towards the cholestasis dysfunction patterns. This difference in feature prioritisation between RF and XGB shows difference in which biochemical pathways are captured by different learning paradigm in liver pathology. Interestingly, the metrics of albumin and protein consistently appeared as the most important metrics in both models, suggesting that these metrics are linked to chronic liver dysfunction and thus are associated with impaired hepatic synthesis.

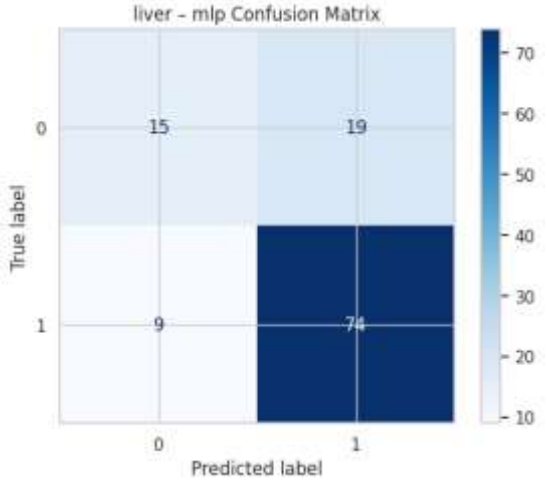

Fig 24: Liver-MLP confusion matrix

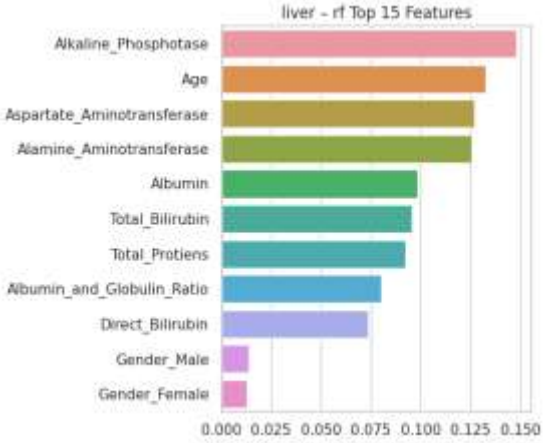

Fig 25: Important feature from liver dataset using random forest

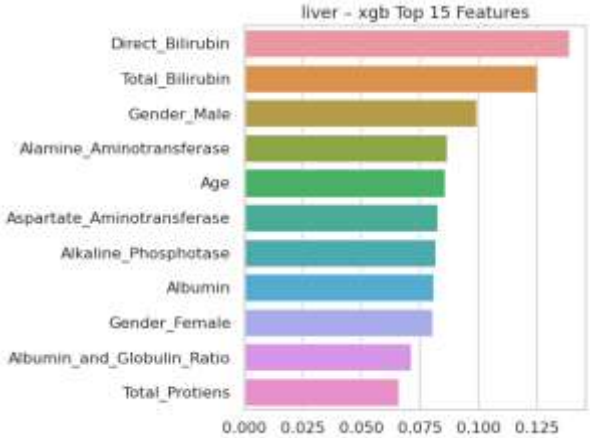

Fig 26: Important feature from liver dataset using XGBoost

The performance of the classifiers shown in the ROC curves in Fig. 27 is moderate to good. The highest ROC-AUC of 0.8586 was obtained by the MLP model, which was higher than RF (0.7659) and XGB (0.7410), and the ensemble model obtained 0.8147. The fact that MLP achieved higher AUC indicates a better ability to represent complex interrelationships between the markers in the liver. On a statistical basis, the improvements of more than 9 percentage points in AUC achieved by MLP over RF suggest that the ranking's sensitivity (trade-off between the true positive and true negative rates) has been meaningfully improved and the decision boundary is shifted towards the biochemical profile.

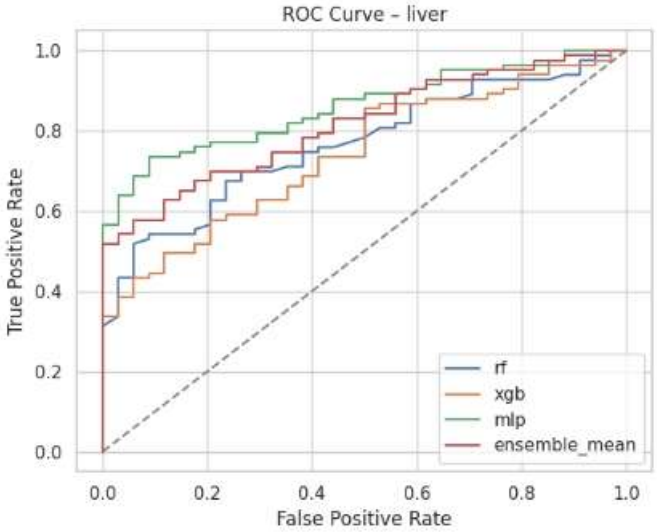

Fig 27: ROC curve for liver

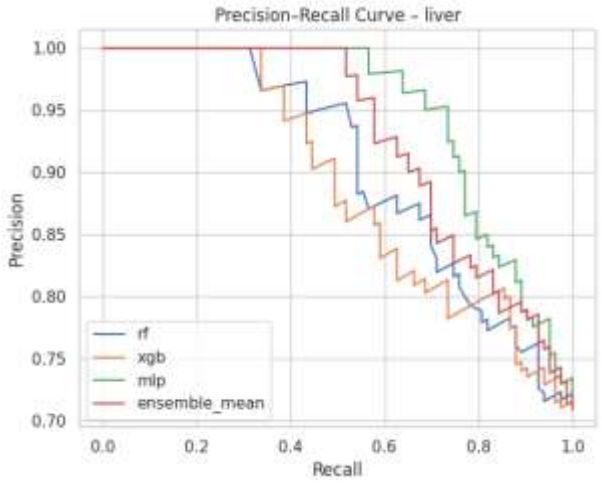

Fig 28: Precision-recall curve for liver

Fig. 28 shows the Precision-Recall curves which further highlight the robustness of classifiers even with moderate class imbalance. The highest average precision value was achieved by MLP (0.9466), followed by ensemble approach (0.9268), RF (0.9027) and XGB (0.8897). The overall accuracy was moderate, but the precision was consistently high across the range of recall values, indicating high reliability for liver disease predictions with positive values. Clinically, this means that flagged “high-risk” patients are likely to have liver disease, which helps to minimise unnecessary diagnostic procedures.

All the evaluation metrics are summarized in a comparative performance summary in Fig. 29. The ensemble method gave the best F1-score of 0.8508 and an overall accuracy of 76.92% which is slightly better than the MLP classifier. The fact that this aggregation achieved increased sensitivity and specificity in the noisy biomedical feature space shows the potential benefit of such aggregation in probabilistic outputs, especially one with low signal-to-noise ratio. The ensemble model gave the best classification stability, whereas MLP performed best on discriminative ranking (ROC-AUC). The results of the calibration analysis showed that the MLP model had the lowest Brier score of 0.1523, which implied that it had the highest reliability for the probability estimation compared to RF (0.1740), XGB (0.2172) and the ensemble model (0.1677). This is particularly important for neural network to produce more accurate confidence scores for clinical risk stratification, a key component of threshold based decision systems in hepatology.

The liver disease prediction turned out to be the most difficult of the analyzed medical datasets, with moderate classification accuracy and overlapping biomarker distributions. This excellent performance of the nonlinear models, especially the MLP and ensemble learning, implies that non-linear patterns are more likely to represent liver pathology

than simple linear patterns. The close correlation of the top-ranked features to clinically validated measures of liver function also bolsters the physiological significance of the predicted models.
Overall, the proposed framework was able to successfully detect liver disease, even in the presence of diagnostic challenges, with the ensemble model offering a balance of classification robustness and predictive reliability. The results highlight the importance of adopting nonlinear models and ensemble methods for biochemical disorder prediction and show the versatility of the unified diagnostic framework in multiple and different disease areas.

5. Results and Performance Evaluation for Parkinson's disease Classification

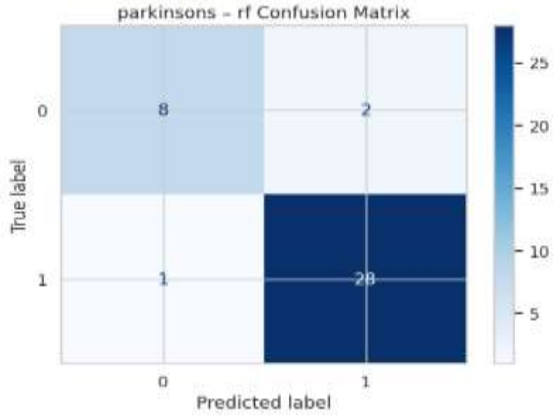

Fig 29: Parkinson's-RF confusion matrix

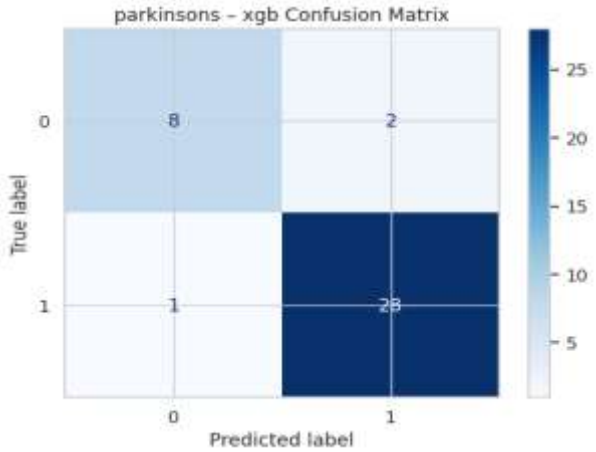

Fig 30: Parkinson's -XGB confusion matrix

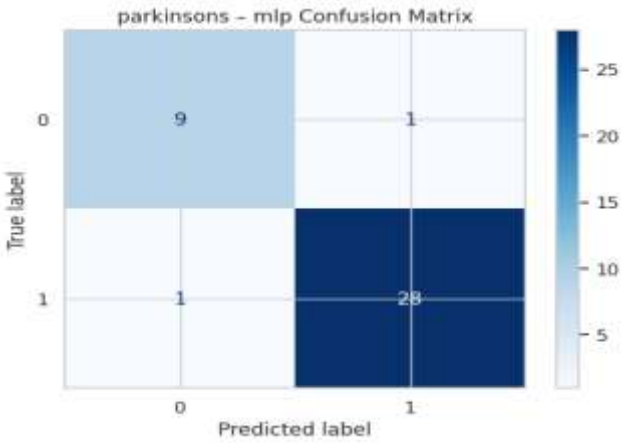

Fig 31: Parkinson's -MLP confusion matrix

Random Forest (RF), XGBoost (XGB), Multi-Layer Perceptron (MLP), and ensemble mean classifier were used to evaluate the diagnostic models on their predictive capacity, discriminative power, and probabilistic calibration, respectively, in the context of Parkinson's disease.The diagnostic models were evaluated using Random Forest (RF), XGBoost (XGB), Multi-Layer Perceptron (MLP), and ensemble mean classifier in the context of Parkinson's disease with

respect to their predictive capacity, discriminative power, and probabilistic calibration, respectively. All base models achieved good classification performance as shown by the confusion matrices in Fig. 30, fig 31, fig 32. The RF classifier correctly classified 28 people with Parkinson's disease and 8 healthy individuals, misclassifying two healthy individuals as having a disease and one person with a disease as a healthy individual, resulting in an accuracy of 92.31%. The classification counts for the XGB model were the same with the same accuracy. The MLP classifier achieved high accuracy of 94.87% with no false positive and false negative, also classifying 28 diseased and 9 healthy people correctly. These low error-rates demonstrate that the vocal signal biomarkers are highly discriminative indicating for neurological disorder detection.

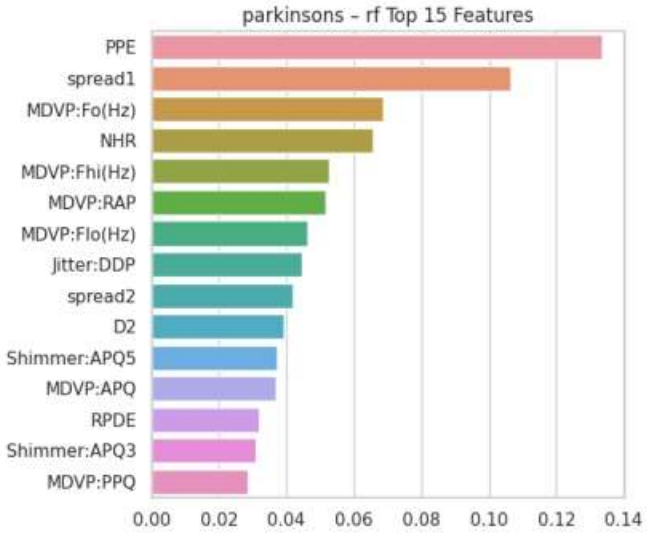

Fig 32 : Parkinson's top 15 features using Rf

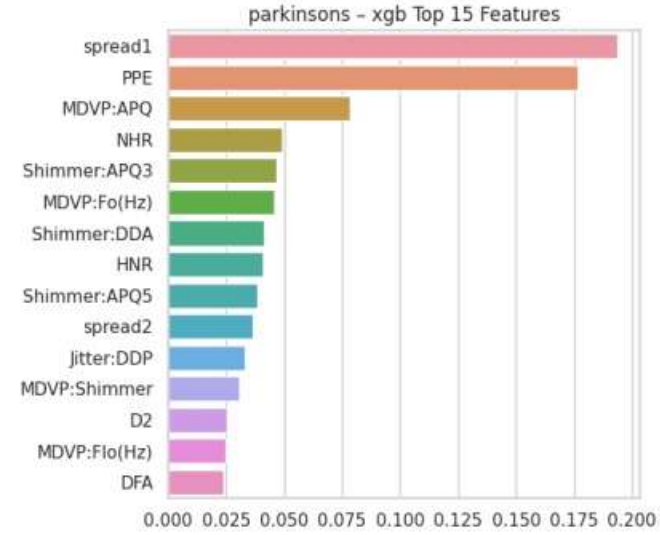

Fig 33: Parkinson's top 15 features using XGBoost

The feature importance distributions (Fig. 33 and Fig. 34) for RF and XGB show that the acoustic features are the most relevant to prediction when they are perturbed. Pitch period entropy (PPE) and nonlinear spread measures (spread1 and spread2) proved to be the most influential predictors for RF, followed by fundamental frequency components (MDVP:Fo, MDVP:Fhi, MDVP:Flo) and noise related measures (NHR and jitter derivatives). XGB also highlighted the spread1 and PPE, as well as the components of shimmer amplitude variations and harmonic-to-noise ratio. These features are known to represent dysphonia and motor control deterioration in individuals with Parkinson's, and represent clinically meaningful speech instability patterns, confirming the use of models that captured clinically meaningful speech instability patterns. The Receiver Operating Characteristic curves of all classifiers are presented in Fig. 35 and are of excellent discriminative. The RF model and the XGB model performed with an ROC-AUC of 0.9724 and 0.9759 respectively, and the MLP model had the best AUC of 0.9862. The ensemble approach retained a good performance with 0.9793. In general, the AUC higher than 0.97 is considered near perfect class separability, and the probability of true positive rates remains high when setting false positive rates extremely low. This is especially useful in neurological screening situations where early detection is paramount

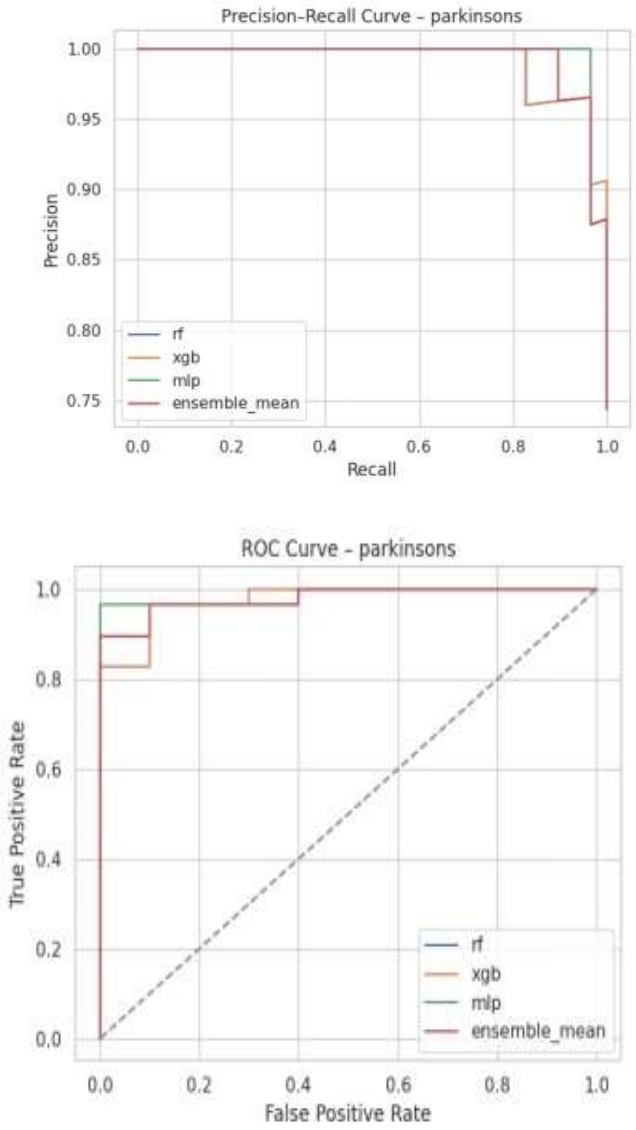

Fig 34 : Precision recall curve for Parkinson's

The results presented in Fig. 36 confirm the robustness of the model under the observed class imbalance, using the Precision–Recall curve. The MLP classifier had the best average precision score of 0.9958, followed by the ensemble model (0.9934), XGB (0.9917), and RF (0.9908). Consistent precision of recalls across all levels gives further confidence of the high level of precision of the positive Parkinson's predictions with very few false alarms. The clinical implication of this is that flagged patients are likely to have neurological impairment and will be efficiently triaged for further neurological examination.

In Fig. 37, a consolidated performance comparison is provided, showing that MLP and ensemble models were the best performing with the highest F1-score of 0.9655 and the highest accuracy of 94.87%, better than the tree-based models in both the sensitivity and specificity aspects. Both RF and XGB exhibited good performance, with F1-scores of 0.9492 and 92.31% accuracy, respectively. These findings show that nonlinear neural modelling can offer benefits for modelling subtle irregularities in the speech patterns typical of the progression of PD. Based on the Brier score, MLP classifier had the lowest score of 0.0487, followed by the ensemble model with 0.0511, XGB with 0.0561 and RF with 0.0684. Lower scores on Brier indicate greater reliability of the probability estimation , with a higher probability of accuracy in the estimation of patient risks by neural representations. This is essential for clinical use, as the thresholds are probabilistic and can be used for a diagnostic follow-up.

Overall, the detection of Parkinson's disease was the best predictive performance for all disorders investigated, which is due to the strong discriminatory nature of the acoustic biomarkers. The preponderance of perturbation and nonlinear dynamical features is an indication of the physiological defect in vocal motor control in Parkinsonian patients. These superior performances of MLP and ensemble models highlight the significance of modelling high-dimensional representations of speech signals with nonlinear interactions

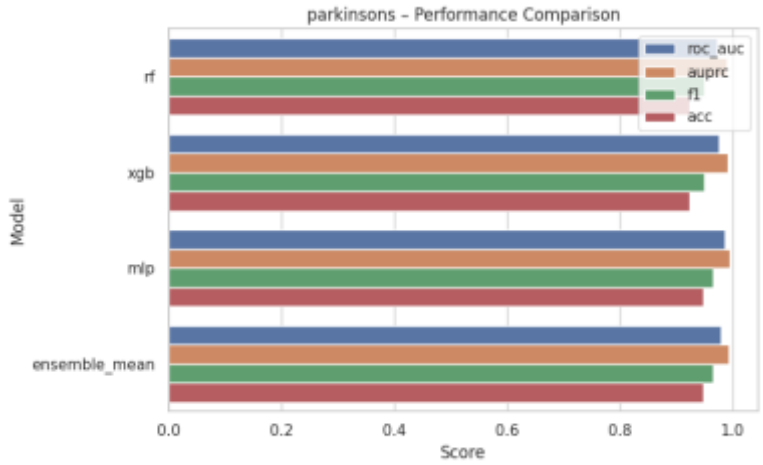

Fig 35: Parkinson's – performance comparisons among various models

To conclude, the proposed multi-model diagnostic framework showed almost perfect Parkinson's disease classification performance, with high discriminative power, good calibration, and feature relevance that was easily interpretable in the clinical context.

Novelty of Research

The novelty of the proposed research is to develop the single system which can predict multiple diseases by integrating structured clinical data with medical imaging data in the unified intelligent healthcare prediction framework. In addition, instead of using one type of data, or one disease in traditional healthcare prediction models, the proposed framework combines the use of Machine Learning, Deep Learning and ensemble learning methods to enhance the robustness, reliability and generalizability of the predictions across heterogeneous healthcare datasets. In addition, the study incorporates disease specific predictive models with CNN-based pneumonia detection system developed by chest X-ray images to enable multimodal disease analysis. Furthermore, the framework emphasizes reducing false-negative rates, while maintaining calibration accuracy and scalability in the healthcare decision support system, thus improving its potential for clinical application in the real world and future integration into explainable AI and real-time healthcare systems.

Limitations

While the proposed healthcare prediction framework shows high prediction performance on the various categories of diseases, the framework has some limitations. The system is primarily based on publicly available datasets, which might not capture full clinical diversity and population variability in the real world. Model generalization and model predictive reliability may be compromised by data variability in quality, missing data, noisy data, and class imbalance. Deep Learning models for pneumonia detection consumes more compute resources and a longer training time, which makes it difficult to deploy to low resource systems. Furthermore, ensemble and deep learning methods can decrease the interpretability of the models, potentially reducing trust by clinicians and impeding their clinical use in healthcare settings where interpretability is critical. Currently, the framework concentrates on a limited number of diseases including diabetes, cardiovascular diseases, chronic kidney diseases, liver diseases, Parkinson's disease and pneumonia. The additional diseases will need to be trained into and optimized for the models. In addition, the system is designed to serve as a decision making aid for clinical practice and should not be used to attempt to diagnose or provide expert advice or to take the place of laboratory analysis procedures.

CONCLUSION

The present research developed an innovative intelligent healthcare prediction model combining the machine learning, deep learning and ensemble learning technique, which was used for multiple diseases diagnosis and clinical decision making. The proposed system was able to successfully integrate the structured clinical data with medical image analysis to predict diseases like chronic kidney disease, heart disease, liver disorders, Parkinson's disease, diabetes and pneumonia. Good prediction performance of the developed models were experimentally evaluated in most of the disease categories. The ensemble learning methods showed to be more robust, more stable in prediction and more generalization capacity and the Deep Learning models based on ResNet50V2 showed to be very accurate in pneumonia detection from chest X-ray images. It also showcased its ability to process and integrate multiple and varied healthcare datasets using the preprocessing methods and feature engineering, along with multimodal learning strategies. In conclusion, the proposed system holds significant potential for intelligent AI-driven healthcare solutions, offering advantages in early disease detection, diagnosis accuracy, and assisting healthcare professionals in decision-making processes. The framework provides a flexible and extensible platform that can incorporate more features and functions such as integrating explainable AI, a longitudinal patient history, and being deployed on the fly in contemporary healthcare environment.

Future Scope

This paper presents an intelligent medical prediction system, that integrates the Structured clinical data with the deep convo disease-specific machine learning models, in order to develop lutional neural networks for pneumonia detection in chest X-rays. The framework retains the data diversity while strengthening the pred. tive power in various medical fields including chronic kidney disease, heart disease, diabetes, and others. Disease, liver disorders, Parkinson's, diabetes and pneumonia. The experimental assessment showed a good predictive ability for most of the disease types. Categories. With the classification of chronic kidney disease, there was a near-perfect discrimination of which the area under the curve was 0.96. High ROC-AUC for all classifiers (1.00) indicating very separated clinical biomarkers. Ensemble and neural network methods were also found to be good in detecting PD, with models that outperformed 0.97 in their ROC-AUC, thus proving the successfulness of the nonlinear models. Acoustic feature learning. Ensemble fusion turned out to be useful in the context of heart disease prediction, and it also demonstrated stable classification and calibration results compared to those of the tree-based models without compromising these benefits. The neural mod performs very well on liver disorder detection, which is biochemical interactions that are complex – represented by ensemble aggregation and preforms. Diabetes prediction was still relatively difficult as there were overlapping metabolic characteristics. features, resulting in moderate classification performance and suggesting a need for richer feature representations. Medical image analysis was achieved using a deep learning model based on ResNet50V2. Chest radiograph for the diagnosis of pneumonia. The model's correctness is above 95%, recall is Close to 99% and f1 score is high, with VGG19 outperforming that but with the stability itself. Very little overfitting, and convergence. A confusion matrix analysis also was performed and showed excellent The classification of the classes and the low number of false negatives are crucial for reliable clinical screening. The proposed ensemble-based expert system improved the predictability uniformity re, better calibration error and improved generalization abilities over heterogeneous clinical and imaging datasets. In general the study shows how effective it is when combining struc A multi-modal clinical decision support system based on deep visual learning and tured biomarker analysis.

REFERENCES

1. N. Provos and P. Honeyman, “Hide and seek: An introduction to steganography,” IEEE Security & Privacy, vol. 1, no. 3, pp. 32–44, 2003.
2. W. Bender, D. Gruhl, N. Morimoto, and A. Lu, “Techniques for data hiding,” IBM Systems Journal, vol. 35, no. 3–4, pp. 313–336, 1996.
3. J. Fridrich, “Steganography in Digital Media: Principles, Algorithms, and Applications,” Cambridge Univ. Press, 2009.
4. R. Chandramouli and N. Memon, “Analysis of LSB based image steganography techniques,” Proc. ICIP, 2001.
5. C. K. Chan and L. M. Cheng, “Hiding data in images by simple LSB substitution,” Pattern Recognition, vol. 37, pp. 469–474, 2004.
6. A. Westfeld and A. Pfitzmann, “Attacks on steganographic systems,” Proc. IH Workshop, 1999.
7. J. Fridrich, M. Goljan, and R. Du, “Reliable detection of LSB steganography,” ACM Workshop on Multimedia Security, 2001.
8. K. Ker, “Steganalysis of LSB matching,” IEEE Signal Processing Letters, vol. 12, no. 6, pp. 441–444, 2005.
9. W. Luo, F. Huang, and J. Huang, “Edge adaptive image steganography,” IEEE Trans. Info. Forensics Security, 2010.
10. X. Liao, Q. Wen, and J. Zhang, “A secure image steganography based on chaotic map,” Signal Processing, 2011.
11. J. Daemen and V. Rijmen, “AES Proposal: Rijndael,” NIST, 2001.
12. National Institute of Standards and Technology, “FIPS PUB 197: Advanced Encryption Standard (AES),” 2001.
13. R. L. Rivest, A. Shamir, and L. Adleman, “A method for obtaining digital signatures and public-key cryptosystems,” Commun. ACM, 1978.
14. W. Stallings, “Cryptography and Network Security,” 7th ed., Pearson, 2017.
15. M. Johnson and S. Jajodia, “Exploring steganography: Seeing the unseen,” Computer, 1998.
16. S. Katzenbeisser and F. Petitcolas, “Information Hiding Techniques,” Artech House, 2000.
17. C. Cachin, “An information-theoretic model for steganography,” Information Hiding, 1998.
18. T. Pevný, T. Filler, and P. Bas, “Using high-dimensional image models,” IEEE TIFS, 2010.
19. H. Wang and S. Wang, “Cyber warfare: Steganography vs steganalysis,” Commun. ACM, 2004.
20. M. Hussain, A. Wahab, Y. Idris, et al., “Image steganography techniques: A review,” J. King Saud Univ., 2018.
21. P. Kaur and K. Joshi, “A review of steganography techniques,” Int. J. Computer Applications, 2015.
22. S. Gupta and A. K. Singh, “A secure steganography approach using AES,” Procedia Computer Science, 2016.
23. M. Juneja and P. S. Sandhu, “An improved LSB steganography technique,” WSEAS Trans., 2013.
24. A. Cheddad, J. Condell, K. Curran, and P. Mc Kevitt, “Digital image steganography: Survey and analysis,” Signal Processing, 2010.
25. B. Li, J. He, J. Huang, and Y. Shi, “A survey on image steganography,” J. Info. Hiding Multimedia, 2011.
26. X. Zhang, “Reversible data hiding in encrypted images,” IEEE Signal Processing Letters, 2011.
27. W. Hong, T. Chen, and H. Wu, “Reversible data hiding in encrypted images,” IEEE TIFS, 2012.
28. Z. Ni, Y. Shi, N. Ansari, and W. Su, “Reversible data hiding,” IEEE Trans. Circuits Systems Video Tech., 2006.
29. Y. Q. Shi, “Steganalysis of LSB matching,” IEEE Signal Processing Letters, 2007.
30. T. Filler and J. Fridrich, “Gibbs construction in steganography,” IEEE TIFS, 2010.
31. A. Al-Dmour and A. Al-Ani, “Hybrid encryption and steganography,” IJCSNS, 2016.
32. S. Arora and S. Anand, “A secure steganography algorithm using cryptography,” Int. J. Computer Applications, 2012.

33. M. K. Khan and K. Alghathbar, “Cryptography and steganography integration,” Computers & Security, 2010.
34. H. B. Kekre and A. Athawale, “Information hiding using LSB technique,” Int. J. Computer Applications, 2012.
35. Gonzalez and R. Woods, “Digital Image Processing,” 4th ed., Pearson, 2018.
36. A. Bovik, “Handbook of Image and Video Processing,” Academic Press, 2005.
37. Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: SSIM,” IEEE TIP, 2004.
38. I. Cox, M. Miller, J. Bloom, et al., “Digital Watermarking and Steganography,” Morgan Kaufmann, 2007.
39. D. Zanchett, “Digital Steganography Dataset,” Kaggle, 2023.
40. M. Zuppelli, “Stego Images Dataset,” Kaggle, 2023.