Prediction Of Multiple Diseases in Rural Populations Using Machine Learning Techniques Through an Accessible Web-Based Application

SIDDHARTH PANDIT1, NITU SIKCHI2, TAMKEEN SUMAIYA3, APEKSHA RAVI4, AASMA VERMA[5], DR. S. SENTHILKUMAR[6]

[1, 2, 3, 4, 5] *Department of CSE, Presidency University, Bangalore*
[6] *Professor & HOD-Department of CSE, Presidency University, Bangalore*

Abstract- *This paper provides an in-depth examination of the transformative impact of machine learning algorithms in healthcare, specifically in predictive disease identification. The introduction highlights the revolutionizing role of machine learning in healthcare, enabling systems to learn from medical data for disease prediction without explicit programming. Support Vector Machine, K-Nearest Neighbours, Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, XGBoost, and AdaBoost are supervised learning algorithms that are investigated for their capacity to examine intricate relationships within datasets and help physicians make well-informed decisions. Then the model is deployed to a web application which is developed using the Django framework.*

Indexed Terms - *Machine learning, healthcare, predictive disease identification, supervised learning algorithms, Support Vector Machine, K-Nearest Neighbors, Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, XGBoost, AdaBoost, medical data, web application, Django framework.*

I. INTRODUCTION

Machine learning algorithms have transformed healthcare by providing illness detection models. These artificial intelligence algorithms allow systems to be explicitly configured to learn from data and generate predictions. They analyze medical data in healthcare to identify patterns and correlations that aid in disease prediction. Algorithms are trained using supervised learning on datasets containing patient information, symptoms, and illness outcomes. These models learn to recognize complicated patterns in data, allowing physicians to make informed assessments and accurately predict diseases. The incorporation of machine learning techniques has had a substantial impact on disease prediction across a variety of medical areas. These algorithms have shown potential in foreseeing diseases such as diabetes, breast cancer, Parkinson's, heart disease, and liver disease. Different models use learned patterns to forecast the likelihood of various disorders by using datasets containing patient histories, genetic information, and clinical features. These approaches aim to improve early diagnosis, treatment options, and patient outcomes across a wide range of diseases.

II. LITERATURE REVIEW

A. Heart Disease

Parise Divyasri, D SreeLakshmi, et al examined different machine learning approaches and combined their predictions into an effective hybrid model to predict if the patient has heart disease based on input factors[1]. The algorithms used were decision tree, KNN, SVM, random forest, XGBoost, and adaptive voting classifier. The Adaptive Voting Classifier gave the highest accuracy of 99.98% with a precision of 98.79%. By merging various algorithms, they found that it provided more accuracy than single models. It also helped to resolve the failures that occurred in individual procedures.

Taher M Ghazal et al focused on predicting which patients have more chances of suffering from heart disease based on their previous medical features[2]. K nearest neighbour and support vector machine algorithms were employed for classification. The confusion matrices of both models were analyzed to compare their respective accuracies and it found that KNN was the better algorithm having a higher accuracy of 86% than SVM.

Various machine algorithms like KNN, decision tree, random forest, naïve Bayes, logistic regression, and gradient boosting were analyzed by Sarah Ibrahim, Nazih Salhab, and Ammar El Falou[3]. The performance of each algorithm was evaluated by calculating metrics like

accuracy, precision, recall, and f1 score. KNN and logistic regression were concluded to have the shortest training time and logistic regression performed better than the other algorithms.

Vijeta Sharma, Shrinkhala Yadav, and Manjari Gupta[4] chose four popular machine learning algorithms to develop a heart disease prediction model using a benchmark dataset of UCI heart disease. Their approach involves the traditional machine learning technique of data collection, data preprocessing, building the model, and evaluating the model performance. Out of the four algorithms – SVM, random forest, decision tree, and naïve Bayes, Random Forest gave the maximum accuracy of almost 99 percent.

B. Parkinson's Disease
Liu Feng et al[5] used three machine learning algorithms to predict success for Parkinson's disease. After comparing the corresponding accuracy and recall rate, an improved KNN algorithm is proposed with the information entropy. The entropy method is applied to the KNN algorithm and it achieved a 93.8% accuracy rate and 97 % recall rate. This method of entropy empowers the sample data in the Parkinson's dataset and makes full use of the data's contribution to location and the actual experimental results show the effectiveness and superiority of the method proposed in this paper.

Brian W. Matthews[6] focused on voice analysis of Parkinson's patients and explored DL and ML algorithms for PD identification, with a CNN network model achieving a 99.34% accuracy. Handwriting and Brain imaging were also analyzed which achieved an accuracy of 84% and 97% respectively with the help of the CNN network model. Deep learning methods yielded 84.7% accuracy in multiclass classification, emphasizing the growing role of AI in early PD identification.

A comparison analysis was presented by Javed Mehedi Shamrat et al. [7] in order to determine the total performance indicators for f1-score, recall, and precision. Random Forest (RF) models offer greater accuracy than KNN models. The use of the RF model, which can handle big datasets, improves the prediction of Parkinson's disease. Compared to other machine learning algorithms, Random Forest requires less training time and achieves a 93% accuracy rate.

C. Diabetes
The study by Radhika Paturi et al[8] delved into diabetes prediction using various machine learning algorithms applied to the Pima Indian Diabetes Dataset. Through meticulous data preprocessing, the team leveraged algorithms such as Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, Gradient Boosting, and ensemble techniques. Impressively, the Logistic Regression algorithm achieved the highest accuracy score of 97%. SVM also showed promising results, reaching an accuracy of 78%. This comprehensive approach demonstrated the potential of machine learning in accurately predicting diabetes risk.

The paper by Kalaiselvi et al[9] focuses on predicting diabetes in PIMA Indian women using machine learning techniques. It utilizes Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) algorithms to analyze clinical data. The SVM classifier achieved an impressive accuracy of 85.6% in predicting diabetes based on the dataset attributes such as pregnancies, diabetes pedigree function, skin thickness, blood pressure, glucose, and age. The study also emphasizes the importance of ML in diagnosing and predicting diabetes accurately. The findings highlight SVM as a superior algorithm for diabetes prediction in this context due to its high accuracy compared to other classifiers used in the study. Rakesh S Raj et al[10] explored diabetes prediction using electronic health records, employing Naive Bayes and Support Vector Machine (SVM) algorithms. They focused on comparing accuracy in predicting diabetes from patient health data. Detailing the risks associated with high glucose levels and the impact of prolonged elevation, the research underscores the significance of leveraging data mining for healthcare insights. The results showcase SVM achieving an accuracy of 82% in predicting diabetes from the dataset, while Naive Bayes scored 62.5%. Ultimately, the study concludes that SVM demonstrates higher efficiency in diabetes prediction compared to Naive Bayes based on the analyzed sample datasets.

The study by E Ramanujam et al[11] explores diabetes prevalence in India, emphasizing the risk among rural populations due to migration and urbanization. It suggests a multilingual decision support system that incorporates a number of machine learning techniques, including J48 (Decision Tree), k-NN, Random Forest, Logistic Regression, and Naive Bayes. Among these algorithms, k-NN achieves the highest accuracy of 98.08% through 10-fold cross-validation and 97.85% mean accuracy over 30 iterations. This system aims to provide cost-effective early diagnosis for diabetes, utilizing a dataset containing symptoms of 520 individuals and user-friendly interfaces in multiple languages to aid rural healthcare.

D. Liver Disease

The research project by R T Umbare et al [12] aims to address the challenge of early detection of liver disease using machine learning. Recognizing the critical role of the liver and the difficulty in early diagnosis, the study proposes a classification approach to distinguish between healthy and diseased individuals, further categorizing the type and level of liver disease. The proposed system anticipates outcomes such as improved disease classification, precautions for symptoms, and a reduction in liver disease-related deaths. The paper concludes by emphasizing the significance of a hybrid classifier for accurate liver disease diagnosis.

In their paper, Dr. Chokka Anuradha et al [13] emphasize the challenge of early liver disease detection and propose a framework using combined machine learning models. Three optimized algorithms - Artificial Neural Networks (ANN), Decision Trees, and K-Nearest Neighbors (KNN) - are employed on the Indian Liver Patient Dataset (ILPD). The combined model, utilizing a majority voting system, outperforms individual models with a maximum accuracy of 96%. This study underscores the effectiveness of machine learning in predicting liver diseases, crucial for early intervention and prevention.

Ketan Gupta, Nasmin Jiwani, Neda Afreen, and Divyarani D focus [14]on predicting liver diseases using machine learning algorithms, employing the Indian Liver Patient Dataset (ILPD). Various algorithms, including logistic regression, decision tree, random forest, k-nearest neighbors, gradient boosting, extreme gradient boosting, and others, are applied and compared for their accuracy. Results suggest that Random Forest, LightGB, and AdaBoost algorithms exhibit better accuracy, with LightGB identified as particularly suitable for liver disease prediction.

Vasan Durai, Suyan Ramesh, and Dinesh Kalthireddy [15] predict liver diseases using machine learning techniques, particularly the J48 algorithm, Naïve Bayes, and Support Vector Machine (SVM). Their study involves the analysis of patient datasets, including the application of the min-max algorithm, Particle Swarm Optimization for feature selection, and the evaluation of different classification algorithms. The J48 algorithm is highlighted for its superior performance with a 95.04% accuracy rate.

E. Breast Cancer

Kobayashi, Harshwardhan Kharpate, and others [16] used SVM, CNN, and KNN algorithms. They separated the task into three stages. The first is to use the patient's MRI report to make a cancer prognosis. The precise location of the tumour or cancer in the MRI is detected in the second phase, and analysis is the final step. SVM and CNN were used as the main algorithms, KNN was used to determine the accuracy of the outcome, and the PyCharm library's TKinter() was used for front-end design. 98% accuracy was the maximum given by SVM.

Aryan Mittal et al used machine learning algorithms like KNN, SVM, and Naïve Bayes for early breast cancer diagnosis and risk prediction with SVM giving the highest accuracy of approx. 90%.

Avneesh Atrey et al performed data classification by implementing different kernels of SVM and measured the classification test accuracy, sensitivity, and specificity values on WDBC dataset. The result depicted that the RBF kernel provides more prominent results by achieving an accuracy of 93.9%.

III. METHODOLOGY

A. Dataset:

The datasets are taken from the Kaggle website and preprocessed to make the data compatible with machine learning algorithms to learn. This step involves cleaning up data and filling in missing values. Once the data is prepared, it is split into training and test sets for implementation and performance evaluation. The final trained model is fed with new data to predict if a person has a disease or not.

No	Feature	Description
1	Age	The patient's age
2	Sex	Patient's gender
3	Chest pain type	Atypical Angina (ATA), Non-Anginal Pain (NAP), Typical Angina (TA), and Asymptomatic Angina (ASY) are the different forms of chest pain.
4	Resting BP	blood pressure at rest [mm Hg]
5	Cholesterol	blood cholesterol level (mm/dl)
6	Fasting BS	blood sugar when fasting [0: otherwise, 1: if Fasting BS > 120 mg/dl]
7	Resting ECG	outcomes of a resting electrocardiogram

8	Max HR	maximal heart rate attained [numerical value in the range of 60–202]
9	Exercise Angina	Angina brought on by exercise
10	Old peak	symbolizes the ST segment's depression brought on by exertion as comparison to rest.
10	ST slope	symbolizes the ST segment's slope during a workout.

Table 1. Heart Disease Dataset

No	Feature	Description
1	Radius	The typical separations between the tumour 's centre and its periphery.
2	Texture	The mean gray-scale values of the pixels within the digital image of the tumor.
3	Perimeter	the perimeter of the tumor's typical size.
4	Area	The area of the tumor at its average size.
5	Smoothness	the typical range of the radius lengths.
6	Compactness	The area is split by twice the average of the squared perimeter.
7	Concavity	the contour's concave sections' average severity.
8	Symmetry	The tumor's average symmetry.
9	Fractal Dimension	a measurement of the tumor's spatial fill-in.

Table 2. Breast Cancer Dataset

No	Feature	Description
1	Age	Patients' ages
2	Gender	Patients' gender
3	Bilirubin level	Total amount of bilirubin, a pigment that becomes yellow when red blood cells break down, in the blood.
4	Alkaline phosphatase	There is an enzyme in the liver, bile ducts, and bones, among other tissues.
5	Alamine aminotransferase(ALT)	A liver-based enzyme is called ALT. Increased values

		could be a sign of inflammation or liver damage.
6	Aspartate aminotransferase(AST)	Another enzyme that is present in the liver and other organs is called AST. Although elevated levels can be a sign of liver disease, the heart and muscles can also contain AST.
7	Total proteins	Total protein content of blood, comprising globulins and albumin
8	Albumin	The liver produces a protein called albumin. It is essential for preserving blood pressure and volume.
9	Albumin and globulin ratio	The albumin level is divided by the total globulin level to get this ratio.

Table 3. Fatty Liver Dataset

| No	Feature	Description
1	Pregnancies	number of pregnancies
2	Glucose	plasma glucose levels two hours after an oral glucose tolerance test
3	Blood Pressure	Blood pressure diastolic (mm Hg)
4	Skin Thickness	Skinfold thickness (mm) of the triceps
5	Insulin	2-Serum insulin (mu U/ml) at 2 hours
6	Diabetes Pedigree Function	indicates a genetic predisposition for diabetes
7	Age	Years of age
8	BMI	Body mass index (kg of weight divided by m of height)^2)

Table 4. Diabetes Disease Dataset

No	Feature	Description
1	Name	Recording number and topic name in ASCII.

2	MDVP : Fo(HZ)	Vocal fundamental frequency average.
3	MDVP : Fhi(HZ)	maximal fundamental frequency of voice.
4	MDVP : Flo(HZ)	Vocal fundamental at minimum.
5	Jitter: DDPR, MDVP: PPQ, MDVP: RAP, MDVP: jitter(%), and MDVP: jitter(ABS)	Multiple measurements of the fundamental frequency variance.
6	MDVP: APQ3, APQ3, MDVP: Shimmer(DB), APQ3, APQ3, APQ3.	Multiple amplitude variation measurements
7	NHR, HNR	There are two measurements for the voice status's ratio of noise to tone components.
8	DFA	Scaling exponents for signal fractals: spread1, spread2.
9	PPE	The basic frequency fluctuation frequency is measured using three nonlinear ways.
10	RPDE ,D2	Two measurements of nonlinear dynamical complexity.

Table 5. Parkinson's Diseases Dataset

B. Implementation:
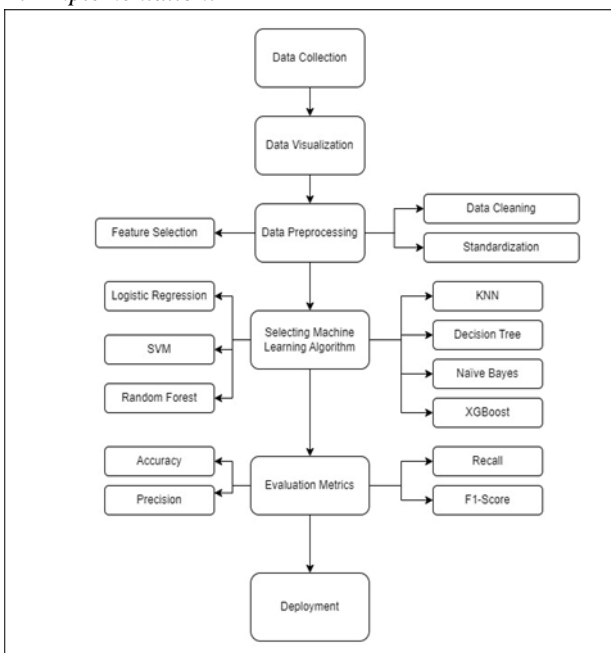
C. Algorithms:

a) SVM: The Support Vector Machine (SVM) algorithm is a supervised learning method that uses machine learning to handle problems with outlier identification, regression, and classification. It works by using effective data transformations to create data point boundaries based on predefined classes, labels, or outputs. Finding the optimal hyperplane in an N-dimensional space to divide data points into discrete classes in the feature space is the main goal of the SVM technique. The hyperplane's purpose is to minimize the distance between the nearest points in each class. The hyperplane's size is determined by the number of features. The hyperplane is only a line if there are only two input features. The hyperplane transforms into a two-dimensional plane when three input features are present.

b) KNN: The KNN method is a popular and adaptive machine-learning technique that is widely utilized due to its ease of implementation. It is also an adaptable approach for a variety of dataset types in classification and regression applications due to its ability to handle both numerical and categorical data. It is a non-parametric forecasting technique that uses the degree of similarity between data points in a particular dataset. The KNN method calculates the K nearest neighbours to a given data point using a distance measure such as Euclidean distance. The majority vote or average of the K neighbours is then utilized to determine the data point's class or value. KNN has applications in recommendation engines, pattern recognition, and data preparation.

c) Logistic Regression: In logistic regression, a supervised machine learning approach, a logistic function—also known as a sigmoid function—is used to produce a probability value between 0 and 1 based on inputs that are independent variables. It is used to predict the categorical dependent variable given a set of independent factors. Rather than fitting a regression line (0 or 1), logistic regression fits a "S" shaped logistic function, which predicts two maximum values. The likelihood of something is displayed on the logistic function's curve.

d) Decision Tree: A decision tree is a diagram that shows a set of decisions and possible outcomes. It is designed to resemble a tree. Each node in the tree represents a decision, and each branch represents the decision's outcome. The tree's leaves represent final decisions or projections. Decision trees are produced by recursively splitting the data into progressively smaller subsets. At each partition, the data is split based on a certain feature, and the division is carried out in a way that maximizes the information gained.

By utilizing the information gain and Gini index, an attribute selected from the dataset is designated as the internal or root node. The Gini Index quantifies the likelihood that a randomly chosen element will be misdiagnosed. Information gain is a measure of this change in entropy.

e) Random Forest: The forecasts from many decision trees are combined using an ensemble learning method known as a random forest to get a single, more accurate forecast. Both regression and classification tasks can benefit from the application of this kind of supervised learning technique. It can handle large, high-dimensional datasets and increases the accuracy of the model to prevent overfitting.

f) Naïve Bayes: For solving classification issues, supervised learning techniques like the Naïve Bayes algorithm are employed. The Bayes theorem forms its foundation. Text categorization utilising a high-dimensional training dataset is its main application case. One of the quickest and easiest classification techniques for developing machine learning models with fast prediction capabilities is the Naïve Bayes classifier. As a probabilistic classifier, it makes predictions based on the likelihood that an object will occur. The Naïve Bayes method is employed in sentiment analysis, spam filtering, and article classification, among other applications.

g) XGBoost: Efficiency and scalability are given priority in the design of XGBoost, a toolset for training machine learning models using distributed gradient boosting. Through the process of ensemble learning, multiple weak models' predictions are combined to produce a more robust forecast.. Extreme Gradient Boosting, or XGBoost, is a machine learning technique that has gained popularity and widespread use due to its capacity to manage massive datasets and provide state-of-the-art performance in a variety of machine learning tasks, such as regression and classification. One of XGBoost's primary characteristics is its effective handling of missing values, which enables it to manage missing values in real-world data with minimal pre-processing. Additionally, XGBoost has parallel processing capabilities which speeds up the process of training models on big datasets.

h) AdaBoost: The machine learning technique for ensembles known as adaptive boosting, or AdaBoost, has numerous applications in classification and regression. In order to categorise data, this supervised learning strategy combines multiple weak or base learners (such as decision trees) to produce a strong learner. The training dataset's instances are weighted based on how well earlier classifications performed in order for AdaBoost to work.

IV. EVALUATION METRICS

1) Accuracy:
The ratio of accurately predicted instances to the total number of instances is used to calculate accuracy, which serves as a gauge for the model's overall correctness.

$$Accuracy = \frac{Number\ of\ Correct\ predictions}{Total\ number\ of\ predictions}$$

2) Precision:
A model's precision informs you how many of its positive predictions come true. Precision is a performance metric. It is computed by dividing the total number of accurate positive forecasts by the total number of inaccurate positive forecasts.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

3) Recall:
A model's recall quantifies its capacity to identify every pertinent instance of a given class. It is the proportion of all real positive instances to all correctly projected positive instances.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

4) F1-Score:
It is the precision and recall harmonic mean. The range of it is [0,1]. This metric frequently shows how stable and accurate our classifier is.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

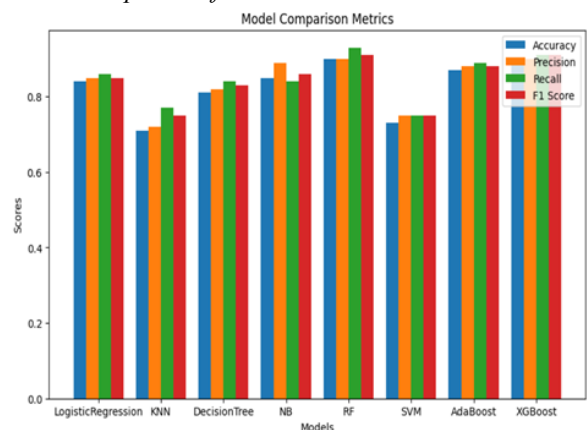*Model Comparison for each disease:*

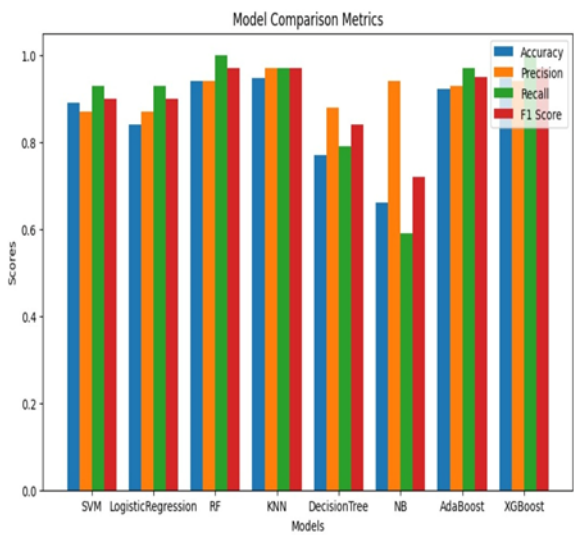Fig 1. Model Comparison for Heart Disease
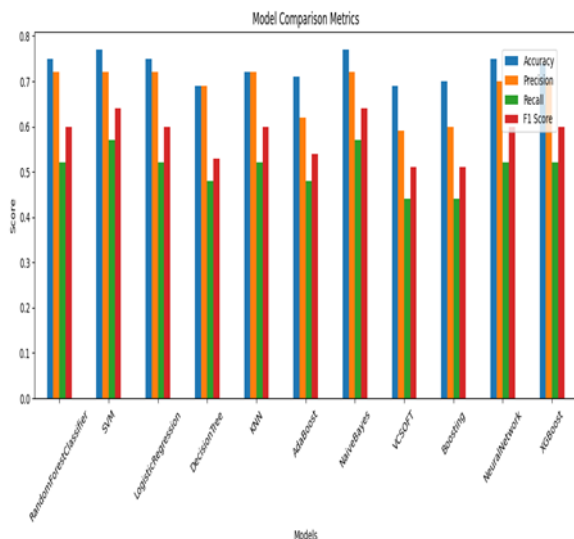
Fig 1.2: Model Comparison for Parkinson's Disease

Fig 1.3: Model Comparison for Diabetes

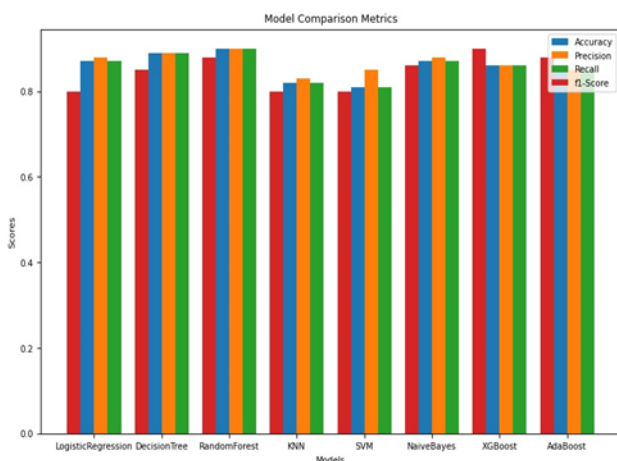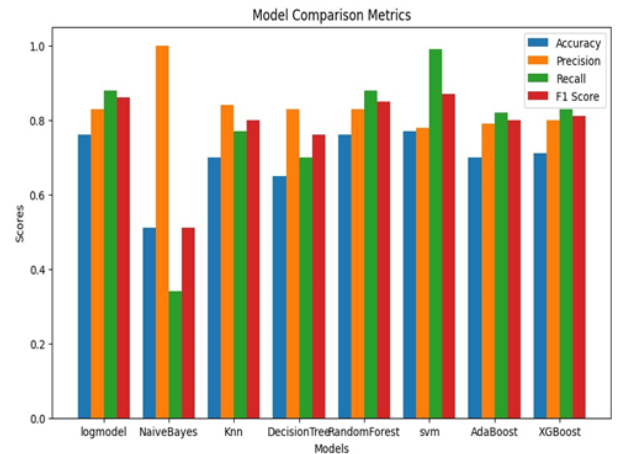Fig 1.4: Model Comparison for Breast Cancer

Fig 1.5: Model Comparison for Liver Disease

No	DISEASE	ALGORITHM	ACCURACY
1.	Heart Disease	XGBoost	0.90
2.	Parkinson's	Random Forest	0.94
3.	Diabetes	SVM	0.77
4.	Fatty Liver Disease	Logistic Regression	0.76
5.	Breast Cancer	Random Forest	0.92

Table 1. Algorithm With Most Accuracy

V. DJANGO FRAMEWORK

Django is a high-level Python web framework that encourages effective development and clear, uncomplicated design. Model-View-Controller (MVC) architecture is followed, and database interactions are handled via an Object-Relational Mapping (ORM) framework. Django is a web development framework that was created with the "batteries-included" mentality. It comes with many pre-built capabilities for basic jobs including templating, URL routing, and authentication. It's a great option for creating dependable and maintainable web apps because of its simplicity and adherence to the DRY (Don't Repeat Yourself) principle. Django is widely used by developers to create effective online applications because of its emphasis on security procedures, scalability, and a thriving community.

All disease prediction models are deployed using the Django framework. Django framework helps in routing the HTML file which makes the task easier for getting the disease prediction results.
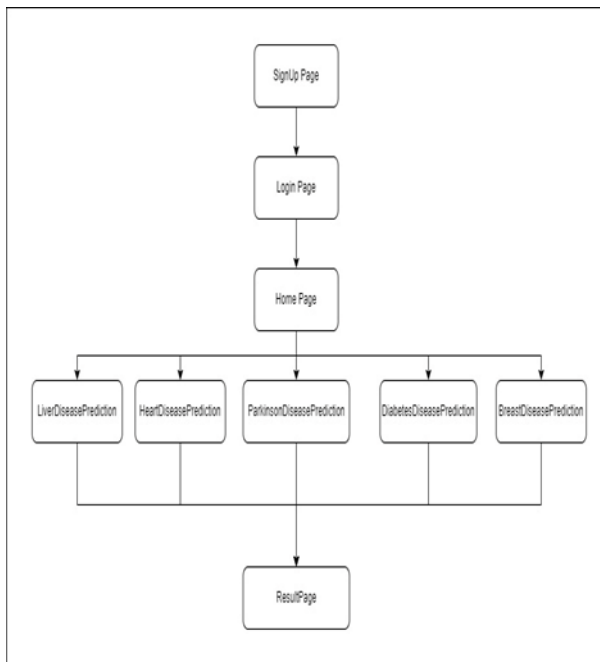
Fig 1. WebApp Flow

CONCLUSION

The creation of hybrid models of each disease is emphasized in this survey paper, along with performance analysis, using machine learning algorithms like Support Vector Machine (SVM), K-Nearest Neighbours (KNN), Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, XGBoost, and AdaBoost. The algorithmic performances varied by disease category; for example, Random Forest was shown to perform well for Parkinson's and heart diseases, while Decision Tree was shown to perform well for liver disease. SVM continuously demonstrated a greater prediction accuracy rate for breast cancer. The evaluation metrics show how well each model predicts the future and reduces false positives. The survey done in machine learning models for multi-disease prediction shows a commitment to continuously improve predictive accuracy in the medical sector. The convergence of machine learning and healthcare holds immense potential for improving early disease detection. A Django-based web app for early disease detection offers a robust and efficient solution. Leveraging Django's security, scalability, and clean design, the app provides a reliable platform for timely health assessments, demonstrating the framework's versatility in critical applications.

ACKNOWLEDGEMENT

REFERENCES

[1] P. Divyasri, D. Sreelakshmi, P. Sathvika, P. Teja, and T. V. Charan, "Cardiovascular Disease Prediction Using Machine Learning," in *2023 6th International Conference on Information Systems and Computer Networks, ISCON 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ISCON57294.2023.10112052.

[2] T. M. Ghazal, A. Ibrahim, A. S. Akram, Z. H. Qaisar, S. Munir, and S. Islam, "Heart Disease Prediction Using Machine Learning," in *2nd International Conference on Business Analytics for Technology and Security, ICBATS 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICBATS57792.2023.10111368.

[3] S. Ibrahim, N. Salhab, and A. El Falou, "Heart disease Prediction using Machine Learning," in *1st International Conference in Advanced Innovation on Smart City, ICAISC 2023 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICAISC56366.2023.10085522.

[4] V. Sharma, S. Yadav, and M. Gupta, "Heart Disease Prediction using Machine Learning Techniques," in *Proceedings - IEEE 2020 2nd International Conference on Advances in Computing, Communication Control and Networking, ICACCCN 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 177–181. doi: 10.1109/ICACCCN51052.2020.9362842.

[5] Z. Fang, "Improved KNN algorithm with information entropy for the diagnosis of Parkinson's disease," in *Proceedings - 2022 International Conference on Machine Learning and Knowledge Engineering, MLKE 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 98–101. doi: 10.1109/MLKE55170.2022.00024.

[6] A. R. Palakayala and P. Kuppusamy, "Survey of Parkinson's Disease Detection using Different Symptoms," in *Proceedings - 2022 Algorithms, Computing and Mathematics Conference, ACM 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 77–82. doi: 10.1109/ACM57404.2022.00020.

[7] N. Alapati, N. Anusha, P. Joharika, N. J. Jerusha, and P. Tanuja, "Prediction of Parkinson's Disease

using Machine Learning," in *Proceedings of the 2023 2nd International Conference on Electronics and Renewable Systems, ICEARS 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 1357–1361. doi: 10.1109/ICEARS56392.2023.10085443.

[8] R. Paturi, D. Sreeja, P. H. Bindhu, B. Ramya, and K. Harika, "Diabetic Prediction System using Machine Learning Model," in *Proceedings - International Conference on Applied Artificial Intelligence and Computing, ICAAIC 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 515–522. doi: 10.1109/ICAAIC53929.2022.9793290.

[9] V. K. G. Kalaiselvi, H. Shanmugasundaram, E. Aishwarya, M. Ragavi, C. Nandhini, and S. J. Bhuvaneshwari, "Analysis of Pima Indian Diabetes Using KNN Classifier and Support Vector Machine Technique," in *Proceedings of the 2022 3rd International Conference on Intelligent Computing, Instrumentation and Control Technologies: Computational Intelligence for Smart Systems, ICICICT 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 1376–1380. doi: 10.1109/ICICICT54557.2022.9917992.

[10] Dayananda Sagar Academy of Technology & Management, Institute of Electrical and Electronics Engineers. Bangalore Section., and Institute of Electrical and Electronics Engineers, *First International Conference on Advanced Technologies in Intelligent Control, Environment, Computing and Communication Engineering (ICATIECE-2019) : 19th-20th March, 2019 : Dayananda Sagar Academy of Technology & Management.*

[11] E. Ramanujam, T. Chandrakumar, K. T. Thivyadharsine, and D. Varsha, "A Multilingual Decision Support System for early detection of Diabetes using Machine Learning approach: Case study for Rural Indian people," in *Proceedings - 2020 5th International Conference on Research in Computational Intelligence and Communication Networks, ICRCICN 2020*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020, pp. 17–21. doi: 10.1109/ICRCICN50933.2020.9296187.

[12] R. T. Umbare, O. Ashtekar, A. Nikhal, B. Pagar, and O. Zare, "Prediction and Detection of Liver Diseases using Machine Learning," in *3rd IEEE International Conference on Technology, Engineering, Management for Societal Impact using Marketing, Entrepreneurship and Talent, TEMSMET 2023*, Institute of Electrical and

Electronics Engineers Inc., 2023. doi: 10.1109/TEMSMET56707.2023.10150135.

[13] C. Anuradha, D. Swapna, B. Thati, V. N. Sree, and S. P. Praveen, "Diagnosing for Liver Disease Prediction in Patients using Combined Machine Learning Models," in *Proceedings - 4th International Conference on Smart Systems and Inventive Technology, ICSSIT 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 889–896. doi: 10.1109/ICSSIT53264.2022.9716312.

[14] K. Gupta, N. Jiwani, and N. Afreen, "1 th IEEE International Conference on Communication Systems and Network Technologies Liver Disease Prediction using Machine learning Classification Techniques", doi: 10.1109/csnt.2022.40.

[15] G. Tomar, Machine Intelligence Research Labs, K. Pranveer Singh Institute of Technology, and IEEE Computer Society, *Proceedings, 2022 IEEE 11th International Conference on Communication Systems and Network Technologies : CSNT 2022.*

[16] A. Chauhan, H. Kharpate, Y. Narekar, S. Gulhane, T. Virulkar, and Y. Hedau, "Breast Cancer Detection and Prediction using Machine Learning," in *Proceedings of the 3rd International Conference on Inventive Research in Computing Applications, ICIRCA 2021*, Institute of Electrical and Electronics Engineers Inc., Sep. 2021, pp. 1135–1143. doi: 10.1109/ICIRCA51532.2021.9544687.