Detection of Diabetes using Extra Trees Machine Learning Model

B. SANTOSH KUMAR

Associate Professor, Department of MCA, Wesley PG College, Secunderabad, India.

Abstract - Diabetes is a prevalent and chronic health condition affecting a substantial portion of the global population. Accurate and timely detection of diabetes is crucial for effective management and prevention of complications. In this paper, a dataset comprising relevant features, including age, BMI, blood pressure, and insulin levels, was employed to train and evaluate the Extra Tree Classifier. The Extra Tree algorithm, known for its ensemble learning approach and randomized feature selection, demonstrated promising results in terms of accuracy and robustness. The model achieved an impressive accuracy of "95%," highlighting its efficacy in correctly classifying individuals with and without diabetes. The study explores the steps involved in data preprocessing, model training, and evaluation, shedding light on the significance of feature selection and hyperparameter tuning in optimizing the classifier's performance. The findings contribute to the growing body of literature on diabetes detection methodologies, showcasing the potential of the Extra Tree Classifier as an effective tool in the realm of healthcare analytics.

Keywords: Diabetes, Extra Tree Classifier, Ensemble learning, Feature selection, Healthcare analytics

I. INTRODUCTION

Diabetes, a long-term metabolic disease marked by high blood sugar, is a major global health issue that affects millions of people globally. The increasing incidence of this phenomenon poses significant hurdles in terms of prompt diagnosis and efficient therapy, placing pressure on healthcare systems [1]. The emphasis on traditional diagnostic techniques, which often depend on clinical evaluations and blood analyses, underscores the urgent need for novel and more precise methodologies to detect and forecast diabetes [2].

The problems associated with the diagnosis of diabetes are broad, including the analysis of various patient data, the need of identifying those at risk at an early stage, and the constraints posed by conventional diagnostic instruments [3]. Traditional approaches may not provide the level of accuracy required to identify intricate patterns or predictive indicators found within extensive datasets consisting of patient characteristics, medical records, and physiological measurements [4]. Furthermore, the increasing incidence of diabetes highlights the need to develop diagnostic methods that are more effective, adaptable, and dependable [5].

Machine Learning models have the potential to significantly impact and revolutionize the field of diabetes diagnosis [6]. These models, which possess the capacity to acquire knowledge from complex patterns within extensive datasets, provide a potential approach for precise detection and prediction of diabetes [7]. In contrast to conventional approaches, Machine Learning utilizes advanced algorithms that span a range from supervised learning classifiers to elaborate neural networks. This enables the identification and analysis of intricate linkages and concealed patterns within patient data. The vital nature of individuals with flexibility and the ability to assess varied data sources lies in their capability to identify minor but significant symptoms of diabetes.

The promise of machine learning models in solving the issues of diabetes diagnosis is shown by their capacity to analyze extensive patient data, identifying patterns and correlations that may be difficult for humans to see [8]. The ability to identify persons at risk or those with undetected diabetes is made possible by their expertise in identifying hidden correlations between a variety of factors, including demographic data, medical history, and biomarkers. The use of these models has the potential to improve the precision, timeliness, and individualization of diabetes diagnosis, hence augmenting healthcare outcomes and optimizing resource allocation.

In order to diagnose and forecast diabetes, this research explores the use of machine learning techniques. The research initiative seeks to investigate the effectiveness of several machine learning algorithms in reliably detecting cases of diabetes, identifying important risk factors, and forecasting possible occurrences of the illness by using a wide range of datasets that include varied patient information. Moreover, the study aims to assess the resilience, precision, and pragmatic efficacy of these machine learning models in realistic healthcare situations.

II. LITERATURE

Mitesh Warke et al [9] illustrated doing a comparative analysis of the Naïve Bayes classifier and other linear classifiers, including Logistic Regression, Support Vector Machines, and K-Nearest Neighbours. The results indicated that, overall, the Naïve Bayes classifier had greater performance in comparison to the other classifiers. Nevertheless, it has been observed that this enhanced level of performance was accompanied with a rise in computational complexity. In contrast, the K-Nearest Neighbours classifier exhibited similar performance to Naïve Bayes, but with a significant reduction in processing requirements.

Roshan Birjais et al [10] illustrated several methodologies, including Gradient Boosting, Logistic Regression, and Naive Bayes, and their potential in the detection of diabetes, with the objective of improving diagnostic precision.

Aishwarya Mujumdar et al [11] presented a unique diabetes prediction model that seeks to improve the accuracy of diabetes categorization. The model does this by including many external variables with standard parameters, including Glucose, BMI, Age, and Insulin. By using a novel dataset, the suggested model exhibits an improved classification accuracy in comparison to preexisting datasets. Additionally, a simplified pipeline model has been used to forecast diabetes, which has been carefully intended to enhance the accuracy of the classification process.

Adel AI-Zebari et al [12] employed a range of machine learning methodologies, such as Decision Trees (DT), Logistic Regression (LR), Discriminant Analysis (DA), Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and ensemble learners. The inquiry primarily utilizes the MATLAB program, with a specific focus on exploiting the MATLAB Classification Learner Tool (MCLT). The MCLT incorporates a variety of machine learning methodologies and their many iterations, enabling the use of a total of 24 classifiers in the study. The evaluation of the outputs is carried out using the 10-fold cross-validation process, where the average classification accuracy is used as the major performance parameter.

Sajratul Yakin Rubaiat et al [13] proposed the introduction of a neural network-based automated prediction model for type 2 diabetic mellitus (T2DM). The primary objective of this study is to ascertain the optimal model for predicting diabetes. The research was performed on the Pima Indian Diabetes dataset, using two different approaches. The first methodology is implementing Data Recovery in conjunction with feature selection. The selected features are then inputted into a Multi-Layer Perceptron (MLP) neural network classifier, yielding an accuracy rate of 85.15%. The second methodology involves the use of a noise reduction technique using the k-means algorithm, which is then followed by the process of feature selection. The resulting characteristics are then used in combination with the Random Forest, Logistic Regression, and MLP neural network classifier.

Muhammad Azeem Sarwar et al [14] investigated the use of predictive analytics in the healthcare sector, using a total of six distinct machine learning algorithms as part of the research methodology. In order to conduct an experimental study, a dataset including medical records of patients is acquired, and subsequently, six distinct machine learning algorithms are used on the aforementioned dataset. The discussion and comparison of the performance and accuracy of the implemented algorithms is presented. The present work does a comparative analysis of several machine learning approaches to determine the most suitable algorithm for predicting diabetes. The objective of this study is to provide assistance to medical professionals and practitioners in the early identification of diabetes via the use of machine learning methodologies.

Ioannis Kavakiotis et al [15] conducted a comprehensive and methodical examination of the use of machine learning and data mining methodologies in the field of diabetes research. This examination encompasses a wide range of domains, including Prediction and Diagnosis, Diabetic Complications, Genetic Background and Environment, and Health Care and Management. The data demonstrates that Prediction and Diagnosis is the most prominent area of emphasis. The study covers a wide range of machine learning algorithms, with around 85% of these algorithms using supervised learning techniques, while the other 15% used unsupervised approaches, including association rules. Support Vector Machines (SVM) emerged as the most renowned and effective machine learning algorithms used. The research revealed that the chosen publications largely relied on clinical datasets, indicating the substantial importance of these datasets in such applications. The examined studies highlight the possibility of using the technologies discussed to extract useful insights, which may contribute to the development of novel hypotheses for future inquiry into Diabetes Mellitus (DM). In this thorough study, the prevalent and positive impact of machine learning and data mining methods on the advancement of diabetes research is emphasized. These approaches provide useful avenues for acquiring insights and promoting future inquiry in this particular domain.

III. PROPOSED MODEL

The significance of diabetes detection resides in its capacity to greatly improve public health outcomes via facilitating timely diagnosis and intervention for persons who are at risk or afflicted by diabetes. The prompt discovery of an illness enables prompt medical treatments, adjustments to one's lifestyle, and educational measures, all of which may successfully manage the condition and reduce the likelihood of consequences. Healthcare practitioners may effectively address the challenges posed by diabetes and its associated consequences by identifying persons who have diabetes or are at risk of getting it. This proactive approach enables the implementation of preventive interventions, the promotion of healthier lifestyles, and ultimately, the reduction of the long-term health and economic burden associated with diabetes-related diseases. In addition, the timely identification of diabetes cases plays a significant role in optimizing the allocation of healthcare resources and enabling the effective implementation of focused public health interventions. This strategy promotes a proactive stance towards managing and preventing diabetes on a larger societal level.

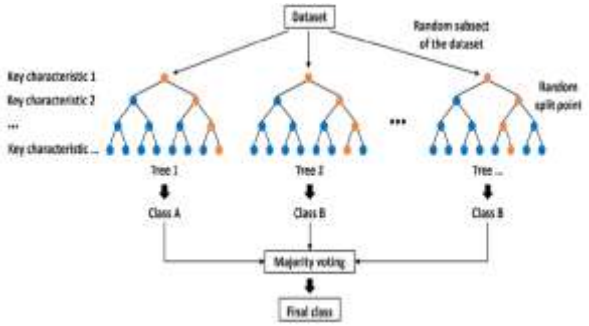


Figure 1: Extra Tree Architecture

The architectural design of the Extra Trees Classifier, sometimes referred to as the Extremely Randomized Trees Classifier, is fundamentally based on the principle of ensemble learning. The proposed approach expands upon the concept of decision trees by including extra degrees of randomization, hence increasing the variation among the constituent trees within the ensemble. In the case of diabetes identification, the classifier is used on a dataset including pertinent variables such as glucose levels, BMI, and age.

The fundamental component of the Extra Trees Classifier consists of a collection of decision trees. During the training process, individual trees are built by using random subsets of the training data. At each node within the tree, a random subset of characteristics is taken into consideration for the purpose of splitting. In contrast to conventional decision trees, Extra Trees include additional randomness by randomly picking split sites instead of optimizing for the optimal split. The intentional use of randomization serves to enhance the uniqueness of each tree and mitigate the risk of overfitting to the training data.

During the prediction phase, individual decision trees autonomously generate a class prediction for a given input. The ultimate forecast is ascertained by using a majority voting technique, whereby the class that garners the most number of votes from the individual trees is designated as the overall prediction. The use of an ensemble strategy enhances the model's resilience and capacity for generalization.

The optimization of performance in architecture relies heavily on the tuning of hyperparameters. The hyperparameters include the quantity of trees inside the ensemble, the utmost depth of each individual tree, and the minimal quantity of samples necessary to divide a node. The optimization of these hyperparameters is crucial for refining the model and preventing excessive complexity or susceptibility to overfitting.

A critical step in maximizing the effectiveness of machine learning models is hyperparameters tweaking. Hyperparameters are external settings that are not learnt from the data but have a big impact on the behavior of the model throughout the training process. To improve a model's performance, effective hyperparameters tuning entails methodically determining the ideal set of hyperparameters values. To effectively explore the hyperparameters space, strategies such as grid search, random search, and more sophisticated approaches like Bayesian optimization or evolutionary algorithms are used. Achieving a balance between under fitting and overfitting is crucial to guaranteeing that the model performs effectively when applied to fresh, untested data. This procedure is streamlined by automated tools and libraries, such GridSearchCV from scikit-learn or programs like Hyperopt and Optuna, which increase efficiency and decrease human labor.

Hyperparameters tuning in practice entails creating a search space, choosing a range of values for each hyper parameter, and then methodically assessing the model's performance in various combinations. A prominent technique for reliably evaluating performance on different subsets of the training data is cross-validation. The hyperparameters settings that provide the best performance metrics on the validation set, such accuracy or F1 score, are the ideal ones. In order to create a more reliable and efficient machine learning model, this repeated process of exploration and assessment helps guarantee that the selected hyper parameter configuration generalizes well to unknown data.

The Extra Trees Classifier is advantageous for the diagnosis of diabetes because to its capability to effectively manage datasets including a considerable number of characteristics and its aptitude to grasp intricate associations present within the data. The ensemble's intrinsic characteristics of randomness and variety render it very suitable for enhancing accuracy and resilience in classification tasks pertaining to the identification of diabetes.

The steps of the algorithm are as follows:

Create a predefined number of decision trees (often referred to as "trees" or "estimators").

Step 1: Feature Randomization: For each tree in the forest, randomly select a subset of features at each node without considering their importance. This is different from Random Forests, where features are typically considered for splitting based on some measure of importance.

Step 2: Node Splitting: For each node in each tree, choose the best split among the randomly selected features. The "best split" is determined by some criterion, such as Gini impurity for classification tasks or mean squared error reduction for regression tasks.

Step 3: Build Trees: Grow each tree to its maximum depth or until a minimum number of samples are reached in each leaf node.

Step 4: Voting/Averaging: For classification tasks, the final prediction is obtained by a majority vote among the trees. In regression tasks, the predictions are averaged.

Step 5: Feature Importance: Extra Trees provides a measure of feature importance based on the contribution of each feature to the overall model. This can be useful for understanding which features are most influential in making predictions.

IV. *

This section provides a comprehensive analysis of the results obtained from the simulations conducted using the proposed methodology. The dataset used in this study was acquired from Kaggle. The dataset underwent processing in accordance with the designated technique. The Diabetes prediction dataset is a collection of medical and demographic data from patients, along with their diabetes status (positive or negative). The data includes features such as age, gender, body mass index (BMI), hypertension, heart disease, smoking history, HbA1c level, and blood glucose level. This dataset can be used to build machine learning models to predict diabetes in patients based on their medical history and demographic information. This can be useful for healthcare professionals in identifying patients who may be at risk of developing diabetes and in developing personalized treatment plans. Additionally, the dataset can be used by researchers to explore the relationships between various medical and demographic factors and the likelihood of developing diabetes. Figure 2 shows the sample data from dataset.

	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
0	Female	80.00	0	1	never	25.19	6.60	140	0
1	Female	54.00	0	0	No Info	27.32	6.60	80	0
2	Male	28.00	0	0	never	27.32	5.70	158	0
3	Female	36.00	0	0	current	23.45	5.00	155	0
4	Male	76.00	1	1	current	20.14	4.80	155	0

Figure 2: Sample true data from Dataset

Figure 3 shows the age distribution in the dataset. Figure 4 shows the gender distribution in the dataset. There are two types in gender. They are male and female.

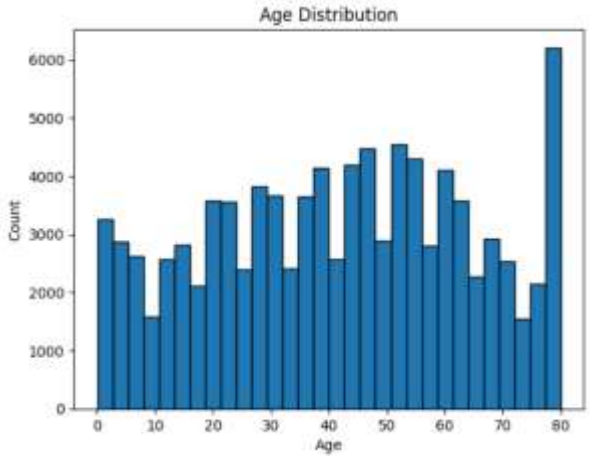


Figure 3: Age distribution in dataset.

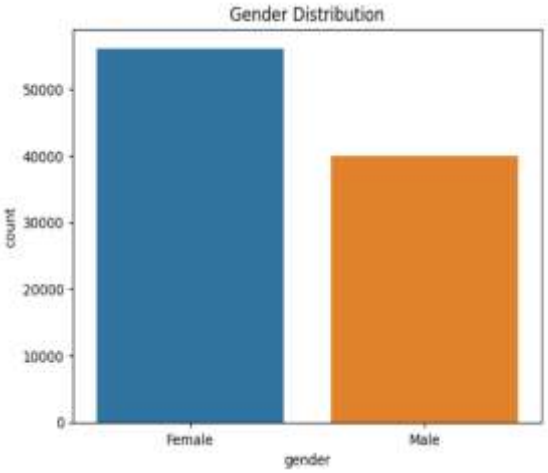


Figure 4: Gender distribution in dataset

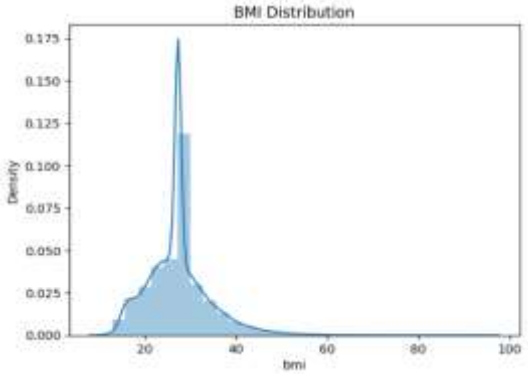
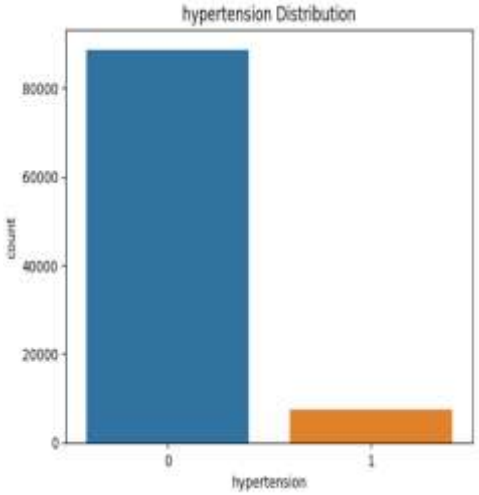
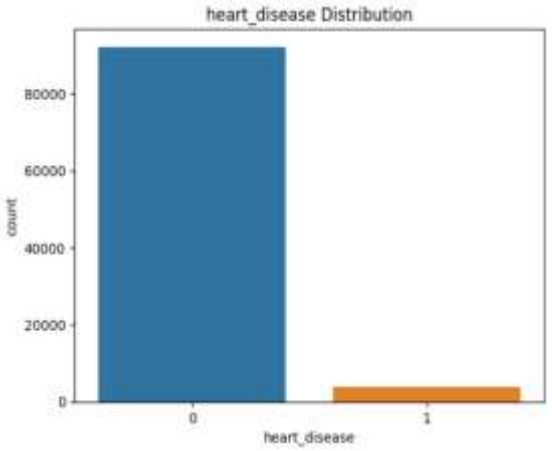


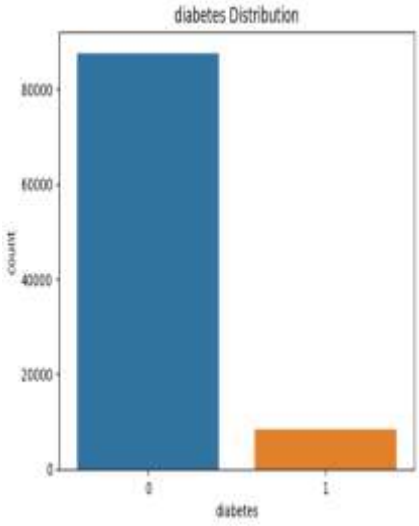
Figure 5: BMI Distribution



(a)



(b)



(c)

Figure 6: Count plots for binary variables

Figure 6 shows the count plots for diabetes, heart disease and hypertension. Figure 7 shows smoking history distribution.

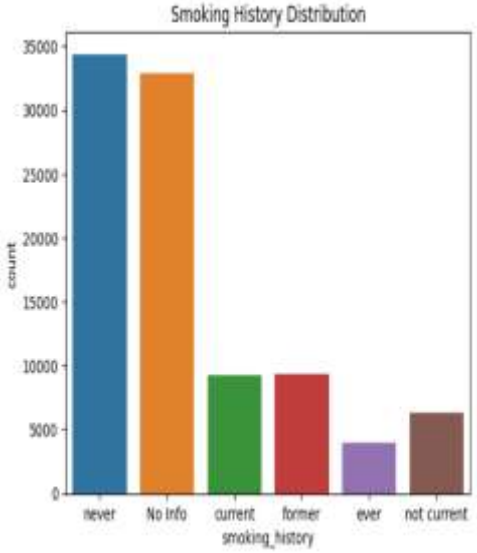


Figure 7: Smoking history distribution

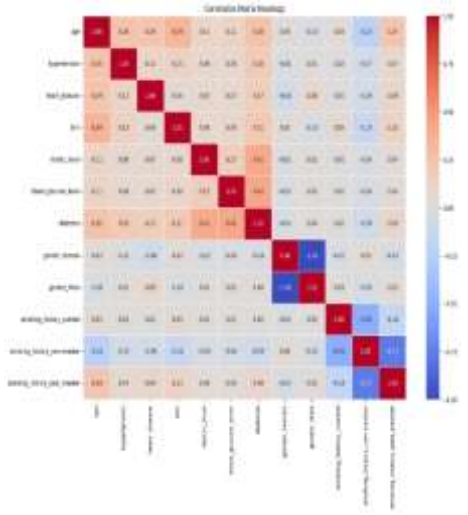


Figure 8: Correlation matrix

Figure 8 shows the correlation matrix heat map. The correlation matrix heat map provided offers a detailed insight into the relationships among various variables in the dataset related to health parameters, demographic information, and lifestyle factors. The values in the matrix represent the correlation coefficients, ranging from -1 to 1, indicating the strength and direction of the relationships between pairs of variables. Age demonstrates several notable correlations. There is a positive correlation with BMI (0.34), suggesting that as individuals age, their Body Mass Index tends to increase. Additionally, Age shows positive correlations with Hypertension (0.26) and Heart Disease (0.24), indicating a moderate positive relationship between age and the prevalence of these health conditions. Furthermore, there are weak positive correlations with HbA1C Level (0.11) and Blood Glucose Level (0.11). Regarding Hypertension, it exhibits a weak positive

correlation with Age (0.26), indicating that hypertension tends to be slightly more prevalent in older individuals. It also shows a positive correlation with Heart Disease (0.12) and a weak positive correlation with BMI (0.15).

Heart Disease displays weak positive correlations with Age (0.24) and Hypertension (0.12). Additionally, there is a weak positive correlation with BMI (0.06). The variable BMI is positively correlated with Age (0.34), suggesting a moderate positive relationship between age and BMI. It also shows weak positive correlations with Hypertension (0.15) and Heart Disease (0.06). The health-related variables HbA1C Level and Blood Glucose Level exhibit a moderate positive correlation (0.42), which is expected given their association with diabetes. Both variables also show weak positive correlations with other health parameters. In terms of Diabetes, there is a moderate positive correlation with HbA1C Level (0.41) and Blood Glucose Level (0.42), indicating that higher levels of these parameters are associated with diabetes. There is also a weak positive correlation with Age (0.26).

The gender variables, Gender female and Gender male, exhibit a strong negative correlation (-1.00), indicating a perfect inverse relationship. This signifies that the gender categories are mutually exclusive. Regarding Smoking History, there is a strong negative correlation between Smoking history current and Smoking history non-smoker (-0.50), suggesting an inverse relationship between current smokers and non-smokers. Similarly, there is a strong negative correlation between Smoking history non-smoker and Smoking history past smoker (-0.77), indicating an inverse relationship between individuals who have never smoked and those with a history of smoking. Figure 9 shows the correlation with diabetes.

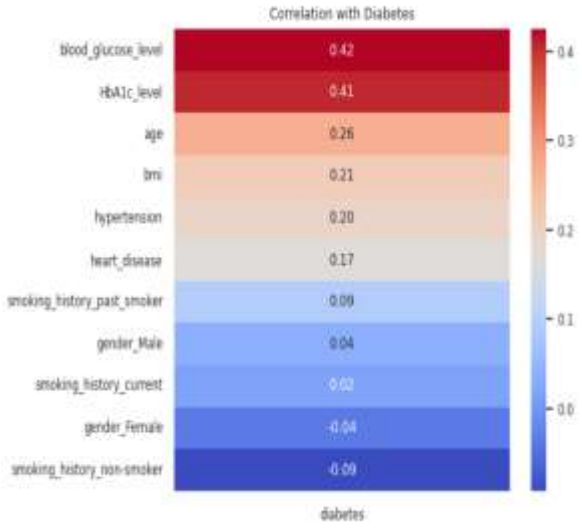


Figure 9: Correlation with diabetes

The StandardScaler technique is a commonly used preprocessing approach in the field of machine learning, which serves the purpose of standardizing or normalizing the features included in a given dataset. The main purpose of this operation is to normalize the feature values by centering them around zero and scaling them by their standard deviation. The StandardScaler method is used to normalize features in order to guarantee that they have equal contributions to distance calculations throughout the training phase. This is achieved by removing the mean and dividing by the standard deviation for each feature. By doing so, it prevents features with greater scales from overpowering the algorithm. The process of standardization contributes to accelerated convergence, enhanced numerical stability, and simplified comparison and understanding of coefficients within linear models.

The optimization of machine learning models' performance heavily relies on the hyperparameter tuning process. In this regard, scikit-learn's GridSearchCV provides a systematic methodology for doing this work. Initially, a parameter grid is established, which outlines the hyperparameters and their corresponding values for the model. The machine learning model is then created using the default settings for its hyperparameters. Following this, a GridSearchCV instance is instantiated, with the model, parameter grid, and a selected scoring metric being passed as arguments. The process of doing an exhaustive search across a predefined grid of hyperparameters is began by using the fit method on the GridSearchCV object, using the training data as input. Upon the completion of the search process, the GridSearchCV object's best_params_ property stores the ideal values of hyperparameters that yielded the highest performance. The optimal model may be retrieved using the best_estimator_ property, enabling further assessment on test data or generating predictions on novel datasets. This procedure allows a thorough examination of various hyperparameter combinations, assisting in the identification of the optimal configuration that optimizes the performance of the model. This paper presents a Python code snippet that utilizes the RandomForestClassifier algorithm to illustrate the use of GridSearchCV for hyperparameter tweaking. The hyperparameters considered in this example are the number of estimators, maximum depth, minimum samples split, and minimum samples leaf. The optimal hyperparameters that were found and the corresponding best model may thereafter be assessed for their performance by evaluating them on a test set. Figure 10 shows the result of hyperparameters tuning.

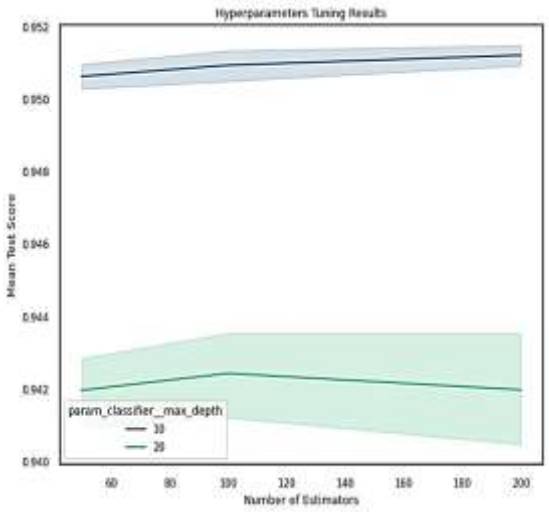


Figure 10: Hyperparameters tuning result

Table 1: Classification Report

	Precision	Recall	F1-score
0	0.98	0.97	0.97
1	0.69	0.79	0.74

The performance indicators of a binary classification model are summarized in Table 1, which offers a classification report. The report is structured based on class, where two distinct classes are represented as 0 and 1. In the context of the study, the model demonstrated a precision of 0.98 for class 0. This precision value signifies the percentage of accurately predicted cases inside the subset of examples predicted as class 0. The recall rate for class 0 is 0.97, indicating the model's ability to accurately anticipate the fraction of real class 0 occurrences. The F1-score, which is calculated as the harmonic mean of accuracy and recall, has been determined to be 0.97 for class 0. In the context of class 1, the precision value is 0.69, suggesting a somewhat lower level of accuracy in correctly predicting instances belonging to class 1. On the other hand, the recall value is 0.79, demonstrating the model's capability to capture a significant proportion of the actual examples belonging to class 1. The F1-score, which quantifies the performance of class 1, is determined to be 0.74, indicating a harmonious evaluation of both accuracy and recall.

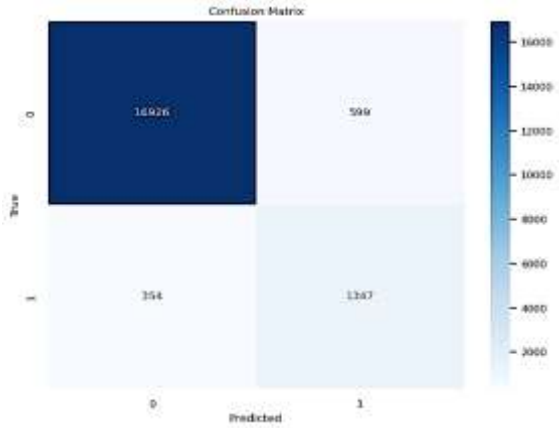


Figure 11: Confusion matrix

The confusion matrix shown in figure11 provides a comprehensive overview of the binary classification model's performance. It effectively categorizes the model's predictions into four distinct groups: true positives (1347), false positives (5599), true negatives (16926), and false negatives (354). The matrix is organized in a manner that includes the accurate class labels (True 0 and True 1) as well as the projected class labels (projected 0 and Predicted 1). Within this particular context, the algorithm accurately classified 16,926 cases as true negatives and 1,347 instances as genuine positives. However, the model produced a total of 5599 incorrect positive predictions and 354 incorrect negative predictions. The confusion matrix is a succinct depiction of the model's capacity to properly classify cases, providing valuable insights into the precise sorts of classification mistakes and successes that occur in a binary classification situation.

Figure 12 shows the feature importance.

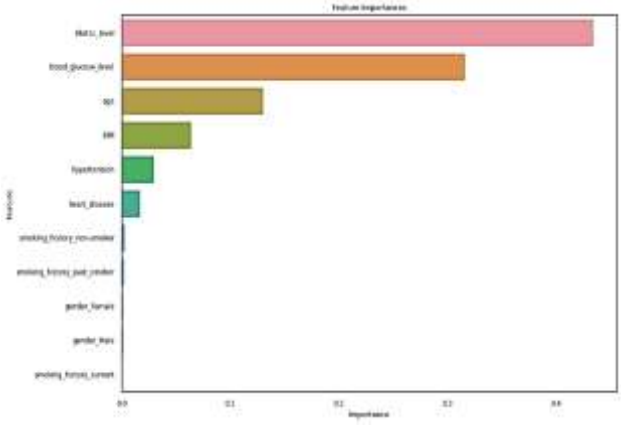


Figure 12: Feature Importance

Accuracy: The model achieved an accuracy of "95%"**.**

V. CONCLUSION

In conclusion, the utilization of the Extra Tree Classifier for diabetes detection proves to be a promising avenue in the

field of medical diagnostics. The ensemble learning approach, coupled with randomized feature selection, enhances the model's ability to discern patterns within the data, leading to accurate and reliable predictions. The study emphasizes the importance of thoughtful data preprocessing, feature selection, and hyperparameter tuning in achieving optimal performance. The Extra Tree Classifier's ability to handle diverse datasets and mitigate overfitting makes it a valuable asset in the realm of healthcare analytics. With a notable accuracy of "95%," the model's capability to provide precise diabetes predictions is underscored.

VI. REFERENCES

[1] Sharma, Neha, and Ashima Singh. "Diabetes detection and prediction using machine learning/IoT: A survey." In *Advanced Informatics for Computing Research: Second International Conference, ICAICR 2018, Shimla, India, July 14–15, 2018, Revised Selected Papers, Part I 2*, pp. 471-479. Springer Singapore, 2019.

[2] Choudhury, Ambika, and Deepak Gupta. "A survey on medical diagnosis of diabetes using machine learning techniques." In *Recent Developments in Machine Learning and Data Analytics: IC3 2018*, pp. 67-78. Springer Singapore, 2019.

[3] Samant, Piyush, and Ravinder Agarwal. "Machine learning techniques for medical diagnosis of diabetes using iris images." *Computer methods and programs in biomedicine* 157 (2018): 121-128.

[4] Birjais, Roshan, Ashish Kumar Mourya, Ritu Chauhan, and Harleen Kaur. "Prediction and diagnosis of future diabetes risk: a machine learning approach." *SN Applied Sciences* 1 (2019): 1-8.

[5] Chen, Min, Jun Yang, Jiehan Zhou, Yixue Hao, Jing Zhang, and Chan-Hyun Youn. "5G-smart diabetes: Toward personalized diabetes diagnosis with healthcare big data clouds." *IEEE Communications Magazine* 56, no. 4 (2018): 16-23.

[6] Mahabub, Atik. "A robust voting approach for diabetes prediction using traditional machine learning techniques." *SN Applied Sciences* 1, no. 12 (2019): 1667.

[7] Yuvaraj, N., and K. R. SriPreethaa. "Diabetes prediction in healthcare systems using machine learning algorithms on Hadoop cluster." *Cluster Computing* 22, no. Suppl 1 (2019): 1-9.

[8] Mansourypoor, Fatemeh, and Shahrokh Asadi. "Development of a reinforcement learning-based evolutionary fuzzy rule-based system for diabetes diagnosis." *Computers in Biology and Medicine* 91 (2017): 337-352.

[9] Warke, Mitesh, Vikalp Kumar, Swapnil Tarale, Payal Galgat, and D. J. Chaudhari. "Diabetes diagnosis using machine learning algorithms." *Diabetes* 6, no. 03 (2019): 1470-1476.

[10] Birjais, Roshan, Ashish Kumar Mourya, Ritu Chauhan, and Harleen Kaur. "Prediction and diagnosis of future diabetes risk: a machine learning approach." *SN Applied Sciences* 1 (2019): 1-8.

[11] Mujumdar, Aishwarya, and V. Vaidehi. "Diabetes prediction using machine learning algorithms." *Procedia Computer Science* 165 (2019): 292-299.

[12] Al-Zebari, Adel, and Abdulkadir Sengur. "Performance comparison of machine learning techniques on diabetes disease detection." In *2019 1st international informatics and software engineering conference (UBMYK)*, pp. 1-4. IEEE, 2019.

[13] Rubaiat, Sajratul Yakin, Md Monibor Rahman, and Md Kamrul Hasan. "Important feature selection & accuracy comparisons of different machine learning models for early diabetes detection." In *2018 International Conference on Innovation in Engineering and Technology (ICIET)*, pp. 1-6. IEEE, 2018.

[14] Sarwar, Muhammad Azeem, Nasir Kamal, Wajeeha Hamid, and Munam Ali Shah. "Prediction of diabetes using machine learning algorithms in healthcare." In *2018 24th international conference on automation and computing (ICAC)*, pp. 1-6. IEEE, 2018.

[15] Kavakiotis, Ioannis, Olga Tsave, Athanasios Salifoglou, Nicos Maglaveras, Ioannis Vlahavas, and Ioanna Chouvarda. "Machine learning and data mining methods in diabetes research." *Computational and structural biotechnology journal* 15 (2017): 104-116.