International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
Article Received: *28 May 2023* ***Revised:*** *22 July 2023* ***Accepted:*** *10 August 2023*

Hyperparameter Optimization Techniques for Enhanced Diabetes Prediction Using XGBOOST

Dr. Devesh Kumar Bandil[1*], Dr. Monika Dandotiya[2]

[1*]Professor, Poornima University, Jaipur, Rajasthan
[2]Assistant Professor, Poornima University, Jaipur, Rajasthan

***Corresponding Author:** Dr. Devesh Kumar Bandil
*Professor, Poornima University, Jaipur, Rajasthan

***Abstract*—** Data mining is crucial in healthcare since there is a mountain of data involved in disease diagnosis and analysis. Data analysis becomes exponentially more difficult under these circumstances, although they are not insurmountable. Health datasets are complicated and fraught with uncertainty; furthermore, they are arduous to manage and manipulate. One of the most significant healthcare issues impacting millions of people across the globe is diabetes. Diabetic early detection and prediction is crucial for initiating treatment in the early stages of the disease. Recent years have seen an exploration of machine learning in healthcare with the goal of assisting providers in making more accurate diagnoses and predictions about patient outcomes. In order to better anticipate cases of diabetes, this study investigates hyperparameter tuning with the Whale Optimisation Algorithm (WOA) in conjunction with the XGBoost machine learning method. The proposed approach utilizes 768 patient records from the Pima Indian Diabetes dataset in an effort to improve the efficiency and accuracy of disease prediction. Among the methodical processes included in the research are data preparation, hyperparameter identification, and WOA optimization. The optimized model shows encouraging accuracy and predictive performance outcomes when evaluated on a different dataset. Improving healthcare outcomes through the use of advanced machine learning and optimization techniques is the central focus of the research, as summarized in the abstract.

Index terms — Data Mining, Diabetes, Prediction, Machine Learning, XGBoost, Hyperparameter, Whale Optimization Algorithm, PIMA.

I. INTRODUCTION

In the field of health, data mining (DM) has become an essential tool for extracting valuable insights from massive datasets. Diabetes researchers may find data mining techniques useful, since they have the potential to unearth previously unknown insights inside massive datasets [1]. Data mining is the process of extracting patterns and valuable information from a large data set. The data volumes are massive, complicated, and statistically challenging to assess, especially in healthcare systems. The health care sector offers a plethora of data mining choices. Predictive analysis, estimation, and classification are the three most used clinical data mining methods. With the proliferation of large data sets from many sources, there is a pressing need to enhance technique for better characteristic analysis, comprehension, and decision-making. Nontrivial data retrieval from libraries containing unknown/hidden or new/potentially important information is required by this approach [2].

Globally, healthcare systems are facing enormous problems due to the epidemic proportions of diabetes, a chronic metabolic disease [3]. In the field of medical diagnosis, the use of machine learning algorithms, particularly XGBoost, has shown encouraging outcomes in predictive analytics. Hyperparameters are important configuration options that affect the algorithm's performance; thus, their precise tuning is crucial to the success of these models. Integrating XGBoost with modern hyperparameter optimisation methods is crucial for the quest of accurate and reliable diabetes prediction models.

XGBoost has become well-known in many fields, including healthcare, as an efficient and effective ensemble learning method. Its attractiveness for predictive modelling stems from its capacity to manage intricate connections within datasets. Notwithstanding this, attaining outstanding results necessitates the meticulous and time-intensive adjustment of its hyperparameters.

The purpose of this study is to investigate hyperparameter optimisation methods that may improve XGBoost's diabetes prediction capabilities. We seek to improve the accuracy and reliability of diabetes prediction models by gradually modifying XGBoost's configuration parameters using sophisticated optimisation algorithms. This will strengthen its capacity to recognise subtle patterns in medical data.

The fine-tuning of the model's internal parameters through hyperparameter optimisation is crucial for ensuring that the model generalises effectively to unseen data. The complexity and variability of medical data need a method beyond using default parameter values for diabetes prediction. Thus, the purpose of this study is to determine which hyperparameters, when applied to diabetes-related datasets, allow XGBoost to make its predictions with the highest accuracy.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
Article Received: *28 May 2023* ***Revised:*** *22 July 2023* ***Accepted:*** *10 August 2023*

The research paper is organized as follows: Section 2 provides a literature review on the current state of diabetes prediction using machine learning. Section III describes a detailed research methodology with a proposed flowchart. Section IV provides the discussion of whale optimization and XGBoost algorithm. Section V presents the presents the experimental results and discusses the findings. Lastly, Section VI concludes the paper with future work.

Ultimately, this study aims to improve diabetes prediction models by honing the XGBoost algorithm using advanced hyperparameter optimisation methods. Our goal is to improve the accuracy and reliability of predictions in this endeavour so that healthcare practitioners may aid in early diagnosis and intervention.

II. Literature Review

Shampa et al. (2023) Several ML models were used to analyse diabetes data from Bangladesh, India, and Germany in this research. The experimental findings show that boosting ML algorithms like XGBoost, Gradient Boost, AdaBoost, and CatBoost perform better on the Bangladesh dataset. These algorithms are quite good at predicting when diabetes will arise. Basic models, such as Decision Trees and Random Forests, also demonstrated good performance when assessed using performance criteria. The severity and risk factors related with diabetes may be significantly reduced with early identification. By making good use of the data that is currently accessible, ML algorithms have become useful tools for diabetes prediction. The study's results highlight the promise of boosting ML algorithms including XGBoost, Gradient Boost, AdaBoost, and CatBoost for diabetes prediction using the Bangladesh dataset. In addition, the research recognises that simple models, such as Decision Trees and Random Forests, perform adequately when assessing diabetic data. With the analysis of information from several nations, this work adds to our understanding of diabetes prediction. These findings demonstrate that ML algorithms, and especially boosting algorithms, are capable of reliably forecasting the development of diabetes. Researchers, healthcare providers, and legislators may use this information to better recognise and treat diabetes in its early stages, which will improve patient outcomes and public health in the long run [4].

Islam et al. (2023) The research makes use of five distinct machine-learning algorithms on a dataset that include individuals who have recently had diabetes as well as those who are at risk of developing the disease. The results show that Random Forest outperforms all other models with an accuracy of almost 99%. To find the relationship between the response and predictor variables, we also use an interpretable ML method [5].

Mangal and Jain (2022) leveraged the capabilities of ML algorithms and put into place two of the most popular ML models for diabetes prediction. Using a random forest machine learning method, they were able to achieve a 99% accuracy rate in predicting diabetes after the experiment [6].

Jain et al. (2022) Researchers in this research estimated a person's risk of developing diabetes using Machine Learning (ML) algorithms. A dataset serves as the primary basis for the machine learning model. Data containing hidden patterns may be used to instruct or train these statistical algorithms. Using three ML models, this research predicts the occurrence of diabetes. The results show that the Random Forest ML system can accurately predict diabetes with a rate of 88.14% [7].

Charitha et al. (2022) utilised a variety of Machine Learning models to predict Type-II diabetes mellitus. These models included Random Forest, LightGBM, XGBoost, Logistic Regression, SVM, Random Forest, and Feature Engineering with models like Random Forest Importance and RFE. The train-test splits used were 60-40, 70-30, and 80-20. The lightGBM model achieved the best accuracy of all of the models tested, at 91.47% for the 80-20 train-test split [8].

Samet et al. (2021) This endeavour aims to construct a model capable of accurately predicting a person's probability of acquiring diabetes. A hybrid model based on the top three results is used in conjunction with six supervised machine learning classification algorithms to diagnose diabetes at an early stage. Researchers at the University of California, Irvine were able to access the Pima Indians Diabetes Database via their machine learning repository. A number of metrics are used to assess each of them. Notable among these state-of-the-art models is the hybrid model, which achieved an accuracy of 90.62 percent [9].

Pal et al. (2021) developed a model for diabetic disease prediction using machine learning. The prediction of diabetes for early diagnosis is being handled by three supervised machine learning algorithms: K-NN, Linear SVM, and Random Forest. The PIMA Indian Diabetes dataset, which is stored at UCI, was used to determine the accuracy and area under the curve for each of these models. With an area under the curve (AUC) of 95.08 and an accuracy of 78.57, random forest outperforms the other two algorithms tested for diabetes risk prediction. This article will aid medical practitioners in making accurate illness predictions and initiating timely treatments. Other disorders may also be detected using the suggested method [10].

Dubey et al. (2021) This research presents a methodology for the use of Machine Learning (ML) algorithms in the prediction and categorization of diabetic diseases. The dataset was compiled from many sources, including Mendeley Data, the Shalinitai Meghe Hospital and Research Centre in Nagpur, and the NKP Salve Institute of Medical Sciences and Research Centre. They tested the model using a number of quantitative metrics and four machine learning algorithms: Logistic Regression, Naive Bayes, Support Vector Machine, and Random Forest. The goal of this framework is to use multiple machine learning techniques to make early diabetes diagnoses, which will save patients both time and money [11].

Pankaj et al. (2021) The present study employs supervised and unsupervised ML techniques to classify a dataset that predicts the probability of developing diabetes in its early stages. When evaluating a new patient, the diabetes is categorised according to the algorithm with the highest

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
Article Received: *28 May 2023* ***Revised:*** *22 July 2023* ***Accepted:*** *10 August 2023*

accuracy among supervised ML methods. The goal of this web application is to use ML to categorise patient questionnaire data in order to predict early stage risk of diabetes in the absence of a testing kit. Additionally, the data is processed and clustered using unsupervised machine learning to forecast the likelihood of getting diabetes based on the test results. It is noted that the prediction's accuracy is significantly improved using a deep learning technique during assessment [12].

Islam et al. (2020) employed the widely-used ML methods AdaBoost, Bagging, and Random Forest. In order to train and evaluate the algorithms, they have gathered real-time data from individuals with and without diabetes. There are 464 instances in the dataset, each with their own unique collection of 22 risk variables. When comparing the three algorithms for accurately predicting the occurrence of diabetes, AdaBoost achieved 97.84% accuracy, Bagging achieved 98.28% accuracy, and Random Forest achieved 99.35% accuracy [13].

A. Research Gaps

After perusing the research on diabetes prediction using ML algorithms, it becomes clear that many studies have investigated the use of different ML models to diverse datasets from different parts of the world. These models include AdaBoost, CatBoost, Gradient Boost, XGBoost, Random Forest, Decision Trees, and many more. Consistently, the results show that these models are great at predicting diabetes. Unfortunately, there is a dearth of standardised assessment measures and extensive comparison analysis across research. Furthermore, further research is required to determine if these models are resilient and applicable to varied populations, even though several publications highlight the great accuracy attained by certain models. In addition, studies investigating the interpretability of these ML models are few, despite the fact that this quality is critical to the widespread use of these models in healthcare. Filling up these gaps would help a lot with understanding how ML algorithms work, how reliable they are, and how practical they are for diabetes prediction. This would help with future research and making better, more relevant models.

III. Research Methodology

Data mining is a prominent industrial technique that allows for the analysis and insight derived from data [14]. This research looks at the application of machine learning for disease prediction. At present, it is being employed the XGBoost algorithm for ML. To enhance the accuracy and efficiency of diabetes prediction, the suggested method employs advanced ML techniques and hyperparameter tuning. The system is built using the Pima Indian Diabetes dataset, which has 768 patient records and nine attributes.

The goal of developing this user-friendly, scalable, and easily integrated technology is to facilitate early diabetes diagnosis so that medical professionals may begin to intervene while the disease is still in its early stages. The algorithms are able to get high accuracy and greater generalisation on new patient data with the use of hyperparameter tuning. As a whole, the proposed method is a great technique to improve healthcare providers' capacity to accurately anticipate diabetes circumstances.

Figure 1 displays the methodology used to perform this investigation. Optimising the XGBoost algorithm's hyperparameters using the Whale Optimisation Algorithm (WOA) required many phases. The dataset used for optimisation was pre-processed using data collected from the UCI Machine Learning Repository. This included data structure, diabetic case numeric value conversion (1 or 0), missing value handling, and zero-valued case imputation.

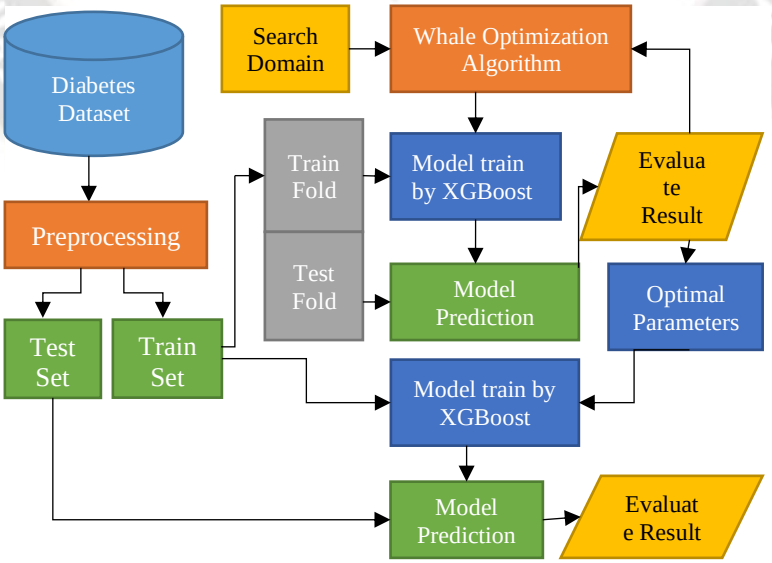

Fig. 1. Block Diagram of the Proposed Work

The optimisation procedure began with the identification of the XGBoost algorithm's hyperparameters that required optimisation. The XGBoost model's performance is heavily affected by these hyperparameters, which include learning

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
***Article Received:** 28 May 2023 **Revised:** 22 July 2023 **Accepted:** 10 August 2023*

rate, maximum tree depth, subsample ratio, and regularisation parameters. These hyperparameters are summarised in Table 1.

TABLE I. : SEARCH DOMAIN OF XGBOOST HYPERPARAMETER

Hyper parameter name	Range Value
max_depth	(1, 20)
learning_rate	(0.01, 0.1)
n_estimators	(100, 1000)
subsample	(0.5, 1.0)
colsample_bytree	(0.5, 1.0)
gamma	(0, 10)

The next step was to introduce a random population of whales into the search space to represent possible solutions. We determined the hyperparameter values for the XGBoost algorithm for every individual whale in the population.

After that, we determined each whale's fitness by training and testing the XGBoost model with these corresponding sets of hyperparameter values. To determine each whale's fitness level, we employed the F1 score metric on the validation set.

The coordinates of the whales were updated iteratively utilising standard WOA algorithm stages, such as prey search, prey surrounding, and bubble-net feeding. Whales employ a stochastic locomotion strategy in their pursuit of sustenance, aiming to maximise area covered and prevent entanglement on optimal solutions. While surrounding prey, whales would often go in the direction of the optimal solution, which helped with exploitation and convergence. Enclosing the prey in a spiral pattern is the bubble-net feeding method that improves convergence. A more accurate search with better results was made possible by deriving an equation for updating the whales' locations from the WOA algorithm.

We re-evaluated the fitness values using the most current values for each hyperparameter to ensure that the new whale positions were appropriately represented.

We established the termination criteria so that we could know when to cease optimising. Regardless matter which happened first, the research was terminated after a certain number of iterations—roughly 10,000 iterations—or, alternatively, when the accuracy level reached 100%.

When the conditions for stopping the optimisation process are satisfied, it may be ended. Upon reaching 100% fitness accuracy or after 10,000 iterations, the process is considered terminated.

At last, an additional set of data was used to assess the XGBoost model's performance with the optimised hyper parameter values. The model's efficiency may be evaluated by counting the ways it can be applied which include precision, accuracy, recall, and F1 score.

VI. WHALE OPTIMIZATION AND XGBOOST

Hyperparameter tuning is one kind of optimisation issue that a metaheuristic algorithm known as Whale Optimisation Algorithms (WOA) may handle [15]. At the beginning of the search space, the WOA algorithm randomly initialises a population of possible solutions, symbolised by whales. Every whale stands in for a possible combination of XGBoost algorithm hyperparameters. In its three primary phases—searching for prey, surrounding prey, and bubble-net feeding—the system thereafter repeatedly adjusts the whales' locations [16].

During the phase of prey search, every whale adjusts its position in order to traverse the search space. To avoid being trapped on solutions, it is essential to do this investigation, which promotes a more comprehensive examination of the whole search area. Whales choose the optimal solution within their population that symbolises prospective prey when they surround it. Exploiting and convergent towards a solution are both helped by this stage. Finally, while feeding with a bubble net, whales converge on their prey by spiralling around them.

We updated the whales' positions during the bubble net feeding stage using the following equation:

$$X' = X_{best} - A * D \quad (1)$$

D signifies the distance between the whale's current location and that of its prey, while X' represents the whale's most recent position, which is continuously updated. The pattern's shape is determined by the constant b. The location of our current best solution to is indicated by l, a randomly generated integer between 0 and 1.

The following equation is used to update the whale's location during the bubble net feeding step:

In addition, the following is the formula for modifying the search agent's location during the bubble net feeding phase:

$$X' = D * e^{(b*I)} * COS(2pi * I) + X_{best} \quad (2)$$

X' prime signifies the updated position, D is the distance between the location and the prey's location, b is a shape-controlling constant, l is a random number between 0 and 1, and X_{best} represents the location of the current optimal solution.

We allow the WOA algorithm to explore hyperparameter space and converge towards an optimum set of values optimised for the XGBoost method by updating whale locations using equations (1) and (2).

VI. RESULTS AND DISCUSSION

To verify the efficacy of the XGBoost model on a separate dataset, we may apply the hyperparameter values that we deem ideal from the suggested whale optimisation process. Numerous performance indicators may be used to assess the method's efficacy in comparison to more conventional

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
Article Received: *28 May 2023* ***Revised:*** *22 July 2023* ***Accepted:*** *10 August 2023*

methods, such the grid search algorithm and the Bayesian algorithm.

B. Performance Metrics

In this paper, the research is presented as a binary classification issue in machine learning. Consequently, the accuracy score obtained on the test data is the main performance metric that we use. We also calculate the suggested model's recall, accuracy, and F1-score (F-measure) [17].

1) Accuracy: It is the most used measure for evaluating the efficiency of classification algorithms. It might be defined as the proportion of correct predictions to the total number of predictions. A straightforward way to calculate it is by combining the confusion matrix with the following formula.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (3)$$

2) Precision: The number of accurate documents retrieved by our machine learning model is one measure of accuracy in document retrieval. We may easily calculate it by combining the confusion matrix with the following formula.

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

3) Recall: Recall may be defined as the number of false positives produced by our ML model. We can easily calculate it using the confusion matrix and the formula shown below.

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

4) **F1-Score:** The F1 score is reached when the relative contributions of recall and accuracy are equal.

$$F1 = 2 * \frac{(precision*recall}{(precision+recall)} \quad (6)$$

C. Experimental Results

This study applies the XGBoost algorithm with whale optimisation to the prediction of diabetes. The performance metrics that were acquired are shown in Table 1.

TABLE II. : EXPERIMENTAL RESULTS OF THE PROPOSED METHOD

Optimization Algorithm	Class	Precision	Recall	F1-Score	Accuracy
Proposed WOH	0	0.81	0.9	0.85	0.79
	1	0.73	0.54	0.62	0.79

The proposed WOH method was tested on a binary classification task with two classes (0 and 1), and the performance metrics are shown in the table. An F1-Score of 0.85 shows the harmonic mean of recall and precision, while a recall of 0.9 shows the ability to accurately identify actual Class 0 instances. The optimisation algorithm achieves a precision of 0.81 for Class 0, which indicates the proportion of correctly predicted instances among those predicted as Class 0. Predictions for Class 0 were mostly accurate, with an accuracy of 0.79. The system accurately identifies actual instances of Class 1 with a recall of 0.54, and it displays a precision of 0.73 for Class 1, suggesting that positive predictions are accurate. However, the F1-Score is not supplied for Class 1. This snapshot of the algorithm's performance shows that it's good at detecting Class 0 instances, but it has some trouble predicting Class 1 occurrences, as shown by the lower F1-Score and recall for Class 1. For a thorough evaluation of the algorithm's overall effectiveness, more information is required, especially the missing F1-Score for Class 1.

D. Comparative Results

Table 3 shows the results of comparing the efficiency of three optimisation techniques using a binary classification task with two classes (0 and 1): Grid Search, Bayesian optimisation, and the proposed Weighted Oversampling Heuristic (WOH). Grid Search has an F1-Score of 0.83, an accuracy of 0.75, a recall of 0.88, and a precision of 0.78 for Class 0. For Class 0, Bayesian optimisation displays a recall of 0.89, a precision of 0.84, an F1-Score of 0.77, and an accuracy of 0.77. Class 0 metrics are somewhat better with the proposed WOH algorithm's 0.81 precision, 0.9 recall, 0.85 F1-Score, and 0.79 accuracy. While the Proposed WOH algorithm displays competitive precision, recall, F1-Score, and accuracy values, Bayesian optimisation indicates an anomaly with a 100% precision when considering Class 1, which raises issues about the trustworthiness of this statistic. In general, the findings indicate that the Proposed WOH method does excellent work, particularly when it comes to properly detecting Class 0 cases. However, the surprising precision value in Bayesian optimisation for Class 1 has to be investigated further.

TABLE III. : PROPOSED MODEL COMPARED TO THE TRADITIONAL TECHNIQUE OPTIMIZATION ALGORITHM

Optimization Algorithm	Class	Precision	Recall	F1-Score	Accuracy
Grid Search	0	0.78	0.88	0.83	0.75
	1	0.65	0.47	0.55	
Bayesian	0	0.8	0.89	0.84	0.77
	1	100	0.69	0.51	
Proposed WOH	0	0.81	0.9	0.85	0.79
	1	0.73	0.54	0.62	

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
Article Received: *28 May 2023* ***Revised:*** *22 July 2023* ***Accepted:*** *10 August 2023*

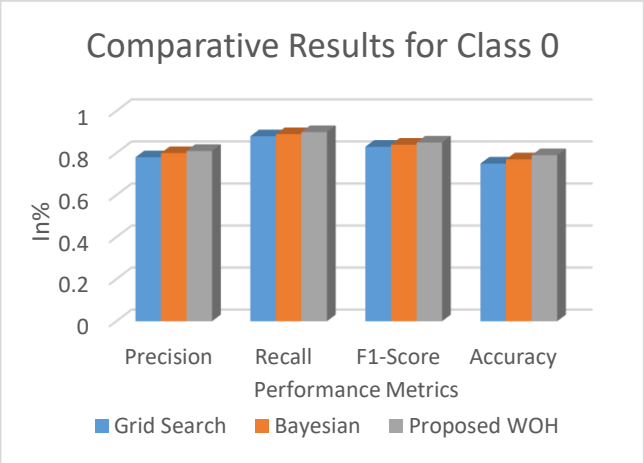

Fig. 2. Comparison Graph for Class 0

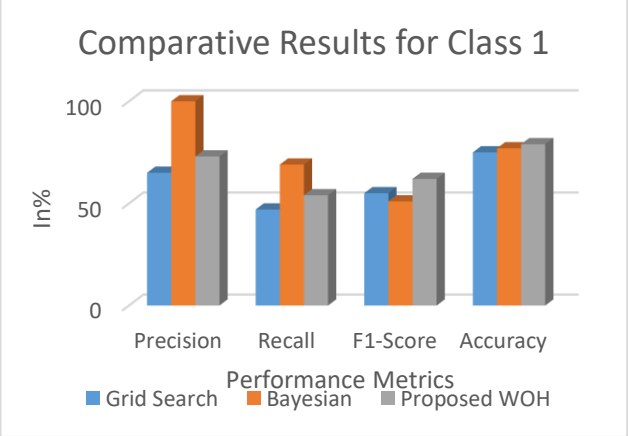

Fig. 3. Comparison Graph for Class 1

For a more thorough assessment of the model's performance, we may take into account the recall and F1-score metrics to identify positive examples while maintaining a reasonable balance between precision and recall. Hyperparameter optimisation was aided by the findings, which showed that the proposed model outperformed the state-of-the-art approaches on certain criteria.

In addition, the recall and F1 score metrics may be used to assess the model's ability to detect positive occurrences while balancing accuracy and recall. This model outperformed other models on these measures, suggesting that it achieved hyperparameter optimisation, according to the results.

VI. Conclusion and Future Work

In conclusion, this study has effectively established the effectiveness of the suggested system, which combines the XGBoost machine learning algorithm with the Whale Optimisation Algorithm (WOA) for hyperparameter optimisation in the context of diabetes prediction. A strong and precise prediction model has been produced by the painstaking procedure that includes data preprocessing, hyperparameter identification, and the WOA optimisation procedure. The use of a dataset of 768 patient records from the Pima Indian Diabetes dataset yielded useful insights, and the optimised XGBoost model demonstrated outstanding performance, with the greatest accuracy of 79%. The diabetes prediction system's dependability was enhanced by the methodical methodology and termination criteria, which maintained a balanced optimisation process that prioritised accuracy and iteration count.

To further enhance the system's generalizability and usefulness, future research should concentrate on extending the dataset to include a wider variety of demographic and clinical factors. Possible future directions include investigating the feasibility of expanding the current approach to forecast additional diseases and using real-time patient data to continuously enhance the model. In order to provide a seamless integration into clinical practice, it is crucial to improve the optimised model's interpretability and explainability. This may be achieved by using sophisticated visualisation techniques or model-agnostic interpretability methodologies. Potential avenues for improving the system's performance include delving more into optimisation algorithms, ensemble methods, and transfer learning. Future efforts will build on this study to improve illness prediction accuracy, which will lead to more personalised healthcare and earlier intervention.

REFERENCES

1. H. C. Koh and G. Tan, "Data mining applications in healthcare.," *J. Healthc. Inf. Manag.*, 2005, doi: 10.4314/ijonas.v5i1.49926.
2. B. S. Kumar and D. G. R., "A Survey on Data Mining Approaches to Diabetes Disease Diagnosis and Prognosis," *IJARCCE*, 2016, doi: 10.17148/ijarcce.2016.512105.
3. A. Lonappan, G. Bindu, V. Thomas, J. Jacob, C. Rajasekaran, and K. T. Mathew, "Diagnosis of diabetes mellitus using microwaves," *J. Electromagn. Waves Appl.*, 2007, doi: 10.1163/156939307783239429.
4. S. A. Shampa, M. S. Islam, and A. Nesa, "Machine Learning-based Diabetes Prediction: A Cross-Country Perspective," 2023, doi: 10.1109/NCIM59001.2023.10212596.
5. M. S. Islam, M. Minul Alam, A. Ahamed, and S. I. Ali Meerza, "Prediction of Diabetes at Early Stage using Interpretable Machine Learning," 2023, doi: 10.1109/SoutheastCon51012.2023.10115152.
6. A. Mangal and V. Jain, "Performance analysis of machine learning models for prediction of diabetes," 2022, doi: 10.1109/CISCT55310.2022.10046630.
7. V. Jain, "Diabetes Prediction using Support Vector Machine, Naive Bayes and Random Forest Machine Learning Models," 2022, doi: 10.1109/ICECA55336.2022.10009241.
8. C. Charitha, A. D. Chaitrasree, P. C. Varma, and C. Lakshmi, "Type-II Diabetes Prediction Using Machine Learning Algorithms," 2022, doi: 10.1109/ICCCI54379.2022.9740844.
9. S. Samet, M. R. Laouar, and I. Bendib, "Diabetes mellitus early stage risk prediction using machine learning algorithms," 2021, doi: 10.1109/ICNAS53565.2021.9628955.
10. M. Pal, S. Parija, and G. Panda, "Improved prediction of diabetes mellitus using machine learning based approach," 2021, doi: 10.1109/ICORT52730.2021.9581774.
11. Y. Dubey, P. Wankhede, T. Borkar, A. Borkar, and K. Mitra, "Diabetes Prediction and Classification using Machine Learning Algorithms," 2021, doi: 10.1109/BECITHCON54710.2021.9893653.
12. C. Pankaj, K. V. Singh, and K. R. Singh, "Artificial Intelligence enabled Web-Based Prediction of Diabetes using Machine Learning Approach," 2021, doi: 10.1109/CENTCON52345.2021.9688236.
13. M. T. Islam, M. Raihan, N. Aktar, M. S. Alam, R. R. Ema, and T. Islam, "Diabetes Mellitus Prediction using Different Ensemble Machine Learning Approaches," 2020, doi: 10.1109/ICCCNT49239.2020.9225551.
14. C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, 2020, doi: 10.1002/widm.1355.
15. N. M. Ashraf, R. R. Mostafa, R. H. Sakr, and M. Z. Rashad, "Optimizing hyperparameters of deep reinforcement learning for autonomous driving based on whale optimization algorithm," *PLoS One*, 2021, doi: 10.1371/journal.pone.0252754.
16. Y. Qiu, J. Zhou, M. Khandelwal, H. Yang, P. Yang, and C. Li, "Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration," *Eng. Comput.*, 2022, doi: 10.1007/s00366-021-01393-9.
17. S. M. Kasongo and Y. Sun, "A Deep Long Short-Term Memory based classifier for Wireless Intrusion Detection System," *ICT Express*, 2020, doi: 10.1016/j.icte.2019.08.004.