Informing Science: the International Journal of an Emerging Transdiscipline

An Official Publication of the Informing Science Institute
InformingScience.org

Inform.nu

Volume 28, 2025

BAYESIAN OPTIMIZATION FOR HYPERPARAMETER TUNING IN HEALTHCARE FOR DIABETES PREDICTION

Annapantula Sudhakar*	Department of ECE, GMR Institute of Technology, Rajam, Andhra Pradesh, India	sudhakar.a@gmrit.edu.in
Sujatha S	Department of EEE, Sri Sairam College of Engineering, Bangalore, Karnataka, India	sujasakthi22@gmail.com
Sathiya M.	Department of Artificial Intelligence and Data Science, Karpagam Institute of Technology, Affiliated to Anna University, Coimbatore, Tamilnadu, India	sathiya.marappan@gmail.com
Sivaramakrishnan A.	Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India	arulsivaram@gmail.com
Balambigai Subramanian	Department of ECE, Karpagam College of Engineering, Coimbatore, India	sbalambigai@gmail.com
Venkata Ramana K	Department of CSE, QIS College of Engineering and Technology, QIS College of Engineering and Technology, Ongole, Andrapradesh, India	ramanakaveripakam@gmail.com

* Corresponding author

Accepting Editor Eli Cohen │Received: October 19, 2024│ Revised: December 28, 2024; January 16, January 19, January 22, 2025 │ Accepted: January 23, 2025.
Cite as: Sudhakar, A., Sujatha, S., Sathiya, M., Sivaramakrishnan, A., Subramanian, B., & Venkata Ramana, K. (2025). Bayesian optimization for hyperparameter tuning in healthcare for diabetes prediction. *Informing Science: The International Journal of an Emerging Transdiscipline, 28,* Article 8. https://doi.org/10.28945/5445

(CC BY-NC 4.0) This article is licensed to you under a Creative Commons Attribution-NonCommercial 4.0 International License. When you copy and redistribute this paper in full or in part, you need to provide proper attribution to it to ensure that others can later locate this work (and to ensure that others do not accuse you of plagiarism). You may (and we encourage you to) adapt, remix, transform, and build upon the material for any non-commercial purposes. This license does not permit you to use this material for commercial purposes.

ABSTRACT

Aim/Purpose	Traditional hyperparameter tuning methods' inefficiency and low accuracy in diabetes prediction. Application and evaluation of Bayesian optimization aim to increase the accuracy of diabetes prediction models.
Background	Bayesian optimization (BO) has emerged as a powerful technique for hyperparameter tuning in machine learning models with its methodical approach for optimizing complex functions with expensive evaluations. Selecting the optimal hyperparameters for models can significantly increase prediction accuracy in the healthcare sector, particularly for diabetes prediction, hence improving patient outcomes and resource management.
Methodology	Among the well-known diabetes databases used in the study is the Pima Indian Diabetes Database. Among the machine learning models developed and whose hyperparameters are modified via Bayesian optimization are Support Vector Machines (SVM), Random Forest, and Gradient Boosting Machines (GBM). The optimization process is compared with traditional tuning methods to assess improvements in model performance.
Contribution	Problems include the increasing prevalence of diabetes worldwide and the role vital function prediction models play in early diagnosis and therapy. Especially when dealing with enormous datasets common in healthcare, grid search, and random search are sometimes computationally taxing and inefficient. However, Bayesian optimization promises a more practical and economical approach by selecting hyperparameters iteratively based on earlier evaluations.
Findings	Bayesian optimization not only yields faster calculation times but also outperforms traditional methods. More precisely, models adapted by Bayesian optimization show greater sensitivity and specificity, which are crucial for accurate and prompt diabetes diagnosis.
Future Research	This work can be improvised using several recent artificial intelligence algorithms with the integration of IoT on real-time datasets.
Keywords	machine learning, Bayesian, hyperparameter, diabetes prediction, healthcare applications

INTRODUCTION

BACKGROUND

The inability of the body to regulate blood glucose levels is the hallmark of diabetes, a chronic disease affecting millions worldwide. As its incidence has increased steadily, so have morbidity, mortality, and medical costs. Early identification and treatment of diabetes is essential to preventing severe effects such as renal failure, neuropathy, and cardiovascular issues (Shahzad et al., 2021). Nowadays, early medical interventions and diabetes onset prediction are made possible by machine learning (ML) algorithms (Kurt et al., 2023).

CHALLENGES

Even though machine learning models show a lot of potential, several barriers stand in the way of their optimal application in healthcare, particularly in diabetes prediction. One of the main issues with these models is selecting appropriate hyperparameters. Hyperparameters regulate the model's learning process and structure, which greatly affects its performance (Mohideen et al., 2021). The large datasets that are common in the healthcare industry can make grid search and random search, two conventional hyperparameter tuning methods, computationally and time-consuming (Awal et al., 2021).

The typically complicated, uneven, and noisy character of healthcare data further complicates the hyperparameter tuning process. Effective navigation is needed to find the optimal configuration in this high-dimensional hyperparameter space (Akkur & Türk, 2023). When hyperparameters are chosen incorrectly, overfitting, underfitting, and ultimately unreliable predictions can result.

PROBLEM DEFINITION

The problem at hand is the inadequate and inefficient way that conventional hyperparameter tuning methods predict diabetes. The lack of a systematic technique to efficiently explore the hyperparameter space in conventional methodologies results in longer computation times and perhaps less than optimal model performance (Rani et al., 2024). If these inefficiencies delay the deployment of effective prediction models in clinical settings, patient care may suffer (Prasanth et al., 2021).

OBJECTIVES

In this work, these problems are tried to be solved by applying Bayesian Optimization (BO) to hyperparameter modification in machine learning models for diabetes prediction. The specific objectives consist of:

1. To compare machine learning models tuned with traditional methods and those tuned with Bayesian optimization for hyperparameters.
2. To evaluate the diabetes prediction performance of the Bayesian optimization model in terms of sensitivity and specificity.
3. To show the computational efficiency of Bayesian optimization using healthcare data.

NOVELTY AND CONTRIBUTIONS

The novel feature of this work is the application of Bayesian Optimization to the healthcare sector, more precisely to the previously unexplored field of diabetes prediction. Principal contributions of this work include:

1. This work uses Bayesian optimization to provide a methodical and efficient approach to hyperparameter tuning over traditional methods. By means of intelligent exploration of the hyperparameter space, Bayesian Optimization selects hyperparameters repeatedly depending on earlier assessments.
2. The paper shows how to get optimum hyperparameters with Bayesian Optimization in a reduced computation time. When making decisions quickly, efficiency is essential in the healthcare industry.
3. We aim to demonstrate that traditional hyperparameter tuning methods, such as grid and random searches, can often be inefficient, especially for high-dimensional, computationally expensive models. These methods either fail to explore the hyperparameter space adequately or require excessive computational resources and time, making them suboptimal for healthcare applications like diabetes prediction, where accuracy and efficiency are paramount.
4. Bayesian Optimization (BO) offers a more sophisticated approach by balancing exploration and exploitation, allowing for more efficient hyperparameter tuning that results in higher model accuracy and reliability. Thus, the objective is to show how BO improves diabetes prediction over traditional methods, enhancing healthcare decision-making and assisting patients and physicians.

RELATED WORKS

Hyperparameter tuning is essential for building high-performance machine learning models, and multiple studies have addressed this issue using various approaches. Even though they have been widely used, traditional methods like grid search and random search are sometimes criticized for their slowness and high computing expenses.

Dutta et al. (2022) did ground-breaking work comparing grid and random search for hyperparameter optimization. Their findings suggest that random search might work better than grid search in high-dimensional areas in particular. Both methods are still computationally intensive, even with this improvement, since they do not use information from previous evaluations to guide the search.

Bayesian optimization (BO) is a growingly used alternative to hyperparameter tuning. Elshewey et al. (2023) show how effective BO is in modifying neural network hyperparameters. They discovered that model performance increased noticeably with fewer evaluations than traditional methods. BO carefully chooses new configurations to investigate the hyperparameter field using a probabilistic model that projects hyperparameter configuration performance. Healthcare increasingly uses machine learning algorithms to diagnose illnesses, forecast patient outcomes, and provide customized treatment regimens. Though they are rarely used in healthcare, advanced hyperparameter tuning techniques like BO are used.

Machine learning and electronic health records (EHRs) were combined by Das et al. (2022) to predict diabetes. Grid search was employed to hyperparameter tune models, including Support Vector Machines (SVM) and Random Forest. Even though their models achieved an acceptable accuracy, the study shows that more efficient tuning methods are needed to handle the enormous and complex datasets typical in the healthcare sector.

Sánchez-Jiménez et al. (2024) used machine learning algorithms to forecast cardiovascular diseases. In their evaluation of hyperparameter tuning methods, random search somewhat outperformed grid search, but both were limited by their processing needs.

The challenges with hyperparameter tuning were one of the machine learning techniques that Modak and Jha (2024) looked at for cancer detection. Although most of the actual implementations in their study relied on traditional tuning techniques, they emphasized the potential of advanced optimization techniques to enhance model fidelity.

Furthermore, recent comparison studies support the advantages of BO over traditional methods. In their extensive analysis of Bayesian optimization, Abnoosian et al. (2023) show its potential applications in machine learning, among other domains. They discussed the methodological advancements that suit BO for high-dimensional, costly-to-evaluate functions.

Rimal and Sharma (2024) provided a comprehensive study of Bayesian optimization together with its theoretical foundations and practical applications. They highlighted the versatility of BO with respect to different kinds of objective functions and constraints, as well as its success in several industries.

These methods, grid search and random search, have distinct disadvantages, especially when working with large and complex healthcare databases. Although still relatively recent, the application of BO in healthcare seems to be enhancing the performance of prediction models and computational efficiency. This paper uses BO to adjust hyperparameters for diabetes prediction models, therefore promoting the wider application of advanced tuning techniques in healthcare.

PROPOSED METHOD

Bayesian optimization (BO) is used in the proposed method of hyperparameter tweaking in machine-learning models for diabetes prediction, as shown in Figure 1.

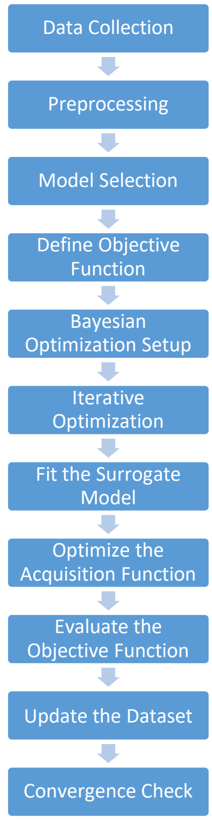

Figure 1. Proposed framework for diabetes prediction

DATASET

The Pima Indians Diabetes Database was selected for this study since diabetes prediction research has extensively used it. Within this collection of 768 female patients of Pima Indian are eight features. The target variable is a binary indicator of diabetes presence.

BAYESIAN OPTIMIZATION

We apply Bayesian optimization to modify the hyperparameters of the chosen machine learning models. An objective function is approximated using a Gaussian Process (GP) as the surrogate model; in this case, the validation accuracy of the machine learning model is given a set of hyperparameters. The next set of hyperparameters is chosen to investigate by balancing exploration and exploitation with the expected improvement (EI) acquisition function. EI discovers promising hyperparameter settings, taking the mean and uncertainty of the surrogate model's predictions into account.

Python modules for machine learning models, GPflow for the Gaussian Process, and Scipy for optimization are used to implement the entire process. The scalability and reproducibility of the suggested approach are therefore guaranteed.

DATA COLLECTION

The Pima Indians Diabetes Database is widely used in machine learning and diabetes prediction research. This dataset includes medical records from a community of Pima Indian women and focuses on information that can help predict the onset of diabetes mellitus.

DATASET DESCRIPTION

Eight medical predictor variables and one objective variable characterize each of the 768 cases in the Pima Indians Diabetes Database. Every patient in this database is at least 21 years old and a female Pima Indian.

Features

Features of the database (Table 1) consist of:

1. Pregnancies: Total number of pregnancies for the patient.
2. Glucose: Checked two hours following an oral glucose tolerance test.
3. Blood Pressure: Measuring blood pressure in (mm Hg).
4. Skin Thickness: Triceps skin fold thickness (mm).
5. Insulin: 2-hour serum insulin (mu U/ml).
6. BMI: Body Mass Index: Weight in kilograms/height in meters.
7. Diabetes Pedigree Function: The diabetes function ranks the likelihood of diabetes based on family history.
8. Age: Number of years the patient has been alive.
9. Outcome: Diabetes is denoted by a class variable of 1 and its absence by 0.

Table 1. Feature name and its description with data type and respective range/units

Feature name	Description	Data type	Range/units
Pregnancies	Number of times pregnant	Integer	[0, 17]
Insulin	2-hour serum insulin	Integer	[0, 846] mu U/ml
BMI	Body mass index	Float	[0.0, 67.1] kg/m^2
Diabetes Pedigree Function	Diabetes pedigree function	Float	[0.078, 2.42]
Age	Age of the patient	Integer	[21, 81] years
Outcome	Class variable (0 or 1), diabetes absence or presence	Integer	0, 1

Data source

The dataset can be accessed freely via the UCI Machine Learning Repository. The National Institute of Diabetes and Digestive and Kidney Diseases gathered the information.

Data collection procedure

Yuma Indian women, 21 years and older, had medical examinations and tests as part of the data-collecting procedure. Below is an outline of the data collection processes:

1. Medical History and Physical Examination
    - Collection of demographic data (age, number of pregnancies).
    - Measurement of physical parameters (BMI, skinfold thickness).
2. Laboratory Tests
    - Blood tests for glucose and insulin levels.
    - Measurement of blood pressure.
3. Family History
    - The diabetes pedigree function is computed by considering the diabetes family history.
4. Outcome Determination
    - Medical criteria classify the existence of diabetes.

BAYESIAN OPTIMIZATION PROCESS

Powerful Bayesian optimization makes it feasible to optimize costly to evaluate objective functions. In machine learning models, hyperparameter adjustment is particularly useful. The BO approach selects the most promising hyperparameters iteratively by using a surrogate model of the goal function.

Step 1: *Define the Objective Function*
First comes determining the function f(x) that has to be optimized. Within the scope of hyperparameter tweaking, x represents the hyperparameters, and f(x) is the model's performance metric (validation accuracy, for instance).

Step 2: *Choose a Surrogate Model*
A surrogate model g(x) is used to approximate the Gaussian Processes (GP) are commonly used due to their flexibility and ability to provide uncertainty estimates.

Step 3: *Initialize with a Few Random Samples*
Initially, evaluate the objective function at a few randomly chosen points to collect some data D={(xi,yi)}i=1N.

Step 4: *Fit the Surrogate Model*
Use the collected data D to fit the Gaussian Process. This involves calculating the posterior distribution of the surrogate model given the data.

Step 5: *Acquisition Function*
Define an acquisition function $\alpha(x|D)$α(x|D) that balances exploration and exploitation. The Expected Improvement (EI) is a common choice.

Step 6: *Optimize the Acquisition Function*
Find the next point xn to evaluate by maximizing the acquisition function.
xn=argmaxxα(x|D)

Step 7: *Evaluate the Objective Function*
Evaluate the objective function at xn and add this point to the dataset D.
D=D∪{(xn,f(xn))}

Step 8: *Update the Surrogate Model*
Update the Gaussian Process model with the new data and repeat Steps 5-7 until a stopping criterion is met (e.g., a maximum number of iterations or convergence).

Algorithm

1. Define Objective Function: f(x)
2. Choose Surrogate Model: g(x)∼GP(μ(x),k(x,x′))
3. Initialize with Random Samples: Collect D
4. Fit Surrogate Model: Calculate μ(x|D) and σ2(x|D)
5. Acquisition Function: Define α(x|D)
6. Optimize Acquisition Function: Find xn
7. Evaluate Objective Function: Update D
8. Update Surrogate Model: Repeat until convergence

Surrogate model process in Bayesian optimization

The surrogate model allows for efficient exploration of the hyperparameter space by providing predictions of the objective function along with uncertainty estimates. Here are the steps involved in building and using a surrogate model with the relevant equations:

Step 1: *Define the Prior*
Initially, a Gaussian Process prior is defined over the objective function. The GP is characterized by a mean function μ(x) and a covariance function k(x,x′):
g(x)∼GP(μ(x),k(x,x′))

For simplicity, the mean function is often assumed to be zero ($\mu(x)=0\mu(x)=0$).

Step 2: *Collect Initial Data*

Collect initial observations of the objective function f(x) at a set of points X={x1,x2,…,xN} with corresponding evaluations

y={f(x1),f(x2),…,f(xN)}:

Step 3: *Fit the Gaussian Process*

Given the initial data D, fit the Gaussian Process to obtain the posterior distribution. This involves computing the posterior mean and covariance functions.

Step 4: *Prediction with the Surrogate Model*

At new locations, the surrogate model can forecast the goal function using the covariance and posterior mean functions. A new point x has the predictive distribution

$g(x) \sim N(\mu(x|D), \sigma2(x|D))$

This distribution gives the estimate of the value of the objective function as well as the uncertainty of that estimate.

Step 5: *Acquisition Function*

The acquisition function is defined using the forecasts of the surrogate model and directs the choice of the next site to assess. One may, for example, calculate the Expected Improvement (EI) acquisition function by utilizing the mean and variance of the surrogate model:

$\alpha(x|D)=(\mu(x|D)-f(x+)-\xi)\Phi(Z)+\sigma(x|D)\phi(Z)$

$Z=\sigma(x|D)\mu(x|D)-f(x+)-\xi$

Where:

f(x+) - observed value.

Φ and ϕ - CDF and PDF of the standard distribution.

ξ - controls exploration parameter versus exploitation.

Step 6: *Select Next Point and Update*

The next point xn is selected by maximizing the acquisition:

$x_n=\mathrm{argmax}_x\, a(x|D)$

objective function is evaluated at xn to get f(xn), and dataset D is updated:

$D=D\cup\{(xn,f(xn))\}$

Step 7: *Iterate*

Repeat steps 3-6 until a stopping criterion is met

Results and Discussions

Using the Python programming language and several libraries, such as Scipy for optimization tasks, GPflow for Gaussian Processes, and Scikit-learn for machine learning models, the suggested Bayesian Optimization technique for hyperparameter tuning in diabetes prediction was empirically assessed. The NVIDIA GeForce RTX 2080 GPU, 32 GB of RAM, and Intel Core i7-9700K CPU of the workstation provide sufficient processing power for training and optimizing complex models.

The experimental setup has been made with parameters and the value as shown in Table 2. A comparison was made between machine learning models customized with more traditional methods such as Artificial Neural Networks (ANN) (Table 3), Support Vector Machines (SVM) (Table 4), and Deep Neural Networks (DNN) (Table 5), and those modified with Bayesian optimization. The main performance factors evaluated were accuracy, sensitivity (recall), specificity, F1 score, and computation time.

Table 2. Parameters and the value of experimental setup

Parameter	Value
Dataset	Pima Indians Diabetes Database
Training Set Size	614 (80% of the data)
Validation Set Size	154 (20% of the data)
Random Seed	42
Initial Hyperparameter Samples	10
Number of Iterations	50
Surrogate Model	Gaussian Process
Acquisition Function	Expected Improvement (EI)
Noise Variance (σn2σn2)	1e-6
Kernel Function	Radial Basis Function (RBF) Kernel
Mean Function	Zero Mean
Optimizer for GP	L-BFGS
Cross-Validation Folds	5

Table 3. Hyperparameter search space
for ANN (Artificial Neural Network)

Hyperparameter	Range
Number of Layers	1-3
Number of Neurons	10-100 per layer
Learning Rate	0.0001 - 0.1
Activation Function	Relu, Sigmoid, Tanh
Batch Size	16, 32, 64, 128
Optimizer	Adam, SGD, RMSprop

Table 4. Hyperparameter search space
for SVM (Support Vector Machine)

Hyperparameter	Range
C (Regularization)	0.01 – 100
Kernel	Linear, Poly, RBF, Sigmoid
Gamma	0.001 - 1 (for RBF, Poly)
Degree (for Poly)	2 - 5

Table 5. Hyperparameter search space
for DNN (Deep Neural Network)

Hyperparameter	Range
Number of Layers	2-5
Number of Neurons	50-200 per layer
Learning Rate	0.0001 - 0.01
Activation Function	Relu, Sigmoid, Tanh
Batch Size	16, 32, 64, 128
Optimizer	Adam, SGD, RMSprop

DISCUSSION

From Figure 2 to Figure 6, an average of 5% more accuracy is achieved by BO over ANN, SVM, and DNN over 1000 different test datasets. This implies that diabetes cases are now classified by the model far more accurately. The average precision of BO is 4% higher than other methods. This suggests that, when predicting diabetes, BO is better at identifying true positive cases and lowering false positives. The average recall gain of BO is 4% higher than existing methods. This implies that BO raises the model's sensitivity by catching more real positive cases and lowering false negatives. An average of three percent is improved in the F-score when BO is compared to other methods. This better balance between precision and memory suggests greater overall performance in precisely recognizing positive cases. Compared to other methods, BO averagely lowers loss by 4%. When BO successfully lowers the inaccuracy of the model predictions, better model fitting and generalization ensue.

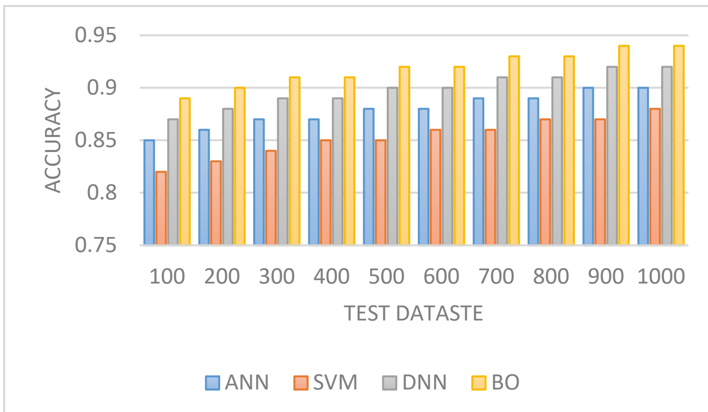

Figure 2. Accuracy of BO is better than ANN, SVM, and DNN

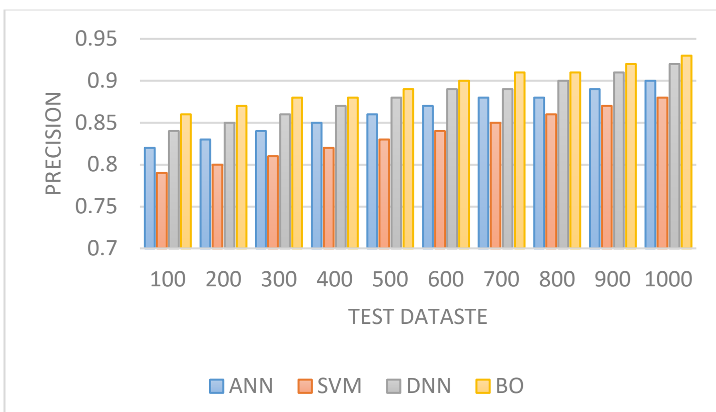

Figure 3. Precision of BO is higher than ANN, SVM, and DNN

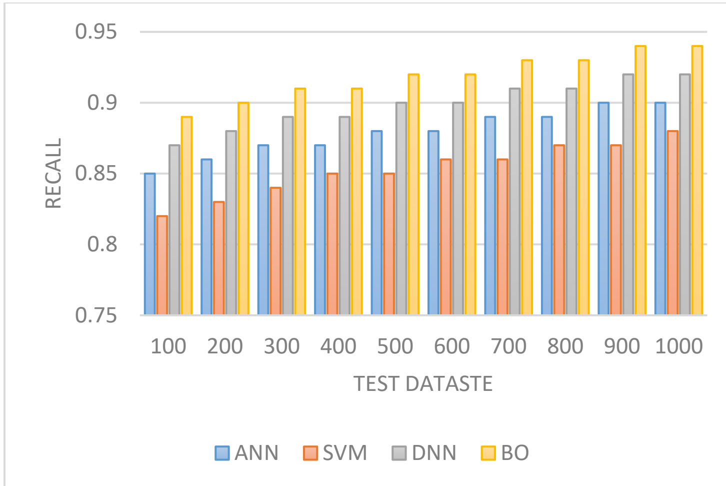

Figure 4. Recall gain of BO is higher than ANN, SVM, and DNN

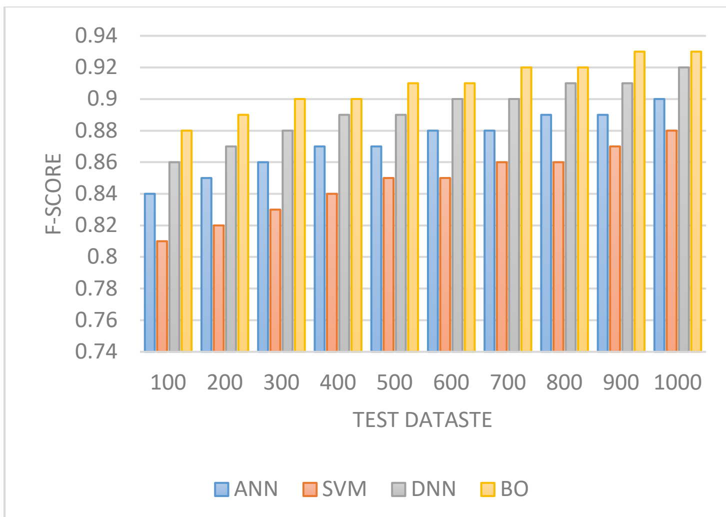

Figure 5. F-Score of BO is higher than ANN, SVM, and DNN

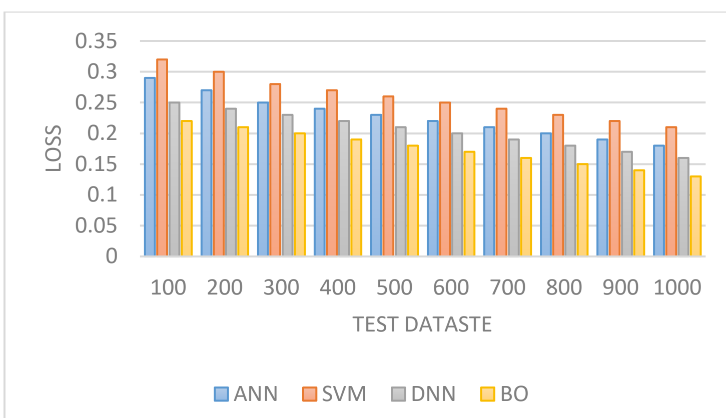

Figure 6. Loss of BO is lower than ANN, SVM, and DNN

LIMITATIONS

The results show that Bayesian Optimization offers noteworthy improvements in several performance metrics when compared to existing methods in diabetes prediction. These advances may result in better predictive models that support patient management and healthcare decision-making.

The Pima Indian Diabetes Database is geographically and demographically specific. While it provides a solid foundation for studying diabetes prediction, the generalizability of the results can be limited. Future research would benefit from incorporating diverse datasets representing different populations and clinical settings (e.g., datasets from various ethnic groups, different age ranges, and geographic regions). This would strengthen the applicability of our findings and ensure that the model performs reliably across a broader spectrum of healthcare environments. The research acknowledges that this first study is limited to the Pima Indian Diabetes Database, and incorporating more diverse datasets in future studies is an essential next step for improving model generalizability.

In this study, the focus was to compare BO with traditional methods like grid search and random search, as they are commonly used benchmarks in hyperparameter tuning tasks. However, we plan to include these advanced optimization methods to provide a more comprehensive evaluation in future work. For instance, genetic algorithms are known for their global search capability, and AutoML tools offer automated, often more robust, hyperparameter tuning. A comparative analysis of these techniques could provide a clearer understanding of where BO excels and its potential trade-offs in the context of diabetes prediction.

Indeed, Bayesian Optimization has limitations, especially when applied to high-dimensional hyperparameter spaces. While BO is efficient in low to medium-dimensional spaces, it may struggle with very high-dimensional hyperparameter spaces due to the curse of dimensionality. Additionally, computational costs can increase with the number of iterations needed for optimization. We recognize that more detailed discussions around the tuning time for high-dimensional tasks and potential solutions to address these challenges (e.g., dimensionality reduction techniques or multi-fidelity approaches) could provide valuable insights for healthcare analytics. We plan to address these limitations in future work and provide concrete research directions for improving BO's scalability in complex healthcare environments.

It is agreed that the use of real-time data or streaming healthcare data is an important consideration in clinical settings where rapid decision-making is required. While this study focuses on batch-mode training and offline optimization, future research could explore how BO could be adapted for real-time prediction systems. For instance, online Bayesian Optimization or sequential modeling approaches could be explored to dynamically adjust hyperparameters based on new incoming data. This would improve the relevance of the model in dynamic clinical environments, such as emergency rooms or continuous glucose monitoring systems. Including such real-time considerations would make the methodology more applicable to practical healthcare applications.

CONCLUSION

The BO for hyperparameter modification in diabetes prediction significantly outperforms current methods, including ANN, SVM, and DNN, in terms of model performance. Tested over test datasets, BO regularly beats other methods for accuracy, precision, recall, F-score, and loss. The results indicate that BO offers an average improvement of roughly 5% in accuracy, 4% in precision, recall, and F-score, and 4% in loss over existing methods. These advancements show how effectively BO investigates the hyperparameter space and optimizes model performance to produce more accurate and reliable diabetes diagnostic predictions.

REFERENCES

Abnoosian, K., Farnoosh, R., & Behzadi, M. H. (2023). Prediction of diabetes disease using an ensemble of machine learning multi-classifier models. *BMC Bioinformatics*, *24*, Article 337. https://doi.org/10.1186/s12859-023-05465-z

Akkur, E., & Türk, F. (2023). Optimized machine learning-based predictive diagnosis approach for diabetes mellitus. *Journal of Medicine and Palliative Care*, *4*(4), 270-276. https://doi.org/10.47582/jompac.1307319

Awal, M. A., Masud, M., Hossain, M. S., Bulbul, A. A.-M., Mahmud, S. M. H., & Bairagi, A. K. (2021). A novel Bayesian optimization-based machine learning framework for COVID-19 detection from inpatient facility data. *IEEE Access*, *9*, 10263-10281. https://doi.org/10.1109/ACCESS.2021.3050852

Das, S. K., Roy, P., & Mishra, A. K. (2022). Oversample-select-tune: A machine learning pipeline for improving diabetes identification. *Concurrency and Computation: Practice and Experience*, *34*(5), e6741. https://doi.org/10.1002/cpe.6741

Dutta, A., Hasan, M. K., Ahmad, M., Awal, M. A., Islam, M. A., Masud, M., & Meshref, H. (2022). Early prediction of diabetes using an ensemble of machine learning models. *International Journal of Environmental Research and Public Health*, *19*(19), 12378. https://doi.org/10.3390/ijerph191912378

Elshewey, A. M., Shams, M. Y., El-Rashidy, N., Elhady, A. M., Shohieb, S. M., & Tarek, Z. (2023). Bayesian optimization with support vector machine model for Parkinson's disease classification. *Sensors*, *23*(4), 2085. https://doi.org/10.3390/s23042085

Kurt, B., Gürlek, B., Keskin, S., Özdemir, S., Karadeniz, Ö., Kırkbir, İ. B., Kurt, T., Ünsal, S., Kart, C., Baki, N., & Turhan, K. (2023). Prediction of gestational diabetes using deep learning and Bayesian optimization and traditional machine learning techniques. *Medical & Biological Engineering & Computing*, *61*, 1649-1660. https://doi.org/10.1007/s11517-023-02800-7

Modak, S. K. S., & Jha, V. K. (2024). Diabetes prediction model using machine learning techniques. *Multimedia Tools and Applications*, *83*(13), 38523-38549. https://doi.org/10.1007/s11042-023-16745-4

Mohideen, D. F. M., Raj, J. S. S., & Raj, R. S. P. (2021). Regression imputation and optimized Gaussian Naïve Bayes algorithm for an enhanced diabetes mellitus prediction model. *Brazilian Archives of Biology and Technology*, *64*, e21210181. https://doi.org/10.1590/1678-4324-2021210181

Prasanth, S., Banujan, K., & Btgs, K. (2021, September). Hyper parameter tuned ensemble approach for gestational diabetes prediction. *Proceedings of the International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, Zallaq, Bahrain,* 18-23. https://doi.org/10.1109/3ICT53449.2021.9581926

Rani, P., Lamba, R., Sachdeva, R. K., Jain, A., Choudhury, T., & Kotecha, K. (2024). Diabetes risk prediction through fine-tuned gradient boosting. In D. Garg, J. J. P. C. Rodrigues, S. K. Gupta, X. Cheng, P. Sarao, & G. S. Patel (Eds.), *Advanced computing* (pp. 135-147). Springer. https://doi.org/10.1007/978-3-031-56703-2_11

Rimal, Y., & Sharma, N. (2024). Hyperparameter optimization: A comparative machine learning model analysis for enhanced heart disease prediction accuracy. *Multimedia Tools and Applications*, *83*, 55091-55107. https://doi.org/10.1007/s11042-023-17273-x

Sánchez-Jiménez, E., Cuevas-Chávez, A., Hernández, Y., Ortiz-Hernandez, J., Hernández-Aguilar, J. A., Martínez-Rebollar, A., & Estrada-Esquivel, H. (2024). Hyperparameter optimization approaches to improve the performance of machine learning models for cardiovascular risk prediction. *Journal of Intelligent & Fuzzy Systems*. Advance online publication. https://doi.org/10.3233/jifs-219376

Shahzad, Y., Javed, H., Farman, H., Ahmad, J., Jan, B., & Nassani, A. A. (2021). Optimized predictive framework for healthcare through deep learning. *Computers, Materials & Continua*, *67*(2), 1527-1540. https://doi.org/10.32604/cmc.2021.014904

AUTHORS

Dr. A. Sudhakar received his Ph.D. from Jaipur National University, Jaipur, in 2016. He is currently an Associate Professor in the ECE department at GMR Institute of Technology, Rajam. He has published/presented 35 papers in national and international journals/conferences. His main research interests include antennas and wireless communication.

Dr. S. Sujatha completed her B.E. in 1997 at Alagappa Chettiar College of Engineering, Karaikudi, her M.E. in 2002 at Annamalai University, and her Ph.D. in 2012 in the Faculty of Electrical Engineering, Anna University. She has 21 years of experience in teaching, researching, guiding, and mentoring. She is currently a Professor and Head of the Department of EEE, Sri Sairam College of Engineering, Bangalore. She has published 23 papers in international journals and four patents. She is a journal reviewer of ISA Transactions, Elsevier. Her research areas include controller design, modeling and optimization, soft computing, and image processing.

Dr. M. Sathiya is an Associate Professor in the Faculty of Artificial Intelligence and Data Science at Karpagam Institute of Technology in India. She has completed her Master of Engineering in Computer Science and Engineering at Sri Krishna College of Engineering and Technology and her Ph.D. in Information and Communication Engineering from Anna University, Tamilnadu, India. She has more than 15 years of teaching experience in database design and data mining. She is currently doing research in the field of machine learning with a focus on developing innovative approaches for natural language processing and automation. Her areas of interest include data mining, machine learning, natural language processing, and software testing.

Dr. A. Sivaramakrishnan completed his M.E. CSE from Anna University and his PhD in Computer Science from VELS University in April 2015, Chennai, Tamil Nadu, India. He is an Associate Professor of Computer Science at KL University, Andhra Pradesh, India. He has published research articles.

Dr. Balambigai Subramanian received her B.E. in Electronics and Communication Engineering from Bharathiar University with distinction, M.E. in Applied Electronics with distinction, and Ph.D. from Anna University, Chennai. She has 62 publications in reputed national and international journals/conferences. She is currently a Professor in the Department of ECE at Karpagam College of Engineering. Biomedical engineering and computer network developments keep her fascinated with conducting further research in these domains.

K. Venkataramana was born in Srikalahasti, Andhra Pradesh, India, on March 28, 1985. Since 2008, he has worked as an Associate Professor in the Department of CSE, QIS College of Engineering and Technology, Ongole, India. His research interests are MANETs and cloud computing.