FILE TERAKHIR - Turnitin.docx

by Jaidyn Sylvester

Submission date: 12-Nov-2025 03:32AM (UTC-0500)
Submission ID: 2811585895
File name: FILE_TERAKHIR_-_Turnitin.docx (418.13K)
Word count: 4360
Character count: 26867

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

COMPARATIVE ANALYSIS OF HYPERPARAMETER OPTIMIZATION TECHNIQUES ON LIGHTGBM FOR ASTHMA PREDICTION

Zulfian Azmi[1*]; Rina Julita[2]; Novica Irawati[3]; Sofyan Pariyasto[4]; Ellanda Purwawijaya[5]

Information Systems, STMIK Triguna Dharma[1]
Informatics , Universitas Dehasen Bengkulu[2]
Information Systems , Universitas Royal [3]
Medical Informatics, Sekolah Tinggi Ilmu Kesehatan Mitra Sejati[4]
Informatics, Universits Battuta[5]
zulfian.azmi@gmail.com[1*]; rinajj72@gmail.com[2]; novicairawati11@gmail.com[3]; spariyasto@gmail.com[4]; ellanda.purwa.wijaya@gmail.com[5]

Abstract— *This study presents a comparative study of hyperparameter optimization methods applied to the Light Gradient Boosting Machine (LightGBM) algorithm for asthma prediction. Traditional machine learning models often face limitations in accuracy and generalization capabilities due to suboptimal hyperparameter configurations. To address these challenges, this study evaluates and compares four approaches: Default LightGBM, RandomizedSearchCV, Optuna Optimization, and Bayesian Optimization. Experimental results show that Bayesian Optimization provides the best performance with an accuracy of 78%, a precision of 0.7778, a recall of 0.7778, an F1-score of 0.7778, and an ROC-AUC of 0.975. These findings emphasize the importance of selecting an appropriate optimization strategy to improve model performance in clinical prediction tasks. Overall, this study confirms the effectiveness of Bayesian Optimization in improving the predictive capabilities of LightGBM and provides an important contribution to the development of decision support systems in healthcare, particularly in the diagnosis and management of asthma.*

Keywords: *Asthma prediction, LightGBM, Hyperparameter optimization, Bayesian optimization, Machine learning.*

Intisari— *Penelitian ini menyajikan studi komparatif mengenai metode optimisasi hyperparameter yang diterapkan pada algoritma Light Gradient Boosting Machine (LightGBM) untuk prediksi penyakit asma. Model machine learning tradisional sering menghadapi keterbatasan akurasi dan kemampuan generalisasi akibat konfigurasi hyperparameter yang tidak optimal. Untuk mengatasi tantangan tersebut, penelitian ini mengevaluasi dan membandingkan empat pendekatan, yaitu Default LightGBM, RandomizedSearchCV, Optuna Optimization, dan Bayesian Optimization. Hasil eksperimen menunjukkan bahwa Bayesian Optimization memberikan kinerja terbaik dengan akurasi sebesar 78%, precision 0.7778, recall 0.7778, F1-score 0.7778, dan ROC-AUC 0.975. Temuan ini menegaskan pentingnya pemilihan strategi optimisasi yang tepat untuk meningkatkan performa model dalam tugas prediksi klinis. Secara keseluruhan, penelitian ini menegaskan efektivitas Bayesian Optimization dalam meningkatkan kemampuan prediktif LightGBM, serta memberikan kontribusi penting bagi pengembangan sistem pendukung keputusan di bidang kesehatan, khususnya dalam diagnosis dan manajemen penyakit asma.*

Kata Kunci: *Prediksi asma, LightGBM, Optimisasi hyperparameter, Bayesian Optimization, Machine learning.*

INTRODUCTION

Artificial Intelligence (AI) has become a highly influential catalyst driving transformation across multiple domains, particularly in healthcare, where machine learning (ML) models are being increasingly employed to enhance clinical decision making and disease prediction [1][2][3]. With the rapid growth of computing power and the availability of large-scale medical data, AI-based methods offer the potential to improve diagnostic accuracy, reduce human bias, and provide more efficient and cost-effective healthcare solutions [4][5][6]. Among various diseases, Globally, asthma impacts over 300 million individuals, making it one of the most common chronic respiratory illnesses [7][8][9][10][11] is a key focus area where advanced computational approaches can make significant contributions to early detection and management.

Asthma involves chronic airway inflammation, variable airflow restriction, and repeated occurrences including symptoms such as wheezing, coughing, and difficulty breathing [12][13][14][15]. Conventional diagnostic methods, such as spirometry and peak expiratory flow measurement, are still considered the gold standard in clinical practice. However, these methods often face challenges, including dependence on patient cooperation, limited sensitivity in detecting changes in the small airways, and potential delays in diagnosis [16][17][18][19][20]. Therefore, researchers are increasingly exploring non-invasive AI-based alternatives, including acoustic analysis of breath sounds and computational modeling of medical records, to improve the efficiency and objectivity of asthma detection.

Several machine learning algorithms, such as Support Vector Machine (SVM), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM), have shown potential in classifying respiratory diseases using both structured and unstructured data [21][22][23][24]. Previous studies have demonstrated that ensemble learning methods combined with acoustic biomarkers or clinical variables can achieve better predictive performance compared to traditional diagnostic tools. For example, Lyu et al. developed an XGBoost-based model for asthma classification using non-invasive sound biomarkers, which demonstrated high accuracy and good generalization ability across multiple validation cohorts [21]. Meanwhile, Wang et al. proposed an ensemble framework for detecting acute exacerbations of chronic obstructive pulmonary disease (AECOPD) by leveraging LightGBM and Bayesian optimization, successfully improving prediction reliability in complex and imbalanced clinical datasets [25]. In another study, Muthevi and Mutyam introduced an enhanced LightGBM model (HY_OptGBM) optimized using the OPTUNA framework for predicting coronary heart disease (CHD). This study demonstrated that hyperparameter tuning and loss function modification can significantly improve prediction accuracy and training efficiency. Furthermore, the integration of a Voting Classifier combining Random Forest and AdaBoost achieved an impressive 99% accuracy, confirming the power of ensemble learning and hyperparameter optimization in medical prediction tasks [26].

Despite these advances, a challenge remains hyperparameter optimization in ML models, which significantly impacts generalizability, accuracy, and computational efficiency. LightGBM, a boosting framework known for its speed and accuracy, is highly sensitive to hyperparameter settings [27][28][29]. Techniques such as Bayesian Optimization, Optuna, and RandomizedSearchCV have emerged as optimal configuration search strategies [30][31], but comparative studies on the effectiveness of these methods in the context of asthma prediction are still limited.

This study attempts to address this gap by conducting a comparative analysis of various hyperparameter optimization techniques in LightGBM for asthma prediction. Using a curated asthma dataset, this research assesses and contrasts the performance of Bayesian Optimization, Optuna, RandomizedSearchCV, and the default configuration of LightGBM. This study not only provides empirical insights into the balance among accuracy, robustness, and computational efficiency of various optimization methods but also confirms the potential of optimized LightGBM as a reliable predictive tool to support clinical decision-making in asthma diagnosis and management.

MATERIALS AND METHODS

This study aimed to investigate the comparison of hyperparameter optimization techniques in the LightGBM algorithm in predicting asthma. The classification results are expected to provide more comprehensive insights into the effectiveness of each optimization method in improving the performance of the LightGBM model to support data-driven diagnosis.

A. *Method of collecting data*

The research data was obtained from an asthma dataset available on Kaggle. This dataset contains several clinical variables relevant to distinguishing patients with and without asthma

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

risk. A total of 900 data samples were used. Before being used for model training, the dataset underwent several preprocessing stages. This process included data processing involving cleansing and imputation of missing values, and normalizing numeric variables to ensure each feature was on a balanced scale. Missing values were handled using median imputation to minimize the influence of outliers and maintain data consistency across features. This ensured data quality and increased accuracy of the results. Although the dataset contains only 900 samples, it provides a balanced and representative subset for benchmarking optimization techniques. However, this sample size may not fully capture the heterogeneity of real-world asthma data, which should be addressed in future studies with larger datasets. The relatively limited sample size could restrict the extent to which the findings can be generalized; thus, future research should validate the findings on larger and more diverse datasets. Examples of sample data used in this study are shown in Table 1.

Table 1. Asthma Data Sample

No.	X1	X2	X3	X4	X5	...	Y
1.	No	28.13	Yes	No	No	...	No
2.	Yes	23.63	Yes	No	No	...	No
3.	Yes	37.42	No	No	No	...	No
4.	Yes	21.95	No	No	No	...	No
5.	No	38.41	Yes	No	No	...	No
6.	No	26.61	No	No	No	...	No
7.	No	23.56	Yes	No	No	...	Yes
8.	No	30.04	No	No	No	...	Yes
9.	No	23.3	No	No	No	...	No
10.	No	33.52	Yes	No	No	...	Yes
11.	No	28.32	No	No	Yes	...	No
12.	No	33.45	No	No	No	...	No
13.	Yes	33.28	Yes	No	No	...	Yes
14.	Yes	24.27	Yes	No	Yes	...	No
15.	No	23.38	Yes	No	No	...	No
16.	No	24.53	No	No	No	...	No
17.	No	23.63	No	No	No	...	No
18.	No	31.09	Yes	No	No	...	Yes
...	...	...	...	...	...	...	...
900.	No	46.56	No	No	No	...	No

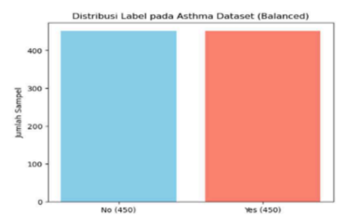
Source : (Research Results, 2025)

Table 2. Asthma Prediction Dataset Structure

	Attribute Name	Description
X1	HeartDisease	History of heart disease (Yes/No).
X2	BMI	Body Mass Index (BMI) is the result of calculating weight/height.
X3	Smoking	Smoking status, whether the individual has ever smoked or not.
X4	AlcoholDrinking	Alcohol consumption status (Yes/No).
X5	Stroke	History of stroke (Yes/No).
X6	PhysicalHealth	Total number of days in the previous month characterized by poor physical condition.
X7	MentalHealth	Total number of days in the previous month characterized by poor Mental condition.
X8	DiffWalking	Difficulty walking or climbing stairs (Yes/No).
X9	Sex	Gender (Male/Female).
X10	AgeCategory	Age groups (e.g.: 18–24, 25–29, etc.).
X11	Race	Respondent's race/ethnicity (e.g.: White, Black, Asian, etc.).
X12	Diabetic	Diabetic status (Yes/No).
X13	PhysicalActivity	Physical activity in the last 30 days (Yes/No).
X14	GenHealth	Overall health evaluation categorized as Excellent, Very Good, Good, Fair, or Poor.
X15	SleepTime	Average sleep time per night (hours).
Y	Asthma	Asthma status (Yes/No) – target label in this study.

Source : (Research Results, 2025)

The data sample taken for this study was balanced, as seen in Figure 1. Figure 1 shows the distribution of asthma and non-asthma cases, confirming a balanced data set (50% positive, 50% negative).



Source : (Research Results, 2025)

Figure 1. Balanced Data

B. Light Gradient Boosting Machine (LightGBM) algorithm

The Light Gradient Boosting Machine (LightGBM) is a gradient boosting decision tree-based algorithm developed to achieve high computational speed while maintaining optimal prediction accuracy. This algorithm uses a leaf-wise growth strategy, allowing for more efficient decision tree branch separation compared to level-wise methods. LightGBM is also capable of handling large-scale datasets, supports categorical data, and relies heavily on hyperparameter configuration to achieve maximum performance. In this study, LightGBM was used as the primary model for asthma classification. The LightGBM workflow is shown in Figure 2.

Source : (Nirajan Khadka, 2023)

Figure 2. LightGBM Architecture and Process Flow

Figure 2 shows the main stages in the LightGBM algorithm. The workflow commences with data preprocessing, a stage that focuses on cleaning, normalizing, and preparing the data. The data is then divided into subsets for training and testing (Splitting the Data). The next stage involves Initialization to initialize the model's initial parameters, followed by Splitting Nodes and Building a Tree to build a decision tree structure based on a leaf-wise growth strategy. Once the tree is formed, the algorithm performs Gradient Boosting, which corrects prediction errors by combining weak tree models. This process is complemented by Regularization to prevent overfitting. The final stage is Prediction, where the model generates predictions for new data, followed by Model Evaluation using performance metrics. If the results are not optimal, Parameter Tuning is performed by adjusting hyperparameters until the model achieves optimal performance.

C. *Hyperparameter Optimization Stages*

Hyperparameter tuning was carried out to achieve the optimal configuration that could improve LightGBM performance. This study used four approaches: the default LightGBM configuration, RandomizedSearchCV, Bayesian Optimization, and Optuna Optimization. The main hyperparameters optimized included the number of leaves (num_leaves), maximum tree depth (max_depth), learning rate (learning_rate), number of boosting rounds (n_estimators), and the minimum number of samples in leaf formation (min_child_samples).

D. *Model Evaluation*

Model performance was assessed through conventional classification metrics such as accuracy, precision, recall, and F1-score, which are defined as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

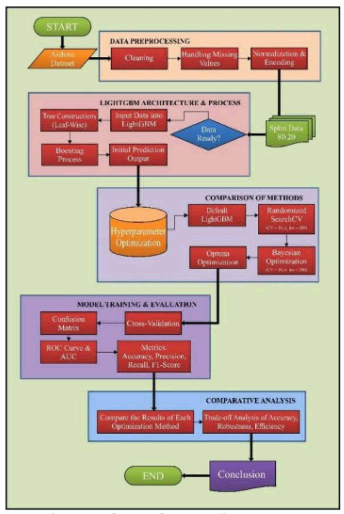
$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F1 - Score} = \frac{2\ x\ Precision\ x\ Recall}{Precision+Recal} \quad (4)$$

In addition, ROC Curve and AUC value are used to measure the model's capability to differentiate between positive and negative categories, while Confusion Matrix is used to analyze the distribution of correct and incorrect predictions.

E. *Research Flow*

This research process begins with data collection and pre-processing, followed by the application of the LightGBM algorithm. Next, hyperparameter optimization is performed using various methods, followed by an evaluation phase using classification metrics. Finally, the results of each optimization method are compared to determine the most effective approach for asthma prediction. The overall research process is visualized in Figure 3.



Source : (Research Results, 2025)

Figure 3. Research Flow Diagram

Diagram 3 illustrates the overall research flow for the asthma prediction study using LightGBM with various hyperparameter optimization techniques. The study began with the collection of the Asthma Dataset, followed by the Data Preprocessing stage, which included data cleaning, handling missing values, normalization, and encoding categorical variables. Categorical variables were encoded using Label Encoding, which is natively supported by LightGBM. This approach ensures efficient computation while maintaining feature interpretability. Subsequently, the data was partitioned into two sets: training and testing.

The next stage is LightGBM Architecture & Process, where data is fed into the LightGBM algorithm to construct a decision tree, run the boosting process, and generate initial predictions. After the base model is formed, Hyperparameter Optimization is performed, which is the focus of this research. Four approaches are compared: Default LightGBM, RandomizedSearchCV, Bayesian Optimization, and Optuna Optimization.

Each optimization method was tested using Model Training & Evaluation using cross-validation techniques. Model evaluation was performed based on accuracy, precision, recall, F1-score, ROC Curve, and AUC metrics. The results of each method were then analyzed in the Comparative Analysis stage, where the performance of the four approaches was compared to examine the trade-offs between accuracy, robustness, and computational efficiency.

The final stage of this research is Conclusion, where conclusions are drawn regarding the most effective hyperparameter optimization method to improve LightGBM performance in asthma prediction.

RESULTS AND DISCUSSION

Based on the results of training, validation, and testing the model using different hyperparameter configurations, varying performance was obtained for asthma prediction. The main objective of this experiment was to evaluate the impact of various hyperparameter optimization techniques on the accuracy and generalization of the LightGBM model.

A. Result

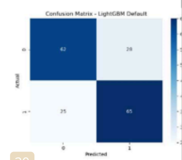
The study results compare the performance of LightGBM across four scenarios: Default, RandomizedSearchCV, Optuna Optimization, and Bayesian Optimization. The evaluation was conducted using accuracy, precision, recall, F1-score, and ROC-AUC metrics.

Under the default configuration, the proposed LightGBM model attained an accuracy of 0.70. These results serve as a baseline demonstrating LightGBM's capabilities without hyperparameter optimization.

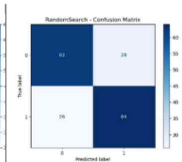
With RandomizedSearchCV, the accuracy value obtained was 0.7 with an ROC-AUC of 0.7527. Although there was an increase in model discrimination in the ROC-AUC, the accuracy and F1-score performance were actually lower than the default configuration.

On Optuna Optimization, the model produced an accuracy of 0.7278 with a precision of 0.7114, a recall of 0.7667, an F1-score of 0.7381, and an ROC-AUC of 0.7698. The higher AUC value compared to RandomizedSearchCV indicates Optuna's superior hyperparameter space exploration capabilities.

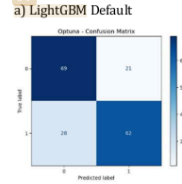
Meanwhile, Bayesian Optimization provided the best performance with achieved an accuracy, precision, recall, and F1-score of 0.78 each, and the highest ROC-AUC among all methods, namely 0.975. In the Confusion Matrix, it can be seen that Bayesian Optimization is superior in balancing the predictions of the "No" and "Yes" classes, although the recall of the "Yes" class is still lower than the "No" class.



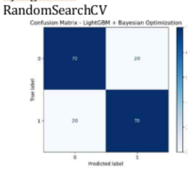
a) LightGBM Default



b) LightGBM + RandomSearchCV



c) LightGBM + Optuna



d) LightGBM + Bayesian Optimization

Source : (Research Results, 2025)

Figure 4. Confusion Matrix LightGBM per Optimization Method

$$Akurasi = \frac{62+65}{62+65+25+28} = \frac{127}{180} \approx 0.7055$$

$$Presisi = \frac{62}{62+25} = \frac{62}{87} \approx 0.7127$$

$$Recall = \frac{62}{62+28} = \frac{62}{90} \approx 0.6889$$

$$F1\ Score = \frac{2.(0.7127\ x\ 0.6889)}{0.7127+0.6889} =$$

$$Akurasi = \frac{62+64}{62+64+26+28} = \frac{126}{180} \approx 0.7$$

$$Presisi = \frac{62}{62+26} = \frac{62}{88} \approx 0.7046$$

$$Recall = \frac{62}{62+28} = \frac{62}{90} \approx 0.6889$$

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

$$\frac{0.9819}{1.4016} \approx 0.7006$$

a) LightGBM Default

$$F1\ Score = \frac{2.(0.7046\ x\ 0.6889)}{0.7046 + 0.6889} = \frac{0.9708}{1.3935} \approx 0.6967$$

b) LightGBM + RandomSearchCV

$$Akurasi = \frac{69+62}{69+62+28+21} = \frac{131}{180} \approx 0.7278$$

$$Presisi = \frac{69}{69+28} = \frac{69}{97} \approx 0.7114$$

$$Recall = \frac{69}{69+21} = \frac{69}{90} \approx 0.7667$$

$$F1\ Score = \frac{2.(0.7114\ x\ 0.7667)}{0.7114 + 0.7667} = \frac{1.0909}{1.4781} \approx 0.7381$$

c) LightGBM + Optuna

$$Akurasi = \frac{70+70}{70+70+20+20} = \frac{140}{180} \approx 0.7778$$

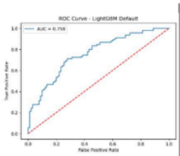
$$Presisi = \frac{70}{70+20} = \frac{70}{90} \approx 0.7778$$

$$Recall = \frac{70}{70+20} = \frac{70}{90} \approx 0.7778$$

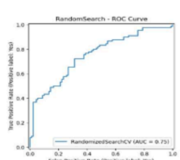
$$F1\ Score = \frac{2.(0.7778\ x\ 0.7778)}{0.7778 + 0.7778} = \frac{1.2099}{1.5556} \approx 0.7778$$

d) LightGBM + Bayesian Optimization

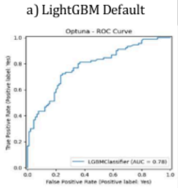
Below is the ROC Curve of all Optimization Methods including Default LightGBM:



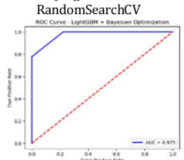
a) LightGBM Default



b) LightGBM + RandomSearchCV



c) LightGBM + Optuna



d) LightGBM + Bayesian Optimization

Source : (Research Results, 2025)

Figure 5. ROC Curve Comparison among Optimization Methods

In addition to the visualization, the performance comparison is shown in the following table:

Table 3. Comparison Results of LightGBM with Different Hyperparameter Optimization Methods

Method	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**ROC-AUC**
Default LightGBM	70%	0.7127	0.6889	0.7006	0.758
RandomSearchCV CV=10, 500 Iterasi	70%	0.7046	0.6889	0.6967	0.7527
Optuna Optimization	72%	0.7114	0.7667	0.7381	0.7698
Bayesian Optimization CV=10 500 Iterasi	78%	0.7778	0.7778	0.7778	0.975

Source : (Research Results, 2025)

Table 3 summarizes the performance of the four hyperparameter optimization methods used in LightGBM. The results show that Bayesian Optimization provides the best performance across all key metrics, specifically with an accuracy of 0.78 and a peak precision of 0.7800 with a training time taken of 1 hour and 28 minutes. Optuna Optimization shows improvement over the default configuration with a higher ROC-AUC value (0.7698) with a training time of 34 minutes, although accuracy is still slightly lower. RandomizedSearchCV produces the lowest performance compared to the other two optimization methods in both accuracy and F1-score, although its ROC-AUC value (0.7527) with a training time of 13 minutes is still relatively good. The default LightGBM configuration is more stable than RandomizedSearchCV with a training time of only 4 minutes, but it is unable to outperform the Bayesian results. Thus, this table confirms that the Bayesian model-based optimization approach is more effective than both random search and the default configuration.

B. Discussion

This study compared four hyperparameter optimization approaches to LightGBM for asthma prediction. Results showed that Bayesian optimization performed best compared to other methods. With an accuracy of 0.78 and relatively balanced precision and recall values, Bayesian optimization proved more effective in finding optimal hyperparameter configurations.

Compared to Optuna, Bayesian performed better, especially in positive class recall, despite both being based on sequential model-based optimization. Meanwhile, RandomizedSearchCV tended to produce less stable results because random search often failed to find the best parameter combination. The default LightGBM actually outperformed RandomizedSearchCV in terms of accuracy, although it was less able to improve ROC-AUC.

Overall, these results confirm that the choice of hyperparameter optimization method significantly impacts model performance. Bayesian optimization emerged as the most consistent and accurate approach for asthma prediction, and therefore can be recommended for the development of machine learning-driven systems for clinical decision-making. Beyond metric comparisons, the differences in performance can be attributed to each optimizer's search mechanism. Bayesian Optimization's probabilistic approach allows it to explore and exploit the hyperparameter space more effectively, leading to better generalization. In contrast, RandomizedSearchCV's random exploration may overlook optimal regions, resulting in less consistent outcomes.

CONCLUSION

Based on the results of a study comparing four methods, one default and three hyperparameter optimization methods in the LightGBM algorithm for asthma prediction, it can be concluded that Bayesian Optimization showed the best performance compared to other methods. Bayesian Optimization with a 10-fold cross-validation configuration and 500 iterations managed to achieve an accuracy of 78%, a precision of 0.7778, a recall of 0.7778, an F1-score of 0.7778, and the highest ROC-AUC value of 0.975.

Meanwhile, the Optuna Optimization method provided quite good results with an accuracy of 72%, a precision of 0.7114, a recall of 0.7667, an F1-score of 0.7381, and an ROC-AUC of 0.7698. This method excels in the recall aspect, although overall it is still below Bayesian Optimization. On the other hand, RandomizedSearchCV with a 10-fold cross-validation configuration and 500 iterations produced an accuracy of 70% with relatively lower performance compared to Bayesian and Optuna. Meanwhile, Default LightGBM showed baseline results with an accuracy of 70% and an ROC-AUC of 0.758.

Bayesian optimization's success in achieving better performance can be attributed to its ability to more efficiently explore and exploit hyperparameter space using a probabilistic model approach. This suggests that selecting the right hyperparameter optimization method substantially enhances the performance of the LightGBM model in predicting asthma. One limitation of this study is the absence of external validation through independent datasets or actual clinical data. Future work should include external validation to confirm the robustness and clinical applicability of the optimized LightGBM models.

Overall, this study confirms that Bayesian optimization is the most effective and consistent hyperparameter optimization method, and therefore can be recommended for use in the development of machine learning-based asthma prediction systems. With its more stable and accurate performance, this approach has the potential to make a significant contribution to supporting decision support systems in the healthcare sector.

REFERENCE

[1] S. Maleki Varnosfaderani and M. Forouzanfar, "The Role of AI in Hospitals and Clinics: Transforming Healthcare in the 21st Century," *Bioengineering*, vol. 11, no. 4, pp. 1–38, 2024, doi: 10.3390/bioengineering11040337.

[2] A. Husnain, S. Rasool, A. Saeed, A. Yousaf Gill, and H. Khawar Hussain, "AI'S Healing Touch: Examining Machine Learning's Transformative Effects On Healthcare," *J. World Sci.*, vol. 2, no. 10, pp. 1681–1695, 2023, doi: 10.58344/jws.v2i10.448.

[3] S. Balakrishna and V. Kumar Solanki, "A Comprehensive Review on AI-Driven Healthcare Transformation," *Ing. Solidar.*, vol. 20, no. 2, pp. 1–30, 2024, doi: 10.16925/2357-6014.2024.02.07.

[4] N. N. Khanna *et al.*, "Economics of Artificial Intelligence in Healthcare: Diagnosis vs. Treatment," *Healthc.*, vol. 10, no. 12, 2022, doi: 10.3390/healthcare10122493.

[5] D. C. Das and M. S. Akter, "AI-Powered Multimodal Diagnostics in Modern Healthcare : A Shift Toward Integrative Intelligence," *Asia Pacific J. Med. Innov.*, vol. 2, no. 3, 2025, doi: https://doi.org/10.70818/apjmi.v02i03.027.

[6] Z. Ahmad, S. Rahim, M. Zubair, and J. Abdul-Ghafar, "Artificial intelligence (AI) in

medicine, current applications and future role with special emphasis on its potential and promise in pathology: present and future impact, obstacles including costs and acceptance among pathologists, practical and philosoph," *Diagn. Pathol.*, vol. 16, no. 1, pp. 1–16, 2021, doi: 10.1186/s13000-021-01085-4.

[7] M. Caminati *et al.*, "Biologics and global burden of asthma: A worldwide portrait and a call for action," *World Allergy Organ. J.*, vol. 14, no. 2, p. 100502, 2021, doi: 10.1016/j.waojou.2020.100502.

[8] P. M. Pitrez, "The challenges of asthma care in low-and middle-income countries: what's next?," *J. Bras. Pneumol.*, vol. 49, no. 3, pp. 2–3, 2023, doi: 10.36416/1806-3756/e20230215.

[9] A. S. Listyoko, R. Okazaki, T. Harada, G. Inui, and A. Yamasaki, "Exploring the association between asthma and chronic comorbidities: impact on clinical outcomes," *Front. Med.*, vol. 11, no. January, 2024, doi: 10.3389/fmed.2024.1305638.

[10] A. A. Faniyi *et al.*, "Addressing the asthma crisis in Africa: challenges, strategies, and recommendations for improved management," *Egypt. J. Intern. Med.*, vol. 36, no. 1, 2024, doi: 10.1186/s43162-024-00340-6.

[11] R. Hurtado-Ruzza *et al.*, "Asthma, much more than a respiratory disease: Influence of depression and anxiety," *Rev. Assoc. Med. Bras.*, vol. 67, no. 4, pp. 571–576, 2021, doi: 10.1590/1806-9282.20201066.

[12] S. Sharma, N. Tasnim, K. Agadi, U. Asfeen, and J. Kanda, "Vulnerability for Respiratory Infections in Asthma Patients: A Systematic Review," *Cureus*, vol. 14, no. 9, pp. 6–11, 2022, doi: 10.7759/cureus.28839.

[13] R. D. Wagh, P. Khade, and Hingane, "Asthma: A Comprehensive Outlook," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 10, no. 1, pp. 691–698, 2022, doi: 10.22214/ijraset.2022.39893.

[14] D. Toumpanakis and O. S. Usmani, "Small airways in asthma: Pathophysiology, identification and management," *Chinese Med. J. Pulm. Crit. Care Med.*, vol. 1, no. 3, pp. 171–180, 2023, doi: 10.1016/j.pccm.2023.07.002.

[15] E. Y. A. Qaid and I. Long, "Asthma unravelled: a comprehensive review of epidemiology, phenotypes, pathophysiology, and emerging therapies," *Egypt. J. Bronchol.*, vol. 19, no. 1, 2025, doi: 10.1186/s43168-025-00443-w.

[16] V. E. Georgakopoulou, D. A. Spandidos, and A. Corlateanu, "Diagnostic tools in respiratory medicine (Review)," *Biomed. Reports*, vol. 23, no. 1, pp. 1–13, 2025, doi: 10.3892/br.2025.1990.

[17] R. Nooreldeen and H. Bach, "Current and future development in lung cancer diagnosis," *Int. J. Mol. Sci.*, vol. 22, no. 16, 2021, doi: 10.3390/ijms22168661.

[18] S. Sarkar, U. Jadhav, B. Ghewade, S. Sarkar, and P. Wagh, "Oscillometry in Lung Function Assessment: A Comprehensive Review of Current Insights and Challenges," *Cureus*, vol. 15, no. 10, 2023, doi: 10.7759/cureus.47935.

[19] M. A. Almeshari, J. Stockley, and E. Sapey, "The diagnosis of asthma. Can physiological tests of small airways function help?," *Chron. Respir. Dis.*, vol. 18, pp. 1–12, 2021, doi: 10.1177/14799731211053332.

[20] M. Xia *et al.*, "Expert consensus on difficult airway assessment," *Hepatobiliary Surg. Nutr.*, vol. 12, no. 4, pp. 545–566, 2023, doi: 10.21037/hbsn-23-46.

[21] Y. Lyu *et al.*, "Non-invasive acoustic classification of adult asthma using an XGBoost model with vocal biomarkers," *Sci. Rep.*, vol. 15, no. 1, pp. 1–28, 2025, doi: 10.1038/s41598-025-14645-1.

[22] Q. Qian, J. Wu, J. Wang, H. Sun, and L. Yang, "Prediction models for aki in icu: A comparative study," *Int. J. Gen. Med.*, vol. 14, pp. 623–632, 2021, doi: 10.2147/IJGM.S289671.

[23] X. Qi, S. Wang, C. Fang, J. Jia, L. Lin, and T. Yuan, "Machine learning and SHAP value interpretation for predicting comorbidity of cardiovascular disease and cancer with

dietary antioxidants," *Redox Biol.*, vol. 79, no. 12, p. 103470, 2025, doi: 10.1016/j.redox.2024.103470.

[24] L. H. Lai *et al.*, "The Use of Machine Learning Models with Optuna in Disease Prediction," *Electron.*, vol. 13, no. 23, pp. 1–20, 2024, doi: 10.3390/electronics13234775.

[25] X. Wang *et al.*, "A Joint Ensemble Framework for the Detection of Acute Exacerbations in Chronic Obstructive Pulmonary Disease," *Res. Sq.*, pp. 1–19, 2023, doi: 10.21203/rs.3.rs-3712629/v1.

[26] A. K. Muthevi and V. M. Mutyam, "Using an Enhanced LightGBM Model to Predict Coronary Heart Disease: Performance Evaluation and Comparison," *Nanotechnol. Perceptions*, vol. 20, no. S14, pp. 2446–2457, 2024, doi: 10.62441/nano-ntp.v20is14.164.

[27] K.-T. Nguyen, T.-N. Tran, and H.-T. Nguyen, "Research on the Influence of Genetic Algorithm Parameters on XGBoost in Load Forecasting," *Eng. Technol. Appl. Sci. Res.*, vol. 14, no. 6, pp. 18849–18854, 2024, doi: 10.48084/etasr.8863.

[28] R. Sibindi, R. W. Mwangi, and A. G. Waititu, "A boosting ensemble learning based hybrid light gradient boosting machine and extreme gradient boosting model for predicting house prices," *Eng. Reports*, vol. 5, no. 4, pp. 1–19, 2023, doi: 10.1002/eng2.12599.

[29] H. Yang, Z. Chen, H. Yang, and M. Tian, "Predicting Coronary Heart Disease Using an Improved LightGBM Model: Performance Analysis and Comparison," *IEEE Access*, vol. 11, no. February, pp. 23366–23380, 2023, doi: 10.1109/ACCESS.2023.3253885.

[30] P. K. Sahu and T. Fatma, "Optimized Breast Cancer Classification Using PCA-LASSO Feature Selection and Ensemble Learning Strategies with Optuna Optimization," *IEEE Access*, vol. 13, no. February, pp. 35645–35661, 2025, doi: 10.1109/ACCESS.2025.3539746.

[31] M. Khaldi, Z. Alilat, H. Bendoubba, and N. Mahammed, "Hyperparameter Optimization for Malicious URL Detection: Leveraging Optuna and Random Search in Machine Learning and Deep Learning Models," *Informatica*, vol. 49, no. 27, pp. 57–68, 2025, doi: 10.31449/inf.v49i27.9106.

FILE TERAKHIR - Turnitin.docx

ORIGINALITY REPORT

16%	12%	12%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Sim University Student Paper	1%
2	Submitted to Universitas Pendidikan Indonesia Student Paper	1%
3	doctorpenguin.com Internet Source	1%
4	Thangaprakash Sengodan, Sanjay Misra, M Murugappan. "Advances in Electrical and Computer Technologies", CRC Press, 2025 Publication	1%
5	www.mdpi.com Internet Source	1%
6	Jing Li, Duan Huang, Beiping Long, Yakui Shao, Mengwei Xiao, Linhao Sun, Xusheng Li, Aiai Wang, Xuanchi Chen, Weike Li. "Prediction of Wildfire Occurrence in the Southern Forest Regions of China in the Future Scenario", Forests, 2024 Publication	1%
7	Sayaka Tomishima, Yoko Komada, Kosuke Tanioka, Isa Okajima, Yuichi Inoue. "A 2-year follow-up study on the course of at-risk delayed sleep-wake phase disorder in the young generation and its impact on daytime function", Sleep Medicine, 2025 Publication	1%

www.frontiersin.org

8	Internet Source	1 %
9	www.nature.com Internet Source	1 %
10	arxiv.org Internet Source	<1 %
11	biosciencetrends.com Internet Source	<1 %
12	bmccancer.biomedcentral.com Internet Source	<1 %
13	www.ewadirect.com Internet Source	<1 %
14	www.ncbi.nlm.nih.gov Internet Source	<1 %
15	dergipark.org.tr Internet Source	<1 %
16	eprints.cihanuniversity.edu.iq Internet Source	<1 %
17	www.genominfo.org Internet Source	<1 %
18	www.jatit.org Internet Source	<1 %
19	mdpi-res.com Internet Source	<1 %
20	Peter V. Dinh, Timothy A. Reiad, Emilio Peveri, Carolyn Marquis, Joseph A. Gil. "A 20-Year National Decline in Wrist Fractures: Unraveling Trends and Persistent Challenges", Hand Surgery and Rehabilitation, 2025 Publication	<1 %

21 Yosua Satria Bara Harmoni, Kartika Maulida Hindrayani, Dwi Arman Prasetya. "Optimizing Categorical Boosting Model with Optuna for Anti-Tuberculosis Drugs Classification", Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics, 2025
Publication
<1 %

22 ir.lib.uwo.ca
Internet Source
<1 %

23 journal.universitasbumigora.ac.id
Internet Source
<1 %

24 Ashish Juneja, Anil Joseph, Dasaka S. Murty. "GeoVadis - The Future of Geotechnical Engineering (Volume 1)", CRC Press, 2025
Publication
<1 %

25 Shihua Luo, Tianxin Chen. "Two Derivative Algorithms of Gradient Boosting Decision Tree For Silicon Content in Blast Furnace System Prediction", IEEE Access, 2020
Publication
<1 %

26 sd.fuv.edu.br
Internet Source
<1 %

27 Gloria A. Odiaga, Newton Masinde, Castro Yoga. "Cybersecurity Modeling: Application of Optimization in Machine Learning-Based Detection Systems", Journal of Computer and Communications, 2025
Publication
<1 %

28 Pathinettampadian Gurusamy, Murugesan Devarajan Vijayakumar, Sathiyamoorthi Ramalingam. "Advances in Additive Manufacturing Technologies - Proceedings of 6th International Conference on Advances in
<1 %

Additive Manufacturing Technologies (ICAAMT 2024), December 19-20th, 2024, Chennai, India Shaping the Future of Digital Fabrication and Smart Manufacturing", CRC Press, 2026
Publication

29	S. Giriprasad, P. Prabhu, T. Saminathan, M. Lalli, T. Malathi, A. Manikandan. "chapter 2 Integrating Multiple ML and Optimization Techniques for Superior Disease Prediction", IGI Global, 2025 Publication	<1 %
30	Saeid Janizadeh, Trang Thi Kieu Tran, Sayed M. Bateni, Changhyun Jun, Dongkyun Kim, Clay Trauernicht, Essam Heggy. "Advancing the LightGBM approach with three novel nature-inspired optimizers for predicting wildfire susceptibility in Kauaʻi and Molokaʻi Islands, Hawaii", Expert Systems with Applications, 2024 Publication	<1 %
31	Utsho Dey, Md. Sakhawat Hossain Rabbi, Md. Abrar Hamim, Md. Tarek Habib. "Chapter 14 Tailored House Price Prediction Insights for Dhaka and Chittagong City", Springer Science and Business Media LLC, 2025 Publication	<1 %
32	Xiangxin Zhao, Xin Wang. "Unsupervised domain adaptation based fracture segmentation method for core CT images", Expert Systems with Applications, 2025 Publication	<1 %
33	fastercapital.com Internet Source	<1 %
34	journal.ugm.ac.id Internet Source	<1 %

35	jutif.if.unsoed.ac.id Internet Source	<1 %
36	openconstructionbuildingtechnologyjournal.com Internet Source	<1 %
37	"Biomedical Imaging", Springer Science and Business Media LLC, 2024 Publication	<1 %
38	ojs3.unpatti.ac.id Internet Source	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		