Journal of Electrical Engineering and Computer (JEECOM)
Vol. xx, No. xx (20xx), DOI: 10.33650/jeecom.v4i2
p-ISSN: 2715-0410; e-ISSN: 2715-6427

Health Risk Level Prediction for Hajj Pilgrims Using Random Forest and Bayesian Optimization (Case Study: Hajj Pilgrims of Balikpapan Embarkation)

Luthfi Bhaktiawan Husag [1], Kusrini Kusrini [2]
[1,2] Department of Informatics Engineering, AMIKOM University of Yogyakarta, Yogyakarta, Indonesia

Article Info

Article history:

Received Agust 9, 2025
Revised Agust 16, 2025
Accepted Oktober 2, 2025

Keywords:

Health Risk Prediction
Hajj Pilgrims
Random Forest
Bayesian Optimization

ABSTRACT

The Hajj pilgrimage is one of the largest religious rituals in the world, involving millions of pilgrims from various countries. The physical condition and health of pilgrims are crucial factors in ensuring the smooth execution of the Hajj. Data from the Ministry of Health indicates that the mortality risk among Hajj pilgrims tends to increase annually, particularly in the elderly age group and among pilgrims with a history of certain diseases, such as hypertension, diabetes, and heart disease. This study aims to compare the performance of a Random Forest model optimized with Bayesian optimization against a Random Forest model without any optimization in predicting the health risk level of Hajj pilgrims at the Balikpapan Embarkation. The research findings show that the Random Forest model optimized with Bayesian Optimization provides superior performance compared to the non-optimized model, using K-Fold Cross-Validation for data splitting to avoid imbalance. The optimized model achieved an average Accuracy of 88.25% and an F1 Score of 88.19%, higher than the standard model which recorded 87.99% and 87.95% on the same metrics. Although their AUC scores were nearly identical (95.46% vs. 95.47%), the improvement in accuracy and F1 Score indicates that Bayesian Optimization can produce a more balanced and accurate classification model. In conclusion, the application of Bayesian Optimization to Random Forest is proven effective for enhancing the predictive accuracy of Hajj pilgrims' health risks, potentially supporting more proactive Hajj healthcare services.

Corresponding Author:
Luthfi Bhaktiawan Husag,
AMIKOM University of Yogyakarta, Jl. Ring Road Utara Condongcatur, Yogyakarta, 55281, Indonesia
Email: luthfi.husag@students.amikom.ac.id

1. INTRODUCTION

The Hajj pilgrimage is one of the largest religious rituals in the world, involving millions of pilgrims from various countries. The physical condition and health of the pilgrims are crucial factors in ensuring the smooth execution of the Hajj. Data from the Ministry of Health indicates that the risk of mortality among Hajj pilgrims tends to increase annually, especially in the elderly group and among pilgrims with a history of certain diseases, such as hypertension, diabetes, and heart disease. The 2024 report from the PPIH Health Sector of the Balikpapan Embarkation shows that the distribution of the top 10 high-risk diseases is dominated by degenerative diseases, with hypertension and senility being the most numerous. High-risk conditions classified as degenerative diseases, such as Diabetes Mellitus (DM) and hypertension, require attention and guidance starting from the initial health examinations in their home regions. The Indonesian Hajj Health Organization, particularly at the Balikpapan Hajj Embarkation, involves a crucial third-stage health examination. This stage aims to determine the pilgrims' fitness to fly based on previous examinations and a reassessment of their health *istithaah* (ability), in accordance with international flight safety and health standards. In an effort to mitigate these health risks, the use of machine learning (ML) technology has shown great potential in predicting health risk levels. Various studies have applied ML algorithms for the early detection and classification of diseases. For instance, a study by Alhazmi [1] developed a triage prediction system for emergency units during the Hajj

period using an ML model, which demonstrated better performance compared to conventional methods. Research conducted by Gao et al [2] show the bayesian hyperparameter optimization technique demonstrated superior stability in comparison to both the grid search and random search methodologies. Within a dataset pertaining to breast cancer diagnosis achieved an accuracy rate of 94.74% alongside a sensitivity of 93.69%. Research by Zhao et al [3] show the primary finding of this study demonstrates that the proposed Bayesian optimization random forest model exhibits superior predictive performance. The model achieved a Mean Absolute Error (MAE) of 0.576 on the test set. This performance is significantly better than that of the three other comparison models, which respectively yielded MAEs of 0.607, 0.605, and 0.581. Furthermore, the study also successfully identified the variable importance of each molecular descriptor, providing deeper insight into the factors that influence the prediction.

In its development, various advancements have been made to improve the accuracy of predictive models, one of which is through hyperparameter optimization. Research by Yang et al. [4] showed that the BO-RF model outperformed other methods in accuracy and anti-noise capability for estimating channel parameters, incorporating K-fold cross-validation to reduce overfitting. Another study by Wang et al. [5] demonstrated that a similar hyperparameter optimization technique significantly improved the predictive performance of Random Forest (RF) and XGBoost models for landslide susceptibility mapping. That approach yielded an increase in AUC of 4% and 3%, respectively. An experiment by Jaya Kusuma et al. [6] also showed that optimal performance was achieved through the integration of XGBoost with a Bayesian method (BOXGB), which recorded an accuracy of 87%. In summary, the application of Bayesian Optimization has proven effective in yielding hyperparameter configurations that enhance the capabilities of machine learning models for various cases, including the prediction of pregnancy risks.

Although many studies have used machine learning (ML) algorithms for health prediction, challenges in achieving optimal detection performance remain. Many ML models, including those that are ensemble-based, often face difficulties in determining the right hyperparameters and are prone to overfitting or underfitting. Several studies show that detection accuracy is still in the 80–90% range, leaving significant room for improvement. This is especially crucial for critical applications like predicting the health risks of Hajj pilgrims, where false negatives can have fatal consequences.

To address these challenges, this study proposes the development of a health risk prediction model for Hajj pilgrims by leveraging the Random Forest algorithm, which is known for its high predictive capabilities. To tackle the challenges of hyperparameter optimization and improve model performance more systematically, this research will integrate Bayesian Optimization. This optimization approach is expected to yield a model with higher and more stable Accuracy, AUC, Precision, and Recall (F1 Score) compared to less efficient methods.

This research aims to test a method for predicting the health risk level of Hajj pilgrims using the Random Forest method with Bayesian optimization, based on data from the Balikpapan embarkation pilgrims in 2024, thereby creating an early detection system to anticipate health issues among the pilgrims.

2. METHOD

This segment delineates the methodological approaches employed to construct and assess the efficacy of the model. Essential details, including the selected methodology for data acquisition, techniques utilized for data preparation, and methodologies for data analysis, are illustrated in figure 1.

ISSN: 2715-6427

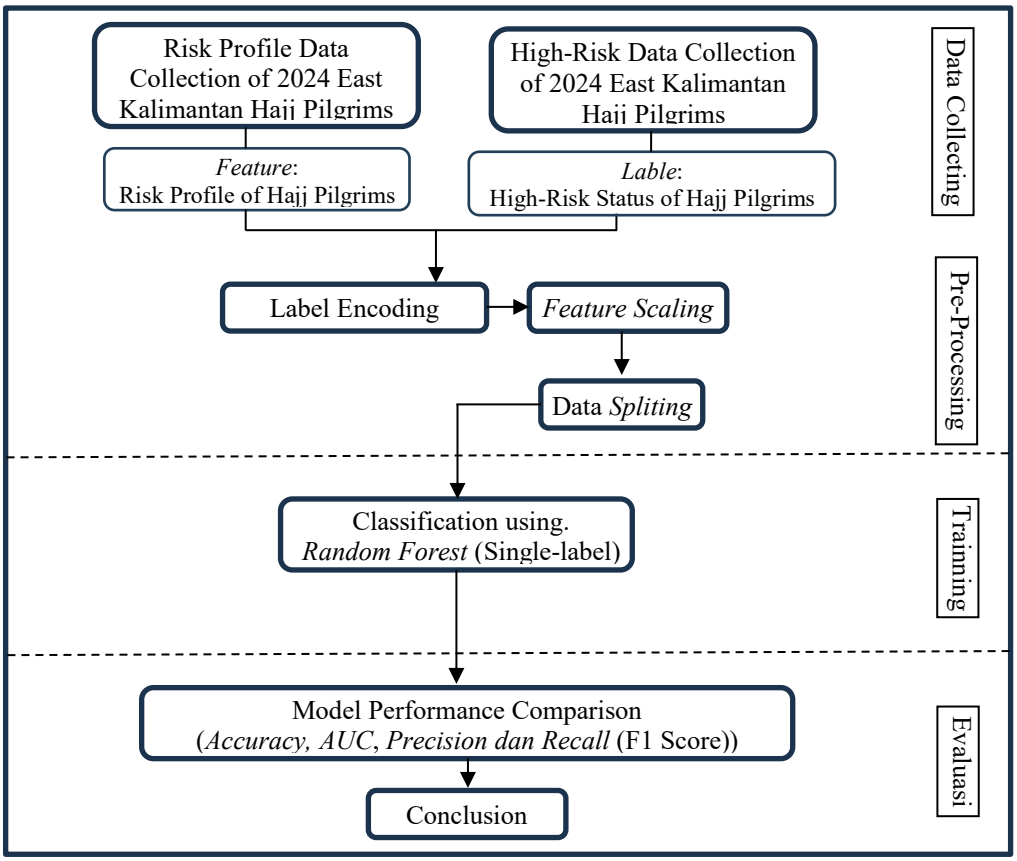

Figure 1 Research Stages

2.1. Data Collection

The data collection process was sourced from the results of the Hajj Health Epidemiological Surveillance. The data was obtained through a phased health examination of the pilgrims, starting from the sub-district, district/city, to the provincial level.

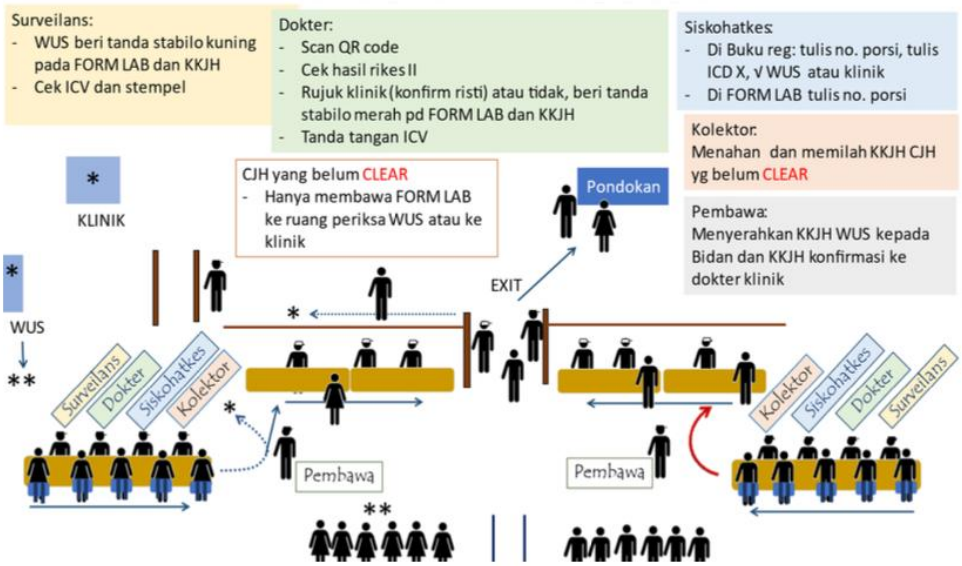

Figure 2 Dataset Collection Process

The collected dataset includes various examination results, Nomor Porsi, Nama Jemaah, Kloter, Rombongan Kloter, Regu Kloter, Umur, No. Paspor, No. Visa, Status Jemaah, Kab/Kota, Jenis Kelamin, WUS (HCG), WUS (PALPASI), ICD 1, ICD 2, ICD 3, ICD 4, ICD 5, STATUS, KR / TK, RISTI / TIDAK, LAIK / TDK LAIK, LANSA / TIDAK. This data represents the health examination results of the Hajj pilgrims for the 2024 departure year, which is the main focus of this study.

2.2. Data Preprocessing

2.2.1. Label Encoding

Label encoding is a fundamental step in preprocessing categorical data[7]. In this study, the dataset consists of twelve variables, with only one being numerical while the remaining variables are represented as strings. Consequently, it becomes essential to convert this categorical data into a numerical format. This transformation renders the categorical data processable by machine learning models that exclusively accept numerical inputs [8]. The results of the dataset process are presented in table 1.

The Label Encoding technique from the scikit-learn library is utilized. This method functions by assigning a unique integer value for each unique category within a feature. For instance, in the 'Jenis Kelamin' feature, the "Pria" category can be represented as 0 and "Wanita" as 1. This process is applied iteratively to all categorical columns identified in the dataset. The results of the dataset process are presented in table 2.

2.2.2. Standarization

Standardization helps to overcome the problem of data having different value ranges, reduces the influence of features with large scales, and facilitates further analysis processes[9]. Numerical features within a dataset often have vastly different value ranges or scales. For example, the 'Umur' feature might range from 20 to 70, while other features that have undergone Label Encoding might only range from 0 to 5. For the ‘Umur’ attribute, a change of scale is required using Z-Score Standardization. Additionally, feature scaling is not necessary for the other attributes, as their values do not represent a meaningful order or distance between categories.

To address this issue, a standardization technique is applied using StandardScaler from scikit-learn. This method transforms the data to have a distribution with a mean of 0 and a standard deviation of 1. This transformation is calculated using the z-score formula as follows Equation (1):

$$z = \frac{(x-\mu)}{\sigma} \quad (1)$$

ISSN: 2715-6427

where z is the value after scaling, x is the original data value, μ is the mean of the feature, and σ is its standard deviation. This scaling ensures that each feature contributes equally during the model training process.

2.2.3. Data Splitting

In this study, data splitting is necessary to divide the data into several folds, either 10 or 5, which serve as the training and testing data. In this process, the labels and features are also randomized into each fold using stratified K-fold, ensuring that the labels are distributed evenly to avoid imbalance.

2.3. Random Forest and Bayesian Optimization

Random forest is an ensemble-based machine learning algorithm frequently used for classification and regression[10]. Bayesian optimization is an efficient method for automatically tuning model hyperparameters, thereby enhancing the model's accuracy and efficiency. Bayesian optimization significantly improves the accuracy of random forest in various applications, such as landslide susceptibility mapping, industrial product quality prediction, stock price prediction, and agricultural parameter estimation[5][11], [12][13], [14]. The accuracy improvement ranges from 1% to 10% compared to random forest without optimization. By selecting important features and adjusting hyperparameters automatically, the model becomes more efficient and reduces computation time without sacrificing accuracy[15]. A Bayesian-optimized random forest model demonstrates better stability and predictive capability on new data or across application domains.

2.4. Model Evaluation and Comparison

Performance evaluation will be conducted by identifying the parameter values used as indicators of the model's performance achievement. The main evaluation metrics include:

Accuracy: Measures the total proportion of correct predictions.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \; X \; 100\% \tag{2}$$

Precision: Measures the proportion of instances identified as positive by the model that are genuinely positive.

$$Precision = \frac{TP}{TP+FP} \; X \; 100\% \tag{3}$$

Recall (Recall / Sensitivity): Measures the percentage of actual positive cases that are correctly identified by the model.

$$Recall = \frac{TP}{TP+FN} \; X \; 100\% \tag{4}$$

F1-score (F1): Is the harmonic mean of Precision and Recall, providing a balanced view between the two measures.

$$F1 = \frac{PRESISI \; X \; RECALL}{PRESISI+RECALL} \tag{5}$$

The accuracy measure (Equation (2)) serves as an indicator of the model's effectiveness in accurately predicting all target categories. Subsequently, the precision metric (Equation (3)) reflects the proportion of instances identified as positive by the model that are genuinely positive. Meanwhile, the recall metric (Equation (4)) measures the percentage of actual positive cases that are correctly identified by the model. The F1-score (Equation (5)) acts as a metric derived from the combination of precision and recall, providing a balanced view between these two measures, and the AUC-ROC (Area Under the Receiver Operating Characteristic Curve) is used to measure the model's ability to distinguish between positive and negative classes at various classification thresholds. This study will compare the evaluation results of the model with optimization intervention against the standard model to identify the differences and impacts of the applied optimization. Furthermore, metrics such as precision, recall, and F1-score will display values for each category or class associated with the target variable (dependent variable).

3. RESULT AND DISCUSSION

3.1. Data Pre-Processing

3.1.1. Label Encoding

Table 1 below shows a sample of the dataset before the Label Encoding process. It is evident that several columns such as 'Status Jemaah' (Pilgrim Status), 'Jenis Kelamin' (Gender), and 'STATUS' are still in string (text) format.

Table 1 Result before Label Encoding

	Umur	Status Jemaah	JK	WUS HCG	WUS PALPASI	ICD 1	ICD 2	ICD 3	ICD 4	ICD 5	STATUS	KRTK
0	42	TPHI	L	Tidak Periksa	General examination and investigation of persons without complaint and reported diagnosis	Negative	Negative	Negative	Negative	Negative	Bukan Lansia	Normal
1	50	TPIHI	L	Tidak Periksa	General examination and investigation of persons without complaint and reported diagnosis	Negative	Negative	Negative	Negative	Negative	Bukan Lansia	Normal
2	36	TKHI	P	NEGATIVE	General examination and investigation of persons without complaint and reported diagnosis	Negative	Negative	Negative	Negative	Negative	Bukan Lansia	Normal
3	43	TKHI	P	NEGATIVE	General examination and investigation of persons without complaint and reported diagnosis	Negative	Negative	Negative	Negative	Negative	Bukan Lansia	Normal
4	47	TKHI	L	NEGATIVE	General examination and investigation of persons without complaint and reported diagnosis	Negative	Negative	Negative	Negative	Negative	Bukan Lansia	Normal

After applying Label Encoding, all these categorical values were successfully converted into integer representations, as shown in Table 2. For example, in the 'Jenis Kelamin' (Gender) column, the value 'Pria' (Male) was changed to 0 and 'Wanita' (Female) to 1. A similar transformation was applied to all other non-numeric columns. The result of this process is a fully numeric dataset ready for the next preprocessing stage.

Table 2 Result after Label Encoding

ISSN: 2715-6427

	Umur	Status Jemaah	JK	WUS HCG	WUS PALPASI	ICD 1	ICD 2	ICD 3	ICD 4	ICD 5	STATUS	KR TK
0	42	6	0	1	2	61	56	43	44	13	0	6
1	50	7	0	1	2	61	56	43	44	13	0	6
2	36	4	1	0	1	61	56	43	44	13	0	6
3	43	4	1	0	1	61	56	43	44	13	0	6
4	47	4	0	1	2	61	56	43	44	13	0	6

3.1.2. Feature Scaling

Table 3 Result after Featuring Scaling

	Umur	Status Jemaah	JK	WUS HCG	WUS PALPASI	ICD 1	ICD 2	ICD 3	ICD 4	ICD 5	STATUS	KR TK
0	-0.908524	6	0	1	2	61	56	43	44	13	0	6
1	-0.246663	7	0	1	2	61	56	43	44	13	0	6
2	-1.404921	4	1	0	1	61	56	43	44	13	0	6
3	-0.825792	4	1	0	1	61	56	43	44	13	0	6
4	-0.494861	4	0	1	2	61	56	43	44	13	0	6
…	…	…	…	…	…	…	…	…	…	…	…	…
6030	-1.404921	0	1	0	1	58	27	43	44	13	0	6
6031	0.332466	0	1	1	2	70	45	43	44	13	0	6
6032	-0.081197	0	1	1	2	58	27	43	44	13	0	6
6033	0.994328	0	0	1	2	107	16	22	44	13	0	6
6034	0.497932	0	1	1	2	25	56	43	44	13	0	5

The result of this scaling in table 3 is a new distribution for the 'Umur' attribute with a mean value close to 0. A positive value indicates that the pilgrim's age is above the average, while a negative value indicates the opposite. The magnitude of the Z-score indicates how far the deviation of a pilgrim's age is from the population average in the dataset. This transformation is crucial for normalizing the contribution of the 'Umur' attribute in the model training process.

3.1.3. Data Splitting

K-fold cross-validation is a popular method for evaluating the performance of machine learning algorithms, particularly in classification and prediction. This method partitions the dataset into k nearly equal, disjoint folds, using each fold sequentially to test the model trained on the remaining k-1 folds [16]. The classification algorithm's performance is determined by averaging the accuracies across the k iterations, with the averaging assumed to occur at the fold level.

In this process, the label and feature data are also randomized into each fold using stratified K-fold, ensuring that the labels are evenly distributed to prevent imbalance.

3.1.4. Training Model

Table 4 Single Label Classification Random Forest

Fold	Single Label Classification: Random Forest		
	Accuracy	AUC Score	F1 Score
Fold 1	0.8742	0.9563	0.8740
Fold 2	0.8808	0.9605	0.8802
Fold 3	0.8775	0.9621	0.8772
Fold 4	0.9007	0.9610	0.9005

Fold 5	0.8825	0.9568	0.8821
Fold 6	0.8839	0.9563	0.8835
Fold 7	0.8756	0.9464	0.8753
Fold 8	0.8922	0.9571	0.8916
Fold 9	0.8574	0.9379	0.8572
Fold 10	0.8740	0.9522	0.8734

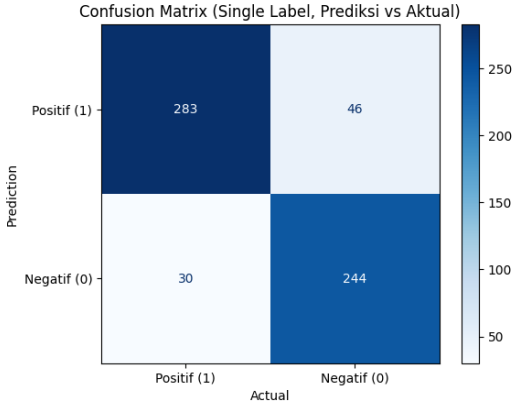

Figure 3 Confusion Matrix Single Label Classification: Random Forest

The model in table 4 and figure 3 shows a high overall accuracy rate (87.4%) and excellent precision (90.4%). High precision indicates that when the model classifies a pilgrim as at-risk (Positive), the prediction has a high level of confidence. However, special attention must be paid to the False Negative value of 46. This value indicates that the model successfully identified 86.0% of all actual Positive cases (recall), but failed to detect the remaining ones.

Table 5 Single Label Classification Random Forest and Bayesian Optimizatiom

Fold	Single Label Classification: Random Forest + Bayesian Optimizatiom		
	Accuracy	**AUC Score**	**F1 Score**
Fold 1	0.8725	0.9498	0.8721
Fold 2	0.8841	0.9650	0.8833
Fold 3	0.8990	0.9629	0.8985
Fold 4	0.8974	0.9642	0.8971
Fold 5	0.8891	0.9506	0.8887
Fold 6	0.8872	0.9559	0.8867
Fold 7	0.8839	0.9515	0.8833
Fold 8	0.8856	0.9531	0.8847
Fold 9	0.8491	0.9364	0.8487
Fold 10	0.8773	0.9564	0.8763

ISSN: 2715-6427

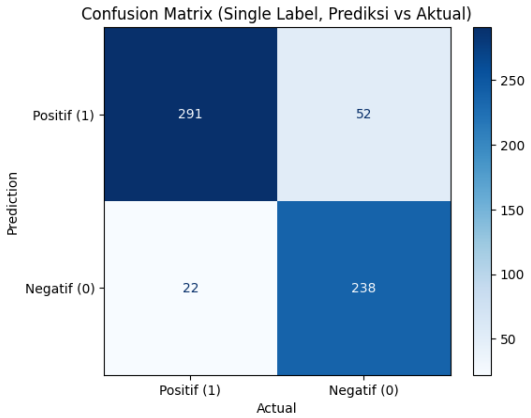

Figure 4 Confusion Matrix Single Label Classification: Random Forest and BO

Despite a minor decline in recall in figure 4, the notable rise in precision and True Positives confirms the efficacy of Bayesian optimization in enhancing model performance. The BO-RF model exhibits increased confidence and accuracy in classifying at-risk pilgrims. Accurate identification of true positive cases while minimizing false positives is vital for health risk management. The significant improvement in precision indicates that interventions based on the BO-RF model will be more precisely directed. Thus, it is evident that Bayesian Optimization has effectively improved the predictive performance of the Random Forest model in this analysis.

Table 6 Average comparison of the two models

Models	AVERAGE		
	Average Subset Accuracy	**Average AUC Score**	**Average F1 Score**
RF	0.8799	0.9547	0.8795
RF and BO	0.8825	0.9546	0.8819

From the table 6, it is evident that the application of Bayesian Optimization provides an improvement in several key performance metrics. The BO-RF model achieves an average accuracy of 88.25%, slightly higher than the standard model, which reaches 87.99%. A similar increase is also observed in the F1-Score, which rises from 87.95% to 88.19%. A higher F1-Score indicates that the optimized model has a better balance between precision and recall.

The AUC Score shows nearly identical results for both models (around 0.954). The AUC Score measures the model's ability to distinguish between positive and negative classes. This very similar result implies that both models have equivalent discriminative capabilities, yet the BO-RF model is able to translate that capability into a slightly more accurate actual classification overall.

4. CONCLUSION

Health screenings for Hajj pilgrims are a crucial step for risk mitigation. Therefore, this study proposes the BO-RF model to predict their health risk level. The study's results show that using Bayesian Optimization (BO) to fine-tune the parameters of the Random Forest (RF) model is effective. Although the improvement is not drastic, this optimization consistently enhances the model's predictive performance. In a context as critical as health risk prediction, even a slight increase in accuracy becomes highly significant. This proves that a systematic parameter search process can yield a more reliable and effective model for this classification task. Future development will focus on a comparison with other potentially superior methods, such as eXtreme Gradient Boosting, using more up-to-date real-world data.

ACKNOWLEDGEMENTS

The author wishes to thank the Health Sector of the Balikpapan Embarkation Hajj Organizing Committee (PPIH) for providing the data for this research. Deep gratitude is also extended to Prof. Dr. Kusrini, M.Kom. for her invaluable guidance, insight, and motivation throughout the process of completing this study.

REFERENCES

[1] H. N. Alhazmi, "A Prediction Triage System for Emergency Department During Hajj Period using Machine Learning Models," *IJCSNS International Journal of Computer Science and Network Security*, vol. 24, no. 7, p. 11, 2024, doi: 10.22937/IJCSNS.2024.24.7.2.

[2] L. Gao and Y. Ding, "Disease prediction via Bayesian hyperparameter optimization and ensemble learning," *BMC Res Notes*, vol. 13, no. 1, Apr. 2020, doi: 10.1186/s13104-020-05050-0.

[3] Y. Zhao and H. Xu, "Prediction of Anti-Breast Cancer Drugs Activity Based on Bayesian Optimization Random Forest," in *2023 42nd Chinese Control Conference (CCC)*, 2023, pp. 3471–3475. doi: 10.23919/CCC58697.2023.10241131.

[4] C. Yang, Y. Wang, A. Zhang, H. Fan, and L. Guo, "A Random Forest Algorithm Combined with Bayesian Optimization for Atmospheric Duct Estimation," *Remote Sens (Basel)*, vol. 15, no. 17, Sep. 2023, doi: 10.3390/rs15174296.

[5] S. Wang, J. Zhuang, J. Zheng, H. Fan, J. Kong, and J. Zhan, "Application of Bayesian Hyperparameter Optimized Random Forest and XGBoost Model for Landslide Susceptibility Mapping," *Front Earth Sci (Lausanne)*, vol. 9, Jul. 2021, doi: 10.3389/feart.2021.712240.

[6] E. Jaya Kusuma *et al.*, "OPTIMASI MODEL EXTREME GRADIENT BOOSTING DALAM UPAYA PENENTUAN TINGKAT RISIKO PADA IBU HAMIL BERBASIS BAYESIAN OPTIMIZATION (BOXGB) MACHINE LEARNING OPTIMIZATION IN DETERMINING THE MATERNAL RISK LEVEL BASED ON BAYESIAN OPTIMIZATION," vol. 12, no. 1, 2025, doi: 10.25126/jtiik.2025129001.

[7] P. Rodríguez, M. A. Bautista, J. Gonzàlez, and S. Escalera, "Beyond One-hot Encoding: lower dimensional target embedding," Jun. 2018, doi: 10.1016/j.imavis.2018.04.004.

[8] E. Jackson and R. Agrawal, "Performance Evaluation of Different Feature Encoding Schemes on Cybersecurity Logs," in *2019 SoutheastCon*, 2019, pp. 1–9. doi: 10.1109/SoutheastCon42311.2019.9020560.

[9] T. A. Runkler, "Data Preprocessing," in *Data Analytics: Models and Algorithms for Intelligent Data Analysis*, T. A. Runkler, Ed., Wiesbaden: Springer Fachmedien Wiesbaden, 2020, pp. 23–36. doi: 10.1007/978-3-658-29779-4_3.

[10] M. Schonlau and R. Y. Zou, "The random forest algorithm for statistical learning," *Stata Journal*, vol. 20, no. 1, pp. 3–29, Mar. 2020, doi: 10.1177/1536867X20909688.

[11] T. Wang *et al.*, "Random Forest-Bayesian Optimization for Product Quality Prediction with Large-Scale Dimensions in Process Industrial Cyber-Physical Systems," *IEEE Internet Things J*, vol. 7, no. 9, pp. 8641–8653, Sep. 2020, doi: 10.1109/JIOT.2020.2992811.

[12] Y. Zhang, X. Zheng, S. Yang, S. Meng, Z. Yang, and X. Fei, "A Random Forest Stock Prediction Model Based on Bayesian Optimization," *2024 7th International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pp. 42–46, 2024, doi: 10.1109/ICAIBD62003.2024.10604441.

[13] G. Rong *et al.*, "Rainfall induced landslide susceptibility mapping based on bayesian optimized random forest and gradient boosting decision tree models—a case study of shuicheng county, china," *Water (Switzerland)*, vol. 12, no. 11, pp. 1–22, Nov. 2020, doi: 10.3390/w12113066.

[14] J. Zhang *et al.*, "Enhanced Crop Leaf Area Index Estimation via Random Forest Regression: Bayesian Optimization and Feature Selection Approach," *Remote Sens (Basel)*, vol. 16, no. 21, Nov. 2024, doi: 10.3390/rs16213917.

[15] P. I. Frazier, "A Tutorial on Bayesian Optimization," Jul. 2018, [Online]. Available: http://arxiv.org/abs/1807.02811

[16] T. T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," *Pattern Recognit*, vol. 48, no. 9, pp. 2839–2846, Sep. 2015, doi: 10.1016/j.patcog.2015.03.009.