Indonesian Journal of Electrical Engineering and Computer Science
Vol. 37, No. 1, January 2025, pp. 443~451
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v37.i1.pp443-451

A predictive model for postpartum depression: ensemble learning strategies in machine learning

Winda Ayu Fazraningtyas[1,2], Bahbibi Rahmatullah[1], Husni Naparin[3], Mohammad Basit[4], Nor Asiah Razak[1]
[1]Faculty of Computing and Meta-Technology, Sultan Idris Education University, Perak, Malaysia
[2]Department of Maternity Nursing, Faculty of Health, Sari Mulia University, Banjarmasin, Indonesia
[3]Department of Information Technology, Faculty of Science and Technology, Sari Mulia University, Banjarmasin, Indonesia
[4]Department of Management in Nursing, Faculty of Health, Sari Mulia University, Banjarmasin, Indonesia

Article Info

Article history:

Received Apr 30, 2024
Revised Sep 9, 2024
Accepted Sep 29, 2024

Keywords:

Ensemble learning
Machine learning
Postnatal depression
Postpartum depression
Predictive model

ABSTRACT

Postpartum depression (PPD) presents a significant mental health challenge for mothers following childbirth. While the precise cause of this condition remains unknown, preventive measures and treatments are available. This study aims to employ ensemble learning techniques, utilizing C4.5 decision tree (DT), gradient boosting tree (GBT), and extreme gradient boosting (XGBoost), to predict the occurrences of PPD in the Banjarmasin, South Kalimantan, Indonesia. The predictive model developed encompasses a dataset comprising 317 records gathered from postpartum mothers in hospitals, community health services, and midwifery clinics (referred to as Model 1). Furthermore, resampling techniques (Model 2) were employed to address class imbalance. Additionally, feature selection including forward selection and backward elimination (Model 3) were implemented to enhance model performance. The findings reveal that XGBoost, combined with resampling methods, achieved the highest accuracy rate at 87.57%. Feature selection identified five crucial factors associated with PPD incidence: marital status, number of living children, history of depression, fear of delivery, and family relationships. The utilization of ensemble learning strategies for PPD prediction yields reliable outcomes that can be applied within clinical settings. Exploring alternative ensemble learning strategies such as random forest and adaptive boosting could further optimize model performance and warrant consideration in future research endeavours.

Corresponding Author:

Bahbibi Rahmatullah
Faculty of Computing and Meta-Technology, Sultan Idris Education University
Perak, Malaysia
Email: bahbibi@meta.upsi.edu.my

1. INTRODUCTION

Reproductive health frequently overlooks mental health considerations [1] despite ongoing surveillance revealing a significant prevalence of mental health issues. This multifaceted health condition encompasses a range from optimal well-being to the initiation of emotional distress and pain. Mental health disorders can inflict prolonged suffering on individuals [2], occurring across various environments and affecting anyone, even mothers, during childbirth. The postpartum phase emerges as the most precarious time for both the mother and the baby, often overlooked and associated with a significant number of fatalities [3]. It marks a crucial period of maternal transformation characterized by physiological and psychological changes. Postpartum depression (PPD) stands out as a severe mental health issue [4], one that can be addressed and treated yet frequently goes unnoticed and undiagnosed [5]. PPD is intricately linked to a combination of physical and emotional factors.

Machine learning (ML) has become an interesting topic in research related to early detection [6], [7]. Early detection of depressive symptoms can significantly improve depression management and reduce its negative impacts [8]. However, routine screening for PPD is often neglected due to the non-economical nature of traditional and time-consuming self-report questionnaires. Based on a systematic literature review and meta-analysis, the prevalence of PPD was found to be 17.22% in the world population, while the rate of PPD was much lower in developed or high-income countries and regions [9]. Meanwhile, in Indonesia, the prevalence of perinatal mental health incidents is not yet known [10]. Research by Fazraningtyas [11] on 88 participants who were hospitalized at the General Hospital in the Banjarmasin area revealed that 17% of postpartum mothers experienced severe depression. This figure is concerning, considering there is no reliable prediction tool for detecting PPD incidents, which impacts the handling of this case.

Studies on predicting PPD cases remain limited and have not yielded significant results. A systematic literature review approach shows that optimizing the performance of ML algorithms varies depending on the dataset and the problem to be addressed [12], [13]. In addition, a systematic literature review study revealed that the random forest (RF) algorithm consistently provided the best prediction model performance [12], with an accuracy of 79.1% in a retrospective cohort study [14]. The extreme randomized forest (XRT) [15] and logistic regression (LR) [16] models achieved a 73% accuracy in PPD prediction. Despite these efforts, the effectiveness of PPD prediction models remains suboptimal, suggesting a need for further research. Ensemble learning, as noted by Raza *et al.* [17], has shown improved performance and potential for predicting PPD.

This study investigates the use of artificial intelligent (AI), particularly ML and ensemble learning, to predict PPD occurrence and evaluate key contributing factors. It optimizes the chosen learning algorithm with various sampling methods and employs forward selection and backward elimination for feature selection. An ensemble learning model, including C4.5 decision tree (DT), gradient boosting tree (GBT), and extreme gradient boosting (XGBoost), is applied to improve model performance by providing stability to the prediction model. The goal is to enhance healthcare development and provide actionable insights for clinical settings using AI.

2. METHOD

The framework illustrated in Figure 1 delineates the consecutive phases and sequential steps employed in this research. The gathered data undergoes processing to prepare it for utilization in ML algorithms. The ML algorithms utilized in this study encompass C4.5 DT, GBT, and XGBoost. The constructed predictive model incorporates data processing, segregation, feature selection, and prediction algorithms [18]. These models are configured to optimize the performance of ML algorithms effectively utilizing resampling and feature selection methodologies.

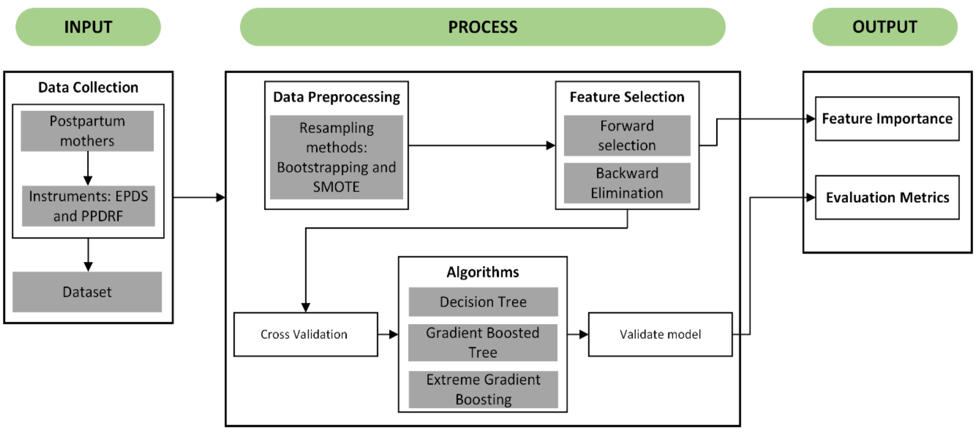

Figure 1. The framework of the study

2.1. Data collection

The dataset for this study was collected from 317 postpartum mothers at hospitals, community health services, and midwifery clinics in Banjarmasin city, using purposive sampling. The edinburgh postnatal depression scale (EPDS) and the postpartum depression risk factors (PPDRF) were utilized as measurement instruments. This questionnaire is available in Indonesian, with construct validation using confirmatory factor analysis (CFA) to obtain a factor loading score of more than 0.5, a comparative fit index (CFI) value of 0.963, and an normed fit index (NFI) of 0.942 [19]. The EPDS is a 10-item questionnaire that is easy to administer and is an effective screening tool. Each question is scored from 0 to 3. For questions 1, 2, and 4, the scoring is from 0 to 3, while for questions 3, 5, 6, 7, 8, 9, and 10, the scoring is from 3 to 0. The maximum score on this questionnaire is 30. According to Levis *et al.* [20], a threshold value of 11 or higher may indicate women who are at risk of PPD.

2.2. Data preprocessing: resampling methods

The resampling technique addresses imbalanced learning data by modifying it to achieve a balanced distribution [21]. This research employed random oversampling, including bootstrapping and SMOTE methods. Bootstrapping involves selecting n observations from an original dataset of size n with replacement, allowing some observations to be duplicated and others omitted. The process is mathematically detailed in the provided:

$$X_i^* = Randomly\ choose\ X_j\ from\ \{X_1, X_2, \ldots, X_n\} \tag{1}$$

where X_i^* represents each element in the bootstrap sample that is replicated from the original dataset.

SMOTE is a data preprocessing approach to tackle class imbalance within a dataset. This technique involves generating new synthetic observations derived from the existing samples of the minority class. Generating these synthetic training records involves randomly selecting one or more K-nearest neighbours for each instance in the minority class. The (2) elucidates the formulation of this method for creating synthetic samples:

$$X_i^* = X_i + \lambda \,.\left(X_{nnj} - X_i\right) \tag{2}$$

The factor λ, ranging between 0 and 1, plays a pivotal role in determining the distance between X_i and X_{nnj} for the creation of synthetic samples. In this study, the number of neighbours was 5.

2.3. Feature selection

Feature selection is a technique to diminish input variables in a prediction model. It achieves this by maximizing the inclusion of pertinent data while eliminating noise from the dataset. This automatic process optimizes relevant features to enhance model performance [21]. Forward selection (FS) is an iterative process that commences by identifying the feature with the highest performance concerning the target feature. In selecting feature subsets, stratification is imperative to ensure adequate representation of each class [22]. Backward elimination (BE) is a valuable tool for selecting pertinent features prior to entering the model testing phase. This algorithm initiates by testing all features and progressively eliminates non-significant features by comparing the results derived from each combination of these features [23], [24].

2.4. Cross validation

Cross-validation (CV) is a model training approach that evaluates prediction accuracy. This technique is particularly advantageous for estimating low-bias models, making it widely adopted in ML algorithms [25]. The present research employs a k-fold CV strategy, dividing the dataset into k subsets. Initially, the sample is randomly partitioned into k subsamples of equal size. Each of the k subsamples functions as validation data exactly once after the cross-validation technique is iteratively applied k times [26]. For this study, a 10-fold CV was implemented with stratified sampling.

2.5. Machine learning algorithms

The fundamental concept behind ensemble learning involves training multiple weak classifiers with training data and then combining these weak classifiers to construct a robust classifier [27], [28]. Ensemble models represent a potent ML approach that has demonstrated tangible advantages in various applications [29], [30]. Generally, the generalization ability of ensembles surpasses that of essential learners [30]. This model was developed using C4.5 DT, GBT, and XGBoost learning algorithms.

2.5.1. C4.5 decision tree

DT constitutes a sequential model that efficiently and cohesively combines a series of basic tests, wherein numerical features are compared with threshold values in each test [31]. DT analysis serves as a technique for categorizing target factor categories by formulating decision-making rules in the structure of a tree [32]. This study employed the C4.5 algorithm, which represents an advancement over the ID3 algorithm. The C4.5 algorithm addresses the limitations inherent in the ID3 algorithm [33].

$$H(T) = -\sum_{i=1}^{c} p_i . log_2 (p_i)$$
$$Gain(T, A) = H(T) - \sum_{v \in Value(A)} \frac{|T_v|}{T} . H(T_v) \quad (3)$$

T denotes the training example, c represents the number of classes, p_i is the proportion of examples in class i, A signifies the feature, $Value(A)$ encompass the potential values of attribute A, T_v denotes the subset of T for which attribute A has the value v, dan $|T|$ indicates the size of the set.

2.5.2. Gradient boosted tree

The GBT algorithm combines regression and classification tree models similar to DT to improve predictive accuracy by iteratively refining estimates. It uses a nonlinear regression approach to increase tree precision and addresses prediction model issues [34] through an iterative boosting process, described by the formula provided.

$$F_m(x) = F_{m-1}(x) + \rho . h_m(x) \quad (4)$$

In this formula, $F_m(x)$ represents the predictive function after including the m^{th} tree, $F_{m-1}(x)$ represents the predictive function after including the $(m\text{-}1)^{th}$ tree, ρ is the learning rate, which is a hyperparameter that determines the influence of each tree, and h_m denotes the m^{th} weak learner that is trained on the negative gradient of the loss function.

2.5.3. Extreme gradient boosting

XGBoost represents a fusion of bagging and boosting algorithms, commencing with constructing a weaker learner model and progressively enhancing its accuracy sequentially [35]. The initial tree in XGBoost tends to be weaker in classification, employing probability initialization. Subsequent weight updates are applied to each constructed tree, resulting in an ensemble of classification trees. The formula employed in this algorithm is as:

$$\hat{y_i} = \sum_{t=1}^{T} f_t(x_i), f_t \in F \quad (5)$$

In the given context, $f_t(x_i)$ symbolizes the predicted value of the t^{th} residual tree for the t^{th} residual, where x_i-y_i denotes the predicted value of the model. The term f_t corresponds to the residual number in the t^{th} round. Additionally, F represents the residual tree function’s vacant space [36].

2.6. Evaluation metrics

Evaluation involves assessing the performance of the constructed prediction model. In this research, the evaluation process depicted in the Table 1 involved examining achievements in accuracy, recall, precision, and specificity.

Table 1. Evaluation metrics of the study

Performance metrics	Equation	Number of equations
Accuracy	$Accuracy = \frac{tp + tn}{tp + fp + tn + fn}$	(6)
Recall	$Precision = \frac{tp}{tp + fp}$	(7)
Precision	$Sensitivity = \frac{tp}{tp + fn}$	(8)
Specificity	$Specificity = \frac{tn}{tn + fp}$	(9)

3. RESULTS AND DISCUSSION

This study compares the effectiveness of three ensemble algorithms C4.5, GBT, and XGBoost for predicting PPD in postpartum mothers. This research involved 317 participants and 32 variables, including 31 independent variables and one dependent variable. The study found that 27.13% of postpartum mothers were at risk of developing PPD (minority class), while 72.87% were not at risk (majority class). This dataset was used as input for the predictive model in Table 2.

Table 2 delineates the experiments conducted, which are employed in three prediction models, each with its unique characteristics. Model 1 centres on the original dataset results, integrated into ensemble learning. Conversely, Model 2 concentrates on refining the model through data preprocessing, employing resampling techniques such as Bootstrapping and SMOTE. Meanwhile, Model 3 incorporates feature selection, specifically FS and BE, yielding feature importance. The outcomes of implementing this structured model are presented in Table 3.

Table 2. Design processes of the PPD model prediction

Type of models	Architecture of the model
Model 1	Dataset + CV + ML algorithms + evaluation metrics
Model 2	Dataset + resampling methods (Bootstrapping and SMOTE) + CV + ML algorithms + evaluation metrics
Model 3	Dataset + feature selection (forward selection and backward elimination) + CV + ML algorithms + evaluation metrics

Table 3. Evaluation of PPD model prediction

Algorithms	Accuracy	Precision	Recall	Specificity
Model 1				
C4.5 DT	68.77	42.02	16.25	88.32
GBT	63.77	38.83	48.19	69.75
XGBoost	67.83	39.39	21.94	84.86
Model 2				
C4.5 DT	69.51	72.10	66.32	72.55
GBT	69.95	68.36	76.71	63.06
XGBoost	87.57	87.46	88.42	86.79
Model 3				
C4.5 DT + FS	78.27	82	28.33	96.96
GBT + FS	72.88	-	0	100
XGBoost + FS	77.94	80.77	24.31	97.83
C4.5 DT + BE	74.16	57.14	19.03	94.84
GBT + BE	69.05	47.80	39.03	80.09
XGBoost + BE	73.20	52.97	26.94	90.49

Table 3 illustrates the outcomes of the experiments that were conducted. Model 1 presents findings from research conducted using original data, wherein C4.5 DT exhibits an accuracy rate of 68.77%, a precision of 42.02%, a recall of 16.25%, and a specificity of 88.32%. In comparison, XGBoost follows with an accuracy of 67.83%, a precision of 39.39%, a recall of 21.94%, and a specificity of 84.86%. GBT trails behind with an accuracy of 63.77%, a precision of 38.83%, a recall of 48.19%, and a specificity of 69.75%. These results still need more satisfaction, as evidenced by the relatively low accuracy values across all learning algorithms. The inconsistent precision, recall, and specificity values also hint at class imbalance issues.

Meanwhile, XGBoost achieved the highest accuracy score with a figure of 87.57%, precision of 87.46%, recall of 88.42%, and specificity of 86.79% in Model 2. Additionally, C4.5 DT and GBT have not shown better results; the accuracy values obtained by both learning algorithms are only 69.51% and 69.95%, respectively. Nevertheless, C4.5 DT and GBT still demonstrate slight improvements, 0.74%, and 6.18%, respectively. This model becomes the most effective model compared to others, as it achieves class balance, as seen from the precision, recall, and specificity figures, which are consistent. This result is different from research conducted by Natarajan *et al.* [37], which shows that gradient boosting is the best model when improved with SMOTE, producing a model performance with an area under the curve (AUC) value of 0.952.

In Model 3, the C4.5 DT algorithm, integrated with forward selection, achieved the highest accuracy value at 78.27%, with precision at 82%, recall at 28.33%, and specificity at 96.96%. Then, XGBoost attained the next highest accuracy level, scoring 77.94%, with precision at 80.77%, recall at 24.31%, and specificity at 97.83%. Conversely, integrated with backward elimination, GBT yielded the poorest results in this model, recording accuracy of 69.05%, precision of 47.80%, recall of 39.03%, and specificity of 80.09%. Using feature selection through this wrapper method failed to deliver satisfactory performance. Moreover, employing this method tends to be more time-consuming than previously established models. Conversely, the anticipated output from operational feature selection is determining the importance of features on the independent variables in this research.

The results revealed that XGBoost, integrated with the resampling method utilizing Bootstrapping and SMOTE, yielded the highest performance, achieving an accuracy score of 87.57%. This figure shows a higher result than Hochman *et al.* [38] investigated XGBoost's capability in predicting PPD incidence, devising a Q-based model, and acquiring an AUC value of 0.712 in a nationwide birth cohort. Correspondingly, Hochman *et al.* [38] also noted a tendency towards low positive predictive value (PPV). This trend arises from the dataset's low incidence of PPD. In our research, employing the resampling method, specifically Bootstrapping and SMOTE, still led to higher accuracy values. While other research [14], [39]–[41] shows RF as the best algorithm in building prediction models, this research provides new insights into the use of ensemble learning with XGBoost, which also provides satisfactory prediction model performance.

The challenge encountered in other models within this study is the presence of class imbalance. For instance, in Model 1 utilizing C4.5 DT, a relatively high accuracy rate of 68.77% was achieved, yet the precision and recall values were notably low at 42.02% and 16.25%, respectively. Accuracy is defined as the proportion of correct predictions, encompassing both true positives and negatives, relative to the total predictions made. Conversely, recall represents the percentage of actual positive cases correctly identified by the algorithm, indicating its effectiveness in classifying positive cases. Additionally, precision denotes the probability of accurately predicting positive class instances, calculated as the ratio of true positives to the sum of true positives and negatives [42], [43]. This performance indicates a failure to correctly predict instances of the minority class despite the relatively high accuracy rate attained. Such class imbalance may lead to classification errors, diminishing the algorithm's accuracy. Consequently, for a more comprehensive evaluation of model performance, considering all elements of the confusion matrix is imperative.

Figure 2 illustrates the features employed in this study, depicting the relationship between independent and dependent variables utilized in the feature selection process through FS and BE. This figure shows that FS tends to yield fewer associated variables compared to BE. This discrepancy arises because FS initiates with each feature and subsequently iterates by gradually incorporating additional features, whereas, BE involves utilizing all features initially and then progressively eliminating them at each iteration. Ultimately, the figure underscores marital status, number of living children, history of depression, fear of delivery, and relationship with the family as pivotal factors in this research; each was selected five times in both feature selection methods.

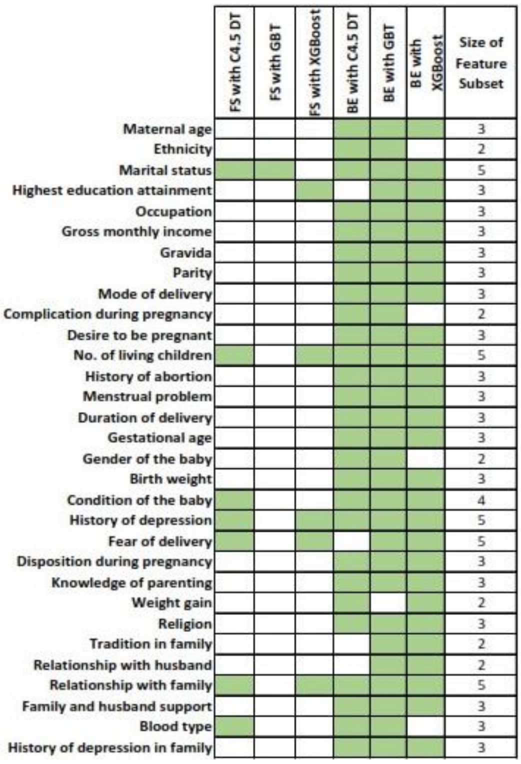

Figure 2. Complete list of features selected by each ML method

Meanwhile, the features consistently selected in the feature selection method employed in this study include marital status, number of living children, history of depression, fear of delivery, and relationship with family. In a marital relationship, the status and quality of the relationship are essential. A high-quality relationship can offer the mother security, as there will be no suspicion within household dynamics. However, partners can also be sources of stress, depending on the relationship's quality. Therefore, focusing on relationship quality rather than just status provides resources and enhances resilience [44]. Additionally, having more children can heighten the risk of postnatal depression. According to Antczak *et al.* [45], mothers with multiple children may experience family conflicts, particularly regarding childcare and education responsibilities. Furthermore, a history of depression experienced by the mother before pregnancy and childbirth also contributes to the likelihood of PPD.

The prediction model developed in this study has yielded commendable outcomes, but certain limitations persist. The dataset comprises primary data, which remains relatively limited in quantity, suggesting the potential consideration of integrating electronic health records (EHR) for predicting PPD incidence. Although the prediction model showcased promising results, exploring alternative ensemble learning strategies, such as RF and adaptive boosting, could enhance model performance. This research holds significant value for healthcare professionals and the community, particularly postpartum mothers, as implementing ML-driven automatic PPD screening can streamline health assessment procedures, saving valuable time. Furthermore, prompt identification of PPD symptoms enables timely and appropriate interventions for affected mothers.

4. CONCLUSION

The combination of physical and emotional exhaustion after childbirth, coupled with hormonal fluctuations, can lead to mental health issues, notably PPD. Severe cases of PPD can escalate to postpartum psychosis, underscoring the critical need for swift and accurate screening methods facilitated by technological advancements. This research employs an ensemble learning model integrated with resampling and feature selection techniques to develop a model demonstrating optimal performance. The most effective prediction model for PPD is achieved through the XGBoost algorithm, integrated with Bootstrapping and SMOTE, yielding an accuracy rate of 87.57% and an AUC of 86.79%. Conversely, the GBT algorithm exhibits less robust performance across all models, with accuracy rates of 63.77% in Model 1, 69.95% in Model 2, 72.88% with forward selection, and 69.05% with backward elimination in Model 3. Furthermore, marital status, number of living children, history of depression, fear of delivery, and relationship with family emerge as independent variables correlated with the dependent variable. This study's PPD prediction model is potentially useful in clinical settings. Nevertheless, we assert that further development and optimization of this model are feasible through utilizing larger datasets and exploring alternative ensemble learning algorithms, as well as various resampling and feature selection methods.

ACKNOWLEDGEMENTS

The authors express heartfelt appreciation to Sultan Idris Education University, Malaysia, and Sari Mulia University for granting the opportunity to undertake this research.

REFERENCES

[1] N. Insan, A. Weke, S. Forrest, and J. Rankin, "Social determinants of antenatal depression and anxiety among women in South Asia: a systematic review & meta-analysis," *PLOS ONE*, vol. 17, no. 2, p. e0263760, Feb. 2022, doi: 10.1371/journal.pone.0263760.

[2] A. F. Nabhan *et al.*, "Women's reproductive span: a systematic scoping review," *Human Reproduction Open*, vol. 2022, no. 2, pp. 1–19, Mar. 2022, doi: 10.1093/hropen/hoac005.

[3] H. Bradshaw, J. N. Riddle, R. Salimgaraev, L. Zhaunova, and J. L. Payne, "Risk factors associated with postpartum depressive symptoms: a multinational study," *Journal of Affective Disorders*, vol. 301, pp. 345–351, Mar. 2022, doi: 10.1016/j.jad.2021.12.121.

[4] J. Slomian, G. Honvo, P. Emonts, J.-Y. Reginster, and O. Bruyère, "Consequences of maternal postpartum depression: a systematic review of maternal and infant outcomes," *Women's Health*, vol. 15, Jan. 2019, doi: 10.1177/1745506519844044.

[5] J. Jayaseelan and Mp. Mohan, "Coping strategies used by postnatal mothers with perceived stress," *Indian Journal of Psychiatry*, vol. 62, no. 4, p. 451, 2020, doi: 10.4103/psychiatry.IndianJPsychiatry_373_19.

[6] S. M. Samuri, T. V. Nova, B. Rahmatullah, S. L. Wang, and Z. . Al-Qaysi, "Classification model for breast cancer mammograms," *IIUM Engineering Journal*, vol. 23, no. 1, pp. 187–199, Jan. 2022, doi: 10.31436/iiumej.v23i1.1825.

[7] D. P. Sinambela, B. Rahmatullah, N. H. C. Lah, and A. W. Selamat, "Machine learning approaches for predicting postpartum hemorrhage: a comprehensive systematic literature review," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 34, no. 3, pp. 2087–2095, Jun. 2024, doi: 10.11591/ijeecs.v34.i3.pp2087-2095.

[8] A. Amanat *et al.*, "Deep learning for depression detection from textual data," *Electronics*, vol. 11, no. 5, p. 676, Feb. 2022, doi: 10.3390/electronics11050676.

[9] Z. Wang *et al.*, "Mapping global prevalence of depression among postpartum women," *Translational Psychiatry*, vol. 11, no. 1, p. 543, Oct. 2021, doi: 10.1038/s41398-021-01663-6.

[10] S. W. N. Amelia and C. P. Kismoyo, "Husband's role in perinatal depression during the new normal period of the COVID-19 pandemic in Sewon Sub-District," *Embrio*, vol. 14, no. 2, pp. 212–220, Nov. 2022, doi: 10.36456/embrio.v14i2.6001.

[11] W. Fazraningtyas, "Psychological Factors in postpartum depression: a study at General Hospitals of Banjarmasin," 2020, doi: 10.4108/eai.23-11-2019.2298367.

[12] W. A. Fazraningtyas, B. Rahmatullah, D. D. Salmarini, S. A. Ariffin, and A. Ismail, "Recent advancements in postpartum depression prediction through machine learning approaches: a systematic review," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 13, no. 4, pp. 2729–2737, Aug. 2024, doi: 10.11591/eei.v13i4.7185.

[13] L. Swastina, B. Rahmatullah, A. Saad, and H. Khan, "A systematic review on research trends, datasets, algorithms, and frameworks of children's nutritional status prediction," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 13, no. 2, pp. 1866–1875, Jun. 2024, doi: 10.11591/ijai.v13.i2.pp1868-1877.

[14] D. Shin, K. J. Lee, T. Adeluwa, and J. Hur, "Machine learning-based predictive modeling of postpartum depression," *Journal of Clinical Medicine*, vol. 9, no. 9, pp. 1–14, Sep. 2020, doi: 10.3390/jcm9092899.

[15] S. Andersson, D. R. Bathula, S. I. Iliadis, M. Walter, and A. Skalkidou, "Predicting women with depressive symptoms postpartum with machine learning methods," *Scientific Reports*, vol. 11, no. 1, p. 7877, Apr. 2021, doi: 10.1038/s41598-021-86368-y.

[16] Y. Park *et al.*, "Comparison of methods to reduce bias from clinical prediction models of postpartum depression," *JAMA Network Open*, vol. 4, no. 4, p. e213909, Apr. 2021, doi: 10.1001/jamanetworkopen.2021.3909.

[17] A. Raza, H. U. R. Siddiqui, K. Munir, M. Almutairi, F. Rustam, and I. Ashraf, "Ensemble learning-based feature engineering to analyze maternal health during pregnancy and health risk prediction," *PLOS ONE*, vol. 17, no. 11, p. e0276525, Nov. 2022, doi: 10.1371/journal.pone.0276525.

[18] J. Kamiri and G. Mariga, "Research methods in machine learning: a content analysis," *International Journal of Computer and Information Technology(2279-0764)*, vol. 10, no. 2, Mar. 2021, doi: 10.24203/ijcit.v10i2.79.

[19] F. Oktamurdiantri and J. I. Saragih, "Use of the edinburgh postnatal depression scale to detect postpartum depression risk," *International Journal of Progressive Sciences and Technologies (IJPSAT)*, vol. 19, no. 1, pp. 31–34, 2020, [Online]. Available: http://ijpsat.ijsht-journals.org.

[20] B. Levis, Z. Negeri, Y. Sun, A. Benedetti, and B. D. Thombs, "Accuracy of the edinburgh postnatal depression scale (EPDS) for screening to detect major depression among pregnant and postpartum women: systematic review and meta-analysis of individual participant data," *BMJ*, p. m4022, Nov. 2020, doi: 10.1136/bmj.m4022.

[21] A. M. Elsobky, A. El. Keshk, and M. G. Malhat, "A comparative study for different resampling techniques for imbalanced datasets," *International Journal for Computers and Information, IJCI*, vol. 10, no. 3, pp. 147–156, 2023.

[22] K. Dissanayake and M. G. Md Johar, "Comparative study on heart disease prediction using feature selection techniques on classification algorithms," *Applied Computational Intelligence and Soft Computing*, vol. 2021, pp. 1–17, Nov. 2021, doi: 10.1155/2021/5581806.

[23] M. Dliyauddin, G. F. Shidik, A. Affandy, and M. A. Soeleman, "Enhancing machine learning accuracy in detecting preventable diseases using backward elimination method," *Jurnal Media Informatika Budidarma*, vol. 8, no. 1, pp. 115–123, Jan. 2024, doi: 10.30865/mib.v8i1.7073.

[24] S. Amri, A. F. Ningrum, and P. R. Arum, "Optimization of Naïve Bayes using backward elimination for heart disease detection," *Jurnal Statistika Universitas Muhammadiyah Semarang*, vol. 6, no. 22, pp. 142–144, 2023.

[25] M. M. Nadzir and J. A. Bakar, "A digital literacy predictive model in the context of distance education," *Journal of ICT in Education*, vol. 10, no. 1, pp. 118–134, 2023.

[26] J. Kaliappan, A. R. Bagepalli, S. Almal, R. Mishra, Y. C. Hu, and K. Srinivasan, "Impact of cross-validation on machine learning models for early detection of intrauterine fetal demise," *Diagnostics*, vol. 13, no. 10, p. 1692, May 2023, doi: 10.3390/diagnostics13101692.

[27] D.-C. Feng *et al.*, "Machine learning-based compressive strength prediction for concrete: an adaptive boosting approach," *Construction and Building Materials*, vol. 230, p. 117000, Jan. 2020, doi: 10.1016/j.conbuildmat.2019.117000.

[28] F. Wang, Z. Li, F. He, R. Wang, W. Yu, and F. Nie, "Feature learning viewpoint of adaboost and a new algorithm," *IEEE Access*, vol. 7, pp. 149890–149899, 2019, doi: 10.1109/ACCESS.2019.2947359.

[29] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: a review," *Engineering Applications of Artificial Intelligence*, vol. 115, p. 105151, Oct. 2022, doi: 10.1016/j.engappai.2022.105151.

[30] A. Mohammed and R. Kora, "A comprehensive review on ensemble deep learning: opportunities and challenges," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 2, pp. 757–774, Feb. 2023, doi: 10.1016/j.jksuci.2023.01.014.

[31] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.

[32] K. M. Kim, J. H. Kim, H. S. Rhee, and B. Y. Youn, "Development of a prediction model for the depression level of the elderly in low-income households: using decision trees, logistic regression, neural networks, and random forest," *Scientific Reports*, vol. 13, no. 1, p. 11473, Jul. 2023, doi: 10.1038/s41598-023-38742-1.

[33] H. Sulistiani and A. A. Aldino, "Decision tree C4.5 algorithm for tuition aid grant program classification (case study: Department of Information System, Universitas Teknokrat Indonesia)," *Edutic - Scientific Journal of Informatics Education*, vol. 7, no. 1, pp. 40–50, Nov. 2020, doi: 10.21107/edutic.v7i1.8849.

[34] M. Ebrahimi, M. Mohammadi-Dehcheshmeh, E. Ebrahimie, and K. R. Petrovski, "Comprehensive analysis of machine learning models for prediction of sub-clinical mastitis: deep learning and gradient-boosted trees outperform other models," *Computers in Biology and Medicine*, vol. 114, p. 103456, Nov. 2019, doi: 10.1016/j.compbiomed.2019.103456.

[35] A. Sharma and W. J. M. I. Verbeke, "Improving diagnosis of depression with XGBOOST machine learning model and a large biomarkers dutch dataset (n = 11,081)," *Frontiers in Big Data*, vol. 3, Apr. 2020, doi: 10.3389/fdata.2020.00015.

[36] L. Wang, M. Han, X. Li, N. Zhang, and H. Cheng, "Review of classification methods on unbalanced data sets," *IEEE Access*, vol. 9, pp. 64606–64628, 2021, doi: 10.1109/ACCESS.2021.3074243.

[37] S. Natarajan, A. Prabhakar, N. Ramanan, A. Bagilone, K. Siek, and K. Connelly, "Boosting for postpartum depression prediction," in *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, Jul. 2017, pp. 232–240, doi: 10.1109/CHASE.2017.82.

[38] E. Hochman *et al.*, "Development and validation of a machine learning-based postpartum depression prediction model: a nationwide cohort study," *Depression and Anxiety*, vol. 38, no. 4, pp. 400–411, Apr. 2021, doi: 10.1002/da.23123.

[39] R. Fischbein, H. L. Cook, K. Baughman, and S. R. Díaz, "Using machine learning to predict help-seeking among 2016–2018 Pregnancy Risk Assessment Monitoring System participants with postpartum depression symptoms," *Women's Health*, vol. 18, Jan. 2022, doi: 10.1177/17455057221139664.

[40] E. Valavani *et al.*, "Data-driven insights towards risk assessment of postpartum depression," in *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies*, 2020, pp. 382–389, doi: 10.5220/0009369303820389.

[41] M. Cai, Y. Wang, Q. Luo, and G. Wei, "Factor analysis of the prediction of the postpartum depression screening scale," *International Journal of Environmental Research and Public Health*, vol. 16, no. 24, p. 5025, Dec. 2019, doi: 10.3390/ijerph16245025.

[42] B. Zhang, B. Rahmatullah, S. L. Wang, G. Zhang, H. Wang, and N. A. Ebrahim, "A bibliometric of publication trends in medical image segmentation: quantitative and qualitative analysis," *Journal of Applied Clinical Medical Physics*, vol. 22, no. 10, pp. 45–65, Oct. 2021, doi: 10.1002/acm2.13394.

[43] B. Zhang, B. Rahmatullah, S. L. Wang, A. A. Zaidan, B. B. Zaidan, and P. Liu, "A review of research on medical image confidentiality related technology coherent taxonomy, motivations, open challenges and recommendations," *Multimedia Tools and Applications*, vol. 82, no. 14, pp. 21867–21906, Jun. 2023, doi: 10.1007/s11042-020-09629-4.

[44] B. F. Hutchens and J. Kearney, "Risk factors for postpartum depression: an umbrella review," *Journal of Midwifery & Women's Health*, vol. 65, no. 1, pp. 96–108, Jan. 2020, doi: 10.1111/jmwh.13067.

[45] R. Antczak, N. T. Quashie, C. A. Mair, and B. Arpino, "Less is (often) more: number of children and health among older adults in 24 countries," *The Journals of Gerontology: Series B*, vol. 78, no. 11, pp. 1892–1902, Nov. 2023, doi: 10.1093/geronb/gbad123.

BIOGRAPHIES OF AUTHORS

Winda Ayu Fazraningtyas received her bachelor's degree in Nursing Science from Lambung Mangkurat University, Banjarmasin, South Kalimantan, Indonesia. She has a strong interest in women's health, which led her to pursue a post-graduate program in the Master of Science in Nursing at St. Paul University Philippines. Currently, she is a Ph.D. student at the Graduate Program Faculty of Computing and Meta-Technology in Tanjong Malim, Perak, Malaysia. In addition to her academic pursuits, she serves as an Assistant Professor at the Department of Maternity Nursing, Faculty of Health, Sari Mulia University. She is also actively involved as a reviewer for indexed international journals. Her research interests i.e., women's mental health, maternity nursing focus on the postpartum period, violence against women, and machine learning. She can be contacted at email: windaayu@unism.ac.id.

Bahbibi Rahmatullah is currently an associate professor in the Department of Computing, Faculty of Computing and Meta-Technology, Sultan Idris Education University, Malaysia. Having received a B.Eng (Electrical) from Vanderbilt University, USA, a M.Eng.Sc from Multimedia University, Malaysia, and a D.Phil in Eng. Science from the University of Oxford, UK, she is keen to apply the technical and research skills gained to improve the quality of research and education in Malaysia. She has authored a wide range of publications and has been invited to review articles for high-impact journals and conferences. Current research interests include image and signal processing, pattern recognition, machine learning, learning analytics, child development, ICT, and education. She can be contacted at email: bahbibi@meta.upsi.edu.my.

Husni Naparin gained his undergraduate degree (S1) from the Department of Technical Information, Technical Faculty, Kalimantan Islamic University, Syekh Muhammad Arsyad Al-Banjary Banjarmasin in 2013 and Graduate Program (S2) from Faculty of Computer Science (Intelligence System), University of Dian Nuswantoro, Semarang, Jawa Tengah in 2015. He graduated Ph.D. program in the Faculty of Computing and Meta-Technology, Universiti Pendidikan Sultan Idris. His main research interest is information system, data mining, data visualization, and multimedia. He can be contacted at email: 1naparinhusni@gmail.com.

Mohammad Basit currently serves as a lecturer in the Nursing Department within the Faculty of Health at Sari Mulia University, located in Banjarmasin, South Kalimantan, Indonesia. His passion lies in utilizing his technical and research expertise to enhance the quality of research and education, particularly in Ward Nursing Management. He has contributed significantly to the field through numerous publications in nationally accredited and international journals and community service initiatives. Leveraging his extensive educational background, he also holds the position of Chair of the Nursing Department at Sari Mulia University. He can be contacted at email: syafabasit@gmail.com.

Nor Asiah Razak is currently a senior lecturer in the Software Engineering and Smart Technology Department at the Faculty of Computing and Meta-Technology, Sultan Idris Education University (UPSI). She earned her Ph.D. in Educational Technology from Universiti Putra Malaysia in 2019, focusing on e-learning. Her current research interests include digitalization in education, focusing on learning strategies for enhancing 21st-century learning skills. She also serves as a reviewer for national and international academic journals. She can be contacted at email: norasiah.razak@meta.upsi.edu.my.