J Surg Med. 2025;9(7):98-103.
https://doi.org/10.28982/josam.8268

Research article
https://jsurgmed.com/

Journal of Surgery and Medicine
e-ISSN: 2602-2079

Stacked hybrid model: Multi-layer perceptron and logistic regression with meta-learning for cesarean section classification

Emre Yalçın [1], Hayriye Tanyıldız [2], Serpil Aslan [3], Süleyman Cansun Demir [1], Mesut Avan [1], Fatma İşlek Uzay [1], Serdar Aykut [1]

[1] Department of Obstetrics and Gynecology, Division of Perinatology, Cukurova University School of Medicine, 01330, Adana, Turkey
[2] Department of Informatics, Malatya Turgut Ozal University, 44210 Malatya, Turkey
[3] Department of Software Engineering, Faculty of Engineering and Natural Sciences, Malatya Turgut Ozal University, 44210, Malatya, Turkey

ORCID of the author(s)

EY: https://orcid.org/0000-0001-7088-6749
HT: https://orcid.org/0000-0002-6300-9016
SA: https://orcid.org/0000-0001-8009-063X
SCD: https://orcid.org/0000-0001-8331-9559
MA: https://orcid.org/0009-0001-8458-6963
FİU: https://orcid.org/0000-0002-8660-4854
SA: https://orcid.org/0000-0002-6590-9982

Corresponding Author
Emre Yalçın
Department of Obstetrics and Gynecology, Division of Perinatology, Cukurova University School of Medicine, 01330, Adana, Turkey
E-mail: dremreyalcin23@gmail.com

Ethics Committee Approval
This study was carried out using the publicly available and anonymized "Caesarian Section Classification Data Set" taken from kaggle.com. As the dataset does not include personal data or involve direct human participation, ethics committee approval was not required. Furthermore, no additional permissions or ethical approvals were necessary for its use.

Conflict of Interest
No conflict of interest was declared by the authors.

Financial Disclosure
The authors declared that this study has received no financial support.

Published
2025 July 4

Abstract

Background/Aim: This study aims to develop an interpretable and practical decision support method for early prediction of the need for cesarean delivery. Although machine learning and deep learning models are prevalent in the literature, their generalization capabilities are often restricted, especially when utilizing small clinical datasets. This limitation underscores the necessity for robust, transparent, and well-regularized models in medical decision-making processes.

Methods: The study proposed a stacking-based hybrid model, which combines the strengths of both classical and modern techniques. The data were normalized using StandardScaler, and feature selection involved principal component analysis (PCA) and SelectKBest to capture global and target-relevant patterns. In the classification phase, two parallel learners – a regularized multi-layer perceptron (MLP) and logistic regression – were used, followed by a random forest meta-learner.

Results: The experimental analysis demonstrated that the proposed model achieved an average accuracy of 96.43% under stratified 5-fold cross-validation. Although this result surpassed the performance of other baseline models within the dataset, it should be regarded as preliminary due to the limited sample size.

Conclusion: The findings indicate that the proposed hybrid approach has potential as a promising direction for future clinical decision support research. Nonetheless, additional validation using larger and more diverse datasets is necessary to adequately assess its generalizability and practical utility.

Keywords: cesarean prediction, clinical decision support, medical data analysis, machine learning in healthcare

How to cite: Yalçın E, Tanyıldız H, Aslan S, Demir SC, Avan M, İşlek Uzay F, Aykut S. Stacked hybrid model: Multi-layer perceptron and logistic regression with meta-learning for cesarean section classification. J Surg Med. 2025;9(7):98-103.

Introduction

The global rate of cesarean deliveries has significantly increased in recent decades. The World Health Organization (WHO) recommends a cesarean section (CS) rate between 10% and 15%; however, many countries have far exceeded this recommendation. For example, before the 1980s, CS rates were generally below 10%, but in the past decade, they have surpassed 30% in many developed countries [1]. Although cesarean delivery is essential when medically justified, its unnecessary use can lead to various complications, prolonged recovery times, and increased burdens on healthcare systems.

Currently, no standardized method supports clinical teams in making objective decisions about CSs in borderline cases with unclear medical indications [2]. In response, artificial intelligence (AI)-based systems have gained traction in medicine, notably for classification tasks. In obstetrics, predictive models for CS can assist clinicians in decision-making. However, challenges such as small sample sizes, class imbalance, clinical data complexity, and the need for interpretability hinder the development of effective models.

To address these limitations, this study proposes a novel stacking-based hybrid classification model for CS prediction. This model integrates classical machine learning algorithms – such as logistic regression, decision trees, and support vector machines – with modern deep learning architectures, including multi-layer perceptron (MLP), convolutional neural network (CNN), and long short-term memory (LSTM). The goal is to leverage both the interpretability of traditional methods and the high representation power of deep learning. Based on experimental observations, a meta-learning approach was employed to harness the strengths of the individual models.

The proposed framework incorporates a feature fusion strategy that integrates principal component analysis (PCA) with SelectKBest. This combination retains global variance and target-specific relevance, thereby reducing overfitting in small clinical datasets. The classification structure utilizes a regularized MLP and logistic regression as base learners, whose outputs are subsequently fed into a random forest meta-learner. This layered design enhances generalization by capturing both linear and non-linear patterns effectively.

Compared to the single-model approaches commonly found in the literature, the proposed hybrid model achieved a superior accuracy of 96.43%, demonstrating its effectiveness in CS classification. The logistic regression component of the model enhances interpretability, thus supporting both practical clinical utility and academic relevance.

Furthermore, this study addresses a gap in the literature by introducing an integrated model that combines stacking, dimensionality reduction, and ensemble learning techniques specifically tailored for cesarean prediction. Although previous studies [3-5] have investigated various feature selection and classification algorithms, few have successfully integrated multiple advanced techniques into a single architecture with demonstrated clinical applicability.

Materials and methods

Dataset and pre-processing

The dataset used in this study was sourced from a publicly available resource on Kaggle [6]. It comprises 26 instances, covering both cesarean and non-cesarean deliveries. Each record encompasses various maternal and fetal attributes, including maternal age, blood pressure, fetal heart rate, and prior delivery history. The target variable specifies whether the delivery was a CS or occurred naturally.

The dataset comprises six input features alongside a binary target variable. The features are defined as follows: Age (represents the mother's age in years, ranging from 22 to 40); Delivery Number (Indicates the total number of previous deliveries, recorded numerically and ranging from 1 to 4); Delivery Time (an ordinal value that specifies the timing of birth, with three categories: 1 for timely, 2 for late, and 3 for early); Blood Pressure (a categorical value representing blood pressure status, where 1 denotes low, 2 normal, and 3 high); Heart Problem (a binary indicator of the presence of a maternal heart condition, coded as 0 for absence and 1 for presence); and Target (Cesarean) (a binary outcome variable that indicates the mode of delivery, where 1 represents a CS and 0 denotes a natural delivery).

Among the 26 cases, 15 were cesarean deliveries and 11 were natural births, which reflects a mildly imbalanced class distribution. Most participants had normal blood pressure and no heart problems, and the majority were first-time mothers. This information is included to enhance transparency about the dataset structure and to enable a more informed assessment of potential biases.

In the pre-processing phase, normalization was applied to standardize feature scales, and incomplete records were removed to ensure data quality. To enhance classification performance, feature selection techniques were utilized. Specifically, PCA and SelectKBest based on the ANOVA F-score were employed. This combined approach reduced dimensionality, eliminated irrelevant or noisy features, and preserved the most informative attributes for model training.

Machine learning models

Logistic regression is a widely used linear classification method for binary problems. It employs the logistic (sigmoid) function to predict the probabilities of the dependent variable, yielding outputs between 0 and 1. Due to its high interpretability, logistic regression is particularly favored in fields such as healthcare, finance, and social sciences [7].

The random forest is an ensemble learning method composed of multiple decision trees. Each tree is trained on a distinct subset of data with randomly selected features. This approach allows random forests to outperform single decision trees in terms of accuracy and generalization, owing to their diversity. Additionally, random forests significantly reduce the risk of overfitting. The algorithm yields effective results in both classification and regression tasks, making it particularly useful for large datasets. A further advantage of random forests is their capability to produce interpretable outputs, such as feature importance rankings, which are valuable to researchers seeking to understand model predictions [8]. However, models with a large number of trees can experience increased training time and computational costs.

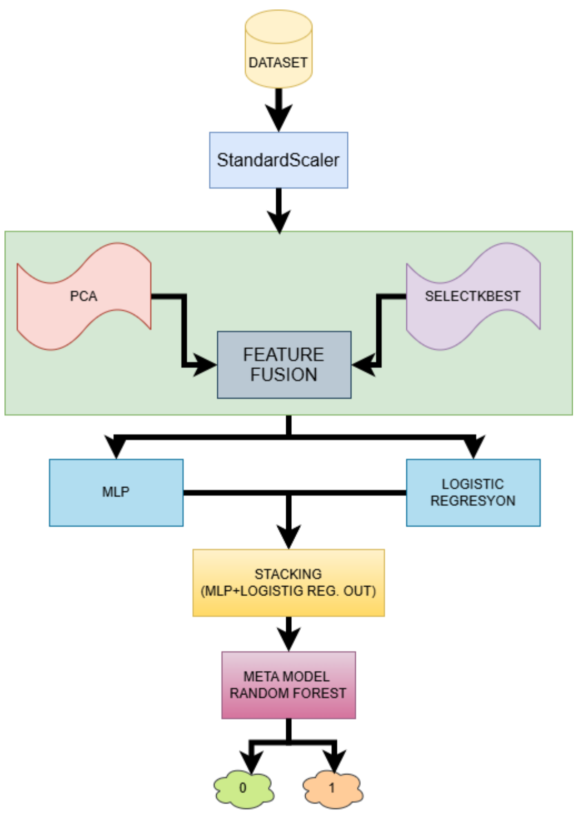
Deep learning models

The multi-layer perceptron (MLP): MLP is one of the most fundamental and widely utilized types of artificial neural networks, primarily used for supervised learning tasks [9]. As a feedforward neural network, an MLP typically consists of an input layer, one or more hidden layers, and an output layer. Each neuron is fully connected to neurons in the preceding and following layers. The use of non-linear activation functions enables the model to learn complex mappings. Variants, including residual MLPs, batch-normalized MLPs, and deeper architectures, are often employed to enhance learning performance and ensure training stability.

Convolutional neural networks (CNNs): CNNs originally developed for image processing tasks, have recently gained popularity in analyzing one-dimensional time series data. By utilizing 1D convolutional layers, CNNs can automatically detect local patterns and short-term dependencies, such as trends or periodic signals, within sequential data. This capability renders them suitable for applications in fields including financial forecasting, sensor signal analysis, and clinical time series interpretation. Compared to traditional statistical methods like ARIMA or standard machine learning algorithms, CNNs often provide superior feature extraction capabilities and exhibit higher predictive performance in many cases [10].

Long short-term memory (LSTM): LSTM networks are a type of recurrent neural network (RNN) designed to model long-range dependencies in sequential data [11]. Distinct from traditional RNNs, LSTMs incorporate memory cells and gating mechanisms – specifically input, forget, and output gates – that maintain information over extended periods and address the vanishing gradient problem. This architecture allows LSTMs to retain relevant patterns from earlier time steps, making them particularly suitable for tasks such as language modeling, energy consumption prediction, and financial time series forecasting.

Figure 1: The architecture of the proposed model.



The proposed hybrid model

The hybrid model developed in this study aims to enhance classification performance by integrating deep learning with traditional machine learning methods. Figure 1 presents the model's overall architecture. It is structured around a multi-stage pre-processing pipeline that optimizes the feature space before model training.

In the initial stage, data normalization is executed using StandardScaler. This ensures that features on varying scales equitably contribute to the learning process. Subsequently, dimensionality reduction is achieved through a feature fusion strategy that integrates PCA with SelectKBest based on the ANOVA F-score. PCA identifies new orthogonal components that maximize variance, while SelectKBest retains statistically significant original features correlated with the target variable. By combining these approaches, the model efficiently captures both global and local patterns, focusing on the most informative attributes and improving learning efficiency.

After feature selection, the dataset is simultaneously processed through two distinct classifiers. The first classifier is a deep neural network, constructed with a multi-layered architecture and equipped with regularization techniques such as L2 regularization, dropout, batch normalization, and the LeakyReLU activation function. This network is designed to capture complex, non-linear relationships within the data. In parallel, a logistic regression model is trained, offering a simpler, interpretable alternative that excels at learning linear decision boundaries and provides transparency in clinical contexts.

The final and most innovative component of the proposed system is the meta-learning stage, which utilizes the stacking ensemble method. In this stage, the output probabilities from both the deep neural network and logistic regression model are regarded as a new feature space. These outputs are then input into a random forest meta-classifier, which generates the final predictions. This ensemble approach capitalizes on the strengths of both base learners, thereby enhancing the model's generalization capability.

Results

To evaluate the performance of the proposed model and baseline algorithms, we conducted a series of experiments in the Google Colab Pro environment, which features an NVIDIA A100 GPU and 84 GB of RAM. The implementations were executed using the Python programming language. In this study, model performance was assessed using two key metrics: accuracy, which measures overall classification success, and specificity, which indicates the model's ability to correctly identify non-cesarean cases. These metrics were selected for their relevance in evaluating both general effectiveness and clinical reliability. Given the small sample size and class imbalance in the dataset, we employed stratified 5-fold cross-validation to ensure balanced representation of classes in each fold. The dataset was divided into five subsets for each iteration; four were used for training and one for testing. Performance results were averaged across all folds to derive a robust and generalizable estimate of model effectiveness.

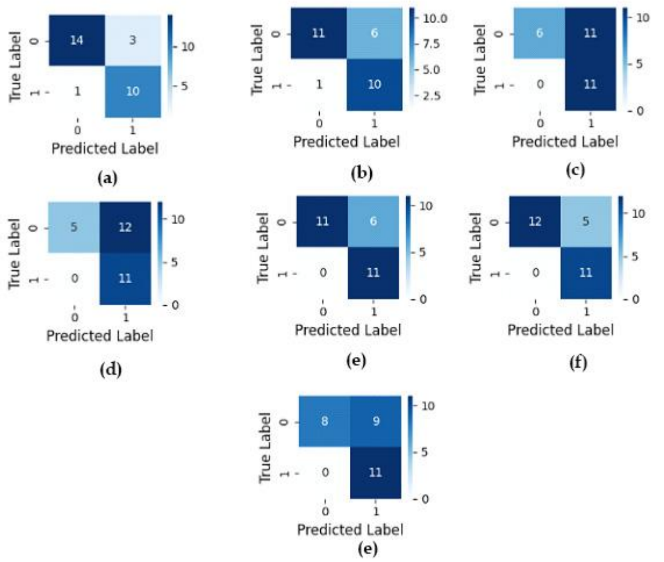
Among deep learning models, the Simple MLP achieved the highest overall accuracy, reaching 85.71%. The Residual MLP demonstrated superior specificity at 88.24%, highlighting its effectiveness in correctly identifying non-cesarean cases. The CNN model also performed well, showing balanced results across accuracy and specificity. In contrast, MLP variants using batch normalization or dropout exhibited weaker performance, especially in specificity, suggesting that such regularization techniques may be less effective in small-sample clinical settings.

Table 1: The performance comparisons of the deep learning models.

Deep Models	Accuracy	Specificity
Simple MLP	85.71	35.29
Deep MLP	75	64.71
MLP with Dropout	60.71	35.29
MLP with Batch Normalization	57.14	29.41
MLP with Residual Connections	78.57	88.24
CNN	82.14	70.59
LSTM	67.86	47

Figure 2 offers further insight into model behavior through confusion matrix analysis. The Simple MLP achieved the highest accuracy, correctly classifying 24 out of 28 instances. The CNN and Residual MLP models followed closely, with 23 and 22 correct predictions, respectively. These models exhibited relatively balanced performance across both classes. In contrast, models such as the Dropout MLP and LSTM, while successful in identifying all cesarean cases, produced a high number of false positives for non-cesarean cases, significantly reducing their specificity. The Batch Normalized MLP demonstrated the lowest overall performance, struggling to identify non-cesarean cases and generating the most misclassifications. These observations underscore the importance of both architecture and regularization in achieving balanced performance. Although deep learning models can achieve high predictive accuracy, their class-wise reliability varies considerably based on design choices.

Figure 2: Confusion matrices for each model architecture: (a) Simple MLP representing a basic multilayer perceptron structure; (b) Deep MLP with increased depth; (c) MLP with Dropout regularization; (d) MLP with Batch Normalization to stabilize learning; (e) MLP with Residual Connections to mitigate vanishing gradient issues; (f) Convolutional Neural Network (CNN) specialized for spatial data; (g) Long Short-Term Memory (LSTM) network tailored for sequential data.



The experimental results of the proposed model

In the proposed model, feature selection for the Deep MLP architecture occurred in two stages: PCA and SelectKBest based on ANOVA F-values. This pre-processing step aimed to reduce dimensionality, eliminate irrelevant features, and enhance the model's learning efficiency. By retaining the most significant variables, we expected the model to achieve improved generalization and predictive performance. The experimental results of this model, developed using the optimized feature set, are presented in the following section.

Figure 3 summarizes the stepwise improvements observed in the proposed model. In the first stage, feature selection using PCA and SelectKBest enhanced the Deep MLP's ability to distinguish between classes. In the second stage, incorporating logistic regression as a meta-learner further improved class-wise balance and overall prediction consistency. This approach resulted in the best-performing hybrid configuration.

Figure 3: Confusion matrices of the proposed hybrid models: (a) Deep MLP with PCA and SelectKBest (ANOVA F-value) feature selection; (b) Hybrid model integrating Logistic Regression into the selected feature structure.

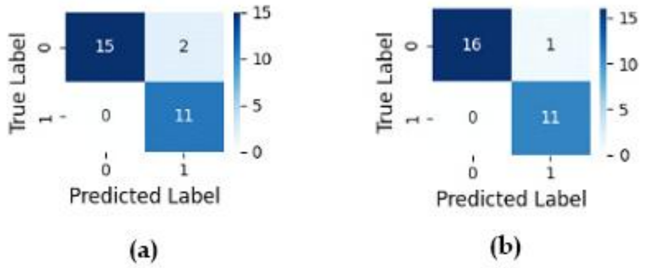


Table 2 presents a comparison between the Deep MLP model after feature selection and the final hybrid model that integrates logistic regression. Feature selection led to substantial performance enhancements, and the hybrid design further improved both accuracy and specificity. The proposed model achieved the best overall results, particularly in accurately identifying non-cesarean cases. These findings suggest that integrating deep learning with a classical learner in a feature-optimized architecture can greatly enhance predictive performance and class-level reliability in limited clinical datasets.

Table 2: Performance comparison of the Deep MLP model with feature selection and the proposed hybrid model integrating logistic regression.

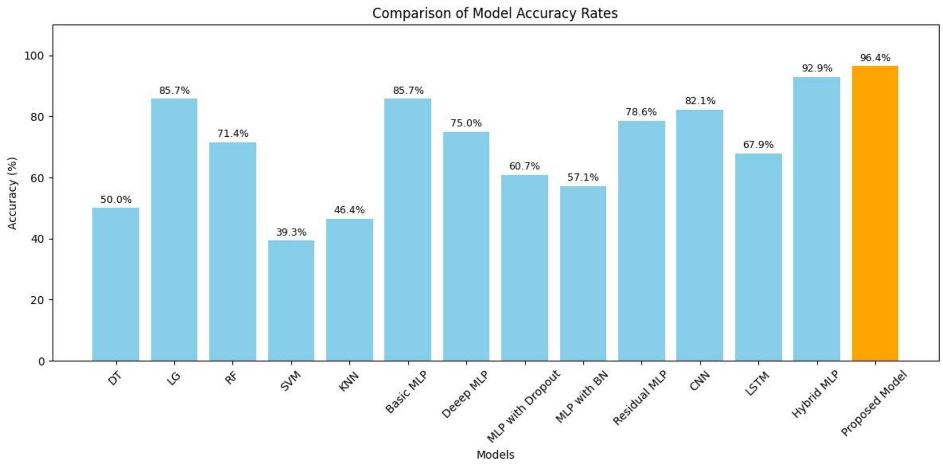
Deep Model	Accuracy	Specificity
(PCA - SelectKBest) + MLP	92.86	88.24
((PCA - SelectKBest) + MLP + Logistic Regression)	96	94.12

Discussion

This study's primary contribution is the development of a novel hybrid model that combines deep learning and traditional machine learning approaches to address the clinically significant and sensitive task of categorizing cesarean deliveries. Previous studies [3-5,12,13] have employed various algorithms individually, achieving accuracy rates between 63% and 89%. In contrast, our proposed model achieves an accuracy of 96.43%, surpassing all existing approaches. By incorporating dimensionality reduction and feature selection techniques, such as PCA and SelectKBest, the model focuses on the most informative inputs, thereby accelerating the training process and reducing the risk of overfitting. While the Deep MLP model is highly capable of learning complex relationships, the logistic regression model enhances interpretability and simplicity in the decision-making process. The stacking architecture, which integrates the outputs of both models using a random forest classifier, effectively harnesses the strengths of each component to enhance overall performance.

Figure 4 compares the accuracy metrics of the proposed hybrid model with various machine learning models (DT, LG, RF, SVM, KNN) and deep learning models (Simple MLP, CNN, LSTM, Residual MLP, etc.). The graph shows that traditional models such as SVM, KNN, and Decision Tree achieved accuracy rates below 50%. Among the deep learning models, only the

Figure 4: Comparison of accuracy rates achieved by the proposed hybrid model and other traditional machine learning (DT, LG, RF, SVM, KNN) and deep learning models (Simple MLP, CNN, LSTM, Residual MLP).



Simple MLP and CNN surpassed 80%. Notably, the proposed hybrid model outperformed all others with an accuracy of 96%. These findings suggest that the proposed model is highly reliable and suitable for integration into clinical decision support systems, especially in critical tasks such as cesarean classification.

Despite the promising performance of the proposed hybrid model, it is crucial to note that the dataset used in this study consisted of only 26 instances. This limited sample size inherently restricts the model's ability to generalize and heightens the risk of overfitting, particularly with the use of complex architectures like deep neural networks. Employing deep learning techniques on such a small dataset may be methodologically questionable due to their tendency to memorize rather than generalize in data-scarce environments. Although various regularization and ensemble techniques were implemented to address these concerns, the results should be viewed as preliminary proof-of-concept. This study aimed to explore whether a hybrid structure, combining interpretability with complexity, could still provide meaningful predictive insights under the real-world limitations often encountered in clinical practice. Further research with larger and more diverse datasets is necessary to confirm the robustness and applicability of the proposed approach.

Study limitations

The primary limitation of this study is the small dataset size (n=26), which diminishes statistical power and limits the generalizability of the findings. Despite using regularization techniques and feature optimization to construct the hybrid model and reduce overfitting, the small sample size presents a significant obstacle to effective model validation. Additionally, applying feature selection methods such as PCA and SelectKBest to such a constrained dataset can produce unstable results, as the identified patterns may not accurately reflect broader population dynamics. Furthermore, although stacking ensemble learning was employed to improve predictive performance, training both the base and meta-learners on the same limited dataset increases the risk of overfitting to sample-specific noise.

Despite implementing various control measures, including dimensionality reduction and regularization strategies, the outcomes should be considered preliminary. This study is best regarded as a feasibility analysis rather than a definitive modeling framework. Future research should aim to validate the approach using larger, multicenter datasets to ensure broader clinical applicability and reliability.

Conclusion

In this study, widely used machine learning and deep learning approaches for predicting the likelihood of cesarean delivery were systematically evaluated. A stacking-based hybrid model was proposed, combining the most effective aspects of these approaches. By applying PCA and SelectKBest, meaningful features were extracted, which both accelerated the learning process and enabled the model to capture complex relationships. The model's multi-stage structure not only aligns with the nature of medical diagnostic problems but also outperforms previous studies in the literature, achieving an accuracy rate of 96.43%.

In conclusion, the proposed hybrid model shows significant potential as a clinical decision support tool. It may be adapted for real-time applications by testing on broader and more diverse datasets. Future studies should focus on integrating explainable AI (XAI) techniques to improve the clinical interpretability of model outputs and facilitate their integration with expert systems.

References

1. Afroja S, Kabir MA, Saleh AB. Machine learning-based algorithms for determining C-section among mothers in Bangladesh. Int J Travel Med Glob Health. 2023;11(4):391-401. doi: 10.34172/ijtmgh.2023.65.
2. Chu K, Cortier H, Maldonado F, Mashant T, Ford N, Trelles M. Cesarean section rates and indications in sub-Saharan Africa: a multi-country study from Médecins Sans Frontières. PLoS One. 2012;7(9):e44484. doi: 10.1371/journal.pone.0044484.
3. Desyani T, Saifudin A, Yulianti Y. Feature selection based on Naive Bayes for caesarean section prediction. IOP Conf Ser Mater Sci Eng. 2020;879(1):012091. doi: 10.1088/1757-899X/879/1/012091
4. de Souza HC, Perdoná GS, Marcolin AC, Oyeneyin LO, Oladapo OT, Mugerwa K, et al. Development of caesarean section prediction models: secondary analysis of a prospective cohort study in two sub-Saharan African countries. Reprod Health. 2019;16:1-11. doi: 10.1186/s12978-019-0832-4.
5. Pereira S, Portela F, Santos MF, Machado J, Abelha A. Predicting type of delivery by identification of obstetric risk factors through data mining. Procedia Comput Sci. 2015;64:601-9. doi: 10.1016/j.procs.2015.08.572.
6. Amir75. Caesarean Section Classification Dataset [Internet]. Kaggle; [cited 2025 April 14]. Available from: https://www.kaggle.com/datasets/amir75/caesarean-section-classification/data
7. Hosmer DW Jr, Lemeshow S, Sturdivant RX. Applied Logistic Regression. 3rd ed. Hoboken (NJ): John Wiley & Sons; 2013.
8. Rigatti SJ. Random forest. Journal of Insurance Medicine. 2017;47(1):31-9.
9. Desai M, Shah M. An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and convolutional neural network (CNN). Clin eHealth. 2021;4:1-11. doi: 10.1016/j.ceh.2020.12.001
10. Toğaçar M, Ergen B. Processing 2D barcode data with metaheuristic based CNN models and detection of malicious PDF files. Appl Soft Comput. 2024;161:111722. doi: 10.1016/j.asoc.2023.111722
11. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735-80. doi: 10.1162/neco.1997.9.8.1735
12. Khan NI, Mahmud T, Islam MN, Mustafina SN. Prediction of cesarean childbirth using ensemble machine learning methods. In: Proceedings of the 22nd International

Conference on Information Integration and Web-based Applications & Services; 2020 Nov; pp. 331-339. doi: 10.1145/3428757.3429136
13. Yimer NB, Mekonnen EG. Development and validation of a risk prediction model for caesarean delivery among multiparous women. Sci Rep. 2025;15(1):5326. doi: 10.1038/s41598-025-12990-9