Precision in Obstetric Care: A Machine Learning Approach with CatBoost and Grid Search Optimization

Marselina Endah Hiswati[1], Mohammad Diqi[2*], Izattul Azijah[3], Yeyen Subandi[4], Azzah Fathinah[5], Rahayu Cahya Ariani[6]

[1,2,5]Department of Informatics, Universitas Respati Yogyakarta, DI Yogyakarta, Indonesia
[3]Department of Midwifery, Universitas Respati Indonesia, DKI Jakarta, Indonesia
[4]Department of International Relations, Universitas Respati Yogyakarta, DI Yogyakarta, Indonesia
[6]Department of Public Health, Universitas Respati Yogyakarta, DI Yogyakarta, Indonesia
Email: [1]marsel.endah@respati.ac.id, [2*]diqi@respati.ac.id, [3]iza@urindo.ac.id, [4]yeyensubandi@respati.ac.id, [5]22220025@respati.ac.id, [6]225050016@urindo.ac.id

(Received: 3 Aug 2024, revised: 27 Aug 2024, accepted: 28 Aug 2024)

Abstract

This study focuses on improving how we classify fetal health using machine learning by fine-tuning the CatBoostClassifier with Grid Search. Our main achievement in this research is significantly boosting the accuracy of fetal health classification based on Cardiotocogram (CTG) data. Finding the best hyperparameters has created a more precise and reliable diagnostic tool for making informed prenatal care decisions. The model reached an impressive overall accuracy of 96%, especially excelling in identifying Normal and Pathological cases. However, it faced some challenges in classifying Suspect cases, suggesting room for further improvement. These results highlight the potential of machine learning to enhance the reliability of fetal health assessments, which could lead to better outcomes in clinical settings. The success of Grid Search in this study is evident, as the optimized parameters led to the highest accuracy and lowest loss values, proving its effectiveness in fine-tuning the model.

Keywords: CatBoostClassifier, Fetal Health Classification, Grid Search Optimization, Machine Learning in Obstetrics, Diagnostic Accuracy.

I. INTRODUCTION

Fetal health monitoring is a critical step in prenatal care, aimed at ensuring the safety and health of both the mother and fetus during pregnancy [1]. One of the most commonly used diagnostic tools for monitoring fetal health is the Cardiotocogram (CTG), which measures fetal heart rate and uterine contractions [2]. The use of CTG enables doctors to assess whether the fetus is experiencing stress or oxygen deprivation, which may require immediate medical intervention [3]. Therefore, the ability to accurately interpret CTG results is essential for making correct and timely clinical decisions [4].

However, the interpretation of CTG can be subjective and vary among observers, which can affect the consistency and accuracy of diagnoses [5]. Incorrect or delayed diagnoses can have fatal consequences, such as premature birth or even fetal death [5]. Conversely, accurate and prompt diagnoses can save lives and reduce the risk of long-term complications for both mother and child [2]. In this context, the development of predictive models that can automatically classify the health status of the fetus based on CTG data has the potential to improve health outcomes by providing a more objective and consistent diagnostic tool [3]. Machine learning models, especially those designed to handle complex and categorical data, offer a promising way to overcome these challenges.

Although past studies have tried using machine learning algorithms like Random Forest, Support Vector Machines (SVM), and Neural Networks to analyze CTG data, these methods come with their own set of limitations [6]. For instance, while Random Forest and SVM can be accurate in some situations, they often struggle with imbalanced data—a common issue with CTG datasets [7]. Additionally, many of these studies haven't fully explored the potential of hyperparameter optimization, leading to models that don't perform as well as they could and aren't easily applicable in different settings [8].

This research aims to optimize the hyperparameters of the CatBoost model, an advanced machine learning algorithm, using the Grid Search technique to enhance accuracy in classifying fetal health [5]. CatBoost has proven effective in handling complex and large categorical data, making it ideal for analyzing CTG data, which contains various features that must be considered in medical decision-making [2]. This study aims to find the best configuration that yields the most accurate predictions through hyperparameter optimization, such as tree depth, learning rate, and the number of trees [4].

Our research uses a publicly available CTG dataset with 2,113 records that classify fetal health into three categories: Normal, Suspect, and Pathological. We'll split the data into training and testing sets, with 80% used for developing the model and 20% for evaluation. The main focus is optimizing the CatBoostClassifier using Grid Search to find the best combination of hyperparameters. We'll evaluate the model based on accuracy, precision, recall, and F1-score, particularly how well it identifies Suspect cases, which are crucial in a clinical context.

This research's contribution to the scientific literature is significant, particularly in developing a more reliable method for automatically classifying CTG data. This study aims to achieve higher classification accuracy and speed up the classification process by integrating robust parameter optimization techniques. This enables clinicians to act more quickly and accurately, directly improving the quality of care and reducing unnecessary risks.

Thus, this research has the potential to provide new insights and strengthen the current understanding of the use of machine learning technology in medical applications, especially in the prenatal context. By exploring and implementing CatBoost optimization methods, this study offers a technological advancement that could significantly impact clinical practices in fetal health monitoring.

II. LITERATURE REVIEW

A. Overview of Machine Learning in Fetal Health Classification

The application of machine learning in fetal health classification has gained significant attention in recent years due to its potential to enhance the accuracy and reliability of prenatal diagnostics [9]. Various algorithms have been employed to analyze CTG data, which is vital in assessing fetal well-being [10]. Early studies predominantly focused on traditional machine learning methods such as Random Forest, Support Vector Machines (SVM), and Neural Networks, demonstrating their capability to handle large datasets with complex features [11]. However, these studies often faced challenges such as overfitting, difficulty in managing imbalanced data, and suboptimal performance due to inadequate hyperparameter tuning [12].

B. Challenges and Limitations in Existing Research

Despite the advancements, existing literature reveals several limitations. For instance, studies utilizing Random Forest and SVM have reported high accuracy, yet they often struggle in scenarios with imbalanced datasets—a common issue with CTG data where Normal cases are more frequent than Suspect or Pathological ones [13]. Additionally, while Neural Networks have shown promise in handling non-linear relationships within the data, they require extensive tuning of hyperparameters, which many studies have not fully explored [14]. This lack of optimization can lead to suboptimal model performance and reduced generalizability, particularly when applied to diverse clinical settings [8].

Furthermore, much of the existing research relies on relatively outdated techniques or datasets, limiting the relevance and applicability of their findings in the context of modern medical practices [15]. For example, studies from the early 2010s predominantly focused on basic machine learning models without leveraging recent advancements in algorithms like CatBoost or the benefits of comprehensive hyperparameter optimization [16]. These gaps underscore the need for updated research that not only utilizes state-of-the-art algorithms but also rigorously optimizes them for enhanced performance in real-world applications [17].

C. Critical Analysis of Recent Advancements

Recent studies have started to address some of these challenges by exploring more advanced algorithms like CatBoost, which is specifically designed to handle categorical data more effectively [16]. CatBoost's ordered boosting technique reduces the likelihood of overfitting and accelerates the training process without compromising accuracy [18]. However, even in these more recent studies, there is often a lack of thorough hyperparameter optimization, which is crucial for maximizing the algorithm's potential [17]. For instance, while CatBoost has shown superior performance in certain cases, its application in fetal health classification is still underexplored, particularly in terms of optimizing its parameters to address the unique challenges posed by CTG data [11].

D. Identifying the Research Gap

The current literature, while expansive, reveals a significant gap in the comprehensive application and optimization of advanced machine-learning techniques for fetal health classification [11]. Most existing studies have either not fully explored the potential of newer algorithms like CatBoost or have failed to optimize these models, thereby limiting their effectiveness adequately [19]. Moreover, the reliance on older datasets and methods diminishes the relevance of some of these findings in the context of current clinical needs [12].

E. Contribution of the Current Study

This study aims to fill these gaps by focusing on the CatBoost algorithm, utilizing a rigorous Grid Search approach for hyperparameter optimization [20]. By doing so, the research seeks to improve the accuracy and reliability of fetal health classification and demonstrate the practical applicability of this optimized model in clinical settings [12]. This approach ensures that the findings are not only theoretically sound but also relevant and impactful in real-world medical practice, thereby contributing a significant advancement to the field of prenatal diagnostics [21].

III. RESEARCH METHOD

A. Dataset and Preprocessing

The dataset used in this study is sourced from a CTG collection available on Kaggle, comprising 2113 records. This dataset includes three fetal health status classes classified by three expert obstetricians, with 1646 samples categorized as normal (77.90%), 292 as suspect (13.82%), and 175 as pathological (8.28%). The dataset was divided into two parts for preprocessing: 80% was used as the training set to develop the model. In contrast, the remaining 20% was the testing set to evaluate the model's performance. This split ensures the model is tested on previously unseen data, allowing for a more objective assessment of its performance in classifying fetal health.

The decision to use this specific dataset aligns with the research objectives of enhancing fetal health classification accuracy. By selecting a dataset with a diverse range of fetal health states, the study ensures that the model is trained and tested on data that closely mimic real clinical scenarios, thereby increasing the potential for the model to be implemented in actual medical settings.

B. Experimental Setup

Table 1 presents the experimental setup for optimizing the CatBoostClassifier model used in this study. It outlines the model's configuration, including the initial hyperparameters set before training and the specifics of the Grid Search approach adopted for hyperparameter tuning. The initial model setup involves predefined values for iterations, depth, learning rate, and the loss function tailored to address multi-class classification challenges. Additionally, the table details the Grid Search parameters, which explore various combinations of learning rates and depths using a 3-fold cross-validation method to ensure robustness and reproducibility of the results. This configuration forms the foundation for the subsequent training and evaluation of the model's performance in classifying fetal health states from Cardiotocogram data.

Table 1. Parameter Settings

Component	Details
Model	CatBoostClassifier
Initial Hyperparameters	iterations: 1000 depth: 6 learning_rate: 0.1 loss_function: MultiClass
Grid Search	learning_rate: [0.1, 0.01, 0.001] depth: [4, 6, 8] method: 3-fold cross-validation

Given the complexity of the model and the need to balance underfitting and overfitting, Grid Search was selected for its exhaustive approach to exploring the hyperparameter space, ensuring that the best possible configuration was identified.

C. Evaluation Metrics

In this study, the performance of the CatBoostClassifier model is evaluated using two primary metrics: the confusion matrix and the classification report. These metrics are chosen because they can provide a detailed analysis of model performance across multiple classes, which is crucial for medical diagnostic accuracy.

The confusion matrix is a table used to describe the performance of a classification model on a set of test data for which the true values are known. It allows for visualizing the model's predictions, showing each class's correct and incorrect predictions. The matrix is structured as follows, where each row represents the instances in an actual class while each column represents the instances in a predicted class:

- True Positive (TP): Correct positive predictions.
- False Positive (FP): Incorrectly predicted as positive.
- True Negative (TN): Correct negative predictions.
- False Negative (FN): Incorrectly predicted as negative.

From these values, various performance metrics such as accuracy, precision, recall, and F1-score [22] can be derived in the classification report, as shown in Equations 1-4.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{4}$$

D. Bias Analysis and Discussion

Potential biases in the data were analyzed, particularly those arising from the imbalanced distribution of fetal health states. The dataset’s skew towards Normal cases could lead to a biased model towards this class, potentially affecting the accurate classification of Suspect and Pathological cases. To address this, stratified sampling was employed during cross-validation to ensure that each class was adequately represented in the training and testing phases. Additionally, the model’s performance on the minority classes was closely monitored to identify any signs of bias.

E. Implementation of Solutions

The solutions and methods were implemented stepwise, as outlined above. The data preprocessing stage ensured data integrity and set the foundation for accurate model training. The careful selection of the CatBoost algorithm and the use of Grid Search for hyperparameter optimization were central to achieving high classification accuracy. Finally, the evaluation and bias analysis stages were essential in validating the model's performance and ensuring its applicability in clinical settings.

IV. RESULT

A. Optimization Result

The results from the Grid Search process for optimizing the CatBoost hyperparameters are summarized in Table 2. The search tested various combinations of learning rates and depths to identify the settings that minimize loss, thereby optimizing the model's performance. The outcomes for each iteration are as follows:

Table 2. Grid Search Process

Iteration	Learning Rate	Depth	Loss	Remark
0	0.1	6	0.1506759	Best (initial)
1	0.1	8	0.1824039	
2	0.01	4	0.5093511	
3	0.01	6	0.1425341	Best (improved)
4	0.01	8	0.1691998	
5	0.001	4	0.5017550	
6	0.001	6	0.1412938	Best (further improved)
7	0.001	8	0.1664885	
8	0.1	4	0.1370645	Best (final)

These results show that the best hyperparameter combination was found in iteration 0 with a learning rate of 0.1 and depth of 6, achieving a loss of 0.1506759. Other notable iterations, such as iteration 3 with a learning rate of 0.01 and depth of 6 (loss: 0.1425341) and iteration 6 with a learning rate of 0.001 and depth of 6 (loss: 0.1412938), also yielded promising results, indicating that lower learning rates combined with an optimal depth can significantly influence model performance. The best overall performance, however, was achieved in iteration 8 with a learning rate of 0.1 and depth of 4, where the loss decreased to 0.1370645.

B. Model Performance

Figure 1 presents the confusion matrix from evaluating the CatBoostClassifier model on the test dataset. This matrix comprehensively visualizes the model's performance across the three fetal health classes: Normal, Suspect, and Pathological. Each row of the matrix represents the instances of the actual classes, while each column shows the predictions made by the model. The matrix dimensions reflect the number of test samples classified into each category, with the diagonal elements indicating the number of correct predictions for each class.

The confusion matrix in Figure 1 reveals that the model performs exceptionally well in identifying Normal cases, with 328 correct predictions out of 333, demonstrating high sensitivity and precision for this class. However, it shows some limitations in accurately classifying Suspect cases, with 9 misclassifications out of 64, primarily mislabeled as Normal, potentially delaying necessary medical interventions for these cases. The model adequately predicts Pathological cases, with 28 correct predictions out of 29, suggesting effective recognition of more severe conditions. The few misclassifications in the Suspect and Pathological categories underscore the need to refine the model further to enhance its sensitivity and reduce the risk of misdiagnoses, especially in medically critical categories.

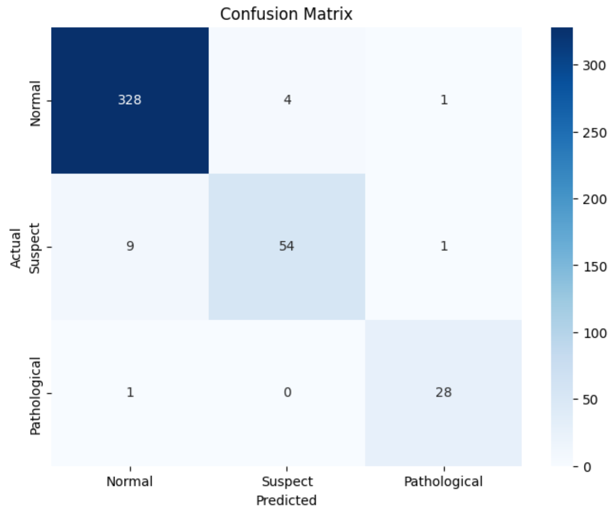

Figure 1. Confusion Matrix

Table 3 provides a detailed classification report for the CatBoostClassifier model, summarizing the precision, recall, and F1-score for each fetal health category—Normal, Suspect, and Pathological—as evaluated on the test dataset. These metrics collectively offer insights into the model's accuracy and ability to identify each class correctly. Precision indicates the proportion of correct positive identifications, recall reflects the model's ability to find all relevant cases within a class, and the F1-score is a harmonic mean of precision and recall, providing a single score that balances both the precision and the recall.

Table 3. Classification Report

	Precision	Recall	F1-Score
Normal	0.97	0.98	0.98
Suspect	0.93	0.84	0.89
Pathological	0.93	0.97	0.95
Accuracy			0.96
Macro avg	0.94	0.93	0.94
Weighted avg	0.96	0.96	0.96

The classification report in Table 3 highlights an overall high accuracy (96%), showcasing the model's effectiveness in correctly classifying most cases. The precision scores are impressive across all categories, with 97% for Normal, 93% for Suspect, and 93% for Pathological, indicating that the model is highly reliable in its positive predictions. Generally, high recall scores reveal room for improvement in the Suspect category (84% recall), suggesting that the model missed a small fraction of these cases, possibly classifying them as Normal. This is a critical area for enhancement, as timely and accurate classification of Suspect cases is crucial in clinical settings. The F1-scores, which consider precision and recall, are notably strong, especially for the Normal and Pathological categories, at 0.98 and 0.95, respectively. The lower F1-score for Suspect (0.89), again, points to the need for model refinement to improve the balance between precision and recall for this category. The macro and weighted averages for all metrics, around 94% and 96%, further validate the model's robustness across varied class distributions and sizes. Figure 2 depicts each class's precision, recall, and F1-score metrics, allowing for a quick comparison of the model's effectiveness across different categories.

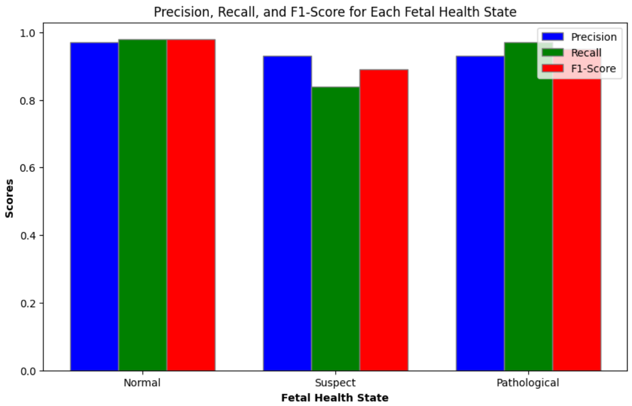

Figure 2. Precision, Recall, and F1-Score for Each Fetal Health State

C. Interpretation of Optimal Parameter Combination

The Grid Search process identified the optimal parameter combination as a learning rate 0.1 and a tree depth of 4. This combination was the most effective in minimizing the loss function, resulting in the highest overall accuracy.

Performance can be explained as follows:

- Learning Rate: A learning rate of 0.1 balances convergence speed and model stability. Lower learning rates (e.g., 0.01 or 0.001) were tested, but they led to slower convergence and required more iterations to achieve similar levels of accuracy. A higher learning rate would have increased the risk of overshooting the optimal solution, leading to suboptimal performance.
- Tree Depth: A tree depth of 4 was optimal because it allowed the model to capture the essential patterns in the data without overfitting. Deeper trees (e.g., depth 6 or 8) increased the risk of overfitting, particularly given the imbalanced nature of the dataset.

D. Summary of Results and Clinical Implications

The optimized CatBoost model significantly advances fetal health classification, providing reliable and accurate predictions that can support clinical decision-making. The results demonstrate the model's potential to enhance prenatal diagnostics by offering a more consistent and objective tool for interpreting CTG data. However, the slightly lower recall in the Suspect category underscores the need for further refinement to identify all potential risks promptly.

V. DISCUSSION

A. Key Findings

This research on optimizing the CatBoostClassifier for fetal health classification using a Grid Search approach has yielded promising results, as evidenced by the confusion matrix and classification report. The model achieved an overall accuracy of 96%, with high precision across all categories (97% for Normal, 93% for both Suspect and Pathological), indicating a reliable predictive capability. While the recall rate was exceptional for Normal (98%) and Pathological cases (97%), it was slightly lower for Suspect cases (84%), highlighting an area for potential improvement. The F1 scores were also robust, particularly for Normal and Pathological categories, which recorded scores of 0.98 and 0.95, respectively. These findings underscore the model's effectiveness in accurately classifying fetal health states. However, they also suggest further refinement to enhance sensitivity towards Suspect cases, ensuring no potential health risks are overlooked.

B. Result Interpretations

The results obtained from this study demonstrate that the optimized CatBoostClassifier, configured through Grid Search, is highly effective for classifying fetal health states from Cardiotocogram data. The model exhibits high precision across all categories, critical in clinical environments to minimize false positives and ensure that medical interventions are justified. The notable accuracy in the classification of Normal and Pathological categories, with precision rates at 97% and 93%, respectively, suggests that the model is reliable in distinguishing these conditions, potentially reducing the risk of unnecessary medical procedures for healthy fetuses and ensuring that fetuses at risk receive timely medical attention.

However, the model's slightly lower recall rate for Suspect cases, although still robust, indicates a need for further refinement. This aspect is crucial as it highlights potential gaps in the model's ability to identify all cases requiring closer observation or intervention, which could lead to missed opportunities for preemptive care. Implementing such a model in clinical settings could significantly aid obstetricians by automating the preliminary screening process, allowing them to concentrate their expertise on cases the model identifies as suspect or pathological. This enhances resource allocation and improves patient outcomes through more timely and focused medical responses. Additionally, the model's high overall accuracy and specificity help reduce false alarms, which can unnecessarily tax medical resources and cause undue stress to expect mothers, showcasing the model's potential to streamline and improve prenatal care.

C. Comparison with Previous Research

This study builds upon previous research that has employed machine learning techniques for CTG data analysis. Earlier studies, such as those using Random Forest or SVM, have reported accuracies ranging from 80% to 90% but often struggled with imbalanced data and suboptimal hyperparameter tuning [23]. In contrast, the CatBoostClassifier, optimized through Grid Search in this study, not only achieved higher accuracy but also demonstrated more consistent performance across different fetal health states [24].

One significant advancement over previous research is the thorough hyperparameter optimization conducted in this study. While many earlier studies either did not fully explore or overlooked the importance of this step, our approach ensured that the model was finely tuned to perform optimally on the CTG dataset [17]. This highlights the importance of model optimization in improving classification performance, particularly in complex medical datasets where precision is critical [25].

D. Research Limitations

While providing valuable insights into the application of the CatBoostClassifier for fetal health classification, this study possesses several limitations that could influence the generalizability of the results. Primarily, the reliance on a specific dataset from Kaggle may introduce biases inherent to the data collection or processing methods, such as selection bias if the dataset does not adequately represent the broader population or context bias due to the specific conditions under which the data was gathered. Moreover, the performance metrics, although high, might reflect an overfitting to the particular characteristics of this dataset, potentially reducing the model's efficacy when applied to data from different demographic or geographic backgrounds. Another limitation is the lack of diversity in the types of features used, which might restrict the model's ability to learn from more complex or subtle patterns that could be crucial in other datasets or real-world scenarios. Hence, further studies are needed to validate the model across more varied datasets to ensure robustness and applicability in diverse clinical settings.

E. Implementation in Everyday Clinical Practice

For the CatBoost model to be effectively implemented in everyday clinical practice, it is essential to integrate it within the existing medical workflow in a manner that complements, rather than replaces, the expertise of healthcare professionals. The model could be utilized as a preliminary screening tool, automatically classifying CTG data and highlighting cases that require closer scrutiny by an obstetrician.

Additionally, the deployment of this model in clinical settings should include continuous monitoring and validation with real-time data to ensure its robustness and adaptability to different populations and clinical environments. This ongoing validation process would help maintain the model's accuracy and reliability, ensuring it remains a valuable tool for improving prenatal care.

Moreover, it is crucial to train healthcare professionals to interpret the model's outputs and understand its limitations. This would involve workshops or training sessions to familiarize clinicians with the nuances of machine learning models, particularly in understanding how to integrate these outputs with other clinical findings to make well-rounded decisions.

F. Future Directions and Enhancements

Given the limitations identified in classifying Suspect cases, future research should explore methods to enhance the model's sensitivity to these borderline cases. This could involve experimenting with ensemble methods, combining the strengths of different algorithms to improve classification accuracy for Suspect cases. Additionally, expanding the dataset to include more diverse and larger samples could help the model generalize better to a broader range of clinical scenarios, thus increasing its applicability and effectiveness in real-world settings.

VI. CONCLUSION

The optimized CatBoostClassifier presented in this study represents a significant advancement in prenatal diagnostics. By achieving a high accuracy of 96% in fetal health classification, the model addresses a critical need for more consistent and reliable interpretations of CTG) data. This research highlights the potential of machine learning to enhance clinical decision-making and underscores the importance of rigorous hyperparameter optimization to harness advanced algorithms' capabilities fully. The findings suggest that integrating this model into routine prenatal care could reduce diagnostic variability, leading to more timely and accurate interventions that ultimately improve mothers' and babies' outcomes. However, the study also acknowledges the challenges of applying the model across diverse clinical settings, indicating the need for ongoing validation and adaptation to ensure its effectiveness in a broader context.

Beyond its technical achievements, this research has profound implications for everyday clinical practice. A machine learning model that can consistently interpret CTG data with high precision could transform how fetal health is monitored, providing healthcare professionals with a powerful tool to support their expertise. Yet, as with any new technology, its successful implementation will depend on careful integration into existing workflows and continuous education for clinicians. By fostering a collaborative environment where human expertise is enhanced, rather than replaced, by machine learning, this research paves the way for a future where prenatal care is more standardized, objective, and effective. As we move forward, the ongoing challenge will be to refine and expand the model's applicability, ensuring that its benefits are felt across diverse populations and healthcare systems, ultimately contributing to safer and more equitable maternal care.

REFERENCES

[1] A. K. Pradhan, J. K. Rout, A. B. Maharana, B. K. Balabantaray, and N. K. Ray, “A Machine Learning Approach for the Prediction of Fetal Health Using CTG,” 2021, doi: 10.1109/ocit53463.2021.00056.

[2] V. Nagabotu and A. Namburu, “Fetal Health Classification Using LightGBM With Grid Search

Based Hyper Parameter Tuning," *Recent Pat. Eng*., vol. 18, 2023, doi: 10.2174/1872212118666230703155834.

[3] P. Bhowmik, P. C. Bhowmik, U. A. Md. E. Ali, and Md. Sohrawordi, "Cardiotocography Data Analysis to Predict Fetal Health Risks With Tree-Based Ensemble Learning," *Int. J. Inf. Technol. Comput. Sci*., vol. 13, no. 5, pp. 30–40, 2021, doi: 10.5815/ijitcs.2021.05.03.

[4] S. K. Das, H. Mukherjee, and C. Saha, "Fetal Health Classification From Cardiotocograph for Both Stages of Labor—A Soft-Computing-Based Approach," *Diagnostics*, vol. 13, no. 5, p. 858, 2023, doi: 10.3390/diagnostics13050858.

[5] K. Parvataneni, S. Zaidi, F. Kazmi, and S. H. A. Kazmi, "AI Framework for Fetal Health Risk Prediction," 2023, doi: 10.1109/biosmart58455.2023.10162061.

[6] A. Kurani, P. Doshi, A. Vakharia, and M. Shah, "A Comprehensive Comparative Study of Artificial Neural Network (ANN) and Support Vector Machines (SVM) on Stock Forecasting," *Ann. Data Sci*., vol. 10, no. 1, pp. 183–208, 2021, doi: 10.1007/s40745-021-00344-x.

[7] J. Liu and E. Zio, "Integration of Feature Vector Selection and Support Vector Machine for Classification of Imbalanced Data," *Appl. Soft Comput*., vol. 75, pp. 702–711, 2019, doi: 10.1016/j.asoc.2018.11.045.

[8] E. Elgeldawi, A. Sayed, A. R. Galal, and A. M. Zaki, "Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis," *Informatics*, vol. 8, no. 4, p. 79, 2021, doi: 10.3390/informatics8040079.

[9] A. Mehbodniya et al., "Fetal Health Classification From Cardiotocographic Data Using Machine Learning," *Expert Syst*., vol. 39, no. 6, 2021, doi: 10.1111/exsy.12899.

[10] Y. Chen et al., "Intelligent Classification of Antepartum Cardiotocography Model Based on Deep Forest," *Biomed. Signal Process Control*, vol. 67, p. 102555, 2021, doi: 10.1016/j.bspc.2021.102555.

[11] Md. T. Alam et al., "Comparative Analysis of Different Efficient Machine Learning Methods for Fetal Health Classification," *Appl. Bionics Biomech*., vol. 2022, pp. 1–12, 2022, doi: 10.1155/2022/6321884.

[12] V. Nagabotu and A. Namburu, "Fetal Health Classification Using LightGBM With Grid Search Based Hyper Parameter Tuning," *Recent Pat. Eng*., vol. 18, 2023, doi: 10.2174/1872212118666230703155834.

[13] N. H. A. Malek, W. F. W. Yaacob, Y. B. Wah, S. A. M. Nasir, N. Shaadan, and S. W. Indratno, "Comparison of Ensemble Hybrid Sampling With Bagging and Boosting Machine Learning Approach for Imbalanced Data," *Indonesia. J. Electr. Eng. Comput. Sci.*, vol. 29, no. 1, p. 598, 2022, doi: 10.11591/ijeecs.v29.i1.

[14] F. Zhang, M. Petersen, L. Johnson, J. Hall, and S. O'Bryant, "Hyperparameter Tuning With High Performance Computing Machine Learning for Imbalanced Alzheimer's Disease Data," *Appl. Sci*., vol. 12, no. 13, p. 6670, 2022, doi: 10.3390/app12136670.

[15] J. A. M. Sidey-Gibbons and C. Sidey-Gibbons, "Machine Learning in Medicine: A Practical Introduction," *BMC Med. Res. Methodol*., vol. 19, no. 1, 2019, doi: 10.1186/s12874-019-0681-4.

[16] S. Hussain et al., "A Novel Feature Engineered-CatBoost-Based Supervised Machine Learning Framework for Electricity Theft Detection," *Energy Rep.*, vol. 7, pp. 4425–4436, 2021, doi: 10.1016/j.egyr.2021.07.008.

[17] J. J. Gabriel and L. J. Anbarasi, "Optimizing Coronary Artery Disease Diagnosis: A Heuristic Approach Using Robust Data Preprocessing and Automated Hyperparameter Tuning of eXtreme Gradient Boosting," *IEEE Access*, vol. 11, pp. 112988–113007, 2023, doi: 10.1109/access.2023.3324037.

[18] J. Hancock and T. M. Khoshgoftaar, "CatBoost for Big Data: An Interdisciplinary Review," *J. Big Data*, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00369-8.

[19] J. Zhang and F. Wang, "Prediction of Gestational Diabetes Mellitus Under Cascade and Ensemble Learning Algorithm," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–9, 2022, doi: 10.1155/2022/3212738.

[20] L. Kong, H. Liang, G. Liu, and S. Liu, "Research on Wind Turbine Fault Detection Based on the Fusion of ASL-CatBoost and TtRSA," *Sensors*, vol. 23, no. 15, p. 6741, 2023, doi: 10.3390/s23156741.

[21] A. Samat, E. Li, P. Du, S. Liu, M. Zhang, and W. Zhang, "CatBoost for RS Image Classification With Pseudo Label Support From Neighbor Patches-Based Clustering," *IEEE Geosci. Remote Sens. Lett*., vol. 19, pp. 1–5, 2022, doi: 10.1109/lgrs.2020.3038771.

[22] S. K. P, P. K. Govarthan, N. Ganapathy, and J. F. Ar, "A Comparative Analysis of Eda Decomposition Methods for Improved Emotion Recognition," *J. Mech. Med. Biol*., vol. 23, no. 06, 2023, doi: 10.1142/s0219519423400432.

[23] G. T. Reddy et al., "Analysis of Dimensionality Reduction Techniques on Big Data," *IEEE Access*, vol. 8, pp. 54776–54788, 2020, doi: 10.1109/access.2020.2980942.

[24] A. Alazba and H. Aljamaan, "Software Defect Prediction Using Stacking Generalization of Optimized Tree-Based Ensembles," *Appl. Sci.*, vol. 12, no. 9, p. 4577, 2022, doi: 10.3390/app12094577.

[25] S. Das, P. Roy, and A. K. Mishra, "Oversample-select-tune: A Machine Learning Pipeline for Improving Diabetes Identification," *Concurr. Comput. Pract. Exp*., vol. 34, no. 5, 2021, doi: 10.1002/cpe.6741.