International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

AN OPTIMIZED MACHINE LEARNING FRAMEWORK FOR ACCURATE FETAL HEALTH CLASSIFICATION USING ACTIVE SAMPLING

Asfia Sultana[1], Dr. I. Samuel Peter James[2]
[1]PG Scholar, Department of CSE, Shadan Women's College of Engineering and Technology, Hyderabad
sultanaasfiya9@gmail.com
[2]Professor & Head, Department of CSE, Shadan Women's College of Engineering and Technology, Hyderabad
drisamuelpeterjames@gmail.com

ABSTRACT
In maternal and foetal healthcare, pregnancy problems continue to be a significant concern. Effective intervention depends on early and precise detection. The time-consuming and frequently subjective nature of traditional manual interpretation of cardiotocography (CTG) data results in inconsistent assessments of foetal health. The current work suggests an effective machine learning-based approach that makes use of the XGBoost algorithm, which is renowned for its exceptional performance in managing structured data, in order to address this. The suggested model, which makes use of a publicly accessible CTG dataset, outperforms previous methods with a noteworthy accuracy of 96%. This illustrates how XGBoost can reliably simulate intricate patterns in foetal monitoring data. The technology facilitates prompt clinical decision-making, improves diagnostic accuracy, and lessens the strain for clinicians. In order to improve maternal and neonatal outcomes and advance automated foetal health evaluation, our work demonstrates the potential of incorporating strong machine learning models into prenatal care workflows.

1. INTRODUCTION
In order to ensure safe pregnancies and lower rates of mother and newborn mortality, foetal health monitoring is essential. One of the most popular methods for keeping an eye on the health of the foetus is cardiotocography (CTG), which measures the foetal heart rate and uterine contractions. However, human interpretation of CTG results might result in delayed or incorrect clinical judgements because it is frequently subjective, time-consuming, and prone to inter-observer variability. Machine learning (ML) approaches present a promising option for automating and enhancing the accuracy of foetal health classification, given the growing availability of foetal health data. Among these, sophisticated algorithms like XGBoost have demonstrated exceptional performance because of their capacity to effectively handle intricate, non-linear data patterns. ML-based methods lessen the workload for medical personnel while simultaneously increasing diagnostic accuracy.In order to improve model performance while lowering the quantity of labelled data needed, we present a system in this paper that combines machine learning with active sampling, a method that chooses the most informative data points for training. With fewer training data, our method guarantees that the model learns from a variety of ambiguous cases, increasing accuracy. Our method uses active sampling and XGBoost to increase foetal health status classification accuracy, allowing for prompt treatments and better clinical results.

2. EXISTING SYSTEM OF PROJECT
- Several conventional machine learning techniques have been investigated in recent years for the interpretation of cardiotocography (CTG) data. However, these models' diagnostic reliability in foetal health monitoring is limited because they frequently fail to reflect the intricate and nonlinear correlations found in CTG data.
- A innovative approach utilising Kolmogorov-Arnold Networks (KANs) has been developed to overcome these constraints. KANs are very good at modelling nonlinearity in medical data because they substitute learnable univariate functions for typical neural network activations. This makes it possible to interpret foetal health issues from CTG records with greater accuracy.
- In both binary and multiclass classification tasks, the study showed that KANs perform better than traditional models. In particular, they demonstrated their efficacy in managing intricate physiological patterns in CTG data by achieving 92.6% for three-class categorisation.

Existing System Disadvantages:
- High Computational Cost
- Overfitting Risk
- Complicated Implementation
- Lack of Standardization

3. RELATED WORK
A study conducted in 2023 explored the use of ensemble learning techniques for fetal health classification using Cardiotocography (CTG) data. The researchers investigated several machine learning algorithms,

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

including Random Forest, AdaBoost, CatBoost, XGBoost, and LightGBM, to classify fetal conditions into healthy, suspected, and pathological categories. To address the issue of class imbalance in CTG datasets, the study employed a random over-sampling technique and developed a soft-voting ensemble model that combined the predictions of individual classifiers. Experimental results demonstrated that the ensemble approach achieved superior classification performance compared to standalone models, highlighting the effectiveness of ensemble learning and boosting algorithms for improving fetal health assessment accuracy and supporting clinical decision-making.

In 2025, researchers proposed a predictive ensemble model for fetal health classification using Cardiotocography data. The study focused on important fetal health indicators such as prolonged decelerations, abnormal short-term variability, long-term variability, histogram characteristics, and acceleration patterns. Machine learning techniques including XGBoost and LightGBM were employed individually and in combination to classify fetal health conditions. The ensemble framework achieved an accuracy exceeding 93%, outperforming individual models and demonstrating the effectiveness of combining multiple boosting algorithms. The research emphasized that machine learning-based fetal monitoring systems can provide reliable diagnostic support for healthcare professionals and contribute to improved pregnancy outcomes.

Another study published in 2023 performed a comparative analysis of machine learning methods for identifying neonatal brain injury risks using perinatal Cardiotocography data. The research evaluated several algorithms, including XGBoost, Random Forest, LightGBM, Decision Tree, and K-Nearest Neighbors on a dataset containing 2,126 fetal monitoring records. The findings revealed that XGBoost achieved the highest classification accuracy of approximately 91%, along with superior precision, recall, and F1-score values. The study demonstrated the feasibility of using machine learning techniques for objective fetal health assessment and highlighted the potential of advanced predictive models to assist clinicians in diagnosing fetal distress and preventing neonatal complications.

Research conducted in 2024 compared multiple machine learning algorithms for fetal health classification using Cardiotocography measurements. The study examined the performance of Artificial Neural Networks, Long Short-Term Memory networks, XGBoost, Support Vector Machines, K-Nearest Neighbors, LightGBM, and Random Forest models across several experimental scenarios. Results indicated that most algorithms achieved high classification accuracy, ranging between 89% and 99%. Among all methods, LightGBM consistently delivered strong performance across different evaluation settings, demonstrating its robustness and suitability for fetal health prediction tasks. The study reinforced the value of machine learning techniques in enhancing the accuracy and reliability of fetal monitoring systems.

A novel active learning framework for fetal health classification was introduced in 2025 using the XGBoost classifier. The proposed approach selectively identified the most informative samples for labeling, thereby reducing annotation costs while maintaining high predictive performance. By combining active learning strategies with the powerful classification capabilities of XGBoost, the model achieved an accuracy of nearly 99% using significantly fewer labeled training samples. The study highlighted the efficiency of active learning in medical data analysis and demonstrated its potential to develop cost-effective and highly accurate fetal health monitoring systems for real-world clinical applications.

4. PROPOSED SYSTEM

- The XGBoost algorithm, a potent and effective gradient boosting technique, is used in the suggested system to increase the accuracy of foetal health categorisation.
- This method offers a scalable and dependable solution for clinical decision-making, acknowledging the limits of human CTG interpretation and the complexity of the data.
- The CTG dataset's complex patterns and non-linear relationships are handled using XGBoost.
- It is perfect for medical datasets because of its integrated regularisation, parallel processing, and capacity to handle missing values.
- The model focuses on identifying three health states: Normal, Suspect, and Pathological.
- It is trained using features taken from foetal heart rate and uterine contraction signals.

Proposed System Advantages:

- Excellent Accuracy and Performance
- Integrated Regularisation.
- Adaptability
- Resilience to multicollinearity

5. MODULES

1) Dataset

A comprehensive collection of maternal and foetal health records, with a particular focus on pregnancies complicated by aberrant cardiotocography (CTG) readings, is the dataset used in this investigation. Several measurements and parameters that are essential for

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

assessing and forecasting pregnancy outcomes are included in this collection.

2) Import the libraries
Sklearn.model_selection.train_test_split: This function divides your data into training and testing sets, which is essential for assessing your model's performance on unseen data. To load the required libraries for preprocessing, data splitting, and applying XGBoost.preparation. **StandardScaler**: This scaler scales to unit variance and eliminates the mean in order to standardise characteristics. Numerical data is frequently preprocessed using it before being fed into machine learning models. **xgboost**: This is the actual XGBoost library, a well-liked gradient boosting library renowned for its effectiveness and quickness.

3) Analyzing
This procedure entails loading your dataset, preparing it to get it ready for modelling, dividing it into training and testing sets, creating an XGBoost model, and assessing its effectiveness using metrics like accuracy and a classification report. Depending on the features of your particular dataset and the goals of your study, changes can be required.

4) Preprocessing
By ensuring that your data is clean, properly encoded, and scaled for efficient modelling using algorithms like XGBoost, preprocessing procedures offer an organised method for getting your dataset ready for machine learning research. Modifications can be made according to the particulars of your dataset and your modelling objectives.

5) Splitting The data
In order to enable an objective assessment of model performance, data splitting is an essential stage in the creation of machine learning models. It gives you insights into your model's capacity for generalisation by enabling you to train it on one subset and validate it on another. Depending on your dataset size and particular needs, modify the test_size parameter.

6) Model
Of course! An XGBoost model can be trained to extract a variety of model-related data, including model parameters and feature importance. The trained XGBoost model can provide feature significance ratings. When creating predictions, feature importance shows how important each feature is in relation to the others. The XGBoost model's training parameters are accessible. These factors include the number of boosting rounds, the maximum tree depth, and the learning rate. Usually made up of several decision trees, XGBoost models are ensemble models. To see how the model generates predictions, you can examine the structure of the trees or visualise them.

7) Accuracy
96% accuracy on Xgboost.

8) Result
There were [number of samples] occurrences with [number of features] features in the foetal health dataset. Foetal health status was represented by the target variable, which was divided into [number of classes] classes.

6. TECHNIQUES USED IN PROJECT

6.1. XGBoost

XGBoost is a durable and powerful algorithm that combines the advantages of gradient boosting methods with a number of improvements. It improves model accuracy repeatedly by minimising the loss when adding new models to the ensemble using a gradient descent optimisation technique.
XGBoost can directly handle missing data and is built to handle both regression and classification jobs.
XGBoost is faster and more accurate than classic boosting techniques because it incorporates a number of sophisticated features, including regularisation (to reduce overfitting), parallel processing, management of missing information, and tree pruning. It is frequently utilised in structured/tabular datasets and supports both regression and classification applications.XGBoost is faster and more accurate than classic boosting techniques because it incorporates a number of sophisticated features, including regularisation (to reduce overfitting), parallel processing, management of missing information, and tree pruning. It is frequently utilised in structured/tabular datasets and supports both regression and classification applications.
XGBoost is a popular choice for competitions and real-world applications in healthcare, finance, and other sectors due to its effectiveness, speed, and resilience.

6.2 SYSTEM ARCHITECTURE:

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

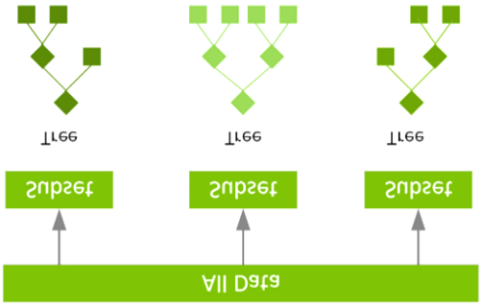


Fig 1: System Architecture

7. RESULTS

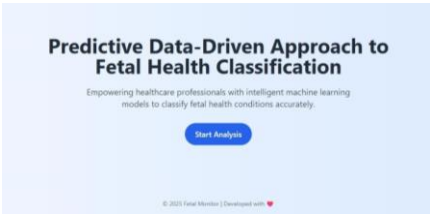


Fig 2: Main website

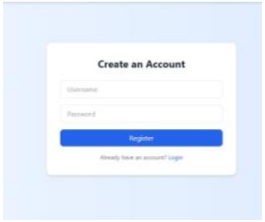


Fig 3: Login page

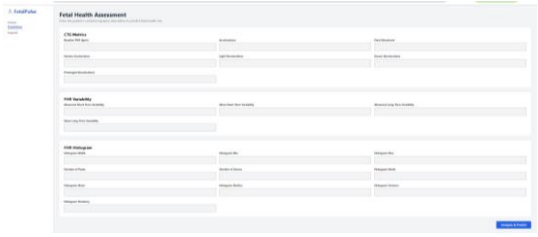


Fig 4: Enter Data

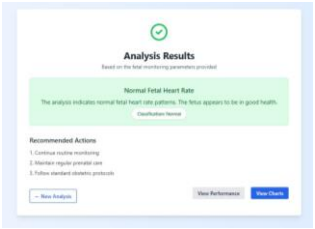


Fig 5: Result – if Normal

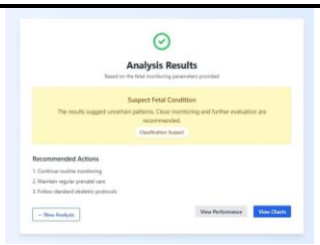


Fig 6: Result – if Suspect

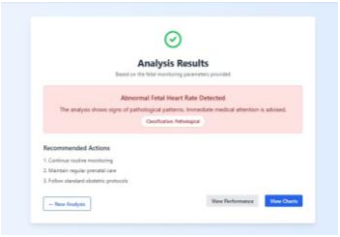


Fig 7: Result – if Pathological

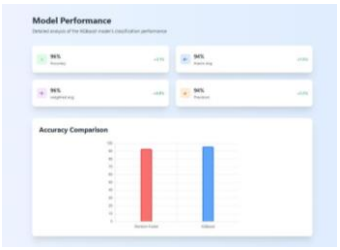


Fig 8: Performance

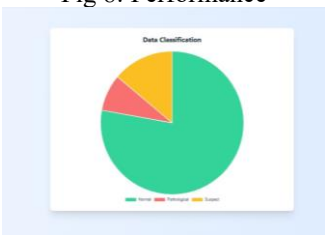


Fig 9: Charts

8. DISCUSSION AND CONCLUSION

To sum up, this study shows how the XGBoost machine learning method can be used effectively to categorise foetal health using cardiotocography (CTG) data. The suggested approach outperformed a number of conventional models with a high accuracy of 96% by utilising its potent gradient boosting capabilities. In addition to improving foetal health assessment reliability, this method lessens reliance on manual interpretation, which is frequently unreliable. The findings demonstrate how machine learning may enhance prenatal diagnosis and facilitate prompt medical interventions. This system has the potential to significantly improve maternal and foetal care with additional developments and incorporation into clinical settings.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

9. FUTURE ENHANCEMENT

In order to enable continuous data analysis and immediate risk detection during labour, this project can be improved in the future by incorporating real-time foetal monitoring devices. The model's diagnostic accuracy can be increased by adding multimodal inputs like maternal vitals, ultrasound images, and other physiological signals to the dataset. Furthermore, the predictions would be more visible if Explainable AI (XAI) were used, which would help medical professionals comprehend and have faith in the system's recommendations. Creating a mobile or cloud-based platform to increase the tool's accessibility in remote and underdeveloped places could be another improvement. Finally, more individualised and accurate medical interventions can be made possible by using adaptive learning processes to customise foetal health projections based on unique maternal and foetal histories.

REFERENCES

[1] B. A. Kamala et al., "Intrapartum fetal heart rate monitoring using a handheld Doppler versus Pinard stethoscope: A randomized controlled study in Dar es Salaam," Int. J. Women's Health, vol. 10, pp. 341–348,.2018.

[2] S. Saleem, S. S. Naqvi, T. Manzoor, A. Saeed, N. Ur Rehman, and J.Mirza, "A strategy for classification of 'vaginal vs. cesarean section'delivery: Bivariate empirical mode decomposition of cardiotocographic recordings," Front. Physiol., vol. 10, 2019, Art. no. 246.

[3] M. Ayat, K. Assaleh, and H. Nashash, "Fetal ECG extraction from a single abdominal ECG signal using SVD and polynomial classifiers," in Proc.2008 IEEE Workshop Mach. Learn. Signal Process., 2008, pp. 250–254.

[4] A. Pinas and E. Chandraharan, "Continuous cardiotocography during labour: Analysis, classification and management," Best Pract. Res. Clin. Obstet. Gynaecol., vol. 30, pp. 33–47, 2016.

[5] M. U. Khan et al., "Cardiotocography: An indicator to predict vaginal and cesarean deliveries," in Proc. 1st Int. Conf. Smart Syst. Emerg. Technol.,2020, pp. 77–82.

[6] P. Fergus, A. Hussain, D. Al-Jumeily, D.-S. Huang, and N. Bouguila, "Classification of caesarean section and normal vaginal deliveries using foetal heart rate signals and advanced machine learning algorithms," BioMedical Eng. OnLine, vol. 16, Jul. 2017, Art. no. 89.

[7] S. T. Nabipour Hosseini, F. Abbasalizadeh, S. Abbasalizadeh, S. Mousavi, and P. Amiri, "A comparative study of CTG monitoring one hour before labor in infants born with and without asphyxia," BMC Pregnancy Childbirth, vol. 23, Oct. 2023, Art. no. 758.

[8] O. Nzelu, E. Chandraharan, and S. Pereira, "Human factors: The dirty dozen in CTG misinterpretation," Glob. J. Reprod. Med., vol. 6, no. 2, pp. 34–39, 2018.

[9] D. K. Donker, H. P. van Geijn, and A. Hasman, "Interobserver variation in the assessment of fetal heart rate recordings," Eur. J. Obstet. Gynecol. Reprod. Biol., vol. 52, no. 1, pp. 21–28, 1993.

[10] Y.-J. Jia, T. Ghi, S. Pereira, A. Gracia Perez-Bonfils, and E. Chandraharan, "Pathophysiological interpretation of fetal heart rate tracings in clinical practice," Amer. J. Obstet. Gynecol., vol. 228, no. 6, pp. 622–644, 2023.

[11] P. N. Ignatov and J. E. Lutomski, "Quantitative cardiotocography to improve fetal assessment during labor: A preliminary randomized controlled trial," Eur. J. Obstet. Gynecol. Reprod. Biol., vol. 205, pp. 91–97,2016.

[12] Y. Salini, S. N. Mohanty, J. V. N. Ramesh, M. Yang, and M. M. V. Chalapathi, "Cardiotocography data analysis for fetal health classification using machine learning models," IEEE Access, vol. 12, pp. 26005–26022, 2024.

[13] R. Zeng, Y. Lu, S. Long, C. Wang, and J. Bai, "Cardiotocography signal abnormality classification using time-frequency features and ensemble cost-sensitive SVM classifier," Comput. Biol. Med., vol. 130, 2021, Art. no. 104218.

[14] Z. Cao et al., "Intelligent antepartum fetal monitoring via deep learning and fusion of cardiotocographic signals and clinical data," Health Inf. Sci. Syst., vol. 11, Mar. 2023, Art. no. 16.

[15] M. Zhong et al., "CTGNet: Automatic analysis of fetal heart rate from cardiotocograph using artificial intelligence," Maternal-Fetal Med., vol. 4, no. 2, pp. 103–112, 2022.

[16] J. Ogasawara et al., "Deep neural network-based classification of cardiotocograms outperformed conventional algorithms," Sci. Rep., vol. 11, Jun. 2021, Art. no. 13367.

[17] Y. Chen et al., "Intelligent classification of antepartum cardiotocography model based on deep forest," Biomed. Signal Process. Control, vol. 67, 2021, Art. no. 102555.

[18] N. M. Hussain, A. U. Rehman, M. T. B. Othman, J. Zafar, H. Zafar, and H. Hamam, "Accessing artificial intelligence for fetus health status using hybrid deep learning algorithm (AlexNet-SVM) on cardiotocographic data," Sensors, vol. 22, no. 14, 2022, Art. no. 5103.

[19] A. Petrozziello, C. W. G. Redman, A. T. Papageorghiou, I. Jordanov,and A. Georgieva, "Multimodal convolutional neural networks to detect fetal compromise during labor and delivery," IEEE Access, vol. 7,pp. 112026–112036, 2019.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

[20] M. Liu, Y. Lu, S. Long, J. Bai, and W. Lian, "An attention-based CNNBiLSTM hybrid neural network enhanced with features of discrete wavelet transformation for fetal acidosis classification," Expert Syst. With Appl., vol. 186, 2021, Art. no. 115714.
[21] E. Spairani, B. Daniele, M. G. Signorini, and G. Magenes, "A deep learning mixed-data type approach for the classification of FHR signals," Front. Bioeng. Biotechnol., vol. 10, 2022, Art. no. 887549.
[22] J. L. Aeberhard et al., "Artificial intelligence and machine learning in cardiotocography: A scoping review," Eur. J. Obstet. Gynecol. Reprod. Biol., vol. 281, pp. 54–62, 2023.
[23] Z. Liu et al., "KAN: Kolmogorov-Arnold networks," 2024, arXiv:2404.19756.
[24] S. Gui, J. Wang, C. Tang, and J. Zhao, "High precision frequency source anomaly monitor using Kolmogorov-Arnold networks," in Proc. 22nd Int. Conf. Opt. Commun. Netw., 2024, pp. 1–2.
[25] L. Pang, J. Li, J. Gan, and J. Zhang, "Dissolved gas analysis of transformers based on Kolmogorov-Arnold network," in Proc. 6th Int. Conf. Energy Syst. Elect. Power, 2024, pp. 1568–1571.
[26] M. H. Sulaiman, Z. Mustaffa, M. S. Saealal, M. M. Saari, and A. Z. Ahmad, "Utilizing the Kolmogorov-Arnold networks for chiller energy consumption prediction in commercial building," J. Building Eng., vol. 96, 2024, Art. no. 110475.
[27] M. Poluektov and A. Polar, "Construction of the Kolmogorov Arnold representation using the Newton-Kaczmarz method," 2023, arXiv:2305.08194.
[28] D. Campos and J. Bernardes, "Cardiotocography," UCI Mach. Learn. Repository, 2000, doi: 10.24432/C51S4N.
[29] D. Ayres-de Campos, J. Bernardes, A. Garrido, J. Marques-de Sa, and L. Pereira-Leite, "SisPorto 2.0: A program for automated analysis of cardiotocograms," J. Maternal-Fetal Med., vol. 9, no. 5, pp. 311–318,2000.
[30] S. P. v. Steinburg et al., "What is the 'normal' fetal heart rate?," PeerJ, vol. 1, 2013, Art. no. e82.
[31] T. Prior and C. Lees, "Control and monitoring of fetal growth," in Encyclopedia of ndocrine Diseases, 2nd ed., I. Huhtaniemi and L. Martini,Eds., Oxford, U.K.: Academic Press, 2019.
[32] J.Warland and P. Glover, "Fetal movements:What are we telling women?,"Women Birth, vol. 30, no. 1, pp. 23–28, 2017.
[33] H. Rosen and Y. Yogev, "Assessment of uterine contractions in labor and delivery," Amer. J. Obstet. Gynecol., vol. 228, no. 5, pp. S1209–S1221, 2023.
[34] I. Ben M'Barek, G. Jauvion, J. Vitrou, E. Holmström, M. Koskas, and P.-F. Ceccaldi, "DeepCTG 1.0: An interpretable model to detect fetal hypoxia from cardiotocography data during labor and delivery," Front. Pediatrics,vol. 11, 2023, Art. no. 1190441.