

Fetal Health Classification based on CTG using Machine Learning

¹Shirin M. Raj, ²M. D. Ingale

¹ Department of Computer Engineering,

¹Jayawantrao Sawant College of Engineering, Pune, India

Abstract - The classification of fetal health plays a pivotal role in prenatal care, with the aim of evaluating and monitoring the well-being of the developing fetus during pregnancy. The accurate and timely classification of fetal health status is essential for identifying potential risks and ensuring optimal maternal and fetal outcomes. This study presents an innovative approach that leverages a comprehensive dataset consisting of ultrasound images and clinical parameters from pregnant women to develop a machine learning model for fetal health classification. The methodology involves the assembly of a substantial dataset that includes ultrasound images and relevant clinical characteristics. This dataset is used to train and evaluate the machine learning model. Machine learning algorithms and image processing techniques are applied to categorize fetal health status, with the goal of achieving a high level of accuracy. The proposed model has the potential to reshape the way we assess and monitor fetal health, leading to more accurate diagnoses and improved outcomes for expectant mothers and their unborn children. This research represents a significant advancement in the field of prenatal care and highlights the synergy between sophisticated machine learning techniques and clinical obstetrics.

Keywords: Machine learning, Genetic programming algorithms, Artificial Neural Network, CNN.

1. INTRODUCTION

There were around 213 million births worldwide in 2025 [1]. In developing nations, 23 million women reported being pregnant, whereas, in poor countries, 190 million women reported being pregnant. In 2019, maternal hemorrhage, abortion complications, high blood pressure, maternal infection, and obstructed labor were directly responsible for the deaths of 293,336 women worldwide [2]. About 830 women die per day from complications connected to pregnancy or childbirth, as reported by the World Health Organization (WHO) [3]. This amounts to nearly 303,000 deaths among pregnant and postpartum women in 2015. Mothers and their unborn children are at risk for serious health complications and even death due to pregnancy in today's modern environment. Indeed, nearly 99 percent of maternal mortality occurs in economically developing nations [3]. The complications of pregnancy and childbirth are the leading cause of death in third-world countries for this reason [2], [3]. Many of these issues manifest themselves during pregnancy, but others are displayed before pregnancy and aggravated during conception. However almost all of these maternal deaths occurred in settings with inadequate access to healthcare, and nearly all of them were preventable or treatable.

Despite a global rise in skilled attendance at births (from 58% to 81% in 1990-2019), maternal health progress remains sluggish. While deaths from pregnancy complications have declined 38% in two decades, the average annual drop of 3% is too slow to reach Sustainable Development Goal (SDG) targets. Unequal access within and across countries hinders progress. Half of maternal deaths occur in fragile settings, and Sub-Saharan Africa and Southern Asia bear the brunt, with 86% of the global total in 2025. [https://www.who.int/health-topics/maternal-health#tab=tab_2]

Complications during pregnancy include high blood pressure, diabetes, infections, preeclampsia, miscarriages, premature labor, and stillbirths. Extreme sickness, vomiting, and anemia from a lack of iron are also possible [4], [5]. Additionally, numerous pregnancies, fetal illness, and intrauterine growth

restriction pose risks to the fetus [6], [7]. Therefore, these abnormalities can create developmental neuron issues throughout infancy, resulting in morbidity or even death in the baby. Cerebral palsy without ambulation, developmental delay, hearing and vision loss, and fetal compromise are a few of these problems.

Cardiotocograms simultaneously gather information from many monitoring methods, including fetal movements in the womb, mother uterine contraction pressure, and fetal heart signals [8], [9], which is essential for assessing the fetus's health. The future potential hazards to the fetus can be averted by studying CTG data. Simple and inexpensive, the clinical CTG test provides insight into the developing baby's health. Fetal well-being is often monitored with an antepartum CTG test beginning around the 28th week of pregnancy (the seventh month) [3]. This test's results can help obstetricians formulate treatment plans in the event of fetal growth abnormalities. In reality, the CTG test assesses the fetus's health by checking whether its tissues are receiving enough oxygen or detecting signs of Hypoxia or Acidosis.

An example of a digitally recorded CTG is shown in Figure 1. The most significant benefit of CTG is its role in the early diagnosis of complications that can arise from a shortage of oxygen, such as cerebral palsy and intrapartum fetal hypoxia. In addition, CTG use has been associated with an uptick in the use of Cesarean sections and instrumental deliveries, although the prevalence of cerebral palsy has remained steady

The CTG is widely used by obstetricians to monitor the fetus's health before, during, and after birth. Automated prediction in various medical applications based on early detection findings has become possible because of the widespread deployment of powerful ML and artificial intelligence approaches in recent years.

2. MATERIALS AND METHODS

2.1. Dataset

The dataset considered in this study contains 2126 fetal CTG records that were automatically processed, and the respective diagnostic features measured. The CTGs were also classified by

three expert obstetricians, and a consensus classification label was assigned to each of them. The classification scheme used three fetal states (N: normal; S: suspect; P: pathological). The dataset was therefore used in three-class experiments in this study.

2.2. Exploratory Dataset

Multiclass FHR classification was performed using the UCI CTG dataset, which includes FHR and UC data. This dataset consists of 2126 pieces of datum with 22 columns. A screenshot of the sample values from the dataset is shown in the form of a data frame, extracted with the Pandas library in the scientific Python environment.

```
In [11]: df.sample(10)
Out[11]:
```

	baseline_value	accelerations	...	histogram_tendency	fetal_health
813	146.0	0.004	...	1.0	0.0
141	154.0	0.003	...	1.0	0.0
2016	125.0	0.000	...	0.0	0.0
1992	135.0	0.008	...	0.0	0.0
1810	143.0	0.000	...	0.0	0.0
237	121.0	0.005	...	0.0	0.0
28	132.0	0.000	...	0.0	1.0
1021	130.0	0.000	...	0.0	0.0

Figure: CTG dataset in the form of a Python Pandas data frame

2.3. Proposed Method

The fetal health outcomes (0: N, 1: S, 2: P) in the dataset were used as the target of the model, and the other data were used as input. A diagram of the overall structure of the pipeline of training and testing of the proposed model is shown in **Figure**

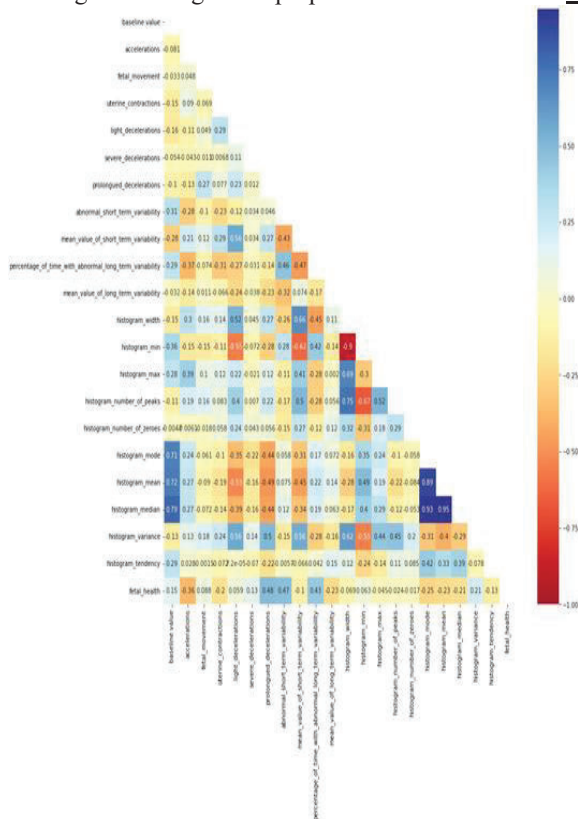


Figure. Workflow of the fetal health classification model using ensemble model

2.4. Data Processing

In the preprocessing step, the numerical input values in the dataset were normalized to the range from -1 to 1 in order to express the correlations in the model's training. Feature extraction was performed, and adaptive class weighting was used to handle the unbalanced dataset. Equation (1) shows the normalization and standardization process for feature extraction, where denotes the new value obtained from the x values.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

$$(x + y)^n = \sum \binom{n}{k} x^k y^{n-k}$$

$$e = \sum_i p_i \log(p_i)$$

2.5. Ensemble Learning (EL)

In ML, ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone [20]. EL algorithms use two different approaches called bagging and boosting. These techniques help in reducing the variance (bagging—RF) and bias (boosting—XGB) and can improve predictions. Parallel learning takes place in the bagging approach, and sequential learning in the boosting approach. After the learning phase, the errors and weights are updated for the next bootstrap. The stacking process combines predictions from two or more classifiers. RF can use another learning approach that optimally combines stacking learning approaches when voting for the final prediction. Another advantage of EL approaches is that the training and testing times are shorter than those of artificial neural networks

$$y'_i = b + \sum_{i=1}^n \omega_i x_i$$

For this purpose, an error analysis is performed based on the difference between the estimation and the actual value, as given in Equation (5), for the mean square error (MSE). The calculated error value should converge to zero. In this direction, the network weights are updated in the following iteration.

$$MSE = \frac{1}{N} \sum_{i=1}^n (y_i - y'_i)^2$$

2.6. Model Evaluation

The evaluation process is carried out using the confusion matrix in figure and is based on a comparison of true classes with estimations made based on the data that are not used in the training process. The confusion matrix allows for not only the accuracy (Acc) but also the F1-score, precision, recall, and Type-1 (FN) and Type-2 (FP) error conditions to be interpreted [24]. In figure true classifications are shown in green, and incorrect classifications are shown in red.

Binary classification		Actuals	
		Class 1	Class 2
Predictions	Class 1	TP	FN
	Class 2	FP	TN

Multiclass classification		Actuals				
		Class 1	2	3	...	N
Predictions	1	True				
	2		True			
	3			True		
	...				True	
	N					True

Figure: Evaluation matrix for binary and multiclass classification.

2.7. Weighted Majority Voting

Soft voting and hard voting are two principal distinct methods for combining the predictions of multiple base classifiers in an ensemble model as shown in . In addition to these, ensemble models can also compute a weighted majority vote by associating a weight w_i with classifier M_i , as shown in Equation (15), which was used in this study.

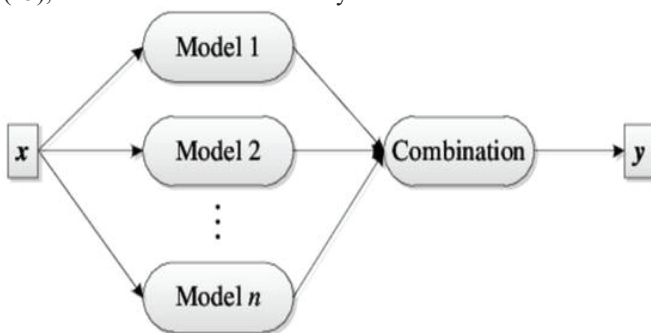
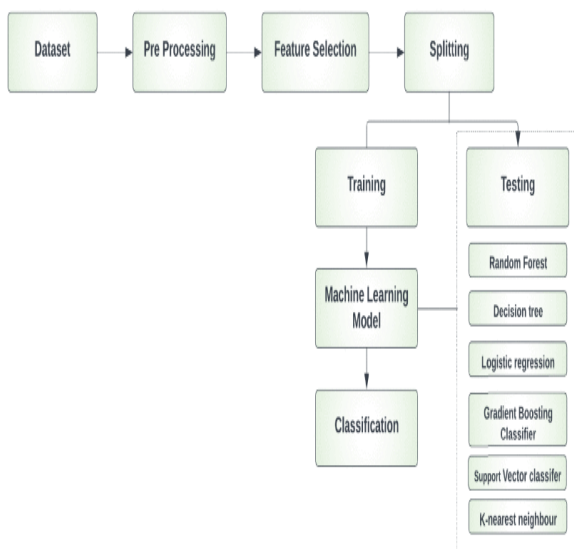


Figure. Combining of ensemble models

2.2 Design



The criteria for classifying CTGs into these three categories were based on the guidelines of the National Institute

of Child Health and Human Development (NICHD). The NICHD guidelines are used by obstetricians and midwives to assess fetal health during pregnancy and labor.

A) Selection Criteria and Demographic Information of the CTG Database

These parameters were used as selection criteria for this database:

- Singleton pregnancy.
- Age at pregnancy
- Each segment lasted 30 minutes, and contained 30 minutes of FHR data.
- Available information on the pH biochemical parameter obtained from the umbilical arterial blood sample.

B. Identifying Fetal Distress in Labor

It is impossible to evaluate the fetus's brain function during labor. Although, the heart of a fetus can be assessed for its unique properties. The fact that shifts in fetal heart rate (FHR) cause brain damage is also crucial. In this way, fetal heart rate (FHR) can serve as a surrogate signal for fetal acid-base status, oxygenation, and blood volume, and a prompt response to aberrant fetal heart rhythms may help prevent brain injury. To prevent perinatal/neonatal morbidity or mortality, fetal monitoring is performed to identify situations in which the health of the fetus may be compromised and to provide prompt, appropriate action.

Before electronic fetal monitoring (EFM) was invented in the late 1960s, intermittent auscultation (IA) was the standard assessment method. For healthy women without risk factors for an unfavorable perinatal outcome, IA is the recommended method of fetal surveillance during labor. A Pinard stethoscope or portable Doppler equipment is used for FHR determination by the below methods:

- Using a pinard stethoscope to monitor a baby's heart rate
- Utilizing a stethoscope to monitor the infant's heart rate
- Using a handheld Doppler to monitor a baby's heart rate

C. Block Diagram of the System

Figure 11 depicts the architectural layout of the ML system. Here are some additional details about each step:

- Pre-Processing: The pre-processing step is important because it can help to improve the accuracy of the ML model. By removing errors and outliers from the data, and by transforming the data into a common scale, the ML model can learn more effectively. Homogenization, also known as data standardization or normalization, is a crucial step in the preprocessing of datasets, including the Cardiotocography (CTG) dataset used in healthcare for fetal monitoring.
- Feature selection: The feature selection step is important because it can help to improve the performance of the ML model. By selecting the most important features, the ML model can be trained more efficiently and can make more accurate predictions.
- Splitting: The splitting step is important because it allows the ML model to be evaluated on data that it has

not seen before. This helps to ensure that the model is not overfitting the training data.

A thorough analysis of numerous recent studies on the classification of fetal health is conducted. This [14] study compared the performance of six different ML models for fetal health classification using CTG data: support vector machines (SVMs), random forests (RFs), decision trees (DTs), logistic regression, k-nearest neighbors, and voting classifier. The [15] study found that RFs had the best performance, with an accuracy of 97.5%. This study used a variety of ML models to classify fetal health status from CTG data, including RFs, DTs, MLPs, and SVMs. The study found that RFs had the best performance, with an accuracy of 96.2%.

While the ensemble learning approach achieved an impressive accuracy of 97.3%, it's important to consider the performance of individual models in the context of this specific study [16]. While the CNN achieved an accuracy of 94.5% [17], other traditional machine learning models like Random Forests (RF) reported an accuracy of 96.2% [15]. This suggests that while the ensemble approach offers a slight advantage in this instance, the performance of individual traditional algorithms remains competitive.

Furthermore, focusing solely on accuracy might not provide a complete picture of model performance. Analyzing other metrics like precision, recall, F1-score, or AUC, and considering the strengths and weaknesses of each model in the context of specific clinical scenarios, could offer a more comprehensive understanding of their true potential.

Ultimately, this study highlights the promise of both ensemble learning and traditional machine learning algorithms for CTG analysis. Further research with larger datasets and diverse algorithms is necessary to draw definitive conclusions about which approach consistently outperforms others in various clinical settings.

Medical experts can use machine learning techniques to make early decisions during complex situations like diagnosis, reducing the risk of maternal mortality and high labor complications. Although ML classification systems have difficulty classifying fetal health stages [18], they can handle them. SVM, RF, and neural networks (NN) are a few of the traditional techniques for classifying data [19] and different techniques are explained in this section.

A. Mechanisms of Fetal Control During Labor

The mother and the developing baby are under a great deal of stress throughout labor and delivery, with the latter being particularly sensitive to the mother's actions and the state of the maternal intrauterine environment. The fetus's reaction to various stimuli may reveal important details about its health. Normal fetuses can endure brief periods of oxygen deprivation throughout this procedure. However, fetus with compromised immunity may suffer from hypoxemia,

hypoxia, or asphyxia, resulting in potentially fatal outcomes such as cell dysfunction, organ failure, developmental delay, disability, or even death [20].

- Hypoxemia corresponds to the earliest stage of oxygen shortage in arterial blood,
- When oxygen depletes peripheral tissues, hypoxia occurs as a second step.
- Asphyxiation is the most crucial stage because major fetal organs like the heart, lungs, liver, gut, and kidneys rely on anaerobic metabolism when oxygen levels drop. An accurate fetal examination and diagnosis during labor are crucial because they provide insight into the health of the fetus, allowing for the avoidance of the complications mentioned above.

B. Monitoring of Fetal Development

The literature [21] suggests that there may be a crucial relationship between a fetus's Fetal heart rate (FHR) and its state as it changes over time. The ANS controls heart rate dynamics by controlling sympathetic and parasympathetic impulses to the heart as seen in Figure 2. This is supported by evidence from adult medical studies [22], [33] The two systems affect heart activity in opposite ways, as seen in Figure 3. In response to danger, the sympathetic nervous system revs up the body for maximum output. While in rest, the parasympathetic nervous system regulates the heart's reaction and helps the body relax.

C. Fetal Distress During Labor: Causes and Symptoms

Although most babies diagnosed with fetal distress are born healthy, prenatal pain has been linked to an increased risk of complications such as cerebral palsy, mental retardation, hypoxia, ischemic encephalopathy, and seizures. Pregnancy complications in two ways [37]:

1) Antepartum

1. Maternal hypotension (epidural anesthesia, supine position)
2. Post maturity
3. Placental insufficiency (pre-eclampsia, IUGR, etc.)
4. Abruptio placenta
5. Chorioamnionitis

2) Intrapartum

1. Hypertonic contractions
2. Scar dehiscence
3. Cord around the neck
4. Cord compression in oligohydramnios
5. Cord prolapses
6. Abnormal uterine contractions

The fetus usually does not react abnormally to minor hypoxia because it can adjust. However, fetal distress will occur in the event of severe fetal hypoxia. During childbirth, the primary goal of fetal monitoring is identifying which fetuses are at risk of hypoxia Clinically applicable signs of fetal distress include:

- Fetal heart rate (FHR) abnormalities
- Meconium stained liquor (MSL) and
- Cord prolapses

D. Interpretation of CTG Traces & Operation of CTG

CTG is the standard method for monitoring fetal heart rate

(FHR) and uterine contractions (UC) during labor. In addition, an external or internal signal recording can be performed during labor, depending on the process stage and the parameters of the procedure [26]. To perform the external CTG, the mother must have a belt placed over her stomach seen in Figure 1. The fetal heartbeat can be extracted from the ultrasound data by having a computer program determine the elapsed time between the cardiac cycle's two loudest peaks and display the result as a percentage.

Continuous CTG monitoring should be initiated if anomalies are observed on intermittent auscultation, and decisions should be based on CTG results. Figure 6 illustrates how the CTG trace is evaluated based on four variables:

- The patient's resting heart rate (FHR): The baseline FHR is the average heart rate of the fetus over a while (usually 10 minutes). A normal baseline FHR is between 110 and 160 beats per minute (bpm).
- Baseline Fetal Heart Rate (Baseline): Baseline variability is the variation in the FHR from beat to beat. A normal baseline variability is between 5 and 25 bpm.
- Declaration: Deceleration is a sudden decrease in FHR of at least 15 bpm below the baseline that lasts for at least 15 seconds. Decelerations can be early, late, or variable.
- Acceleration: An acceleration is a sudden increase in FHR of at least 15 bpm above the baseline that lasts for at least 15 seconds. Accelerations are usually a sign of a healthy fetus.

E. Monitoring of Fetal Development

Given the wide range of circumstances under which Indian women give birth, no standardized protocols for fetal monitoring during labor have been developed. Moreover, there may be various reasons, from mothers giving birth alone and unsupervised in their own homes to inadequate emergency obstetric care facilities. So let's think about how to implement fetal surveillance effectively across India's three-tiered system of hospitals for giving birth represented in the below steps :

- Tier I: Primary health care centers (PHCs)/ small nursing homes (no CTG machine, only IA available, Cesarean delivery not possible)
- Tier II: District hospitals/Private nursing homes (both IA and CTG available, no facilities for FBS, Cesarean delivery possible)
- Tier III: Tertiary care institutes/ Corporate hospitals and research centers (all facilities for fetal surveillance and delivery available)

The findings of this study suggest that the parameters used in the study are a valuable tool for assessing fetal health and identifying fetuses that are at risk for developing problems. The study from Table 2 also suggests that the FHR signal can be used to assess the development of the ANS in the fetus.

2.3 Identification and Selection of Studies

By following these steps, it is possible to improve the accuracy of the ML model and make better predictions about fetal health

status. The system utilizes the complete CTG data set, encompassing all attributes and their corresponding values. Initially, we scrutinized the data set for categorical values, identifying only one such value. Consequently, we delved into the relationships among fetal state characteristics using the functionality of ML models and subsequently visualized our findings.

The model was furnished with the target value and requisite parameters for predictive analysis. Subsequently, we partitioned the data set into distinct training and testing subsets. Despite utilizing random sampling to establish the division, an inherent imbalance persisted between the training and testing sets. The training subset comprised 77% of the data, while the testing subset encompassed 33%, ultimately resulting in a stratified sampling approach.

In light of this, the features underwent standardization for scaling purposes. It is calculated using the formula:

$$z = (x - \text{mean}) / \text{std}$$

where x is the original value, mean is the mean of the feature, and std is the standard deviation. The need for ML models in classifying fetal health arises from the complexity of CTG data and the limitations of traditional methods. These models contribute to resource and time savings by providing efficient, objective, and standardized analyses, ultimately improving the overall quality of fetal health classification and potentially saving lives through early detection and intervention

To enable the model to make predictions, we have assigned the features it needs and set the target value. Subsets of the data set were then used for training and evaluation. Although a random sample was used to determine the split, the consequence is an inequitable distribution of participants between the training and testing groups. In this study, we use the CTG data set to evaluate six popular ML algorithms for the recurrent categorizing of fetal states.

- Random Forest
- Decision Tree
- K-Nearest Neighbour
- Logistic Regression
- Support Vector Classifier
- Gradient Boosting Classifier

1) Random Fore

This type of ensemble method aims to improve generalization by combining multiple learning models.

2) K-Neares Neighbors (KNN)

It is a memory-based model in which predictions are made by comparing the current sample to the nearest elements in the training set based on the distance metric provided explained from Eq (2). One of the key advantages of this method is that it is very straightforward, but it is difficult to robustly determine which similarity function is optimal and which meta-parameters should be used.

$$y_i = \text{argmax}_{j \in K} \text{sim}(x_i, x_j)$$

where:

- y_i is the predicted class label for the the data point.
- K is the number of neighbors.
- x_i and x_j are the feature vectors of the the and the data

points.

3) Logistic Regression (L)

is the basis of this kernel. One of the key advantages of this model is its simplicity, scalability, and interpretation in terms of how changes in an input feature influence a linear parameter's log odds to see

$$P(Y=1|X)=\frac{1}{1+e^{-(\beta_0+\beta_1X_1+\beta_2X_2+\dots+\beta_pX_p)}}$$

where:

- $P(Y=1|X)$ is the probability of the fetus being healthy, given the values of the independent variables X_1, X_2, \dots, X_p .
- X_1, X_2, \dots, X_p are the independent variables, such as the fetal heart rate, the number of accelerations per.

2.5 System Architecture

multi-tier intelligent healthcare framework that integrates frontend user interaction, backend API services, Machine Learning prediction modules, and database management systems into a unified platform. The system architecture mainly consists of four major components: Frontend Interface, Backend API Server, Machine Learning Service, and Database Layer. These components communicate with each other using HTTP/REST APIs and JSON-based data exchange mechanisms to ensure efficient and real-time healthcare prediction services.

The frontend layer acts as the user interaction module of the system and is developed using ReactJS, Vite, and TailwindCSS technologies. This layer provides an interactive chatbot-based healthcare interface where users can register, log in, answer PHQ-9 and GAD-7 assessment questionnaires, view previous assessment history, download PDF reports, and access personalized recommendations. The frontend interface is designed to provide a responsive and user-friendly healthcare environment that allows users to communicate with the system easily. After completing the mental health assessment, the frontend sends the collected user responses and behavioral information to the backend API server through HTTP/REST requests in JSON format.

The backend API server is implemented using Node.js and ExpressJS and functions as the central communication and processing layer of the architecture. The backend handles core functionalities such as user authentication and authorization, session management, report generation, administrative monitoring, and communication with the Machine Learning prediction service. Once assessment data is received from the The Machine Learning service is developed using FastAPI and Python-based Deep Learning frameworks including TensorFlow and Keras. This component acts as the intelligent prediction engine of the system. The ML service loads the trained Deep Learning model and performs preprocessing operations on the received assessment data. After preprocessing, the model analyzes PHQ-9 depression indicators, GAD-7 anxiety indicators, and demographic behavioral attributes to predict mental health severity levels.

The system classifies users into four categories: None, Mild, Moderate, and Severe. The database layer uses PostgreSQL for storing system information including user credentials, assessment history, chat sessions, reports, and prediction results. The backend server continuously performs read and write operations on the database to maintain healthcare records and user activity logs. This database management system ensures secure storage, scalability, and efficient retrieval of healthcare data required for system functionality.

The architecture also includes a complete prediction workflow that demonstrates the end-to-end communication process within the system. Initially, the user interacts with the chatbot interface and completes PHQ-9 and GAD-7 assessments. The frontend sends assessment responses to the backend API server, which forwards the data to the Machine Learning service for prediction analysis. The ML model processes the input features and generates mental health severity predictions along with probability scores. The backend then stores the session details and prediction reports in the database before returning the final prediction results and healthcare recommendations to the frontend interface for user display.

Overall, the proposed architecture provides a scalable, modular, and intelligent healthcare framework capable of performing automated mental health assessment, emotional risk prediction, report generation, and healthcare monitoring. The integration of Artificial Intelligence, Machine Learning, chatbot interaction, and real-time API communication makes the system efficient, user-friendly, and suitable for modern digital mental healthcare applications.

3. R ULT

3.1 Summary of Reviewed Results

To determine whether the best ML model is better than the current commercialized system, one would need to consider the specific models, their training datasets, validation studies, and real-world clinical performance. It's also essential to consider factors such as user-friendliness, integration with existing healthcare systems, and regulatory compliance. However, we can analyze the potential advantages of this specific ML approach compared to some limitations of existing systems:

- The study's ML model is better than all existing commercial systems requires further information and rigorous comparisons involving diverse data sets and real-world validation.
- It's more likely that the best approach lies in collaboration and integration between researchers developing advanced ML models and commercial system developers with their market experience and regulatory compliance.
- The performance of both ML models and commercial systems can vary depending on factors like data quality, specific clinical setting, and patient population.
- Continuous research and development are crucial for improving the accuracy and reliability of both research-based and commercial CTG interpretation tools.

- Ultimately, the choice of CTG interpretation system should be based on a careful evaluation of its performance, clinical relevance, ease of use, and integration with existing healthcare practices

It's important to note: The absence of CNNs in this specific study doesn't necessarily indicate their unsuitability for CTG analysis. Further research with larger datasets and optimized architectures might unlock the potential of CNNs in this domain.

- Choosing the right model for a specific task depends on various factors, including data characteristics, resource constraints, and desired outcomes.

Moreover, ML models cannot replace clinical judgment but serve as valuable decision-support tools, empowering healthcare providers to make more informed and timely decisions, ultimately improving fetal health outcomes. It's important to note that while ML holds great promise, its integration into clinical practice should be approached with caution. The interpretability of ML models, ethical considerations, and validation through rigorous clinical studies are crucial aspects to address before widespread adoption.

3.2 Visualising the Selection of Features

The feature selection method is visualized in Figure 13. Understanding the correlation between features is helped by feature selection [27]. The scores are calculated using the Pearson correlation index, which determines the covariance between variables. Scores closer to 1 will indicate a strong positive correlation, i.e. as the value of one feature increases. Conversely, scores close to -1 will indicate a negative correlation (as the value of one feature increases, the value of the other feature decreases). Scores close to 0 indicate no correlation.

3.3 A. Comparison and Evaluation of Models

The learning curve, which represents the model's capacity for learning, is derived from the training data set. On the other hand, a validation data set is used to create the validation learning curve, which shows how well the model generalizes. Figure 14 shows the curve plots of Precision, Recall, and F1-Score for Different Classes and Classifier Models.

B. Visualizing Data

Graphs of the recurrence dispersion of unlimited classes are called histograms. It is a representation of the area between types based on square shapes with bases at the intervals between the borders and regions proportional to the frequency of the two classes. Using this representation, all squares are linked since the bottom fills the spaces between class boundaries. A histogram of the entire data set is shown in Figure 12. The proportions of the data set can be depicted using a histogram.



FIGURE: Histograms of data set.

such as “gad7_q1”, “gad7_q2”, and “gad7_q3” significantly contribute toward prediction performance. These parameters represent nervousness, excessive worrying, and anxiety symptoms, which are important indicators for identifying moderate and severe mental health conditions. Clinical demographic features including prior mental health history, sleep issues, age, and social withdrawal also contributed toward prediction performance. Among these, prior mental health history showed the highest importance among demographic variables, indicating that users with previous psychological conditions have higher risk probabilities.

Overall, the feature importance analysis validates that combining depression indicators, anxiety parameters, and behavioral attributes improves the efficiency and reliability of the Deep Learning prediction model.

5.Limitation

While machine learning models offer promising potential for fetal health classification using CTG data, there are several limitations to consider: Data-related limitations

- Data quality and heterogeneity: CTG data can be noisy, incomplete, and vary significantly between individuals and pregnancies. This can lead to inaccurate model predictions and difficulty in generalizing results.
- Limited data availability: Large, high-quality datasets with accurate labeling of fetal health outcomes are crucial for training and validating ML models. However, such datasets are often scarce and expensive to collect.
- Bias and confounding factors: Biases in data collection, labeling, and selection can lead to models that unfairly discriminate against certain groups or fail to capture important relationships between features and outcomes.

- Multicenter Research: It can be a powerful tool to overcome data-related limitations and significantly advance fetal health analysis. By carefully considering the benefits, challenges, and necessary resources, you can assess whether this approach is suitable for your specific research goals and contribute to improved understanding and outcomes in fetal health.

Model-related limitations

- Overfitting and underfitting: ML models can overfit to the training data, performing well on the training set but poorly on unseen data. Conversely, underfitting can occur if the model is too simple and fails to capture the complexity of the data.
- Interpretability and explainability: Black-box models, while often highly accurate, can be difficult to understand and interpret. This can hinder trust in their predictions and limit their clinical utility.
- Computational resources: Training and deploying complex ML models can require significant computational resources, which may be unavailable in all healthcare settings.

Clinical and ethical limitations:

- False positives and negatives: ML models can misclassify healthy fetuses as distressed or vice versa, leading to unnecessary interventions or missed opportunities for treatment.
- Ethical considerations: Issues like data privacy, bias, and potential misuse of models need careful consideration and ethical guidelines.
- Clinician acceptance and integration: Clinicians may be hesitant to trust and rely on ML models for decision-making, requiring careful integration into existing workflows and training.

Additional limitations

- Limited understanding of fetal physiology: The complex relationship between CTG features and fetal health is not fully understood, which can limit the accuracy and interpretability of models.
- Evolving clinical practices and guidelines: CTG interpretation and management of pregnancy complications are constantly evolving, requiring models to adapt and stay relevant.

CONCLUSION

In conclusion, it is imperative to underscore the theoretical and practical implications of our research on the impact of pregnancy complications, particularly focusing on the potential repercussions for both maternal and fetal health. Our investigation has brought to light crucial insights that can significantly contribute to the field, emphasizing the need for early identification and intervention to mitigate the severity of risks associated with fetal anomalies.

Research contributions emanate from our revelation that among women undergoing Cesarean sections due to worrisome cardiotocography (CTG) readings, a substantial 19.5% exhibited neonatal acidemia a clear indicator of fetal distress. Notably, acidemia emerged as a robust predictive factor in both scenarios, highlighting the importance of educational background in understanding and interpreting CTG data effectively [35].

In practical terms, our findings emphasize the potential advantages of judicious decision-making based on specific fetal heart rate patterns intricately linked with acidosis. By circumventing delayed and unnecessary interventions, our approach preserves the well-being of newborns, reducing the need for aggressive resuscitation efforts and minimizing prolonged hospital stays. The precision of the models employed in our analysis, surpassing that of previous research endeavors, enhances the reliability of our conclusions.

However, it is essential to acknowledge certain limitations in our research. The current approach to CTG data analysis involves manual scrutiny by obstetricians, introducing potential inaccuracies and hazards

REFERENCES

- [1] I.G. Sedgh, S. Singh, and R. Hussain, "Intended and unintended pregnancies worldwide in 2012 and recent trends," *Stud. Family Planning*, vol. 45, no. 3, pp. 301–314, Sep. 2014
- [2] Z.C. J. Murray, "Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: A systematic analysis for the global burden of disease study 2013," *Lancet*, vol. 385, no. 9963, 1990, Art. no. 117171.
- [3] 3.(2016). World Health Organization, *Maternal Mortality: Fact Sheet*. [Online]. Available:
- [4] 4.National Institutes of Health, *What Are Some Common Complications of Pregnancy?* U.S. Department of Health and Human Services. [Online].
- [5] 5.Office on Women's Health, *Pregnancy Complications* US Department of Health and Human Services. [Online]. Available: <https://www.womenshealth.gov/pregnancy/your-pregnant-now>
- [6] 6.G. Magenes, R. Bellazzi, A. Malovini, and M. G. Signorini, "Comparison of data mining techniques applied to fetal heart rate parameters for the early identification of IUGR fetuses," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Orlando, FL, USA, Aug. 2016, pp. 916–919, doi: 10.1109/EMBC.2016.7590850.
- [7] 7.N. M. Fisk and R. P. Smit, "Fetal growth restriction; small for gestational age," in *Turnbull's Obstetrics*, 3rd ed., G. Chamberlain, and P. Steer, Eds. Edinburgh, Scotland: Churchill Livingstone, 2001, pp. 197–209.
- [8] 8.S. C. R. Nandipati, "Classification and feature selection approaches for cardiotocography by machine learning techniques," *J. Telecommun., Electron. Comput. Eng.*, vol. 12, no. 1, pp. 7–14, 2020.
- [9] 9.S. Das, K. Roy, and C. K. Saha, "Establishment of automated technique of FHR baseline and variability detection using CTG: Statistical comparison with expert's analysis," *Int. J. Inf. Eng. Electron. Bus.*, vol. 11, no. 1, pp. 27–35, Jan. 2019, doi: 10.5815/ijieeb.2019.01.04.
- [10] 10.W. W. Stead, "Clinical implications and challenges of artificial intelligence and deep learning," *J. Amer. Med. Assoc.*, vol. 320, no. 11, p. 1107, Sep. 2018, doi: 10.1001/jama.2018.11029.
- [11] 11.K. H. Miao, J. H. Miao, and G. J. Miao, "Diagnosing coronary heart disease using ensemble machine learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 10, p. 3039, 2016.
- [12] 12.M. G. Signorini, N. Pini, A. Malovini, R. Bellazzi, and G. Magenes, "Integrating machine learning techniques and physiology-based heart rate features for antepartum fetal monitoring," *Comput. Methods Programs Biomed.*, vol. 185, Mar. 2020, Art. no. 105015.
- [13] 13.J. H. Miao, K. H. Miao, and G. J. Miao, "Breast cancer biopsy predictions based on mammographic diagnosis using support vector

- machine learning, *Multidisciplinary Journals in Science and Technology*, J. Sel. Areas Bioinf., vol. 5, no. 4, p. 19, 2015.
- [14] 14.M. T. Alam, M. A. I. Khan, N. N. Dola, T. Tazin, M. M. Khan, A. A. Albraikan, and F. A. Almalki, "Comparative analysis of different efficient machine learning methods for fetal health classification," *Appl. Bionics Biomechanics*, vol. 2022, pp. 1–12, Apr. 2022.
- [15] 15.A. Mehbodniya, A. J. P. Lazar, J. Webber, D. K. Sharma, S. Jayagopalan, K. P. Singh, R. Rajan, S. Pandya, and S. Sengan, "Fetal health classification from cardiocotographic data using machine learning," *Expert Syst.*, vol. 39, no. 6, Jul. 2022, Art. no. e12899.
- [16] 16.A. Kuzu and Y. Santur, "Early diagnosis and classification of fetal health status from a fetal cardiocotography dataset using ensemble learning," *Diagnostics*, vol. 13, no. 15, p. 2471, Jul. 2023.
- [17] 17.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," *Biomed. Eng. OnLine*, vol. 16, no. 1, pp. 1–26, Dec. 2017.
- [18] 18.E. J. Quilligan and R. H. Paul, "Fetal monitoring: Is it worth it?" *Obstetrics Gynecol.*, vol. 45, no. 1, pp. 96–100, 1975.
- [19] 19.F. G. Esposito, "Fetal heart rate monitoring and neonatal outcome in a population of earlyand lateonset intrauterine growth restriction," *J. Obstetrics Gynaecology Res.*, vol. 45, no. 7, pp. 1343–1351, 2019.
- [20] 20.A. Ugwumadu, "Are we (MIS) guided by current guidelines on intrapartum fetal heart rate monitoring? Case for a more physiological approach to interpretation," *BJOG, Int. J. Obstetrics Gynaecol.*, vol. 121, no. 9, pp. 1063–1070, 2014.
- [21] 21.C. P. Anaruma, M. Ferreira, C. H. G. Sponton, M. A. Delbin, and A. Zanesco, "Heart rate variability and plasma biomarkers in patients with type 1 diabetes mellitus: Effect of a bout of aerobic exercise," *Diabetes Res. Clin. Pract.*, vol. 111, pp. 19–27, Jan. 2016.
- [22] 22.F. Andreotti, J. Behar, S. Zaunseeder, J. Oster, and G. D. Clifford, "An open-source framework for stress-testing non-invasive foetal ECG extraction algorithms," *Physiological Meas.*, vol. 37, no. 5, pp. 627–648, May 2016.
- [23] 23.G. D. Clifford, I. Silva, J. Behar, and G. B. Moody, "Non-invasive fetal ECG analysis," *Physiological Meas.*, vol. 35, no. 8, pp. 1521–1536, Aug. 2014.
- [24] 24.T. P. Richardson, M. C. Peters, A. B. Ennett, and D. J. Mooney, "Polymeric system for dual growth factor delivery," *Nature Biotechnol.*, vol. 19, no. 11, pp. 1029–1034, Nov. 2001.
- [25] 25.D. Maulik, *Doppler Ultrasound in Obstetrics and Gynecology*, I. Zalud, Ed. Berlin, Germany: Springer, 2005, pp. 363–374.
- [26] 26.D. Goswami, S. Bhasin, and S. Batra, "Obstetric & gynecological emergencies," in *Wife Goes*, 2011.
- [27] 27.P. F. F. Ortiz, "Automatic fetal distress assessment during labor based on modal and parametrical analysis of the cardiocotographic recording," *Otto-von-Guericke-Universität, Magdeburg, Germany, Tech. Rep.*, 2020.
- [28] 28.Z. Cömert and A. Fatih, "Evaluation of fetal distress diagnosis during delivery stages based on linear and nonlinear features of fetal heart rate for neural network community," *Int. J. Comput. Appl.*, vol. 156, no. 4, pp. 26–31, Dec. 2016.
- [29] 29.V. Paladugu, S. Sreedhar, R. Chitra, S. T. Mannava, S. Sreekumar, and J. Mangalakanthi, "Association of CTG diagnosis of intrapartum fetal distress and immediate postpartum acidemia in foetal umbilical artery," *J. Obstetrics Gynecol. India*, vol. 73, no. 1, pp. 28–35, Feb. 2023.
- [30] 30.N. Krupa, M. Ma, E. Zahedi, S. Ahmed, and F. M. Hassan, "Antepartum fetal heart rate feature extraction and classification using empirical mode decomposition and support vector machine," *Biomed. Eng. OnLine*, vol. 10, no. 1, p. 6, 2011.