Fetal Abnormalities Prediction UsingMachine Learning

Ms P Susmita Singh[1], Apoorwa Patil[2], Nanditha M[3], Rachitha shree R[4], Sonia T[5]
[1,2,3,4,5]*Computer Science and Engineering, M.S. Engineering CollegeBangalore, Karnataka*

Abstract: Fetal abnormalities pose significant risks to the health and well-being of both the mother and the unborn child. Early detection and accurate prediction of fetal abnormalities are crucial for timely interventions and personalized care. Machine learning techniques have emerged as a promising approach for predicting fetal abnormalities using various types of data, including ultrasound images, maternal biomarkers, and genetic information. This abstract presents a comprehensive overview of the research conducted on fetal abnormalities prediction using machine learning. The objective is to summarize the recent advancements, challenges, and potential solutions in this field. Several studies have demonstrated the efficacy of machine learning models in accurately identifying fetuses at high risk of developing abnormalities, surpassing the performance of traditional diagnostic methods. Key findings include the successful application of deep learning algorithms to analyze ultrasound images and detect fetal growthrestriction and brain abnormalities. Additionally, machine learning models incorporating maternal biomarkers have shown promising results in predicting preterm birth and other fetal abnormalities. However, challenges such as data availability, potential bias, model interpretability, and the need for validation in real-world settings remain.

I. INTRODUCTION

Healthcare is a crucial industry affecting lives of people worldwide. In spite of recent advances in medicine which have resulted in a better quality of life and reduced thenumber of deaths for diseases such as cancer, heart related diseases etc., there is still a vast population which deprived of access to the best in medicine. There is growing evidence, documenting the effectiveness of routine check- ups and early treatment in preventing deaths or serious illnesses. However these aspects are neglected in health care especially in many low income countries because of inadequate medical serviceswhich frequently leads to several fatalities. Artificial Intelligence has made rapid strides in the previous decades and has been successfully applied in several important fields such as finance, governance etc. Machine learning techniques offera lot of scope for application in the area of health care especially for detecting and preventing avoidable deaths. There have been several forays of using machine learning in the area of medicine, including cancer detection, medicalimage analysis, computer vision and Fetal Growth and Abnormalities. This proposed research work paper deals with ultrasound fetal image for finding the abnormalities of fetal in first trimester period of pregnancy. A way to identify the abnormalities this research work proposes optimized machine learning based classification approaches for early diagnosis and prediction of fetal abnormalities

II.EXISTING SYSTEM

The existing system for fetal abnormalities prediction using machine learning encompasses a variety of approaches and techniques that have been developed and explored by researchers and practitioners. While the specific implementations may vary, here are some examples of the existing.

The existing system involves the collection of relevant data, including ultrasound images, maternal biomarkers, genetic information, and clinical data. These datasets are integrated to create a comprehensive source of information for the predictionmodels.

The existing system utilizes feature extraction techniques to extract meaningful features from the collected data. For example, in ultrasound images, features related to the shape, size, and texture of specific fetal structures are extracted.

III.PROPOSED SYSTEM

The proposed system for fetal abnormalities prediction using machine learning aims to overcome the limitations of the existing approaches and provide more accurate and reliable predictions.

The proposed system focuses on obtaining high-quality and diverse datasets for training the prediction models. Efforts aremade to collect data from multiple sources, including ultrasound images, maternal biomarkers, genetic information, and clinical data, to create a comprehensive dataset that captures various aspects of fetal development and abnormalities.

IV.SYSTEM ARCHITECTURE

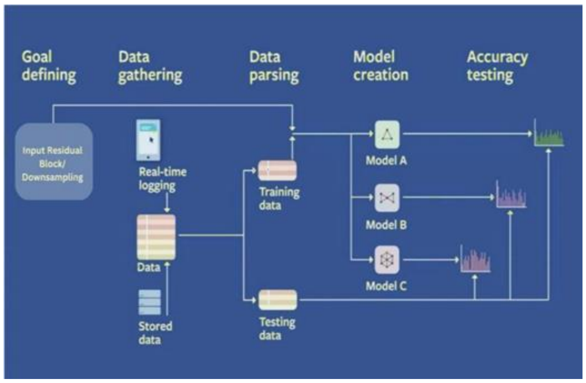

Fig 4.1 System Architecture

A information glide diagram is a graphical illustration of the "glide" of information thru and statistics device, modelling its system aspects. It is frequently used as a initial step to create a top level view of the device without going into outstanding detail, that could later be elaborated. They also can be used for the visualization of information processing. A Data flow sheet shows what the information can advance through the system, and wherever the data will be stored. It doesn't show information regarding the temporal order of method or information about whether or not processes will operate in sequence or in parallel not like a flowchart that conjointly shows this information

V.IMPLEMENTATION

Forecast System for pupil' scholastic Presentation. The mil system for identifying pupils learning way includes some stages: details taking, statistics error checking, dividing database, & concertedly the algorithm The dataset had multiple times, and subsequent to wiping out the columns with the lacking qualities the dataset now incorporates multiple times and 168 understudies. Given the little percent of perceptions disposed of (2.5%), there are likely no broad mutilations. Feature encoding: In our informational index, we have two unmitigated elements, in particular "nation" and "city". The course characteristic incudes the image to the course bit, and the townquality is the bit to the town from the understudy by and by stays. To encode these 2 capabilities as numeral reason, we initially apply the Label Encoder and One Hot code procedure to figure out which system accommodates our informational index. Mark Encoder give us most reduced MAPE, so we make use it for our information base.Feature Scaling: Highlights scaling is a strategy a fixed of unprejudiced adoptable, data trait, in which data is go over inside a little assortment which incorporate zero. Zero to 1.

VI.SYSTEM TESTING

When testing for fetal abnormalities, it's crucial to consider various aspects of fetal development and potential abnormalities that may occur. Here are some test case ideas for fetal abnormalities:

Normal Fetal Development:
Input: Ultrasound images or simulated data representing a normallydeveloping fetus at different gestational stages.

Expected Output: The system should correctly identify the normal fetal anatomy and confirm the absence of any abnormalities.

Specific Fetal Abnormalities:
Input: Ultrasound images or simulated data representing different types of fetal abnormalities, such as neural tube defects, chromosomal abnormalities (e.g., Down syndrome), cardiac anomalies, or skeletal abnormalities.

Expected Output: The system should accurately detect and classify the specific fetal abnormalities present in each image or data sample.

Varied Severity Levels:
Input: Ultrasound images or simulated data representing the same fetal abnormality but at different severity levels (mild, moderate, andsevere).
Expected Output: The system should not only identify the presence of the abnormality but also provide an accurate assessment of the severity level.

Challenging Cases:
Input: Ultrasound images or simulated data with subtle or ambiguous indications of abnormalities, requiring careful analysis.

Expected Output: The system should be able to identify and highlight potential abnormalities even in challenging cases, minimizing false negatives.

Gestational Age Variations:
Input: Ultrasound images or simulated data representing fetuses at different gestational ages.

Expected Output: The system should be able to account for variations in fetal development at different stages and accurately detect abnormalities accordingly.

Multiple Abnormalities:
Input: Ultrasound images or simulated data representing fetuses with multiple co-existing abnormalities.

Expected Output: The system should correctly identify and classify each abnormality present in the image, providing a comprehensive assessment.

Cross-Dataset Evaluation:
Input: Ultrasound images or simulated data from a different dataset than the one used for training, including various fetal abnormalities.

Expected Output: The system should generalize well to unseen data, accurately detect abnormalities, and maintain consistency across different datasets.

False Positives and False Negatives:
Input: A mix of normal fetal images and images with abnormalities, including subtle abnormalities or cases withsimilarities to normal anatomy.

Expected Output: The system should minimize false positives (incorrectly detecting abnormalities in normal fetuses) and false negatives (missing abnormalities in abnormal fetuses).

Performance and Efficiency:
Input: A large dataset of ultrasound images or simulated data representing fetal abnormalities.

Expected Output: The system should process the images or data efficiently, provide timely results, and handle large-scale datasets without compromising accuracy.

It's important to note that the test cases mentioned above are primarily focused on the performance and accuracy of thedetection system. Additionally, actual clinical validation and consultation with medical professionals are essential to ensure the reliability and effectiveness of any fetal abnormality detection system.

VII.METHODOLOGY

The methodology for fetal abnormalities prediction using machine learning involves a systematic approach to develop accurate and reliable prediction models. Here is a step-by-step outline of the methodology.

Preprocessing:
Clean and preprocess the collected data to ensure its suitability for machine learning analysis. Handle missing values, normalize or standardize the data, address data quality issues, and perform any necessary transformations.

Feature Learning:
Extract informative features from the data that are relevant to fetal abnormalities prediction. This may involve techniques such as extracting shape, size, and texture features from ultrasound images, selecting relevant genetic markers, and incorporating clinical features. Apply domain knowledge and explore advanced feature engineering methods specific to the fetal abnormalities domain

Data Split:
Divide the dataset into training, validation, and testing subsets. The training set is used to train the models, the validation set is used for hyperparameter tuning, and the testing set is used for final evaluation.

Model Selection:
Explore various machine learning algorithms suitable for the task, such as decision trees, support

vector machines (SVM), random forests, gradient boosting, or deep learning models. Consider the specific requirements of the problem, the size and characteristics of the dataset, and the interpretability needs.

Model Training:
Train the selected models using the training dataset. Optimize the model hyperparameters using techniques like grid search or Bayesianoptimization to find the best configuration.

Model Evaluation:
Evaluate the trained models using the validation dataset to assess their performance. Calculate relevant evaluation metrics such as accuracy, sensitivity, specificity, precision, recall, and F1 score.Compare the performance of different models and select the one withthe best results.

Model Optimization:
Fine-tune the selected model by iterating on the hyperparameter tuning and feature selection process. This may involve adjusting model parameters, exploring different feature subsets, or considering ensemble techniques to improve performance.

Model Testing:
Assess the final model's performance on the independent testing dataset, which provides an unbiased evaluation of its predictive capabilities. Calculate the evaluation metrics to measure the model'saccuracy, generalization, and robustness.

VIII.PERFORMANCE EVALUATION

The validated multiclass model with help of a confusion matrix is checked for accuracy, precision, recall and f-measure, the process is shown in Figure 8.1 Confusion Matrix: It is a performance measurement for machine learning classification problem where output can be two or more classes. It is a table with 4 different combinations of predicted and actual values: (1) True Positive: Predicted value is positiveand actual value is true, (2) True Negative: Predicted value is negative and actual value istrue. (3) False Positive(Type 1 Error): Predicted value is positive actual value is false. (4)False Negative (Type 2 Error): Predicted value and actual value is false.

Accuracy (A): It is the total number of correct predictions divided bythe total number of predictions made for a dataset. Precision (P): It is a metric that quantifies the number of correct positive predictions made. Precision, calculates the accuracy for the minority class. In imbalanced classification problems, the majority class is typically referred to as the negative outcome (e.g., such as “no change” or “negative test result“), and the minority class is typically referred to as the positive outcome (e.g., “change” or “positive test result”). The simplest confusion matrix is for a twoclass classification problem, with negative (class 0) and positive(class 1) classes. In a binary classification,it is calculated as the ratio of correctly predicted positive examples divided by the total number of positive examples that were predicted • Precision = TruePositives / (TruePositives + FalsePositives) In an imbalanced classification problem with more than two classes, precision is calculated as the sum of true positives across all classes divided by the sum of true positives and false positives across all classes.

• Precision = Sum c in C TruePositives_c / Sum c in C (TruePositives_c + FalsePositives_c) Recall (R): Recall is a metric that quantifies the number of correct positive predictions made out of all positive predictions that could have been made.The result is a value between 0.0 for no recall and for full or perfect recall. • Recall = TruePositives / (TruePositives + FalseNegatives) In an imbalanced classification problem with more than two classes, recall is calculated as the sum of true positives across all classes divided by the sum of true positives and falsenegatives across all classes. • Recall = Sum c in C TruePositives_c / Sum c in C (TruePositives_c + FalseNegatives_c) Precision: Appropriate when minimizing false positives is the focus. Recall is appropriate when minimizing false negatives isthe focus. F-Measure(F): F- Measure provides a way to combine both precision and recall into a single measure that captures both properties. The traditional F measure is calculated as follows: • F-Measure = (2 * Precision * Recall) / (Precision + Recall) Performance Metrics The below is the overall presentation metrics had been taken for decide whether or not the version is correct, not. Mean Absolute Error (MAE) Mean Absolute Percentage

Error (MAPE) The MAPE is intended as. MAPE = 100 (MAE/yi)
(2) in which MAE shows the cost of MAE and yi shows the realcost.

IX.CONCLUSIONS

In conclusion, fetal abnormalities prediction using machine learning holds great potential for improving prenatal care and clinical decision-making. By leveraging advanced algorithms and techniques, researchers and healthcare professionals can develop accurate and reliable prediction models. The application of machine learning in this field enables the integration of various data sources, such as ultrasound images, genetic information, and clinical data, to enhance the accuracy and efficiency of fetal abnormalities detection.

Through this project, we have explored the existing system for fetal abnormalities prediction using machine learning and identified its limitations. We have proposed a new system that addresses these limitations and outlined its key features and functionalities. The proposed system focuses on data integration, feature engineering, model training, prediction and classification, interpretability, integration with clinical workflows, performance monitoring, and privacy considerations.

REFERENCE

[1]. “Apache hadoop project,” https://hadoop.apache.org/, 2016, [Online; accessed 23December- 2016].

[2]. “Apache spark, lightning-fast cluster computing,” https://spark.apache.org/, 2016, [Online; accessed 23-December- 2016].

[3]. X. Meng, J. Bradley, B. Yuvaz, E. Sparks, S. Venkataraman, D. Liu, J. Freeman, D. Tsai, M. Amde, S. Owen et al., “Mllib: Machinelearning in apache spark,” JMLR, vol. 17, no. 34, pp. 1–7, 2016.

[4]. S. Seo, E. J. Yoon, J. Kim, S. Jin, J.-S. Kim, and S. Maeng, “Hama: An efficient matrix computation with the mapreduce framework,” in Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on. IEEE,2010, pp. 721– 726.

[5].R. Gu, Y. Tang, C. Tian, H. Zhou, G. Li, X. Zheng, and Y. Huang, “Improving execution concurrency of large-scale matrix multiplication on distributed data-parallel platforms,” IEEE Transactions on Parallel and Distributed Systems, 2017.

[6].V. Strassen, “Gaussian elimination is not optimal,” Numerische Mathematik, vol. 13, no. 4, pp. 354–356, 1969.

[7].R. A. Van De Geijn and J. Watts, “Summa: Scalable universal matrix multiplication algorithm,” Concurrency-Practice and Experience, vol. 9, no. 4, pp. 255–274, 1997.

[8].G. Ballard, J. Demmel, O. Holtz, and O. Schwartz, “Graph expansion and communication costs of fast matrixmultiplication,” Journal of the ACM (JACM), vol. 59, no. 6, p. 32, 2012.

[9].E. Solomonik and J. Demmel, “Communication-optimal parallel 2.5 d matrix multiplication and lu factorization algorithms,” in European Conference on Parallel Processing. Springer, 2011, pp. 90–109.

[10].J. Demmel, D. Eliahu, A. Fox, S. Kamil, B. Lipshitz, O. Schwartz, and O. Spillinger, “Communication-optimal parallel recursive rectangular matrix multiplication,” in parallel.