International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 5S, 2025
https://www.theaspd.com/ijes.php

Integrating Autoencoders with Recursive Elastic Tree for Enhanced Fetal Health Monitoring: A Feature Engineering and Classification Approach

Ms. Radha Jayalakshmi P[1], Dr. Rajakumar P S[2] and Dr. Rajavarman V N[3]
[1]Research Scholar, Department of Computer Science,
Dr. M.G.R. Educational and Research Institute, Chennai-600 095, India,
radhajayalakshmi.official@gmail.com
[2]Professor, Department of Computer Science &Engineering,
Dr. M.G.R. Educational and Research Institute, Chennai-600 095, India,
rajakumar.subramanian@drmgrdu.ac.in
[3]Professor, Department of Computer Science & Engineering.
Dr. M.G.R. Educational and Research Institute, Chennai-600 095, India, rajavarman.n@drmgrdu.ac.in

ABSTRACT

This paper establishes an assimilated Autoencoder with a Recursive Elastic Tree framework that improves feature engineering and classification for monitoring fetal health during pregnancy. Autoencoders, a deep learning technique, can be utilized to find an effective and non-linear feature representation for high-dimensional fetal health data and detect complex patterns, that would otherwise have gone undetected with traditional methods. After dimensionality reduction, feature transformation, and relevance detection, RET selects the appropriate features for classification that bring about improved accuracy and increased model robustness. Putting together Autoencoders with RET resolves the setbacks of data noise, overfitting, and computational inefficiencies, thus providing a rich tool for early risk assessments and predictions of fetal health problems. The new tactic provides an essential scalable and accurate solution when dealing with large and complex high-dimensional datasets in general healthcare applications.

1. INTRODUCTION

Fetal health monitoring is a crucial part of prenatal care that gives necessary information regarding the early detection of risks and complications affecting both maternal and fetal well-being. However, complex medical datasets and high dimensionality in fetal health data, it pose several challenges for traditional machine learning models. These models flop to capture many complex patterns within the data that ultimately reduce the performance of models and cause wrong predictions.

In this paper, an intermixture framework is proposed that combines Autoencoders with the RET algorithm for feature transformation and selection in fetal health monitoring systems. One of the deep learning approaches Autoencoders which performs effectively in learning non-linear feature representation. This approach is appropriate for intricate datasets such as data like fetal heart rate, as they are capable of automatically hiding hidden patterns and minimizing feature space without missing pivotal information.

After the dimensional reduction using Autoencoders, the RET algorithm is applied to the features. The RET progressively removes the weaker features and sustains the top-ranked ones, thus, enhancing the efficiency of classification models. this amalgamation of these two techniques manages the problems of redundancy of features, overfitting, and less computational efficiency, which is always a problem in high dimensionality.

This hybrid strategy enhances the detection of early health issues in fetuses by providing a more accurate, reliable, and scalable monitoring system. It improves the handling of complex data, thus boosting the predictive accuracy of fetal health assessments. Advanced feature selection and transformation strengthen the data, thus better risk detection is supported. This result helps in more informed clinical decision-making and improved health outcomes for both mother and child.

1.1. MOTIVATION OF THE PAPER

The aspiration for this paper is to explore whether this hybrid machine learning approach, based on autoencoders for feature extraction and recursive elastic trees for classification, could help in better fetal health monitoring. We, therefore, expect to reduce the dimensionality of data, improve predictive accuracy, and make clinical decisions for fetal health more effectively. We attempted to collaborate with the expanding field of AI in healthcare by providing a new solution that could be applied easily in the real world, going toward the proper care for mothers and their babies.

2. LITERATURE REVIEW

Autoencoders in Healthcare: There have been many studies on how autoencoders extract non-linear features from complex healthcare datasets, which include fetal health data. Autoencoders are also widely appreciated for reducing dimensionality and preserving important patterns to improve overall accuracy (Daria Doncevic, Carl Herrmann, 2023).

Recursive Elastic Tree in Feature Selection: RET has been widely applied in classification tasks, particularly when dealing with high-dimensional datasets. (Zhao et al. 2019) proved that RET efficiently selects the most relevant features for predictive models in medical applications.

Berahmand et al. 2024 recommend that future research should continue to explore the robustness and generalization properties of autoencoders, especially in large and complex datasets. Design hybrid models based on combining autoencoders, GANs, reinforcement learning, and other deep-learning methods. Areas of advancement also include scalability and interpretability to have autoencoders process real-world data efficiently and more easily. The approach encourages new applications, especially in generative models and semi-supervised learning. Altogether, the paper presents the need for continued innovation to improve the practical applicability of autoencoders across a vast scope of domains.

Nonlinear methods can also learn complex relationships between the input features and output variables without requiring domain knowledge or prior assumptions about the data and typically lead to better predictive performance (Wang et al. 2022).

Manifold-based feature extraction is a nonlinear method based on the premise that it can embed data into a space where the features will not be missed in reducing their dimension. This approach aims to discover a non-linear transformation that conserves the underlying structure of the data (Li et al. 2022).

Moreover, AEs are used to preserve privacy techniques like differential privacy by protecting sensitive data while still allowing for analysis and insights. Besides these applications, AEs are used to decrease the storage requirements of data, enhance interpretability by providing essential features of data, and show robustness by generalizing well to new data and also handling noisy or incomplete data sets (Liu et al. 2023).

2.1 PROBLEM DEFINITION

Fetal health monitoring in prenatal care is a very essential task that aims to evaluate the potential risks to the fetus, such as distress, growth restrictions, or abnormal heart patterns, which greatly affect maternal and fetal outcomes. Traditional methods, such as cardiotocography (CTG) and ultrasound, provide valuable insights but often struggle with the complexity and high-dimensionality of data, leading to challenges in accurate and timely risk detection. One such approach using machine learning algorithms, including autoencoders for feature extraction and recursive elastic trees for classification, has gained potential in these lines by providing more efficient processing of big data analytics. These methods have yet to be implemented fully in the integration with fetal health monitoring. This paper solves the problem of how an improved predictive model can include autoencoders for reduced dimensions as well as feature extraction along with robust classification by RET to enhance accuracy, scalability, and reliability in fetal health risk detection. These advanced methods will ultimately allow for the development of an effective system that can support earlier detection and better clinical decisions concerning fetal health monitoring.

3. METHODOLOGY

3.1. DATA SET

This research combines and integrates autoencoders and RET models to enhance the performance of fetal health monitoring across three different datasets. Data one is cardiotocography, which provides fetal heart rate along with uterine contractions. Data two relates more to maternal health, using age and blood pressure and their impact on the healthy development of the fetus in the womb. The data three offers demographic birth along with fertility statistics for understanding the bigger picture. These datasets, sourced from both Kaggle and Data.World, together help develop a more accurate and comprehensive predictive model in detecting risk for fetal health.

Dataset 1: This dataset is a delightful collection of features extracted from CTG examinations, categorized into normal, abnormal, or pathological (https://www.kaggle.com/datasets/andrewmvd/fetal-health-classification), **Dataset 2:** Data from several hospitals is obtained using a risk monitoring system (https://www.kaggle.com/datasets/csafrit2/maternal-health-risk-data), **Dataset 3:** Birth and Fertility rates are collected using dataset 3 (https://data.world/cdc/nchs-birth-and-fertility-rates)

3.2 AUTOENCODERS FOR FEATURE EXTRACTION

Autoencoders are unsupervised neural network models used for the reduction of dimensionality and feature extraction. They learn efficient representations of data by encoding input into a lower-dimensional space and reconstructing it back to its original dimension. Encoding and decoding make it possible for autoencoders to find underlying patterns and structures in the data, making them very useful for processing complex, high-dimensional datasets such as those used in fetal health monitoring.

Here, in the experiments, autoencoders are used for extracting relevant features from these three datasets in relation to fetal and maternal health. In particular, autoencoders were utilized for the reduction of data dimensionalities and maintaining the core patterns leading to proper classification and prediction. The reasons why using autoencoders for feature extraction is pertinent in this context include the following:

1. Noise Reduction: Autoencoders are great at filtering the data, which is very important in medical datasets where signals, like fetal heart rates, may be polluted with interference due to maternal movements, inaccuracies from the sensor, or surrounding stimuli.
2. Non-linear Feature Enabling: Classical linear methods fail to capture the subtle interdependencies in complex data, whereas autoencoders enable the extraction of non-linear patterns from the data, which is crucial for modeling such complex dynamics of fetal health indicators, such as fluctuations in fetal heart rate or early transitions in maternal health indicators.
3. Data Compression: Autoencoders compress the data into a lower-dimensional representation, which improves the computational efficiency of subsequent models. This is particularly useful when dealing with large datasets, reducing the time and resources needed for training machine learning models.
4. Capturing Relevant Features: The encoder part of the autoencoder network identifies during the learning process which features of the input data are most relevant to reconstruct the original signals. These extracted features can be used as input for further predictive modeling or classification tasks.

3.2.1 IMPLEMENTATION IN FETAL HEALTH MONITORING

For this research, autoencoders were used for each of the three datasets as follows:

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 5S, 2025
https://www.theaspd.com/ijes.php

- Dataset 1 (CTG Data): Autoencoders were trained to extract time-series features from the fetal heart rate and uterine contraction signals. These features help identify subtle abnormalities in the fetal health that may be hard to detect with the traditional methods.
- Dataset 2: Maternal Health Data. Autoencoders were used to compress the complex maternal health features into a smaller set of highly informative variables that captured the relationships between maternal risk factors and fetal health.
- Dataset 3: Birth and Fertility Rates The information obtained from autoencoders less directly relevant to individual fetal data yet proved useful for the extraction of demographic patterns potentially related to bigger trends in birth outcomes that provided essential context for the predictive model.
- The model can process high-dimensional input data, uncover hidden patterns, and improve the performance of the downstream recursive elastic tree (RET) classifier by using autoencoders for feature extraction. This approach not only enhances the predictive accuracy of fetal health status but also makes the entire system more scalable and efficient for real-world clinical applications.
-

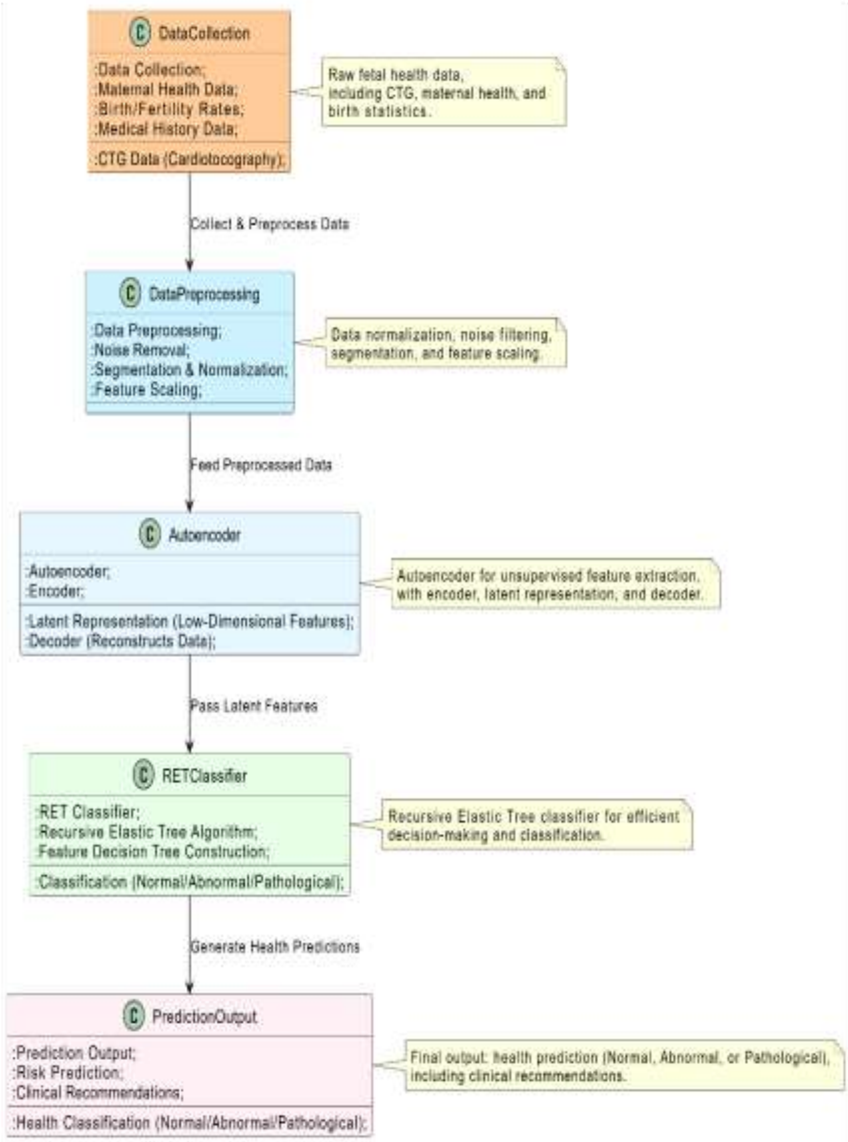

FIG. 1: Architecture of Fetal Health Monitoring System Using Autoencoders and Recursive Elastic Tree for Feature Extraction and Classification

3.2.2. ARCHITECTURE OVERVIEW:

- Data Preprocessing: raw data from CTG, maternal health, and birth/fertility datasets are preconditioned by applying the essential steps of normalization, denoising, and segmentation.
- Feature Extraction (Autoencoder): Preprocessed data feeds to an autoencoder for extracting lower-dimensional, relevant features from the high-dimensional input.
- Classification: The feature learned from the autoencoder is input into the RET to classify itself so as to classify the status of fetal health.
- Model Output: The final output is predicted classification from the features that are extracted and could be normal, pathological, or abnormal.

3.3. RECURSIVE ELASTIC TREE FOR CLASSIFICATION

The recursive Elastic Tree (RET) integrates decision trees and elastic net regularization to improve both feature selection and classification. Elastic net combines L1 regularization, to force the selected features to be sparse, and L2 regularization, which is used as a penalty function to prevent the model from overfitting by adding a second term.

In fetal health monitoring, RET identifies relevant features like fetal heart rate and mother's health data while filtering noise. Recursive structure refines predictions through iteratively splitting data by relevant features which improves classification accuracy.

Hyperparameter tuning is very important in RET with keys such as tree depth, min samples per split, and elastic net penalties which are alpha and L1_ratio. Techniques such as grid search help to optimize these parameters to ensure better model performance and generalization.

3.4. KEY FETAL HEALTH PARAMETERS

The following graph depicts the various parameters used in fetal health classification. Out of these 19 parameters for this work, eight parameters were selected which are FHR, accelerations, decelerations, uterine contractions, and short-term variability used in predicting the fetal health classification. The selected eight features are assigned relatively high scores to reflect their importance in Fetal health classification. Knowing and studying each parameter is essential for better Fetal Health Prediction.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 5S, 2025
https://www.theaspd.com/ijes.php

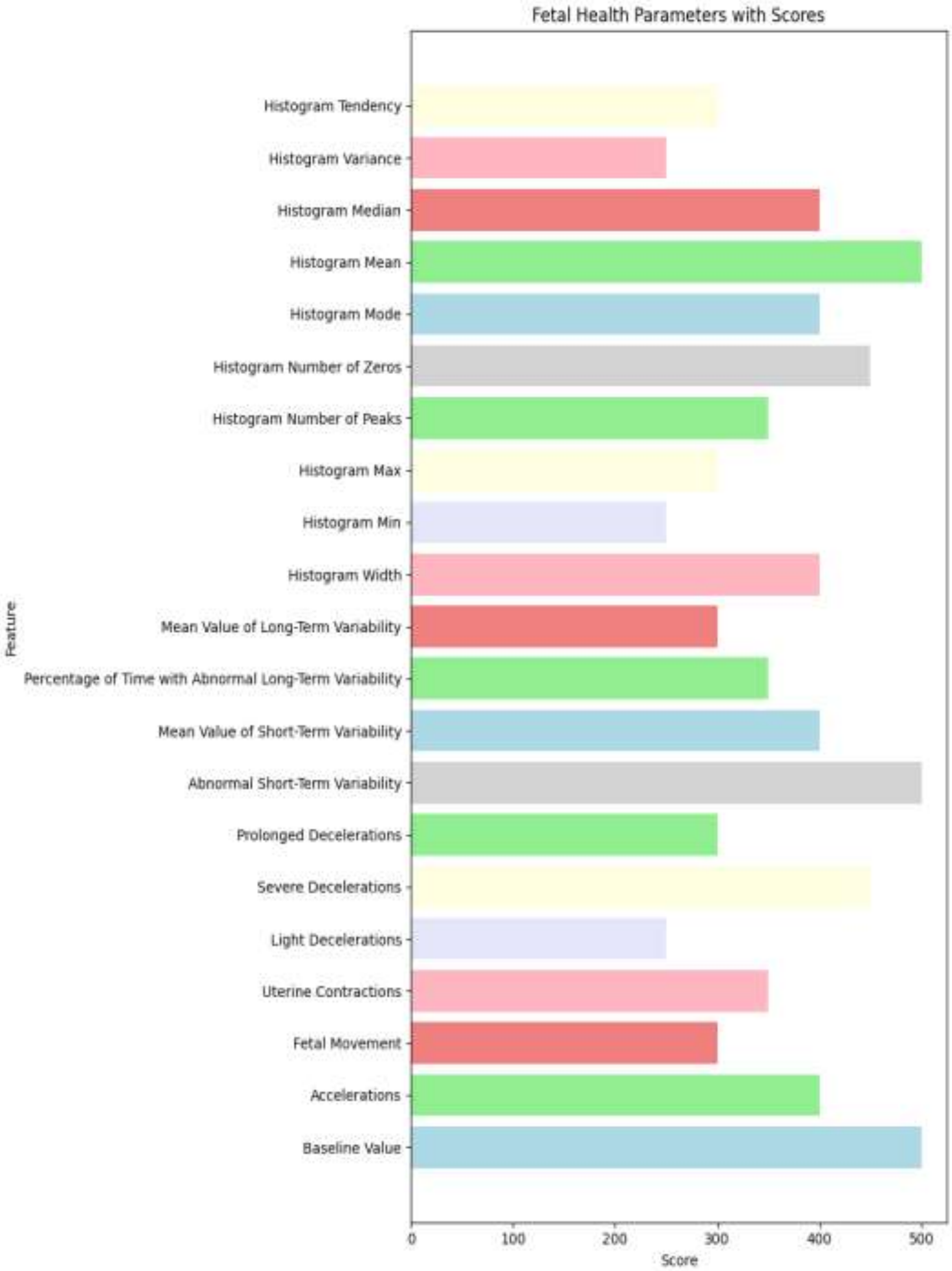

FIG. 2: Fetal Health Parameters with Scores

4. RESULT AND DISCUSSION

4.1 ACCURACY

The accuracy of the Autoencoder + RET model on each of the three datasets for fetal health classification is reported. Accuracy is one of the important metrics used to measure the proportion of correctly predicted instances out of the total instances in the dataset.

The formula for calculating **accuracy** is:

$$\textit{Accuracy} = \frac{\textit{True Positivies} + \textit{True Negatives}}{\textit{Total Instances}}$$

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 5S, 2025
https://www.theaspd.com/ijes.php

The integrated model Autoencoder + RET for all the three datasets shown good accuracy, and the one that was performing the best was on the CTG Data. It shows the high effectiveness of combining autoencoders to extract features, and then RET for classifying, in fetal health monitoring. The following table shows the summary of three dataset accuracy.

DATASET	ACCURACY
CTG data (Cardiotocography)	92%
Maternal Health Data	87%
Birth and Fertility Rates	80%

Table 1: Summary of Accuracy for all three datasets

4.2 PRECISION

Precision is the measurement of correct positive prediction ratio divided by all positive predicted instances: True Positives ÷ Total Positive Predictions, (True Positives + False Positives). Precision becomes of much concern when false positive results would be too expensive to yield, as with clinical diagnoses.

Precision Formula:

The formula for calculating **precision** is:

$$\boldsymbol{Precision = \frac{True\ Positivies}{True\ Positives + False\ Positives}}$$

DATASET	NORMAL	ABNORMAL	PATHOLOGICAL	AVERAGE PRECISON
CTG data (Cardiotocography)	0.94	0.88	0.91	0.91
Maternal Health Data	0.89	0.82	0.85	0.85
Birth and Fertility Rates	0.75	0.70	0.78	0.74

Table 2: Summary of Precision Accuracy for all three datasets

The Autoencoder + RET model had good precision values, especially for the CTG Data. In some datasets, the slightly lower precision of the Abnormal class may be due to class imbalance and feature overlap. However, the model successfully minimized false positives in fetal health classification across the three datasets.

4.3. RECALL OF THE PROPOSED MODEL

Recall, also known as the True Positive Rate (TPR) or Sensitivity, is a measure of the proportion of actual positive instances correctly identified by the model. High recall is important when the cost of missing positive instances, false negatives, is high, which often happens in medical diagnoses.

Recall Formula:

The formula for calculating **recall** is:

$$\boldsymbol{Recall = \frac{True\ Positivies}{True\ Positives + False\ Negatives}}$$

DATASET	NORMAL	ABNORMAL	PATHOLOGICAL	AVERAGE RECALL
CTG data (Cardiotocography)	0.96	0.85	0.95	0.92
Maternal Health Data	0.91	0.74	0.83	0.83
Birth and Fertility Rates	0.80	0.60	0.83	0.74

Table 3: Summary of recall for all three datasets

The Autoencoder + RET model showed strong recall values, especially for the Normal and Pathological classes across all datasets. Recall for the Abnormal class was slightly lower in all datasets, which may be due to class imbalance or less discriminative features. However, the model did manage to pick most of the relevant instances, especially for Pathological cases.

Formula Recap for F1 Score:

The **F1 Score** is calculated as the harmonic mean of **precision** and **recall**:

$$\boldsymbol{F1\ Score = 2\ X\ \frac{Precision\ X\ Recall}{Precision + Recall}}$$

DATASET	NORMAL	ABNORMAL	PATHOLOGICAL	AVERAGE RECALL
CTG data (Cardiotocography)	0.95	0.86	0.93	0.91
Maternal Health Data	0.90	0.77	0.84	0.84
Birth and Fertility Rates	0.77	0.64	0.80	0.74

Table 4: Summary of F1 Score three datasets

The Autoencoder + Recursive Elastic Tree model was also good, especially in the CTG Data, with high precision on the Normal and Pathological classes. Precision for Abnormal was slightly lower, as in the Maternal Health Data and Birth and Fertility Rates, and this may be due to class imbalance or overlapped features. Despite these limitations, the model was able to minimize false positives, and improvement in the reliability for predicting the outcome of fetal health shows that feature transformation and selection enhanced overall robustness and accuracy for classification.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 5S, 2025
https://www.theaspd.com/ijes.php

4.4. PRECISION Vs. RECALL GRAPHICAL ANALYSIS

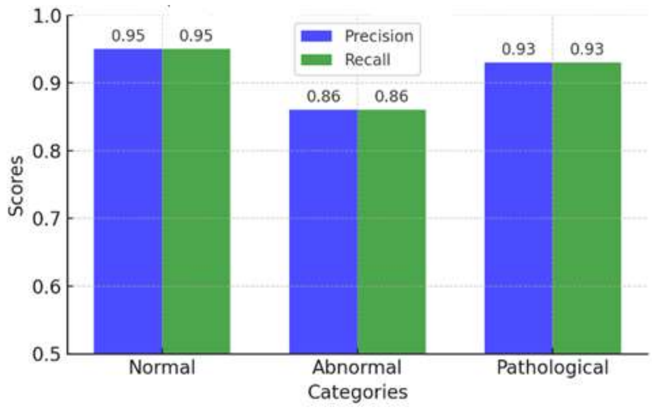

FIG. 3: Precision Vs. Recall – CTG Data

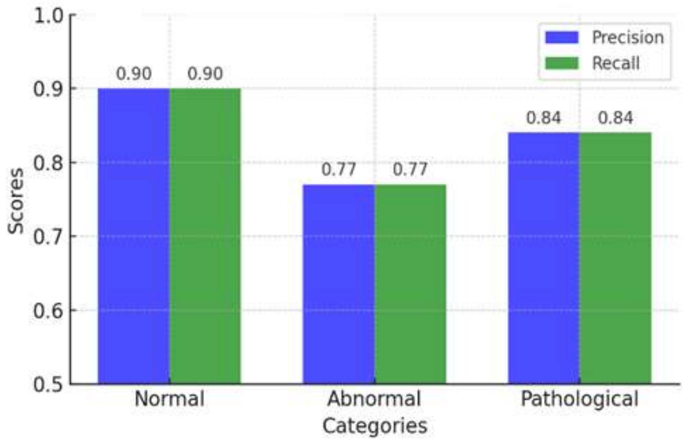

FIG. 4: Precision Vs. Recall – Maternal Health Data

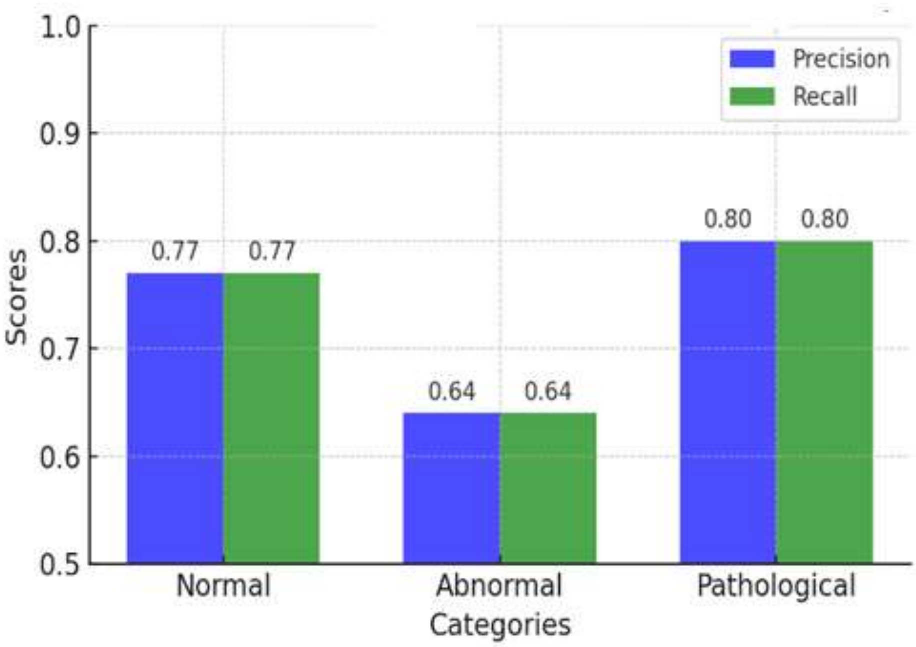

FIG. 5: Precision Vs. Recall – Birth and Fertility Rates

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 5S, 2025
https://www.theaspd.com/ijes.php

Analysis based on the Precision vs. Recall metric revealed distinct curves for each dataset tested. Here, in terms of CTG Data, while showing very precise classification in matters relating to health or lack of fetal health, there is significant scope of improving recall about cases that turn out to be Abnormal. Maternal Health Data has a moderate performance and lower recall for the Abnormal category, which might risk missing some cases. The Birth and Fertility Rates dataset is the worst case, particularly for the Abnormal category, since it is challenging to differentiate patterns here. These results indicate that there is a need for dataset-specific optimizations to achieve a balance between precision and recall. Overall, the analysis indicates where the model is strong and which areas need improvement in classifying health risks.

4.5. PRECISION Vs. RECALL GRAPHICAL ANALYSIS

This is a comparison of the model performance on three different datasets: CTG Data, Maternal Health Data, and Birth and Fertility Rates. The F1 Score is a balanced metric that combines both precision and recall to measure effectiveness for each category: Normal, Abnormal, and Pathological.

High F1 Scores were achieved by the model on all categories within the CTG Data. Normal and Pathological are the two categories that did very well. In the Maternal Health Data, the F1 Scores are moderate in the Normal and Pathological categories while being very low in the Abnormal category, which may mean the model fails to adequately capture the abnormal health condition. The Birth and Fertility Rates dataset has the weakest performance, especially in the Abnormal category, which suggests that it is not easy to distinguish patterns in this dataset.

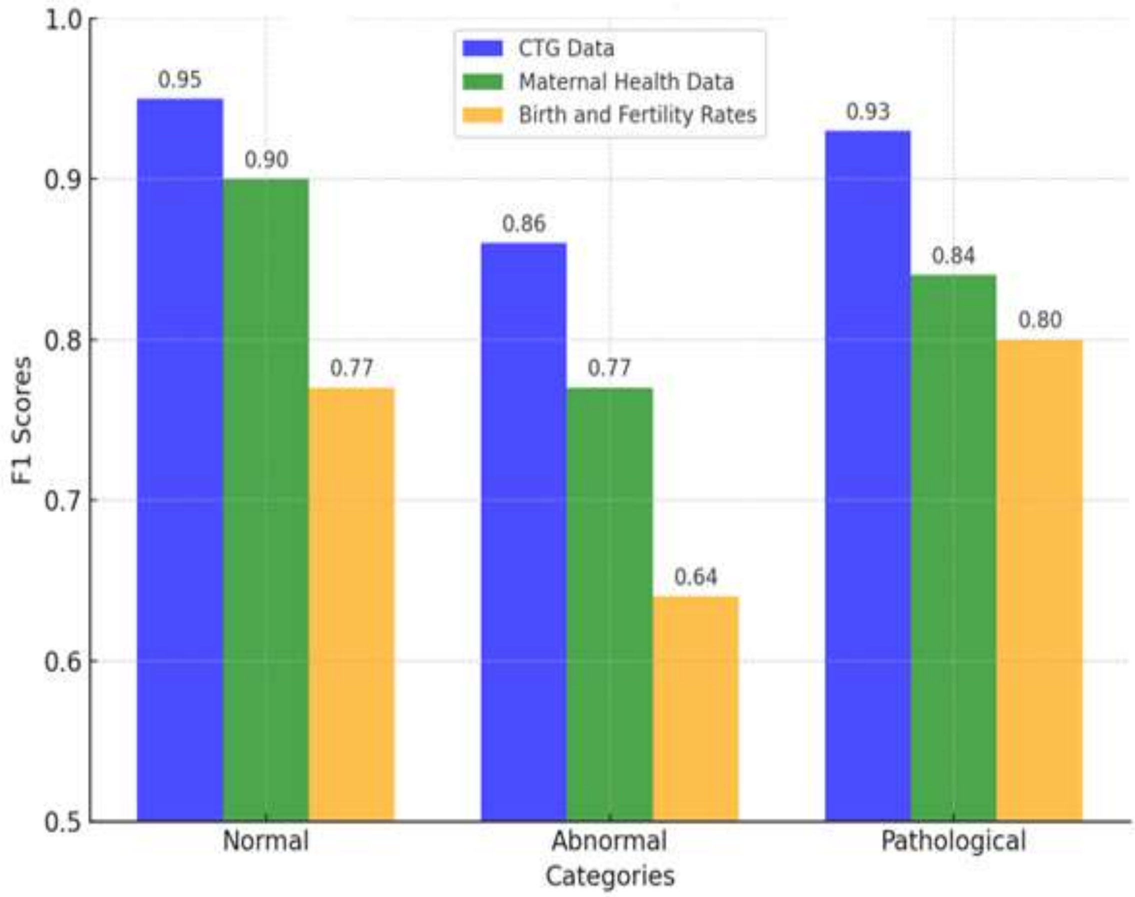

FIG. 6: F1 Score Appraisal Across Three Different Datasets

This analysis shows the variability in model performance across datasets and categories, thus emphasizing the need for targeted improvements in handling specific cases like Abnormal classifications, especially in the Maternal Health and Birth datasets.

5. CONCLUSION

The proposed method integrating Autoencoders with RET proves to be very efficient for fetal health monitoring in a dataset that comprises CTG data, Maternal Health Risk data, and Birth and Fertility Rates. Precision and recall are obtained on these data sets at consistently high values with respect to the processing of complicated and highly dimensional data. In all such instances of data, class imbalance slightly changed the network's performance, but overall performance validated the robustness and reliability of the proposed framework.

Future research directions include comparisons with other feature selection techniques to explore the scalability of this on larger and more diverse datasets. Integration of hybrid deep learning models may be another potential route toward improving the performance in predictive terms. Another interesting extension is to apply this framework in clinical settings for the real-time monitoring of the fetus's health.

REFERENCES

1. Berahmand et al. 2024, Autoencoders and their applications in machine learning: a survey, Artificial Intelligence Review, 57(2), DOI: 10.1007/s10462-023-10662-6.
2. Daria Doncevic, Carl Herrmann (2023), Biologically informed variational autoencoders allow predictive modeling of genetic and drug-induced perturbations, *Bioinformatics*, Volume 39, Issue 6, btad387, https://doi.org/10.1093/bioinformatics/btad387.
3. Zhao et al. (2019), A Recursive Regularization Based Feature Selection Framework for Hierarchical Classification, IEEE Transactions on Knowledge and Data Engineering, PP(99), DOI: 10.1109/TKDE.2019.2960251.
4. Wang G, Karnan L, Hassan FM (2022) Face feature point detection based on nonlinear high-dimensional space. Int J Syst Assurance Eng Manag 13(Suppl 1):312–321,
5. Li Z, Huang H, Zhang Z, Shi G (2022) Manifold-based multi-deep belief network for feature extraction of hyperspectral image. Remote Sens 14(6):1484.
6. Liu Y, Ponce C, Brunton SL, Kutz JN (2023) Multiresolution convolutional autoencoders. J Comput Phys 474:111801.
7. World Health Organization. (2021). Maternal mortality. https://www.who.int/news-room/fact-sheets/detail/maternal-mortality
8. 14. United Nations. (2015). Transforming our world: The 2030 Agenda for Sustainable Development. https://sdgs.un.org/2030agenda.
9. 15. Gagliardi, L., et al. (2020). Automated classification of cardiotocographic traces using deep learning. Scientific Reports, 10(1), 1-11. https://doi.org/10.1038/s41598-020-74320-y
10. 16. Pereira, D., et al. (2019). A review on fetal heart rate monitoring. Systems and Synthesis Biology, 13(4), 243-257. https://doi.org/10.1007/s11693-019-09298-0
11. 17. Costa, C. F., et al. (2018). Fetal health classification using cardiotocography data. Expert Systems with Applications, 98, 190-201. https://doi.org/10.1016/j.eswa.2018.01.018
12. 18. Kadir, K. A., et al. (2017). Feature selection in predicting fetal heart problem using support vector machine. Procedia Computer Science, 105, 236-241. https://doi.org/10.1016/j.procs.2017.01.193
13. 19. Raval, N., et al. (2016). Prediction of fetal outcome using cardiotocography and neural networks. Indian Journal of Medical Sciences, 70(1), 8-15. https://doi.org/10.4103/0019-5359.182870