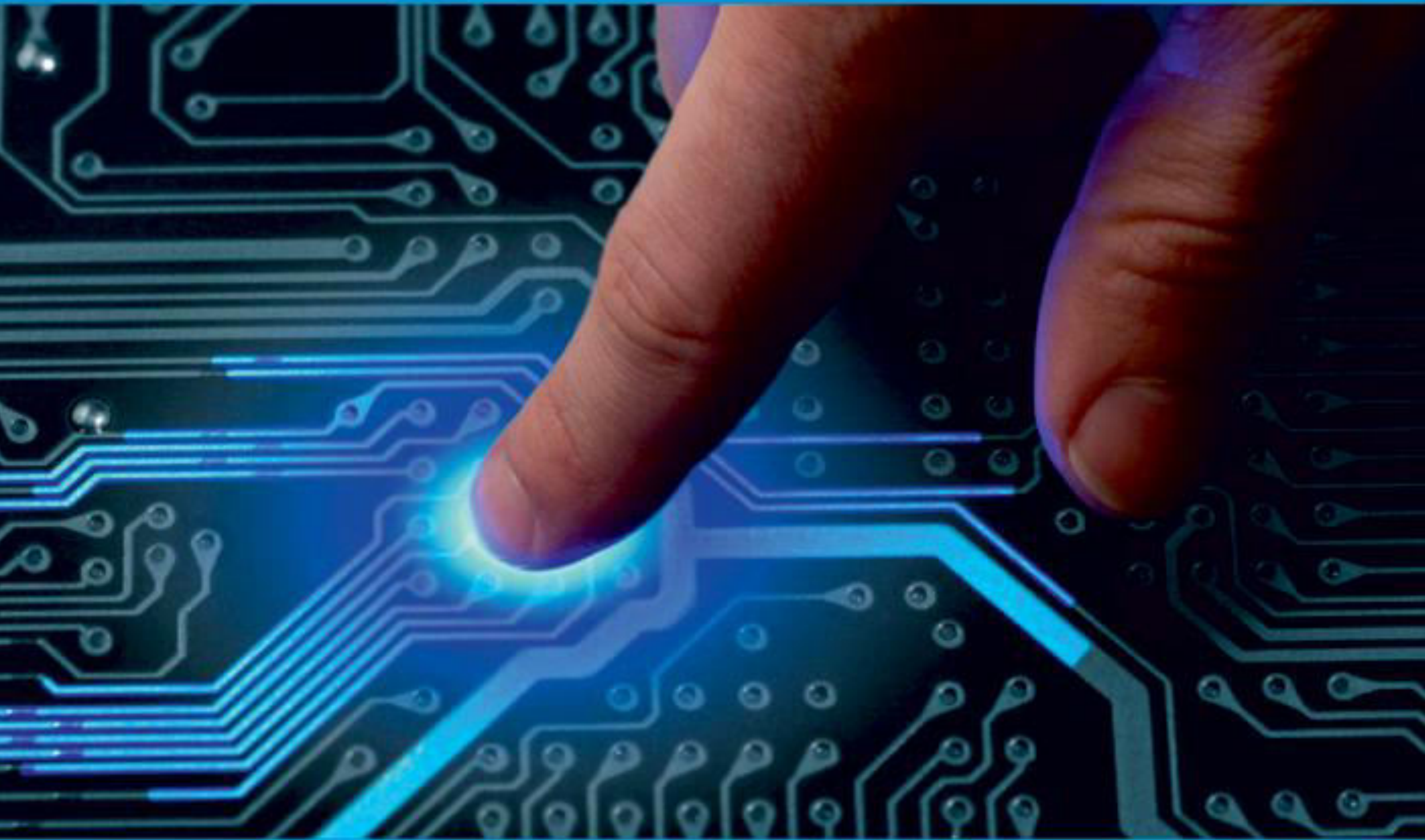
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 1, January 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

9940 572 462 6381 907 438 ijircce@gmail.com www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1201012 |

Design and Implementation of Fetal Distress Detection Model using Machine Learning and Deep Learning Approaches

Vaishnavi[*1], Rohini A[*2], Irala Gnapika Reddy[*3], Sanjana N[*4]

UG Student, Dept. of CSE, Presidency University, Bengaluru, Karnataka, India[*1,3,4]

Assistant Professor, Dept. of CSE, Presidency University, Bengaluru, Karnataka, India[*2]

ABSTRACT: Cardiotocography (CTG) is a technique which monitors and determines the level of fetal distress. Even though, CTG is mostly used for monitoring fetus health, the presence of high number of false positive results in inappropriate surgical delivery and delayed intervention. The paper aims to determine how CTG data can be used to detect fetal distress using machine learning and deep learning models. The CTG data is acquired from CTU-UHB Database, from which the Fetal heart rate (FHR) and Uterine Contractions (UC) signals are extracted. After the extraction, the signals are resampled and preprocessed. The study concentrated on finding the important combinations of neonatal characteristics, such as umbilical cord pH and Apgar5 (5-minute) score, with optimum thresholds for classifying the samples accurately. Furthermore, to address challenges associated with small dataset size and class imbalance, we apply data augmentation. Various classifiers, such as Support Vector Machine (SVM), Random Forest (RF), and Convolutional Neural Network (CNN) are used, and they are assessed in terms of accuracy, precision, recall and f1-score. Experimental results reveal that CNN achieves the highest accuracy of 99.5%.

KEYWORDS: Fetal Distress, Deep Learning, Cardiotocography, Machine Learning, Support Vector Machine, Random Forest, Convolutional Neural Networks

I. INTRODUCTION

Several challenges are faced by doctors while delivering a baby, and one of the most crucial concern is ensuring the unborn baby's well-being. Fetal distress [1] arises when insufficient oxygen reaches the baby's tissues leading to an elevation in the acidity of the baby's body fluids. Without immediate intervention, it can cause severe damage to the baby's brain or, in extreme cases, prenatal demise. Similar to diagnosing other health conditions, detecting fetal distress is usually subjective and depends on the expertise of healthcare professionals. Fetal distress is a complex condition influenced by various clinical factors like neonatal parameters and maternal risk factors. One of the widely used technique to detect fetal distress is Cardiotocography. CTG was first introduced in the mid-twentieth century by E. Hon [4]. It evaluates fetal health based on two key parameters: fetal heart rate (FHR) and uterine contractions (UC). The Dr. C Bravado approach is followed for the manual interpretation of CTG. Fetal Distress is a challenging condition as there are many factors which can influence it. Fetal distress when combined with the subjective interpretation of CTG, shows inconsistency in how different medical experts assess the same graph. While CTG was introduced to assess fetal distress, it hasn't significantly improved the prediction rates of abnormal patterns. Therefore, it's important to aim for creating a very objective method to detect fetal distress. This could be crucial in accurately recognizing the cases of fetal distress, lowering the instances of false positives, reducing the rate of unnecessary caesarean sections, and, in the end, helps reducing mortality rates of the infant. Detecting fetal distress in an early stage, using Machine learning and Deep learning technologies can help the patient's well-being. Conventional algorithms and Deep Learning algorithms are used in this study to detect fetal distress.

II. RELATED WORK

Extensive research has been conducted, and is still ongoing in the area of machine learning for identifying fetal distress. Several papers were reviewed to understand the performance of existing models on various datasets using machine learning and deep learning techniques. Further, the studied paper are discussed in brief.

Y.D.Daydulo et al. [4], developed deep learning model for the classification of Fetal Distress, it was stated that only samples with pH greater than 7.15 were labelled "Normal" and pH less than or equal to 7.15 were labelled as "Distress". The authors have not evaluated Apgar5 score as it was considered as a subjective labelling criterion. In

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1201012 |

this model, 439 samples out of 552 recordings were labelled as “Normal” and the rest 113 samples were labelled as “Distress”. Authors have used Morse Wavelet for signal processing and amalgamation of transfer learning with a ResNet50 Model for the classification of fetal distress. They acquired high accuracies of 98.7% for the first stage of labor and 96.1% for the last stage of labor.

A similar labelling approach was followed in a study by H.Liang et al. [7], where samples were classified as distress only when there pH was less than 7.15. The model produced 447 Normal cases which is exactly 8 more cases than the initially reviewed paper. The authors have used one dimensional CNN Model with bidirectional GRU to create a highly effective model which provides an accuracy of 96%.

In a study by M.E O’Sullivan et al. [8], it was observed that their criteria for labelling the samples as “Distress” was pH value to be less than or equal to 7.0 and low Apgar5 score between 5 and 6. The authors have obtained the CTG data from the CTU-UHB database [2], which comprised of 552 samples. The dataset was then filtered by excluding samples which had over 30 percent missing traces in CTG. In the model, 310 samples were labelled as “Normal”, 23 samples as “Distress”, and 99 samples did not belong to either. The authors have used machine learning algorithms such as SVM and Logistic Regression and their objective is to enhance the features within the CTU-UHB database concentrating on complexities of feature engineering.

Another study by Sahana et al. [9], proposed a machine learning model which was applied to the first and second stage of labor, using SVM, RF, Bagging and Multi-layer perceptron (MLP) for the classification of CTG data. The Sensitivity and Specificity ranged from 92.7-96.4% and 92.4-98.4%. It was seen that SVM and RF had outperformed Bagging and MLP, as they exhibited better performance, exclusively in the second stage. SVM and RF showed a great level of agreement between the model and medical experts.

In a study conducted by Liang and Li [10], the criteria for labelling the samples as “Normal” was pH to be less than or equal to 7.05. The authors incorporated a CNN model and acquired an accuracy of 89.3% and 79.95% at different threshold values. In this model, 508 samples were labelled as “Normal” and the rest 44 as “Distress”. The authors have also obtained the CTG data from the CTU-UHB database.

In another study by Zeng et al. [11], the criteria for labelling the samples as “Normal” was pH less than or equal to 7.05 and Base Excess (BE) less than or equal to -10. The authors have also obtained the CTG data from the CTU- UHB database. They have used techniques such as Continuous Wavelet Transform (CWT), Wavelet Transform Coherence (WTC), and Cross Wavelet Transform (XWT) combined with Ensemble Cost-Sensitive Support Vector Machine (ECSVM) for the classification of fetal distress. The model achieved an accuracy of 67.2%, Sensitivity of 85.2% and Specificity of 66.1%.

In a study by Liu et al. [12], the criteria for labelling the samples as “Normal” was pH to be less than or equal to 7.15. The authors had worked on CTU-UHB Database and the methodology involves signal preprocessing and down sampling, using CNN for spatial feature extraction combined with Bi-LSTM to address long-term dependencies. The authors also used DWT for obtaining transformation coefficient features from the FHR Signals. Their proposed method has achieved an accuracy of 71.71%, Sensitivity of 75.23%, and Specificity of 70.82%.

In a study by Astik et al. [13], the authors used various machine learning algorithms such as Logistic Regression, KNN, RF and Gradient Boosting Machines. The models are evaluated based on precision, recall, f-1 score and accuracy. The authors worked on a dataset from Kaggle, which had 2126 samples. The samples were labelled as “Normal”, Suspect” and “Pathological”. The Random forest model achieved an accuracy of 93%, which was highest when compared to other models.

In a study by V.Chuda‘ cek et al. [2], which introduced the CTU-UHB database had over 21 publications on the fetal distress classification criteria, which were applied to many databases, are shown in Table 2 of the study. In which, thirteen studies used pH as a criterion and five studies used Apgar score as a criterion.

According to the literature survey, Authors have proposed various Conventional and Deep Learning models on various datasets. By understanding each phase of the proposed models, authors have incorporated various techniques and methodologies which includes, data labelling, data augmentation, machine learning and deep learning models. For labelling the dataset into two classes (i.e. Normal and Distress), authors have used pH and Apgar score [3][6] as a criterion. For Data augmentation, methods such as jittering, rotating, scaling, window slicing and so on were used to increase the number of samples in the dataset. To address the semantic gap between the existing models various machine learning and deep learning models are incorporated in the proposed model.

III. PROPOSED ALGORITHM

The proposed fetal distress classification model is described in this section. The main objective of this study is to use effective methods to reduce the misclassification of fetal distress. Initially, the samples present in the dataset are labelled, extracted and pre-processed. Labelling of the samples is done based on the pH value and Apgar5 score. Data augmentation is then performed to increase the size of the dataset. Finally, Support Vector Machine (SVM), Random

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1201012 |

Forest (RF) and Convolutional neural network (CNN) models are trained and assessed based on their performance. The steps involved for developing our proposed model are explained in the upcoming sections.

A. **Data Labeling**

The Labelling Phase involves assigning labels to the 552 samples of CTU-UHB database. The parameters that highly indicate fetal distress are opted. By examining the header file, each sample's umbilical cord pH and Apgar 5 minute [6] were compared with the predefined thresholds.

The criteria to label the samples as "Normal" is as follows:

- pH < 7.15
- 7< Apgar5 < 9

The samples whose pH and Apgar5 score falls below or above the criteria are labelled as "Distress". The classification resulted in 439 Normal samples and 113 Distress samples as shown in Fig. 1.

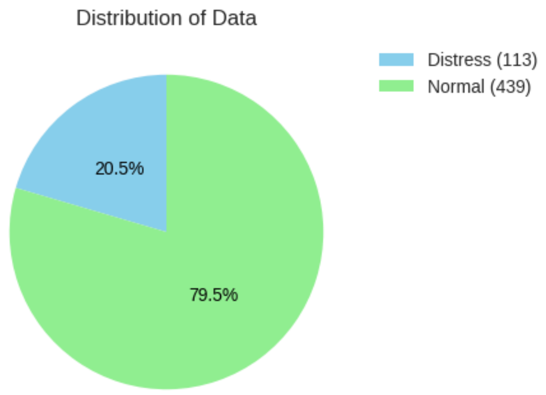


Fig 1: Distribution of classes after labeling.

B. **Data Preprocessing**

The preprocessing phase involves extraction of FHR and UC signals using the wfdb python library from the CTU-UHB Database. Raw signals for each sample are acquired and based on the signal names FHR and UC are detached. The extracted FHR and UC signals are resampled to thousand features each, to handle time series data. Since, the unit of FHR (bpm) and UC signals (mmHg) differ, Min-Max Normalization is performed for each sample. Each value of the FHR signal is scaled by maximum FHR value and same is done with UC signal, which denotes that the values are within the uniform range of 0 and 1. The preprocessed FHR and UC features are then concatenated and appended to a python list, which consists of preprocessed signals for all 552 samples. The shape of this list is (552, 2000), which is used for further analysis and model training.

C. **Data Augmentation**

When SVM and RF model were trained with the 552 samples, it was understood that both SVM and RF are not providing best outcomes and can do better. Challenges arse while training deep learning models like CNN due to database's limited size. Data Augmentation is done to address these limitations for which tsaug python library is used. Gaussian Noise transformation with a standard deviation of 0.01 is applied, which adds noise to the original signals. To expand the dataset, 20 augmented samples are generated for each original sample. By vertically stacking the original and augmented data, the augmented dataset is reshaped. To enhance sparsity and variability, the augmented dataset is shuffled before training and classification. The total number of samples increased from 552 to 11592 after augmentation. The augmentation is not only applied to the original dataset but also is demonstrated using SVM and RF models. The class distribution after augmentation is shown by Fig. 2. After enlarging the dataset through data

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1201012 |

augmentation, an issue of class imbalance persisted. The class distribution which is shown in Fig. 2 depicts that out of 11595 samples, 2373 samples were labelled “Distress” and rest (9219) samples were labelled as “Normal”. It is observed that number of samples in distressed class were very less than samples in the Normal class. To address this issue, 2373 normal samples were randomly selected for balancing the dataset ensuring 50 to 50 class distribution [7]. Then, the balanced dataset was divided into training and testing sets where 70% of dataset was used for training (3322 samples) and 30% for testing (1424 samples).

To facilitate the Model’s understanding, the labels (’Normal’ and ’Distress’) were encoded into numerical values. The Training and Testing sets were reshaped by transforming the two-dimensional matrix into three-dimensional tensor. This preprocessing was incorporated into SVM, RF and CNN Models, ensuring uniform data representation and input compatibility for different classifiers.

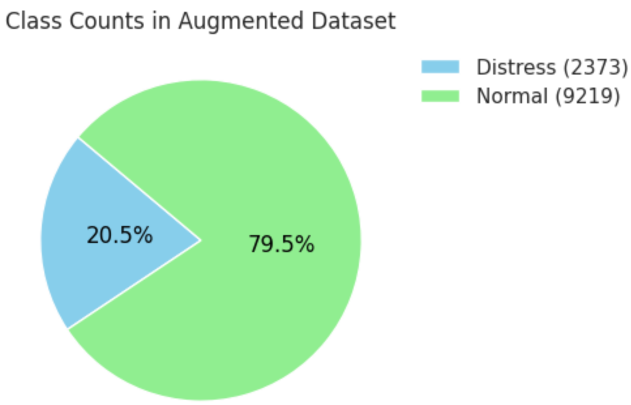


Fig 2: Distribution of classes after data augmentation.

D. **Classification**

Classification is a process of extracting a collection of features from the samples and then mapping them to a specified class. As this study is done on a two-class problem, the Conventional models (SVM and RF) and deep learning model (CNN) are used to assign the samples into Normal and Distress classes. Further, the considered classifiers are evaluated based on their performance.

1. Support Vector Machine
   The linear SVM classifier is widely used for binary classification. The linear SVM is incorporated as it classifies the given input samples into normal and distress classes. The classifier analyzes the samples and trains the model to minimize misclassifications of Fetal distress. The trained SVM Model is later tested by providing samples which were not trained.

2. Random Forest
   Random Forest is an ensemble learning algorithm that combines multiple decision trees to get a single result which enhances classification accuracy. RF Classifier plays a crucial role in the classification process, specifically for binary classification of the samples into “Normal” and “Distress” classes. In the proposed model, RF is initialized with parameters such as 50 estimators, a maximum depth of 5, and a random state of 42. The ensemble learning approach allows the model to provide best insights from multiple decision trees. The RF classifier carefully processes samples and trains the model to optimize its ability to classify fetal distress accurately. After training the RF model is tested where its effectiveness is evaluated using untrained samples. The RF consists of diverse decision trees, which contributes to generalization and robustness capabilities of the model.

3. Convulational Neural Networks
   Convulational Neural Networks is a deep learning algorithm which is effective for classifying time-series data. It is well-suited for image recognition and processing tasks. It consists of multiple layers, which includes convolutional layers, pooling layers, and fully connected layers. The proposed model consists a sequence of layers, each designed for a specific operation. The following layers were used in the proposed model:

International Journal of Innovative Research in Computer and Communication Engineering

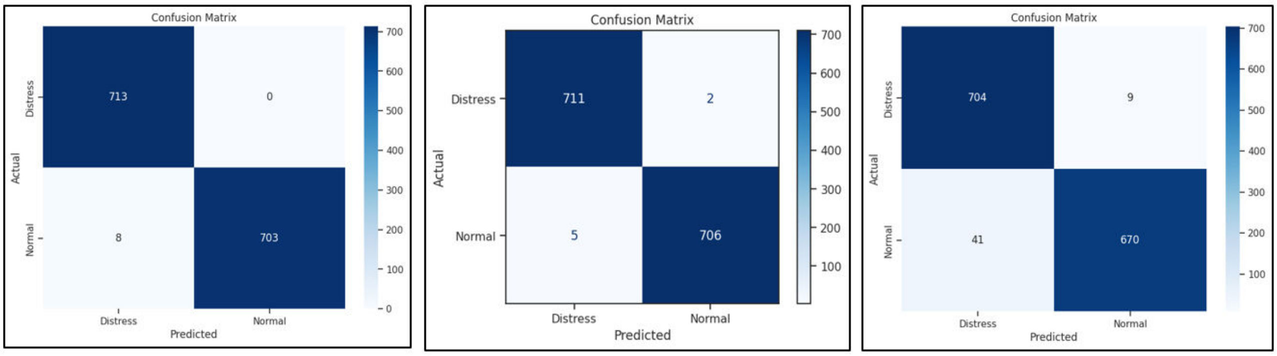
| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1201012 |

1) Convolutional 1D Layer: The first convolutional layer with 32 and 64 filters of size 3. It extracts essential features and transforms it into a tensor.
2) 1D Max Pooling Layer: This layer performs one dimensional max pooling over 2-unit window, which is used for downsampling the spatial dimensions.
3) Batch Normalization: This layer enhances the training stability by normalizing the output of the previous layer.
4) Dropout Layer: This layer prevents overfitting by randomly deactivating the neurons during training.
5) Flatten Layer: This layer flattens the three-dimensional output into a one-dimensional vector which is used for further processing.
6) Dense Layer: It is a fully connected layer with 128 neurons. It processes the flattened input. The batch size is set to 64 and the learning rate to 0.00011. Before creation of the second fully connected layer, the dropout rate is increased from 50% to 70%. The model is compiled using the Adam optimizer, incorporating the previously mentioned learning rate and utilizing binary cross-entropy as loss function also the model was trained for 50 epochs. Then the model's ability of classifying was evaluated based on its performance.

IV. SIMULATION RESULTS



a) SVM b) CNN c) RF

Fig. 3: Confusion Matrix

The performance of the proposed model is examined by conducting experimentation on the CTG samples of CTU-UHB database. The database consists of 552 samples, the proposed model is balanced with 2373 Normal samples and 2373 Distress samples, later trained with 3322 samples. The model is then tested with 1424 untrained samples. The performance measures of the proposed model is evaluated based on accuracy and confusion matrix which measures precision, recall and F1 score for each model. Accuracy (Equation 1): Identifying distressed samples from a pool of samples accurately. Precision (Equation 2): Samples which are correctly identified over the correctly and wrongly categorized samples. Recall (Equation 3): Samples which are accurately categorized samples over the correctly identified samples and wrongly rejected samples. F1-Score (Equation 4): It is the harmonic mean of Recall and Precision. These measures use specific parameters such as True Positives (TP) for correctly identified samples, False Positives (FP) for wrongly identified samples, True Negatives (TN) for correctly rejected samples and False Negatives (FN) for wrongly rejected samples.

$$\text{Accuracy} = (TP + TN) / (TP + FP + FN + TN) \quad (1)$$
$$\text{Precision} = TP / (TP + FP) \quad (2)$$
$$\text{Recall} = TP / (TP + FN) \quad (3)$$
$$\text{F1 Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision}) \quad (4)$$

The parameters such as pH value and Apgar5 score play an important role in the classification of the CTG recordings into normal and distress classes. The classifiers are assessed based on their performance. The model which outperforms the measuring terms and provides accurate classification is selected by analyzing the performance of individual classifiers. The performance analysis of the classifiers is represented by Fig. 4. It is found that Convolutional Neural (CNN) is providing better performance when compared with other two models.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1201012 |

Classifiers	TP	TN	FP	FN	Accuracy	Precision	Recall	FScore
SVM	713	703	0	8	99.43	99.44	99.43	99.43
RF	704	670	9	41	96.48	96.58	96.48	96.48
CNN	711	706	2	5	99.50	99.71	99.29	99.50

Fig. 4: Performance Analysis

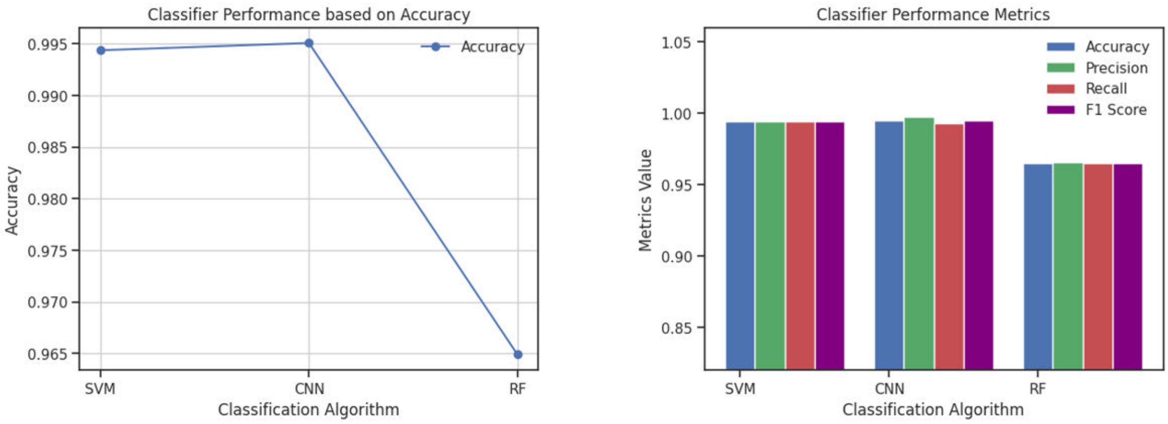


Fig. 5: ROC performance of each classifier

The accuracy score of CNN is 99.5, which is highest among the other conventional algorithms. The CNN and SVM model had the best performance when compared with RF. The Confusion Matrix of each classifier is represented by Fig. 3. The Representation of classification performance which visually presents the classification metrics - Accuracy, Precision, Recall, and F-score for each classifier is represented by Fig. 5, which ensures a comprehensive understanding of the classifier's capabilities.

V. CONCLUSION AND FUTURE WORK

In this study, we created an effective model for analyzing the samples and classifying CTG samples into Normal and Distress classes. The pH of the umbilical cord and Apgar 5-minute score was used to determine the presence of fetal distress in the CTG sample. The issue of small size dataset and class imbalance were solved by employing data augmentation and balancing CTU-UHB database. The performance of the conventional classifiers such as SVM and RF with deep learning model CNN was tested. The performance of each classifier has been examined using the confusion matrix and ROC curve. This study reveals that CNN is the efficient model for detecting normal and distress cases. In the future, the proposed model will be employed to detect fetal distress on different datasets and the effectiveness of the model will be tested.

REFERENCES

[1] J.T. Parer, E.G. Livingston. "*What is fetal distress?*". American Journal of Obstetrics and Gynecology, vol. 162, no. 6, pp. 1421-1427, 1990.

[2] V. Chuda´cek, J. Spilka, M. Bur ˇ sa, P. Jank, L. Hruban, M. Huptych, and ˇ L. Lhotska, "*Open access intrapartum ctg database*," ´ BMC pregnancy and childbirth, vol. 14, pp. 1–12, 2014.

[3] U.S. National Library of Medicine. (n.d.). "*Apgar score*". https://medlineplus.gov/ency/article/003402.htm

[4] H. Hon, "*The electronic evaluation of the fetal heart rate: Preliminary report*," American Journal of Obstetrics & Gynecology, vol. 75, no. 6, pp. 1215–1230, 1958.

[5] Y. D. Daydulo, B. L. Thamineni, H. K. Dasari, and G.T. Aboye, "*Deep learning based fetal distress detection from time frequency representation of cardiotocogram signal using morse wavelet: research study*," BMC Med- ical Informatics and Decision Making, vol. 22, no. 1, p. 329, 2022.

[6] L. Hoffmann, J. Ø. Hjortdal, N. J.Secher, B. Weile "*The relationship between Apgar score, umbilical arterypH and operative delivery for fetal distress in 2778 infants born at term*". European Journal of Obstetricsand Gynecology and Reproductive

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1201012 |

Biology, vol. 38, no. 2, 97-101, 1991.

[7] H. Liang, Y. Lu, Q. Liu, and X. Fu, "*Fully automatic classification of cardiotocographic signals with 1d-cnn and bi-directional gru*," in 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2022, pp. 4590–4594.

[8] M. O'Sullivan, T. Gabruseva, G. B. Boylan, M. O'Riordan, G. Light body, and W. Marnane, "*Classification of fetal compromise during labour: signal processing and feature engineering of the cardiotocograph*,"in 2021 29th European Signal Processing Conference (EUSIPCO). IEEE, 2021, pp. 1331–1335.

[9] S. Das, H. Mukherjee, K. Roy, & C. K. Saha. "*Fetal Health Classification from Cardiotocograph for Both Stages of Labor—A Soft-Computing-Based Approach*".Diagnostics, vol. 13, no. 5, 858, 2023.

[10] S. Liang and Q. Li, "*Automatic Evaluation of Fetal Heart Rate Based on Deep Learning*," 2nd Information Communication Technologies Conference (ICTC), Nan-jing, China, pp. 235-240, 2021.

[11] Zeng R., Lu Y., Long S., Wang C., Bai J. "*Cardiotocography signal abnormality classification using time-frequency features and ensemble cost-sensitive SVM classifier*". Computers in Biology and Medicine, vol. 134, 2021, 104466.

[12] Liu M., Lu Y., Long S., Bai J., Lian W. "*An attention-based CNN-BiLSTM hybrid neural network enhanced with features of discrete wavelet transformation for fetalacidosis classification*". Expert Systems with Applications, vol. 186, 2021, 115714.

[13] A. K. Pradhan, J. K. Rout, A. B. Maharana, B. K. Balabantaray and N. K. Ray, "*A Machine Learning Approach for the Prediction of Fetal Health using CTG*," 2021 19th OITS International Conference on Information Technology (OCIT), Bhubaneswar, India, 2021, pp. 239-244, doi: 10.1109/OCIT53463.2021.00056.

[14] N. Baghel, R.Burget, and M.K. Dutta. "*1D-FHRNet: automatic diagnosis of fetal acidosis from fetal heart rate signals*". Biomedical Signal Processing and Control, 71, p.102794, 2022.

[15] Deng Y, Zhang Y, Zhou Z, Zhang X, Jiao P, Zhao Z. "*A lightweight fetal distress-assisted diagnosis model based on a cross-channel interactive attention mechanism*". Front Physiol. 2023; 14:1090937

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.379

doi crossref

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details