EasyChair Preprint

№ 13557

NeuroFetalNet: Advancing Remote Electronic Fetal Monitoring with a New Dataset and Comparative Analysis of FHR and UCP Impact

Black Sun, Jiaqi Zhao, Xinrong Miao, Yanqiao Wu and Min Fang

June 5, 2024

NeuroFetalNet: Advancing Remote Electronic Fetal Monitoring with a New Dataset and Comparative Analysis of FHR and UCP Impact

Black Sun
School of Computer Science and Technology
Harbin Institute of Technology, Shenzhen
Shenzhen, China
blackthompson770@gmail.com

Jiaqi Zhao
School of Computer Science and Technology
Harbin Institute of Technology, Shenzhen
Shenzhen, China
teabaobo@gmail.com

Xinrong Miao
School of Computer Science and Technology
Harbin Institute of Technology, Shenzhen
Shenzhen, China
stefenmiao@gmail.com

Yanqiao Wu
School of Computer Science and Technology
Harbin Institute of Technology, Shenzhen
Shenzhen, China
210110410@stu.hit.edu.cn

Min Fang
Education Center of Experiments and Innovations
Harbin Institute of Technology, Shenzhen
Shenzhen, China
fangmin@hit.edu.cn

***Abstract*—This study explores the evolution of electronic fetal monitoring (EFM) and introduces remote electronic fetal monitoring (REFM). A new REFM dataset is constructed to address gaps in existing datasets. The study compares the impact of fetal heart rate (FHR), uterine contraction pressure (UCP), and their simultaneous presence on prediction accuracy. By leveraging deep learning techniques, our study introduces NeuroFetalNet, which has shown the best performance in our proposed REFM datasets. The results emphasize the multi-scale feature extractor's potential in NeuroFetalNet to improve the accuracy and efficiency of remote fetal monitoring. These findings have implications for enhancing pregnant and fetal healthcare.**

***Index Terms*—remote electronic fetal monitoring, multi-scale feature extraction, dataset, deep learning, classification**

I. Introduction

During pregnancy, if the fetus lacks sufficient oxygen, it can lead to permanent brain damage, slow fetal development, and even fetal death in severe cases [1]. Cardiotocography (CTG) is a very effective method for fetal monitoring. The signals obtained from fetal heart rate (FHR) and uterine contraction pressure (UCP) can reflect the health status of the fetus inside the uterus. Therefore, through CTG, the fetal abnormalities can be detected and intervened early [2]. Continuous improvement in fetal monitoring methods undoubtedly holds significant clinical significance. Electronic fetal monitoring (EFM) systems have made giant progress with the development of electronics and imaging in medicine. EFM has many advantages, such as real-time, continuous monitoring, and easy interpretation of results [3]. EFM has become the most widely used method for fetal monitoring [4]. As shown in Fig. 1, it is an electronic fetal heart monitor. However, EFM is still an auxiliary tool, and doctors must determine the final results.

Fig. 1. An electronic fetal heart monitor.

For pregnant women, FHR monitoring can start from 32 weeks, while high-risk pregnant women can start from 26 to 28 weeks of pregnancy [5]. However, even low-risk pregnant women are advised to undergo monitoring in the hospital once or twice a week, which results in quite a commuting burden for pregnant women. Therefore, remote electronic fetal monitoring (REFM) is a common practice, where pregnant women use fetal heart monitoring devices at home to collect data. Then, the data will be uploaded to the doctor for analysis. Further

diagnosis can be conducted at the hospital if any abnormalities are detected.

With the development of artificial intelligence in the medical field, many deep learning algorithms can approach or even exceed the level of human experts in disease diagnosis [6]. Therefore, applying deep learning algorithms to fetal electronic monitoring systems can help achieve accurate and automated fetal monitoring analysis. The analysis of FHR and UCP has become increasingly critical for determining the health status of the fetus. The contributions of this research mainly include the following:

- Constructing a new REFM dataset, filling the gaps in existing datasets.
- Comparing the effects of FHR, UCP, and simultaneous FHR and UCP on prediction accuracy, precision, and recall.
- Proposing NeuroFetalNet, which achieved the best performance on the REFM dataset. And our code is available at https://github.com/BlackThompson/NeuroFetalNet.

II. RELATED WORK

A. Dataset of Electronic Fetal Monitoring

There is a scarcity of publicly available Electronic Fetal Monitoring (EFM) datasets. The commonly used dataset is the Cardiotocography dataset from the UC Irvine Machine Learning Repository, published in 2010 [7]. This dataset consists of 2126 fetal cardiotocograms (CTGs) that were automatically processed, and diagnostic features were measured. Additionally, three expert obstetricians classified the CTGs, and a consensus classification label was assigned to each. The classification was based on both morphologic patterns and fetal states. Another dataset is the CTU-CHB [8], a prenatal CTG database comprising 552 data. The CTG data in this dataset were evaluated by nine expert obstetricians based on the annotation of the signals using FIGO guidelines. The CTG recordings started no more than 90 minutes before actual delivery. Each data in the CTG diagram includes fetal heart rate and uterine contraction signal, and the sampling rate of the data is 4Hz. However, these datasets have some limitations: the datasets are relatively old and small. Additionally, The data was collected by specialized medical professionals in hospitals, who are more skilled at operating fetal monitors than average pregnant women.

In contrast, our dataset focuses on remote electronic fetal monitoring (REFM), where pregnant women themselves use fetal monitoring devices to collect data. The data from REFM has many disconnections and abnormalities, which require preprocessing. The methods used to process publicly available EFM datasets are unsuitable for our task. Therefore, we obtained data from cooperating hospitals and constructed the REFM dataset.

B. Predicting Fetal Health Conditions Based on Cardiotocography

Analyzing fetal heart rate (FHR) using data analysis and machine learning or deep learning algorithms is becoming increasingly crucial. Previous studies by Dash et al. [9], Georgoulas et al. [3], and Tsui et al. [10] proposed methods to improve automatic FHR classification. Still, they faced limitations such as small test datasets, reliance on accurate feature extraction, and low true positive rates. Feature engineering is crucial for effective deep learning and machine learning. Spilka et al. [11], Fergus et al. [12], Subasi et al. [13], and Das et al. [14] used various feature extraction and selection methods, applied machine learning models like support vector machine (SVM), random forest (RF) and XGBoost to classify fetal conditions based on selected features.

While CTG feature extraction in previous studies relied on manual or semi-automatic methods, deep learning offers automatic feature extraction with minimal data processing. Convolutional neural networks (CNNs) have succeeded in image classification and have been applied to time series analysis [15], [16]. Researchers like Zhao et al. [17] converted FHR into frequency domain images using the Short-time Fourier transform (STFT). One-dimensional CNNs are commonly applied to time series tasks, as Li et al. [18] demonstrated by dividing the FHR signal into windows for classification. Recurrent neural networks (RNNs), such as LSTM and GRU, have also been utilized for time series analysis, including FHR data. Although its training is time-consuming, Lv et al. [19] successfully predicted ECG diseases using LSTM. Compared to LSTM, GRU provides faster training speeds with streamlined architecture [20]–[22]. Moreover, Liang proposed a method combining 1D-CNN and BiGRU [5].

In summary, while previous studies relied on classical machine learning models and manual feature extraction, utilizing techniques like SVM, RF, and XGBoost, recent advancements have shown the effectiveness of deep learning approaches, such as CNNs and RNNs (LSTM, GRU) for FHR classification. These techniques enable automatic feature extraction from FHR and showcase their potential to improve the accuracy and efficiency of fetal health monitoring. However, previous studies faced a common issue. When inputting data into deep learning models, only fetal heart rate (FHR) or uterine contraction pressure (UCP) was used as independent inputs, without comparing the respective performance for prediction. Hence, in our study, we compared the results of using only FHR as an input, only UCP as an input, and both FHR and UCP as inputs.

III. PROPOSED METHOD

A. Data Collection

Pregnant women collected the data using a fetal heart rate monitor at home. The measurements were uploaded to the cloud, where they were reviewed and assessed by doctors at the hospital to determine the fetus's health condition. The assessments were categorized into normal and abnormal. A total of 38,142 data entries were collected. The gestational age ranged from 30 to 40 weeks, as shown in Fig. 2. The pregnant women ranged from 20 to 42 years old, as shown in Fig. 3. The collected data included fetal heart rate (FHR) and

uterine contractions (UCP), with sampling times ranging from 25 minutes to 60 minutes and a sampling rate of 4Hz.

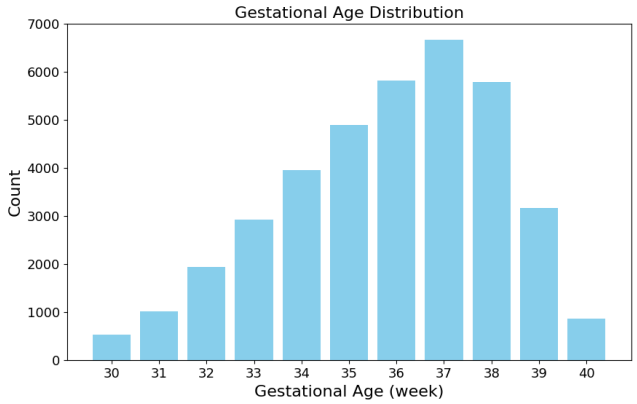

Fig. 2. Gestational age distribution.

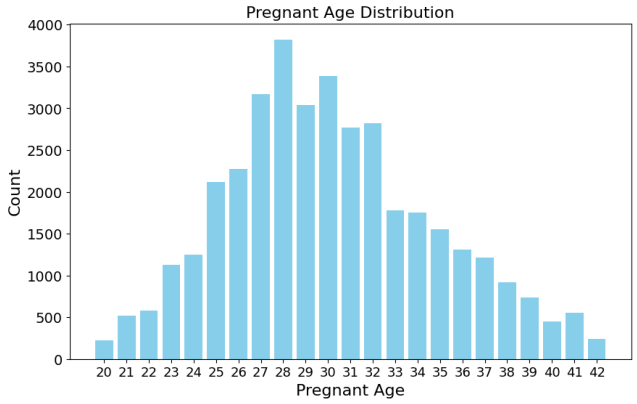

Fig. 3. Pregnant Age distribution.

B. Preprocessing

Initially, there were 38,142 raw data entries. Entries were excluded if they met any of the following criteria: (1) missing data (e.g., missing gestational age information, partial FHR or UCP measurements); (2) obvious data abnormalities (e.g., abnormal maternal age, abnormal gestational age); and (3) unable to determine fetal health.

After excluding these entries, 20,484 data entries remained. 2,790 were categorized as abnormal fetal health, and 17,694 as normal fetal health. The proportion of abnormal values was 0.136, as shown in Fig. 4. Due to the lack of proficiency in using the fetal heart rate monitor for pregnant women during self-measurements, the measured FHR and UCP data had considerable noise, discontinuity, and jitter. After consulting with medical professionals and referring to the approach taken by Liang [5], the following methods were applied to preprocess the FHR and UCP data:

- Data points with FHR values less than 50 or greater than 200 were removed. Corresponding UCP data points were also removed to ensure consistency in data length between FHR and UCP.
- A sliding window of size 8 was used for smoothing. As shown in (1), this window smoothed FHR and UCP.

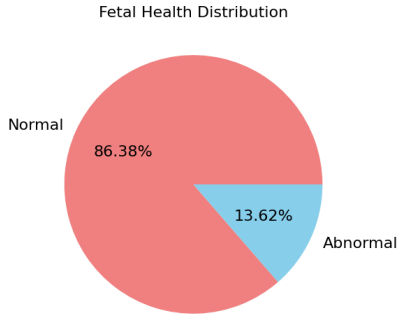

Fig. 4. Fetal health distribution.

Assume we have a time series x_t, where t represents the time step. The formula for calculating the moving average [23] is given by:

$$MA_t = \frac{1}{k}\sum_{i=0}^{k-1} x_{t-i} \tag{1}$$

here, MA_t represents the moving average at time step t. k is the moving window size, indicating the number of time steps considered. A comparison of the original and processed FHR and UCP data is shown in Fig. 5 and Fig. 6.

After the smoothing process, we followed the cutting method proposed by Cui et al. [24] to ensure consistent input data length. Because doctors require a minimum of 20 minutes for assessments, equivalent to 4,800 sampling points, we align our input with the doctors' needs and segment the duration into 20-minute intervals as input for the model. The segmentation method is as follows.

For a time series [23], the length is n:

$$T = \{t_1, t_2, ..., t_n\} \tag{2}$$

A slicing of the time series can be defined as:

$$S_{i:j} = \{s_i, s_{i+1}, \ldots, s_j\}, \quad 1 \le i, j \le n \tag{3}$$

The slicing set can be defined as:

$$\text{slicings}(T, m, l) = \{S_{km:km+l}\} \cup \{S_{n-l:n}\}, \quad 0 \le k \le \left\lfloor \frac{n-l}{m} \right\rfloor \tag{4}$$

here, l represents the slicing length, and m represents the stride.

Each record was divided into segments of length 4,800 (20 minutes), with an overlap of 1,200 sampling points for data augmentation. Thus, the slicings of FHR and UCP are $\text{slicings}(T, 3600, 4800)$. The labels of the segmented data remained the same as the original data. After segmentation, our dataset increased from 20,484 to 29,045. Among them, there were 5,400 abnormal cases and 23,654 normal cases. To balance the dataset and prevent the model from overfitting to normal data [18], we randomly selected 5,400 entries from the normal cases. We combined them with the abnormal

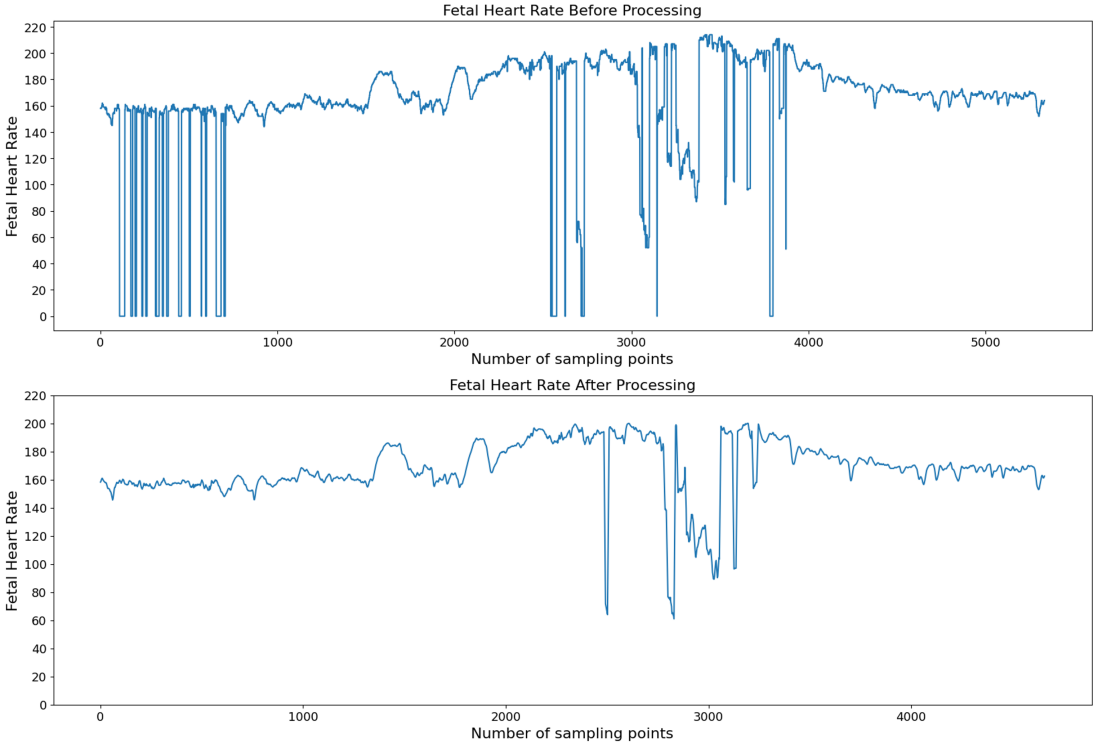

Fig. 5. FHR contrast before and after processing.

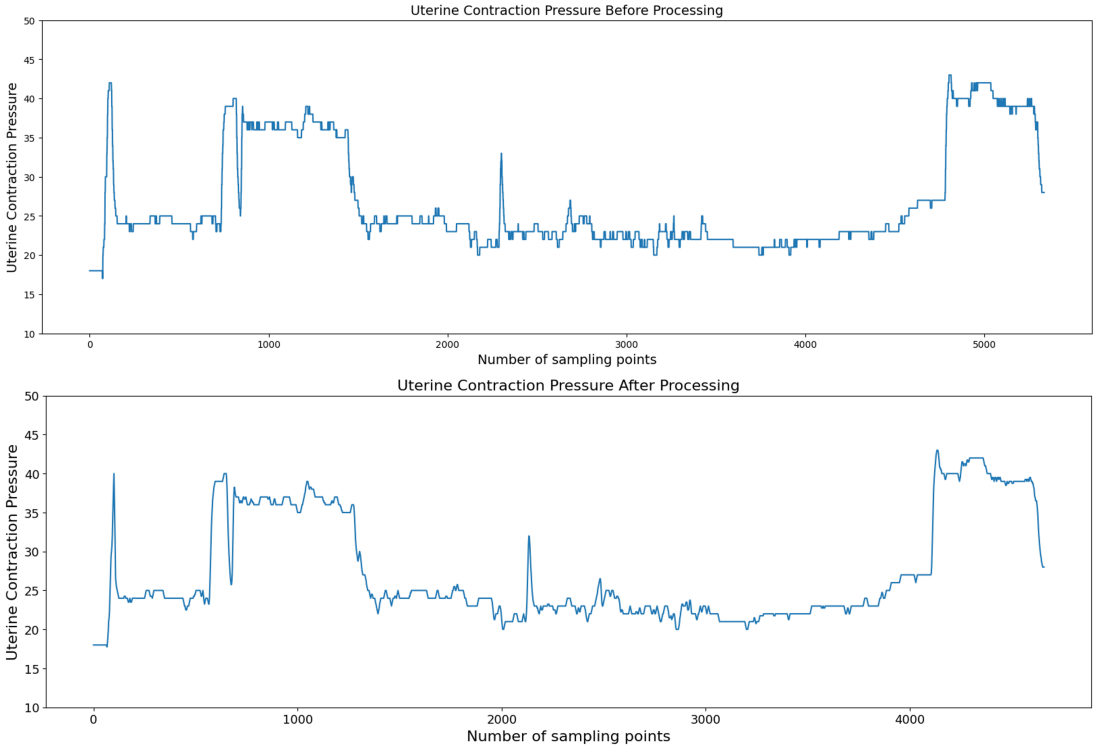

Fig. 6. UCP contrast before and after processing.

cases, resulting in our final REFM dataset containing an equal number of normal and abnormal data entries.

C. Multi-scale Feature Extractor

When doctors diagnose a fetus's health condition based on CTG, they rely on features such as fetal heart rate accelerations, decelerations, baseline, and variability. These features encompass overall characteristics, such as the overall baseline of the fetal heart rate, and local characteristics, such as accelerations and decelerations [25]. They also include smaller-scale features, such as baseline variability. However, previous models used for this task have not been able to extract features at different scales effectively. To address this issue, we propose NeuroFetalNet, as illustrated in Fig. 7. After the FHR and UCP are inputted, they undergo multi-scale feature extraction. Each feature extractor outputs features of 128 dimensions, which are then fused to obtain features of 512 dimensions. Finally, average pooling and fully connected layers are applied for the final output.

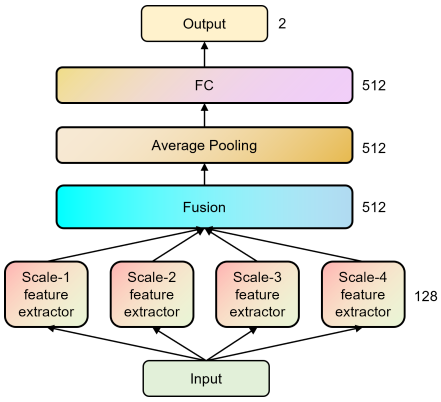

Fig. 7. Structure of NeuroFetalNet model.

NeuroFetalNet is a convolutional neural network (CNN) architecture. The model incorporates multi-scale residual blocks and channel attention mechanisms to enhance feature extraction and representation capabilities. The core component of NeuroFetalNet comprises four parallel ResNet modules, each utilizing different kernel sizes (3, 5, 7, and 9). This multi-scale approach allows the model to capture features at varying resolutions, thereby improving its ability to recognize intricate patterns in the input data. Each ResBlock within the ResNet modules consists of two convolutional layers, followed by Batch Normalization and ReLU activation functions. A shortcut connection facilitates gradient flow and mitigates the vanishing gradient problem.

As shown in Fig. 8, given an input tensor x with dimensions (C, L), where C is the number of channels, and L is the length of the sequence, the output y of a ResBlock is:

$$y = \text{ReLU}(\text{BN}(\text{Conv1d}(x))) \tag{5}$$

$$y = \text{BN}(\text{Conv1d}(y)) \tag{6}$$

$$y = \text{ReLU}(y + \text{Shortcut}(x)) \tag{7}$$

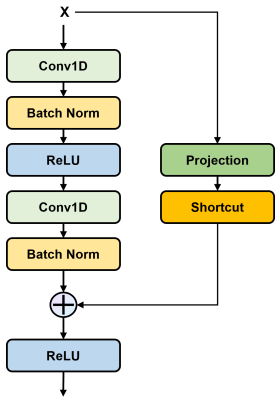

Fig. 8. Structure of residual block.

The shortcut connection can either be an identity mapping or a projection using 1×1 convolutions when the input and output dimensions differ. Each ResNet module consists of multiple ResBlocks followed by average pooling layers to reduce the dimensionality of the feature maps. This hierarchical structure produces an output tensor with reduced dimensionality, facilitating efficient feature extraction.

The NeuroFetalNet employs an Average Channel Attention (ACA) Block to enhance the extracted features' discriminative power. This mechanism calculates a weighted sum of the feature maps, enabling the model to focus on the most relevant channels. The ACA mechanism computes channel-wise attention weights using global average pooling followed by a two-layer fully connected network and a sigmoid activation:

$$w = \sigma(\text{FC}(\text{ReLU}(\text{FC}(\text{AvgPool1d}(x))))) \tag{8}$$

The input tensor x is then multiplied by these attention weights to yield the refined feature maps:

$$y = x \odot w \tag{9}$$

where $\odot$ denotes element-wise multiplication.

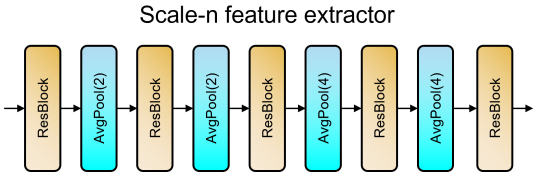

Fig. 9. Structure of each multi-scale feature extractor.

The NeuroFetalNet combines a multi-scale feature extractor with channel attention mechanisms. This architecture leverages residual connections and attention mechanisms to capture complex patterns and dependencies in the input data, enhancing its ability to perform accurate and robust classifications. To explore whether better model performance can be achieved by inputting only FHR, only UCP, or both FHR and UCP, we implemented all three variations and compared their performances to determine the most optimal approach.

TABLE I
COMPARISON OF METRICS FOR EACH MODEL ON THE REFM DATASET.

Model	Feature	Accuracy	Precision	Recall	F1-score
CNN	FHR	**83.17%**	**83.45%**	**84.83%**	**84.13%**
	UCP	55.96%	53.89%	70.59%	61.12%
	FHR + UCP	80.58%	79.65%	81.21%	80.43%
CNN + BiGRU	FHR	**87.12%**	**87.95%**	86.96%	**87.45%**
	UCP	53.46%	53.74%	75.79%	62.88%
	FHR + UCP	84.04%	80.62%	**88.02%**	84.16%
ResNet	FHR	**84.90%**	**87.79%**	82.82%	**85.23%**
	UCP	50.67%	52.86%	57.40%	55.04%
	FHR + UCP	82.88%	81.84%	**83.76%**	82.79%
ResNet + BiGRU	FHR	**89.81%**	**90.22%**	**89.17%**	**89.69%**
	UCP	50.38%	50.27%	67.79%	57.73%
	FHR + UCP	88.85%	89.45%	88.08%	88.76%
NeuroFetalNet	FHR	93.17%	92.67%	93.90%	93.28%
	UCP	61.63%	60.00%	72.00%	65.45%
	FHR + UCP	**94.23%**	**94.38%**	**94.02%**	**94.20%**

TABLE II
COMPARISON OF METRICS FOR ABLATION EXPERIMENTS ON THE REFM DATASET.

Model	Feature	Accuracy	Precision	Recall	F1-score
M1	FHR	84.90%	87.79%	82.82%	85.23%
	UCP	50.67%	52.86%	57.40%	55.04%
	FHR + UCP	82.88%	81.84%	83.76%	82.79%
M2	FHR	91.15%	92.02%	89.86%	90.93%
	UCP	52.21%	51.82%	44.44%	47.85%
	FHR + UCP	87.12%	86.73%	88.66%	87.68%
M3	FHR	92.69%	92.12%	93.52%	92.82%
	UCP	50.77%	50.83%	76.00%	60.92%
	FHR + UCP	93.08%	92.47%	93.88%	93.17%
M4	FHR	93.17%	93.24%	**94.08%**	93.66%
	UCP	53.37%	55.81%	62.12%	58.79%
	FHR + UCP	92.40%	90.18%	93.20%	91.67%
NeuroFetalNet	FHR	93.17%	92.67%	93.90%	93.28%
	UCP	61.63%	60.00%	72.00%	65.45%
	FHR + UCP	**94.23%**	**94.38%**	94.02%	**94.20%**

IV. EXPERIMENT

A. Dataset

The dataset used in our study is the REFM dataset, which we constructed ourselves. It consists of 10,800 data entries, 5,400 labeled as positive (indicating abnormal fetal health) and 5,400 labeled as negative (indicating normal fetal health). We use 1 to represent positive (abnormal) and 0 to represent negative (normal). The dataset includes two features, FHR and UCP, which have a length of 4,800.

B. Experiment Settings

The experimental environment settings are as follows: the GPU used is an NVIDIA GeForce RTX 4090, with CUDA version 12.2, Python version 3.8.17, and PyTorch version 2.0.1. The parameter configurations for the experiment include the number of training epochs set to 100, a batch size of 16, and an early stopping setting of 15. The split ratio for the training, validation, and testing sets is 0.8:0.1:0.1. The optimizer used is the Adam optimizer, with an initial learning rate of 0.001. The learning rate schedule follows a cosine decay, as shown in (10):

$$new\_lr = lr \times 0.5 \times (1 + cos(\frac{epoch + 1}{num\_epochs} \times \pi)) \quad (10)$$

here, new_lr represents the new learning rate, lr represents the original learning rate, $epoch$ represents the current epoch number, and num_epochs represents the total number of epochs.

The experimental results will be evaluated using the following metrics: Accuracy, Precision, Recall, and F1-score. The prediction results can be divided into four categories: true negatives (TN), false negatives(FN), true positives (TP), and false positives (FP).

Accuracy is the ratio of correctly classified samples to the total number of samples:

$$\text{Accuracy} = \frac{TP + TN}{TN + FN + TP + FP} \quad (11)$$

Precision focuses on positive prediction accuracy:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (12)$$

Recall focuses on actual positives captured:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (13)$$

F1-score focuses on the balance between Precision and Recall:

$$\text{F1-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (14)$$

C. Results and Discussion

Table I shows the results obtained for different models and feature inputs in the training set of 8,640 instances and the testing set of 1,080. The compared models include CNN, CNN+BiGRU, ResNet, ResNet+BiGRU, and our NeuroFetalNet. The input feature conditions include only FHR, only UCP, and both FHR and UCP. In Table I, the **bold font** indicates the best performance for a specific metric within the same model, while the underline represents the best performance across all models. From the Table I, the following conclusions can be drawn:

1) The model's prediction performance is the worst when only UCP is used as the input feature.
2) Except for NeuroFetalNet, using only FHR as the input feature produces better prediction performance after training in most cases.
3) NeuroFetalNet outperforms the SOTA models in all metrics, achieving the best performance.

The bad performance when UCP is used as the sole input feature can be attributed to its high noise and inaccurate measurement due to the difficulty in accurately measuring UCP at home using a fetal heart rate monitor, as advised by medical professionals. When the model lacks robustness, introducing excessive noise from UCP as an input feature leads to decreased performance. NeuroFetalNet extracts features at different scales through the multi-scale feature extractor and fusion network. It is robust and can extract the available information from UCP, which has fewer usable features than FHR. Therefore, NeuroFetalNet performs best when FHR and UCP are input features.

To validate the effectiveness of our method, we conducted an ablation study. As shown in Table II, M1, M2, M3, and M4 correspond to models using only feature extractors with kernel sizes of 3, 5, 7, and 9, respectively. In the ablation study, NeuroFetalNet still achieved the best Accuracy, Precision, and F1-score performance, with Recall performance being consistent with the best performers.

V. Conclusion

In this study, we constructed a new dataset specifically for remote electronic fetal monitoring (REFM) and performed the necessary preprocessing steps on the dataset. We have also thoroughly compared the model performance when using only FHR, only UCP, or both FHR and UCP as input features. Our results demonstrate that, in situations where the model lacks strong robustness, using only FHR as an input feature yields better performance. However, our proposed NeuroFetalNet effectively extracts useful information from UCP and performs best when both FHR and UCP are input features. Furthermore, we have introduced NeuroFetalNet, which utilizes a multi-scale feature extractor. Through comparative experiments and ablation studies, we have demonstrated the superiority of our model, which outperforms the current state-of-the-art models.

Acknowledgment

This research was supported by the National Natural Science Foundation of China under grant numbers 62176164 and 62203134. The authors would like to express their sincere gratitude for the support provided by the College Students' Innovative Entrepreneurial Training Plan Program, with project number IETP2023S0006.

References

[1] D. Hutter, E. Jaeggi *et al.*, "Causes and mechanisms of intrauterine hypoxia and its impact on the fetal cardiovascular system: a review," *International journal of pediatrics*, vol. 2010, 2010.

[2] A. Pinas and E. Chandraharan, "Continuous cardiotocography during labour: Analysis, classification and management," *Best practice & research Clinical obstetrics & gynaecology*, vol. 30, pp. 33–47, 2016.

[3] G. Georgoulas, D. Stylios, and P. Groumpos, "Predicting the risk of metabolic acidosis for newborns based on fetal heart rate signal classification using support vector machines," *IEEE Transactions on biomedical engineering*, vol. 53, no. 5, pp. 875–884, 2006.

[4] A. Fanelli, G. Magenes, M. Campanile, and M. G. Signorini, "Quantitative assessment of fetal well-being through ctg recordings: a new parameter based on phase-rectified signal average," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 5, pp. 959–966, 2013.

[5] H. Liang and Y. Lu, "A cnn-rnn unified framework for intrapartum cardiotocograph classification," *Computer Methods and Programs in Biomedicine*, vol. 229, p. 107300, 2023.

[6] X. Liu, L. Faes, A. U. Kale, S. K. Wagner, D. J. Fu, A. Bruynseels, T. Mahendiran, G. Moraes, M. Shamdas, C. Kern *et al.*, "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis," *The lancet digital health*, vol. 1, no. 6, pp. e271–e297, 2019.

[7] V. Subha, D. Murugan, J. Rani, K. Rajalakshmi, and T. Tirunelveli, "Comparative analysis of classification techniques using cardiotocography dataset," *International Journal of Research in Information Technology*, vol. 1, no. 12, pp. 274–280, 2013.

[8] S. Romagnoli, A. Sbrollini, L. Burattini, I. Marcantoni, M. Morettini, and L. Burattini, "Annotation dataset of the cardiotocographic recordings constituting the "ctu-chb intra-partum ctg database"," *Data in brief*, vol. 31, p. 105690, 2020.

[9] S. Dash, J. G. Quirk, and P. M. Djurić, "Fetal heart rate classification using generative models," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 11, pp. 2796–2805, 2014.

[10] S.-Y. Tsui, C.-S. Liu, and C.-W. Lin, "Modified maternal ecg cancellation for portable fetal heart rate monitor," *Biomedical Signal Processing and Control*, vol. 32, pp. 76–81, 2017.

[11] J. Spilka, J. Frecon, R. Leonarduzzi, N. Pustelnik, P. Abry, and M. Doret, "Sparse support vector machine for intrapartum fetal heart rate classification," *IEEE journal of biomedical and health informatics*, vol. 21, no. 3, pp. 664–671, 2016.

[12] P. Fergus, A. Hussain, D. Al-Jumeily, D.-S. Huang, and N. Bouguila, "Classification of caesarean section and normal vaginal deliveries using foetal heart rate signals and advanced machine learning algorithms," *Biomedical engineering online*, vol. 16, no. 1, pp. 1–26, 2017.

[13] A. Subasi, B. Kadasa, and E. Kremic, "Classification of the cardiotocogram data for anticipation of fetal risks using bagging ensemble classifier," *Procedia Computer Science*, vol. 168, pp. 34–39, 2020.

[14] S. Das, H. Mukherjee, S. M. Obaidullah, K. Santosh, K. Roy, and C. K. Saha, "Recurrent neural network based classification of fetal heart rate using cardiotocograph," in *Recent Trends in Image Processing and Pattern Recognition: Second International Conference, RTIP2R 2018, Solapur, India, December 21–22, 2018, Revised Selected Papers, Part II 2*. Springer, 2019, pp. 226–234.

[15] P. Hewage, A. Behera, M. Trovati, E. Pereira, M. Ghahremani, F. Palmieri, and Y. Liu, "Temporal convolutional neural (tcn) network for an effective weather forecasting using time-series data from the local weather station," *Soft Computing*, vol. 24, pp. 16 453–16 482, 2020.

[16] I. E. Livieris, E. Pintelas, and P. Pintelas, "A cnn–lstm model for gold price time-series forecasting," *Neural computing and applications*, vol. 32, pp. 17 351–17 360, 2020.

[17] Z. Zhao, Y. Deng, Y. Zhang, Y. Zhang, X. Zhang, and L. Shao, "Deepfhr: intelligent prediction of fetal acidemia using fetal heart rate signals based on convolutional neural network," *BMC medical informatics and decision making*, vol. 19, pp. 1–15, 2019.

[18] J. Li, Z.-Z. Chen, L. Huang, M. Fang, B. Li, X. Fu, H. Wang, and Q. Zhao, "Automatic classification of fetal heart rate based on convolutional neural network," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1394–1401, 2018.

[19] Q.-J. Lv, H.-Y. Chen, W.-B. Zhong, Y.-Y. Wang, J.-Y. Song, S.-D. Guo, L.-X. Qi, and C. Y.-C. Chen, "A multi-task group bi-lstm networks application on electrocardiogram classification," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 8, pp. 1–11, 2019.

[20] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (gru) neural networks," in *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*. IEEE, 2017, pp. 1597–1600.

[21] R. Fu, Z. Zhang, and L. Li, "Using lstm and gru neural network methods for traffic flow prediction," in *2016 31st Youth academic annual conference of Chinese association of automation (YAC)*. IEEE, 2016, pp. 324–328.

[22] A. Shewalkar, D. Nyavanandi, and S. A. Ludwig, "Performance evaluation of deep neural networks applied to speech recognition: Rnn, lstm and gru," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 9, no. 4, pp. 235–245, 2019.

[23] S. Hansun, "A new approach of moving average method in time series analysis," in *2013 conference on new media studies (CoNMedia)*. IEEE, 2013, pp. 1–4.

[24] Z. Cui, W. Chen, and Y. Chen, "Multi-scale convolutional neural networks for time series classification," *arXiv preprint arXiv:1603.06995*, 2016.

[25] V. Sinai Talaulikar and S. Arulkumaran, "Medico-legal issues with ctg interpretation," *Current Women's Health Reviews*, vol. 9, no. 3, pp. 145–157, 2013.