

RESEARCH

Open Access



Hybrid-FHR: a multi-modal AI approach for automated fetal acidosis diagnosis

Zhidong Zhao^{1*}, Jiawei Zhu², Pengfei Jiao¹, Jinpeng Wang¹, Xiaohong Zhang², Xinmiao Lu² and Yefei Zhang¹

Abstract

Background In clinical medicine, fetal heart rate (FHR) monitoring using cardiotocography (CTG) is one of the most commonly used methods for assessing fetal acidosis. However, as the visual interpretation of CTG depends on the subjective judgment of the clinician, this has led to high inter-observer and intra-observer variability, making it necessary to introduce automated diagnostic techniques.

Methods In this study, we propose a computer-aided diagnostic algorithm (Hybrid-FHR) for fetal acidosis to assist physicians in making objective decisions and taking timely interventions. Hybrid-FHR uses multi-modal features, including one-dimensional FHR signals and three types of expert features designed based on prior knowledge (morphological time domain, frequency domain, and nonlinear). To extract the spatiotemporal feature representation of one-dimensional FHR signals, we designed a multi-scale squeeze and excitation temporal convolutional network (SE-TCN) backbone model based on dilated causal convolution, which can effectively capture the long-term dependence of FHR signals by expanding the receptive field of each layer's convolution kernel while maintaining a relatively small parameter size. In addition, we proposed a cross-modal feature fusion (CMFF) method that uses multi-head attention mechanisms to explore the relationships between different modalities, obtaining more informative feature representations and improving diagnostic accuracy.

Results Our ablation experiments show that the Hybrid-FHR outperforms traditional previous methods, with average accuracy, specificity, sensitivity, precision, and F1 score of 96.8, 97.5, 96, 97.5, and 96.7%, respectively.

Conclusions Our algorithm enables automated CTG analysis, assisting healthcare professionals in the early identification of fetal acidosis and the prompt implementation of interventions.

Keywords Fetal heart rate, Fetal acidosis, Cardiotocography, Cross-modal feature fusion, Multi-modal, Temporal convolutional network, Attention mechanisms

Background

Fetal acidosis is an imbalance in the acid-base balance of the fetus's body that causes the fetus's blood to become too acidic [1]. Fetal acidosis caused by hypoxia can lead

to multiple organ damage, and even death. Therefore, we need a safe and effective method for early detection of fetal acidosis to assist obstetricians in determining whether intervention measures during childbirth are required.

Cardiotocography (CTG), also known as electronic fetal monitoring (EFM), is a common monitoring technique wherein clinicians assess the fetal health by analyzing signals related to the Fetal Heart Rate (FHR) and uterine contractions (UC) obtained from CTG. While CTG has become the most widely employed fetal monitoring method [2], its utility remains a subject of debate

*Correspondence:

Zhidong Zhao
zhaozd@hdu.edu.cn

¹ School of Cyberspace, Hangzhou Dianzi University, Hangzhou, China

² College of Electronics and Information Engineering, Hangzhou Dianzi University, Hangzhou, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

due to high interobserver (different specialists at the same time) and intraobserver (same specialist at different times) variability. Furthermore, CTG may lead to an increase in false positives and a higher rate of planned deliveries [3, 4]. Consequently, there is an urgent need to develop an automated diagnostic technique to address these limitations.

Previously, researchers employed morphological time domain, frequency domain, and nonlinear domain parameters of FHR signals for feature extraction, feature selection, and classification. Georgieva et al. [5] extracted 12 clinical parameter features, and researchers obtained a sensitivity of 60.3% and a specificity of 67.5% using a feedforward artificial neural network (ANN). Spilka et al. [6, 7] extracted a total of more than 50 features including the above three domain features, and used the Adaptive Boosting (AdaBoost) classifier and Random Forest classifier, respectively. Cömert et al. [8] used the short-time Fourier transform (STFT) and gray level co-occurrence matrix (GLCM) to extract the image-based time-frequency features (IBTF) from the FHR signal. Zhao et al. [9] extracted 47 expert features from FHR signals and utilized statistical testing (ST) and PCA for dimensionality reduction. Pini et al. [10] extracted 23 expert features and applied the recursive feature elimination (RFE) method to select the most relevant subset of features. These methods rely on expert features. Although they are highly reliable and interpretable, feature extraction can be complex and limited by the quality of the signal and domain-specific knowledge.

In the past decade, with the development of deep learning (DL), numerous studies demonstrated that deep neural networks have a wide range of applications in healthcare [11, 12]. Compared to traditional machine learning (ML) methods, these algorithms can learn important features automatically from the original input signal. This self-learning ability allows them to discover complex patterns in time series signals without the need for human feature engineering. Bursa et al. [13] and Cömert et al. [14] conducted research on two-dimensional convolutional neural network (2D-CNN) models. Bursa et al. utilized Continuous Wavelet Transform (CWT) on 1-dimensional fetal heart rate signals and contraction signals, and authors achieved a high classification accuracy of 94.1%. Cömert et al. used the STFT with transfer learning to analyze FHR signals. Li et al. [15] used a one-dimensional convolutional neural network (1D-CNN) and compared it with traditional feature extraction methods, demonstrating that 1D-CNN outperforms traditional methods. Liang et al. [16] proposed a one-dimensional convolutional neural network - gated recurrent unit (1D-CNN-GRU) model, and authors obtained an accuracy of 95.15%. Fei et al. [17] integrated

three signals - FHR, UC, and fetal movement (FetMov) - by using an embedding layer to combine the features at the input level. Spairani et al. [18] proposed a hybrid method based on neural structures, where they converted FHR signals into the image domain, and researchers then parallelly input a set of expert features and finally perform decision fusion at the classification level.

However, most existing studies in FHR signal analysis are based on a single modality feature, which may not provide sufficient information to fully describe and analyze complex FHR signals. Moreover, FHR signals are often subject to various types of noise and interference, making single-modal features less stable and reliable. In contrast, multimodal features can capture a richer representation of potential features, and different modalities may have varying importance in different scenarios. By fusing multimodal features, the weights of each modality can be learned adaptively, thereby improving the accuracy of diagnosis.

Based on the analysis presented, we propose a novel framework called Hybrid-FHR to diagnose fetal acidosis, assist doctors in identifying pathological fetuses, and reduce the rate of stillbirths. The algorithm utilizes multimodal features and combines the advantages of deep learning with expert prior knowledge. The overall framework of the Hybrid-FHR algorithm is depicted in Fig. 1.

The contributions and innovations of this study are listed as follows:

1. Our proposed fetal acidosis diagnostic framework (Hybrid-FHR) incorporates multimodal features and effectively leverages the information provided by various features. Through our experiments, we have demonstrated that our approach achieves significant performance gains in the diagnosis of fetal acidosis.
2. We designed a lightweight backbone network SE-TCN for extracting spatio-temporal representations of FHR signals, which utilizes dilated casual convolutions to effectively enhance the global perception capability of the entire network. Furthermore, a cross-modal feature fusion (CMFF) method based on multi-head attention mechanism is proposed to achieve optimal weighted fusion of different modalities.
3. We designed three types of expert features (morphological time domain, frequency domain, and nonlinear) by incorporating expert prior knowledge, which further improved the performance of the model.

Methods

In this section, we first introduce the three types of expert features based on prior knowledge. Next, we elaborate on the SE-TCN backbone network for extracting

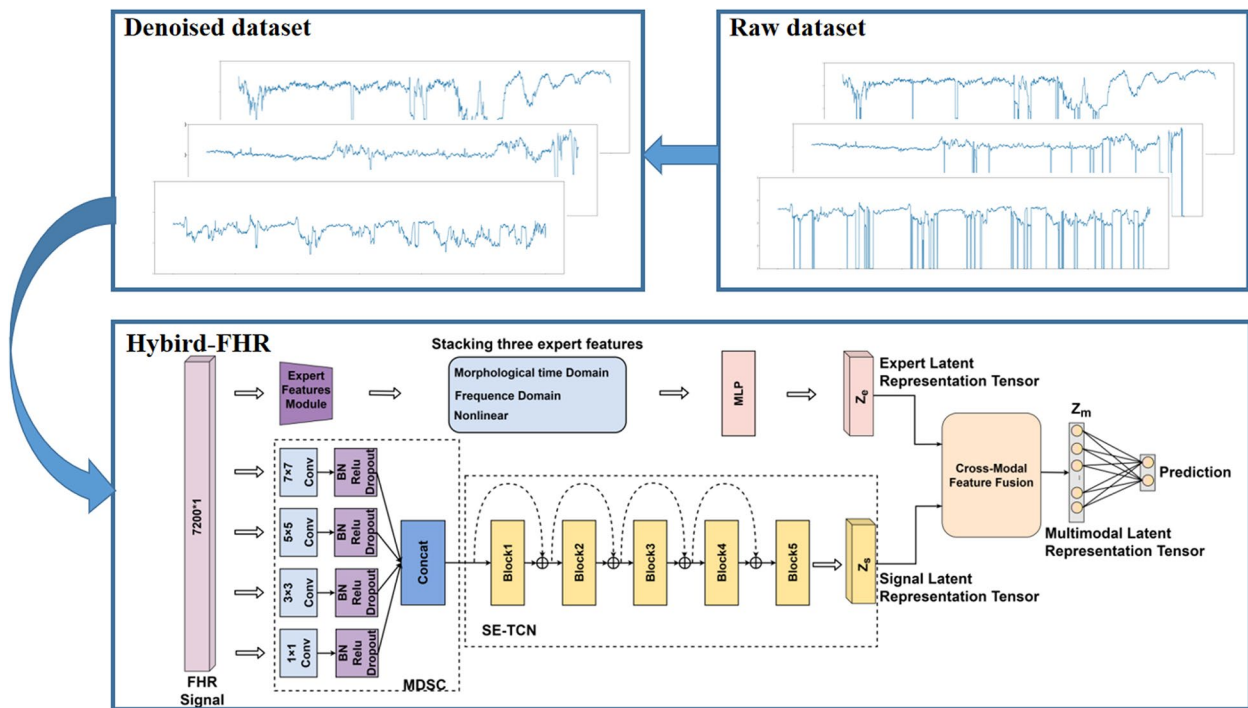


Fig. 1 Overview of proposed method (Hybrid-FHR)

features from one-dimensional FHR signals, and finally, we introduce the cross-modal feature fusion (CMFF) approach, which uses a multi-head attention mechanism to adaptively weight different modal features.

Expert features module

Based on expert prior knowledge, we carefully designed a set of 45 features from the pre-processed FHR signals, including 21 morphological time domain, 14 frequency domain, and 10 nonlinear features. The specific formulas and details of these parameters can be found in (Additional files 1, 2, 3). These 45 features were processed through two layers of linear projection to obtain the expert latent representation tensor, denoted as Z_e .

A. Morphological time domain

In this study, we calculated several morphological time domain characteristics following the International Federation of Gynecology and Obstetrics (FIGO) guidelines [19], including baseline (BL), number of accelerations (nACC), and number of decelerations (nDEC).

Time domain characteristics are mainly derived from the fetal heart rate variability (FHRV), which is the variability of the heartbeat cycle variation. To analyze the HRV, we must convert the FHR to RR

(heartbeat-by-heartbeat) interval sequences with the following conversion equation:

$$RR = \frac{60000}{FHR} \tag{1}$$

The time difference between two consecutive RR intervals is called NN, which is calculated as follows:

$$NN = diff(RR) \tag{2}$$

In this study, we have referred to commonly used parameters for adult HRV and calculated various statistical measures to analyze the fetal heart rate variability signal in the time domain. These measures include basic parameters such as the maximum, minimum, mean, median, standard deviation, kurtosis, and skewness of the RR interval. Other parameters include the standard deviation of NN (SDNN), the root mean square of successive differences of RR intervals (RMSSD), NN50, and pNN50, which determine the number and percentage of NN that differ by more than 50 ms. Short-term variability and long-term variability (STV and LTV [20]), the triangular index (Tri [21]), and the triangular interpolation of the NN interval histogram (TINN [22]) were also calculated.

The morphological time domain features of the FHR are therefore as follows:

Morphological time domain: {mean_baseline, max_baseline, min_baseline, std_baseline, nACC, nDEC, max_rr, min_rr, mean_rr, median_rr, std_rr, skew_rr, kurt_rr, SDNN, RMSSD, NN50, pNN50, STV, LTV, Tri, TINN}.

B. Frequency domain

The spectral analysis of FHRV examines changes in the fetal autonomic nervous system (ANS) activity [23], which can be observed in the periodic changes in FHRV. We followed the suggestion in [24] to divide the frequency range into four bands: very low frequency (VLF, 0–0.03 Hz), low frequency (LF, 0.03–0.15 Hz), medium frequency (MF, 0.15–0.5 Hz), and high frequency (HF, 0.5–1 Hz).

We used the Fast Fourier Transform (FFT) to convert the signal into the frequency domain and divided it into four frequency bands. We extracted the power spectral density, power spectral ratio, peak frequency, and total power spectral density of each frequency band. We also calculated the LF/(MF + HF) energy ratio. Therefore, the frequency domain features of FHR are as follows:

Frequency Domain: {rr_VLF, rr_LF, rr_MF, rr_HF, rr_Total_Power, rr_percent_VLF, rr_percent_LF, rr_percent_MF, rr_percent_HF, rr_peak_VLF, rr_peak_LF, rr_peak_MF, rr_peak_HF, rr_ratio,}.

C. Nonlinear

In recent years, nonlinear measurements for studying FHR kinetics have become increasingly available and have shown promising results [25–27]. We perform nonlinear feature extraction using the NeuroKit2 library in Python. The nonlinear methods used in this study include Poincare plot parameters [28], approximate entropy (ApEn, [29]), sample entropy (SampEn, [29]), Shannon entropy (ShannEn, [30]), fuzzy entropy (FuzzyEn, [29]), Lempel-Ziv complexity index (LZC,

[31]), fractal dimension (FD, [32]), and Hurst index (Hurst, [33]), as follows:

Nonlinear: {SD1, SD2, SD_Ration, ApEn, SampEn, ShannEn, FuzzyEn, LZC, FD, Hurst}.

Where SD1 and SD2 represent the short-axis and long-axis deviations of the Poincare plot, respectively, and SD_Ratio represents the ratio of the two.

Signal backbone

This paper proposes a SE-TCN backbone network to extract latent feature representations of FHR signals. The network comprises a Multi-scale Depthwise Separable Convolution (MDSC) module and five SE-TCNBlocks. Table 1 presents the detailed hyperparameter settings and output dimensions of each layer of the proposed signal backbone.

A. MDSC

Assuming that $X_s \in \mathbb{R}^{B \times N \times C}$ is a whole representation of a set of continuous one-dimensional FHR signals, where subscript s is an abbreviation for signal, and \mathbb{R} denotes the real numbers set, B represents the batchsize, N represents the signal length, and C represents the number of channels.

Before the FHR signal passes through the SE-TCN-Blocks, we designed a MDSC module for capturing signal features at different scales. In MDSC, we adopt depthwise separable convolution (DSC [34]) to replace ordinary convolution. DSC decomposes the convolution operation into depthwise convolution and pointwise convolution. The former performs convolution only on each input channel, while the latter performs convolution on the output channels. Compared to ordinary convolution, DSC can effectively reduce the number of parameters and computation, thereby improving model efficiency.

MDSC combines multiple DSCs of different scales. Different-sized convolution kernels move along the one-dimensional direction to extract features from the entire signal, gradually obtaining features that can fully

Table 1 The detailed hyperparameter settings and output dimensions of each layer of the SE-TCN

Layer	Output shape	Kernel size	Stride	Padding	Dilation factor	Activation
Input	$B \times 7200 \times 1$	–	–	–	–	–
MDSC	$B \times 7200 \times 64$	[1, 3, 5, 7]	1	Same	1	ReLU
SE-TCNBlock1	$B \times 2400 \times 64$	15	3	Casual	2	ReLU
SE-TCNBlock2	$B \times 2400 \times 64$	15	1	Casual	4	ReLU
SE-TCNBlock3	$B \times 800 \times 128$	15	3	Casual	8	ReLU
SE-TCNBlock4	$B \times 800 \times 128$	15	1	Casual	16	ReLU
SE-TCNBlock5	$B \times 400 \times 256$	15	2	Casual	32	ReLU

B Batchsize

represent the sequence in a locally-aware manner. The four different channels of convolution kernels in MDSC have sizes of 1, 3, 5, and 7, with dilation factors of 1 and channel numbers of 16. Finally, by fusing the outputs of different convolution kernel channels, a tensor with a channel number of 64 is obtained.

B. SE-TCNBlock

The 1D convolution method is often used for feature extraction in time series data. For long-series problems such as FHR signals, the normal convolutional approach (dilation factors $d=1$) is prone to phenomena such as gradient disappearance, which is not satisfactory. To increase the long time dependence of the network and to improve its ability to reach into the past for prediction, a temporal convolutional network (TCN) was proposed [35, 36].

TCN combines causal with dilated convolution, and Fig. 2 depicts the dilated causal convolution with dilation factors $d=1, 2, 3$ and a convolution kernel size $k=3$. The output at a certain moment is only related to the current and the past moments, using a zero-padding approach with the number of paddings per layer equal to $d \times (k-1)$. Furthermore, the receptive field size (RFS) of the network increases exponentially with the number of layers. For a one-dimensional time series X and a convolution kernel w of size k , the dilated convolution can be expressed as follows:

$$Y(t) = (X *_d w)(t) = \sum_{i=0}^{k-1} w(i) \cdot X(t - d \cdot i) \quad (3)$$

Where $Y(t)$ represents the t -th element in the output sequence, $*_d$ denotes the convolution operator with

dilation factors d , and $w(i)$ is the weights of convolution kernel w .

As shown in Fig. 3, we use the residual connection [37] in SE-TCNBlock to effectively train the deep neural network, which alleviates the gradient disappearance problem to some extent. Each SE-TCNBlock contains two channels, where the main channel of the residual connection contains two dilated causal convolution layers, and each convolution layer is activated after using batch normalization [38] and a rectified linear unit (ReLU) [39]. The dropout rate is set to 0.1, the dilated convolution factor d in the SE-TCNBlock is equal to 2^L , where $L = (1, 2, 3, 4, 5)$, and the RFS of the network is exponentially related to the number of layers, which is computed as follows:

$$RFS = 1 + 2 \cdot (k - 1) \cdot \sum_{i=1}^5 d_i \quad (4)$$

Therefore, we enhance the RFS of the network by choosing a larger convolution kernel size k , increasing the dilation factor d or the number of the network layers L .

The sub-channel of the residual connection includes a downsampled convolutional layer with a convolutional kernel size of 1 (1×1 Conv) and an SEBlock.

SEBlock is a channel-wise attention mechanism module within SENet [40], that aims to capture the interdependencies of each channel in the feature map.

To capture dependencies between different lengths and time steps, this network employs varying dilation factors within each SE-TCNBlock. The blocks are hierarchically connected, with the output of each block

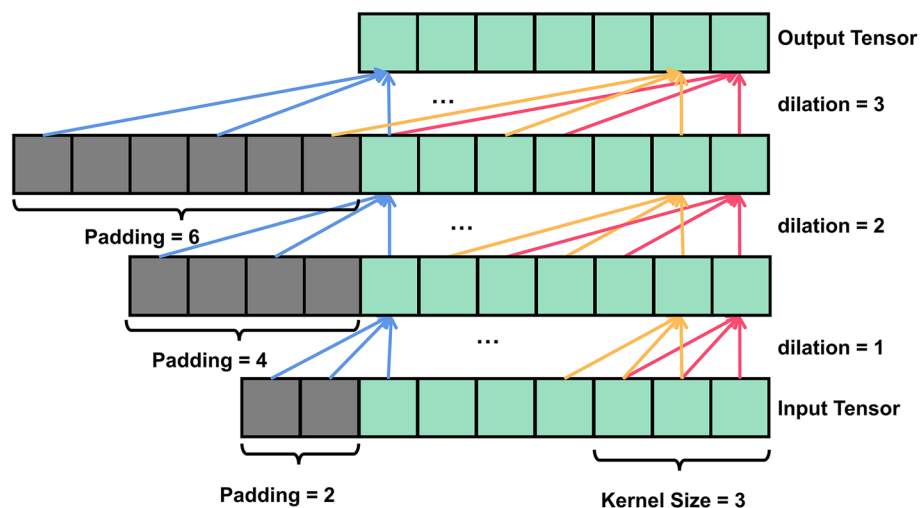


Fig. 2 A Dilated Causal Convolution with dilation factors $d=1, 2, 3$ and kernel size $k=3$

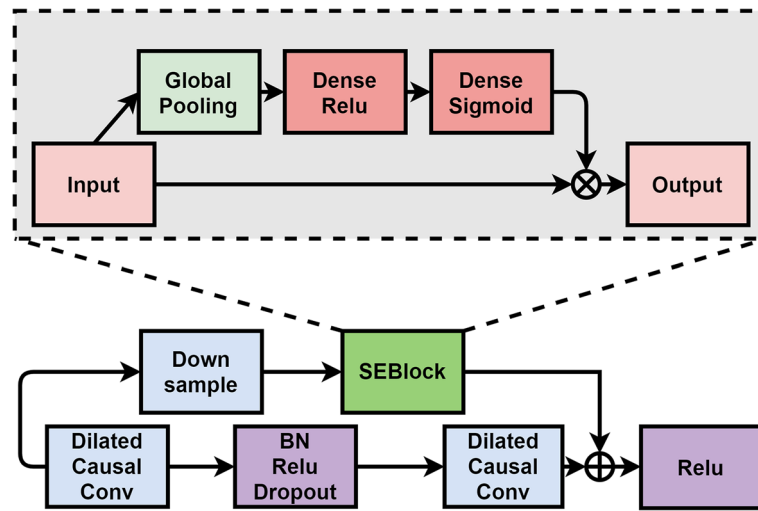


Fig. 3 The internal structure of SE-TCNBlock

feeding into the input of the next. The final output of the last SE-TCNBlock is the signal latent representation tensor denoted as Z_s , which represents the feature-extracted representation of the original signal.

Cross-modal feature fusion

The ordinary feature fusion approach can be divided into two types: early fusion and late fusion, depending on where the fusion occurs. Early Fusion or Feature-level Fusion, combines features from different modalities at the input level to obtain a richer representation. Late

In the multi-head attention mechanism, each representation tensor is linearly projected to a set of vectors with different semantics, denoted as $Q_i = Z_i * W^{Q_i}$, $K_i = Z_i * W^{K_i}$, and $V_i = Z_i * W^{V_i}$. where $i \in \{e, s\}$, W^{Q_i} , W^{K_i} , and W^{V_i} denote the query matrix, key matrix and value matrix respectively. Then, these vectors are divided into 8 attention heads, and each head performs self-attention calculation independently. The weight matrices of each head are then concatenated together. Finally, the output tensor of the multi-headed attention mechanism is computed as follows:

$$MultiHeadAttention(Q_i, K_i, V_i) = Concat(head_1^1, \dots, head_1^8) W^{O_i}, i \in \{e, s\} \tag{5}$$

Fusion or Decision-level Fusion, involves using different models to extract features from different modalities and then integrating the prediction results of these models at the decision level.

Both early fusion and late fusion have their advantages and limitations. Early fusion can provide a holistic representation of information from different modalities but may not effectively capture the relationships between features. Late fusion, on the other hand, can model the relationships between features more flexibly but may require more computational resources and time.

In the CMFF module presented in Fig. 4, a multi-head attention mechanism [41] is utilized to measure the similarity between the latent representation tensors of the signal (denoted as Z_s) and the expert (denoted as Z_e). The purpose of this module is to fuse the features from different modalities and capture the cross-modal interactions for improved performance in the given task.

$$head_i^n = softmax\left(\frac{Q_i^n (K_i^n)^T}{\sqrt{d_{model}}}\right) V_i^n, i \in \{e, s\}, d_{model} = 32 \tag{6}$$

$$softmax(x_i) = \frac{\exp(x_i)}{\sum_{j=1}^n \exp(x_j)} \tag{7}$$

where W^{O_i} denotes the output weight matrix. In $head_i^n$, the superscript n belongs to the set $\{1, 2, \dots, 8\}$ and indicates the number of attention heads, the subscript i belongs to the set $\{e, s\}$.

We denote the outputs of Z_e and Z_s after multi-headed self-attention (intra-modal) as Z_e' and Z_s' respectively. We then calculate the cosine similarity between Z_e' and Z_s' , and normalize them using the softmax function to obtain the cross-modal attention score (CMAS, inter-modal). Next, we weight the output of the multi-headed self-attention with

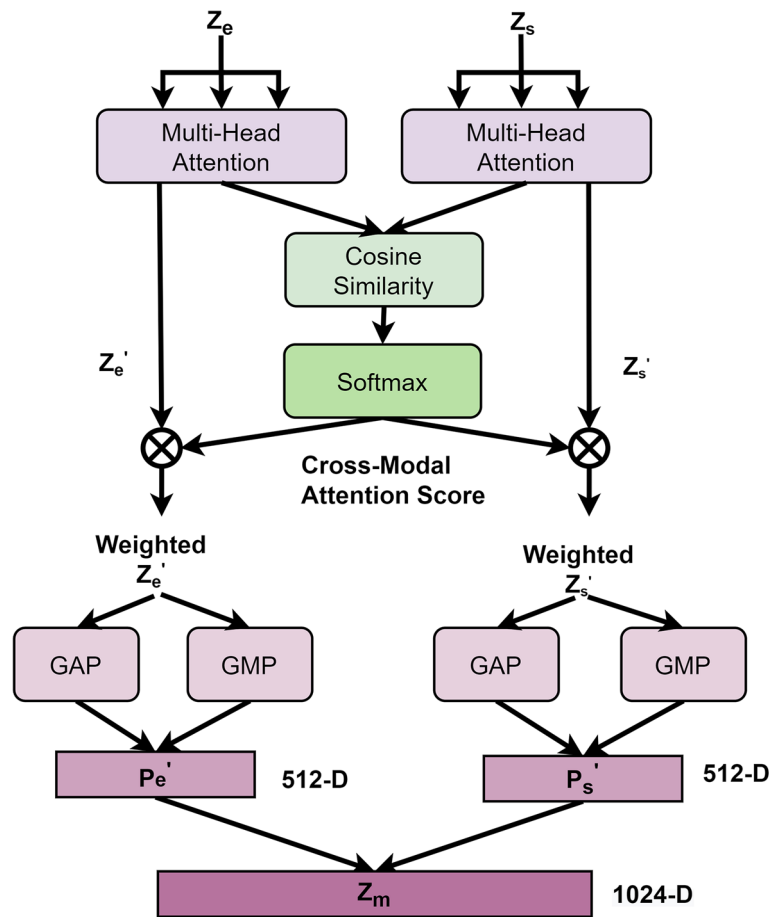


Fig. 4 The internal structure of CMFF

the CMAS to obtain a weighted representation denoted as $Weighted_Z_i'$, $i \in \{e, s\}$. We also apply global average pooling (GAP) and global maximum pooling (GMP) on the weighted representation $Weighted_Z_i'$, and concatenate the resulting vectors to obtain a 512-dimensional tensor denoted as $P_i \in \mathbb{R}^{B \times 512}$, $i \in \{e, s\}$. Finally, we concatenate P_e and P_s , resulting in a multimodal fusion latent representation tensor denoted as $Z_m \in \mathbb{R}^{B \times 1024}$. This fusion tensor contains information from both Z_e and Z_s , combined through the CMAS and the pooling operations, which can be further used for downstream tasks or analyses.

Z_m (m is an abbreviation for multimodal) is calculated as follows:

$$cosine_similarity(Z_e', Z_s') = \frac{Z_e' \cdot Z_s'}{\|Z_e'\| \|Z_s'\|} \quad (8)$$

$$CMAS = softmax(cosine_similarity(Z_e', Z_s')) \quad (9)$$

$$Weighted_Z_i' = CMAS \cdot Z_i', i \in \{e, s\} \quad (10)$$

$$P_i = concat(GAP(Weighted_Z_i'), GMP(Weighted_Z_i')), i \in \{e, s\} \quad (11)$$

$$Z_m = concat(P_e, P_s) \quad (12)$$

Overall, the CMFF module combines the strengths of different modalities and captures their complementary information, which can improve the performance of subsequent classification tasks (as discussed in Experiment three).

Results

In this section, we conducted three main experiments. Firstly, we performed hyperparameter analysis by tuning the hyperparameters of the model to study their impact on the experimental results. Secondly, we compared different signal backbone models

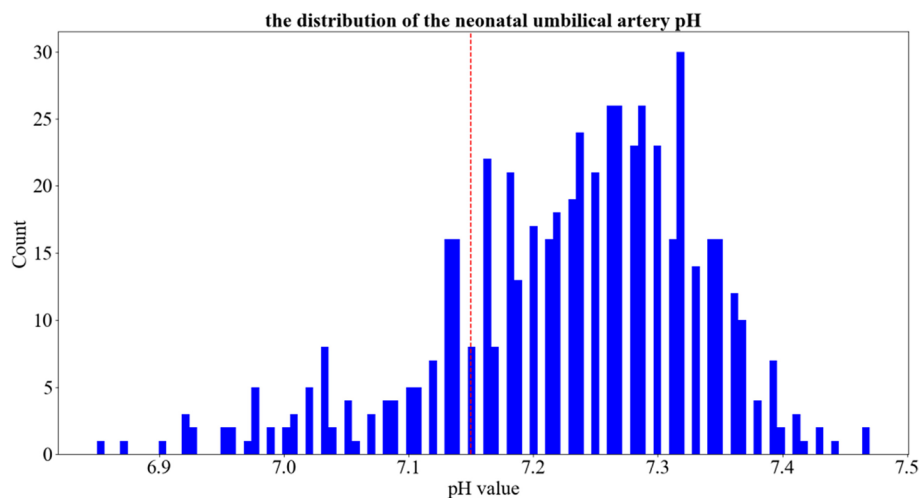


Fig. 5 The distribution of the neonatal umbilical artery pH

to investigate their performance differences in the cross-modal feature fusion task. Finally, we conducted ablation experiments by comparing the performance of single-modal and multi-modal inputs to validate the effectiveness of the CMFF method. The results confirmed that our proposed model achieved the best accuracy (96.8%).

Experimental setup

Dataset

The data used in this study were obtained from CTU-UHB [42, 43], a database of CTG recordings, containing a total of 552 samples with a sampling frequency of 4 Hz. Each CTG recording contains a set of FHR signals and a set of UC signals. In order to accurately assess intrauterine fetal acidosis, it is crucial to integrate these signals with clinical indicators. One such indicator is the neonatal umbilical artery pH, which serves as one of the gold standards for evaluating the presence of acidosis in the intrauterine environment. The lower the pH value, the more severe the fetal hypoxia. Different clinical doctors or research institutions may use different pH thresholds to determine whether the fetus is hypoxic, depending on their clinical experience and actual situation. We referred to the most commonly used criteria for delineation at this stage [8, 13, 14, 16, 26] and used 7.15 as a threshold value, with a pH value below 7.15 considered pathological and one greater than or equal to 7.15 considered normal, yielding a total of 447 normal and 105 pathological records. The distribution of pH values in the umbilical artery of newborns in the dataset is shown in Fig. 5.

Data preprocessing

Noises during recording may disrupt the FHR signal, compromising its quality and impacting diagnostic tasks. Additionally, the imbalance between positive and negative samples poses a challenge, requiring data augmentation to increase the number of pathological samples. To overcome the challenges mentioned above, we adopted the preprocessing and data augmentation methods previously proposed by our group [44, 45] to enhance the original signal, and the denoised signal is shown in Fig. 6. Firstly, to ensure high integrity and quality of the signals used, signals of effective lengths below 10,000 (severely incomplete) were discarded and a total of 524 samples (pathological: 95, normal: 429) were used. Secondly, noise disturbances such as missing values are removed using a mini-batch-based minimized sparse dictionary learning approach [43], and all 524 records had an effective length greater than or equal to 10,000. Thirdly, since fetal distress mainly occurs before delivery, we focused on the last 30 minutes of each sample in our experiments, which corresponds to a sample length of 7200 (4 Hz). Finally, the pathological FHR signals were synthesized using a Generative Adversarial Network (GAN) [45] to make the sample distribution balanced. GAN is used only for the training set, and the information in the test set is not used to synthesize data samples; therefore, the evaluation process is reliable and generalized.

Evaluation

We first randomly sampled 444 samples (pathological: 55, normal: 389) from the original 524 samples for training, and used 80 samples (pathological: 40, normal: 40)

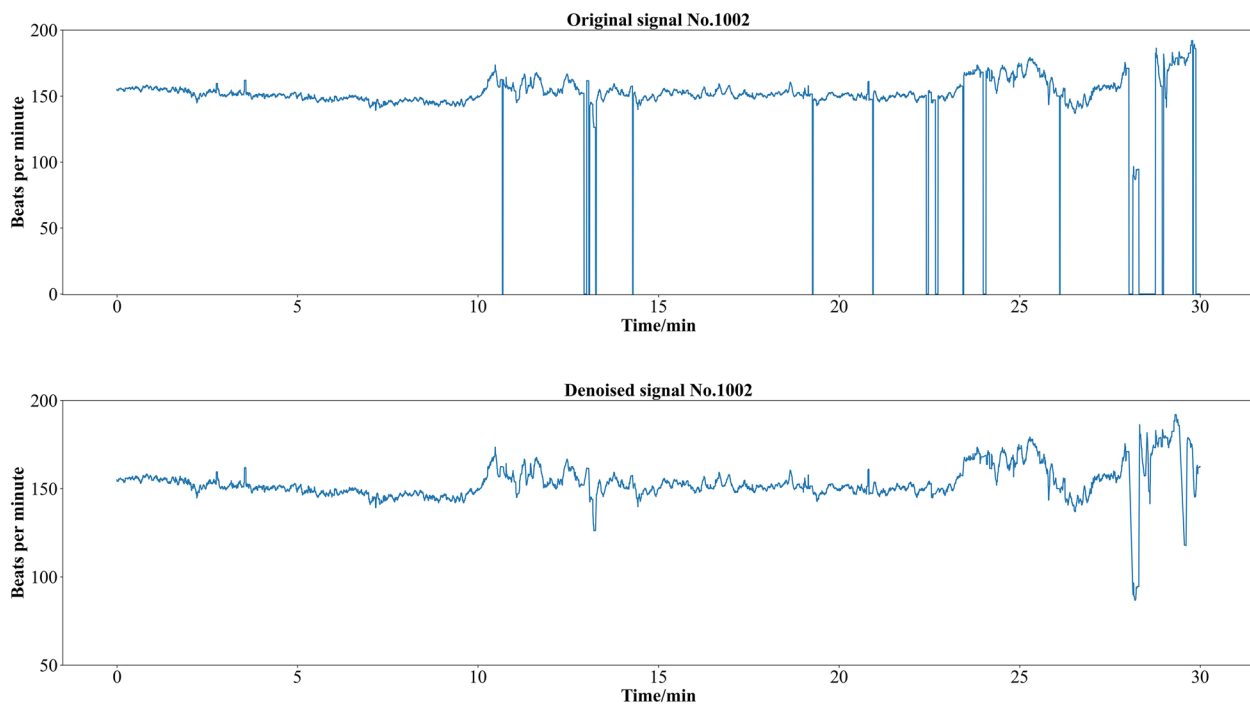


Fig. 6 Comparison of original signal (Top) and denoised (Bottom) signal. Outliers and missing values are removed from FHR signals using a mini-batch-based minimized sparse dictionary learning approach

for testing. Next, we used GAN for data augmentation on the training set to balance the positive and negative samples (pathological: 389, normal: 389). During the training process, we further divided the training set into training and validation sets using 5-fold cross-validation, and evaluated the model on the original test set. The average of the predictions from 5-fold cross-validation was used as the final prediction result. We calculated several metrics including accuracy (Acc), precision (Pre), sensitivity (Sen), specificity (Spe), and F1-Score (F1).

Experiment one: Hyperparameter optimization

To achieve optimal model performance, we conducted a thorough analysis of different hyperparameter settings and their impact on classification results. Our experimental findings revealed that the kernel size in the SE-TCNBlock and the number of heads in the multi-headed attention mechanism significantly influenced the classification performance, as illustrated in Fig. 7. The remaining hyperparameters were set to their default values, as follows: the cross-entropy loss function and the Adam optimizer [46] were utilized in the training process. The batch size was set to 16, and the training duration was configured for 120 epochs, with early stopping [47]. The learning rate strategy employed cosine annealing with an initial learning rate of $2.5e-4$ and a decay factor set to 0.8.

When the kernel size was increased from 3 to 15, a remarkable improvement in F1 scores was observed for both the validation set and the test set, indicating superior performance. However, when the kernel size was further increased to 19, a slight drop in the F1 score for the validation set and a more significant drop to 0.93 for the test set were observed, implying that larger kernel sizes may not always yield better results. Similarly, increasing num_heads from 4 to 8 resulted in a successive improvement in F1 scores for both the validation and test sets, suggesting that incorporating more attentional heads can enhance the model's performance. Nevertheless, when num_heads continued to increase to 16, a slight decrease in the F1 score for the validation set and a more substantial drop to 0.92 for the test set were observed. This suggests that excessively large num_heads may lead to over-complexity and overfitting, ultimately negatively impacting the model's performance.

In summary, the experimental results suggest that a moderate kernel size and num_heads may help to improve the performance of the model, but too large a kernel size and num_heads may have a negative impact on performance. Therefore, in this paper we set the kernel size and num_heads to 15 and 8 respectively.

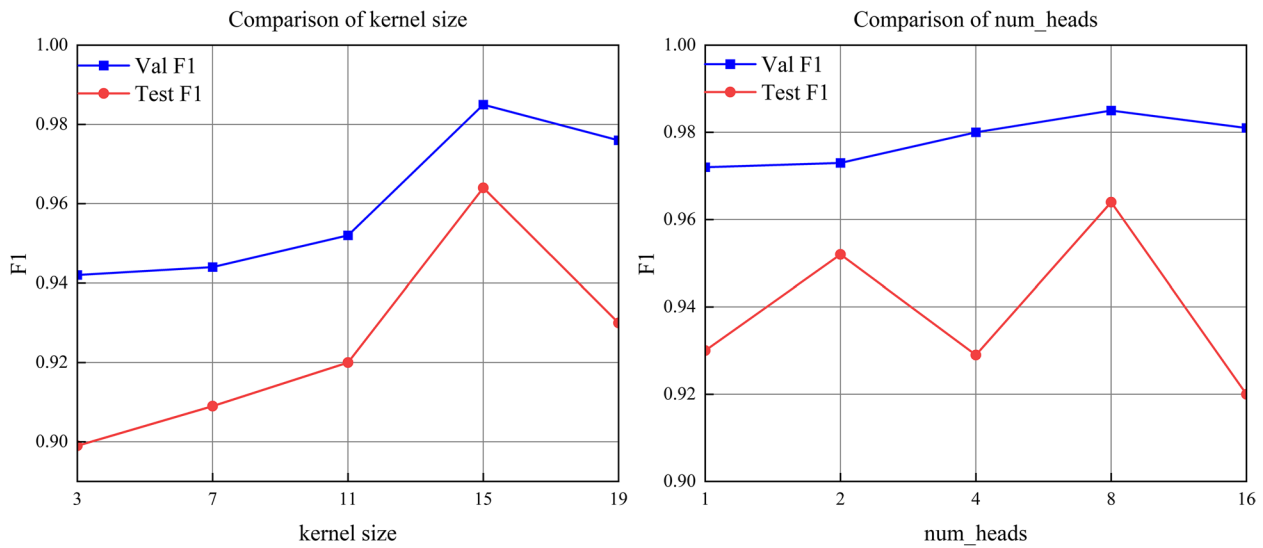


Fig. 7 Effect of different kernel sizes (left) and num_heads (right) on the model

Experiment two: comparing different signal backbone

In order to substantiate the superiority of the SE-TCN model, Experiment two involved a meticulous comparison of various Signal Backbone models, including ResNet18, ResNext18, Inception, VGG16, SE-ResNet18, and SE-ResNext18. Notably, we exclusively replaced the signal backbone component while keeping the expert feature module and cross-modal feature fusion module unchanged. Moreover, consistent datasets were employed for training and

testing, and identical hyperparameter settings, were utilized to ensure utmost fairness and reliability of the experiments.

Figure 8 represents the average accuracy curves of different signal backbone models on the validation set during the training process, while Fig. 9 depicts the average accuracy of different signal backbone models on the test set. The experimental results clearly demonstrate that the SE-TCN model exhibits a significant advantage, achieving an average accuracy of 0.968 on

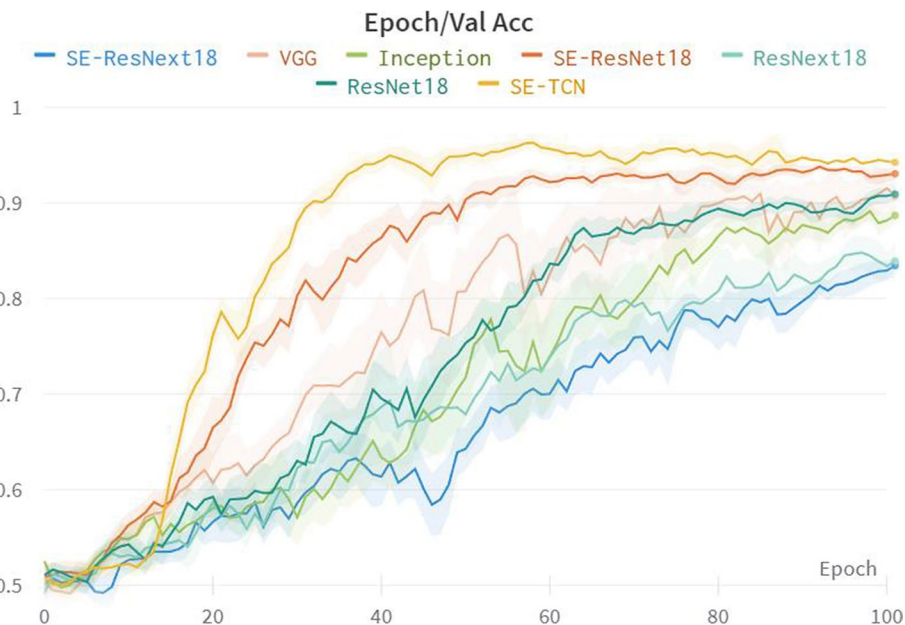


Fig. 8 The comparison results of the Accuracy of different Signal Backbone models on the validation set

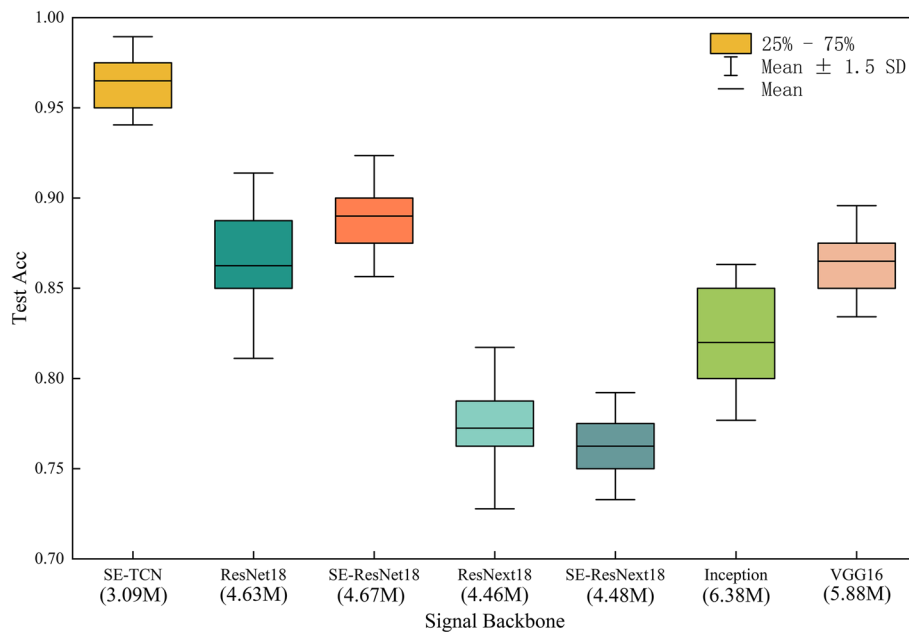


Fig. 9 The boxplots of the accuracy of different Signal Backbone models on the test set. The numbers in brackets on the x-axis indicate the total number of parameters for each signal backbone model. SD stands for standard deviation

the test set, compared to the accuracy range of 0.7725 to 0.89 for other models. Notably, the SE-TCN model surpasses the SE-ResNet18 and SE-ResNext18 models by 7.5 and 20.2% in terms of accuracy, respectively. This indicates that the SE-TCN model excels in feature extraction and cross-modal fusion, resulting in a noteworthy improvement in model accuracy. Furthermore, the SE-TCN model boasts a smaller total number of parameters, totaling at 3.09M, which makes it more lightweight compared to other models.

In summary, the SE-TCN model holds promising potential for applications in multi-modal signal processing tasks, as it demonstrates high accuracy while minimizing the number of parameters, making it a favorable choice for a high-performance and low-complexity model.

Experiment three: ablation experiments

In Experiment three, we conducted ablation experiments to thoroughly investigate the effects of different components in the Hybrid-FHR architecture. Specifically, we compared the performance of (1) using only expert features, (2) using only the signal backbone model (SE-TCN), and (3) using the complete Hybrid-FHR architecture. Furthermore, to demonstrate the importance of the proposed CMFF module, we compared early and late fusion approaches. In early fusion, the expert latent representation tensor and signal latent representation tensor are fused through simple concatenation. In late fusion, the two different modality tensors are each passed through their respective classification heads and then fused with a 1:1 decision weight.

Table 2 Performance comparison of different modal features and different fusion methods

Method	Acc%	Spe%	Sen%	Pre%	F1%
Morphological time domain features only	79.3	76	82.5	77.5	79.9
Frequency domain features only	80.8	77.5	84	78.9	81.4
Nonlinear features only	75.5	80.5	70.5	78.3	74.2
All expert features	86.8	83.5	90.0	84.5	87.1
Signal features only	92.0	95.0	89.0	94.7	91.8
Early fusion (concatenation fusion)	93.5	97.0	90.0	96.8	93.3
Late fusion (decision-level fusion)	95.8	96.0	95.5	96.0	95.7
Hybrid-FHR (Expert features + Signal features + CMFF)	96.8	97.5	96.0	97.5	96.7

According to the results of ablation experiments (Table 2), when considering only a single type of expert features, the order of the three expert feature types is: frequency domain > morphological time domain > nonlinear. The performance of single-modal features decreased to a certain extent compared to using the complete Hybrid-FHR architecture. When using all expert features, the accuracy decreased by 10 to 86.8% compared to the complete architecture, and when using signal features, the accuracy decreased by 4.8 to 92.0%. This indicates that the fusion of multimodal information is of great significance for improving the diagnostic accuracy and efficiency in medical diagnosis. Furthermore, in the comparison of different fusion methods, the late fusion performed slightly better than the early fusion, but still lower than our proposed CMFF method. This indicates that the CMFF method can better fuse different modal

information and improve the classification performance of the model.

In Table 3, we compared the generalization error of the model in two scenarios: with and without expert features. we can see that the generalization error of the model is reduced from 4.9 to 3% after incorporating the expert features. This indicates that incorporating expert features helps to reduce the generalization error of the model and prevents the risk of overfitting.

We plotted a t-distribution stochastic neighbor embedding (t-SNE) to visualize the output of each layer, as shown in Fig. 10. Initially, the raw data distribution appears scattered and lacks clear decision boundaries. However, as the network undergoes successive layers of feature extraction, the t-SNE plot gradually reveals distinct and separable clusters. This suggests that the network progressively learns and captures informative representations, leading to more discriminative features. Remarkably, the fusion latent representation output by CMFF forms visually distinct and well-separated clusters in the t-SNE plot. These evident clusters showcase the ability of CMFF to accurately capture and differentiate underlying patterns within the data.

Table 3 Comparison of generalization error with and without expert features

Method	Train Acc%	Test Acc%	Generalization error%
Without expert features	96.9	92.0	4.9
With expert features	99.8	96.8	3

Discussion

Table 4 offers a comprehensive overview of the various approaches proposed by researchers over the last few decades for fetal acidosis diagnosis. As show in Table 4

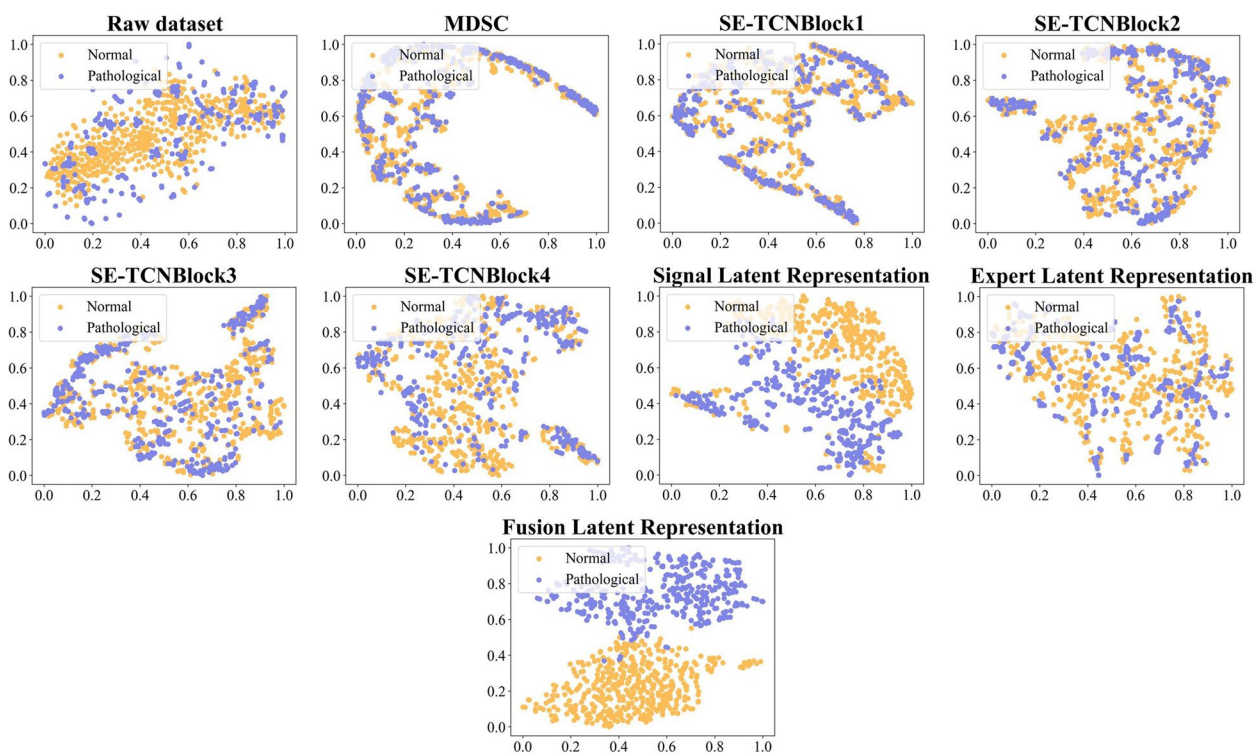


Fig. 10 The Visualization Output of each Layer in the Hybrid-FHR using t-SNE

Table 4 Comparison of different algorithms for fetal acidosis diagnosis

Reference	Method	Database	Division criterion	Performance%
[1]	Expert +1D-CNN	CTU-UHB	pH < 7.15	Sen: 75.23; Spe: 70.82
[5]	Expert + PCA + ANN	Private	pH < 7.1	Sen:60.3; Spe:67.5
[6]	Expert + RELIEF + AdaBoost	CTU-UHB	pH < 7.05	Sen:64.1; Spe:65.2
[7]	Expert + Random Forest	CTU-UHB	Expert annotation	Sen:72.41; Spe:78.4
[16]	1D-CNN-GRU	CTU-UHB	pH < 7.15	Acc: 95.15; Sen: 96.20 Spe: 94.09; Pre: 94.21
[26]	Expert + SVM / Naive Bayes	CTU-UHB	pH < 7.15	Sen: 73.4 / 72.3; Spe: 76.3 / 75.6
Current work	Hybrid-FHR (Expert + SE-TCN + CMFF)	CTU-UHB	pH < 7.15	Acc: 96.8 ; Sen: 96.0 Spe: 97.5 ; Pre: 97.5 F1: 96.7

AdaBoost Adaptive Boosting, GRU gated recurrent units, SVM support vector machine

Most of the existing studies use only single modal features (e.g., expert features, or 1D signal features).

To ensure fairness, we only compared algorithms [1, 16, 26] that utilized the CTU-UHB dataset and employed a pH threshold of 7.15 as the division criterion. We can draw several conclusions from Table 4. Firstly, our algorithm outperforms the state-of-the-art algorithms reported in previous literature, achieving the best performance on five different metrics. Secondly, comparing [16, 26], algorithms based on 1D signal features perform much better than algorithms based on expert features, demonstrating the advantage of DL over traditional ML methods. Thirdly, we notice some similarities between our approach and [1], which also incorporates expert features that are fused with 1D signal features. Nevertheless, it is worth noting that [1] employs a simple 1D-CNN model for extracting 1D signal features, followed by a late fusion at the decision level. In contrast, we utilized the SE-TCN as backbone network, which boasts superior long sequence signal feature extraction capabilities compared to conventional CNNs. Additionally, we introduced the CMFF module at the feature level, which explicitly models the correlation and difference between different modalities and further improves the classification effect.

In this work, we present an intelligent analysis algorithm Hybrid-FHR for diagnosing fetal acidosis. This algorithm can be integrated into clinical practice to aid obstetricians in making accurate medical decisions by considering the extracted expert feature parameters and the final predicted probability results. Based on the experimental results, we draw the following conclusions: (a.) Multimodal features lead to better classification results than using signal features or expert features alone. (b.) SE-TCN can effectively extract complex features from FHR signals, and outperforms six different baseline models in terms of convergence speed and parameter size. (c.)

Both late fusion and early fusion methods achieve satisfactory results, but they are still inferior to our proposed CMFF method in terms of accuracy.

Our algorithm in obstetrics and perinatal care holds significant practical implications by providing accurate and timely assessments of fetal distress. It facilitates early identification, leading to timely clinical interventions and preventing complications for both the mother and fetus. The algorithm reduces the diagnostic burden on healthcare professionals, automating aspects of diagnosis and allowing them to focus on critical patient care. Additionally, its computational nature makes it suitable for telemedicine applications, enabling remote monitoring and diagnosis, especially in areas with limited access to specialized healthcare facilities. In conclusion, our fetal distress diagnosis algorithm has the potential to enhance diagnostic efficiency, accuracy, and timeliness, positively impacting patient outcomes and overall perinatal care quality.

Conclusions

In this study, we propose a novel artificial intelligence algorithm called Hybrid-FHR for fetal acidosis diagnosis using multimodal features of the FHR signal. Our algorithm consists of three key components. First, we designed the SE-TCN backbone network to extract one-dimensional spatiotemporal representations from the FHR signal. Second, we incorporated three types of expert features including morphological time domain, frequency domain, and nonlinear parameters based on expert prior knowledge. Finally, we developed a cross-modal feature fusion (CMFF) method, which employs a multi-headed attention mechanism for fusing signal representations with expert feature representations.

We evaluate our algorithm against six baseline models and two fusion approaches on a publicly available dataset of FHR recordings. Our results demonstrate that

Hybrid-FHR outperforms the existing methods in terms of accuracy (96.8%) and efficiency. With the increasing number of publicly available datasets, we will apply the algorithm proposed in this study to different datasets to increase the robustness and generalizability of the model, while considering interpretable analysis to help clinicians make more objective and accurate decisions.

Abbreviations

1D-CNN	One-dimensional convolutional neural network
Acc	Accuracy
ANN	Artificial neural network
ANS	Autonomic nervous system
CMAS	Cross-modal attention score
CMFF	Cross-modal feature fusion
CWT	Continue wavelet transform
CTG	Cardiotocography
EFM	Electronic fetal monitoring
F1	F1-Score
FIGO	the International federation of gynecology and obstetrics
FFT	Fast fourier transform
FHR	Fetal heart rate
FHRV	Fetal heart rate variability
FetMov	Fetal movement
GAN	Generative adversarial network
GAP	Global average pooling
GLCM	Gray level co-occurrence matrix
GMP	Global maximum pooling
GRU	Gate recurrent unit
MDSC	Multi-scale depthwise separable convolution
PCA	Principal component analysis
Pre	Precision
RBF-SVM	Radial basis function support vector machine
RELIEF	RElevance in estimating features
ReLU	Rectified linear unit
RFE	Recursive feature elimination
SE-TCN	Squeeze and excitation temporal convolutional network
Sen	Sensitivity
Spe	Specificity
ST	Statistical testing
STFT	Short-time fourier transform
t-SNE	t-distribution Stochastic neighbor embedding
UC	Uterine contraction

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12911-024-02423-4>.

Additional file 1.

Additional file 2.

Additional file 3.

Acknowledgements

The authors wish to express their gratitude to the CTU-UHB database for its availability as a public resource.

Authors' contributions

Z.Z., P.J., J.W. and J.Z. conceived and designed the analysis; Z.Z. and J.Z. performed the analysis; J.Z. wrote the original draft; Z.Z., Y.Z., P.J., J.W., X.Z. and X.L. reviewed and edited the draft; Z.Z. and Y.Z. acquired the funding. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (Grant 62071162 and 62301205) and the Zhejiang Provincial Natural Science

Foundation of China (Grant No. LDT23F01012F01 and LDT23F01015F01). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The CTU-UHB database is a publicly available resource (<https://physionet.org/content/ctu-uhb-ctgdb/1.0.0/>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 26 April 2023 Accepted: 10 January 2024

Published online: 22 January 2024

References

- 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 fetal acidosis classification. *Expert Syst Appl.* 2021;186:115714.
- Fanelli A, Magenes G, Campanile M, Signorini MG. Quantitative assessment of fetal well-being through CTG recordings: a new parameter based on phase-rectified signal average. *IEEE J Biomed Health Inform.* 2013;17(5):959–66. <https://doi.org/10.1109/JBHI.2013.2268423>.
- Steer PJ. Has electronic fetal heart rate monitoring made a difference. *Semin Fetal Neonatal Med.* 2008;13(1):2–7. <https://doi.org/10.1016/j.siny.2007.09.005>.
- Costa Santos C, Costa Pereira A, Bernardes J. Agreement studies in obstetrics and gynaecology: inappropriateness, controversies and consequences. *BJOG.* 2005;112(5):667–9. <https://doi.org/10.1111/j.1471-0528.2004.00505.x>.
- Georgieva A, Payne SJ, Moulden M, Redman CWG. Artificial neural networks applied to fetal monitoring in labour. *Neural Comput Appl.* 2013;22(1):85–93. <https://doi.org/10.1007/s00521-011-0743-y>.
- Spilka J, Georgoulas G, Karvelis P, Oikonomou VP, Chudáček V, et al. Automatic evaluation of FHR recordings from CTU-UHB CTG database. *Inf Technol Bio Med Inform.* 2013;8060:47–61. https://doi.org/10.1007/978-3-642-40093-3_4.
- Spilka J, Georgoulas G, Karvelis P, Chudáček V, Stylios CD, et al. Discriminating Normal from abnormal pregnancy cases using an automated FHR evaluation method. *Lect Notes Artif Intell.* 2014;8445:21–31. https://doi.org/10.1007/978-3-319-07064-3_45.
- Cömert Z, Kocamaz AF, Subha V. Prognostic model based on image-based time-frequency features and genetic algorithm for fetal hypoxia assessment. *Comput Biol Med.* 2018;99(1):85–97. <https://doi.org/10.1016/j.combiomed.2018.06.003>.
- Zhao Z, Zhang Y, Deng Y. A comprehensive feature analysis of the fetal heart rate signal for the intelligent assessment of fetal state. *J Clin Med.* 2018;7(8):223. <https://doi.org/10.3390/jcm7080223>.
- Pini N, Lucchini M, Esposito G, Tagliaferri S, Campanile M, et al. A machine learning approach to monitor the emergence of late intrauterine growth restriction. *Front. Artif Intell.* 2021;4 <https://doi.org/10.3389/FRAI.2021.622616>.
- Naylor C. On the prospects for a (deep) learning health care system. *JAMA.* 2018;320(11):1099–100. <https://doi.org/10.1001/jama.2018.11103>.
- Hinton G. Deep learning—a technology with the potential to transform health care. *JAMA.* 2018;320(11):1101–2. <https://doi.org/10.1001/jama.2018.11100>.
- Bursa M, Lhotska L. The use of convolutional neural networks in biomedical data processing. *Proc ITBAM.* 2017;57:100–19. https://doi.org/10.1007/978-3-319-64265-9_9.
- Cömert Z, Kocamaz AF. Fetal hypoxia detection based on deep convolutional neural network with transfer learning approach. In: *In: Proceedings*

- of 7th computer science on-line conference, software engineering and algorithms in intelligent systems, April. Springer International Publishing; 2018. https://doi.org/10.1007/978-3-319-91186-1_25.
15. Li J, Chen ZZ, Huang L, Fang M, Li B, et al. Automatic classification of fetal heart rate based on convolutional neural network. *IEEE Internet Things*. 2018; <https://doi.org/10.1109/JIOT.2018.2845128>.
 16. Liang H, Lu Y. A CNN-RNN unified framework for intrapartum cardiotocograph classification. *Comput Methods Prog Biomed*. 2022;229 <https://doi.org/10.1016/j.cmpb.2022.107300>.
 17. Fei Y, Chen F, He L, Chen J, Hao Y, et al. Intelligent classification of antenatal cardiotocography signals via multimodal bidirectional gated recurrent units. *Biomed. Signal Process*. 2022;78 <https://doi.org/10.1016/j.bspc.2022.104008>.
 18. Spairani E, Daniele B, Signorini MG, Magenes G. A deep learning mixed-data type approach for the classification of FHR signals. *Front Bioeng Biotechnol*. 2022;10 <https://doi.org/10.3389/fbioe.2022.887549>.
 19. Ayres-de-Campos D, Spong CY, Chandraran E, Panel FIFMEC. FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography. *Int J Gynaecol Obstet*. 2015;131(1):13–24. <https://doi.org/10.1016/j.ijgo.2015.06.020>.
 20. Gatellier MA, De Jonckheere J, Storme L, Houfflin-Debarge V, Ghesquiere L, et al. Fetal heart rate variability analysis for neonatal acidosis prediction. *J Clin Monit Comput*. 2021;35(4):771–7. <https://doi.org/10.1007/s10877-020-00535-6>.
 21. Hämmerle P, Eick C, Blum S, Schlageter V, Bauer A, et al. Heart rate variability triangular index as a predictor of cardiovascular mortality in patients with atrial fibrillation. *J Clin Monit Comput*. 2020;9(15):e016075. <https://doi.org/10.1161/JAHA.120.016075>.
 22. Signorini MG, Magenes G, Cerutti S, Arduini D. Linear and nonlinear parameters for the analysis of fetal heart rate signal from cardiotocographic recordings. *IEEE Trans Biomed Eng*. 2003;50(3):365–74. <https://doi.org/10.1109/TBME.2003.808824>.
 23. David M, Hirsch M, Karin J, Toledo E, Akselrod S. An estimate of fetal autonomic state by time-frequency analysis of fetal heart rate variability. *J Appl Physiol*. 2007;102(3):1057–64. <https://doi.org/10.1152/jappphysiol.00114.2006>.
 24. Warmerdam GJJ, Vullings R, Bergmans JWM, Oei SG. Reliability of spectral analysis of fetal heart rate variability. Chicago: In: Proceedings of IEEE EMBC; 2014. p. 2817–20. <https://doi.org/10.1109/EMBC.2014.6944209>.
 25. Ribeiro M, Monteiro-Santos J, Castro L, Antunes L, Costa-Santos C, et al. Non-linear methods predominant in fetal heart rate analysis: a systematic review. *Front Med (Lausanne)*. 2021;8 <https://doi.org/10.3389/fmed.2021.661226>.
 26. Spilka J, Chudáček V, Koucký M, Lhotská L, Huptych M, et al. Using non-linear features for fetal heart rate classification. *Biomed Signal Process*. 2012;7(4):350–7. <https://doi.org/10.1016/j.bspc.2011.06.008>.
 27. Tetschke F, Schneider U, Schlessner E, Witte OW, Hoyer D. Assessment of fetal maturation age by heart rate variability measures using random forest methodology. *Comput Biol Med*. 2016;70:157–62. <https://doi.org/10.1016/j.compbimed.2016.01.020>.
 28. Fang B, Chen J, Liu Y, Wang W, Wang K, et al. Dual-channel neural network for atrial fibrillation detection from a single Lead ECG wave. *IEEE J Biomed Health Inform*. 2023;27(5):2296–305. <https://doi.org/10.1109/JBHI.2021.3120890>.
 29. Shi B, Zhang Y, Yuan C, Wang S, Li P. Entropy analysis of short-term heart-beat interval time series during regular walking. *Entropy*. 2017;19(10) <https://doi.org/10.3390/e19100568>.
 30. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J*. 1948;27(3):379–423. <https://doi.org/10.1145/584091.584093>.
 31. Lempel A, Ziv J. On the complexity of finite sequences. *IEEE Trans Inf Theory*. 1976;22(1):75–81. <https://doi.org/10.1109/TIT.1976.1055501>.
 32. Higuchi T. Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena*. 1988;31(2):277–83. [https://doi.org/10.1016/0167-2789\(88\)90081-4](https://doi.org/10.1016/0167-2789(88)90081-4).
 33. Hurst HE. Long-term storage capacity of reservoirs. *Trans Am Soc Civ Eng*. 1951;116(1):770–99. <https://doi.org/10.1061/TACEAT.0006518>.
 34. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv. <https://doi.org/10.48550/arXiv.1704.04861>.
 35. van den Oord A, Dieleman S, Zen H, Simonyan K, Vinyals O, et al. WaveNet: A generative model for raw audio. arXiv. 2016. <https://doi.org/10.48550/arXiv.1609.03499>.
 36. Bai S, Zico Kolter J, Koltun V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. arXiv. <https://doi.org/10.48550/arXiv.1803.01271>.
 37. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: In: 2016 IEEE conference on computer vision and pattern recognition (CVPR). IEEE; 2016. p. 770–8. <https://doi.org/10.1109/CVPR.2016.90>.
 38. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. arXiv. <https://doi.org/10.48550/arXiv.1502.03167>.
 39. Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on International Conference on Machine Learning. 2010; pp. 807–14. <https://doi.org/10.5555/3104322.3104425>.
 40. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: In: 2018 IEEE conference on computer vision and pattern recognition (CVPR). IEEE; 2018. p. 132–41. <https://doi.org/10.1109/CVPR.2018.00745>.
 41. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, et al. Attention Is All You Need. arXiv; 2017. <https://doi.org/10.48550/arXiv.1706.03762>.
 42. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*. 2000;101(23):e215–ee20. <https://doi.org/10.1161/01.cir.101.23.e215>.
 43. Chudáček V, Spilka J, Burša M, Janků P, Hruban L, et al. Open access intrapartum CTG database. *BMC Pregnancy Childbirth*. 2014;14(1):16. <https://doi.org/10.1186/1471-2393-14-16>.
 44. Zhang Y, Zhao Z, Deng Y, Zhang X, Zhang Y. Reconstruction of missing samples in antepartum and intrapartum FHR measurements via mini-batch-based minimized sparse dictionary learning. *IEEE J Biomed Health Inform*. 2022;26(1):276–88. <https://doi.org/10.1109/JBHI.2021.3093647>.
 45. Zhang Y, Zhao Z, Deng Y, Zhang X. FHRGAN: Generative adversarial networks for synthetic fetal heart rate signal generation in low-resource settings. *Inf Sci*. 2022;594:136–50. <https://doi.org/10.1016/j.ins.2022.01.070>.
 46. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. arXiv. <https://doi.org/10.48550/arXiv.1412.6980>.
 47. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. arXiv. <https://doi.org/10.48550/arXiv.1603.04467>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.