

PAPER

Fetal Health Assessment Using Multimodal Time-Series Data and Attention Neural Networks

Hao Wang, Yin Zhao, Yajie Gu,
Fengliang Huang  

Nanjing Normal University,
Nanjing, China

huangfengliang@njnu.edu.cn

ABSTRACT

To address the issues of high subjectivity in traditional interpretation methods, the widespread neglect of gestational age and maternal age in existing research, small dataset sizes, class imbalance, and limited model performance, this paper proposes a convolutional long short-term memory network based on a multimodal attention mechanism (MACLM) to assist in the assessment of fetal health status. Based on 22,813 fetal monitoring records, the study embeds gestational age information using sinusoidal positional encoding and combines an attention mechanism with a convolutional neural network–long short-term memory (CNN-LSTM) architecture to achieve multimodal feature fusion and temporal modeling. Experimental results show that positional encoding improves the model's accuracy by 4%–7% and reduces the loss by 46%–50%. MACLM ultimately achieved a classification accuracy of 97.55%, with an interpretation time of only 0.0234 seconds for 100 fetal monitoring cases. Its overall performance significantly outperformed other comparison models, with a maximum improvement of 92.7%.

KEYWORDS

fetal health monitoring, multimodal, neural networks, attention mechanisms

1 INTRODUCTION

Fetal safety has always been the top priority in perinatal care. During pregnancy, fetal hypoxia may cause complications such as cerebral palsy and fetal growth restriction and, in severe cases, can result in fetal death [1]. According to World Health Organization data, as of 2019, the global number of stillbirths exceeded two million; half of these deaths occurred during the delivery process [2]. The 2022 Statistical Bulletin on the Development of China's Health and Wellness Sector report indicates that the neonatal mortality rate was 1.9‰ in urban areas and approximately 3.6‰ in rural areas [3]. It is evident that the uneven distribution of medical resources, with some regions experiencing a shortage of doctors, is also a significant factor contributing to stillbirths. Currently, cardiotocography

Wang, H., Zhao, Y., Gu, Y., Huang, F. (2026). Fetal Health Assessment Using Multimodal Time-Series Data and Attention Neural Networks. *International Journal of Online and Biomedical Engineering (iJOE)*, 22(10), pp. 161–178. <https://doi.org/10.3991/ijoe.v22i10.62248>

Article submitted 2026-05-10. Revision uploaded 2026-06-17. Final acceptance 2026-06-17.

© 2026 by the authors of this article. Published under CC-BY.

(CTG) serves as the primary monitoring method for assessing fetal health. Through accurate interpretation of CTG tracings, the condition of the fetus within the uterus can be promptly evaluated [4]. When interpreting CTG results, physicians rely on subjective clinical judgment, which inevitably leads to flaws in interpretation and significant individual variation, particularly in pathological cases [5, 6]. This has created an urgent need for fetal monitoring technology to enhance physician interpretation efficiency and accuracy, enabling timely access to highly accurate and objective CTG results.

In recent years, with advancements in deep learning, machine learning, computer hardware, and database structures, artificial intelligence has been widely applied in the medical field [7, 8]. Research over the past few years has demonstrated its significant potential for practical implementation [9]. To enhance the reliability of clinical decision-making in fetal heart monitoring, the academic community has progressively integrated machine learning with CTG classification. Machine learning-based intelligent CTG diagnosis enables rapid and accurate assessment of fetal health status, effectively mitigating subjective errors and reducing inter-observer variability. This technology shows broad application prospects in the field of fetal heart rate monitoring [10]. Dash et al. extracted feature sequences from local patterns in fetal heart rate evolution. By integrating generative models and Bayesian theory, they achieved 82% specificity and 61% sensitivity across 83 fetal heart rate recordings [11]. Georgoulas et al. achieved an 81.25% classification accuracy for fetal health status by extracting time-domain features, frequency-domain features, and morphological features, then applying feature dimension reduction methods combined with support vector machines (SVM). They noted that classification using fetal heart rate alone is insufficient to attain very high model performance [12]. Chamidah et al. performed clustering on 21 features, including fetal heart rate (FHR) and uterine contractions (UC), using the K-means algorithm. They constructed seven fuzzy features through a fuzzy membership reconstruction method and fed these fuzzy features into an SVM for classification, achieving a fetal health status classification accuracy of 90.64% [13]. Feng et al. achieved a fetal health classification accuracy of 96.08% and an average F1 score of 0.9336 using a stacked model based on extreme gradient boosting and random forest (XGB-RF) feature selection. Simultaneously, XGB-RF demonstrated that different features in CTG exhibit varying importance coefficients, with a maximum difference of 0.168836 [14]. Zhang et al. employed Principal Component Analysis to select five features from 21 in the CTG dataset and used Adaptive Boosting, achieving a 98.6% classification accuracy for fetal health status [15]. The aforementioned studies have focused on extracting effective features from fetal heart rate monitoring data and employing dimension reduction or feature selection methods to eliminate less significant features. However, the differences in importance among various features after extraction have not been given sufficient attention or fully utilized. This oversight may result in under-reporting or misreporting by models.

The foundation of research based on artificial intelligence and machine learning algorithms lies in constructing extensive datasets containing a vast array of diverse data, while the determination of categories must be authoritative [16]. Current research predominantly utilizes CTG datasets sourced from the UCI Repository, categorized into three groups: 1,655 normal cases, 295 suspected cases, and 176 pathological cases. These datasets suffer from limited sample sizes, insufficient testing capabilities, and imbalanced category distributions, which may compromise

model performance and reliability [17]. To address the challenges of limited data volume and uneven distribution of data categories, the academic community has employed oversampling techniques to augment training data and enhance model performance [18, 19]. However, oversampling cannot substitute for authentic clinical data. Research by Tarawneh et al. indicates that this sampling technique generates samples with an error rate of 40–60%, with some of the generated data belonging to the majority group [20]. Therefore, it is crucial to construct a fetal health monitoring dataset that is large in volume, authentic in content, and valuable for practical use, while employing various methods to mitigate the impact of imbalanced data category distribution.

2 MATERIALS AND METHODS

2.1 Data collection and preprocessing

The data for this study were sourced from the China Perinatal Maternal and Infant Health Surveillance and AI Risk Early Warning Thematic Scientific Database within the data warehouse of the China National Population Health Science Data Center [21]. From January 1, 2015, to December 31, 2022, the National Institute of Health Science and Technology of the National Health Commission of China collected fetal heart rate variation data from 14,000 pregnant women across representative administrative regions (Hebei, Liaoning, Shandong, Hubei, Guangdong, Chongqing, and Shaanxi) using remote electronic fetal heart rate monitoring devices. The dataset includes normal fetal heart rate patterns, monitoring data under different maternal conditions, abnormal data from fetuses with varying degrees of hypoxia, and post-intervention recovery data, totaling 305,075 cases. Among these, 231,446; 61,870; 2,275; and 9484 cases were classified as normal, suspicious, pathological, and uninterpretable, respectively. The data collection process followed relevant clinical guidelines, such as the ACOG guidelines, and monitoring was performed using the eFM-60 Doppler fetal monitor. The data verification and consistency evaluation study was conducted by the Center for Maternal and Child Health of the Chinese Center for Disease Control and Prevention.

Madow noted that when samples are positively correlated with their neighboring units, equal-distance sampling can maintain sample representativeness while achieving a smaller sample mean variance than simple random sampling of the same size [22]. In accordance with the research objectives, this study examined the raw data and excluded records with an interpretation result of “uninterpretable,” as well as data with missing fields such as maternal age, gestational week, gestational day, key fetal heart rate monitoring characteristics, interpretation results, or UID. Due to the positive correlation between adjacent samples in the original data and the presence of numerous similar samples, this issue easily leads to model overfitting. Therefore, we employed an equidistant sampling strategy (with a stride set to 15) for samples classified as normal, suspicious, and pathological. To maintain consistent maternal age and pregnancy duration distributions before and after sampling, the sample underwent calibration: by sequentially comparing the proportions of different gestational ages and pregnancy durations in pre- and post-sampling data, we supplemented data with significant deviations using equidistant sampling with a stride of 5 from the remaining eligible data. A total of 22813 fetal health

monitoring data points were ultimately obtained. This corresponds to 3,841 unique pregnant women.

Given that a single pregnant woman may have multiple fetal heart rate monitoring records from different examination times, this paper further employs UID-based data validation during the dataset splitting process to prevent records from the same subject from appearing in both the training and testing sets, thereby avoiding potential data leakage. Specifically, multiple records corresponding to the same pregnant woman share the same UID identifier; prior to splitting the dataset into training and testing sets, samples are first grouped by UID, ensuring that all records with the same UID are assigned to only one data subset. After the split is complete, the UIDs in the different data subsets are verified one by one to ensure that there are no overlapping records of the same pregnant woman between the training and testing sets. This approach minimizes the impact of data from the same subject on model performance, allowing the test results to more accurately reflect the model's generalization ability to unseen pregnant women. Table 1 presents the distribution of individuals by gestational age and maternal age.

Table 1. Distribution of gestational and maternal age

Characteristic	Range	Number	Ratio
Maternal Age (Year)	18–23	1261	5.53%
	24–29	11640	51.02%
	30–35	7781	34.11%
	36–41	1946	8.53%
	42–46	185	0.81%
Gestational age (Day)	155–187	36	0.16%
	188–220	502	2.20%
	221–243	3982	17.45%
	244–266	11352	49.76%
	267–295	6941	30.42%

Based on fetal heart rate monitoring interpretation standards [23] and the potential effects of gestational age and maternal age, only relevant features were retained from the 22,813 data points, as detailed in Table 2.

Table 2. Feature selection results list

Characteristic	Description	Retain
Data number (Id)	Unique Data ID	Yes
Individual number (UID)	32-character string; user unique identifier	No
Guardianship start time	Year-Month-Day-Hour-Minute-Second	No
End of custody time	Year-Month-Day-Hour-Minute-Second	No
Age of the foetus at the time of monitoring	Gestational weeks + gestational days	No

(Continued)

Table 2. Feature selection results list (*Continued*)

Characteristic	Description	Retain
Gestational week (Mk)	Unit: Week	Yes
Maternal age (Ma)	Unit: year	Yes
Gestational day (Md)	Unit: Day	Yes
Filing hospitals	32-bit string; hospital identification code	No
Fetal cardiac monitoring without stress test monitoring results (Mr)	Range of values: 1–3. 1: reactive; 2: suspicious; 3: non-reactive; 4: unreadable.	Yes
Is there a contraction picture	Value range 0–1. 0: no contraction picture; 1: contraction picture.	No
Detailed Data Pictures	Image resource	No
Type of mutation	Range of values: 0–3. 0: unknown; 1: variation missing; 2: minor variation; 3: normal variation; 4: significant variation.	No
Fetal heart rate baseline (Fh)	Unit: bmp	Yes
Accelerations (Ac)	Number of fetal heart rate accelerations	Yes
Variable decelerations (Vd)	Number of variable decelerations	Yes
Late decelerations (Ld)	The number of late decelerations detected	Yes

Given the scarcity of pathological cases, pathological and suspected cases were categorized as abnormal cases. Among the 22,813 data points, the number of normal and abnormal cases was 16,491 and 6,322, respectively, represented by 0 and 1. Partial data is shown in Table 3. Min–max normalization was performed on features (excluding Id, Mk, Md, and Mr).

Table 3. Partial input data

Id	Mk	Md	Mr	Ma	Fh	Ac	Vd	Ld
jhwgfi...	40	4	1	28	144	6	2	0
kjkhk...	39	1	0	31	135	1	1	0
akfba...	38	2	1	37	135	4	1	0
....	...	...	...	...	...	...	...	...
cljdho...	35	6	1	24	148	7	1	0

2.2 Method

This paper proposes a multimodal temporal attention convolutional long short-term memory neural network model (MACLM) for classifying fetal health monitoring results during pregnancy. The overall model architecture is shown in Figure 1.

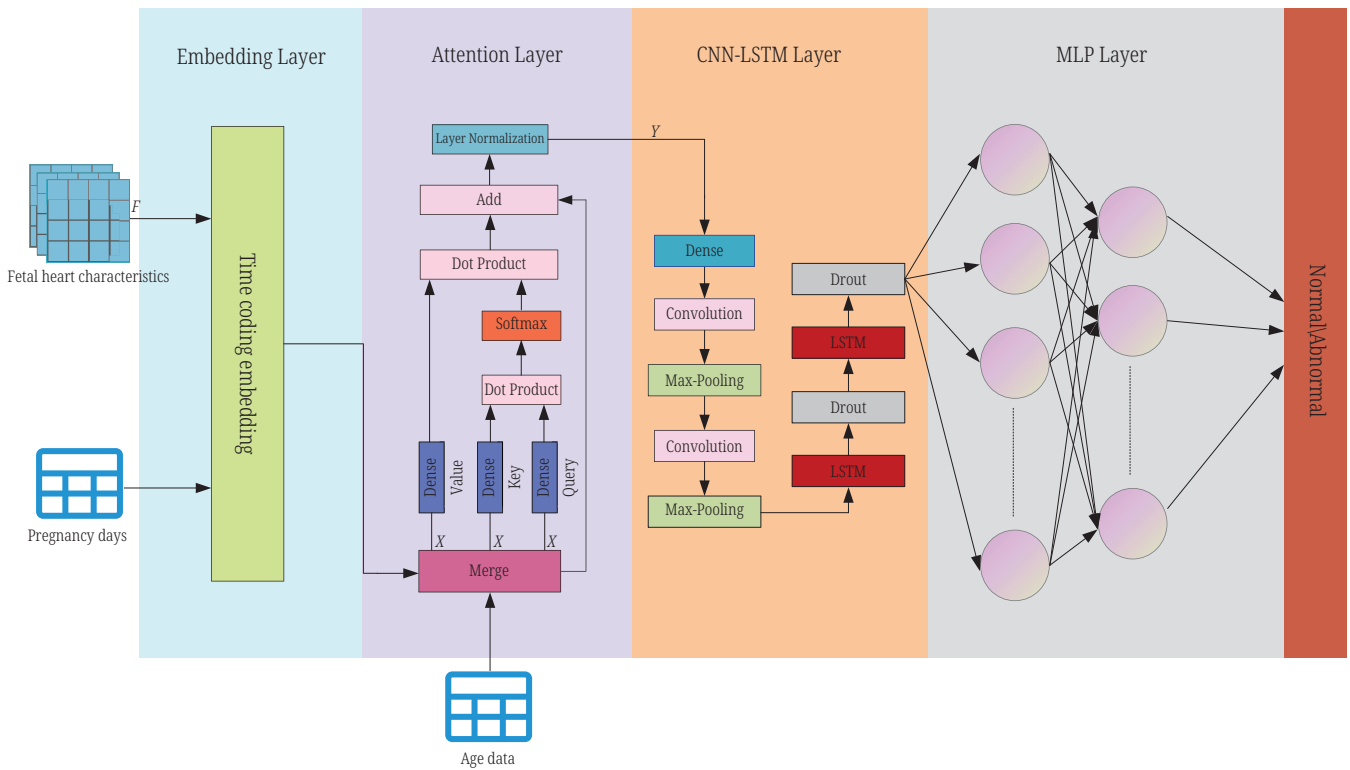


Fig. 1. MTACLM model structure

Note: The model processes multimodal fetal heart rate monitoring features sequentially through gestational age encoding, attention-based feature representation, CNN-LSTM temporal feature extraction, and an MLP classification module.

Experimental setup. All model training, testing, and inference time evaluations in this study were conducted in a standardized hardware and software environment. The experimental platform consisted of an Intel® Core™ i7-10750H processor, an RTX 2060 graphics card, and 16 GB of RAM. The deep learning frameworks used in this study were TensorFlow 2.13.0 and PyTorch 2.4.1.

Time embedding layer. The temporal embedding layer is constructed based on the positional encoding method proposed by Vaswani et al. [24], aiming to convert gestational age position information into a gestational age position-time vector **PE** through encoding formula (1), thereby achieving fusion of multimodal features across the temporal dimension.

$$\begin{cases} PE(pos, 2i) = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \text{ if } i \text{ is even} \\ PE(pos, 2i+1) = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \text{ if } i \text{ is odd} \end{cases} \quad (1)$$

In equation (1), pos is the position index of gestational age, where i represents the dimensional index of the position encoding vector, ranging from 0 to d_{model} . Here, d_{model} represents the model's dimensionality.

Calculate gestational age using Mk and Md, then sort all gestational ages in ascending order to form a one-dimensional ordered array. Insert missing values to ensure continuous gestational age. In the array, each unique gestational age is

assigned a unique positional index. The encoded vector representing gestational age is added to the fetal heart monitoring data matrix F , enabling the model to capture changes in fetal heart characteristics over gestational age. This approach overcomes the limitations of traditional neural network models constrained by gestational age restrictions.

Attention layer. The construction of this layer is based on a multi-head self-attention mechanism [24]. Based on formula (2), compute the query(Q), key(K), and value(V) vectors for the input features.

$$\begin{cases} Q = XW_Q \\ K = XW_K \\ V = XW_V \end{cases} \quad (2)$$

Where W_Q , W_K , and W_V represent their respective learnable weight matrices, the weighted sum of the value matrix is computed using formula (3). This enables MACLM to identify critical time steps that may foreshadow or explain the current fetal heart event, thereby adjusting the attention weights of different features to enhance the detection capability of abnormal fetal heart conditions.

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \cdot V \quad (3)$$

Through residual connections and layer normalization (4), the model is made more amenable to learning identity mappings and mitigating gradient vanishing.

$$Y = LayerNorm(X + Attention(Q, K, V)) \quad (4)$$

CNN-LSTM layer. The CNN-LSTM layer in MACLM, as shown in Figure 2, consists of a fully connected layer, a convolutional layer, a max pooling layer, an LSTM layer, and a forget layer.

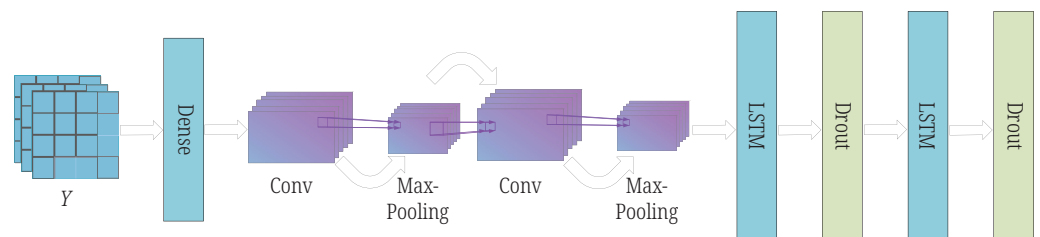


Fig. 2. Convolutional neural network structure

Note: The CNN module extracts local temporal patterns from the fused fetal heart rate feature sequences through convolutional and pooling operations, providing local feature representations for subsequent LSTM-based modeling of temporal dependencies.

Projecting the attention mechanism's output through a fully connected layer enhances the expressive power of features. Feature extraction and refinement are achieved using convolutional layers and max-pooling layers based on formula (5) [25, 26].

$$C_k^l = f(W_i^l \cdot G^{l-1} + B_i^l) \quad (5)$$

Where C_k^l is the k -th feature of layer l ; B_i^l and W_i^l represent the weights and biases of the i -th convolution kernel in the l -th layer, respectively; G^{l-1} is the output feature map of layer $l-1$ th; f denotes the ReLU activation function.

The formula for the maximum pooling layer is formula (6):

$$P_k^l = \text{Max}(C_k^l) \quad (6)$$

Where P_k^l is the max-pooling output for region l ; C_k^l denotes the k -th region within the l -th feature map; Max represents the max-pooling function.

The output of the pooling layer serves as input to the LSTM. Within the LSTM, the memory cell is central to handling long-term dependencies, with its state controlled by the input gate, forget gate, and output gate. A dropout layer is introduced after each LSTM layer to prevent overfitting [27]. The components of LSTM are computed as formula (7) [28]:

$$\begin{cases} f_t = s \cdot (W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t = s \cdot (W_i \cdot [h_{t-1}, x_t] + b_i) \\ \bar{C}_t = \tanh[W_c \cdot (h_{t-1}, x_t) + b_c] \\ C_t = f_t \cdot C_{t-1} + i_t \cdot \bar{C}_t \\ o_t = [W_o \cdot (h_{t-1}, x_t) + b_o] \\ h_t = o_t \cdot \tanh(C_t) \end{cases} \quad (7)$$

Where f_t is the output vector for the forgetting gate; h_{t-1} is the previous input; x_t is the current input; b_f is the gated bias vector; s is the sigmoid activation function; i_t is the input gate; $\bar{C}_t$ is the new candidate vector generated by the tanh layer; c_t is the time step memory state; o_t is the activation vector for the output gate; and h_t is the output result.

Multi-layer perceptron layer. A multi-layer perceptron (MLP) is introduced, utilizing the LSTM output h as input. Through three linear transformations and a nonlinear activation function, higher-order features are progressively extracted, yielding the classification result y via the equation (8).

$$y = s(W_3(\text{ReLU}(W_2(\text{ReLU}(W_1 h_t + b_1)) + b_2)) + b_3) \quad (8)$$

Where W_1 , W_2 , and W_3 are linear transformations; b_1 , b_2 , and b_3 are bias terms.

Parameter learning. In the proposed MACLM, the trainable parameters of all components are learned using a custom binary cross-entropy loss function (9). The dataset used in this study has a ratio of approximately 2.6:1 between normal and abnormal samples, which may adversely affect classifier performance. Therefore, greater weights are assigned to abnormal cases in the loss function to minimize missed diagnoses.

$$\text{Loss} = -\frac{1}{N} (d \cdot (y \cdot \log y') + ((1-d) \cdot (1-y) \cdot \log(1-y'))) \quad (9)$$

In the loss function, d is a hyperparameter real number between 0 and 1 used to define the relative weight of predictions; y represents the true value; y' represents the predicted value; N represents the total number of samples. In this study, d is set to 0.7. For the optimization process, training is performed using the Adaptive Moment Estimation (Adam) optimizer [29] with a default learning rate of 0.001.

2.3 Evaluation metrics

To validate the effectiveness of MACLM classification, this study introduces accuracy (Acc), precision (Pre), recall, specificity (Spe), normal category F1 score (NF1), abnormal category F1 score (AF1), and Matthews correlation coefficient (MCC) based on the confusion matrix to evaluate model performance. The calculation methods for these indicators are function (10) [30]:

$$\left\{ \begin{array}{l} \text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \\ \text{Precision} = \frac{TP}{TP + FP} \\ \text{Recall} = \frac{TP}{TP + FN} \\ \text{Specificity} = \frac{TN}{TN + FP} \\ \text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \\ \text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \end{array} \right. \quad (10)$$

The area under the curve (AUC) is introduced to intuitively reflect the model's classification capability. AUC represents the area under the receiver operating characteristic curve (ROC). The ROC curve is defined by the false positive rate (FPR) and true positive rate (TPR), with FPR on the horizontal axis and TPR on the vertical axis. The calculation formula is function (11):

$$\left\{ \begin{array}{l} \text{FPR} = \frac{FP}{FP + TN} \\ \text{TPR} = \frac{TP}{TP + FN} \end{array} \right. \quad (11)$$

To more intuitively quantify the model's interpretation efficiency, the Average Processing Time (APT) required to process 100 data points is calculated. APT calculation formula such as formula (12). $Time_{total}$ represents the total time required to compute all test sets (in seconds), and Sum represents the total number of data.

$$APT = 100 \cdot \frac{Time_{total}}{Sum} \quad (12)$$

3 RESULTS

3.1 Sensitivity analysis of the d-value and analysis of attention weights for different features

Given that the dataset contains more normal samples than abnormal ones, if the weights are set improperly, the model may favor the majority class or unduly increase the weights of the minority class, thereby affecting the overall classification stability. Therefore, this study compared the impact of different d values on model performance while keeping other training parameters constant. In the experiments, d was set to 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9, and the model was evaluated using metrics

such as accuracy, recall, precision, specificity, F1-score, MCC, and AUC. The experimental results are shown in Table 4. The experimental results indicate that $d = 0.7$ achieves a good balance between the ability to identify anomalous samples and the ability to distinguish normal samples, and the model achieved the best overall performance. Therefore, this paper ultimately selects $d = 0.7$ as the weight parameter in the weighted loss function.

Table 4. Comparison of MACLM performance at different d values

d	Acc (%)	Recall (%)	Pre (%)	Spe (%)	NF1 (%)	AF1 (%)	MCC (%)	AUC
0.4	91.56	91.35	96.89	92.12	94.22	85.82	83.40	0.7951
0.5	93.77	94.21	97.18	92.66	95.63	89.19	85.36	0.8582
0.6	96.14	97.25	97.45	93.23	97.3	92.98	91.31	0.9285
0.7	97.55	98.61	97.71	95.46	96.91	95.06	93.27	0.9689
0.8	96.22	97.85	96.95	91.97	97.4	93.17	91.28	0.9363
0.9	80.17	92.87	82.14	47.22	87.12	56.89	46.26	0.7144

To further improve the interpretability of the MACLM model's classification results, this paper conducts a statistical analysis of the attention weights within the model. After model training is complete, the weights corresponding to different input features in the attention layer are extracted to analyze the relative contributions of each input feature to the model's classification process. The results of the attention weight analysis are shown in Table 5. Notably, the weights of indicators directly related to fetal heart rate monitoring are all higher than those of maternal age, indicating that the model's classification process primarily relies on the fetal heart rate monitoring signal itself, while maternal age serves as a background factor providing supplementary discriminative information. The weight of the number of accelerations is the highest, suggesting that the model focuses more on fetal reactivity information during classification; indicators related to fetal heart rate baseline and decelerations also have high weights, indicating that the model is capable of monitoring the baseline fetal heart rate status as well as potential abnormal deceleration information.

Table 5. Feature importance weights

Feature	Importance Weight
Ac	0.2980
Fh	0.2228
Ld	0.2128
Vd	0.1747
Ma	0.0917

3.2 Comparison of classification results for fetal health monitoring data across different models

To validate the performance of the proposed model and its advantages over other models, comparisons were conducted with SVM, random forest (RF),

LSTM, gated recurrent units (GRU), InceptionTime (IT), and MultiROCKET (MRT). Simultaneously, we constructed an Attention-CNN-LSTM model (ACNNLSTM) with gestational age encoding removed and a CNN-LSTM model with both attention mechanisms and gestational age encoding removed. These formed ablation experiments to validate the performance of the ensemble model. All models were trained and tested under identical conditions. Quantitative results are presented in Table 6. Each model employed a series of evaluation metrics based on confusion matrices to calculate diagnostic efficiency, with optimal results highlighted in bold black.

Table 6. Comparison of results from different models

Model	Acc	Recal	Pre	Spe	NF1	AF1	MCC	AUC	APT
MACLM	97.55%	98.61%	97.71%	95.46%	96.91%	95.06%	93.27%	0.9689	0.0234
LSTM	85.15%	96.12%	82.75%	92.95%	89.23%	78.31%	69.05%	0.8781	0.0209
SVM	75.04%	85.63%	78.70%	65.46%	82.43%	59.42%	41.89%	0.7208	0.2073
RF	88.25%	93.88%	89.54%	84.99%	91.58%	80.26%	72.11%	0.8726	0.0011
DNN	68.15%	79.05%	73.47%	56.81%	76.24%	52.67%	29.21%	0.6514	0.0041
GRU	84.06%	93.83%	83.26%	86.13%	88.18%	75.48%	65.23%	0.8469	0.0106
IT [31]	89.84%	95.21%	88.74%	88.52%	92.76%	82.20%	76.33%	0.9150	0.0461
MRT [32]	92.77%	95.81%	89.53%	88.60%	93.11%	83.06%	79.51%	0.9225	0.0714
CNN-LSTM	86.61%	97.86%	84.02%	94.52%	90.45%	77.96%	71.00%	0.8926	0.0156
ACNNLSTM	89.93%	97.44%	88.92%	92.93%	92.97%	82.08%	76.10%	0.9091	0.0188

Comparative experiments demonstrate that the MACLM model achieved outstanding performance on the test set: accuracy of 97.55%, recall of 98.61%, precision of 97.71%, specificity of 95.46%, NF1 score of 96.91%, AF1 of 95.06%, MCC of 93.27%, AUC of 0.9689, and ATP of 0.0234 seconds. The results conclusively demonstrate that the proposed model not only achieves exceptionally high classification accuracy in fetal health monitoring data but also exhibits strong recognition capabilities for both positive and negative categories. Its overall discrimination ability is remarkably robust, proving that MACLM possesses strong generalization capabilities and robustness in practical applications. Both precision and recall rates remain at high levels, further demonstrating that MACLM can reduce false positives and false negatives while accommodating samples across different categories, ensuring the reliability, practicality, and speed of classification results.

3.3 Comparison of fetal health monitoring data classification results across different models using gestational age position coding

Gestational age positional encoding is applied to two traditional machine learning models, RF and SVM, as well as three deep learning models—LSTM, DNN (Deep Neural Network), and GRU—to validate the impact of positional encoding on model classification performance. The results are shown in Figure 3.

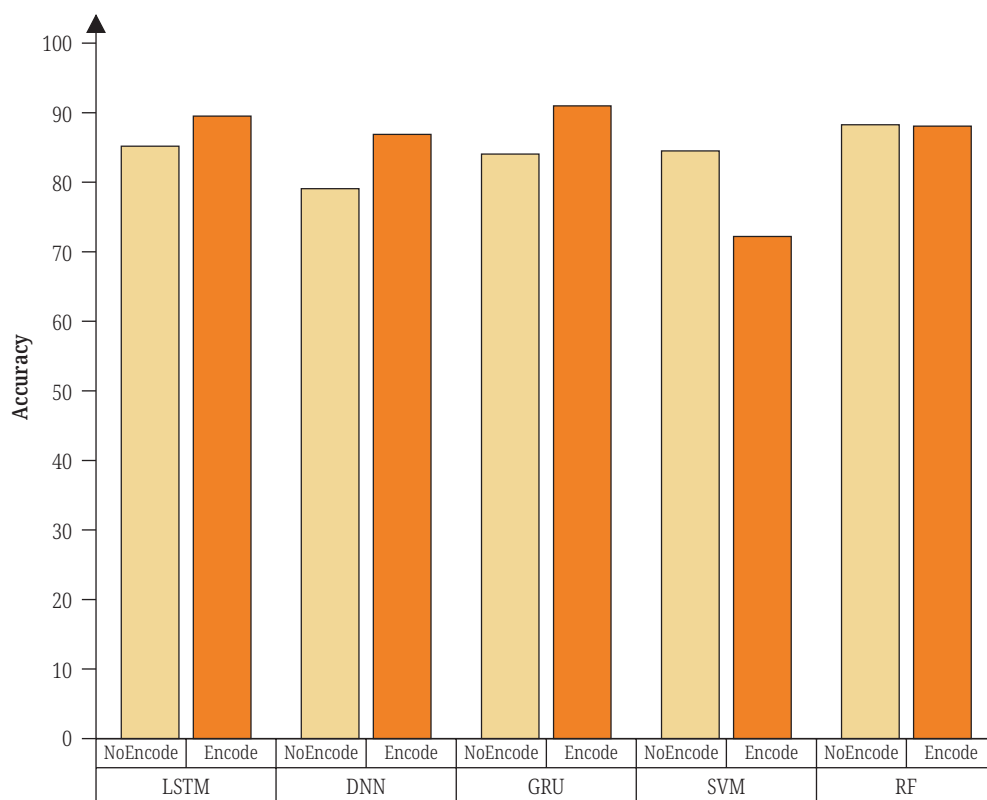


Fig. 3. Comparison of accuracy with or without position coding

To further validate the feasibility of incorporating positional encoding into deep learning models, we computed the loss for each deep learning model. The results are shown in Table 7.

Table 7. The performance improvement effect of position coding in different models

Model	Position Code	Loss	Decline in Losses
GRU	No	0.3927	46.05%
	Yes	0.2119	
LSTM	No	0.3991	45.97%
	Yes	0.2156	
DNN	No	0.6425	50.75%
	Yes	0.3165	

4 DISCUSSION

This study proposes a multi-modal attention neural network model for classifying fetal health status and constructs a multimodal dataset comprising 22,813 samples for model training and validation. We proposed embedding gestational age into fetal heart rate characteristics through positional encoding. However, the scarcity of abnormal category data remains an urgent issue that artificial intelligence must address in applications for monitoring fetal health status.

The results of the comparative tests show that MACLM significantly outperforms other models in the confusion matrix-based performance evaluation index. This overall advantage stems from the introduction of the maternal age, gestational age coding mechanism, and self-attention mechanism in the model design to dynamically regulate the different importance of multimodal features so that the model can more adequately capture the pattern of change of fetal heart features with gestational age and the differences in maternal age. The median time for a physician to confirm an electrocardiogram (ECG) is 2.8 minutes [33]. An experienced radiologist takes an average of 30 seconds to interpret a chest X-ray [34]. These benchmarks indicate that even with rapid interpretation, human analysis of fetal health monitoring data requires tens of seconds to several minutes. The MACLM model proposed in this study requires only 0.0234 seconds to interpret 100 fetal health monitoring data points. MACLM demonstrates that its efficiency far surpasses manual interpretation, significantly reducing labor costs and improving diagnostic timeliness for fetal health monitoring.

Evaluate the contribution of each component in the MACLM model using ablation experiments. Compared to CNN-LSTM, the introduction of the attention mechanism resulted in improvements across all metrics except specificity. It is evident that the attention mechanism enables the model to focus on more significant fetal heart rate features, overcoming the limitation of the CNN-LSTM network architecture that disregards the varying importance of features. This enhancement significantly improves model performance. This result validates the application potential and practical value of attention mechanisms in analyzing fetal health status. Due to the predominance of normal samples in the dataset, the model prioritized identifying healthy fetuses, resulting in a decrease in specificity to 92.93%. After incorporating gestational age position encoding, the model's performance achieved a significant improvement. The accuracy rate improved from 89.93% to 97.55%, the AUC increased from 0.9091 to 0.9689, and key metrics such as F1 score and MCC also showed significant progress. This enhancement primarily stems from incorporating gestational age as supplementary physiological background information. This enables the model to dynamically adjust feature calculations based on varying gestational ages, thereby better distinguishing samples with different health statuses and yielding more accurate and reliable classification results. This result also demonstrates that the synergistic interaction between fetal age position encoding and attention mechanisms further enhances the model's discriminative power and generalization capability.

As shown in Figure 3, after introducing positional encoding, the performance of the traditional machine learning model SVM declined significantly, while RF remained largely unaffected. This is because SVM treats each sample as a static point in feature space, assuming no temporal relationships exist between features by default. When position encoding is applied to SVM, the originally approximately separable dataset is infused with complex nonlinear features, disrupting the original feature distribution. RF, however, does not rely on linear relationships between features. When constructing decision trees for node splitting, each tree seeks only the optimal splitting feature. This characteristic makes it difficult for RF to accurately capture dynamic patterns over the temporal dimension. Position encoding disperses gestational age information across multiple feature dimensions, causing each individual encoded feature to exhibit neither linear growth nor monotonic variation. Consequently, the newly added gestational age encoding features neither introduce significant patterns into the original data nor effectively enhance the information gain capacity of decision tree nodes. Thus, they fail to substantially impact the overall model performance.

In contrast, deep learning models demonstrated further improvements in classification accuracy after incorporating positional encoding. This phenomenon indicates that for data processing tasks involving temporal characteristics, positional encoding of gestational age provides models with richer, more dynamic temporal information. This enables models to more accurately capture changes in fetal heart characteristics during the developmental process.

As shown in Table 7, the results indicate that after incorporating gestational age position encoding, the training loss of all deep learning models decreased significantly. This was particularly evident in temporal models such as GRU and LSTM, which exhibited lower loss values. Even models such as DNN, lacking time series modeling capabilities, were able to effectively capture information differences corresponding to different gestational ages after adding position encoding. This enabled them to learn more complex nonlinear relationships, significantly improving performance.

The results indicate that when MACLM is applied to interpreting fetal health status monitoring data, it demonstrates high accuracy and reliability. At present, although artificial intelligence diagnostic technology has not yet been widely adopted, developing a neural network model capable of comprehensively considering multiple characteristics and promptly assessing fetal health status is of great significance. In actual clinical practice, the assessment of fetal health status often relies on specialists for interpretation. This leads to situations where timely and accurate information about fetal health cannot be obtained in areas with scarce medical resources or when pregnant women are at home. Therefore, MACLM has important applications in assisting physicians with fetal status interpretation and in telemedicine.

Neural network models hold broad application prospects in fetal health monitoring; gestational age information is crucial for the model's assessment of fetal health status. Future research may consider incorporating more multimodal data to develop more comprehensive models for fetal health monitoring.

For the first time, a CNN-LSTM network-based time series modeling approach was applied in fetal health monitoring, jointly extracting local temporal patterns and long-term dependencies from fetal health monitoring data. Additionally, an attention mechanism was applied to process multimodal features. Moreover, fetal age is encoded into fetal heart rate characteristics using position-based coding, enabling the model to detect changes in fetal heart rate characteristics caused by fetal age. This study also has some limitations. For example, factors such as the mother's medical history (e.g., gestational diabetes, previous adverse pregnancy outcomes, hypertension, etc.) were not included. Furthermore, despite efforts to enhance the model's recognition capability for minority classes by assigning higher weights to outlier cases in the loss function, data imbalance remains a significant challenge.

The MACLM model proposed in this study is primarily intended as a decision-support tool for physicians in fetal heart rate monitoring analysis, rather than as a substitute for physicians in making final diagnoses. In the future, the model can be integrated into clinical workstation software, hospital local server interfaces, or web-based auxiliary interpretation modules, enabling physicians to obtain auxiliary classification results after fetal heart rate monitoring data is uploaded. It holds potential for future deployment on standard clinical workstations, hospital local servers, secure cloud platforms, or general-purpose edge devices. It should be noted that this study has not yet conducted real-world clinical deployment testing. Further validation is required when applying the model to actual medical settings, in accordance with medical data security, privacy protection, access control, ethical review, and relevant medical software regulatory requirements. At the same time, this study

primarily focuses on evaluating model performance based on electronic fetal heart rate monitoring data from pregnant women in China. Therefore, the results presented here primarily reflect the classification performance of the MACLM within this specific data context. Validation of the model's generalizability and clinical validation across other regions, populations, and healthcare systems—which involve differences in data sources, collection conditions, and population distributions—will be addressed in future research.

5 CONCLUSIONS

This study proposes a MACLM model for classifying fetal health status and constructs a multimodal dataset comprising 22,813 samples. We innovatively propose embedding gestational age into fetal heart rate features via positional encoding. Model comparison experiments demonstrate that this method can increase the accuracy of time-series neural network models by up to 7%. Ablation studies show that combining attention mechanisms with convolutional neural networks for feature enhancement and extraction effectively improves model performance. By accounting for the effects of gestational age and maternal age, the MACLM model—which employs attention mechanisms, a CNN-LSTM architecture combined with an MLP—achieved an accuracy of 97.55%, a recall of 98.61%, a precision of 97.76%, a specificity of 95.46%, and F1 scores of 96.91% and 95.06% for normal and abnormal cases, respectively, an MCC of 93.27%, and an AUC of 0.9689, with a processing time of just 0.0234 seconds for 100 data points. This demonstrates that MACLM offers strong classification performance and inference efficiency and can serve as a useful reference for physicians interpreting fetal heart rate monitoring results. Future work will further refine the model and evaluate the operational efficiency, stability, and clinical applicability of MACLM on cloud platforms and edge AI devices.

6 ACKNOWLEDGMENTS

6.1 Funding

This study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

6.2 Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

6.3 Research data

Data link: <https://www.ncmi.cn/phda/dataDetails.do?id=CSTR:17970.11.A001U.202204.282.V1.0>

7 REFERENCES

- [1] D. Hutter, J. Kingdom, and E. Jaeggi, “Causes and mechanisms of intrauterine hypoxia and its impact on the fetal cardiovascular system: A review,” *International Journal of Pediatrics*, vol. 2010, pp. 401323–401323, 2010. <https://doi.org/10.1155/2010/401323>
- [2] W.H.O, “Ending preventable maternal mortality (EPMM): A renewed focus for improving maternal and newborn health and well-being,” World Health Organization, Geneva, Switzerland, 2021. <https://www.who.int/publications/i/item/9789240040519> [Accessed: Sep. 14, 2025].
- [3] Department of Planning, Development and Information Technology, National Health Commission of the People's Republic of China, “2022 Statistical bulletin on the development of China's health and wellness sector,” *Chinese Journal of Viral Diseases*, vol. 14, no. 4, pp. 317–325, 2024. <https://doi.org/10.16505/j.2095-0136.2024.0066>
- [4] R. M. Grivell, Z. Alfirevic, G. M. Gyte, and D. Devane, “Antenatal cardiotocography for fetal assessment,” *Cochrane Database of Systematic Reviews*, no. 12, Art. no. CD007863, 2012. <https://doi.org/10.1002/14651858.CD007863.pub3>
- [5] L. Hruban *et al.*, “Agreement on intrapartum cardiotocogram recordings between expert obstetricians,” *Journal of Evaluation in Clinical Practice*, vol. 21, no. 4, pp. 694–702, 2015. <https://doi.org/10.1111/jep.12368>
- [6] G. D. Jones, W. R. Cooke, M. Vatish, and C. W. G. Redman, “Computerized analysis of antepartum cardiotocography: A review,” *Maternal-Fetal Medicine*, vol. 4, no. 2, pp. 130–140, 2022. <https://doi.org/10.1097/FM9.000000000000141>
- [7] F. Salvetti, B. Bertagni, R. Gardner, and J. Rudolph, “AI-powered avatars in medical education: Advancing virtual coaching for clinical readiness,” *International Journal of Advanced Corporate Learning (ijAC)*, vol. 19, no. 2, pp. 80–102, 2026. <https://doi.org/10.3991/ijac.v19i2.58755>
- [8] A. M. Khataan, G. Surendra Bapat, S. Rahaman, and A. Kumar, “Designing AI-enabled engineering courses: E-AIP framework for learning outcomes, process evidence, and integrity,” *International Journal of Engineering Pedagogy (ijEP)*, vol. 15, no. 7, pp. 6–15, 2025. <https://doi.org/10.3991/ijep.v15i7.59045>
- [9] P. Garcia-Canadilla, S. Sanchez-Martinez, F. Crispi, and B. Bijmens, “Machine learning in fetal cardiology: What to expect,” *Fetal Diagnosis and Therapy*, vol. 47, no. 5, pp. 363–372, 2020. <https://doi.org/10.1159/000505021>
- [10] M. E. O'Sullivan, E. C. Considine, M. O'Riordan, W. P. Marnane, J. M. Rennie, and G. B. Boylan, “Challenges of developing robust AI for intrapartum fetal heart rate monitoring,” *Frontiers in Artificial Intelligence*, vol. 4, Art. no. 765210, 2021. <https://doi.org/10.3389/frai.2021.765210>
- [11] S. Dash, J. G. Quirk, and P. M. Djuric, “Fetal heart rate classification using generative models,” *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 11, pp. 2796–2805, 2014. <https://doi.org/10.1109/TBME.2014.2330556>
- [12] 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. <https://doi.org/10.1109/TBME.2006.872814>
- [13] N. Chamidah and I. Wasito, “Fetal state classification from cardiotocography based on feature extraction using hybrid K-Means and support vector machine,” in *2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, Depok, Indonesia, 2015, pp. 37–41. <https://doi.org/10.1109/ICACSIS.2015.7415166>
- [14] J. Feng *et al.*, “Effective techniques for intelligent cardiotocography interpretation using XGB-RF feature selection and stacking fusion,” in *2021, IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Houston, TX, USA, 2021, pp. 2667–2673. <https://doi.org/10.1109/BIBM52615.2021.9669694>

- [15] Y. Zhang and Z. Zhao, "Fetal state assessment based on cardiotocography parameters using PCA and AdaBoost," in *2017, 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, Shanghai, China, 2017, pp. 1–6. <https://doi.org/10.1109/CISP-BMEI.2017.8302314>
- [16] K. Barnova, R. Martinek, R. Vilimkova Kahankova, R. Jaros, V. Snasel, and S. Mirjalili, "Artificial intelligence and machine learning in electronic fetal monitoring," *Archives of Computational Methods in Engineering*, vol. 31, no. 5, pp. 2557–2588, 2024. <https://doi.org/10.1007/s11831-023-10055-6>
- [17] J. Piri, P. Mohapatra, and R. Dey, "Multi-objective ant lion optimization based feature retrieval methodology for investigation of fetal wellbeing," in *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, 2021, pp. 1732–1737. <https://doi.org/10.1109/ICIRCA51532.2021.9544860>
- [18] G. Georgoulas, P. Karvelis, J. Spilka, V. Chudáček, C. D. Stylios, and L. Lhotská, "Investigating pH based evaluation of fetal heart rate (FHR) recordings," *Health and Technology*, vol. 7, nos. 2–3, pp. 241–254, 2017. <https://doi.org/10.1007/s12553-017-0201-7>
- [19] C. Ricciardi, G. Improta, F. Amato, G. Cesarelli, and M. Romano, "Classifying the type of delivery from cardiotocographic signals: A machine learning approach," *Computer Methods and Programs in Biomedicine*, vol. 196, Art. no. 105712, 2020. <https://doi.org/10.1016/j.cmpb.2020.105712>
- [20] A. S. Tarawneh, A. B. Hassanat, G. A. Altarawneh, and A. Almuhaimeed, "Stop oversampling for class imbalance learning: A review," *IEEE Access*, vol. 10, pp. 47643–47660, 2022. <https://doi.org/10.1109/ACCESS.2022.3169512>
- [21] National Research Institute for Family Planning, "Database of perinatal maternal and infant health monitoring and AI risk early warning in China," Population Health Data Archive (PHDA), 2022. <https://doi.org/10.12213/11.A001U.202204.282.V1.0>
- [22] W. G. Madow, "On the theory of systematic sampling, II," *The Annals of Mathematical Statistics*, vol. 20, no. 3, pp. 333–354, 1949. <https://doi.org/10.1214/aoms/1177729988>
- [23] American College of Obstetricians and Gynecologists, "Antepartum fetal surveillance," *Obstetrics & Gynecology*, vol. 137, no. 6, pp. e116–e127, 2021. <https://doi.org/10.1097/AOG.0000000000004410>
- [24] A. Vaswani *et al.*, "Attention is all you need," in *31st Annual Conference on Neural Information Processing Systems (NIPS)*, Long Beach, CA, 2017, pp. 5998–6008.
- [25] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. <https://doi.org/10.1109/5.726791>
- [26] H. Chen, K. Wang, M. Zhao, Y. Chen, and Y. He, "A CNN-LSTM-attention based seepage pressure prediction method for Earth and rock dams," *Scientific Reports*, vol. 15, no. 1, Art. no. 12960, 2025. <https://doi.org/10.1038/s41598-025-96936-1>
- [27] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.
- [28] J. Ke, H. Zheng, H. Yang, and X. Chen, "Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach," *Transportation Research Part C: Emerging Technologies*, vol. 85, pp. 591–608, 2017. <https://doi.org/10.1016/j.trc.2017.10.016>
- [29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations (ICLR)*, San Diego, CA, USA, *arXiv preprint arXiv.1412.6980*, 2015. <https://doi.org/10.48550/arXiv.1412.6980>
- [30] J. B. Brown, "Classifiers and their metrics quantified," *Molecular Informatics*, vol. 37, nos. 1–2, Art. no. 1700127, 2018. <https://doi.org/10.1002/minf.201700127>

- [31] H. Ismail Fawaz *et al.*, “InceptionTime: Finding AlexNet for time series classification,” *Data Mining and Knowledge Discovery*, vol. 34, no. 6, pp. 1936–1962, 2020. <https://doi.org/10.1007/s10618-020-00710-y>
- [32] C. W. Tan, A. Dempster, C. Bergmeir, and G. I. Webb, “MultiRocket: Multiple pooling operators and transformations for fast and effective time series classification,” *arXiv preprint arXiv:2102.00457*, 2021. <https://doi.org/10.48550/arXiv.2102.00457>
- [33] N. A. Villarroel, C. J. Houghton, S. C. Mader, K. E. Poronsky, A. L. Deutsch, and T. J. Mader, “A prospective analysis of time to screen protocol ECGs in adult emergency department triage patients,” *The American Journal of Emergency Medicine*, vol. 46, pp. 23–26, 2021. <https://doi.org/10.1016/j.ajem.2020.03.025>
- [34] D. Manning, S. C. Ethell, and T. Crawford, “Eye-tracking AFROC study of the influence of experience and training on chest x-ray interpretation,” in *Proc. SPIE 5034, Medical Imaging 2003: Image Perception, Observer Performance, and Technology Assessment*, 2003. <https://doi.org/10.1117/12.479985>

8 APPENDIX

8.1 Model parameters

Model parameters as shown in Table A:

Table A. Model parameter

Structure Name	Parameters
First CNN layer	Number of filters = 64; convolution kernel size = 2; activation function = relu
First maximum pooling layer	Pooling window size: 2
Second CNN layer	Number of filters = 32; convolution kernel size = 2; activation function = relu
Second-largest pooling layer	Pooling window size: 2
First LSTM layer	Number of units: 32; Regularizer = 0.01;
Second LSTM layer	Number of units:16; Regularizer = 0.01;
First and second Drout layers	The forgetting rates were all 0.2
Batch	10
first full connectivity layer	Number of units: 16; Activation function: relu.
The second fully connected layer	Number of units: 4.
Output layer	Activation function: sigmoid

9 AUTHORS

Hao Wang is with the College of Electrical and Automation Engineering, Nanjing Normal University, Nanjing, China.

Yin Zhao is with the College of Electrical and Automation Engineering, Nanjing Normal University, Nanjing, China.

Yajie Gu is with the College of Electrical and Automation Engineering, Nanjing Normal University, Nanjing, China.

Fengliang Huang is with the College of Electrical and Automation Engineering, Nanjing Normal University, Nanjing, China (E-mail: huangfengliang@njnu.edu.cn).