IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY SECTION

Received 30 December 2025, accepted 25 January 2026, date of publication 9 February 2026, date of current version 18 February 2026.

Digital Object Identifier 10.1109/ACCESS.2026.3662271

RESEARCH ARTICLE

Classification of Hypoxic Ischemic Encephalopathy (HIE) in Neonates Using ECG and a Transformer-Based Model

KIMIA REZAEI[1,2], SEAN R. MATHIESON[2], GORDON LIGHTBODY[1,2], GERALDINE B. BOYLAN[2,3], AND WILLIAM P. MARNANE[1,2]

[1]Electrical and Electronic Engineering, School of Engineering, University College Cork, Cork, T12K8AF Ireland
[2]INFANT Research Centre, University College Cork, Cork, T12K8AF Ireland
[3]Department of Pediatrics and Child Health, University College Cork, Cork, T12K8AF Ireland

Corresponding author: Kimia Rezaei (121101892@umail.ucc.ie)

This work was supported by Taighde Eireann—Research Ireland under Grant 19/FFP/6782.

ABSTRACT Hypoxic-Ischemic Encephalopathy (HIE) is caused by inadequate blood flow and oxygen delivery to the brain, typically occurring around the time of birth. In the Neonatal Unit, neonates with moderate to severe HIE can be treated by employing therapeutic hypothermia. The increasing use of electroencephalogram (EEG) for supporting clinical decisions, such as initiating therapeutic hypothermia, seizure detection and prognostication, led to growing interest in developing automated systems to assist clinicians in grading HIE-EEG. Electrocardiogram (ECG) signals offer a potential alternative to EEG, being widely accessible and routinely recorded. This study investigates three different approaches for binary classification of infant HIE, utilizing a large ECG dataset from neonates. The traditional method involves deriving the heart rate (HR) from the ECG signal, followed by feature extraction in both time and frequency domains, and classification using classical machine learning techniques such as XGBoost. In contrast, two new approaches are proposed: one extracts spectrograms directly from the ECG signal to input into a convolutional neural network (CNN), while the other uses the raw ECG signal as input to transformer-based models. This paper proposes Trend-Augmented Multiscale Tokenization Transformer (TMFormer), a transformer-based model that introduces a tokenization approach derived from raw ECG signal attributes, capturing spatial patterns at multiple scales and tracking the signal's long-term progression. Additionally, the transformer's output on the ECG is analyzed in relation to the corresponding NN intervals. The proposed TMFormer using raw ECG outperform a standard transformer and other models using HRV based features and ECG spectrogram. By eliminating the ECG preprocessing step and leveraging more robust deep learning models, computational overhead can be reduced, enabling real-time processing, and enhancing the practicality of supporting clinicians in desicion-making regarding the HIE binary category even in resource-limited hospital settings.

INDEX TERMS ECG, HIE, multiscale tokenization, spectrogram, transformer, neonatal.

I. INTRODUCTION

Neonatal encephalopathy is a clinical syndrome of neurological dysfunction in term and near-term infants. It is characterized by altered consciousness and abnormal tone and reflexes. When it is caused by a perinatal hypoxic-ischemic event, it is termed Hypoxic ischaemic encephalopathy (HIE). HIE is a leading cause of neonatal mortality and long-term complications. HIE can result in a variety of neurological and developmental conditions, including epilepsy, cerebral palsy, visual and hearing impairments, learning disabilities, and behavioural and emotional disorders [1].

HIE is clinically graded as normal, mild, moderate, or severe shortly after birth using Sarnat scoring [2], [3]. Neonates with moderate to severe HIE can be treated with therapeutic hypothermia (TH), which improves neurological

The associate editor coordinating the review of this manuscript and approving it for publication was Cheolsoo Park.

TABLE 1. Classification of EEG background activity. IBI indicates interburst interval.

Grade	Description
0	Normal: Continuous background pattern with normal physiologic features such as anterior slow waves
1	Normal/mild abnormalities: Continuous background pattern with slightly abnormal activity (eg, mild asymmetry, mild voltage depression, or poorly defined SWC)
2	Moderate abnormalities: Discontinuous activity with IBI of <10 s, no clear SWC, or clear asymmetry or asynchrony
3	Major abnormalities: Discontinuous activity with IBI of 10–60 s, severe attenuation of background patterns, or no SWC
4	Inactive: Background activity of <10 muV or severe discontinuity with IBI of >60 s

outcomes and reduces mortality in many infants [4]. TH is a time sensitive treatment in which the infant is cooled within six hours of birth to a target core body temperature of 33.5°C for 72 hours. After this period, the infant is gradually rewarmed to normal body temperature over 7 to 14 hours. Early and accurate assessment of HIE severity is essential in the Neonatal Intensive Care Unit (NICU) to support immediate diagnosis and treatment. Currently, continuous electroencephalography (cEEG) or amplitude-integrated EEG (aEEG) is commonly used to help assess HIE. Background EEG can be graded for severity, such as the scheme proposed by Murray et al. [5], EEG background is graded to 5 grades of normal, normal/mild, moderate, major and inactive as shown in Table 1.

The EEG grades of normal and normal/mild (grades 0 and 1) are generally seen in infants with a clinically mild encephalopathy who do not receive TH. Moderate, major and inactive EEG grades (grades 2, 3 and 4) are usually seen in those infants with a clinically moderate to severe encephalopathy and requiring TH. However EEG and Sarnat grades do not always align and in infants recruited to the TOBY cooling trial [6], an abnormal aEEG (i.e., discontinuous activity, equivalent to moderate or worse EEG grade) was a requirement for cooling.

EEG/aEEG, if available, is considered a valuable adjunct to neurological examination in determining the HIE-EEG severity. However, EEG monitoring presents significant challenges in the neonatal setting [7]. Interpreting neonatal EEG requires specialized expertise, which may not always be available in the NICU, and the high cost and complexity of EEG interpretation in the immediate postnatal period can hinder timely diagnosis. These combined limitations have driven ongoing research into automated and accessible methods for grading HIE-EEG to support clinicians in timely clinical assessment.

One potential alternative is the electrocardiogram (ECG), which is widely accessible, routinely monitored in neonates, and feasible even in smaller neonatal units and resource-limited settings. By analysing and processing ECG signals, valuable diagnostic information can be extracted. However, research on ECG-based methods regarding the HIE assessment remains limited, highlighting the need for further investigation in this area. Existing research focused on using EEG signals [8], [9], [10], [11], [12], [13] or analysing ECG signals indirectly through heart rate (HR) extraction [14], [15], [16], [17]. In general, the findings showed that as the severity of HIE increased, most HRV measures decreased. Methods based on heart rate variability (HRV) require several preprocessing steps, including detecting the QRS complex, extracting NN intervals, and computing features. Traditional HR calculation techniques, such as the Pan-Tompkins algorithm [18], are more computationally demanding. Additionally, accurately extracting NN intervals continues to be a challenge, limiting the practical effectiveness of these methods in clinical settings [19], [20].

This study introduces two alternative approaches for analysing ECG signals in neonates with varying degrees of HIE-EEG severity, aimed to overcome the limitations of traditional methods. These systems are developed to classify ECG segments into two categories: Category 0, comprising normal and normal/mild HIE-EEG cases, and Category 1, including moderate, major, and inactive HIE-EEG cases. The proposed systems aim to support clinicians in the timely assessment of HIE in neonates. To the best of our knowledge, this is the first study to utilize raw ECG signals and ECG spectrograms for this purpose. Our main contributions are summarized as follows:

- The first proposed method directly extracts time-frequency information from the ECG signal in the form of a spectrogram and uses a 2D convolutional neural network (CNN) for classification. This eliminates the HR calculation processes, addressing their limitations. This experiment trained and tested on a large-scale neonatal dataset collected shortly after birth across multiple clinical settings, enabling a more comprehensive analysis.
- The second proposed method directly utilizes raw ECG signals from the same dataset, eliminating preprocessing steps like HR extraction and spectrogram generation. This approach preserves the temporal information of the data while reducing computational overhead, enabling real-time processing and enhancing its practicality for clinical use. Additionally, for classification, we propose TMFormer (Trend-Augmented Multiscale Tokenization Transformer), a transformer-based model that introduces a tokenization approach derived from the ECG signal attributes, capturing spatial patterns at multiple scales and tracking the signal's long-term progression.
- Attention rollout is utilized to investigate the relationship between variations in NN intervals, reflecting classical heart rate measures, and the regions of the ECG signal identified as most important by proposed TMFormer model.

II. METHOD

A. DATASET

In this work, two datasets called ANSeR1 and ANSeR2 were used [21], [22]. A multichannel EEG, as well as a single or two lead ECG where available, was recorded from

neonates at risk of seizures shortly after birth and was continued for up to 72 hours where possible. All infants had a gestational age of at least 36 weeks. The signals were recorded in eight tertiary-level neonatal intensive care units across Europe (Ireland, United Kingdom, Sweden and Netherlands) using Nihon Kohden Neurofax (EEG-1200), NicoletOne ICU Monitor (Natus, Middleton, WI, USA) or XLTek (Natus).

A subset of EEG epochs from infants with HIE was graded at five different postnatal ages (6, 12, 24, 36, and 48 hours), with each epoch lasting 1 hour. The grades assigned to background EEG are based on the work by Murray et al. [5], in which they graded the EEGs as normal, normal/mild, moderate, major and inactive. For each neonate, there was more data available but unlabelled. To enhance model performance, additional 1-hour epochs were extracted from the intervals between these five time points and weakly labelled based on the preceding and following epoch classifications. The weak labelling was only employed on the ANSeR2 dataset which was used as a training dataset. The EEG-based labels were applied to the corresponding time segments of the ECG signal for model training. For binary classification, 1-hour ECG epochs were categorized into two groups. Category 0 comprised normal and normal/mild HIE-EEG cases, while Category 1 included moderate, major, and inactive HIE-EEG cases.

Recordings from patients without an ECG signal channel were entirely excluded from the analysis. For patients with available ECG data, 1-hour epochs that were heavily corrupted by noise were removed. Consequently, 58 out of 91 neonates from ANSeR1 and 73 out of 90 neonates from ANSeR2 were included in this study. The ANSeR1 dataset includes 216 1-hour epochs (130 class 0, 86 class 1). The ANSeR2 dataset comprises 254 strongly labelled and 1286 weakly labelled 1-hour epochs (792 class 0, 748 class 1). All ECG recordings were originally sampled at 256 Hz. The ANSeR1 and ANSeR2 datasets are fully patient-independent, and no neonate appears in both datasets.

B. CLASSIFICATION USING HRV FEATURES AND XGBoost AND MLP

As a baseline for comparison with the techniques introduced in this work, a classical approach involving extracting heart rate (HR) signals from the ECG signal, computing HR features, and employing binary classification using XGBoost and multilayer perceptron (MLP) is implemented.

The Pan-Tompkins algorithm was employed to detect R-peaks in the ECG signal and extract NN intervals [18]. While this approach is generally reliable, it can still result in inaccuracies or missed peaks during QRS detection. The main reason for these issues is that the method is optimized for detecting peaks in normal heart rhythms, but real-world ECG recordings are often contaminated with different types of noise and artifacts such as 50 Hz power line interference, baseline wander and muscle artifact. In HIE, fluctuations in heart rate can introduce additional

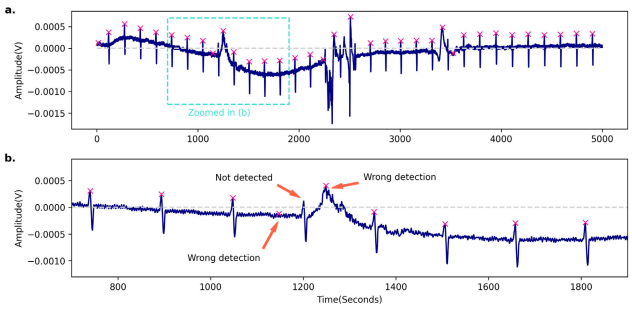


FIGURE 1. (a) An example of a noisy ECG signal where Pan-Tompkins missed some peaks and falsely detected others. (b) A zoomed-in segment of (a) showing missed and incorrectly detected peaks.

irregularities in the ECG signal, which further complicates accurate peak detection [23], [24]. These challenges can affect the precision of the values calculated during feature extraction. An example of peak detection on a noisy ECG signal segment using the Pan-Tompkins method is shown in Fig. 1. As depicted, some peaks were either missed or incorrectly identified during the signal spikes.

Here, ECG signal peaks were detected using the Pan-Tompkins method and NN intervals were calculated. Afterwards, the Hampel filter was used to remove outliers from NN interval data, preparing it for further feature extraction. In adults, the average resting heart rate (HR) typically falls between 60 and 100 BPM, while neonates generally have a higher HR. To achieve a balance between optimal HR data analysis and a sufficient duration for detecting abnormalities and signal variations, a 4-minute segment length was selected [25], [26]. An 80% overlap was also used to reduce information loss at segment boundaries, providing a more comprehensive representation of signal patterns. Features were then extracted from each 4-minute segment based on heart rate or NN intervals.

Sixteen frequency and time domain features were extracted from each segment [17]. Frequency domain features were derived by estimating the power spectral density (PSD) of the NN series using Welch's method, which requires resampling the NN sequences at uniform intervals. Therefore, a spline interpolation technique was employed, fitting low-degree polynomials to small subsets of the data. The four frequency domain features extracted include Very Low Frequency (VLF) power, Low Frequency (LF) power, and High Frequency (HF) power, corresponding to frequency bands of (0.01–0.04 Hz), (0.04–0.2 Hz), and (0.2–2 Hz), respectively, along with the LF/HF ratio. Extracted time domain features include the standard deviation of the NN series (SDNN), the mean of the NN interval (muNN), the root mean square of successive NN interval differences (RMSSD), kurtosis, and skewness of the NN distribution. Additionally, a geometric parameter, the triangular index, which represents the ratio between the number of NN intervals (NNI) and the maximum of the NNI histogram, was calculated. Nonlinear features such as Poincaré parameters SD1, SD2, and the SD2/SD1 ratio for self-similarity assessment, as well as Approximate

Entropy (ApEn), which measures the complexity and regularity of a stationary signal were also computed [27]. Lastly, the modified Cardiac Sympathetic Index (CSI) and the Cardiac Vagal Index (CVI) [28] were extracted.

Supervised classification was performed on 16 extracted features using two models: XGBoost (eXtreme Gradient Boosting) and a Multi-Layer Perceptron (MLP). For XGBoost model, a 10-fold cross validation was performed through grid search to find the best hyperparameters. The final selected parameters were a subsample ratio of 0.7, 200 estimators, a maximum depth of 4, a minimum child weight of 13, a learning rate of 0.01, and a column sampling ratio by tree of 0.5.

The MLP classifier applied for this study consists of three fully connected layers. ReLU activation function was used to introduce nonlinearity. The model takes 16 features as input, with the first layer configured to have 16 neurons. In the second layer, the number of neurons is increased to 64 to capture a broader range of information. The third layer reduces the number of neurons to 32, and the final output layer has 2 neurons, corresponding to the two classes. This architecture is designed to effectively balance model complexity and computational efficiency for an input vector of 16 features.

C. CLASSIFICATION USING SPECTROGRAMS AND CNN

This method is proposed to overcome the limitations and shortcomings of the heart rate calculation, feature extraction and traditional classification methods. In this method, the spectrogram is used to directly compute the time-frequency information of the ECG signal. Spectrograms are a widely used tool for signal representation, capturing the changes in a nonstationary signal's frequency content over time through consecutive Fourier transforms particularly in fields such as audio and speech recognition [29]. ECG spectrograms have been shown to be effective in cardiovascular disease classification [30], arrhythmia detection [31] and classifying the severity of tetanus [32]. However infections and arrhythmia detection may present a relatively simpler classification challenge via ECG compared to the classification of neurological conditions. Spectrogram-based representations enable the application of deep learning models such as CNNs, which offer significant advantages over classical methods in terms of accuracy. Unlike traditional approaches that rely on manual feature engineering, CNNs automatically learn features directly from raw data and are specifically designed to capture spatial patterns. CNNs exhibit translation invariance, allowing them to recognize shifts in input data and making them more resilient to noise.

Initially, a notch filter was applied to the ECG signals to eliminate power line interference. Following this, Low-frequency suppression was performed using a second-order high-pass Butterworth filter with a 0.5 Hz cutoff frequency to remove baseline wander. Spectrograms were generated for each 4-minute segment of the filtered ECG signal with

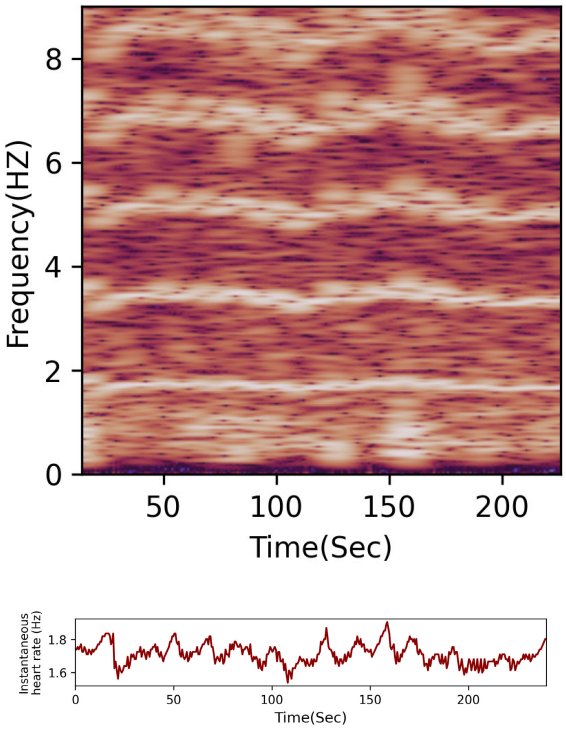


FIGURE 2. Upper image is the log-scaled spectrogram of a 4-minute ECG segment, cropped to a maximum frequency of 9 Hz, with a size of 253 × 213. Lower image is instantaneous heart rate of the same ECG segment extracted using Pan-Tompkins.

80% overlap similar to the HR feature extraction method. A 28-second Hamming window with a 27-second overlap was used for the transformation. To emphasize lower frequencies, the power spectral density of spectrogram was converted to a logarithmic scale and the spectrogram was then cropped to a maximum frequency of 9 Hz, capturing the primary harmonics of the ECG signal [33]. This processing reshaped the spectrogram from 3585 × 213 to a more balanced 253 × 213, optimizing it for CNN processing. The reshaped spectrogram is better suited for convolutional and pooling layers, which perform more efficiently when the aspect ratio is more uniform. Fig. 2 shows the log-scaled spectrogram and instantaneous hear rate for the same 4-minute ECG segment. The low-frequency components in the spectrogram correspond to the slow variations in heart rate, also evident in the heart rate signal.

The resulting spectrogram matrix is a 2D array where each element represents the log power spectral density of the signal at a given frequency and time. These matrices were standardized using the global mean and standard deviation to have zero mean and unit variance before being used as input for classification. The input tensor has a shape of 1 × 253 × 213 where the first dimension represents the number of channels, while the second and third correspond to the height and width, respectively. Given the simplicity of the input data, a shallow yet efficient CNN architecture was designed to optimize both performance and computational efficiency.

TABLE 2. Feature map dimensions at each layer of the proposed CNN for an input spectrogram matrix of size 1 × 253 × 213.

Layer Name	Output Shape (Channels × Height × Width)
Conv1	$8 \times 247 \times 207$
MaxPool1	$8 \times 123 \times 103$
Conv2	$16 \times 119 \times 99$
MaxPool2	$16 \times 59 \times 49$
Conv3	$32 \times 57 \times 47$
MaxPool3	$32 \times 28 \times 23$
Conv4	$64 \times 26 \times 21$
MaxPool4	$64 \times 13 \times 10$
Global Avg Pool	$64 \times 1 \times 1$

The proposed CNN architecture consists of four sequential convolutional blocks, each followed by batch normalization, ReLU activation, and max pooling to progressively extract hierarchical features while reducing spatial dimensions. The network begins with a convolutional layer that applies 8 filters of size 7×7 to the input, followed by a second layer with 16 filters of size 5×5, a third with 32 filters of size 3×3, and a final convolutional layer with 64 filters, also of size 3×3. All convolutional layers use a stride of 1 and no padding. To maintain low computational cost, the network begins with a small number of filters (8 here), since the number of filters increases in deeper layers. As the network deepens, this design increases the receptive field [34], defined as the region in the input that contributes to each feature, while also helping to mitigate overfitting. Max pooling with a 2×2 kernel size and batch normalization were applied after each block to downsample the feature maps. After the final convolutional stage, a global average pooling layer reduces the spatial dimensions to 1×1, resulting in a 64-dimensional feature vector. This vector is flattened and passed through a fully connected layer with 32 units, followed by batch normalization, ReLU activation, and dropout with a probability of 0.5. The final classification is performed using a dense layer with 2 output units, corresponding to the target classes. An overview of the proposed CNN architecture is depicted in Fig. 3. Table 2 provides a detailed summary of the changes in feature map dimensions and the number of channels at each stage of the network.

D. CLASSIFICATION USING RAW ECG SIGNAL AND TMFORMER

To eliminate additional preprocessing steps and reduce computational cost, this method utilizes raw ECG signals, preserving valuable temporal information that might be discarded during feature extraction or spectrogram calculation. Using raw ECG have been shown to be effective in areas such as in cardiac disease diagnostic [35] and arrhythmia diagnosis [36]. Transformers have proven to be highly effective in capturing long-range dependencies in data through their self-attention mechanism. This capability is particularly valuable for accurately analyzing the ECG patterns in neonates with HIE, where understanding the context from earlier or later parts of the signal, as well as the time differences between beats is essential for detecting subtle changes in the ECG signal. Unlike CNNs or classical methods, which often rely on signal representations or feature extraction (such as spectrograms or statistical features), transformers can process raw ECG time-series data directly. This allows the model to automatically learn the most relevant patterns from the signal, without the need for hand-crafted features.

To ensure both signal quality and computational efficiency, the ECG recordings were first filtered to remove common noise artifacts. A notch filter eliminated 50 Hz powerline interference, and a high-pass Butterworth filter with a 0.5 Hz cutoff removed baseline wander, as previously used for spectrogram calculation. Due to the quadratic complexity of self-attention, transformers become computationally expensive with long sequences. Shortening the sequence length optimizes memory usage and training speed, enabling larger batch sizes and deeper models. To achieve this, we down sampled the ECG signal from 256 Hz to 80 Hz. This is particularly effective since most of the meaningful components of ECG signals, such as the P, QRS, and T waves, lie below the 40 Hz range. By down sampling to 80 Hz, we preserve these critical features while eliminating unnecessary high-frequency noise, ensuring that the signal remains informative and computationally efficient. Further, the ECG signal was segmented into 4-minute windows with 80% overlap, consistent with the preprocessing used for the HR features and spectrogram calculation. Each segment was then standardized using the global mean and standard deviation to achieve zero mean and unit variance before being fed into the model.

1) OVERALL MODEL ARCHITECTURE

The overall design of the proposed model consists of three main components, a tokenizer, two transformer encoder blocks and a classification head, as illustrated in Fig 4. The tokenizer processes the ECG signal without explicit feature extraction, allowing the model to automatically learn and identify relevant patterns. The transformer encoder blocks then refine these learned representations, capturing complex temporal and frequency-based relationships in the ECG signal to enhance contextual understanding. Finally, the classification head analyzes the refined representations to predict the corresponding binary class (category 0 or category 1). Our code is available at https://github.com/kimiarezaei/TMFormer.

2) TREND-AUGMENTED MULTI-SCALE TOKENIZATION

Here, a novel Tokenization technique specialized for ECG is proposed. The first stage of the designed tokenization has a multi-scale architecture consisting of 2 layers. The first layer tokenizes the ECG signal into large segments with minimal redundancy. A large kernel is used to capture global patterns, such as overall waveform morphology and long-term dependencies. Thus, a convolutional layer with a large kernel of size 160 is used to extract tokens. The reason

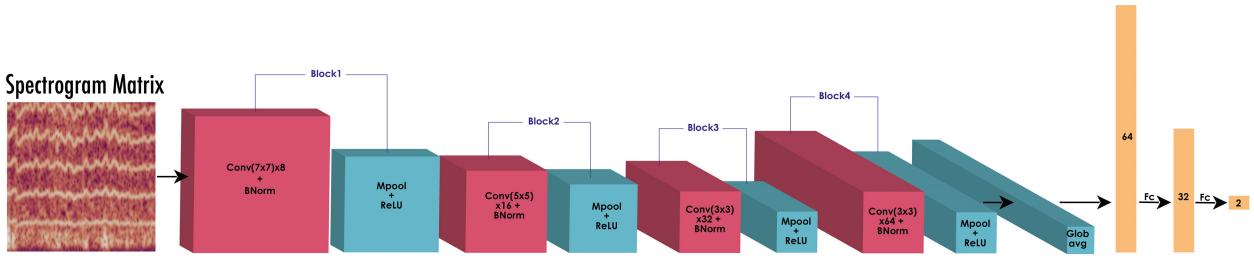


FIGURE 3. Architecture of the proposed CNN model. The input is a two-dimensional spectrogram matrix generated from a 4-minute ECG segment. The network comprises four convolutional blocks, each including a convolutional layer (Conv), batch normalization (BNorm), and max pooling (MPool). A global average pooling layer (Glob-Ave) is used to aggregate the spatial features, followed by a fully connected classification head consisting of two linear layers.

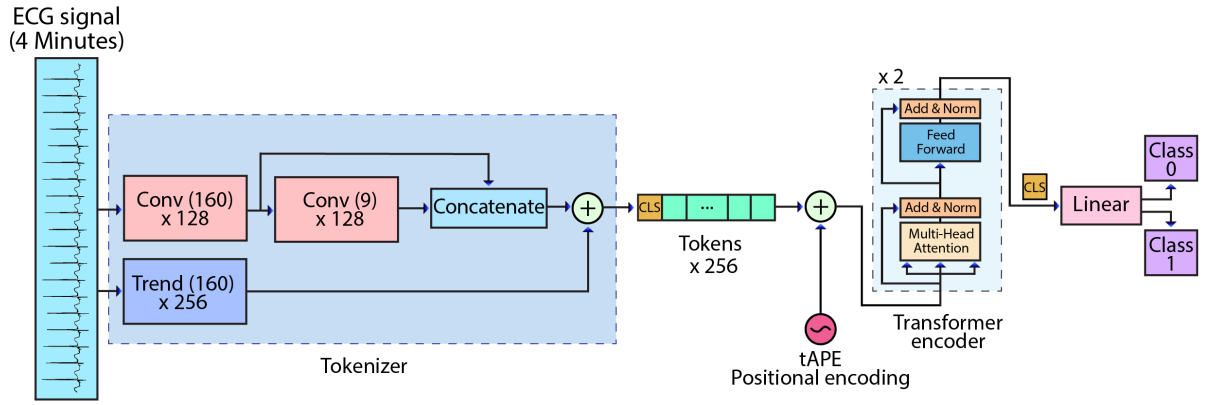


FIGURE 4. The proposed TMFormer model for binary classification. The input is a 4-minute raw ECG segment. The framework consist of a Trend-Augmented Multi-scale Tokenization, two transformer encoder blocks and a classification head.

for choice of size 160 is that it is equal to 2 seconds of ECG which normally contains four to five consecutive P, QRS, and T wave forms. A small overlap of 11 is used to ensure minimal redundancy while preserving continuity between segments. The 1D ECG data is tokenized into 128 tokens and each is then embedded into a 128-dimensional vector space. This transformation enriches the representation, enabling the model to capture more complex and informative features. The second layer apply a smaller kernel of size 9 on each token, extracting finer details from them. This allows the model to capture local features, such as peaks, slopes, and subtle variations in the signal. Additionally, the output of the second layer preserves both the number of tokens and the embedding dimension at 128, consistent with the first convolutional layer's output. Ultimately, the tokens extracted by both kernels are concatenated channel-wise, maintaining the number of tokens at 128 while increasing the embedding dimension to 256. This allows the model to learn both contextual and detailed features simultaneously. Additionally, the computational load is lower compared to processing the entire ECG with small kernels.

Time series decomposition is a technique that breaks down the time series into multiple components, each representing a distinct pattern or underlying structure [37]. One key component is the trend, which captures the long-term progression of the series. Since the trend component reflects changes in the signal's pattern over extended periods, and ECG pattern alterations are of interest in the context of HIE, analyzing the trend component of the ECG signal can be effective for detecting these changes. In this paper, a refined approach utilizing the ECG trend component in the second stage of the tokenization is proposed. This was achieved using an average pooling layer to extract the trend, which is then incorporated into the model's convolution output in a residual manner. This, enhances the model's ability to capture long-term variations in ECG signals, providing a more comprehensive representation of HIE effects on ECG. Here, the trend can be calculated as (1):

$$T_j = \frac{1}{k}\sum_{i=0}^{k-1} x_{s\times j+i} \tag{1}$$

The estimate of the trend T at position j, is obtained by averaging values of the ECG signal x, within k samples of i. x starts at index $s \times j$, where s is the stride. Here, k is set to 160, to match the kernel size of the first convolutional layer. A stride of 149 corresponding to an overlap of 11 samples is used, consistent with the overlap in the first layer. This ensures that the output contains exactly 128 tokens, allowing for proper alignment during residual addition. Finally, the residual trend calculation is performed across all 256 tokens dimensions.

Time Absolute Position Encoding (tAPE) [38] enhances traditional sinusoidal-based absolute position encoding,

which was initially developed for language models which usually use high embedding dimensions. Unlike the original approach, tAPE integrates sequence length into the frequency terms of sine and cosine functions, ensuring a smoother and more consistent dot product between positional embedding regardless of series length. By adjusting the frequency scaling with the embedding dimension, tAPE maintains distance awareness even in lower-dimensional embedding, which is mostly the case in time-series tasks. The original transformer frequency term (2) and the modified tAPE version (3) are presented below.

$$\omega_k^{original} = 10000^{-2k/d_{model}} \tag{2}$$

$$\omega_k^{tAPE} = \frac{\omega_k^{original} \times d_{model}}{L} \tag{3}$$

where L is the sequence length and d_{model} is the embedding dimension.

Integrating the proposed tokenizer as the first component of our architecture not only enables the capture of both local and global patterns within the raw ECG signal but also, reduces the sequence length, optimizing the data for tAPE position encoding while lowering computational complexity. Therefore, tAPE positional encoding is added to the extracted tokens by the proposed tokenizer, to pass to the encoder layer. Following this, a dropout rate of 0.3 is applied to improve regularization, preventing the model from over-relying on specific feature embedding and promoting a more comprehensive understanding of the entire 4-minute ECG segment.

3) TRANSFORMER ENCODER LAYERS

The model comprises two Transformer encoder layers, each containing layer normalization, multi-head self-attention, and a feedforward neural network (MLP). A classification token (CLS token) is prepended to the input sequence and serves as a summary representation for classification tasks. The attention mechanism uses 8 heads with a dropout rate of 0.2 to promote regularization. The MLP consists of two linear layers that expand the embedding dimension from 256 to 512, enhancing the model's capacity to learn complex representations. A GELU activation function is applied between the layers to introduce non-linearity.

4) CLASSIFICATION HEAD

After the Transformer encoders process the input sequence including the CLS token, the output corresponding to the CLS token is used for classification. This output is passed through a feedforward network comprising a layer normalization followed by a linear layer to produce the final class prediction.

5) EXPLAINABILITY WITH ATTENTION ROLLOUT

Attention rollout [39] multiplies attention weights across all transformer layers to track how information flows from the input tokens to the CLS token. This process shows how information from each token flows through the network and highlights which parts of the input the model attends to at each layer. It produces a cumulative attention vector that reflects the overall influence of each token on the final decision.

Here, the goal is to identify the portions of the ECG signal that contain the most relevant physiological features, which the model relies on to make its final decision. To achieve this, both the attention flow at each layer and the cumulative attention flow across all layers (attention rollout) are calculated. Since heart rate, which is derived from NN intervals, fluctuates during HIE, attention flow is compared with the corresponding NN intervals. This comparison explores how variations in NN intervals correspond to the regions of the ECG signal that the model attends to across layers.

III. EXPERIMENTS

Experiments were conducted to evaluate classification performance, using all applied methods, including XGBoost and MLP with HRV features, CNN with ECG spectrograms, and Transformer and TMFormer models with raw ECG signals.

Additionally, the proposed TMFormer model is compared to a transformer using a simple tokenization method, with a kernel size of 160 and the same number of parameters. TMFormer is also compared to an architecturally identical model in which the trend component is excluded during tokenization, to evaluate the contribution of trend information to overall performance.

In all methods, classifiers were trained on the ANSeR2 dataset and tested on the completely unseen ANSeR1 dataset, which consisted of unedited clinical data collected from multiple hospitals, to ensure generalization across domains. The ANSeR1 and ANSeR2 datasets include different patients, ensuring that no patient or associated records appear in both training and testing sets. For each 1-hour epoch, the predicted probabilities of all test samples were averaged, and this mean probability determined the category 0 or category 1 for that epoch.

XGBoost model was configured using optimized parameters detailed in section II-B. The MLP model weights were initialized using the Xavier uniform method. The model was trained using the cross-entropy loss function with a batch size of 1024 for a maximum of 150 iterations. Optimization was performed using the Adam optimizer with an initial learning rate of 0.0005, and a weight decay of 0.001 was applied for regularization. Internal cross-validation was used to monitor performance and enable early stopping.

For all other models (CNN, Transformer and TMFormer), the ANSeR2 dataset was randomly split into training (80%) and validation (20%) subsets for the purpose of loss optimization. Training was conducted with a batch size of 1024 for over 100 iterations, and was terminated early once the validation loss began to increase, indicating potential overfitting. The model achieving the highest validation AUC was selected and subsequently evaluated on the unseen

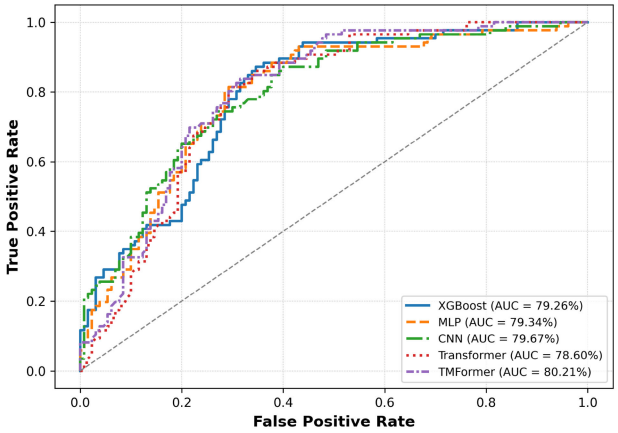


FIGURE 5. ROC curves and corresponding AUC values for: the XGBoost and Multilayer Perceptron models using extracted features; the Convolutional Neural Network using spectrograms; Transformer and TMFormer models using raw ECG signal. Results are based on the 1-hour epoch from the unseen test set of patients.

TABLE 3. AUC values for binary classification using all models and transformer variations used in this paper.

Model	Input data	AUC score (%)
XGBoost	HRV features	79.26
MLP	HRV features	79.34
CNN	ECG Spectrogram	79.67
Transformer	Raw ECG	78.6
TMFormer	Raw ECG	**80.21**
TMFormer without trend	Raw ECG	79.02

TABLE 4. True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) metrics calculated from binary classification on all 4-minute segments across all applied models.

model	TP	FP	TN	FN
XGBoost	4875	3086	6144	1183
MLP	4767	3471	5759	1291
CNN	4763	3525	5705	1338
Transformer	5001	3792	5438	1105
TMFormer	4517	3094	6136	1589

ANSeR1 test dataset. The models were initialized using the Kaiming method to ensure efficient weight distribution at the start of training. Training was conducted using a cosine annealing learning rate scheduler, beginning with an initial learning rate of 10^{-5}. Optimization was performed using the AdamW optimizer, with cross-entropy loss employed as the objective function.

All experiments were implemented using the PyTorch 2.3 framework in Python. Training was executed on a system equipped with a single Tesla V100-SXM2 GPU (16 GB VRAM) and a Common KVM processor running at 2.60 GHz, featuring 8 physical cores and 8 logical threads.

IV. RESULTS

The classification performance on the test data was assessed using the area under the receiver operating characteristic curve (AUC).

Among the models using heart rate (HR) features, XGBoost and MLP achieved AUC scores of 79.26% and 79.34%, respectively, with MLP showing a slight performance advantage. The CNN model trained on spectrogram representations outperformed both, reaching an AUC of 79.67%, highlighting the benefit of time-frequency input representations. For methods based on raw ECG signals, the proposed TMFormer model achieved the AUC of 80.21%, outperforming the standard Transformer, which yielded an AUC of 78.6%. While the differences in AUC values across models are relatively modest, the TMFormer using raw ECG signals achieved the highest AUC among all methods, demonstrating its ability to capture the dynamic temporal structure inherent in ECG data. The ROC curves and corresponding AUC values for all models on the unseen test set are visualized in Fig. 5.

Additionally, the performance of the TMFormer was evaluated with and without the trend component, yielding AUCs of 80.21% and 79.02%, respectively. The results demonstrate that incorporating the trend component in the tokenization process leads to a 1.19% improvement in AUC. table 3 presents the AUC score of all models and various transformer-based configurations tested during the experiments.

Since the classification results for all applied models were based on the average predicted probability across all test samples within each one-hour ECG epoch, the corresponding True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) metrics, calculated from predictions on all 4-minute segments, are provided in Table 4.

The execution time required for input data generation across all methods is summarized in Table 5. The measurement was performed for each 1-hour ECG epoch and the per-segment execution time was calculated based on the number of 4-minutes segments within a 1-hour epoch. Using raw ECG eliminates HRV feature extraction and spectrogram computation, making it suitable for real-time or low-resource scenarios. Since no computational time is required for raw ECG, it was not included in the table. Spectrogram generation is highly efficient, requiring just 0.022 milliseconds per segment, which makes it over 4700 times faster than HR feature extraction which, takes over 100 milliseconds per segment.

Fig. 6 shows the classification results of all models which are compared across the five HIE-EEG grades previously explained in Table 1. For each 1-hour ECG epochs the corresponding HIE-EEG grade is used. Each bar represents the correct and incorrect binary classification of 1-hour of ECG epochs by the respective classifier. For each grade, the bars show the number of incorrect classifications stacked on top of the correct ones. The chart highlights the misclassification patterns across five different HIE-EEG severity levels, with the total number of 1-hour of ECG epochs per grade annotated above each bar. A vertical line marks the borderline between category 0 and category 1.

TABLE 5. The execution time for data generation of all applied methods.

Generated data type	1-hour ECG epoch (ms)	4-minutes ECG segment (ms)
Spectrogram	1.55	0.022
NN Interval + HR Features	7453.0	104.972

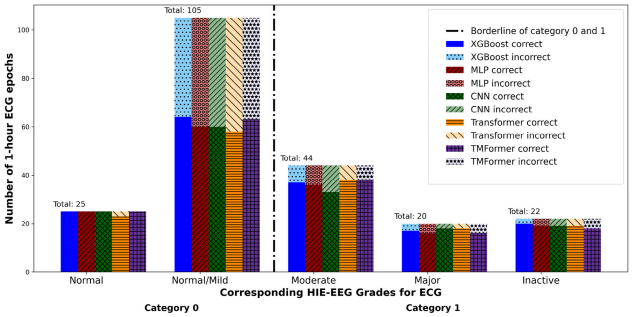


FIGURE 6. The classification results of all models for each 1-hour ECG epoch across the corresponding five HIE-EEG grades for ECG signal. Each bar shows correct (bottom) and incorrect (top) classifications of 1-hour ECG epochs. A vertical line marks the threshold between category 0 and catergory 1, used for binary classification.

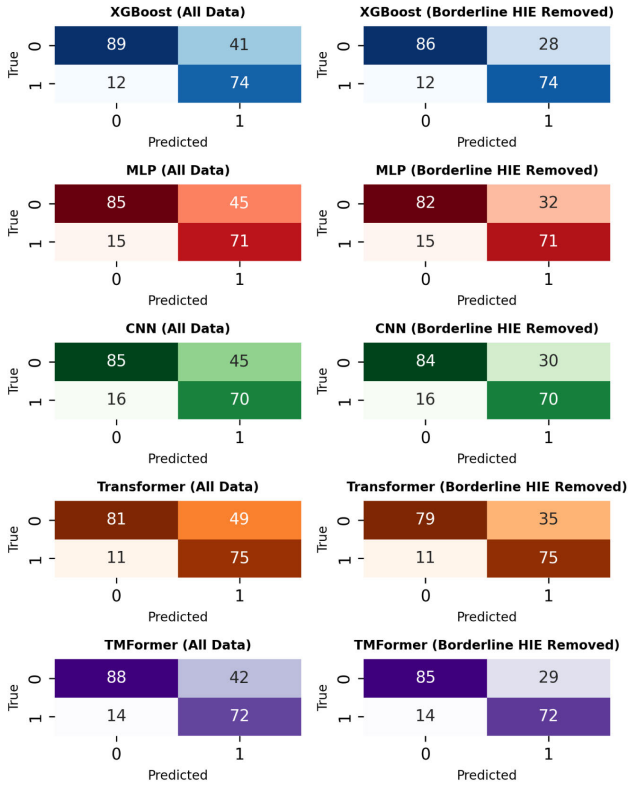


FIGURE 7. Confusion matrices of all models using the full dataset and after excluding corresponding borderline 1-hour ECG epochs.

An important finding was the consistent misclassification of Normal/mild HIE-EEG cases as category 1. Upon further expert review by a neurophysiologist, 17 one-hour EEG segments were flagged as ambiguous, falling between mild and moderate classifications. Given the subjective nature of HIE-EEG grading and the signal complexity, these ambiguous cases were difficult to label. Therefore, an analysis of the models performances were varied when the corresponding 1-hour ECG epochs excluded. As illustrated in the confusion matrices in Fig. 7, this exclusion led to a marked drop in misclassification rates across all models. Precision also improved for all models, with XGBoost achieving the highest precision (72.55%) and TMFormer reaching 71.3%. Additionally, accuracy, specificity, and F1 scores improved for all models, however sensitivity remained unchanged (See table 1 in supplementary materials).

Fig. 8 presents a 4 minute input ECG signal, overlaid with the attention flow from the CLS token at the first layer, the second layer, and the cumulative attention flow across both layers, along with the extracted NN intervals derived from the ECG signal. Each layer's attention flow is visualized as a heatmap, where darker pink regions indicate areas that the model focuses on more strongly during classification, based on the attention weights of that specific layer. The cumulative attention flow (attention rollout) is shown as the combined (multiplied) attention from both layers. The extracted NN intervals, calculated from the peaks of the ECG signal, are included for comparison. The plots show that the model's attention consistently emphasizes regions of the ECG signal that align with fluctuations in the NN interval pattern, suggesting a connection between the model's focus and heart rate dynamics, which are often altered in cases of HIE.

Since the aim of this study was to develop an automatic real-time system for clinical support in hospitals to assist clinicians in decision-making regarding TH, the model's predictions are shown here for a continuous ECG recording from an example neonate between hour 6 and hour 48 after birth with variation in HIE category. Fig. 9 shows the prediction probability of binary classes (category 0 and category 1) produced by XGBoost and MLP models using HR features, CNN using ECG spectrogram and TMFormer using raw ECG. Predictions were generated for each segment, and a rolling average window was applied to smooth the probability curves. Vertical lines indicate the boundaries of five 1-hour ECG epochs labelled by a neurophysiologist, with the assigned target categories displayed above each epoch. As shown the predicted probabilities increase when the category is 1 and decrease when the category is 0. When comparing the prediction curves, the probabilities generated by TMFormer using the raw ECG method more closely align with the target categories, indicating greater confidence and accuracy in distinguishing between category 0 and category 1 compared to other methods. An important observation is that, the predictions from XGBoost and MLP using extracted features show missing segments between hours 6–9

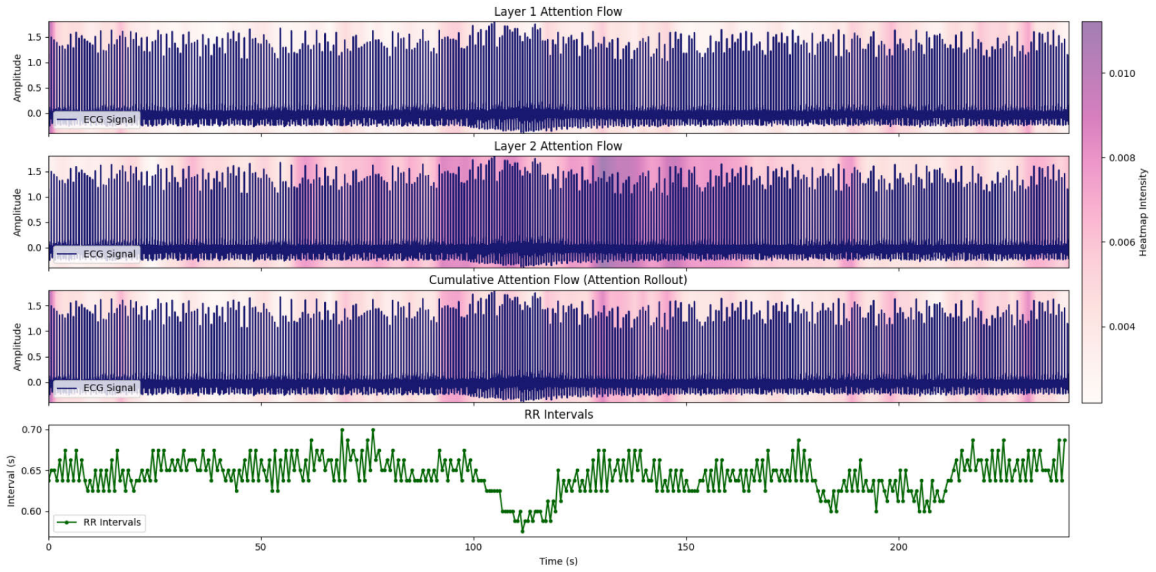


FIGURE 8. A 4-minute ECG segment with overlaid TMFormer attention flow heatmaps from the CLS token at the first and second layer and the cumulative attention flow across both layers, alongside NN intervals for comparison. The darker pink regions highlight areas of the ECG signal that the model attends to more strongly during classification. The ECG recording is from hour 6 of a neonate with target category 1.

and 30–33. These gaps correspond to noisy portions of the ECG signal, where QRS peak detection using the Pan-Tompkins algorithm failed. As a result, heart rate features could not be computed for those segments, and they were subsequently discarded. This issue is not observed in other methods, highlighting a limitation of HRV-based approaches for real-time applications.

To enhance visual understanding and interpretation, the model predictions for continuous ECG recordings from additional neonates are presented. Fig. 10 illustrates the predicted probabilities for a neonate with target category 1, covering the period from hour 6 to 48 after birth. Notably, the TMFormer model using raw ECG, produces a consistently strong prediction, showing a high degree of certainty in predicting the correct target category. Also, the predictions from XGBoost and MLP using extracted features show missing segments between hours 9–12 and 18–21, corresponding to noisy intervals in the signal. See figure 1 in supplementary materials for visualization of prediction probabilities for a neonate with target category 0.

As the result of each classifier varied across patients, an ensemble analysis was performed using all four methods: XGBoost and MLP using HR features, CNN using spectrogram and TMFormer using raw ECG. Since the models were trained on different feature types, their predicted probabilities were combined without additional training or fine-tuning, using simple averaging. This ensemble analysis was conducted to assess the relative contribution of each model to overall performance. A leave-one-out approach was employed, where each method was removed in turn and the AUC of the remaining ensemble was calculated. The results presented in Table 6 show that the full ensemble

TABLE 6. AUC scores for the full ensemble and after removing each model. XGBoost and MLP using HR features, CNN using spectrogram and TMFormer using raw ECG methods were used and the impact of each model on ensemble performance is shown.

Models	AUC (%)
XGBoost + MLP + CNN + TMFormer	82.03
XGBoost removed	82.35
MLP removed	81.88
CNN removed	81.57
TMFormer removed	81.08

achieved an AUC of 82.03%. Notably, the exclusion of TMFormer led to the largest decrease in performance, indicating its substantial contribution to the ensemble's predictive capability. Conversely, the removal of XGBoost resulted in a slight performance improvement, suggesting that it may introduce some degree of redundancy or noise in the combined predictions.

Since timely assessments of infants with HIE who may benefit from therapeutic hypothermia (TH) is critical, we evaluated the performance of TMFormer using ECG recordings from progressively earlier time windows after birth. Specifically, we assessed model performance using the first 12 hours and the first 24 hours of postnatal ECG data (see table 2 in supplementary materials for more details). Using 84 one-hour ECG epochs from the initial 12 hours of life, the model achieved an AUC of 85.91%. When extended to the first 24 hours, the model achieved a comparable AUC of 84.9%. Notably, both results exceed the performance obtained using the full 48-hour dataset, 216 one-hour epochs, which yielded an AUC of 80.21%. These findings suggest that TMFormer is capable of classifying

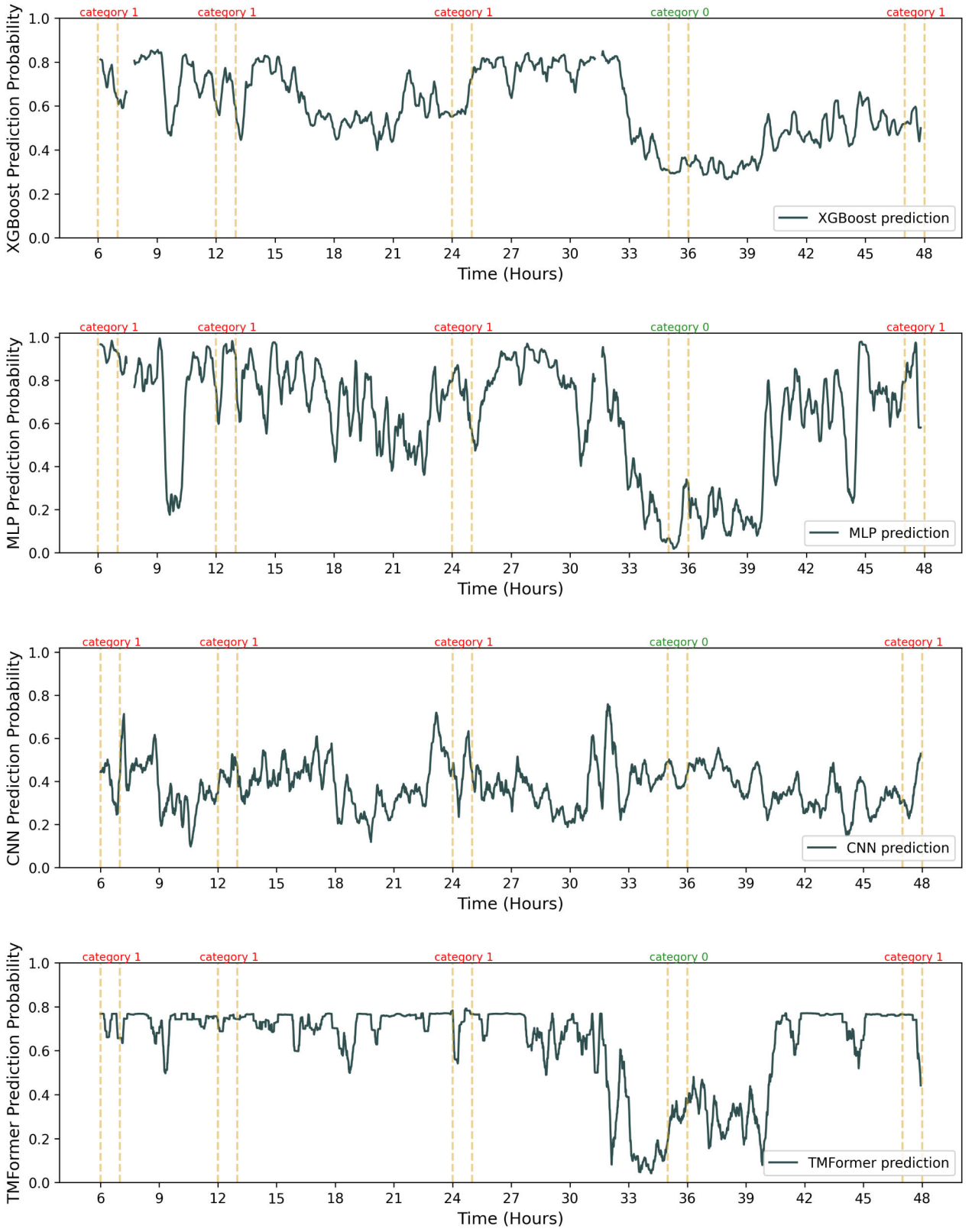


FIGURE 9. The model predictions of a continuous ECG recording from a single neonate between hour 6 and hour 48 after birth with fluctuating target category. The plots, from top to bottom, show: XGBoost predictions using HR features, MLP predictions using HR features, CNN predictions using ECG spectrograms, and TMFormer model predictions using raw ECG data.

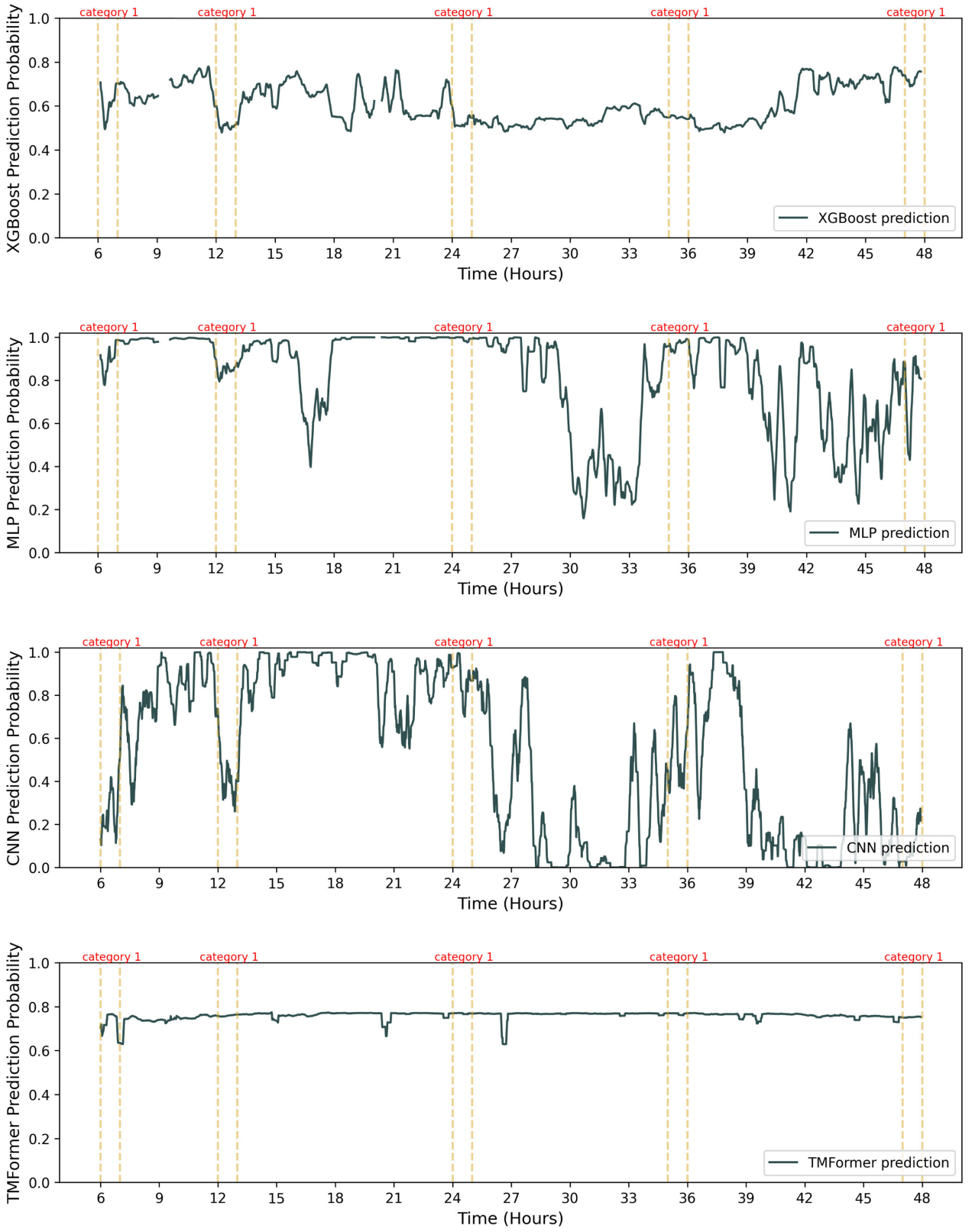


FIGURE 10. The model predictions of a continuous ECG recording from a single neonate between hour 6 and hour 48 after birth with target category 1. The plots, from top to bottom, show: XGBoost predictions using HR features, MLP predictions using HR features, CNN predictions using ECG spectrograms, and TMFormer model predictions using raw ECG data.

infants using substantially less data, highlighting its potential value for as a clinical support tool.

V. CONCLUSION

This study demonstrates the potential of utilizing raw ECG signals and ECG spectrograms in combination with state-of-the-art deep learning models for binary classification of ECG signal segments in neonates with HIE. While the performance shows only a slight improvement over traditional HRV-based methods, the use of raw ECG signals offers key advantages, such as eliminating the need for preprocessing and enabling real-time analysis, enhancing its clinical practicality. Additionally, the lightweight TMFormer used in this study is efficient enough to run on portable devices and remains robust under noisy conditions.

Furthermore, achieving an AUC of 80.21% highlights the potential of this approach to assist clinicians in assessments of infants with HIE, although further research is warranted. These findings represent progress toward developing supportive tools that complement clinical evaluation. By leveraging easily obtainable ECG signals, this method reduces reliance on EEG-based assessments by neurophysiologists in lower-resource healthcare settings.

In addition to inherent noise in the ECG signals, a significant limitation of this study is the signal interference in some patients occurring during therapeutic hypothermia and the effects of medication, which may affect data quality. Another limitation includes variability of inter-rater label agreement specially on borderline cases. While the datasets employed were relatively large, the model's performance could be further improved by incorporating even larger and more diverse datasets that better capture the variability seen across different clinical scenarios. The ECG signal was extracted from EEG recordings; using ECG routinely monitored by the clinical team, rather than captured as part of the EEG setup, may improve signal quality and robustness.

REFERENCES

[1] P. Greco, G. Nencini, I. Piva, M. Scioscia, C. A. Volta, S. Spadaro, M. Neri, G. Bonaccorsi, F. Greco, I. Cocco, F. Sorrentino, F. D'Antonio, and L. Nappi, "Pathophysiology of hypoxic–ischemic encephalopathy: A review of the past and a view on the future," *Acta Neurol. Belgica*, vol. 120, no. 2, pp. 277–288, Apr. 2020.

[2] H. B. Sarnat and M. S. Sarnat, "Neonatal encephalopathy following fetal distress: A clinical and electroencephalographic study," *Arch. Neurol.*, vol. 33, no. 10, pp. 696–705, 1976.

[3] J. M. Conway, B. H. Walsh, G. B. Boylan, and D. M. Murray, "Mild hypoxic ischaemic encephalopathy and long term neurodevelopmental outcome—A systematic review," *Early Human Develop.*, vol. 120, pp. 80–87, May 2018.

[4] G. Wassink, J. O. Davidson, S. K. Dhillon, K. Q. Zhou, L. Bennet, M. Thoresen, and A. J. Gunn, "Therapeutic hypothermia in neonatal hypoxic-ischemic encephalopathy," *Current Neurol. Neurosci. Rep.*, vol. 19, no. 1, p. 2, 2019.

[5] D. M. Murray, G. B. Boylan, C. A. Ryan, and S. Connolly, "Early EEG findings in hypoxic-ischemic encephalopathy predict outcomes at 2 years," *Pediatrics*, vol. 124, no. 3, pp. e459–e467, Sep. 2009.

[6] D. V. Azzopardi, B. Strohm, A. D. Edwards, L. Dyet, H. L. Halliday, E. Juszczak, O. Kapellou, M. Levene, N. Marlow, E. Porter, and M. Thoresen, "Moderate hypothermia to treat perinatal asphyxial encephalopathy," *New England J. Med.*, vol. 361, no. 14, pp. 1349–1358, 2009.

[7] D. Neubauer, D. Osredkar, D. Paro-Panjan, A. Skofljanec, and M. Derganc, "Recording conventional and amplitude-integrated EEG in neonatal intensive care unit," *Eur. J. Paediatric Neurol.*, vol. 15, no. 5, pp. 405–416, 2011.

[8] R. Ahmed, A. Temko, W. Marnane, G. Lightbody, and G. Boylan, "Grading hypoxic–ischemic encephalopathy severity in neonatal EEG using GMM supervectors and the support vector machine," *Clin. Neurophysiol.*, vol. 127, no. 1, pp. 297–309, Jan. 2016.

[9] S. A. Raurale, G. B. Boylan, S. R. Mathieson, W. P. Marnane, G. Lightbody, and J. M. O'Toole, "Grading hypoxic-ischemic encephalopathy in neonatal EEG with convolutional neural networks and quadratic time–frequency distributions," *J. Neural Eng.*, vol. 18, no. 4, Aug. 2021, Art. no. 046007.

[10] S. Yu, W. P. Marnane, G. B. Boylan, and G. Lightbody, "Neonatal hypoxic-ischemic encephalopathy grading from multi-channel EEG time-series data using a fully convolutional neural network," *Technologies*, vol. 11, no. 6, p. 151, Oct. 2023.

[11] M. E. O'Sullivan, G. Lightbody, S. R. Mathieson, W. P. Marnane, G. B. Boylan, and J. M. O'Toole, "Development of an EEG artefact detection algorithm and its application in grading neonatal hypoxic-ischemic encephalopathy," *Expert Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 118917.

[12] L. Twomey, S. Gomez, E. Popovici, and A. Temko, "Long-term neonatal EEG modeling with DSP and ML for grading hypoxic–ischemic encephalopathy injury," *Sensors*, vol. 25, no. 10, p. 3007, May 2025.

[13] D.-H. Shih, F.-I. Chung, T.-W. Wu, S.-Y. Huang, and M.-H. Shih, "Advanced trans-EEGNet deep learning model for hypoxic-ischemic encephalopathy severity grading," *Mathematics*, vol. 12, no. 24, p. 3915, Dec. 2024.

[14] R. Ahmed, A. Temko, W. P. Marnane, G. Boylan, and G. Lightbody, "Classification of hypoxic-ischemic encephalopathy using long term heart rate variability based features," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 2355–2358.

[15] R. M. Goulding, N. J. Stevenson, D. M. Murray, V. Livingstone, P. M. Filan, and G. B. Boylan, "Heart rate variability in hypoxic ischemic encephalopathy: Correlation with EEG grade and 2-y neurodevelopmental outcome," *Pediatric Res.*, vol. 77, no. 5, pp. 681–687, May 2015.

[16] A. M. Pavel, S. R. Mathieson, V. Livingstone, J. M. O'Toole, R. M. Pressler, L. S. de Vries, J. M. Rennie, S. Mitra, E. M. Dempsey, D. M. Murray, W. P. Marnane, G. B. Boylan, and A. Consortium, "Heart rate variability analysis for the prediction of EEG grade in infants with hypoxic ischaemic encephalopathy within the first 12 H of birth," *Frontiers Pediatrics*, vol. 10, Jan. 2023, Art. no. 1016211.

[17] K. Rezaei, K. Yu, S. R. Mathieson, A. Flynn, G. Lightbody, G. B. Boylan, and W. P. Marnane, "Assessing the effectiveness of heart rate variability as a diagnostic tool for brain injuries in infants," in *Proc. 46th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2024, pp. 1–4.

[18] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vols. BME–32, no. 3, pp. 230–236, Mar. 1985.

[19] M. Bolanos, H. Nazeran, and E. Haltiwanger, "Comparison of heart rate variability signal features derived from electrocardiography and photoplethysmography in healthy individuals," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2006, pp. 4289–4294.

[20] R. Gutiérrez-Rivas, J. J. García, W. P. Marnane, and Á. Hernández, "Novel real-time low-complexity QRS complex detector based on adaptive thresholding," *IEEE Sensors J.*, vol. 15, no. 10, pp. 6036–6043, Oct. 2015.

[21] A. M. Pavel, J. M. Rennie, L. S. de Vries, M. Blennow, A. Foran, D. K. Shah, R. M. Pressler, O. Kapellou, E. M. Dempsey, S. R. Mathieson, E. Pavlidis, A. C. van Huffelen, V. Livingstone, M. C. Toet, L. C. Weeke, M. Finder, S. Mitra, D. M. Murray, W. P. Marnane, and G. B. Boylan, "A machine-learning algorithm for neonatal seizure recognition: A multicentre, randomised, controlled trial," *Lancet Child Adolescent Health*, vol. 4, no. 10, pp. 740–749, Oct. 2020.

[22] J. M. Rennie, L. S. de Vries, M. Blennow, A. Foran, D. K. Shah, V. Livingstone, A. C. van Huffelen, S. R. Mathieson, E. Pavlidis, L. C. Weeke, M. C. Toet, M. Finder, R. M. Pinnamaneni, D. M. Murray, A. C. Ryan, W. P. Marnane, and G. B. Boylan, "Characterisation of neonatal seizures and their treatment using continuous EEG monitoring: A multicentre experience," *Arch. Disease Childhood-Fetal Neonatal Ed.*, vol. 104, no. 5, pp. F493–F501, Sep. 2019.

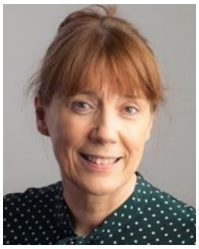
[23] M. Hammad, A. Maher, K. Wang, F. Jiang, and M. Amrani, "Detection of abnormal heart conditions based on characteristics of ECG signals," *Measurement*, vol. 125, pp. 634–644, Sep. 2018.

[24] H. Li and P. Boulanger, "Structural anomalies detection from electrocardiogram (ECG) with spectrogram and handcrafted features," *Sensors*, vol. 22, no. 7, p. 2467, Mar. 2022.
[25] O. M. Doyle, A. Temko, W. Marnane, G. Lightbody, and G. B. Boylan, "Heart rate based automatic seizure detection in the newborn," *Med. Eng. Phys.*, vol. 32, no. 8, pp. 829–839, Oct. 2010.
[26] G. Volpes, C. Barà, A. Busacca, S. Stivala, M. Javorka, L. Faes, and R. Pernice, "Feasibility of ultra-short term analysis of heart rate and systolic arterial pressure variability at rest and during stress via time-domain and entropy-based measures," *Sensors*, vol. 22, no. 23, p. 9149, Nov. 2022.
[27] M. Toichi, T. Sugiura, T. Murai, and A. Sengoku, "A new method of assessing cardiac autonomic function and its comparison with spectral analysis and coefficient of variation of R–R interval," *J. Autonomic Nervous Syst.*, vol. 62, nos. 1–2, pp. 79–84, Jan. 1997.
[28] J. Jeppesen, S. Beniczky, P. Johansen, P. Sidenius, and A. Fuglsang-Frederiksen, "Detection of epileptic seizures with a modified heart rate variability algorithm based on Lorenz plot," *Seizure*, vol. 24, pp. 1–7, Jan. 2015.
[29] J. Feng, M. H. Erol, J. Son Chung, and A. Senocak, "From coarse to fine: Efficient training for audio spectrogram transformers," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2024, pp. 1416–1420.
[30] S. Daphin Lilda and R. Jayaparvathy, "Enhancing cardiovascular disease classification in ECG spectrograms by using multi-branch CNN," *Comput. Biol. Med.*, vol. 186, Mar. 2025, Art. no. 109737.
[31] Y. D. Daydulo, B. L. Thamineni, and A. A. Dawud, "Cardiac arrhythmia detection using deep learning approach and time frequency representation of ECG signals," *BMC Med. Informat. Decis. Making*, vol. 23, no. 1, p. 232, Oct. 2023.
[32] P. Lu, C. Wang, J. Hagenah, S. Ghiasi, V. Consortium, T. Zhu, L. Thwaites, and D. A. Clifton, "Improving classification of tetanus severity for patients in low-middle income countries wearing ECG sensors by using a CNN-transformer network," *IEEE Trans. Biomed. Eng.*, vol. 70, no. 4, pp. 1340–1350, Apr. 2023.
[33] L. Cabrera-Quiros, G. Varisco, Z. Zhan, X. Long, P. Andriessen, E. J. E. Cottaar, and C. van Pul, "Estimation of heart rate directly from ECG spectrogram in neonate intensive care units," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 320–323.
[34] W. Luo, Y. Li, R. Urtasun, and R. S. Zemel, "Understanding the effective receptive field in deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2017, pp. 4898–4906.
[35] J. Wang and M. Hauskrecht, "Augmentation-free contrastive learning for EKG classification," in *Proc. Int. Conf. Artif. Intell. Med.* Cham, Switzerland: Springer, 2025, pp. 468–479.
[36] E. C. Chukwu and P. A. Moreno-Sánchez, "Enhancing arrhythmia diagnosis with data-driven methods: A 12-lead ECG-based explainable AI model," in *Proc. Nordic Conf. Digit. Health Wireless Solutions*, 2024, pp. 242–259.
[37] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed., Waterloo, ON, Canada: OTexts, 2018.
[38] N. M. Foumani, C. W. Tan, G. I. Webb, and M. Salehi, "Improving position encoding of transformers for multivariate time series classification," *Data Mining Knowl. Discovery*, vol. 38, no. 1, pp. 22–48, Jan. 2024.
[39] S. Abnar and W. Zuidema, "Quantifying attention flow in transformers," 2020, *arXiv:2005.00928*.

KIMIA REZAEI received the B.S. and M.S. degrees in electrical and electronic engineering from Islamic Azad University, Iran, in 2012 and 2014, respectively. She is currently pursuing the joint Ph.D. degree in electrical and electronic engineering with University College Cork and the INFANT Research Centre, Cork, Ireland. From 2015 to 2020, she was a Telecommunications Engineer at Sahand Parsian Gharb Telecom Company, Shiraz, Iran. Her research interests include neonate HIE grading and seizure detection by leveraging the ECG signal as an alternative to EEG.

SEAN R. MATHIESON received the B.Sc. degree in biology in Bath, U.K., the M.Sc. degree in neuroscience from UCL, London, and the Ph.D. degree in NICU from UCLH, London, in 2016, with a focus to test an automated seizure detection algorithm for neonatal EEG. He trained and worked as a Clinical Physiologist at the Clinical Neurophysiology Department, Great Ormond Street Hospital, from 2001 to 2009. Since then, he has been at the INFANT Research Group, Cork, Ireland, supporting various EEG-based studies.

GORDON LIGHTBODY received the M.Eng. (Hons.) and Ph.D. degrees in electrical and electronic engineering from Queen's University Belfast, in 1989 and 1993, respectively. He is currently a Professor of control engineering with University College Cork, Ireland, and a Principal Investigator of the Infant (www.infantcentre.ie) Research Centre. His current research interests include artificial intelligence techniques for intelligent control and signal-processing, focusing on energy/power and biomedical applications. He is an Associate Editor of *Control Engineering Practice* (Elsevier).



GERALDINE B. BOYLAN is currently a Professor of neonatal physiology with University College Cork (UCC) and the Director of INFANT, the Irish centre for maternal and child health research-based at UCC. An expert in neonatal neurophysiology, she leads the Neonatal Brain Research Group, INFANT—a multidisciplinary team focused on improving neuromonitoring for neonates in intensive care, particularly for seizure detection and early diagnosis of brain injury. She has published extensively on neonatal seizures and neonatal EEG. Her research has resulted in patents, licensed technologies, and spinouts.

WILLIAM P. MARNANE received the B.E. degree in electrical engineering from University College Cork (UCC), in 1984, and the D.Phil. degree from the University of Oxford, U.K., in 1989.

His D.Phil. thesis was on "Testing VLSI Regular Arrays." He was appointed as a Lecturer of VLSI design at the School of Electronic Engineering Science, University of Wales, Bangor, in 1989. In 1993, he was appointed as a Lecturer of digital signal processing at the Department of Electrical and Electronic Engineering, UCC, where he was a Senior Lecturer, in 1999, and a Professor, in 2014. He was the Vice Head of the School of Engineering, UCC, from 2010 to 2013. He was the Dean of Graduate Studies for UCC, from 2013 to 2016. He was the Head of the School of Engineering, UCC, from 2016 to 2019. He is currently a Professor of electrical and electronic engineering with UCC. He was a Co-PI in the Science Foundation Ireland (SFI) funded Claude Shannon Institute for Discrete Mathematics, Coding and Cryptography, from 2006 to 2012. He was the Director of the SFI-funded Strategic Research Cluster on Efficient Embedded Digital Signal Processing for Mobile Digital Health, from 2007 to 2011. He is the PI of the INFANT Research Centre—Irish Centre for Fetal and Neonatal Translational Research. His research interests include biomedical signal processing and machine learning. This research has focused on neonatal EEG signal analysis and processing.

Dr. Marnane was awarded the Postdoctoral Marie Curie Fellowship from European Union to investigate the design for test of VLSI arrays at the IRISA University of Rennes, in 1992.