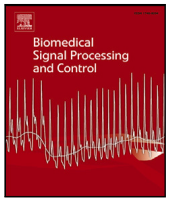
Biomedical Signal Processing and Control 78 (2022) 103998

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



Interpretable seizure detection with signal temporal logic neural network

Ruixuan Yan *, A. Agung Julius

Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY, United States

ARTICLE INFO

Keywords:
Epileptic seizure detection
Electroencephalogram (EEG)
Multi-view feature extraction
Interpretable machine learning
Signal temporal logic neural network

ABSTRACT

In this work, we develop a novel neuro-symbolic model for automated seizure detection using multi-views of data representation. Firstly, the spectral and line length features are extracted using a multi-view feature extraction technique. Next, a signal temporal logic neural network (STONE) that combines the benefits of neural networks and temporal logics is constructed to classify the seizure and nonseizure data. STONE is designed in such a way that each neuron has a symbolic representation corresponding to a component in a weighted signal temporal logic (wSTL) formula. Compared with traditional STL inference algorithms, STONE is end-to-end differentiable such that the learning can be accomplished through back-propagation. In addition, STONE improves the interpretability of seizure detection models as the outcome of STONE is a wSTL formula that is interpretable and human-readable. Importantly, the wSTL formula reveals the reasoning behind seizure as a description of the evolution of EEG signals. STONE is tested on two popular EEG databases and demonstrated to achieve promising detection performance in terms of accuracy, sensitivity, and specificity when compared with existing state-of-the-art models. Furthermore, STONE can provide a human-readable formula as a description of the seizure characteristics, and the formula is also visualizable for easy interpretation of the classifier, which is a missing property in existing seizure detection methods.

1. Introduction

Epileptic seizure, also known as seizure, refers to a period of symptoms that are caused by abnormally excessive or synchronized neuronal activity in the brain [1]. Seizure is a severe neurological disorder with the characteristics of recurrency. It is reported that nearly 10% of the population worldwide has experienced the epileptic seizure at least once [2]. Epileptic seizures can be classified into two categories based on their potential causes. One type is called the provoked seizure caused by a temporary incident such as low blood sugar, fever, alcohol intake, etc [1]. Another type of seizure is called the unprovoked seizure, which has no fixable causes, and ongoing seizures follow a similar pattern [3]. Nearly 3.5 of 10000 people have experienced provoked seizures in a year, and nearly 4.2 of 10000 people have experienced unprovoked seizures in a year [2]. Accurate detection and timely treatment of seizures could reduce the risk of recurrence and enhance human well-being. One of the most popular techniques for seizure detection is through analyzing electroencephalogram signals. Electroencephalogram (EEG) is a technique for recording an electrogram of electrical activity on the scalp, which has been proven to reflect the macroscopic activity of the brain's surface layer underneath [4]. Traditional seizure detection approaches rely on neurologists observing the long-term EEG recordings and manually discovering the happening of an epileptic seizure. These approaches are obviously inefficient and not accurate enough. To overcome the limitations of traditional seizure diagnostic methods and improve the diagnosing efficiency, automated seizure detection from EEG data has become a research hotspot in the field.

With the advancement of machine learning (ML) techniques, incorporating machine learning models into automated seizure detection frameworks has attracted lots of interest. These frameworks mainly apply machine learning models to features that are extracted from the EEG signals. In particular, statistical features have been extracted through time-domain transformations such as empirical mode decomposition (EMD) [5] and frequency or time frequency-based transformation techniques such as fast Fourier transform (FFT), discrete wavelet transform (DWT), discrete cosine transform (DCT), intrinsic mode function (IMF), wavelet packet decomposition (WPD), etc [6–10]. These features have been validated to be suitable for seizure detection and other brain-related activities [11]. It was also noted that *"line length"* could be used as a powerful feature for seizure detection and generate satisfactory outcomes [8]. Combining *"line length"* and other features into machine learning models could obtain promising outcomes as well [11]. Although machine learning techniques have proven their contribution in seizure detection, most classification models are not in a human-readable, interpretable format [12–16], such as hyperplanes in a higher-dimensional parameter space [17–20], or ensemble

* Corresponding author.
E-mail address: yanr5@rpi.edu (R. Yan).

https://doi.org/10.1016/j.bspc.2022.103998
Received 30 March 2022; Received in revised form 8 June 2022; Accepted 11 July 2022
Available online 25 July 2022

models that combine the outputs of multiple base classifiers [5,21]. The lack of interpretability and transparency [22] prevents users or neurologists from understanding and trusting the outcomes of the above models [23]. In addition to classifying data into a certain category, enhancing the classifier to be intuitive to provide rules that are interpretable can promote the classifier's utilization in clinical practice as these rules can better assist in decision making, especially for neurologists who need to understand the reasoning behind the classifier. Recently, approaches toward improving the interpretability of machine learning models in analyzing EEG signals have been proposed in [24,25]. In this paper, we tackle the interpretability issue from a novel perspective by combining the capabilities of neural networks and symbolic logic to build a neuro-symbolic model that can learn logical rules as interpretable descriptions of the properties of EEG signals.

Temporal logics are formal languages that can express specifications about the temporal properties of systems. Compared with traditional ML models, temporal logic formulas can express temporal properties of EEG signals in a format similar to natural language, allowing for a better understanding of EEG signals' evolution through intuitive visualization. Using temporal logic formulas as outcomes of the classifier offers a deeper level of interaction with the users. With these benefits, temporal logics have been widely exploited to model temporal data [26,27] using their expressiveness of data properties.

Signal temporal logic (STL), a branch of temporal logics for analyzing continuous-time data, has been applied to analyze signals from cyber–physical systems [28–32]. Traditional approaches for learning STL formulas face the limitations of low computation efficiency, and the learned STL formula cannot mirror the importance of signals at different time slots in describing the data characteristics. This paper utilizes weighted STL (wSTL) [33] as the target formula that takes advantage of weights to reflect the importance of subformulas and signals at various time slots. Additionally, a neuro-symbolic model called Signal Temporal lOgic Neural nEtwork (STONE) is proposed to accomplish seizure detection tasks, which can also learn a wSTL formula that describes the characteristics of seizures. The benefits of STONE are mainly two folds. STONE offers an end-to-end differentiable way of learning wSTL formulas. Also, STONE improves the interpretability and transparency of seizure detection models by associating each neuron with a symbolic component in a wSTL formula and producing an interpretable wSTL formula, which takes a step further toward constructing trustworthy seizure detection models.

Many existing seizure detection models are designed as cross-patient classifiers that assume one patient's EEG seizure signal resembles another patient's EEG seizure signal [34,35]. In practice, the patterns of EEG seizure signals vary across different patients. The variability may cause cross-patient seizure detection models not accurate enough or missing detection of seizure onsets. STONE takes the above factors into consideration and is designed in a way that it can be used as a cross-patient or patient-dependent detector. While STONE is exploited for patient-dependent seizure detection, it can both declare the onset of a seizure and provide a patient-dependent wSTL formula describing the temporal patterns underlying a particular patient's seizure data.

The contributions of this paper are summarized as follows: (1) we propose a neuro-symbolic model called Signal Temporal lOgic Neural nEtwork (STONE) by merging the benefits of neural networks with STL to accomplish seizure detection tasks; (2) STONE is designed in such a way that it can be utilized as a cross-patient or patient-dependent seizure detection model, with a patient-specific wSTL formula learned while being used for patient-dependent purposes; (3) STONE is tested on the CHB-MIT scalp EEG database and the Bonn University database, and the detector is shown to exhibit accurate detection of seizures in terms of high sensitivity and specificity and acquire impressive performance compared to state-of-the-art models; (4) STONE is capable of learning human-readable formulas that reveal how the EEG data from various channels evolves over time will cause seizure, allowing the patterns to be visualized such that the classifier can be better understood from the perspectives of speech and vision.

The remaining sections of this paper are organized as follows. The description of the databases used in this paper, including the CHB-MIT and the Bonn University databases, is presented in Section 2. Section 3 presents the proposed methodology to tackle the seizure detection problem, which includes the multi-view feature extraction through FFT and line length extraction, the classification model that is designed as a signal temporal logic neural network (STONE), the learning process of STONE, and the description of the evaluation metrics. The experiment design and results are presented in Section 4, which includes the performance of STONE and six conventional ML models on the CHB-MIT database and the Bonn University database, the comparison of STONE with the existing state-of-the-art models, and the ablation study of STONE, and the interpretability analysis of STONE. Finally, the conclusion and the contributions of this work are summarized in Section 5.

2. EEG dataset description

This paper relies on the Children's Hospital Boston - Massachusetts Institute of Technology (CHB-MIT) scalp electroencephalography (EEG) database [36,37] from the Physionet repository [38] and the Bonn University database [39] to accomplish the seizure detection task. The above datasets are considered as the target datasets because of their public availability, patient richness, and widespread acceptance in prior studies [5,12–16,20,40].

2.1. CHB-MIT database

The CHB-MIT database contains 24 cases of EEG recordings from 22 pediatric patients with stubborn seizures. The pediatric patients include five males aged three to twenty-two and seventeen females aged one year and a half to nineteen. The EEG signals are monitored when the anti-seizure drug is withdrawn, with the goal of determining their surgical intervention capabilities.

Each set of the recordings comprises between 9 and 42 continuous files in the European data format (EDF) from the same patient. The majority of the EDF files include digitized EEG signals that are exactly one hour long, while some recordings are up to four hours long and some are shorter. There are 664 EDF files in total, representing nearly 962 h of recordings in the CHB-MIT database. The recordings are obtained using the International 10–20 system of EEG electrode positions and nomenclature, where each recording is sampled at a frequency of 256 Hz with 16-bit resolution. Within the total 664 EDF files, 192 EDF files contain one or more seizures. The 664 EDF files contain 198 seizures in total that are annotated with the beginning and end locations of seizures.

2.2. Bonn university database

The Bonn University database contains five sets of EEG recordings, namely "Z", "O", "N", "F", and "S". Sets "Z" and "O" consist of EEG recordings from five healthy volunteers in a relaxed condition with their eyes open and closed. Sets "N", "F", and "S" comprise EEG recordings that are acquired from five epileptic patients using the electrodes placed intracranially. The recordings in sets "N" and "F" are collected during seizure-free intervals, while the recordings in the set "S" are collected from intervals corresponding to seizures. Each set is composed of 100 recordings with one channel of EEG signal. Each recording is with duration of 23.6 s. The recordings are digitalized using the International 10–20 system at a frequency of 173.6 Hz with 12-bit resolution. Hence the length of each recording is $173.6 \times 23.6 \approx 4097$ sample points. All the five sets of EEG recordings are exploited in this work, and the classification problems (CPs) considered in this paper are shown in Table 1.

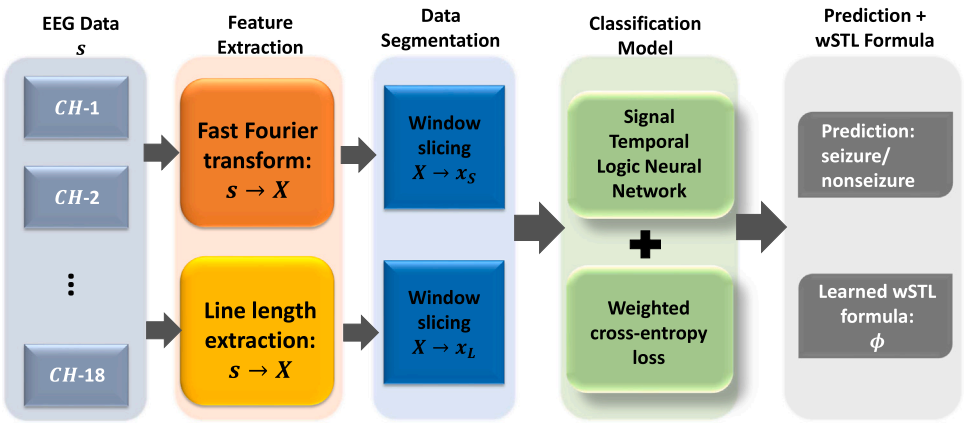


Fig. 1. Workflow of the overall seizure detection framework: feature extraction, data segmentation, classification, and prediction.

Table 1
Classification problems for the Bonn University database.

CPs	Class pair	Description	# EEG signal
1	Z–S	Normal, eyes open vs Ictal	200
2	O–S	Normal, eyes closed vs Ictal	200
3	N–S	Interictal vs Ictal	200
4	F–S	Interictal vs Ictal	200

3. Methodology

3.1. Multi-view feature extraction

The EEG signals are often complex, non-stationary, nonlinear, and time-dependent. As a result, directly applying classification models on the raw EEG data may not yield ideal results. Techniques such as extracting meaningful and significant features from the raw data have been proposed to improve the performance of seizure detection models. Ubiquitous data transformation techniques such as Fourier transform (FT), discrete wavelet transform (DWT), and singular value decomposition [6,41,42] have been applied to EEG signals to extract informative features. The classifier in this study uses spectral features and the line length feature as feature inputs. As an illustration, the workflow of the seizure detection framework is shown in Fig. 1. Firstly, we apply feature extraction techniques to collect the spectral features and the line length feature. The whole-series data is then segmented into temporal data with equal time lengths using a sliding window. Finally, the data segments are fed into the seizure detection model (STONE) to train the classifier such that it can make a prediction on the data and generate a wSTL formula. Here the CHB-MIT database is chosen to explain the feature extraction process, which can be equivalently applied to the Bonn University database.

3.1.1. Spectral feature

Using spectral characteristics to identify seizures has been demonstrated to be effective in several studies [6,42]. Here we adopt the fast Fourier transform (FFT) to acquire the spectral features. Let s denote the original EEG signal, and s_i denote the ith channel EEG signal. Most of the signals in the database contain 18 or 23 channels, and a few contain 24 to 26 channels. Similar to [37], we select 18 channels of the EEG signal for classification, hence $i \in \{1, 2, \dots, 18\}$. The fast Fourier transform (FFT) transforms the data in the time domain to a representation in the frequency domain by computing the discrete Fourier transform (DFT) of s. In the CHB-MIT database, each recording is sampled at a frequency of 256 Hz, which means a one-second recording comprises 256 samples. Hence the FFT of the EEG signal, denoted by $X_i(t,k)$, is expressed as

$$\begin{aligned} X_i(t,k) &= \sum_{n=0}^{N-1} s_i(t,n) e^{-\mathbb{j}\frac{2\pi kn}{N}}, k = 0, \dots, N-1, \\ &= X_i^a(t,k) + \mathbb{j} X_i^b(t,k), \end{aligned} \tag{1}$$

where $X_i(t,k)$ denotes the spectral data for $s_i(t)$ at frequency k, $s_i(t,n)$ denotes the nth sample at t, and $N = 256$, and $\mathbb{j}$ is the imaginary unit. For simplicity, the energy information is utilized as the spectral feature, which is expressed as

$$X_i(t,k) = \sqrt{(X_i^a(t,k))^2 + (X_i^b(t,k))^2}. \tag{2}$$

As most seizure and nonseizure activities fall in the frequency range between 0.1 Hz to 30 Hz [43], we retain the frequency data within the frequency range 0–30 Hz, and the remaining part is neglected. The detailed description of the bandwidths of EEG data and the associated abnormal activities are shown in Table 2. Based on the bandwidth information, we extract seven features for each one-second signal, including the energy within 0–30 Hz, the energy within 0–2 Hz, the energy within 3–4 Hz, the energy within 5–8 Hz, the energy within 9–16 Hz, the energy within 17–30 Hz, and the peak frequency within 0–30 Hz. Specifically, the features are derived via the following calculation:

$$\begin{aligned} X_{i,1}(t) &= \sum_{k=0}^{30} X_i(t,k), X_{i,2}(t) = \sum_{k=0}^{2} X_i(t,k), \\ X_{i,3}(t) &= \sum_{k=3}^{4} X_i(t,k), X_{i,4}(t) = \sum_{k=5}^{8} X_i(t,k), \\ X_{i,5}(t) &= \sum_{k=9}^{16} X_i(t,k), X_{i,6}(t) = \sum_{k=17}^{30} X_i(t,k), \\ X_{i,7}(t) &= \arg \max_{k \in [0,30]} X_i(t,k), \end{aligned} \tag{3}$$

where $X_{i,j}(t)$ denotes the jth feature extracted from the ith channel at t. With the above spectral feature selection procedure, we can extract spectral features at time slot t as

$$X_t = [X_{1,1}(t), X_{1,2}(t), \dots, X_{1,7}(t), \dots, X_{18,1}(t), X_{18,2}(t), \dots, X_{18,7}(t)]^T \in \mathbb{R}^{18\times 7}.$$

Due to the nonstationarity of EEG data, we extract temporal segments of EEG data using a sliding window with size $w = 10$ to obtain seizure or nonseizure segments for classification, which corresponds to the data segmentation module in Fig. 1. Thus for each spectral classification data x_S, it comprises ten time slots, each of which has $126 = 18 \times 7$ features, i.e., $x_S = [X_t, X_{t-1}, \dots, X_{t-9}]$. For better illustration, the spectral feature extraction process is also depicted in Fig. 2.

3.1.2. Line length feature

The "line length" feature, proposed in [8], has been used for studying normalization techniques in discriminating the EEG data [42].

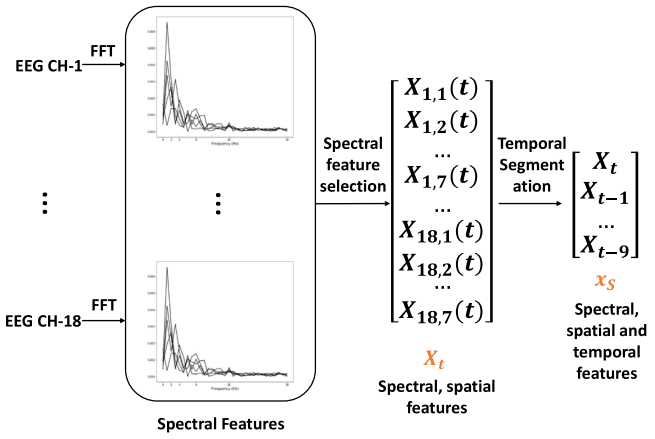


Fig. 2. Illustration of the spectral feature extraction process.

Table 2
Description of EEG bandwidths and corresponding frequency range and abnormal activities [43].

Bandwidth	Frequency	Abnormal activities
Delta (δ)	0.1–4 Hz	Structural lesion, encephalopathy, seizures
Theta (θ)	5–8 Hz	Encephalopathy, seizures
Alpha (α)	9–16 Hz	Coma, seizures
Beta (β)	17–30 Hz	Drug overdose, seizures

Later, it was observed that combining line length and other features into machine learning models could obtain promising seizure detection results as well [11]. As a result, we combine the spectral features and the line length feature as inputs to the seizure detection model. The line length (LL) feature is calculated on the 1-second non-overlapping segments $s_i(t)$, which is defined as

$$X_i(t) = \sum_{n=1}^{N-1} |s_i(t,n) - s_i(t,n-1)|, N = 256. \tag{4}$$

In other words, the line length feature can also be considered as the total variation within one second of EEG signal. An example of the line length feature extracted from the EEG data of patient 1 is depicted in Fig. 3, which shows clear discrimination between the feature within the seizure segment and the feature within the nonseizure segment. The line length feature extracted from one-second EEG signal is denoted as $X_t = [X_1(t), X_2(t), \dots, X_{18}(t)]^T$. Similar to the spectral feature extraction, we concatenate the line length features within the 10-second window as the final line length feature for classification and denote it as $x_L = [X_t, X_{t-1}, \dots, X_{t-9}]$. The process of extracting the line length feature is shown in Fig. 4. The label for temporal segmented data x_L (x_S) is determined as follows. If $s_i(t)$ is within the seizure segment, the label is 1. Otherwise, the label is 0. Through the above processes, we could obtain a dataset $D = \{(x_i, y_i)\}_{i=1}^{\mathcal{N}}$, $y_i \in \{0, 1\}$ for the classification model, where x_i represents the ith data composed of the spectral feature and the line length feature, $y_i = 0$ represents the nonseizure label, and $y_i = 1$ represents the seizure label.

3.2. Classification model

3.2.1. Weighted signal temporal logic

The notion of weighted signal temporal logic (wSTL) was first introduced by [44]. Similar to natural language, wSTL is a formal language in which syntax and semantics define how a wSTL formula is expressed and interpreted.

Definition 1. The syntax of wSTL is defined as follows [44].

$$\phi := \top | \pi | \neg\phi | \phi_1^{w_1} \wedge \phi_2^{w_2} | \phi_1^{w_1} \vee \phi_2^{w_2} | \Diamond_{[t_1,t_2]}^{\boldsymbol{w}} \phi | \Box_{[t_1,t_2]}^{\boldsymbol{w}} \phi,$$

where $\top$ is Boolean TRUE, π denotes an atomic predicate in the form of $f(x) > 0$, e.g., if $f(x) = \boldsymbol{a}^T x - c, \|\boldsymbol{a}\| = 1$, then π represents a half-space in $\mathbb{R}^d$, and $x \in \mathbb{R}^d$ is the temporal signal that can be either the spectral feature or the line length feature. In addition, $\neg$ is the negation operator, $\wedge$ and $\vee$ denote logical conjunction and logical disjunction, respectively. $\Diamond$ and $\Box$ are temporal operators read as "Eventually" and "Always", respectively, and $t_1, t_2 \in \mathbb{Z}_{\geq 0}, t_1 \leq t_2$, and w_1, w_2 denote the nonnegative weights associated with the subformulas ϕ_1 and ϕ_2, and $\boldsymbol{w} \in \mathbb{R}_{\geq 0}^{t_2 - t_1 + 1}$ denotes the nonnegative weights associated with $\Diamond$ or $\Box$ operators.

With the syntax in Definition 1, we could derive other forms of wSTL formulas recursively. Taking the nested "Eventually" and "Always" operators as an example, $\phi = \Diamond_{[0,5]}^{\boldsymbol{w}^1} \Box_{[0,3]}^{\boldsymbol{w}^2} (x_1 - 1 > 0)$ reads as "During 0 to 5 s, there will be at least one second such that x_1 is larger than 1 and x remains larger than 1 for 3 s".

The semantics of wSTL are divided into two categories, one is called Boolean semantics, and another is called bounded quantitative semantics. The Boolean semantics can evaluate the Boolean satisfaction of wSTL formulas over signals, i.e. $(x,t) \vDash \phi$ represents x satisfies ϕ at t, and $(x,t) \nvDash \phi$ represents x violates ϕ at t. In addition to the Boolean semantics, wSTL embodies bounded quantitative semantics, i.e., truth degrees, to reflect the degree of satisfaction or violation quantitatively.

Definition 2. Let $p(x,\phi,t)$ denote the truth degree of ϕ over x at t, which can be calculated through the bounded quantitative semantics defined as follows.

$$\begin{aligned}
&p(x,\pi,t) = g\left(f(x(t))\right),\\
&p(x,\neg\phi,t) = 1 - p(x,\phi,t),\\
&p(x,\phi_1^{w_1} \wedge \phi_2^{w_2},t) = \otimes^{\wedge}\left([w_j, p(x,\phi_j,t)]_{j=1,2}, \beta\right),\\
&p(x,\phi_1^{w_1} \vee \phi_2^{w_2},t) = \oplus^{\vee}\left([w_j, p(x,\phi_j,t)]_{j=1,2}, \beta\right),\\
&p(x,\Diamond_{[t_1,t_2]}^{\boldsymbol{w}}\phi,t) = \oplus^{\Diamond}([w_{t'}, p(x,\phi,t+t')]_{t'\in[t_1,t_2]}, \beta),\\
&p(x,\Box_{[t_1,t_2]}^{\boldsymbol{w}}\phi,t) = \otimes^{\Box}([w_{t'}, p(x,\phi,t+t')]_{t'\in[t_1,t_2]}, \beta),
\end{aligned} \tag{5}$$

where g is the activation function for the atomic predicates, $\pi := (f(x) > 0)$. and $\otimes^{\wedge}$, $\oplus^{\vee}$, $\otimes^{\Box}$, and $\oplus^{\Diamond}$ are activation functions for the $\wedge$, $\vee$, $\Box$ and $\Diamond$ operators, respectively.

3.2.2. Signal Temporal Logic Neural Network (STONE)

The design of STONE aims to seamlessly associate each neuron in a neural network with a symbolic representation in the wSTL. In such manner, STONE can enjoy the advantages of both neural networks and wSTL so that it can generate interpretable and human-readable wSTL formulas through learning as a neural network. In this subsection, the explicit activation functions for (5) are given to accomplish the above goals.

The activation function for the predicate π, $g(f(x(t)))$, is defined as the sigmoid function

$$g(f(x(t))) = \frac{1}{1 + e^{-f(x(t))}}. \tag{6}$$

Note that the motivation of (6) is that $g(f(x(t)))$ represents the degree of truth of π over x at t, thus $g(f(x(t))) \in [0,1]$. If $p(x,\pi,t)$ is close to 1, then it reflects that x robustly satisfies π at t, and if $p(x,\pi,t)$ is close to 0, then it reflects that x robustly violates π at t. Additionally, the choice of $g(f(x(t)))$ reflects that the truth degree of π over x at t becomes larger when $g(f(x(t)))$ increases. Throughout this paper, we simplify $p(x,\phi,0)$ as $p(x,\phi)$. The relationship between the bounded quantitative semantics and the Boolean semantics is that $p(x,\phi,t) > 0.5$ means $(x,t) \vDash \phi$.

From the definition of $\Diamond$ and $\Box$ operators, we know that if $\Diamond_{[t_1,t_2]}\phi$ is evaluated on x at t, then it represents "ϕ is eventually satisfied between $t+t_1$ and $t+t_2$", which is the same as "ϕ is satisfied at least one time slot in $[t+t_1, t+t_2]$". With this interpretation, $\Diamond_{[t_1,t_2]}$ can be written in terms of a sequence of $\vee$ operators as follows

$$\Diamond_{[t_1,t_2]}\phi = \phi_{t_1}^{w_{t_1}} \vee \phi_{t_1+1}^{w_{t_1+1}} \vee \dots \vee \phi_{t_2}^{w_{t_2}}, \tag{7}$$

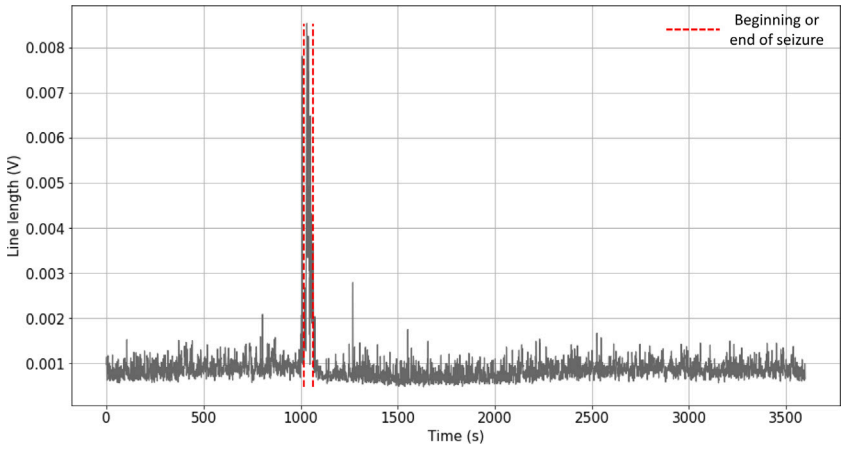


Fig. 3. Illustration of the LL feature extracted from patient 1's EEG data with annotated beginning and end of the seizure.

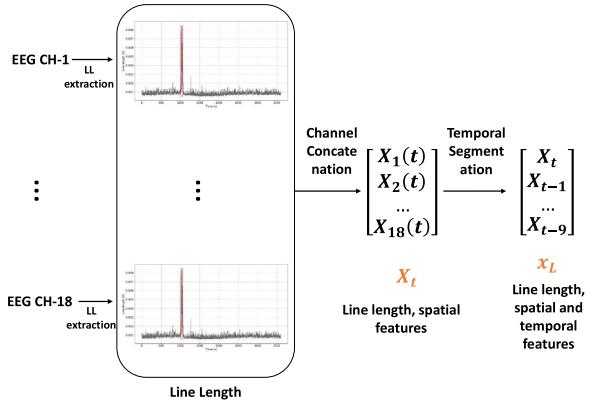


Fig. 4. Illustration of the line length extraction process.

where $\phi_{t'}$ denotes the formula ϕ evaluated on the data instance $x(t'), t' \in [t_1, t_2]$. Similarly, $\square_{[t_1,t_2]}\phi$ can be written in terms of a sequence of $\wedge$ operators. As a consequence, the activation functions for $\square$ and $\Diamond$ operators can be derived from the activation functions for $\wedge$ and $\vee$ operators. Also, the activation functions for $\wedge$ and $\vee$ operators should satisfy the De Morgan's law [45], which is a traditional rule in propositional logic and Boolean algebra. The wSTL activation functions should be defined in such a way that $p(x, \phi, t)$ measures the degree of satisfaction of ϕ over x at t, considering the importance of subformulas and time. For simplicity, we only discuss the design principles for the activation function for the $\wedge$ operator, and these principles also hold for the other operators due to the aforementioned reasons. The activation function for the $\wedge$ operator should satisfy the following properties:

- Nonimpact for zero weights: subformulas with zero weights have no impact on the truth degree of the overall formula, i.e., if $w_j = 0$, $(j = 1, 2)$, then $p(x, \phi_j, t)$ has no impact on $p(x, \phi_1^{w_1} \wedge \phi_2^{w_2}, t)$, where w_j are the weights defined in (5).
- Ordering of impact: impact of truth degree of subformulas on the truth degree of the overall formula follows the order of their weights, i.e., if $p(x, \phi_1, t) = p(x, \phi_2, t)$ and $w_1 \geq w_2$, then

$$\frac{\partial p(x, \phi_1^{w_1} \wedge \phi_2^{w_2}, t)}{\partial p(x, \phi_1, t)} \geq \frac{\partial p(x, \phi_1^{w_1} \wedge \phi_2^{w_2}, t)}{\partial p(x, \phi_2, t)}. \tag{8}$$

- Monotonicity: The truth degree, $p(x, \phi_1^{w_1} \wedge \phi_2^{w_2}, t)$, increases monotonically with $p(x, \phi_j, t), j = 1, 2$, i.e., for $0 \leq \Delta p \leq 1$, the following holds:

$$\otimes^{\wedge}\left([w_j, p(x, \phi_j, t)]_{j=1,2}, \beta\right) \leq \otimes^{\wedge}([w_j, p(x, \phi_j, t) + \Delta p]_{j=1,2}, \beta). \tag{9}$$

With the above consideration, we define the activation functions for wSTL operators that exhibit the above properties as follows.

$$\otimes^{\wedge}\left([w_j, p(x, \phi_j, t)]_{j=1,2}, \beta\right) \triangleq h(\beta - \sum_{j=1}^{2} \bar{w}_j(1 - p(x, \phi_j, t))), \tag{10}$$

$$\oplus^{\vee}\left([w_j, p(x, \phi_j, t)]_{j=1,2}, \beta\right) \triangleq h(1 - \beta + \sum_{j=1}^{2} \bar{w}_j p(x, \phi_j, t)),$$

$$\oplus^{\Diamond}\left([w_{t'}, p(x, \phi, t + t')]_{t' \in [t_1,t_2]}, \beta\right) \triangleq$$

$$h(1 - \beta + \sum_{t' \in [t_1,t_2]} \bar{w}_{t'} p(x, \phi, t + t') \mathbb{1}(t + t' < T)),$$

$$\otimes^{\square}\left([w_{t'}, p(x, \phi, t + t')]_{t' \in [t_1,t_2]}, \beta\right) \triangleq$$

$$h(\beta - \sum_{t' \in [t_1,t_2]} \bar{w}_{t'}\left(1 - p(x, \phi, t + t')\right) \mathbb{1}(t + t' < T)),$$

where $\beta \geq 0$ is a parameter to learn, and $h(z) = \max\{0, \min\{z, 1\}\}$ such that the truth degree can be clamped into $[0, 1]$, and $\bar{w}_j$ or $\bar{w}_{t'}$ denote the normalized weight, and

$$\mathbb{1}(t + t' < T) = \begin{cases} 1 & \text{if } t + t' < T, \\ 0 & \text{otherwise,} \end{cases} \tag{11}$$

which is to ignore the impact of signals beyond T.

With the activation functions defined in (10), we could construct a Signal Temporal lOgic Neural nEtwork (STONE) to perform wSTL formula learning tasks, where each neuron has an associated symbolic expression that represents a component of a wSTL formula. For example, the STONE for the wSTL formula $\phi = \Diamond_{[0,2]}^{\boldsymbol{w}^3} \square_{[0,2]}^{\boldsymbol{w}^2}(\boldsymbol{a}^T x > c)$ is composed of neurons representing atomic predicates $\pi := (\boldsymbol{a}^T x > c)$, and neurons representing the $\Diamond$ and $\square$ operators. Specifically, the architecture of STONE for ϕ is depicted in Fig. 5, where the parameters to learn are $\boldsymbol{a}, c, w_1^2, w_2^2, w_3^2, \beta^2, w_1^3, w_2^3, w_3^3, \beta^3$. Note that the three neurons in the third layer essentially represent the same component $\square_{[0,2]}$ and share the same parameters $(w_1^2, w_2^2, w_3^2, \beta^2)$, but the neurons are plotted separately so that the propagation process is clear. The first layer represents the input data x that can be either spectral features x_S or the line length feature x_L, and the second layer represents the atomic predicates, and the third layer represents the $\square$ operator, and the fourth layer represents the $\Diamond$ operator, and the output layer represents the truth degree of ϕ over x at $t = 0$. The above example indicates that for any ϕ, we could construct a corresponding STONE whose structure is determined by ϕ.

3.2.3. Learning of STONE

Unweighted Objective Function. The classification problem tackled in this paper is a binary classification problem. Note that the output of STONE is the truth degree of ϕ over x, which is a value between 0

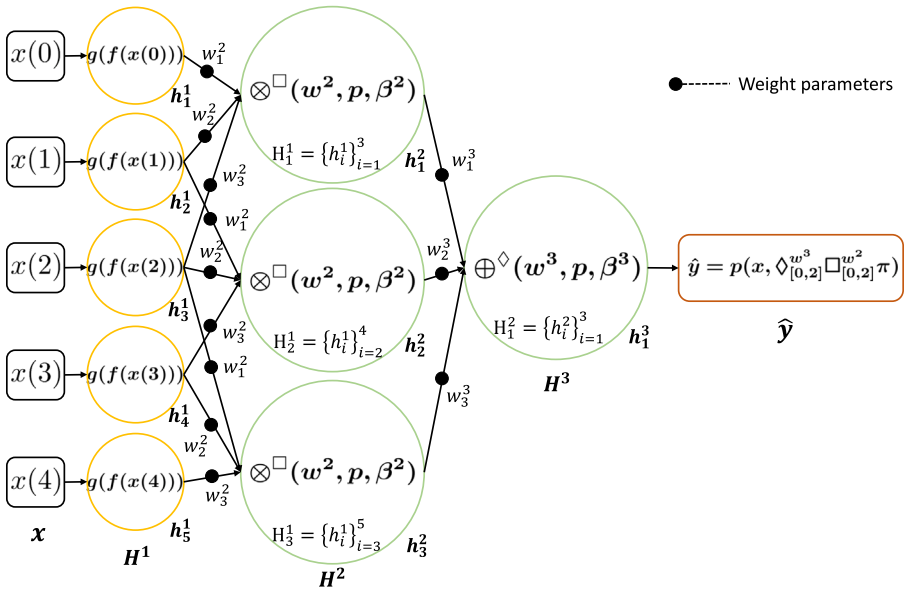


Fig. 5. Model structure of STONE for $\phi = \Diamond_{[0,2]}^{w^3} \square_{[0,2]}^{w^2} (\boldsymbol{a}^T x > c)$.

and 1. Alternatively, we can consider it as the probability of a data instance belonging to the seizure class. Hence the following standard cross-entropy loss function can be considered as a candidate objective function:

$$\mathcal{L}_f(D, \phi) = \sum_{i=1}^{\mathcal{N}} -y_i \log p(x_i, \phi) - (1 - y_i) \log(1 - p(x_i, \phi)). \tag{12}$$

Weighted Objective Function. Due to the fact that epileptic seizure only happens in certain recordings, and in most cases, the duration of epileptic seizure is less than two minutes [2], the number of seizure data is much less than the number of nonseizure data. This would cause a data imbalance issue in the dataset, which means the class distribution is biased or skewed instead of balanced or close to balanced. Data imbalance is a long-existing challenge for machine learning models as many machine learning models are designed based on the assumption that the size of data in each class are more or less equal. For example, the objective function in (12) assigns the same weights to the two classes of data. If we adopt this objective function in the learning process, then the model will have poor prediction performance, especially for the seizure class, which is the minority class. Several popular techniques for tackling data imbalance issues have been proposed, such as oversampling for the minority class and undersampling for the majority class. One of the most effective oversampling techniques is the synthetic minority over-sampling technique (SMOTE) [46]. Although these approaches could balance out the dataset, they would introduce new issues. Oversampling generates more samples in the minority class, which will slow down the learning process and may cause overlapping of classes resulting in noises in the data. Undersampling removes a certain number of observations, which may raise the issues of missing valuable information, underfitting, and poor generalization on the validation data.

In contrast, a more reasonable approach is to modify the objective function. By introducing class weights into the objective function, the model could realize penalizing more on the minority class. Accordingly, manipulating the objective function could achieve better results without adjusting the dataset. To quantify the penalization, we introduce the notion of imbalance ratio as

$$I_R = \frac{N_n}{N_p}, \tag{13}$$

where $N_p = |\{i | y_i = 1\}|$ is the number of seizure data, and $N_n = |\{i | y_i = 0\}|$ is the number of nonseizure data. Consequently, the objective function becomes the following weighted cross-entropy function

$$\mathcal{L}(D, \phi) = \sum_{i=1}^{\mathcal{N}} -I_R \cdot y_i \log p(x_i, \phi) - (1 - y_i) \log(1 - p(x_i, \phi)). \tag{14}$$

Forward Propagation. Note that STONE is a typical neural network where each neuron is endowed a symbolic representation within a wSTL formula. The forward propagation of STONE is similar to a classical NN, where every neuron in the hidden layers takes the outputs from the preceding layer as inputs and computes the outputs using the activation function of the component that the neuron represents. Suppose $x = \{x(0), x(1), \ldots, x(T-1)\}$ is the input layer with T nodes, $\hat{y}$ represents the output layer, L denotes the number of hidden layers, H^l represents the lth hidden layer with N_l nodes, h_m^l represents the mth node in the lth layer, and $H^0 = x$ and $H^{L+1} = \hat{y}$, and $H_{n_l}^{l-1}$ represents the outputs from the $(l-1)$-th hidden layer that are used as inputs of the n_l-th node in the lth hidden layer, and $\boldsymbol{w}_{n_l}^l$ represents the weights for the n_l-th node in the lth hidden layer, $\beta_{n_l}^l$ is the bias of the n_l-th node in the lth layer. Specifically, Fig. 5 gives a detailed explanation on the above notations. The output of the n_l-th node in the lth layer is computed by

$$h_{n_l}^l = \boldsymbol{\Phi}_{n_l}^l (H_{n_l}^{l-1}, \boldsymbol{w}_{n_l}^l, \beta_{n_l}^l), \tag{15}$$

where $\boldsymbol{\Phi}_{n_l}^l$ can be any activation function in (10) depending on the component that the n_l-th node in the lth layer represents. The output of STONE is the same as the output of the Lth hidden layer, which is computed by

$$\hat{y} = h_1^L = \boldsymbol{\Phi}^L (H_1^{L-1}, \boldsymbol{w}_1^L, \beta_1^L). \tag{16}$$

Backward Propagation. For a data instance x with label y, the loss function is defined as

$$\mathcal{L}(y, \hat{y}) = -I_R y \log(\hat{y}) - (1 - y) \log(1 - \hat{y}). \tag{17}$$

The derivative of $L(y, \hat{y})$ with respect to $\hat{y}$ is

$$\nabla \hat{y} = \frac{(1 - y)}{1 - \hat{y}} - \frac{I_R y}{\hat{y}}. \tag{18}$$

For the output layer, $\hat{y}$ is obtained via (16), hence the gradients of $\mathcal{L}(y, \hat{y})$ with respect to $H_1^{L-1}, \boldsymbol{w}_1^L, \beta_1^L$ are computed by the chain rule as

follows:

$$\nabla H_1^{L-1} = \frac{\partial \Phi^L}{\partial H_1^{L-1}} \nabla \hat{y}, \nabla \boldsymbol{w}_1^L = \frac{\partial \Phi^L}{\partial \boldsymbol{w}_1^L} \nabla \hat{y},$$
$$\nabla \beta_1^L = \frac{\partial \Phi^L}{\partial \beta_1^L} \nabla \hat{y}. \tag{19}$$

Similarly, from the $L-1$-th hidden layer to the 1st hidden layer, the gradients for $H_{n_l}^{l-1}, \boldsymbol{w}_{n_l}^l, \beta_{n_l}^l$ are computed using the same procedures as (19). Specifically, the gradients of the activation functions with respect to the parameters are given as follows (assuming the truth degree before clamping falls in the range of $(0, 1)$).

- For the activation function of the $\wedge$ or the $\square$ operator, the gradients are

$$\frac{\partial \Phi_{n_l}^l}{\partial \beta_{n_l}^l} = 1, \quad \frac{\partial \Phi_{n_l}^l}{\partial w_{n_l,j}^l} = -(1 - H_{n_l,j}^{l-1}),$$
$$\frac{\partial \Phi_{n_l}^l}{\partial H_{n_l,j}^{l-1}} = w_{n_l,j}^l,$$

  where $H_{n_l,j}^{l-1}$ is the jth element in the $H_{n_l}^{l-1}$, and $w_{n_l,j}^l$ is the weight associated with $H_{n_l,j}^{l-1}$.

- For the activation function of the $\vee$ or the $\Diamond$ operator, the gradients are

$$\frac{\partial \Phi_{n_l}^l}{\partial \beta_{n_l}^l} = -1, \quad \frac{\partial \Phi_{n_l}^l}{\partial w_{n_l,j}^l} = H_{n_l,j}^{l-1},$$
$$\frac{\partial \Phi_{n_l}^l}{\partial H_{n_l,j}^{l-1}} = w_{n_l,j}^l.$$

With the above gradient derivation, the gradients of all the parameters in STONE can be computed through the chain rule. Hence the parameters can be learned through backward propagation. In this paper, we employ the *AdamW* optimizer [47] in the *Pytorch* library to optimize the parameters in STONE.

3.2.4. Evaluation Metrics

We adopt a stratified k-fold cross-validation strategy to carry out the experiment. Stratified k-fold cross-validation refers to the original dataset is randomly partitioned into k folds of data with equal size and each fold contains approximately the same proportions of two classes of data (seizure and nonseizure data). The cross-validation process is conducted with k experiments, where in the k_i-th experiment, the k_i-th fold is retained as the validation set to evaluate the model, and the remaining $k-1$ folds are utilized as training data. In this manner, each fold will be utilized as a validation set exactly once, and the results from the k experiments are averaged as the final estimation. Here k is set as five, and the average result of the five experiments on each patient is reported.

For evaluation of the proposed STONE, six commonly used evaluation measures are adopted here [16,48], including Accuracy (*Acc*), Sensitivity (*Sen*), Specificity (*Spe*), F-Score (*FS*), AUC score (*AUC*), and Kappa score (*Kap*). Suppose the seizure class is the positive class, and the nonseizure class is the negative class. The evaluation metrics are computed by the following four outcomes from STONE: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). The explanation of these four outcomes is presented in Table 3.

The evaluation metrics of *Acc*, *Sen*, *Spe*, *FS*, *AUC*, and *Kap* have been exploited in many state-of-the-art (SOTA) methods [13,15,16]. Using these evaluation metrics provides a fair comparison with SOTA methods as well. Specifically, the evaluation metrics are defined as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}.$$

Acc measures the fraction of predictions that a model got correct. For an imbalanced dataset such as the CHB-MIT database, the nonseizure group size is much larger than the seizure group size, which may lead to a high *Acc* if all the data is predicted to be nonseizure. Therefore, the evaluation measures of *Sen* and *Spe* are introduced to reflect the categorical prediction accuracy. *Sen* is defined as

$$Sen = \frac{TP}{TP + FN}.$$

Sen reflects the fraction of data that are correctly predicted in the positive class. In other words, *Sen* measures the fraction of seizure data that are correctly detected, which mirrors a model's capability of detecting seizures. *Spe* is defined as

$$Spe = \frac{TN}{TN + FP}.$$

Spe reflects the fraction of data that are correctly predicted in the negative class. Equivalently, *Spe* measures how much nonseizure data is correctly identified, which mirrors a model's false alarm rate.

From the above definition, it is evident that the two key factors for evaluating a seizure detection model are *Sen* and *Spe*. Here we define a *FS* different from the conventional notion, which is based on the *Sen* and *Spe* such that *FS* considers the sensitivity and specificity of a model simultaneously. *FS* is defined as

$$FS = \frac{2 \times Sen \times Spe}{Sen + Spe}. \tag{20}$$

AUC is another evaluation metric exploited in this work, which reflects the area under the Receiver Operating Characteristics (ROC) curve. ROC curve is a performance measurement for the seizure detection problem at various threshold settings.

Kap is a statistic measurement for inter-rater reliability, which is considered to be a more robust measurement than *Acc*, as *Kap* takes the possibility of the correct prediction occurring by chance into consideration. *Kap* is defined as

$$Kap = \frac{p_o - p_e}{1 - p_e}, \tag{21}$$

where $p_o = Acc$, and

$$p_e = \frac{(TP + FN) \times (TP + FP) + (TN + FP) \times (TN + FN)}{(TP + TN + FP + FN)^2}. \tag{22}$$

Table 3
Explanation of four classification outcomes from STONE.

Terminology	Explanation
TP	The number of seizure data that is classified correctly
FN	The number of seizure data that is classified as nonseizure
FP	The number of nonseizure data that is classified as seizure
TN	The number of nonseizure data that is classified correctly

4. Experimental Results

This section tests STONE on the CHB-MIT and the Bonn University databases to demonstrate its performance for seizure detection tasks. Meanwhile, the interpretability analysis of the rules learned by STONE is carried out through visualization. Aside from STONE, conventional machine learning (ML) models, including an artificial neural network (ANN) that is implemented as a multi-layer perceptron (MLP), a K-nearest neighbors (KNN) classifier, a supported vector machine (SVM) classifier, a decision tree (DT) classifier, a bagging classifier (BA) with SVMs as base classifiers, and an AdaBoost (AB) classifier with decision trees as base estimators, are also tested on the two databases. The comparison of STONE with existing works is presented, which manifests the advantage of STONE over the existing seizure detection models.

4.1. Formula Structures

As seen in Fig. 5, STONE's structure is determined by the structure of the formula it represents. In this paper, we consider using the following formula structures that have been widely used in prior STL inference works [27]:

(1) Conjunctive patterns: $\phi = \phi_S^{w_1^1} \wedge \phi_L^{w_2^1}$, where

$$\phi_S = \pi_{t,1}^{w^{t,1}} \wedge \pi_{t,2}^{w^{t,2}} \wedge \cdots \wedge \pi_{t,7}^{w^{t,7}} \wedge \pi_{t-1,1}^{w^{t-1,1}} \wedge \pi_{t-1,2}^{w^{t-1,2}} \wedge \cdots \wedge \pi_{t-1,7}^{w^{t-1,7}} \wedge \cdots \wedge \pi_{t-9,1}^{w^{t-9,1}} \wedge \pi_{t-9,2}^{w^{t-9,2}} \wedge \cdots \wedge \pi_{t-9,7}^{w^{t-9,7}},$$

and $\pi_{t',j}$ is an atomic predicate describing the jth spectral feature at time t'. Similarly, ϕ_L is defined as

$$\phi_L = \pi_t^{w^t} \wedge \pi_{t-1}^{w^{t-1}} \wedge \cdots \wedge \pi_{t-9}^{w^{t-9}}. \tag{23}$$

This formula structure provides information about what conjunctive patterns of the spectral features and the line length hold will cause seizures.

(2) Disjunctive patterns: $\phi = \phi_S^{w_1^1} \vee \phi_L^{w_2^1}$, where

$$\phi_S = \pi_{t,1}^{w^{t,1}} \vee \pi_{t,2}^{w^{t,2}} \vee \ldots \vee \pi_{t,7}^{w^{t,7}} \vee \pi_{t-1,1}^{w^{t-1,1}} \vee \pi_{t-1,2}^{w^{t-1,2}} \vee \ldots \vee \pi_{t-1,7}^{w^{t-1,7}} \vee \ldots \vee \pi_{t-9,1}^{w^{t-9,1}} \vee \pi_{t-9,2}^{w^{t-9,2}} \vee \ldots \vee \pi_{t-9,7}^{w^{t-9,7}}.$$

The subformula ϕ_L is defined as

$$\phi_L = \pi_t^{w^t} \vee \pi_{t-1}^{w^{t-1}} \vee \ldots \vee \pi_{t-9}^{w^{t-9}}. \tag{24}$$

This formula structure provides information about what disjunctive patterns of the spectral features and the line length hold will cause seizures.

(3) Consistent patterns: $\phi = \phi_S^{w_1^1} \wedge \phi_L^{w_2^1}$, where

$$\phi_S = \Box_{[1,10]}^{\boldsymbol{w}^2}(\pi_1^{w^{3,1}} \wedge \pi_2^{w^{3,2}} \wedge \cdots \wedge \pi_7^{w^{3,7}}), \tag{25}$$

and π_j is an atomic predicate defined on the jth spectral feature from 18 channels. Also, ϕ_L is defined as $\phi_L = \Box_{[1,10]}^{\boldsymbol{w}^4}(\pi)$, where π is an atomic predicate defined on the line length feature from 18 channels. This formula structure describes what patterns of the spectral features and the line length hold consistently will cause seizures.

(4) Alternative patterns: $\phi = \phi_S^{w_1^1} \vee \phi_L^{w_2^1}$, where

$$\phi_S = \Diamond_{[1,10]}^{\boldsymbol{w}^2}(\pi_1^{w^{3,1}} \wedge \pi_2^{w^{3,2}} \wedge \cdots \wedge \pi_7^{w^{3,7}}). \tag{26}$$

Also, ϕ_L is defined as $\phi_L = \Diamond_{[1,10]}^{\boldsymbol{w}^4}(\pi)$. This formula structure describes what patterns of the spectral features and the line length hold alternatively will cause seizures.

(5) Persistent patterns: $\phi = \phi_S^{w_1^1} \wedge \phi_L^{w_2^1}$, where

$$\phi_S = \Box_{[0,t_1]}^{\boldsymbol{w}^2} \Diamond_{[0,t_2]}^{\boldsymbol{w}^3}(\pi_1^{w^{4,1}} \wedge \pi_2^{w^{4,2}} \wedge \cdots \wedge \pi_7^{w^{4,7}}), \tag{27}$$

and $\phi_L = \Box_{[0,t_1]}^{\boldsymbol{w}^5} \Diamond_{[0,t_2]}^{\boldsymbol{w}^6}(\pi)$. This formula structure describes what patterns of the spectral features and the line length feature hold persistently will cause seizures.

(6) Eventually consistent patterns: $\phi = \phi_S^{w_1^1} \vee \phi_L^{w_2^1}$, where

$$\phi_S = \Diamond_{[0,t_1]}^{\boldsymbol{w}^2} \Box_{[0,t_2]}^{\boldsymbol{w}^3}(\pi_1^{w^{4,1}} \wedge \pi_2^{w^{4,2}} \wedge \cdots \wedge \pi_7^{w^{4,7}}), \tag{28}$$

and $\phi_L = \Diamond_{[0,t_1]}^{\boldsymbol{w}^5} \Box_{[0,t_2]}^{\boldsymbol{w}^6}(\pi)$. This formula structure describes what patterns of the spectral features and the line length feature hold eventually consistently will cause seizures.

4.2. Classification Results

In this section, the classification results for STONE and the six conventional ML classifiers are presented, and a comparison between STONE and the existing classifiers' performance is carried out.

4.2.1. Classification Results on the CHB-MIT Database

Classification Results of STONE. In the experiment, the parameters defining the predicates and the weight parameters are initialized as samples from the normal distribution with mean 0 and variance 1, and β's are initialized as 1. Before the data is given to the classifier, it is normalized by removing the mean and scaling to unit variance. The Kappa score is not an appropriate criterion for the CHB-MIT database as the CHB-MIT database is an imbalanced dataset. The Kappa score may exhibit undesired behavior in imbalanced datasets [49]. The average *Acc*, *Sen*, *Spe*, *FS*, and *AUC* of the cross-validation experiments are reported in Table 4, where the evaluation measures for each patient are reported as the best classification result from the four formula structures defined in Section 4.1. It can be observed from Table 4 that the classification accuracy, sensitivity, and specificity of higher than 99.00% are achieved for patients 2, 10, 11, 21, 22, 23. Sensitivities for patients 7, 17, and 23 all achieve 100%, which means that STONE can detect seizures perfectly for these three patients. The best classification result is observed on patient 23, for whom the *Acc, Sen, Spe, and FS* are 99.97%, 100%, 99.97%, and 0.9999, respectively. It can also be observed that STONE does not have an ideal result on patient 6 because of misclassification of nonseizures, and patient 14 and patient 16 due to missing detection of seizures.

Classification Results of ML Models. To demonstrate the advantage of STONE over the conventional ML models, the classifiers of BA, AB, DT, KNN, ANN, and SVM are also implemented on the two databases. The average performance of the above ML models over the 24 patients in the CHB-MIT database is presented in Table 5.

Table 5 shows that the Bagging classifier can achieve an accuracy of 99.34%, the AdaBoost classifier can achieve an accuracy of 99.40%, the DT classifier can achieve an accuracy of 98.90%, and the KNN classifier can achieve an accuracy of 99.60%, and the ANN classifier can achieve an accuracy of 99.85%, and the SVM classifier can achieve an accuracy of 99.44%. The classifiers of BA, AB, DT, KNN, ANN, and SVM all achieve a classification accuracy higher than STONE. Nevertheless, the BA, AB, DT, KNN, ANN, and SVM classifiers attain a sensitivity of 66.83%, 80.24%, 75.67%, 80.21%, 92.51%, and 73.09%, respectively, which means they cannot attain a fairly comparable sensitivity result with STONE. Equivalently, the conventional ML models did not perform well in detecting seizures. As shown in Table 5, the high classification accuracy of the conventional ML models is caused by the high sensitivity. Due to the fact that the dataset is imbalanced, where the number of non-seizure data is much larger than the number of seizure data, the classifiers are trained to recognize the nonseizures more accurately and recognize the seizures less accurately. By introducing the weighted cross-entropy loss into STONE, we observe that STONE can address this limitation and achieve both high sensitivity and specificity. More importantly, STONE can both provide decent seizure detection results and generate signal temporal logic (STL) formulas that are interpretable and human-readable.

Comparison With Existing Methods. To better demonstrate the performance of STONE on seizure detection tasks, some existing works that use the CHB-MIT database with different feature extraction techniques and various machine learning models are considered. The comparison of *Acc, Sen, Spe* of STONE and some recent works is presented in Table 6. It is evident from Table 6 that STONE could achieve better or comparable performance in terms of *Acc, Sen, Spe* with the existing models. Even though some related works achieve slightly better results, STONE enjoys the advantages of interpretability and transparency, which is a missing attribute of the existing classifiers. Chen et al. [13] exploited DWT to extract features from the EEG data, based on which statistical features are acquired as inputs to the SVM classifier. As shown by Table 6, STONE exhibits better performance than the existing works [12–14,19] in terms of higher values of *Acc, Sen*, and *Spe*. Although STONE attains a sensitivity score slightly lower than the 98.28% sensitivity score reported in [18], STONE achieves

Table 4
Performance of STONE on the CHB-MIT database in terms of *Acc*, *Sen*, *Spe*, *FS*, *AUC*.

Subject	*Acc* (%)	*Sen* (%)	*Spe* (%)	*FS*	*AUC*
1	98.94 ± 0.49	97.96 ± 1.50	98.96 ± 0.48	0.9845 ± 0.0090	0.9962 ± 0.0040
2	99.22 ± 0.35	99.43 ± 1.14	99.22 ± 0.35	0.9932 ± 0.0071	0.9982 ± 0.0015
3	98.92 ± 0.14	98.75 ± 0.79	98.92 ± 0.13	0.9884 ± 0.0043	0.9934 ± 0.0058
4	98.36 ± 0.26	98.47 ± 1.43	98.36 ± 0.25	0.9841 ± 0.0081	0.9960 ± 0.0035
5	99.23 ± 0.28	97.85 ± 0.99	99.26 ± 0.28	0.9854 ± 0.0055	0.9961 ± 0.0033
6	94.18 ± 1.26	97.10 ± 2.39	94.13 ± 1.26	0.9589 ± 0.0163	0.9787 ± 0.0158
7	98.75 ± 0.67	100.0 ± 0.00	98.73 ± 0.68	0.9931 ± 0.0035	0.9986 ± 0.0019
8	98.51 ± 0.22	98.15 ± 0.65	98.55 ± 0.18	0.9835 ± 0.0040	0.9923 ± 0.0059
9	99.57 ± 0.39	98.54 ± 1.95	99.59 ± 0.39	0.9905 ± 0.0105	0.9931 ± 0.0071
10	99.75 ± 0.22	99.11 ± 0.83	99.77 ± 0.25	0.9944 ± 0.0035	0.9981 ± 0.0021
11	99.65 ± 0.19	99.38 ± 0.55	99.66 ± 0.18	0.9952 ± 0.0034	0.9994 ± 0.0003
12	97.83 ± 0.53	98.16 ± 1.05	97.81 ± 0.59	0.9798 ± 0.0051	0.9911 ± 0.0049
13	98.93 ± 0.33	95.46 ± 2.55	99.01 ± 0.33	0.9719 ± 0.0135	0.9821 ± 0.0075
14	96.46 ± 1.41	93.22 ± 4.02	96.50 ± 1.42	0.9501 ± 0.0181	0.9691 ± 0.0199
15	98.66 ± 0.42	97.46 ± 0.82	98.77 ± 0.45	0.9811 ± 0.0047	0.9957 ± 0.0010
16	96.62 ± 2.40	94.55 ± 6.80	96.64 ± 2.38	0.9548 ± 0.0445	0.9776 ± 0.0190
17	98.45 ± 0.15	100.0 ± 0.00	98.41 ± 0.16	0.9920 ± 0.0008	0.9986 ± 0.0005
18	97.25 ± 1.48	97.78 ± 2.37	97.24 ± 1.50	0.9750 ± 0.0144	0.9911 ± 0.0061
19	98.88 ± 0.26	99.15 ± 1.04	98.87 ± 0.25	0.9900 ± 0.0060	0.9954 ± 0.0041
20	97.89 ± 0.88	97.07 ± 0.46	97.90 ± 0.90	0.9745 ± 0.0237	0.9968 ± 0.0007
21	99.34 ± 0.44	99.50 ± 0.00	99.35 ± 0.44	0.9946 ± 0.0022	0.9974 ± 0.0033
22	99.51 ± 0.47	99.51 ± 0.97	99.51 ± 0.46	0.9951 ± 0.0069	0.9994 ± 0.0006
23	99.97 ± 0.05	100.0 ± 0.00	99.97 ± 0.05	0.9999 ± 0.0002	0.9989 ± 0.0017
24	98.90 ± 0.50	99.50 ± 1.06	98.90 ± 0.53	0.9920 ± 0.0040	0.9963 ± 0.0024
Average	98.49 ± 0.57	98.17 ± 1.56	98.50 ± 0.58	0.9834 ± 0.0091	0.9929 ± 0.0051

Table 5
ML models' performance on the CHB-MIT database in terms of *Acc*, *Sen*, *Spe*, *FS*, *AUC*.

Classifier	*Acc* (%)	*Sen* (%)	*Spe* (%)	*FS*	*AUC*
BA	99.34 ± 0.09	66.83 ± 4.38	99.97 ± 0.01	0.8010 ± 0.0386	0.8710 ± 0.0278
AB	99.40 ± 0.12	80.24 ± 4.64	99.84 ± 0.06	0.8897 ± 0.0346	0.9847 ± 0.0079
DT	98.90 ± 0.16	75.67 ± 5.25	99.42 ± 0.13	0.8593 ± 0.0391	0.8755 ± 0.0262
KNN	99.60 ± 0.07	80.21 ± 3.58	99.98 ± 0.01	0.8901 ± 0.0274	0.9619 ± 0.0135
ANN	99.85 ± 0.06	92.51 ± 3.24	99.98 ± 0.01	0.9610 ± 0.0202	0.9959 ± 0.0037
SVM	99.44 ± 0.09	73.09 ± 4.88	99.95 ± 0.01	0.8443 ± 0.0391	0.9947 ± 0.0044

the same specificity score and higher accuracy. Furthermore, STONE is much easier to interpret compared with SVM. Despite of the fact that the specificity score of 98.57% reported in [15] is slightly higher than STONE's 98.50%, STONE acquires an accuracy of 98.49% and a sensitivity of 98.17%, which are higher than the accuracy of 98.21% and sensitivity of 97.85% in [15]. In other words, STONE is better at diagnosing seizures than the model in [15]. In addition, STONE can provide human-readable formulas as outputs. Even though [16] develops a seizure detection model that can achieve better accuracy and sensitivity than STONE, the random forest (RF) classifier is an ensemble learning model whose classification result depends on the predictions from multiple decision trees, resulting in the model to be complicated and hard to understand. By contrast, STONE is simpler than RF in that an STL formula akin to natural language can be obtained from STONE. In summary, the major advantage of STONE over existing works is that STONE can provide wSTL formulas as interpretable and transparent rules for determining the seizure class and for a better understanding of the classifier through visualization, as shown later, which is a property that the existing seizure detection models do not possess.

4.2.2. Classification Results on the Bonn University Database

Classification Results of STONE. The parameter initialization of STONE on the Bonn University database is the same as the setting for the CHB-MIT database. As the Bonn University database is a balanced dataset, the Kappa score (*Kap*) is a proper evaluation metric for this dataset. Hence *Kap* is introduced into the evaluation process of the Bonn University database. The *Acc, Spe, Sen, FS, AUC,* and *Kap* for different classification problems is shown in Table 7. It can be observed that classification accuracy of 99.00%, 99.50%, 99.50%, and 100.0% are achieved for Cases of F–S, N–S, O–S, and Z–S, respectively, using the five fold cross validation approach. STONE also achieves a sensitivity score of 100% for cases O–S and Z–S, and a sensitivity score of 99.00% for cases F–S and N–S. In terms of specificity, all the nonseizures are correctly identified for the cases of N–S and Z–S. The cases of F–S and O–S achieve a sensitivity of 99.00%. The results in Table 7 demonstrate that the proposed STONE can handle all four cases of classification problems satisfactorily.

Table 6
Comparison of performance of STONE and recent works on the CHB-MIT database.

Authors	Features & classifiers	*Acc*	*Sen*	*Spe*
[13]	DWT, statistical features + SVM	92.3	91.7	92.89
[14]	Local Mean Decomposition + LSTM	92.66	93.61	91.85
[18]	FFT, deep features through PCANet + SVM	98.47	98.28	98.50
[15]	Stein-kernel based sparse representation	98.21	97.85	98.57
[19]	DWT, statistical features + SVM	97.09	96.81	97.26
[16]	Multiscale spectral features + RF	98.97	98.12	99.17
[12]	Entropy, variance, energy + SVM, NB	95.63	96.55	95.70
This paper	FFT, LL + STONE	98.49	98.17	98.50

Classification Results of ML Models. The average results of the four classification problems (F–S, N–S, O–S, Z–S) for the ML models on the Bonn University database is summarized in Table 8. Table 8 demonstrates that the classifiers of BA, AB, DT, KNN, ANN, and SVM can achieve decent results on the Bonn University database. Specifically, BA and AB can achieve higher than 98% accuracy, and DT, KNN, and SVM can achieve higher than 97% accuracy. In terms of sensitivity, ANN achieves the highest score compared with the other ML models, and BA, KNN, and SVM can achieve higher than 98% sensitivity. With regard to specificity, AB can achieve 98.75% sensitivity score, and BA, DT, and SVM classifiers achieve a specificity of higher than 97%. Though ANN acquires a sensitivity higher than the other ML models, it exhibits

Table 7
Performance of STONE on the Bonn University database for four classification problems.

Cases	*Acc* (%)	*Sen* (%)	*Spe* (%)	*FS*	*AUC*	*Kap*
F–S	99.00 ± 1.22	99.00 ± 2.00	99.00 ± 2.00	0.9897 ± 0.0125	0.9923 ± 0.0131	0.9800 ± 0.0244
N–S	99.50 ± 1.00	99.00 ± 2.00	100.0 ± 0.00	0.9950 ± 0.0102	0.9995 ± 0.0010	0.9900 ± 0.0200
O–S	99.50 ± 1.00	100.0 ± 0.00	99.00 ± 2.00	0.9950 ± 0.0102	0.9945 ± 0.0010	0.9900 ± 0.0200
Z–S	100.0 ± 0.00	100.0 ± 0.00	100.0 ± 0.00	1.0000 ± 0.0000	1.0000 ± 0.0000	1.0000 ± 0.0000
Average	99.50 ± 0.80	99.50 ± 1.00	99.50 ± 1.00	0.9949 ± 0.0083	0.9966 ± 0.0037	0.9900 ± 0.0161

Table 8
Average performance of the conventional ML models on the Bonn University database.

Classifier	*Acc* (%)	*Sen* (%)	*Spe* (%)	*FS*	*AUC*	*Kap*
BA	98.00 ± 0.93	98.50 ± 1.50	97.50 ± 0.86	0.9799 ± 0.0094	0.9905 ± 0.0057	0.9600 ± 0.0272
AB	98.00 ± 0.93	97.25 ± 1.09	98.75 ± 1.09	0.9799 ± 0.0097	0.9833 ± 0.0110	0.9600 ± 0.0550
DT	97.12 ± 0.54	97.25 ± 1.64	97.00 ± 0.70	0.9712 ± 0.0060	0.9712 ± 0.0054	0.9425 ± 0.0675
KNN	97.12 ± 1.34	98.00 ± 2.45	96.25 ± 0.43	0.9711 ± 0.0138	0.9974 ± 0.0023	0.9425 ± 0.0355
ANN	90.62 ± 2.46	99.50 ± 0.86	81.75 ± 4.09	0.8975 ± 0.0209	0.9957 ± 0.0049	0.8125 ± 0.0684
SVM	97.87 ± 0.89	98.50 ± 1.50	97.25 ± 1.09	0.9787 ± 0.0090	0.9991 ± 0.0012	0.9575 ± 0.0261

a lower specificity, which means it is inferior to the other models in detecting the nonseizures. It can also be observed that STONE's performance on the Bonn University database is superior to the conventional ML models. Though ANN can achieve the same sensitivity as STONE, its specificity is the lowest among all the classifiers.

Comparison with Existing Methods. The comparison of *Acc, Sen*, and *Spe* of STONE and the existing works on the four classification problems in the Bonn University database is presented in Table 9. It is clearly shown that STONE could achieve comparable performance with the existing methods in the four classification problems. Tuncer et al. [40] proposed a local graph structure technique to extract features from the raw EEG signals, which is then used as the input to the ANN model for seizure classification. The *Sen* and *Spe* score of [40] are not available in the original paper. Considering the *Acc*, STONE can achieve a higher *Acc* for the classification problems of Z–S, O–S, and F–S. While [5,21] utilized the same complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) technique to extract features from the raw EEG signals, and they all achieved a precise classification result for the case of Z–S, distinct results were acquired for the cases of F–S and N–S because the seizure detection models are designed in different manners. Compared with these two models, STONE achieves the same result on the case of Z–S, and STONE achieves a higher *Acc*, *Sen*, and *Spe* for the case of F–S in comparison with [5]. Although [16] obtains the same or higher *Acc*, *Sen*, and *Spe* than STONE for the four classification problems, STONE carries the advantage of learning interpretable STL formulas as outputs. With this benefit, STONE can better assist the neurologists in decision making, which also promotes its acceptance in practical operations. However, the random forest (RF) model cannot provide such an asset. A more detailed analysis of the interpretability of STONE will be exposed in Section 4.4.

Table 9
Comparison of the classifiers' performance on the Bonn University database.

Authors	Features & classifiers	CPs	Acc (%)	Sen (%)	Spe (%)
[40]	Local graph structure + ANN	Z–S	99.68	–	–
		Z–F	99.88	–	–
		O–S	98.25	–	–
		F–S	98.85	–	–
[5]	CEEMDAN + LPBoost	Z–S	100.0	100.0	100.0
		F–S	97.00	97.40	98.25
		N–S	100.0	100.0	100.0
[21]	CEEMDAN + AdaBoost	Z–S	100.0	100.0	100.0
		F–S	100.0	100.0	100.0
		N–S	99.0	100.0	98.0
[16]	MSSFs + RF	Z–S	100.0	100.0	100.0
		O–S	100.00	100.0	100.0
		N–S	99.75	99.55	100.0
		F–S	100.0	100.0	100.0
This work	Spectral, LL + STONE	Z–S	100.0	100.0	100.0
		O–S	99.50	100.0	99.00
		N–S	99.50	99.00	100.0
		F–S	99.00	99.00	99.00

4.3. Ablation Study

In this subsection, we perform an ablation study on the CHB-MIT database to investigate the necessity of each module in the seizure detection framework, including the spectral feature extraction module and the line length feature extraction module. To study the effectiveness of a module, we exclude that particular module from the seizure detection framework and retrain the classifier to recalculate the *Acc, Sen, Spe, FS, AUC*, respectively.

4.3.1. Spectral component

To study the effect of the spectral feature, we exclude the fast Fourier transform module of the Feature Extraction stage in Fig. 1. This means we only keep the line length feature and classify the seizure and nonseizure data using the line length feature. The evaluation results for excluding the spectral feature are presented in Table 10. As described in Table 10, the average *Acc, Sen, Spe*, and *FS* drop to 94.33%, 95.93%, 94.32%, and 0.9505, respectively. This demonstrates that incorporating the spectral features into the seizure detection framework is necessary as the performance will degrade without it.

Table 10
Performance of STONE on the CHB-MIT database without the spectral or the LL component.

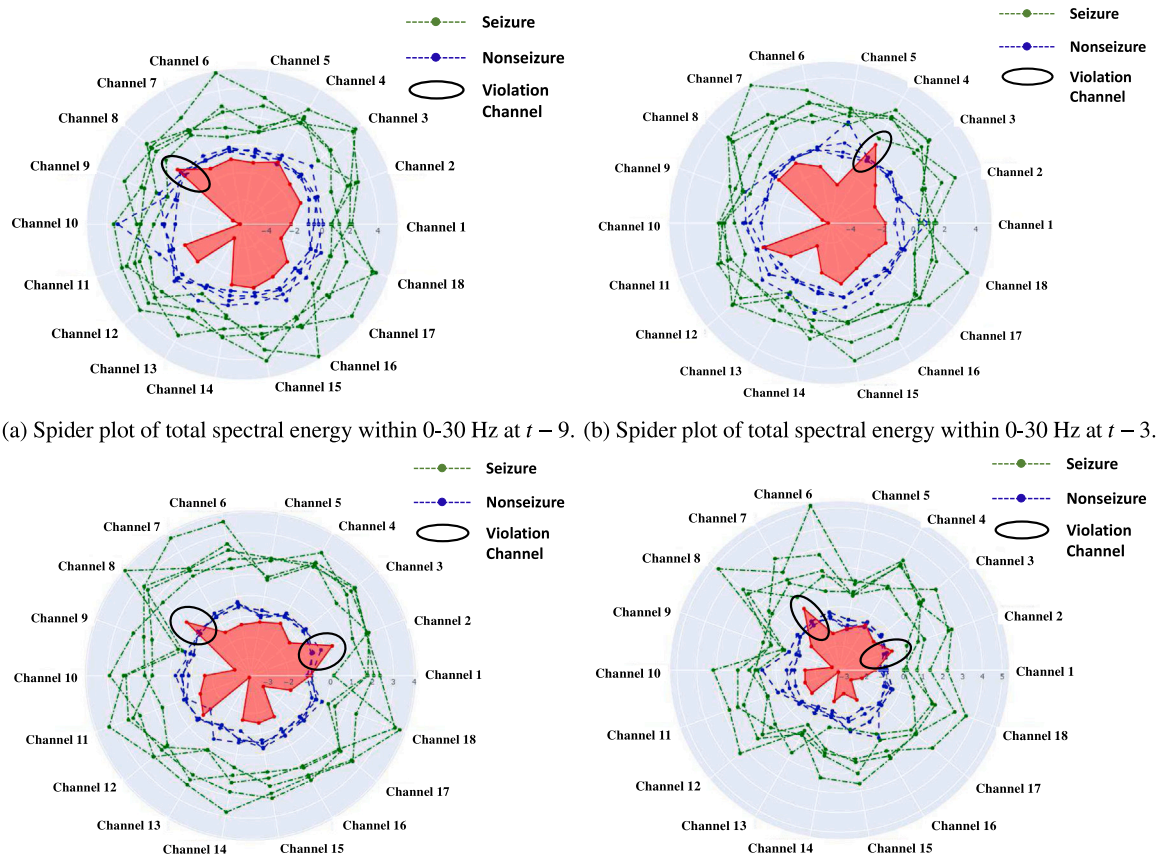
Exp setting	*Acc* (%)	*Sen* (%)	*Spe* (%)	*FS*	*AUC*
Full model	98.49	98.17	98.50	0.9834	0.9929
No spectral	94.33	95.93	94.32	0.9505	0.9827
No LL	97.58	96.64	97.61	0.9716	0.9905

4.3.2. Line Length component

To study the effect of the line length feature, we exclude the line length extraction module from the feature extraction stage in Fig. 1. The evaluation results for each patient under this experimental setting are presented in Table 10. From Table 10 we can observe that the *Acc, Sen, Spe, FS* drop to 97.58%, 96.64%, 97.61%, and 0.9716, respectively. The above results demonstrate that incorporating the line length feature into the seizure detection framework is significant in improving the detection performance.

4.4. Interpretability Analysis

Previous experimental results tell that STONE can achieve a good performance on seizure detection. In addition, a seizure detection



(a) Spider plot of total spectral energy within 0-30 Hz at $t-9$. (b) Spider plot of total spectral energy within 0-30 Hz at $t-3$.

(c) Spider plot of total spectral energy within 0-30 Hz at $t-2$. (d) Spider plot of total spectral energy within 0-30 Hz at $t-1$.

Fig. 6. Interpretable visualization of wSTL formula learned from seizure and nonseizure data of patient 23.

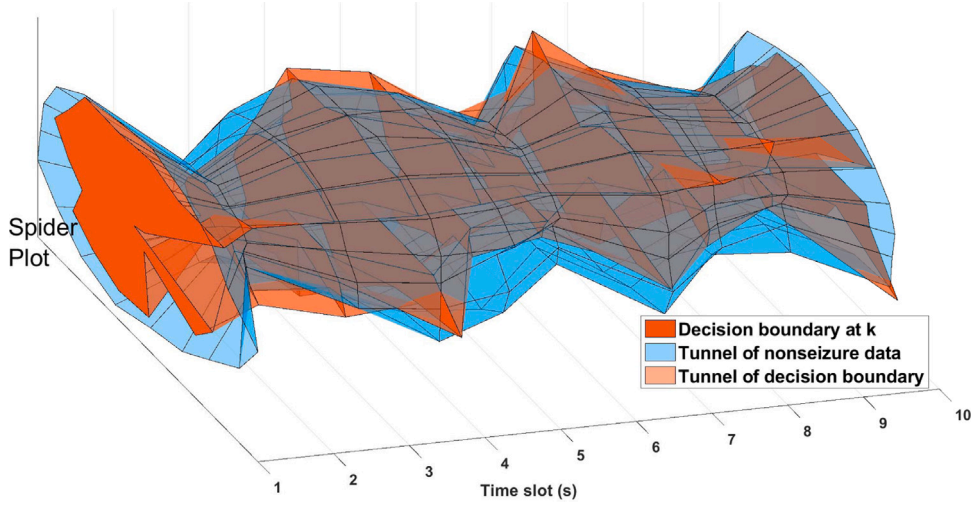


Fig. 7. The evolution of the decision boundary learned by STONE and the nonseizure data of patient 23 in the CHB-MIT database.

model is more trustworthy and applicable to clinical practice if it could provide interpretable rules to neurologists for a better understanding of the decisions made by the model [23]. STONE is such an interpretable model in that it can learn a human-readable and interpretable wSTL formula that is homogeneous to natural-language. In this section, we will analyze the interpretability of STONE by visualizing the formulas discovered by STONE that describe how the EEG features evolve over time will cause seizure. Here we select four snapshots of the energy within 0–30 Hz from patient 23 in the CHB-MIT database to demonstrate the interpretability. For better interpretation and visualization, we select the formula structure as the conjunctive patterns in (1), and the predicates describe single-channel features, i.e., $\pi_{t,j} = (\pi_{t,j,1} \wedge \pi_{t,j,2} \wedge \cdots \wedge \pi_{t,j,18})$, and $\pi_{t,j,i} \triangleq (\boldsymbol{a} X_{i,j}(t) > c)$, $\boldsymbol{a}, c \in \mathbb{R}$.

As a showcase, the evolution of the spectral features described by the formula is interpreted. At $t-9$, the formula learned on the spectral feature of the total energy within 0–30 Hz from all the 18 channels

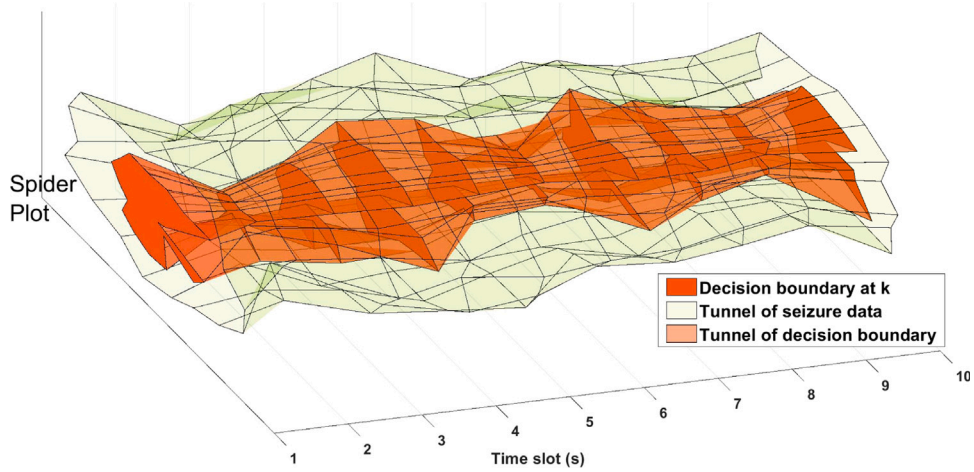


Fig. 8. The evolution of the decision boundary learned by STONE and the seizure data of patient 23 in the CHB-MIT database.

is expressed as $\phi_{t-9} = (X_{1,1} > -2.22) \wedge (X_{2,1} > -1.26) \wedge (X_{3,1} > -1.25) \wedge (X_{4,1} > 0.57) \wedge (X_{5,1} > -1.24) \wedge (X_{6,1} > -0.99) \wedge (X_{7,1} > -1.11) \wedge (X_{8,1} > 0.36) \wedge (X_{9,1} > -4.97) \wedge (X_{10,1} > -5.54) \wedge (X_{11,1} > -1.33) \wedge (X_{12,1} > -1.52) \wedge (X_{13,1} > -4.55) \wedge (X_{14,1} > -1.29) \wedge (X_{15,1} > -1.07) \wedge (X_{16,1} > -1.35) \wedge (X_{17,1} > -1.50) \wedge (X_{18,1} > -2.75)$. ϕ_{t-9} describes what properties that the total energy within 0–30 Hz from 18 channels should satisfy at $t-9$ will contribute to classifying the data as seizure. In the meantime, the corresponding decision boundary is plotted as the red line in the spider plot of Fig. 6(a), in which the green dotted lines represent the seizure data, and the blue dashed lines represent the nonseizure data, and the black circle highlights the channel for which the formula is violated by the nonseizure data. The intuitive interpretation is that if data from all the 18 channels lie outside of the boundary, then the data is classified as seizure. If at least one channel of the signal lies in the red region, then the data is classified as nonseizure. Similarly, we could interpret the wSTL formula learned on the spectral feature of the total energy within 0–30 Hz from all the 18 channels at $t-3$, $t-2$, and $t-1$ through visualization, as shown in Fig. 6(b), Fig. 6(c), and Fig. 6(d), respectively. Specifically, the wSTL formula for $t-3$ is $\phi_{t-3} = (X_{1,1} > -1.77) \wedge (X_{2,1} > 2.26) \wedge (X_{3,1} > -1.53) \wedge (X_{4,1} > 0.51) \wedge (X_{5,1} > -2.80) \wedge (X_{6,1} > -1.62) \wedge (X_{7,1} > -0.86) \wedge (X_{8,1} > -0.98) \wedge (X_{9,1} > -4.72) \wedge (X_{10,1} > -5.14) \wedge (X_{11,1} > -0.71) \wedge (X_{12,1} > 1.98) \wedge (X_{13,1} > -3.59) \wedge (X_{14,1} > -2.05) \wedge (X_{15,1} > -1.33) \wedge (X_{16,1} > -1.94) \wedge (X_{17,1} > -2.06) \wedge (X_{18,1} > -1.51)$, and the wSTL formula for $t-2$ is $\phi_{t-2} = (X_{1,1} > -0.94) \wedge (X_{2,1} > 0.34) \wedge (X_{3,1} > -1.38) \wedge (X_{4,1} > -0.94) \wedge (X_{5,1} > -1.22) \wedge (X_{6,1} > -1.34) \wedge (X_{7,1} > -1.43) \wedge (X_{8,1} > 0.16) \wedge (X_{9,1} > -3.02) \wedge (X_{10,1} > -1.64) \wedge (X_{11,1} > -1.27) \wedge (X_{12,1} > -0.88) \wedge (X_{13,1} > -3.7) \wedge (X_{14,1} > -1.66) \wedge (X_{15,1} > -1.52) \wedge (X_{16,1} > -1.55) \wedge (X_{17,1} > -3.01) \wedge (X_{18,1} > -1.77)$, and the wSTL formula for $t-1$ is $\phi_{t-1} = (X_{1,1} > -1.40) \wedge (X_{2,1} > -0.39) \wedge (X_{3,1} > -0.97) \wedge (X_{4,1} > -0.58) \wedge (X_{5,1} > -1.07) \wedge (X_{6,1} > -1.33) \wedge (X_{7,1} > 0.38) \wedge (X_{8,1} > -1.35) \wedge (X_{9,1} > -2.80) \wedge (X_{10,1} > -1.47) \wedge (X_{11,1} > -1.36) \wedge (X_{12,1} > -1.40) \wedge (X_{13,1} > -3.10) \wedge (X_{14,1} > -1.60) \wedge (X_{15,1} > -2.01) \wedge (X_{16,1} > -1.47) \wedge (X_{17,1} > -2.45) \wedge (X_{18,1} > -2.01)$. In principle, we could consider the red region as a safe region. If there is at least one channel of which the data lies in the safe region during $t-9$ to t, then the data will be nonseizure data. Otherwise, if the data of all the channels lies outside of the safe region during $t-9$ to t, then the data will be seizure data. The nonseizure and the seizure data shown in Fig. 6(a) to Fig. 6(d) precisely match the properties described above. On the other hand, we could interpret the wSTL formula through picking a particular spectral feature and visualizing the evolution of that feature and the decision boundary from $t-9$ to t. Here we select the feature of the total energy within 0–30 Hz ($X_{i,1}, i = 1, 2, \ldots, 18$), and interpret the wSTL formula discovered by STONE by visualizing the evolution of $X_{i,1}$ from $t' = 1$ to $t' = 10$ (corresponding to $t-9$ to t) as Fig. 7 to Fig. 8. In Fig. 7, we plot the evolution of the decision boundary and the nonseizure data, where the subplot at each time slot represents a spider plot, as shown in Fig. 6(a)–6(d). We could see at any $t' \in [1, 10]$, there exists at least one channel of which the data is inside of the safe region at the corresponding time slot. Likewise, by connecting the seizure data and the decision boundary at different $t' \in [1, 10]$, we could interpret the wSTL formula through visualizing the evolution of the decision boundary and the seizure data, as shown in Fig. 8. At each time slot, the seizure data always lies outside of the safe region. From the above analysis, we observe that STONE could learn human-readable formulas that are also easily interpretable through visualizing the evolution of the EEG signal characteristics as described by the formulas.

5. Conclusion

In this paper, we developed a signal temporal logic neural network (STONE) as an interpretable seizure detector to accomplish seizure detection tasks. Multi-view features, including the spectral features and the line length feature, are extracted from the raw EEG signals such that the patterns underneath the seizure data can be discovered from different perspectives. STONE is a neuro-symbolic model that combines the characteristics of neural networks and weighted signal temporal logic (wSTL), where each neuron has a corresponding symbolic expression in a wSTL formula, and the model's outcome is a human-readable formula that expresses the rules of seizures in an interpretable format. STONE is implemented on the CHB-MIT database and the Bonn University database and is demonstrated to achieve competitive performance with state-of-the-art models. Simultaneously, we utilize the weighted cross-entropy loss to tackle the data imbalance issue such that both good sensitivity and specificity can be acquired. The STL formulas provided by STONE could also be visualized for a better understanding and interpretation of the model. The experimental results demonstrate that STONE is effective in detecting seizures and providing better interaction with users.

CRediT authorship contribution statement

Ruixuan Yan: Methodology, Software, Formal analysis, Visualization, Writing – original draft. **A. Agung Julius:** Conceptualization, Methodology, Supervision, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

One or more of the authors of this paper have disclosed potential or pertinent conflicts of interest, which may include receipt of payment, either direct or indirect, institutional support, or association with an entity in the biomedical field which may be perceived to have potential conflict of interest with this work. For full disclosure statements refer to https://doi.org/10.1016/j.bspc.2022.103998. A. Agung Julius reports financial support was provided by National Science Foundation.

Acknowledgments

This work was supported by the National Science Foundation, United States (NSF) through grant CMMI-1936578 and the Rensselaer-IBM Artificial Intelligence Research Collaboration (AIRC). The authors would like to thank IBM researchers: Alexander Gray, Achille Fokoue, Tengfei Ma, Maria Chang, and Rosario Uceda-Sosa for the discussions on neuro-symbolic models in machine learning used in this paper.

References

[1] R.S. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J.H. Cross, C.E. Elger, J. Engel Jr., L. Forsgren, J.A. French, M. Glynn, et al., ILAE official report: a practical clinical definition of epilepsy, Epilepsia 55 (4) (2014) 475–482.
[2] F. F.F., Ferri's Clinical Advisor 2019 E-Book: 5 Books in 1, Elsevier Health Sciences, 2018.
[3] K.E. Misulis, E.L. Murray, Essentials of Hospital Neurology, Oxford University Press, 2017.
[4] W.O. Tatum IV., Handbook of EEG Interpretation, Springer Publishing Company, 2021.
[5] A.R. Hassan, A. Subasi, Automatic identification of epileptic seizures from EEG signals using linear programming boosting, Comput. Methods Programs Biomed. 136 (2016) 65–77.
[6] L. Logesparan, A.J. Casson, E. Rodriguez-Villegas, Optimal features for online seizure detection, Med. Biol. Eng. Comput. 50 (7) (2012) 659–669.
[7] E. Alickovic, J. Kevric, A. Subasi, Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction, Biomed. Signal Process. Control 39 (2018) 94–102.
[8] R. Esteller, J. Echauz, T. Tcheng, B. Litt, B. Pless, Line length: an efficient feature for seizure onset detection, in: 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 2, IEEE, 2001, pp. 1707–1710.
[9] A. Subasi, Application of adaptive neuro-fuzzy inference system for epileptic seizure detection using wavelet feature extraction, Comput. Biol. Med. 37 (2) (2007) 227–244.
[10] H. Albaqami, G.M. Hassan, A. Subasi, A. Datta, Automatic detection of abnormal EEG signals using wavelet feature extraction and gradient boosting decision tree, Biomed. Signal Process. Control 70 (2021) 102957.
[11] M.K. Siddiqui, R. Morales-Menendez, X. Huang, N. Hussain, A review of epileptic seizure detection using machine learning classifiers, Brain Inf. 7 (2020) 1–18.
[12] R.S. Selvakumari, M. Mahalakshmi, P. Prashalee, Patient-specific seizure detection method using hybrid classifier with optimized electrodes, J. Med. Syst. 43 (5) (2019) 1–7.
[13] D. Chen, S. Wan, J. Xiang, F.S. Bao, A high-performance seizure detection algorithm based on Discrete Wavelet Transform (DWT) and EEG, PLoS One 12 (3) (2017) e0173138.
[14] X. Hu, S. Yuan, F. Xu, Y. Leng, K. Yuan, Q. Yuan, Scalp EEG classification using deep Bi-LSTM network for seizure detection, Comput. Biol. Med. 124 (2020) 103919.
[15] H. Peng, C. Lei, S. Zheng, C. Zhao, C. Wu, J. Sun, B. Hu, Automatic epileptic seizure detection via Stein kernel-based sparse representation, Comput. Biol. Med. 132 (2021) 104338.
[16] M. Chakraborty, D. Mitra, et al., A computationally efficient automated seizure detection method based on the novel idea of multiscale spectral features, Biomed. Signal Process. Control 70 (2021) 102990.
[17] H.I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, P.-A. Muller, Deep learning for time series classification: a review, Data Min. Knowl. Discov. 33 (4) (2019) 917–963.
[18] M. Li, W. Chen, FFT-based deep feature learning method for EEG classification, Biomed. Signal Process. Control 66 (2021) 102492.
[19] A. Zarei, B.M. Asl, Automatic seizure detection using orthogonal matching pursuit, discrete wavelet transform, and entropy based features of EEG signals, Comput. Biol. Med. 131 (2021) 104250.
[20] S. Mian Qaisar, A. Subasi, Effective epileptic seizure detection based on the event-driven processing and machine learning for mobile healthcare, J. Ambient Intell. Humaniz. Comput. (2020) 1–13.
[21] A.R. Hassan, A. Subasi, Y. Zhang, Epilepsy seizure detection using complete ensemble empirical mode decomposition with adaptive noise, Knowl.-Based Syst. 191 (2020) 105333.
[22] G. Rätsch, S. Sonnenburg, C. Schäfer, Learning interpretable SVMs for biological sequence classification, BMC Bioinform. 7 (1) (2006) 1–14.
[23] A.F. Markus, J.A. Kors, P.R. Rijnbeek, The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies, J. Biomed. Inform. 113 (2021) 103655.
[24] M. Mortaga, A. Brenner, E. Kutafina, Towards interpretable machine learning in EEG analysis, in: German Medical Data Sciences 2021: Digital Medicine: Recognize–Understand–Heal, IOS Press, 2021, pp. 32–38.
[25] C. Joucla, D. Gabriel, E. Haffen, J.-P. Ortega, Three Simple Steps to Improve the Interpretability of EEG-SVM Studies, Cold Spring Harbor Laboratory, 2021, BioRxiv.
[26] C. Yoo, C. Belta, Rich Time Series Classification Using Temporal Logic, in: N.M. Amato, S.S. Srinivasa, N. Ayanian, S. Kuindersma (Eds.), Robotics: Science and Systems XIII, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, July 12-16, 2017.
[27] G. Bombara, C.-I. Vasile, F. Penedo, H. Yasuoka, C. Belta, A decision tree approach to data classification using signal temporal logic, in: Proceedings of the 19th International Conference on Hybrid Systems: Computation and Control, 2016, pp. 1–10.
[28] A. Donzé, O. Maler, Robust satisfaction of temporal logic over real-valued signals, in: International Conference on Formal Modeling and Analysis of Timed Systems, Springer, 2010, pp. 92–106.
[29] N. Mehdipour, C.-I. Vasile, C. Belta, Arithmetic-geometric mean robustness for control from signal temporal logic specifications, in: 2019 American Control Conference, ACC, IEEE, 2019, pp. 1690–1695.
[30] R. Yan, A. Julius, A Decentralized BB Algorithm for Motion Planning of Robot Swarms With Temporal Logic Specifications, IEEE Robot. Autom. Lett. 6 (4) (2021) 7389–7396.
[31] M. Charitidou, D.V. Dimarogonas, Signal Temporal Logic Task Decomposition via Convex Optimization, IEEE Control Syst. Lett. (2021).
[32] D. Gundana, H. Kress-Gazit, Event-based signal temporal logic synthesis for single and multi-robot tasks, IEEE Robot. Autom. Lett. 6 (2) (2021) 3687–3694.
[33] N. Mehdipour, C.-I. Vasile, C. Belta, Specifying user preferences using weighted signal temporal logic, IEEE Control Syst. Lett. 5 (6) (2020) 2006–2011.
[34] P. Thodoroff, J. Pineau, A. Lim, Learning robust features using deep learning for automatic seizure detection, in: Machine Learning for Healthcare Conference, PMLR, 2016, pp. 178–190.
[35] R. Meier, H. Dittrich, A. Schulze-Bonhage, A. Aertsen, Detecting epileptic seizures in long-term human EEG: a new approach to automatic online and real-time detection and classification of polymorphic seizure patterns, J. Clin. Neurophysiol. 25 (3) (2008) 119–131.
[36] A. Shoeb, H. Edwards, J. Connolly, B. Bourgeois, S.T. Treves, J. Guttag, Patient-specific seizure onset detection, Epilepsy Behav. 5 (4) (2004) 483–498.
[37] A. Shoeb, J. Guttag, Application of Machine Learning to Epileptic Seizure Detection, in: Proceedings of the 27th International Conference on International Conference on Machine Learning, in: ICML'10, Omni Press, Madison, WI, USA, 2010, pp. 975–982.
[38] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) e215–e220.
[39] R.G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, C.E. Elger, Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state, Phys. Rev. E 64 (6) (2001) 061907.
[40] T. Tuncer, S. Dogan, F. Ertam, A. Subasi, A novel ensemble local graph structure based feature extraction network for EEG signal analysis, Biomed. Signal Process. Control 61 (2020) 102006.
[41] A. Subasi, J. Kevric, M. Abdullah Canbaz, Epileptic seizure detection using hybrid machine learning methods, Neural Comput. Appl. 31 (1) (2019) 317–325.
[42] L. Logesparan, E. Rodriguez-Villegas, A.J. Casson, The impact of signal normalization on seizure detection using line length features, Med. Biol. Eng. Comput. 53 (10) (2015) 929–942.
[43] J. Birjandtalab, M.B. Pouyan, D. Cogan, M. Nourani, J. Harvey, Automated seizure detection using limited-channel EEG and non-linear dimension reduction, Comput. Biol. Med. 82 (2017) 49–58.
[44] N. Mehdipour, C.-I. Vasile, C. Belta, Specifying User Preferences Using Weighted Signal Temporal Logic, IEEE Control Syst. Lett. 5 (6) (2021) 2006–2011.
[45] P.J. Hurley, A Concise Introduction to Logic, Cengage Learning, 2014.

[46] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artificial Intelligence Res. 16 (2002) 321–357.

[47] I. Loshchilov, F. Hutter, Decoupled Weight Decay Regularization, in: 7th International Conference on Learning Representations, ICLR 2019, New Orleans, la, USA, May 6-9, 2019, 2019, OpenReview.net.

[48] I. Ullah, M. Hussain, H. Aboalsamh, et al., An automated system for epilepsy detection using EEG brain signals based on deep learning approach, Expert Syst. Appl. 107 (2018) 61–71.

[49] R. Delgado, X.-A. Tibau, Why Cohenś Kappa should be avoided as performance measure in classification, PLoS One 14 (9) (2019) e0222916.