IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Enhanced Multiple Sound Event Detection and Classification Using Physical Signal Properties in Recurrent Spiking Neural Networks

ZAHRA ROOZBEHI [1], AJIT NARAYANAN[1], MAHSA MOHAGHEGH[1,2], and SAMANEH-ALSADAT SAEEDINIA[3]

[1]School of Electrical and Computer Engineering, Auckland University of Technology, Auckland, New Zealand
[2]Faculty of Design and Creative Technologies, Auckland University of Technology, Auckland, New Zealand
[3]Department of Electrical and Computer Engineering, Iran University of Science and Technology, Tehran, Iran

Corresponding author: Zahra Roozbehi (e-mail: zahra.roozbehi@aut.ac.nz).

ABSTRACT Sound event detection and classification present significant challenges, particularly in noisy environments with multiple overlapping sources. This paper introduces an innovative architecture for multiple sound event detection and classification utilizing recurrent spiking neural networks (SNNs). Our method uniquely leverages temporal data to detect and classify multiple sound sources simultaneously, integrating the physical concept of signal power matching with neuronal output power and employing a binaural strategy to enhance detection accuracy in real-world scenarios. The architecture processes spatiotemporal data to dynamically update synaptic weights, enabling precise identification of sound event categories and their occurrences. Our simulations reveal substantial performance improvements, achieving the highest precision of 73% in classification tasks, including multilayer perceptrons (MLP), convolutional recurrent neural networks (CRNN), and recurrent neural networks (RNN). Statistical analysis indicates that these improvements are significant (p-value < 0.05). These findings suggest practical applications in various fields such as surveillance, autonomous vehicles, and smart home systems, where robust sound event detection is critical.

INDEX TERMS Sound Classification, Spiking Neural Network,Tempotron, Audio signal power, detection, multiple sound sources.

I. INTRODUCTION

SOUNDevent detection and classification have become integral to our daily lives, powering technologies like smart home assistants and industrial applications. Imagine walking into your home and saying, "Hey, assistant, play some jazz and dim the lights," or a factory where machinery sounds are continuously monitored for fault detection. These scenarios require sophisticated sound event recognition across multiple acoustic dimensions.

While deep learning methods, such as Convolutional Recurrent Neural Networks (CRNNs) and other types of Recurrent Neural Networks (RNNs), have demonstrated promising results, they face specific limitations. These methods often require extensive, well-labeled datasets to achieve high performance, which can be a significant challenge in practical scenarios. Additionally, they struggle with detecting multiple overlapping sound events in temporal patterns and tend to neglect the physical aspects of input signals. This omission can limit their reliability and analytical interpretability, as they often operate as black-box models that lack transparency [32]–[34].

In contrast, Spiking Neural Networks (SNNs), particularly Recurrent Spiking Neural Networks (RSNNs), offer a compelling alternative for sound event detection and classification. SNNs process sounds with high temporal precision by leveraging their event-triggering capabilities, which allow them to efficiently handle asynchronous data. This feature makes them computationally efficient compared to traditional deep learning architectures [2]. This is due to the event-triggering capabilities inherent in SNNs, which allow for more efficient processing compared to the traditional deep-learning approaches [2].

An SNN is a type of artificial neural network that is based on communicating neurons that emit spikes, which are electrical pulses generated when a spiking neuron fires [3], [4]. Unlike conventional neural networks that rely on error backprop-

agation and gradients for supervised learning, SNNs face distinct challenges due to the non-differentiable nature of spike events, operating instead on spike timing and time differences, necessitating alternative training methods [5]. Recent advancements in SNN training methodologies, such as Spike-propagation [6] and Superspike [7] techniques, have broadened the training capabilities of SNNs beyond traditional approaches. Saeedinia et al. [8] have also introduced an error backpropagation learning law for SNNs that uses neuron voltage outputs, which offers a solution to the challenge of training SNNs. Contrary to supervised learning, SNNs can also be trained with unsupervised learning techniques such as Hebbian learning or spike-timing-dependent plasticity [9]. In Hebbian learning, a postsynaptic neuron is strengthened when its presynaptic neuron spikes. In contrast, spike-timing-dependent plasticity (STDP) takes both the time delay between the spikes of the pre-and postsynaptic neuron into account and strengthens the connection if the pre-synaptic neuron spikes before the post-synaptic neuron. STDP operates on the principle that neurons fire more quickly and accurately when an input stimulus 'reinforces' the firing, and conversely, the neuron fires less when it is 'inhibited'. This enables the neuron to learn and adapt the firing patterns according to the reinforcement and inhibition it receives from the input stimuli [9].

Recent advancements in semi-supervised learning for Spiking Neural Networks (SNNs) combine supervised learning with Spike-Timing-Dependent Plasticity (STDP) to improve performance, especially with limited labeled data. This method involves supervised training followed by unsupervised refinement using STDP, leveraging both labeled and unlabeled data to enhance classification accuracy without extensive labeling. STDP not only aids self-organization but also boosts learning efficiency in SNNs [35], [36].

In the domain of classification utilizing Spiking SNNs, the Tempotron, which was introduced by Gütig and Sompolinsky in 2006 [10], emerges as a crucial supervised learning algorithm specifically designed for this task. It enhances classification accuracy by updating weights iteratively, capitalizing on the concept of heightened firing precision and velocity in neurons under positive stimulation, while reducing firing in response to inhibitory stimuli. This mechanism allows neurons to modify their firing sequences based on the reinforcement and inhibition from stimuli. Through the utilization of Spike-Timing-Dependent Plasticity (STDP), the Tempotron fine-tunes weights incrementally during training loops, enhancing classification efficacy by exposing various stimuli until achieving peak performance [11] [12]. The Tempotron, characterized by a biologically plausible learning principle, excels in categorizing spatiotemporal patterns, rendering it valuable in diverse applications such as sound event classification [5]. By leveraging Spike-Timing-Dependent Plasticity (STDP), the Tempotron dynamically modifies synaptic weights depending on the timing of pre- and postsynaptic spikes, reinforcing connections when presynaptic spikes precede postsynaptic ones. This feature empowers the Tempotron to effectively grasp and identify temporal patterns in input data [5]. Within this framework, the Tempotron learning principle adheres to a supervised learning methodology based on gradients for spiking neuron models. It adapts synaptic weights by assessing their influence on the postsynaptic membrane potential, amplifying weights for firing responses and diminishing for silence until achieving precise classification of all patterns within the training dataset [11]. These weights undergo gradual modifications through training loops, exposing the neuron to diverse stimuli until reaching an optimal level of classification efficiency [10] [11].

In Hebbian learning, the intensity of a postsynaptic neuron is elevated by the spiking activity of its presynaptic counterpart. In contrast, STDP takes both the time delay between the spikes of the pre- and postsynaptic neuron into account and strengthens the connection if the presynaptic neuron spikes before the postsynaptic neuron [11]. The Tempotron uses this principle to adjust the weights and improve the classification accuracy [10] [11].

The Tempotron learning rule is a gradient-based supervised learning approach for spiking neuron models, adjusting synaptic weights according to their impact on the postsynaptic membrane potential. Synaptic weights are increased for firing responses and decreased for silence until accurate classification of all patterns in the training set is achieved [8]. This helps to tune the temporal coding of the sounds to better match the required sound classification. These weights are adjusted iteratively through a series of training cycles, in which the neuron is presented with a series of varying sounds and the weights are adjusted until the efficiency of the sound classifications reaches its highest point [10] [11].

The Tempotron and STDP algorithms have been used in a variety of sound classification tasks, with highly impressive results [13] [14] [15] [16]. They have been employed in speech recognition tasks, achieving accuracies of up to 97.5% [16]. Furthermore, they have been used to classify complex soundscapes, such as those from birdcalls and other wildlife, which would otherwise be too difficult for a human to classify [14]. As such, the Tempotron and STDP algorithms are invaluable tools for sound classification. By leveraging the precise temporal coding of the neurons, the Tempotron is able to analyze the sounds and rapidly reach the desired accuracy in a variety of applications. Additionally, the STDP algorithm is used to optimize the network and ensure the highest accuracy in sound classification. Populations of spike neurons with competing STDP-based learning rules have been proven to self-organize and understand whole repetitive patterns with the aid of firing continuously [18]. They have also been utilized in more complex feedforward networks for image classification tasks with encouraging results [19], [20]. Moreover, [21] utilized Tempotron-STDP rules in a multilayer Spiking Neural Network structure to analyze the frequency spectrum of the received signals in order to detect and classify a single sound event. In conclusion, the Tempotron and STDP algorithms are powerful tools for sound classification tasks,

providing accurate and reliable results.

The Tempotron-STDP approach in sound classification, especially in the context of multiple sound classification by SNNs, encounters challenges related to robustness against noise, efficiency in learning, and effectiveness in representing features. Various techniques such as training with noise-corrupted samples, making early decisions, and employing competitive STDP-based feature learning have been suggested by researchers to improve the performance of SNNs [22] [23] [24].

State-of-the-art Spiking Neural Networks (SNNs) have demonstrated enhanced performance in noisy environments [37], especially in applications such as Voice Activity Detection (VAD) and content-based image retrieval (CBIR). For example, the sVAD model enhances noise robustness with an attention mechanism, while a bioinspired SNN-based edge detection method improves CBIR precision by better mimicking human edge perception [38].

Despite the progress in sound processing, there are persistent challenges in addressing noisy environments with multiple sound origins. Ongoing research is concentrated on the real-time configuration of SNNs to enhance the comprehension of temporal data implications. The imperative task lies in devising SNN architectures capable of managing real-time data processing with precise event identification. An adaptive and dynamic structure for SNNs has been previously recommended for the detection and localization of individual sound sources within real-world environments [24]. However, detecting and classifying multiple sound sources is still a challenge.To address these challenges, this study explores a novel SNN architecture designed to enhance robustness and accuracy in multi-source sound classification, leveraging principles of temporal coding and dynamic adaptation to improve real-time processing capabilities.

The central focus of this research lies in exploring the implications of exclusively considering time-domain real-time data in practical environments and applications for configuring Spiking Neural Network (SNN) structures. Additionally, the study delves into the design aspects of SNN structures, aiming to address the challenges associated with real-time data processing while ensuring satisfactory results in sound event detection and classification. This inquiry is crucial in the context of sound event detection, as it grapples with the complexities of real-time data processing and the need to adeptly design SNN structures to handle such data without compromising performance.

In the realm of multiple sound event detection and classification by SNNs, the utilization of multiple layers proves pivotal. These layers empower the network to acquire intricate representations of sound [25]. The versatility offered by SNNs in detecting and classifying multiple sound events presents a promising approach, potentially surpassing the capabilities of traditional neural networks. Through the incorporation of sophisticated architectures, an SNN can be trained to accurately detect and classify sound events in real-time while minimizing power consumption [1]. Moreover, the application of a matched learning law aligned with the physics of the events can significantly enhance accuracy and learning rates.

The emerging field of Physic-Informed (PI) Neural Network Classification of Sound Events focuses on developing systems capable of classifying sound events by incorporating physical principles. Such systems leverage attributes like duration, pitch, loudness, and contextual information such as sound source location to achieve accurate classifications [27], [28]. PINNs demonstrate effectiveness across various applications, including animal classification based on vocalizations. Beyond enhanced accuracy, these networks prove more efficient, processing large datasets rapidly for real-time sound event classification, particularly beneficial in applications like security monitoring [28]. Despite promising results in Sound Event Classification (SEC), SNNs employing frequency-based processing methods often face computational constraints, limiting their online applications. In response, this paper proposes a novel SNN framework employing a time-domain power-based learning methodology within a recurrent structure for the detection and classification of multiple sound events. This recurrent framework integrates the Tempotron STDP law with a PI learning strategy, aligning synaptic weights with the measured power of sound signals. By fusing power intensity with temporal features, this approach tackles simultaneous sound event detection and classification issues, utilizing time-domain raw data without the need for frequency analysis features. The proposed structure demonstrates the capability for online deep detection and classification, discerning the occurrence of multiple sound events simultaneously. This paper is organized with the method described in Section II, followed by results in Section III, and concluding remarks in the final section.

II. MATERIALS AND METHODS

In this section, we propose to combine the STDP learning strategy based on the Homeostatic Plasticity law with an innovative PI Power-Matched learning strategy (PIPM) in a recurrent spiking neural network (RSNN) structure. The adaptive threshold [12] helps to maintain a balanced firing rate among the neurons, thus allowing for a greater diversity of receptive fields to form in the feature extraction layer within a different recurrent SNN structure and cost function. The proposed architecture is composed of Leaky-Integrate and-Fire (LIF) neurons; whose membrane potential V evolves according to the following equation:

$$\frac{dV_i(t)}{dt} = \frac{1}{\tau_m}\left(-V_i(t) + I_i(t)\right) \tag{1}$$

Where τ_m denotes the membrane decaying time constant. $I_i(t)$ is the synaptic current. The overall structure of the proposed architecture is demonstrated in Fig.1.

As illustrated in Fig.1, the proposed strategy consists of three main layers: an input layer responsible for the online encoding of measured signals, a processor layer which is an RSNN structure of LIF neurons with two types of neurons

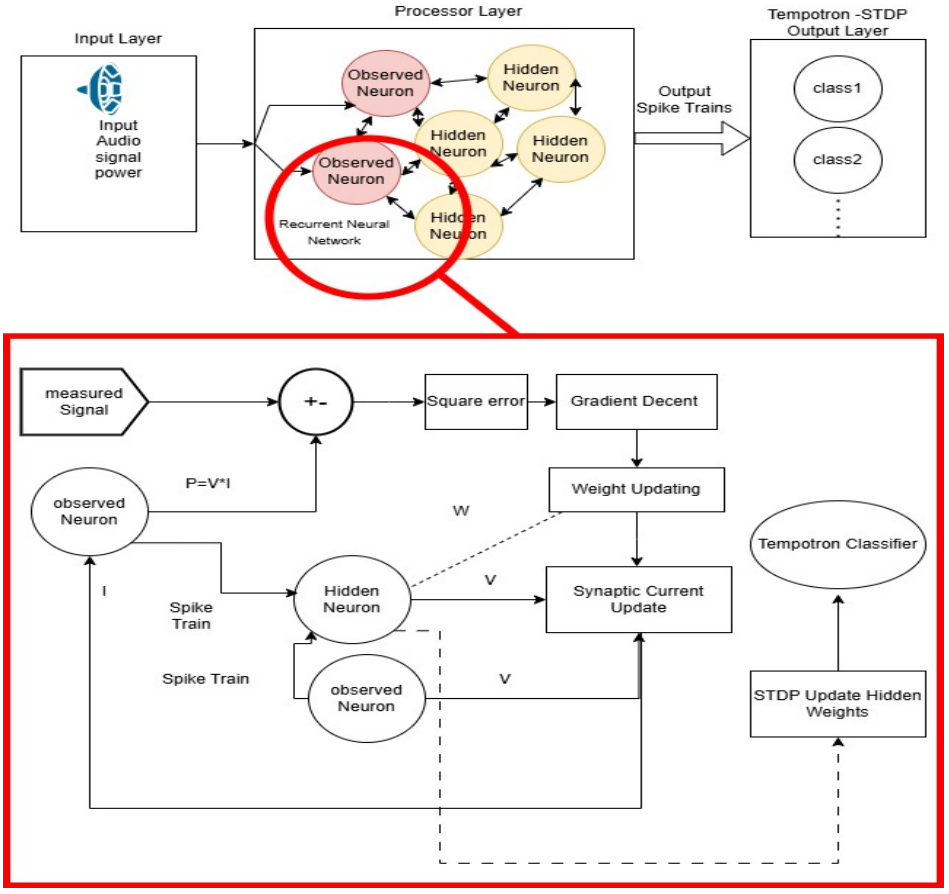


FIGURE 1. Overall Structure of Adaptive PI RSNN classifier and the detail of the calculation process

(Hidden and Observed), and an output layer composed of Tempotron-STDP classifier. The observed neurons receive data as their desired potential and transmit data through firing to other lateral neurons, spiking based on the proposed Power-matched strategy. The hidden neurons, however, do not receive external input; they are connected to the observed neurons and other lateral neurons in the output layer, spiking based on the Tempotron-STDP law [13]. Finally, the output layer is linked to the distinguished high-power neurons. The pipeline of the proposed strategy is also depicted in Fig.2.

According to the given pipelines, the first stage involves sending raw data to the online encoder module, i.e. the initialization part. The second step entails estimating the electrical power outputs of observed neurons concerning the received signals' power. The third and fourth stages involve updating the parameters of the network, while the fifth stage eliminates low-power neuron connections to reduce the computational costs of the network and optimize its size. Finally, the Tempotron classifier is employed to detect high-power neuron firing patterns. This strategy facilitates the detection of multiple sound events by their power patterns and classifies them according to their firing patterns. By leveraging this strategy, it is possible to quickly and accurately identify and categorize sound events. The next section describes the STDP-Tempotron Law.

A. TEMPOTRON-STDP LAW

Tempotron [5] is a temporal learning rule that works based on the membrane potential of the neurons to train binary classifiers. In this learning law, a trained neuron fires whenever it observes patterns from its class. Otherwise, its membrane potential remains under . In this classification module, each class is assigned to a specific output neuron. In other words, if the input pattern belongs to the *i*th class, the *i*th output neuron is a mere neuron that fires, and the rest of the neurons are silent. The STDP-Tempotron for the ith output neuron is proposed to be modified as follows:

$$\Delta W_{ij} = \begin{cases} \lambda_0 \exp(-\frac{t_i - t_j}{\tau_+}) \sum\limits_{t_i < t_{\max}} K(t_{\max} - t_i) & if \quad t_i > t_j \\ -\lambda_0 \exp(\frac{t_i - t_j}{\tau_-}) \sum\limits_{t_i < t_{\max}} K(t_{\max} - t_i) & t_i < t_j \\ 0 & else \end{cases} \tag{2}$$

Where λ_0 is a constant learning rate. t_{max} is the time sample when the neuron potential peak is maximized and t_i is the last

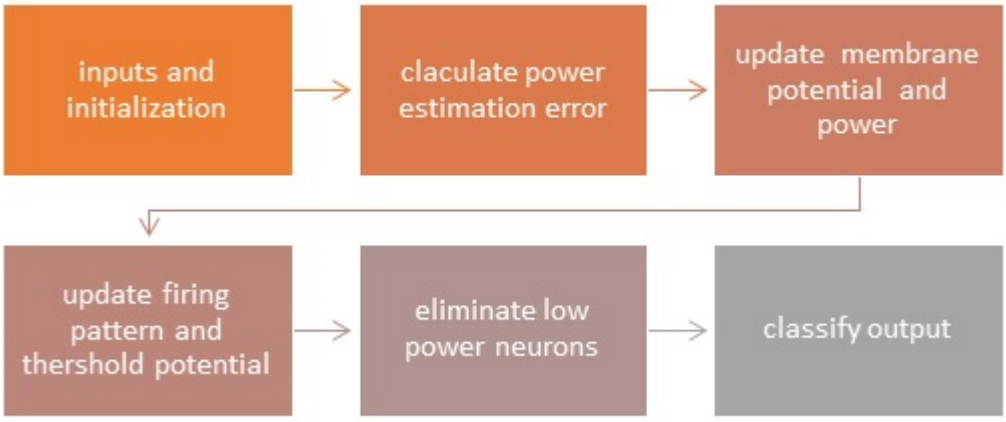


FIGURE 2. Pipeline of PIPM- RSNN sound event classifier

spike time of neuron i. τ_+ and τ_- are the time constants. $K(.)$ is the kernel function which is calculated by:

$$K(t-t_i) = K_0\left(\left(\exp\left(-\frac{t-t_i}{\tau_m}\right)\right) - \left(\exp\left(-\frac{t-t_i}{\tau_s}\right)\right)\right) \tag{3}$$

where K_0 is a normalization factor, guarantees that the maximum amplitude of the kernel, i.e. $K(t-t_i)$, is 1. τ_m and τ_s are respectively the membrane and synaptic decay time constants to determine the shape of the kernel function. regards only spikes before time t. The membrane potential of the output neurons can be updated according to (1) and based on the STDP-Tempotron learning law. In this regard, the synaptic current of (1) is defined as follows:

$$I_i(t) = \sum_j W_{ij} K(t-t_j) \tag{4}$$

In this strategy, the optimal firing rate of f_d for neurons in the feature extraction layer is predefined. The firing rate of the f_t is calculated after presenting a predetermined number of audio frames and the rate of the neuron membrane potential threshold is updated as follows:

$$\frac{dv_{th}}{dt} = \lambda_h(f_d - f_t) \tag{5}$$

where λ_h is a constant learning rate. This can be set to 0.5, in order to match quickly in thousands of frames. This feature is employed to detect events, by reducing noise effects by updating the firing threshold. This law is utilized for hidden neurons. The next section presents the proposed supervised PIPM learning law for observed neurons.

B. ADAPTIVE PI RSNN CLASSIFIER

To embed the Tempotron-STDP approach to the newly suggested structure, we define the error of the power estimation to match the physical concepts of the electrical power definition for a spike neuron circuit model with the measured power of input audio signal:

$$E = (X_{in} - V_{ob}(k) I_{oh}(k)) \tag{6}$$

In which X_{in} is the input measured signal, $V_{ob}(k)$ is the observed neuron membrane potential, and is the input synaptic current to the observed neurons, calculated by:

$$I_{oh}(k) = W_{oh}(k) V(k) \tag{7}$$

where $W_{oh}(k)$ is the weight of the observed neuron, and $V(k)$ is the vector of membrane potential of all neurons. The updating law for the weights of the observed neurons is determined by Equation (8), with a detailed proof provided in Appendix IV.

$$\begin{aligned}\Delta W_{oh}(k) = \eta(X_{in}(k) - V_{ob}(k) \cdot I_{oh}(k)) \\ \cdot (V_{ob}(k) \times V(k)^T)\end{aligned} \tag{8}$$

In the PIPM-RSNN structure, we have considered LIF neurons with I inputs, and its postsynaptic membrane potential represented by $V_j(t)$ stays at the resting potential $V_{\text{rest}} = 0$ when no spikes are received. Keeping abreast of a spike produced at pre-synaptic neuron i, a postsynaptic potential (PSP) is induced in the LIF neuron. By integrating the PSPs resulting from a number of the spike inputs, the LIF neuron fires a spike when its membrane potential $V(t)$ reaches the firing threshold v_{th}. Therefore,(8) can be the result, based on the dynamics of the neuron postsynaptic membrane $V(t)$.

The output layer of the proposed structure is the classification layer as well as an indicator of power hidden neuron heatmap to determine the quantities of the detected events simultaneously.

C. EVALUATION METHODOLOGY

To assess the efficacy of the newly designed SNN classifier, power patterns and firing rates of various sound events are detected and classified using two different benchmark datasets,

first the TAU-NIGENS Spatial Sound Events 2020 database [27] and the second ESC dataset [39], [40].

The TAU-NIGENS dataset can be accessed on Zenodo, a platform dedicated to research data storage. This database consists of multiple spatial sound scene recordings containing different classes of sound events embedded in various acoustic spaces and various directions and distances of sound sources from the recording locations. All sound events are filtered through real spatial impulse responses (RIRs) recorded in various rooms of different shapes, sizes, and sound absorption characteristics. Additionally, each scene recording is provided in two spatial recording formats: Microphone Array (Microphone) and First Order Ambisonics (FOA). Static or moving sound sources in a room are spaces that use time-varying RIRs. Each audio event in an audio scene is associated with its entry path to the save point, as well as a start time and time offset. The individual tapes used to build the audio scene come from the NIGENS public audio event database [27].

The ESC-50 dataset has emerged as a widely -used benchmark for environmental sound classification research [39]. This comprehensive collection comprises 2000 labeled environmental audio recordings, each spanning 5 seconds, categorized into 50 semantic classes with 40 examples per class [39], [40]. The dataset encompasses a diverse range of sounds, including animal vocalizations, natural phenomena, human non-speech sounds, domestic noises, and urban soundscapes. All audio samples are sourced from public field recordings available through the Freesound.org project. The recordings are standardized with a 44.1 kHz sample rate and provided as single-channel audio.

To comprehensively evaluate the performance of our PI-RSNN model, we compared it with three well-established baseline architectures: the Multilayer Perceptron (MLP), the Recurrent Neural Network (RNN), and the Convolutional Recurrent Neural Network (CRNN). The MLP served as a basic feedforward network, allowing us to assess the advantages of temporal processing. The RNN provided a standard recurrent architecture for sequence modeling, which enabled us to emphasize the benefits of spiking neuron dynamics and adaptive thresholds. Finally, the CRNN, recognized for its state-of-the-art performance in audio tasks, provided a strong benchmark against existing deep learning techniques for sound event detection.

Simulations are conducted in Python 3.10, and the performance of the proposed Adaptive PI RSNN method is compared to MLP [29], CRNN [30], and RNN [28]. The parameters and specifications of these networks are provided in Table 1.

As illustrated in Fig.3, the data inputs to the network models are raw waveforms. The key distinction between our proposed method and other methods, namely MLP, CRNN, and RNN, lies in our approach of utilizing the first 350 samples online for training and the remaining samples for testing. This methodology enables real-time processing in various

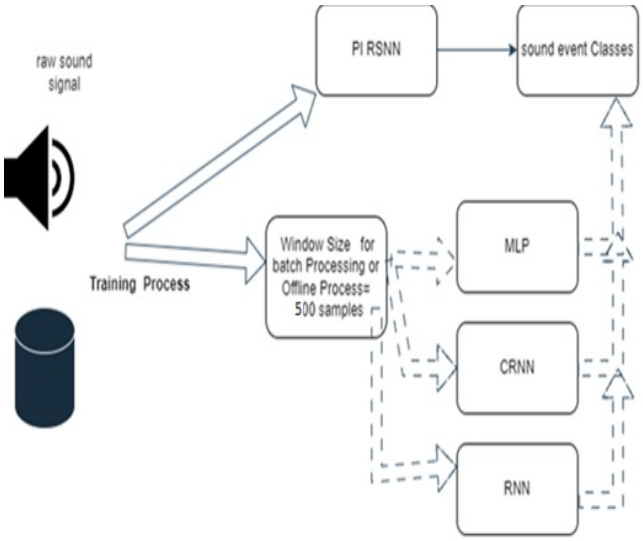


FIGURE 3. Workflow of data processing for classification and comparison models

applications. In contrast, MLP, CRNN, and RNN employ an offline process, where 90% of the data is allocated for training and the remaining 10% for testing purposes using cross-validation. This difference in training methodology highlights the adaptability and potential real-time capabilities of our proposed method compared to the batch processing approach used in traditional methods.

The scalability of the proposed model, given the parameters in 1, can be enhanced by incorporating spatiotemporal data if available. When the locations of sound sources are established through prior knowledge or sound localization methods, we can effectively adjust the number of hidden neurons in the model based on the complexity of the environment containing all sound sources. Specifically, increasing the number of hidden neurons can improve classification accuracy; however, beyond 300 neurons, the performance gains are minimal for online classification and detection tasks. In batch processing scenarios, it is advisable to match the number of hidden neurons to the input data size. For instance, if 500 sample data points are utilized for each input, we recommend employing 500 observed neurons alongside 500 hidden neurons. Our studies indicate that fewer hidden neurons can still yield promising results, suggesting that the optimal number can be fine-tuned based on specific sound detection and classification tasks. Furthermore, regarding the determination of f_d as the optimal firing rate, considering the frequency range of sounds can aid in fine-tuning this parameter. It is essential to select a moderate value for f_d; in our case, we have opted for 0.5, avoiding extremes such as 1 or 0 to maintain effective model performance.

III. RESULTS AND DISCUSSION

This section investigates the performance of the proposed structure and its classification results. Our evaluation is

TABLE 1. Networks' Specifications

Model	Parameters and Layers	Value
Adaptive PI-RSNN	τ_n:Membrane decaying time constant of the LIF Neuron	0.5
	τ_m: Membrane decaying time constant of Kernel function	2.5
	τ_s: Synaptic decaying time constant of Kernel function	0.5
	$\tau_{\pm}$: Time constant of STDP-Tempotron updating weight law	1
	λ_h: Cconstant learning rate of neuron membrane potential threshold	20
	K_0 : Normalization Factor	100
	λ_0: Constant learning rate of Tempotron output layer	10
	Hidden neurons numbers	50
	f_d:Optimal firing rate	30
	Initial values for membrane potential	- 65
	Initial values for the synaptic current	0
	Initial wights	Random:N(0,1)
MLP [26]	Hidden Layer	2
	Hidden neurons numbers	24
	Output layer neurons	3
	Input layer Neurons (Input samples)	500
	Learning rate	0.95
CRNN2 (0.25M) [27]	Input samples	500×1
	Convolutional layer	C(64, 80/4)
	Max Pooling	4×1
	Convolutional layer	C(64,3)
	Max Pooling	4×1
	RNN(128) and Output layer	Fully Connected (activation: softmax)
RNN [28]	Number of Layers	3
	Size of Hidden state vector	256
	The learning rate for Adam optimizer	0.0001
	Dropout rate	0.1
	Regularization parameter	0.001

categorized into two steps. The first step focuses on event detections with consideration of the neurons' arrangements (52 neurons) for online training located in the area matched with the location of the closed environment where our sound sources are located. To effectively perform this task, prior knowledge of the sound source locations is not essential, however, it can improve detection accuracy in cases where we have multiple sound events in the environment. For this purpose, we leverage our previous work on ART-rSNN, as presented in [24], to arrange the neurons.
The second evaluation focuses solely on the classification approach, conducted on a 500-sample batch process training with 500 observed neurons with random arrangements without any tendency to map the location of neurons in the exact location of arrangements of neurons or source localization approach, and 500 hidden neurons.

Figures 4,5, and 6 indicate event sound detection by knowledge about the location of the sound sources and locating input neurons in those locations. The second structure considers inputs without considering the spatial data of sound sources. Figures 4, 5, and 6 depict the power patterns of the different sound event samples generated by the neurons of the Adaptive PI-RSNN.

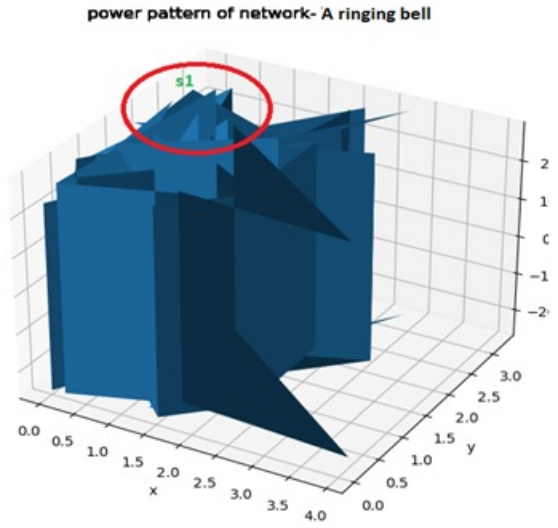


FIGURE 4. Network Power Pattern for Ringing Bell Sound Event: Areas exhibiting peak tendencies indicate the detection of distinct sound events.

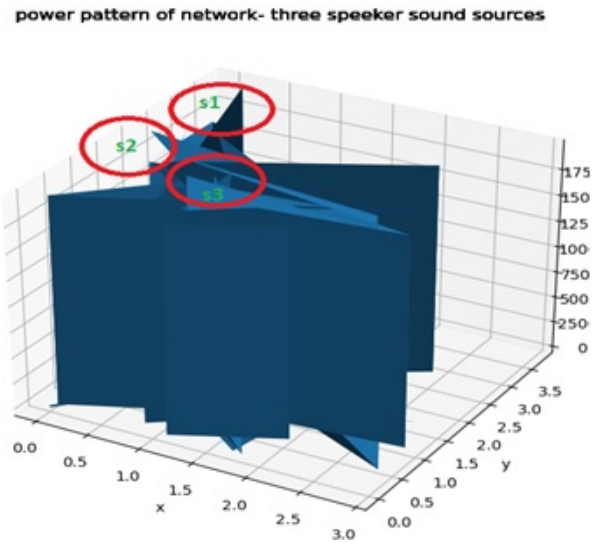


FIGURE 5. power pattern example of three sound sources of human speaking sound in PI-RSNN: Areas exhibiting peak tendencies indicate the detection of distinct sound events

Figures 4,5, and 6 indicate that the new strategy is able to detect the number of sound sources by illustrating power peaks in the hidden network structure. The subsequent figure depicts the firing patterns of the proposed network in 20 time samples. This data demonstrates the efficacy of the new strategy in accurately detecting sound sources. As illustrated in Fig.4,Fig.5, and Fig.6, we can observe that this structure determines the number of sound sources available. Figure 7

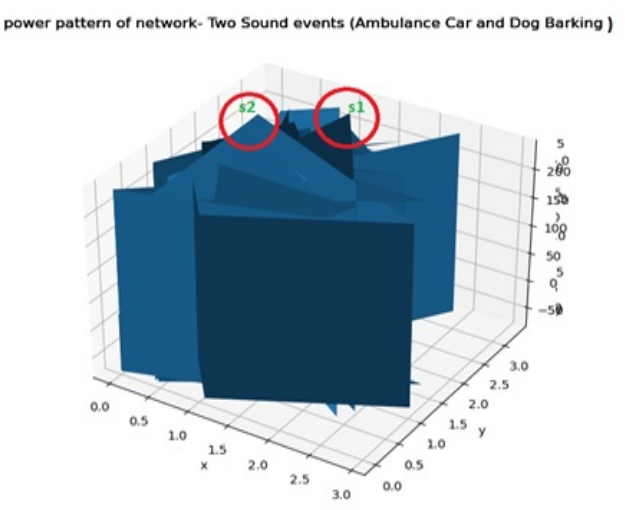


FIGURE 6. Power pattern of two mixed sound events, including Dog barking and Ambulance Car sound, the mixed sound signals are overlapped: Areas exhibiting peak tendencies indicate the detection of distinct sound events

depicts the firing patterns of various samples.

Figure 8 indicates the firing pattern of the online training process with error band criteria for training samples, and the rest are considered as the test dataset. Figure 7 illustrates how the proposed strategy is able to detect ringing bells, dog barks, and ambulance car sounds.

Inspection of the above figure reveals that the proposed system can detect events with almost negligible delay (0.2 seconds). To further compare this structure with the well-known Machine Learning (ML) algorithms, the classification performance of the proposed strategy is compared with Multilayer Perceptron (MLP), Convolutional Recurrent Neural Network (CRNN), and Recurrent Neural Network (RNN) classification algorithms in a second scenario, including a batch process comparison with 500 samples, ensuring a more fair comparison between online and offline processes.
Figures9, 10, and 11 illustrate the firing patterns generated by the proposed methods in the second scenario (batch process) with a 10-cross-fold validation of three classes of sounds in the network with 500 observed neurons and 500 hidden neurons, highlighting the network's ability to effectively process and classify complex data. Table2 compares the performance of MLP, CRNN, RNN, and PI SNN classification.

Table 2 and Table 3 Comparison of averaged classification accuracy of the proposed structure against other neural network systems for 10 random partitions of TAU-NIGENs and ESC datasets, containing 500 samples, with 90% of the data used for training and 10% for testing. Statistically significant differences between PI-RSNN metrics and other networks using ranked T tests for the 10 runs are indicated by asterisks (*=p0.05, **=p0.01).

TABLE 2. Model Performance Comparison in TAU-NIGENs Dataset

Model	Accuracy	Precision	Recall	F1 Score
MLP [29]	0.571	0.321^{*}	0.528	0.439
CRNN [30]	0.651^{**}	0.728	0.636^{**}	0.611^{**}
RNN [31]	0.335^{**}	0.108^{**}	0.333^{**}	0.163^{**}
PI RSNN	0.622	0.738	0.604	0.557

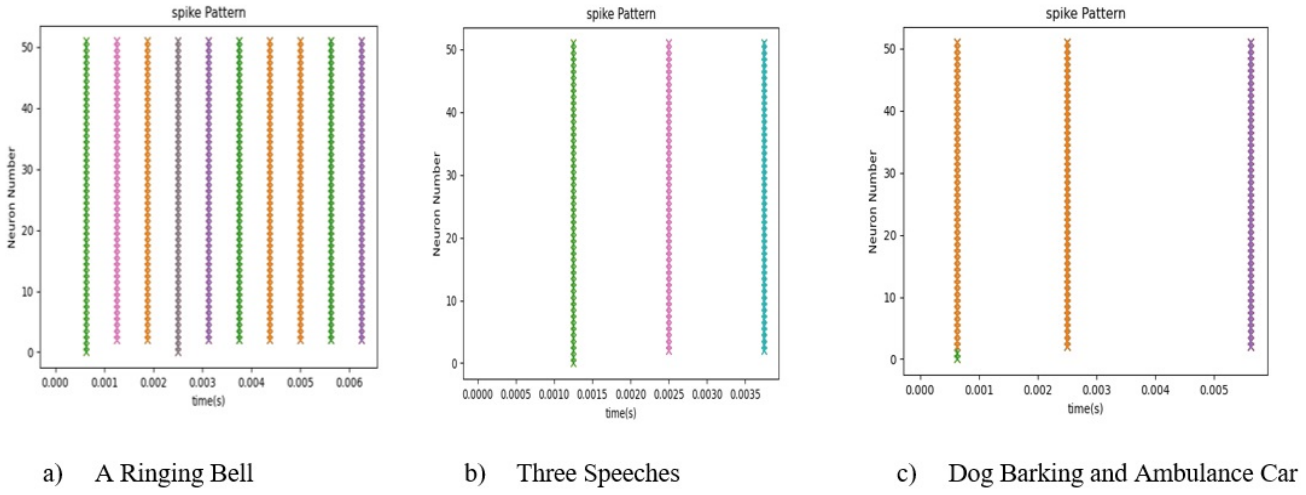
*= p$\leq$ 0.05, and **= p$\leq$ 0.01 denote statistical significance levels.

TABLE 3. Model Performance Comparison in Esc50 Dataset

Model	Accuracy	Precision	Recall	F1 Score
MLP [29]	0.620^{*}	0.571	0.501	0.478
CRNN [30]	0.841^{**}	0.898	0.823^{**}	0.879^{**}
RNN [31]	0.413^{*}	0.342^{**}	0.407^{*}	0.263^{**}
PI RSNN	0.91	0.870	0.881	0.847

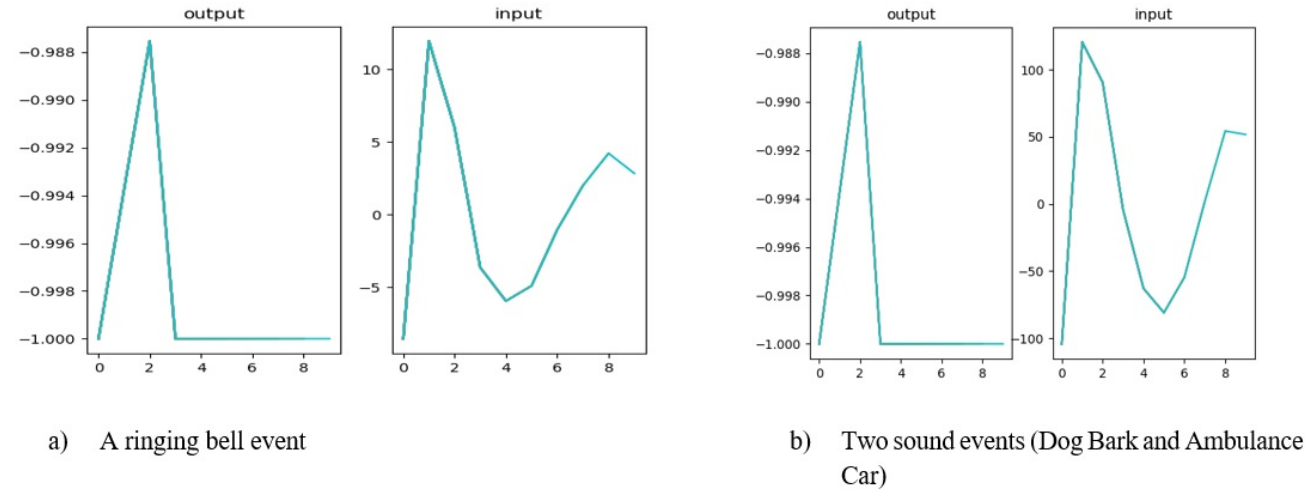
*= p$\leq$ 0.05, and **= p$\leq$ 0.01 denote statistical significance levels.

Table2 indicates the PI RSNN model achieves the highest precision of 73%, underscoring its effectiveness in classification tasks. Each model, namely MLP, CRNN, RNN, and PI RSNN exhibits distinct classification capabilities. Notably, CRNN demonstrates a precision of 72%, accompanied by balanced precision, recall, and F1 score values, indicating its robustness in accurately identifying and classifying instances. Conversely, RNN lags with a precision of 11% and the lowest precision, recall, and F1 score values, suggesting limitations in its classification effectiveness. MLP metrics, while not significantly different from PI RSNN, display a trend to perform at a lower level. The comparison underscores the importance of selecting appropriate neural network architectures tailored to specific classification tasks, as different models exhibit varying performance levels. Further analysis and experimentation may be necessary to understand the underlying factors contributing to the observed performance differences among the models.
The performance comparison on the ESC-50 dataset (Table 3) demonstrates the effectiveness of the PI RSNN model in environmental sound classification. With an accuracy of 91.0%, the PI RSNN outperforms traditional architectures and compares favorably with state-of-the-art (SOTA) results, which range from 46% to 98% [40]. It's important to note that most SOTA approaches utilize time-frequency features or are pretrained on large datasets, unlike our time-domain signal approach. While our method may not achieve the highest reported accuracy, it provides a feasible accuracy of 91.0% with lower density and reduced training costs. The CRNN model shows promising results with 84.1% accuracy, while MLP and RNN models achieve 62.0% and 41.3% accuracy, respectively. These findings are statistically significant, as indicated by the asterisks. Notably, the PI RSNN's performance



a) A Ringing Bell b) Three Speeches c) Dog Barking and Ambulance Car

FIGURE 7. **Three firing pattern samples of multiple events a) A ringing Bells, b) Three Speech sounds, and c) Dog Barking and Ambulance Car**



a) A ringing bell event b) Two sound events (Dog Bark and Ambulance Car)

FIGURE 8. **Sound Event Detection by the observed neuron activation at the embarking of the process a) a ringing bell and b) two sound events, including dog barking and Ambulance Car**

surpasses the reported human accuracy of 81.3% on this dataset. The model's balanced performance across all metrics, including precision (87.0%), recall (88.1%), and F1 score (84.7%), further underscores its robustness and effectiveness in tackling complex environmental sound classification tasks, especially considering its efficient use of resources.

In this evaluation, within the same processor, the MLP and our proposed method exhibit comparable computational costs(almost 0.216 sec/ epoch), addressing real-world feasibility concerns. In contrast, the RNN and CRNN models incur approximately 2.5 and 3.2 times higher computational costs, respectively. This improved efficiency, while maintaining high accuracy, positions our model as a practical solution for audio classification tasks, especially when considering resource constraints.

In Fig.12, we illustrate the mechanism of potential threshold and its updating process, showcasing three samples with varying learning rates of firing adaptation and desired mean firing rates.

Fig.13 demonstrates the impact of these parameters, namely the learning rate and f_d, on the resulting spike patterns.

Selecting appropriate parameters for tuning the threshold is crucial. Low or high values for the learning rate and f_d can lead to insufficient threshold adaptation. As shown in Fig.13, these scenarios may result in a lack of firing, hindering the network's proper function. Moderate values, on the other hand, enhance efficiency.

Figure 12 indicates that, across three examples of f_d and learning rate combinations, the spiking threshold converges within approximately the first 10 samples. The firing patterns, as depicted in different threshold adapting values, reveal that while the spike threshold of the neuron's membrane potential

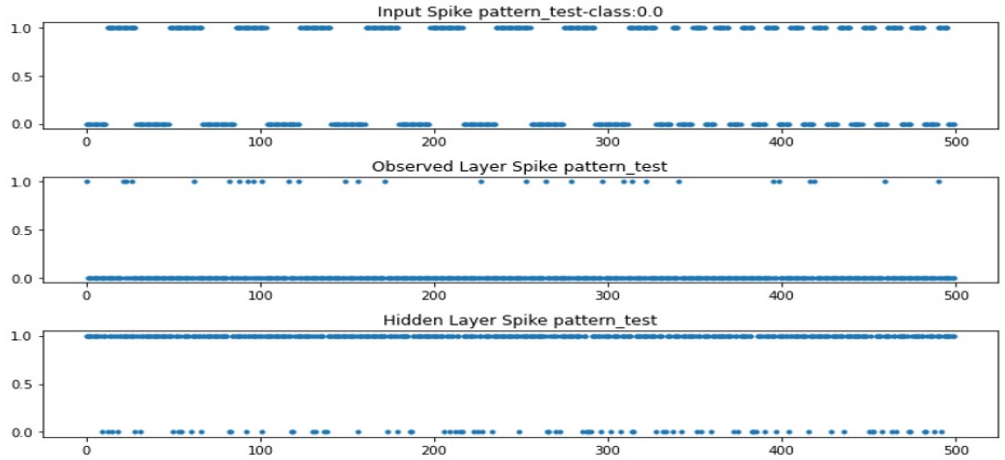


FIGURE 9. **Firing pattern of encoded input and generated output spike trains by the proposed method for class 0**

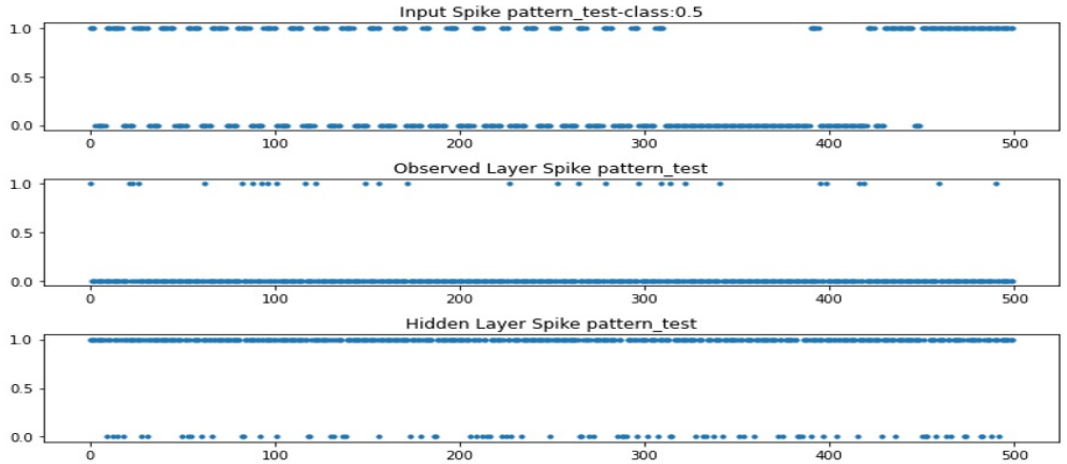


FIGURE 10. **Firing pattern of encoded input and generated output spike trains by the proposed method for class 0.5**

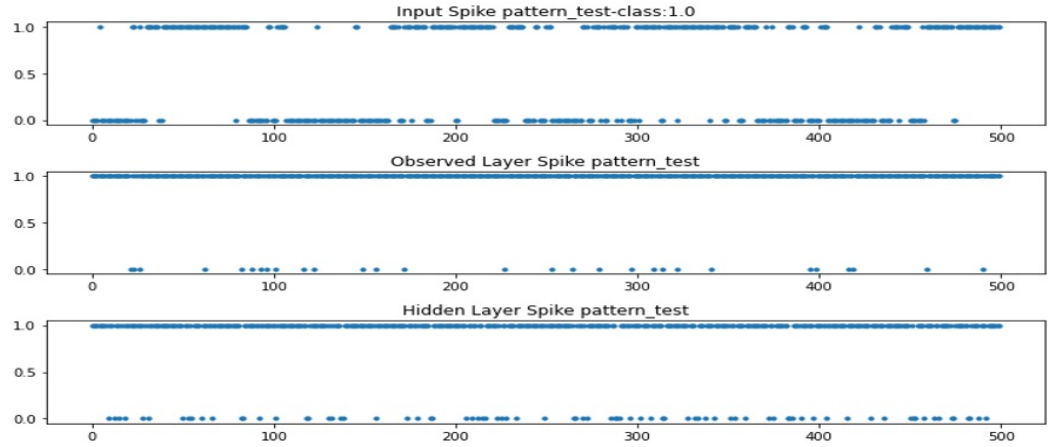


FIGURE 11. **Firing pattern of encoded input and generated output spike trains by the proposed method for class 1**

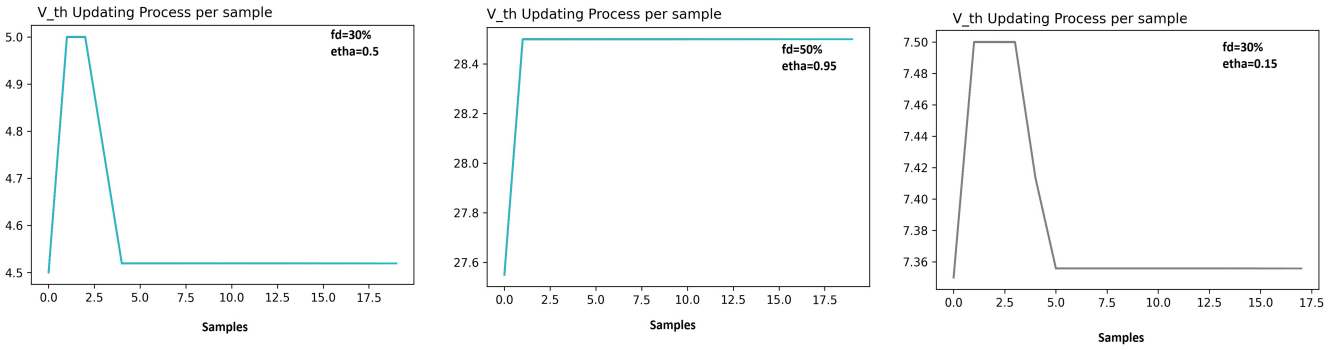


FIGURE 12. **Mechanism of potential threshold and its updating process.**

converges to a specific value, the value itself is critical. These parameters suggest that convergence to a specific value does not guarantee an optimal firing pattern. We recommend a learning rate between 0.3 and 0.7 for the firing adapting equation, with f_d adjusted between 30% and 70% to achieve optimal values.

Low f_d values may cause the omission of certain events, while high values may lead to the incorporation of noise effects. Therefore, the selection of these parameters is of significant importance.

A question may arise: Why choose an adaptation process that involves tuning two parameters instead of just one (the threshold)?

The response is that we can select a wide range between 0.3 to 0.7 for these parameters, which helps the network handle different scales and environments more effectively without requiring a re-tuning of the threshold.

By and large, the proposed PI-RSNN architecture shows strong potential for real-time applications. Its spiking neural network core allows for continuous, streaming online detection of sound events. While real-time detection is readily achievable, accurate sound event classification typically benefits from a windowing approach, where a short segment of audio is processed as a batch. This presents a trade-off: applications prioritizing immediate event awareness can leverage online detection, while those requiring higher classification accuracy can employ batch processing with appropriately sized windows to balance latency and performance.

A key challenge remains in the detection and classification of mixed, overlapping sound events. The superposition of simultaneous sound sources can lead to misclassifications. While our method effectively utilizes temporal patterns for sound detection, these overlapping patterns present a potential limitation. To mitigate this, we propose future investigation into the application of Empirical Mode Decomposition (EMD). EMD could decompose complex audio signals into their intrinsic mode functions, isolating distinct components of the mixed sounds. Furthermore, incorporating spatio-temporal data, such as the spatial location of each sound source alongside its temporal power pattern, could provide an additional layer of differentiation. This would enable the system to analyze both the timing and spatial origins of sound events, enhancing classification accuracy in complex auditory scenes. However, fully addressing this challenge necessitates extensive future research, including testing across diverse scenarios and evaluating the effectiveness of these methods under varying conditions. Such investigations will be crucial for improving the robustness and reliability of our framework in real-world applications involving multiple mixed sound sources.

IV. CONCLUSION

This study presents an innovative learning algorithm for recurrent Spiking Neural Networks (rSNNs) that combines the power-matched approach with STDP-Tempotron principles. This integration enables the detection of multiple sound events by analyzing subtle fluctuations in the power of hidden neurons. Unlike conventional Tempotron classifiers, which are limited to recognizing single sound events, this method relies solely on time-domain information, distinguishing it from most neural network classifiers that depend on time-frequency data.

The proposed adaptive thresholded spiking neural network (PI-RSNN) has achieved a sound source classification precision of 73%, surpassing traditional methods such as Convolutional Recurrent Neural Networks (CRNNs), Recurrent Neural Networks (RNNs), and Multilayer Perceptrons (MLPs) in TAU-NIGEN Dataset. While it achieves comparable F-scores to CRNN models, the PI-RSNN has distinct advantages in terms of reduced computational complexity and its suitability for low-power, real-time processing. These features make rSNNs a promising option for resource-constrained applications, such as wearable devices and edge computing platforms, where energy efficiency is crucial.

The significance of this work lies in its potential to bridge the gap between biological auditory processing and machine learning, paving the way for a new generation of efficient, robust, and intelligent sound event detection systems. By leveraging the temporal dynamics of spiking neurons, the proposed framework offers a biologically plausible approach to complex sound event detection, which could inspire further exploration of spiking neural networks in various applica-

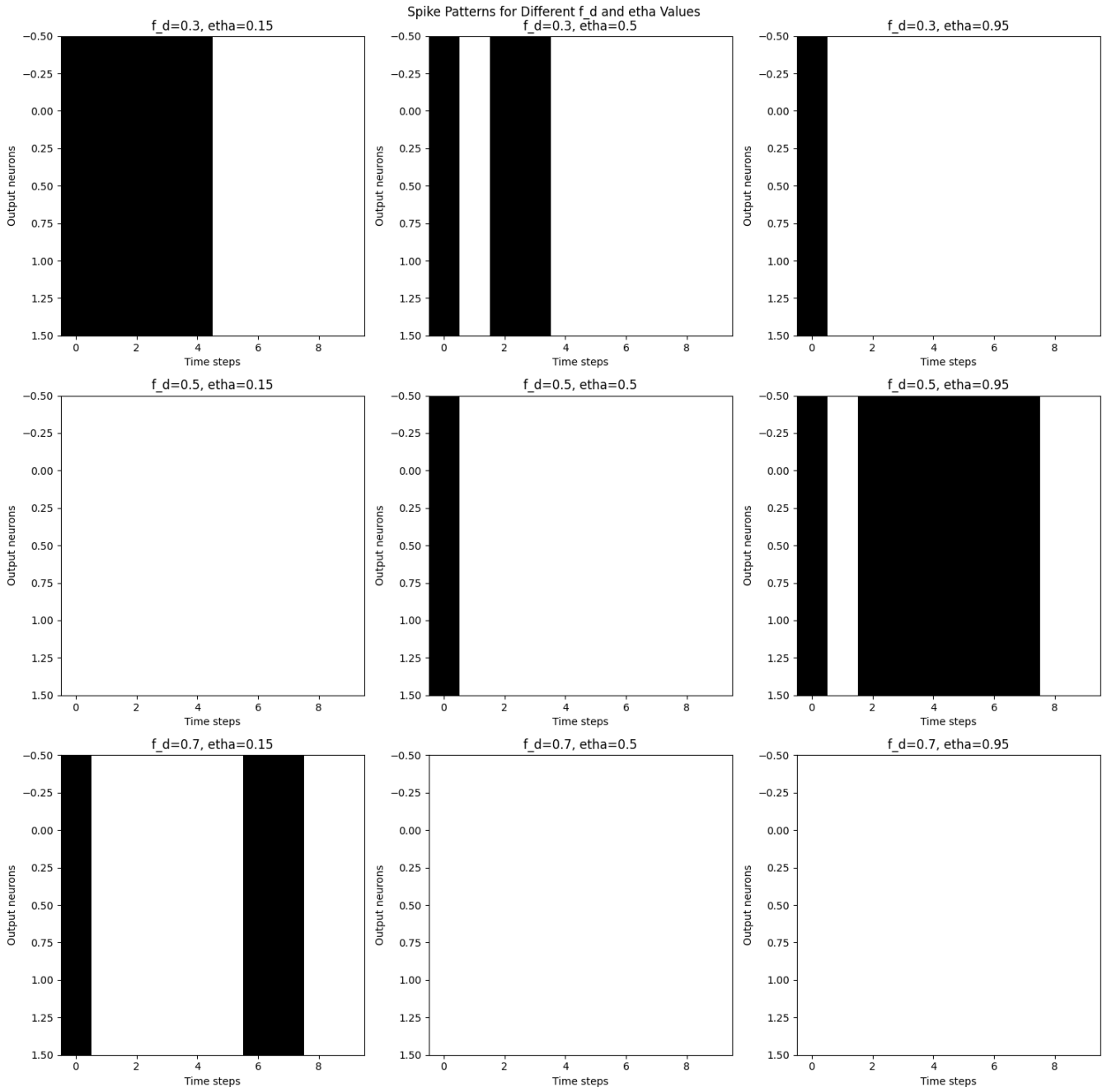


FIGURE 13. **Impact of firing learning rate and f_d on spike patterns.**

tions.

Furthermore, the architecture has substantial potential across various real-world applications. In smart home technology, it can improve voice recognition in devices like smart speakers and home assistants, effectively differentiating between multiple sound sources even in noisy environments. In hearing aids, it can help users focus on specific sounds while filtering out background noise. Additionally, this architecture could enhance industrial monitoring systems by detecting and classifying operational sounds, thereby improving safety and efficiency. Its application in environmental monitoring can aid wildlife conservation efforts by classifying natural sounds in diverse habitats. Overall, the versatility of this architecture highlights its broader impact across multiple sectors, paving the way for more efficient auditory processing technologies that operate reliably under challenging conditions. Continued research will determine whether further improvements in accuracy and reliability can be achieved through more complex samples and comprehensive training-testing strategies

Future research should investigate the potential of RSNNs in

more challenging acoustic environments, particularly those with overlapping sound events, a known limitation in sound event localization and detection (SELD). Addressing this challenge could involve exploring sound separation techniques as a pre-processing step to disentangle mixed audio signals. Additionally, there is a need to explore advanced encoding strategies that utilize the temporal dynamics of spiking neurons for improved feature representation. Specifically, integrating attention mechanisms within the RSNN framework could help selectively process relevant acoustic features, and developing hybrid architectures that combine the strengths of both spiking and traditional artificial neural networks could enhance performance further. Given the difficulty in effectively utilizing unlabeled data, future work should also explore semi-supervised or unsupervised learning approaches to leverage the abundance of unlabeled audio data and improve the generalization ability of RSNNs.

APPENDIX

To prove the updating law for observed LIF neurons,given in (8), consider the square error cost function as follows:

$$J = \frac{1}{2} E^T E \tag{9}$$

where E is defined in (6). Equation (9) describes the square error on normalized sound signal's power estimation.
To obtain the observed neurons' weight updating law, the gradient decent of J with respect to W_{oh} is calculated, using chain rule as follows:

$$\Delta W_{oh} = -\eta \frac{\partial J}{\partial W_{oh}} =$$

$$-0.5\eta . E^T \times \left(\frac{\partial E}{\partial W_{oh}} \right) - 0.5\eta(k) . \left(\frac{\partial E}{\partial W_{oh}} \right)^T \times E \tag{10}$$

where η is learning rate, a constant value between 0 and 1.

$$\frac{\partial E}{\partial W_{oh}} = \frac{\partial (X_{in} - \left({V_{ob}}^T \times I_{oh} \right)^T)}{\partial W_{oh}} = -\left(\frac{\partial I_{oh}}{\partial W_{oh}} \right)^T \times V_{ob} \tag{11}$$

$$\frac{\partial I}{\partial W_{oh}} = V_{ob} \times V^T \tag{12}$$

The substitution of equations (11) and (12) into equation (10) results in equation (8).

ACKNOWLEDGMENT

We would like to extend our heartfelt thanks and profound appreciation to Auckland University of Technology for granting us the opportunity to conduct our research.

REFERENCES

[1] J. Wu, Y. Chua, M. Zhang, H. Li, and K. C. Tan, "A spiking neural network framework for robust sound classification," Frontiers in Neuroscience, vol. 12, p. 836, 2018.
[2] K. Yamazaki, V.-K. Vo-Ho, D. Bulsara, and N. Le, "Spiking neural networks and their applications: A Review," Brain Sciences, vol. 12, no. 7, p. 863, 2022.
[3] A. Taherkhani, A. Belatreche, Y. Li, G. Cosma, L. P. Maguire, and T. M. McGinnity, "A review of learning in biologically plausible spiking neural networks," Neural Networks, vol. 122, pp. 253-272, 2020.
[4] Z. Pan, H. Li, J. Wu, and Y. Chua, "An event-based cochlear filter temporal encoding scheme for speech signals," in 2018 International Joint Conference on Neural Networks (IJCNN), 2018: IEEE, pp. 1-8.
[5] E. O. Neftci, H. Mostafa, and F. Zenke, "Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks," IEEE Signal Processing Magazine, vol. 36, no. 6, pp. 51-63, 2019.
[6] R. E. Suri and T. J. Sejnowski, "Spike propagation synchronized by temporally asymmetric Hebbian learning," Biological cybernetics, vol. 87, no. 5, pp. 440-445, 2002.
[7] F. Zenke and S. Ganguli, "Superspike: Supervised learning in multilayer spiking neural networks," Neural computation, vol. 30, no. 6, pp. 1514-1541, 2018.
[8] S. A. Saeedinia, M. R. Jahed-Motlagh, A. Tafakhori, and N. Kasabov, "Design of MRI structured spiking neural networks and learning algorithms for personalized modelling, analysis, and prediction of EEG signals," Scientific Reports, vol. 11, no. 1, p. 12064, 2021.
[9] N. Caporale and Y. Dan, "Spike timing–dependent plasticity: a Hebbian learning rule," Annu. Rev. Neurosci., vol. 31, pp. 25-46, 2008.
[10] L. R. Iyer and Y. Chua, "Classifying neuromorphic datasets with tempotron and spike timing dependent plasticity," in 2020 International Joint Conference on Neural Networks (IJCNN), 2020: IEEE, pp. 1-8.
[11] R. V. Florian, "Tempotron-Like Learning with ReSuMe," Berlin, Heidelberg, 2008: Springer Berlin Heidelberg, in Artificial Neural Networks - ICANN 2008, pp. 368-375.
[12] R. Gütig and H. Sompolinsky, "The tempotron: a neuron that learns spike timing–based decisions," Nature neuroscience, vol. 9, no. 3, pp. 420-428, 2006.
[13] R. Rubin, R. Monasson, and H. Sompolinsky, "Theory of spike timing-based neural classifiers," Physical review letters, vol. 105, no. 21, p. 218102, 2010.
[14] J. Wu, Y. Chua, M. Zhang, H. Li, and K. C. Tan, "A spiking neural network framework for robust sound classification," Frontiers in neuroscience, vol. 12, p. 836, 2018.
[15] R. Gütig, R. Aharonov, S. Rotter, and H. Sompolinsky, "Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity," Journal of Neuroscience, vol. 23, no. 9, pp. 3697-3714, 2003.
[16] M. Dong, X. Huang, and B. Xu, "Unsupervised speech recognition through spike-timing-dependent plasticity in a convolutional spiking neural network," PloS one, vol. 13, no. 11, p. e0204596, 2018.
[17] L. R. Iyer and Y. Chua, "Classifying neuromorphic datasets with tempotron and spike timing dependent plasticity," in 2020 International Joint Conference on Neural Networks (IJCNN), 2020: IEEE, pp. 1-8.
[18] T. Masquelier, R. Guyonneau, and S. J. Thorpe, "Competitive STDP-based spike pattern learning," Neural computation, vol. 21, no. 5, pp. 1259-1276, 2009.
[19] S. R. Kheradpisheh, M. Ganjtabesh, S. J. Thorpe, and T. Masquelier, "STDP-based spiking deep convolutional neural networks for object recognition," Neural Networks, vol. 99, pp. 56-67, 2018.
[20] A. Amir et al., "A low power, fully event-based gesture recognition system," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 7243-7252.
[21] T. Serrano-Gotarredona, B. Linares-Barranco, F. Galluppi, L. Plana, and S. Furber, "ConvNets experiments on SpiNNaker," in 2015 IEEE International Symposium on Circuits and Systems (ISCAS), 2015: IEEE, pp. 2405-2408.
[22] Y. Liu, H. Fu, Y. Wei, and H. Zhang, "Sound event classification based on frequency-energy feature representation and two-stage data dimension reduction," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 31, pp. 1290-1304, 2023.
[23] C. Ganguly, S. S. Bezugam, E. Abs, M. Payvand, S. Dey, and M. Suri, "Spike frequency adaptation: bridging neural models and neuromorphic applications," Communications Engineering, vol. 3, no. 1, p. 22, 2024.

[24] Z. Roozbehi, A. Narayanan, M. Mohaghegh, and S.-A. Saeedinia, "Dynamic-Structured Reservoir Spiking Neural Network in Sound Localization," IEEE Access, 2024.
[25] A. Bittar and P. N. Garner, "A surrogate gradient spiking baseline for speech command recognition," Frontiers in Neuroscience, vol. 16, p. 865897, 2022.
[26] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," Nature Reviews Physics, vol. 3, no. 6, pp. 422-440, 2021.
[27] A. A. S. . V. T. Politis., "A Dataset of Reverberant Spatial Sound Scenes with Moving Sources for Sound Event Localization and Detection," In Proceedings of the Detection and Classification of Acoustic Scenes and Events 2020 Workshop (*DCASE*2020), Tokyo, Japan, 2020.
[28] M. Elhamod et al., "Hierarchy-guided neural network for species classification," Methods in Ecology and Evolution, vol. 13, no. 3, pp. 642-652, 2022.
[29] R. Pahuja and A. Kumar, "Sound-spectrogram based automatic bird species recognition using MLP classifier," Applied Acoustics, vol. 180, p. 108077, 2021.
[30] J. Sang, S. Park, and J. Lee, "Convolutional recurrent neural networks for urban sound classification using raw waveforms," in 2018 26th European Signal Processing Conference (EUSIPCO), 2018: IEEE, pp. 2444-2448.
[31] H. Phan, P. Koch, F. Katzberg, M. Maass, R. Mazur, and A. Mertins, "Audio scene classification with deep recurrent neural networks," arXiv preprint arXiv:1703.04770, 2017.
[32] Y. Qiu, L. Ma, and R. Priyadarshi, "Deep learning challenges and prospects in wireless sensor network deployment," Archives of Computational Methods in Engineering, pp. 1-24, 2024.
[33] A. Qurthobi, R. Damasevicius, V. Barzdaitis, and R. Maskeliunas, "Robust Forest Sound Classification Using Pareto-Mordukhovich Optimized MFCC in Environmental Monitoring," IEEE Access, 2025.
[34] M. Orosoo et al., "Transforming English language learning: Advanced speech recognition with MLP-LSTM for personalized education," Alexandria Engineering Journal, vol. 111, pp. 21-32, 2025.
[35] S. A. Saeedinia, M. R. Jahed-Motlagh, A. Tafakhori, and N. K. Kasabov, "Diagnostic biomarker discovery from brain EEG data using LSTM, reservoir-SNN, and NeuCube methods in a pilot study comparing epilepsy and migraine," Scientific Reports, vol. 14, no. 1, p. 10667, 2024.
[36] D. G. Peterson, T. Nawarathne, and H. Leung, "Modulating stdp with back-propagated error signals to train snns for audio classification," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 7, no. 1, pp. 89-101, 2022.
[37] Y. Zheng, J. Xue, J. Liu, and Y. Zhang, "Biologically Inspired Spatial–Temporal Perceiving Strategies for Spiking Neural Network," Biomimetics, vol. 10, no. 1, p. 48, 2025.
[38] Q. Yang, Q. Liu, N. Li, M. Ge, Z. Song, and H. Li, "SVAD: A Robust, Low-Power, and Light-Weight Voice Activity Detection with Spiking Neural Networks," in ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2024: IEEE, pp. 221-225.
[39] K. J. Piczak, "ESC: Dataset for environmental sound classification," in Proceedings of the 23rd ACM international conference on Multimedia, 2015, pp. 1015-1018.
[40] K. J. Piczak, "ESC: Dataset for Environmental Sound Classification," GitHub repository, 2015. [Online]. Available: https://github.com/karolpiczak/ESC-50. [Accessed: Feb. 6, 2025].

ZAHRA ROOZBEHI has a degree in Ba Honar Kerman University and an M.Sc. degree from Alzahra University. She is a Ph.D. of computer science, researching the spatio-temporal behavior of the brain. Her expertise includes data analysis, mathematics, neuroscience, and research tool implementation. She enjoys generating new ideas and finding solutions. Her colleagues describe her as motivated and resourceful. She is currently working on mimicking the mammalian hearing system and applying the model in technology. Her interests include data analytics, computational neuroscience, brain-inspired modeling, fractal geometry, and spiking neural networks.

AJIT NARAYANAN received his B.Sc. (Hons.) degree in communication science and linguistics from the University of Aston, Birmingham, U.K. in 1973, and his Ph.D. degree in philosophy from the University of Exeter, Philosophy, Exeter, U.K. in 1976. He is presently a Professor at the Auckland University of Technology, School of Engineering, Computer Science, and Mathematics. He has been a lecturer, professor, and dean at universities in the United Kingdom before arriving in New Zealand in 2007. He has published over 100 articles and conducted reviews for various journals and conferences on artificial intelligence and its applications in the medical field. His research interests include artificial intelligence, nature-inspired computing, machine learning, computational statistics, and machine ethics.

MAHSA MOHAGHEGH is a Senior Lecturer and Director of Women in Tech in AUT's School of Computer, Engineering, and Mathematical Sciences. She is a well-recognized leader in AI and machine learning. She is also the founder of the charitable trust She Sharp, a women's technology networking and learning group, where she works to encourage young New Zealand girls to consider what a career in technology offers. She was named the Emerging Leader category winner in the 2013 Westpac Women of Influence Awards and was one of ten finalists for the 2018 Kiwibank New Zealander of the Year. In 2019 she was the Champion Award winner of the YWCA Equal Pay awards, and in 2020 presented the Massey University Distinguished Alumni Award.

SAMANEH-ALSADAT SAEEDINIA holds a Bachelor's degree from Imam Khomeini International University and a Master's degree from Iran University of Science and Technology. Her research interests encompass intelligent control and modeling, neuroscience, and neural computations. In 2006, she was awarded third prize in the Kharazmi Scientific Competition in Tehran Province. Saeedinia published a book on electrical instruments in 2018. She is currently a Ph.D. in Control Electrical Engineering.