Energy-Efficient Drowsiness Detection with Spiking Neural Networks

Satya Sairam Varma Manthena*[1], Peng Kang*[1]
*Department of Computer Science, University of Illinois Springfield, Springfield, IL, USA

***Abstract*—Driver drowsiness detection is a critical component of intelligent transportation safety systems, where energy-efficient, real-time performance is essential for reliable in-vehicle deployment. In this work, we propose a Convolutional Spiking Neural Network (CSNN) that leverages event-driven computation to achieve high accuracy with low power consumption. A custom eye–mouth image dataset was constructed to train and evaluate both the CSNN and a conventional Convolutional Neural Network (CNN) under identical conditions. The proposed CSNN achieved a best testing accuracy of 97.94%, slightly outperforming the CNN baseline (97.79%), while reducing estimated energy consumption by approximately 61%. These results demonstrate the strong potential of spiking computation for efficient visual recognition in resource-constrained embedded systems, paving the way for neuromorphic implementations of real-time fatigue monitoring systems on edge devices.**

***Index Terms*—Spiking Neural Networks, Convolutional Neural Networks, Drowsiness Detection**

I. Introduction

Artificial Neural Networks (ANNs) have led to remarkable progress in machine cognition, enabling systems to achieve human-like performance in tasks speech recognition [1], gesture classification [2], computer vision [3], [4], [5], and natural language processing [6], [7], [8], [9]. Despite their success, ANNs only capture limited aspects of biological brain function, primarily focusing on dense connectivity and hierarchical organization [10]. In contrast, the mammalian brain also leverages intricate time-dependent signaling and event-driven processing for information flow, capabilities that conventional ANNs generally lack [11], [12].

To address this gap, researchers have introduced Spiking Neural Networks (SNNs) [13]. Different from artificial neural networks (ANNs) that rely on continuous-valued activations and dense matrix multiplications, spiking neural networks (SNNs) process information through discrete 0–1 spikes. This event-driven computation replaces intensive dot-product operations with sparse summations, substantially reducing the overall computational load. Moreover, neurons in SNNs remain quiescent unless triggered by incoming spikes, resulting in asynchronous and energy-efficient processing. In addition, SNNs are particularly well-suited for capturing temporal dynamics and sequential dependencies in time-varying signals. These advantages make SNNs a promising framework for low-power, real-time processing on emerging neuromorphic hardware platforms [14]. Figure 1 illustrates a detailed comparison between ANNs and SNNs.

Recent work has demonstrated that SNNs can perform competitively on sequential event-based learning tasks such as speech recognition [15], gesture classification [16], tactile recognition [17], and event-based vision [18], [19], [20]. While SNNs have shown strong performance in these event-driven domains, their potential in drowsiness detection remains largely underexplored. This gap is particularly important because drowsiness detection systems are often deployed in resource-constrained and energy-sensitive environments, such as embedded modules within vehicles. In such settings, continuous monitoring must be achieved with minimal power consumption and low-latency response to ensure real-time safety intervention. The inherent event-driven and energy-efficient nature of SNNs makes them a promising candidate for low-power, real-time drowsiness detection in vehicular and wearable applications.

In this work, we develop and evaluate a Convolutional Spiking Neural Network (CSNN) for the task of drowsiness detection, comparing its performance to a conventional CNN baseline under identical datasets and evaluation protocols. Our experimental results demonstrate that CSNNs can achieve comparable accuracy while substantially reducing computational cost, highlighting their potential as an energy-efficient

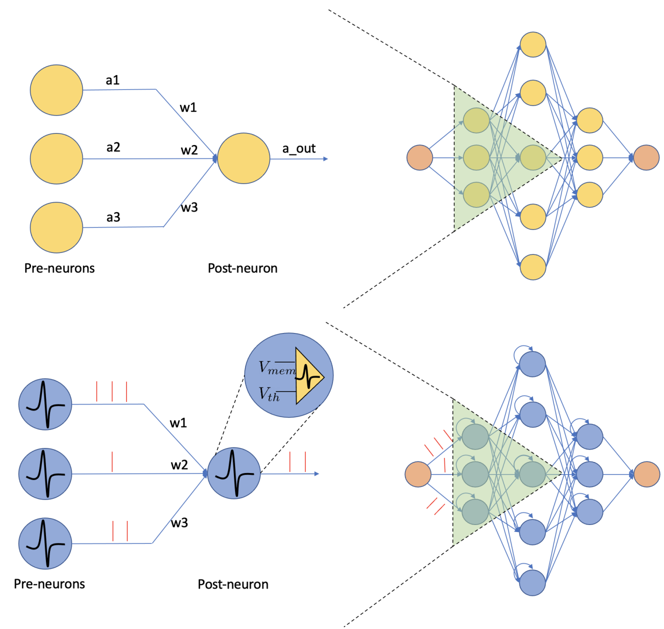

Fig. 1. Comparison between an Artificial Neural Network (ANN) and a Spiking Neural Network (SNN) architectures. The upper diagram illustrates the ANN, where information is transmitted by continuous-valued activations and computation is performed as weighted sums followed by nonlinear functions. The lower diagram shows the SNN, where neurons communicate via discrete spikes (events), and computation is event-driven with neuron activation depending on reaching a membrane threshold.

[1]Both authors contributed equally to this work. Peng Kang is the corresponding author. Email: pkang8@uis.edu

solution for real-world fatigue monitoring where both precision and low-power operation are critical.

II. Methods

We designed an experimental framework to evaluate the effectiveness of spiking neural networks in drowsiness detection. The methodology includes constructing a representative dataset of alert and drowsy states, developing a Convolutional Spiking Neural Network (CSNN) model, and training it under consistent protocols alongside a CNN baseline to compare accuracy and efficiency.

A. Dataset Construction

For this study, we compiled and published a custom dataset based on Kaggle[1]. Raw facial images, originally sourced from two public repositories[2][3], were systematically combined and preprocessed into a unified collection. Each image was converted to grayscale and resized to $(1, 28, 28)$ pixels. Paired eye and mouth cues were then merged into $(2, 28, 28)$ tensors by concatenating the processed eye and mouth images.

Labeling was performed as follows:

- **Open Eyes + Yawn**: classified as drowsy (1)
- **Open Eyes + No Yawn**: classified as not drowsy (0)
- **Closed Eyes + Yawn**: classified as drowsy (1)
- **Closed Eyes + No Yawn**: classified as drowsy (1)

Each sample's label was determined by this rule, marking the presence of yawning or closed eyes as indicators of drowsiness. To ensure reliability, the dataset was stratified into 4000 training samples and 1360 testing samples, with strict guarantees of zero overlap between the two sets. This separation prevents data leakage and ensures that evaluation metrics reflect genuine generalization performance.

Class balancing was also enforced by limiting the number of samples per combination (eye/mouth pair and label) per set, which guards against model bias induced by imbalanced data distributions. The final dataset thus provides a robust, unbiased baseline for comparative training and evaluation of both the CNN and CSNN architectures.

B. Model Architecture

To investigate the advantages of spiking computation in drowsiness detection, we developed a Convolutional Spiking Neural Network (CSNN) inspired by the structure of a conventional CNN baseline. The CSNN maintains the same hierarchical convolutional organization but replaces continuous-valued activations with spiking neurons that communicate through discrete binary spikes. This event-driven mechanism enables sparse and energy-efficient computation while retaining comparable representational power.

The overall architecture is summarized in Figure 3. The proposed CSNN, implemented using the SpikingJelly framework, consists of three convolutional blocks followed by two fully connected layers. Each convolutional block contains a convolution layer, batch normalization, an Integrate-and-Fire (IF) neuron layer, and a max-pooling layer for spatial downsampling. The final classifier module includes a flatten layer, a linear projection, dropout regularization, and an output spiking layer producing class-level firing rates. The entire network operates over multiple discrete timesteps T, during which the same input frame is repeatedly presented to simulate temporal dynamics. Finally, the class-level average firing rate across timesteps is then used as the network's output prediction.

Fig. 2. Example combined eye–mouth images, where label 1 indicates drowsy, and label 0 indicates not drowsy.

Mathematically, for each layer l, the membrane potential u_t^l at timestep t is updated as:

$$u_t^l = \lambda u_{t-1}^l + W^l o_t^{l-1}, \tag{1}$$

where λ denotes the membrane decay factor, W^l represents the synaptic weights, and o_t^{l-1} is the spike output from the previous layer. A spike is emitted whenever u_t^l exceeds the firing threshold θ, i.e.,

$$o_t^l = H(u_t^l - \theta), \tag{2}$$

with $H(\cdot)$ being the Heaviside step function.

C. Training Details

Following the work [21], we adopt direct training with surrogate gradients as our approach to optimize SNNs. During backpropagation, the non-differentiable Heaviside step

[1]https://www.kaggle.com
[2]https://www.kaggle.com/datasets/prasadvpatil/mrl-dataset
[3]https://www.kaggle.com/datasets/davidvazquezcic/yawn-dataset/data

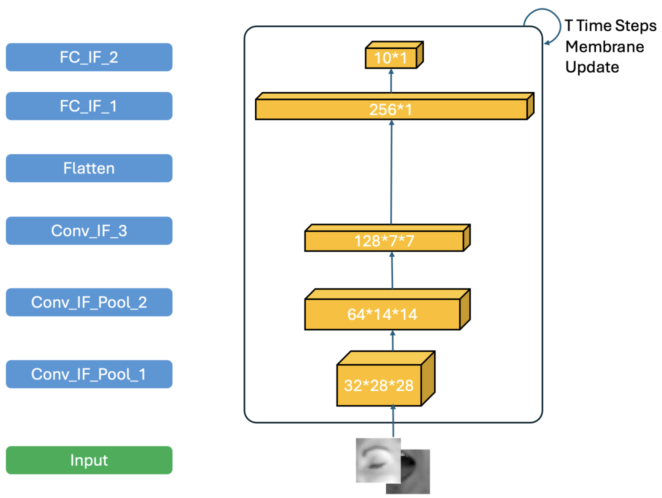

Fig. 3. Architecture of the proposed Convolutional Spiking Neural Network (CSNN).

function is replaced with an arctangent-based surrogate that provides a smooth gradient approximation. Specifically, the surrogate function is defined as:

$$O_t \approx \frac{1}{\pi} \arctan\left(\frac{\pi}{2}\alpha U_t\right) + \frac{1}{2}, \tag{3}$$

where α is a hyperparameter controlling the steepness of the arctangent curve. A larger α results in a sharper transition and more closely resembles the ideal step function.

The derivative of O_t with respect to the membrane potential U_t is given by:

$$\frac{\partial O_t}{\partial U_t} = \frac{\alpha/2}{1 + \left(\frac{\pi}{2}\alpha U_t\right)^2}, \tag{4}$$

which yields a continuous surrogate gradient suitable for back-propagation through time (BPTT). This formulation enables end-to-end training of the CSNN in a manner analogous to conventional CNNs while maintaining the temporal characteristics of spiking activity.

The same training process was applied to both models (CNN and CSNN) to ensure experimental fairness. Each model was trained for 100 epochs with a batch size of 128, using the Adam optimizer and a learning rate of 0.001. Cross-entropy loss was employed as the training objective, and model performance was evaluated on the held-out test set after each epoch. This consistent setup allows direct comparison between the CNN and CSNN in terms of both accuracy and computational efficiency.

III. Experiments

In this section, we present a comprehensive evaluation of the proposed Convolutional Spiking Neural Network (CSNN) in comparison with a conventional Convolutional Neural Network (CNN) baseline. Both models were trained and tested on the custom drowsiness detection dataset described earlier under identical experimental conditions to ensure a fair comparison. We assess their performance in terms of classification accuracy and energy consumption. In particular, we analyze how the event-driven nature of the CSNN contributes to reduced energy usage while maintaining competitive accuracy, highlighting its potential for real-time, low-power deployment in embedded fatigue monitoring systems.

A. Accuracy Comparison

Figure 4 and Figure 5 show the training and testing accuracy profiles of the CNN and CSNN models, respectively, over 100 epochs. Both models exhibit rapid convergence within the first 20 epochs, followed by stable accuracy with minimal signs of overfitting.

The CNN achieved a best testing accuracy of 97.79%, while the CSNN reached a slightly higher best testing accuracy of 97.94%. The CSNN was trained with a temporal window of $T = 10$ timesteps, allowing it to integrate temporal information through spike-based computation. Despite its discrete and event-driven nature, the CSNN achieved comparable—and slightly superior—performance to the CNN, demonstrating its capability to capture temporal dynamics inherent in fatigue-related visual cues. The overall stability of both training profiles indicates well-optimized learning behavior under identical experimental conditions.

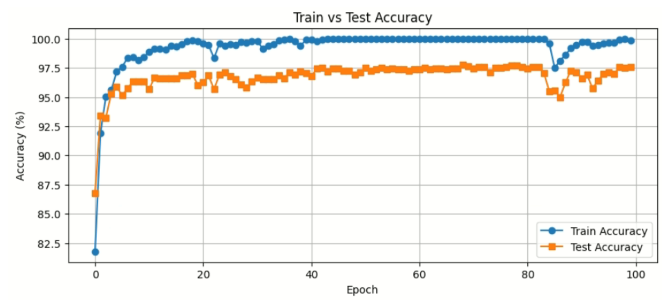

Fig. 4. Training and testing accuracy curves of the CNN model over 100 epochs.

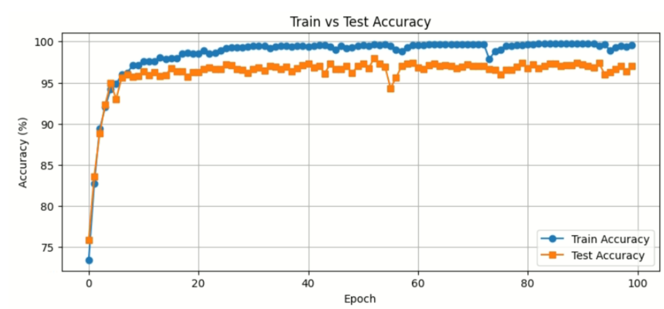

Fig. 5. Training and testing accuracy curves of the CSNN model over 100 epochs ($T = 10$).

To further examine the effect of temporal resolution on model performance, we conducted an ablation study by varying the number of simulation timesteps $T \in \{2, 4, 6, 8, 10\}$. As shown in Table I, the CSNN achieved progressively higher accuracy as T increased, reflecting improved temporal integration capability with longer spike sequences.

TABLE I
ABLATION STUDY ON THE NUMBER OF TIMESTEPS T FOR THE CSNN.

Timesteps (T)	**Test Accuracy (%)**
2	97.06
4	97.57
6	97.43
8	97.79
10	**97.94**

These results indicate that incorporating moderate temporal depth enhances the model's ability to encode temporal dependencies while keeping computational overhead manageable.

B. Energy Efficiency Analysis

To quantitatively evaluate the computational and energy efficiency of the proposed CSNN relative to the CNN baseline, we analyzed both the arithmetic operation counts and their estimated energy costs. In conventional CNNs, the computational load is dominated by multiply–accumulate (MAC) operations, whereas in spiking networks, computations are event-driven and consist solely of accumulations (ACs) triggered by spike events, eliminating multiplications altogether.

For the CNN, the total number of floating-point operations (FLOPs) can be expressed as:

$$\text{FLOPs}_{\text{CNN}} = 2 \times \text{MACs}_{\text{CNN}}, \tag{5}$$

where one multiplication and one addition correspond to two FLOPs per MAC. With the help of ptflops[4], we obtain that the CNN contains $\text{MACs}_{\text{CNN}} = 9.53\text{M}$. Thus, the total FLOPs are 19.06M.

For the CSNN, the total number of additions depends on the average firing rate f and the number of simulation timesteps T:

$$\text{ACs}_{\text{CSNN}} = \text{MACs}_{\text{CNN}} \times f \times T. \tag{6}$$

Since no multiplications occur in the spiking computation, the total FLOPs for the CSNN equal its number of additions. By evaluation the average firing rate over the testing samples in our dataset, we obtain $f = 0.1986$ and $T = 10$, the total number of accumulations for the CSNN is 18.91M.

In terms of hardware energy, it has been reported that one accumulation (AC) is approximately $\mathbf{5.1}\times$ more energy-efficient than one MAC operation in the 45 nm CMOS process [22]. Based on this, the normalized energy consumption of the CSNN relative to the CNN can be found in Table II.

These results clearly demonstrate that the proposed CSNN achieves substantial energy savings while preserving comparable accuracy (97.94% vs. 97.79%). By replacing energy-expensive multiplications with lightweight additions and leveraging sparse, event-driven activations, the CSNN provides an energy-efficient alternative.

[4]https://pypi.org/project/ptflops/

IV. CONCLUSION

In this work, we developed and evaluated a Convolutional Spiking Neural Network (CSNN) for energy-efficient drowsiness detection. A custom eye–mouth image dataset was constructed to support fair comparison with a conventional CNN baseline. Experimental results demonstrate that the proposed CSNN achieves a best testing accuracy of 97.94%, slightly outperforming the CNN while reducing estimated energy consumption by approximately 61%. These findings confirm that spiking computation can effectively capture temporal dynamics in visual fatigue cues while offering substantial energy savings. Future work will focus on extending this approach to multimodal inputs and deploying the model on neuromorphic hardware to further validate its real-world efficiency and responsiveness in embedded systems.

ACKNOWLEDGMENT

This work was supported in part by the *UIS AI Campus Server* provided by the Information Technology Services (ITS) at the University of Illinois Springfield. We gratefully acknowledge the computing resources and technical assistance that made this research possible.

REFERENCES

[1] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE international conference on acoustics, speech and signal processing*. Ieee, 2013, pp. 6645–6649.

[2] Okan Köpüklü, Ahmet Gunduz, Neslihan Kose, and Gerhard Rigoll, "Real-time hand gesture detection and classification using convolutional neural networks," in *2019 14th IEEE international conference on automatic face & gesture recognition (FG 2019)*. IEEE, 2019, pp. 1–8.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[4] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[5] Alexey Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[6] Ilya Sutskever, Oriol Vinyals, and Quoc V Le, "Sequence to sequence learning with neural networks," *Advances in neural information processing systems*, vol. 27, 2014.

[7] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, 2019, pp. 4171–4186.

[9] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al., "Language models are unsupervised multitask learners," *OpenAI blog*, vol. 1, no. 8, pp. 9, 2019.

[10] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[11] João D Nunes, Marcelo Carvalho, Diogo Carneiro, and Jaime S Cardoso, "Spiking neural networks: A survey," *IEEE access*, vol. 10, pp. 60738–60764, 2022.

[12] Simon Davidson and Steve B Furber, "Comparison of artificial and spiking neural networks on digital hardware," *Frontiers in Neuroscience*, vol. 15, pp. 651141, 2021.

TABLE II
COMPARISON OF COMPUTATIONAL COMPLEXITY AND ESTIMATED ENERGY EFFICIENCY BETWEEN CNN AND CSNN.

Model	**MACs / ACs (M)**	**FLOPs (M)**	**Relative Energy Cost**
CNN	9.53 (MACs)	19.06	$1.0\times$
CSNN (f=0.1986, T=10)	18.91 (ACs)	18.91	$\mathbf{0.39}\times$

[13] Kaushik Roy, Akhilesh Jaiswal, and Priyadarshini Panda, "Towards spike-based machine intelligence with neuromorphic computing," *Nature*, vol. 575, no. 7784, pp. 607–617, 2019.

[14] Michael Pfeiffer and Thomas Pfeil, "Deep learning with spiking neurons: opportunities and challenges," *Frontiers in neuroscience*, vol. 12, pp. 774, 2018.

[15] Jibin Wu, Emre Yılmaz, Malu Zhang, Haizhou Li, and Kay Chen Tan, "Deep spiking neural networks for large vocabulary automatic speech recognition," *Frontiers in neuroscience*, vol. 14, pp. 199, 2020.

[16] Yannan Xing, Gaetano Di Caterina, and John Soraghan, "A new spiking convolutional recurrent neural network (scrnn) with applications to event-based hand gesture recognition," *Frontiers in neuroscience*, vol. 14, pp. 590164, 2020.

[17] Peng Kang, Srutarshi Banerjee, Henry Chopp, Aggelos Katsaggelos, and Oliver Cossairt, "Boost event-driven tactile learning with location spiking neurons," *Frontiers in Neuroscience*, vol. 17, pp. 1127537, 2023.

[18] Lin Zhu, Xiao Wang, Yi Chang, Jianing Li, Tiejun Huang, and Yonghong Tian, "Event-based video reconstruction via potential-assisted spiking neural network," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 3594–3604.

[19] Peng Kang, Hao Hu, Srutarshi Banerjee, Henry Chopp, Aggelos Katsaggelos, and Oliver Cossairt, "Human vision-like robust object recognition," in *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2021, pp. 709–713.

[20] Peng Kang, Srutarshi Banerjee, Henry Chopp, Aggelos Katsaggelos, and Oliver Cossairt, "When event-based shape from polarization meets spiking neural networks," in *2025 IEEE Conference on Artificial Intelligence (CAI)*. IEEE, 2025, pp. 767–770.

[21] Wei Fang, Yanqi Chen, Jianhao Ding, Ding Chen, Zhaofei Yu, Huihui Zhou, and Yonghong Tian, "other contributors. spikingjelly," 2020.

[22] Mark Horowitz, "1.1 computing's energy problem (and what we can do about it)," in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*. IEEE, 2014, pp. 10–14.