International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Novel Emotion Recognition Framework from Facial Expressions Using Spiking Neural Networks on Wearable Edge Devices

A.Jasvant Ram[1], K. Prema[2], N. Susila[3], K. Jaya Deepthi[4], C. Sakthi Lakshmi Priya[5], S. Manikandan[6]

Submitted: 14/05/2024 Revised: 27/06/2024 Accepted: 07/07/2024

Abstract: Real-time emotion recognition from facial expressions holds significant promise for enhancing human-computer interaction and personalizing user experiences. Harnessing the potential of advanced technologies, the research presents a novel approach to real-time emotion recognition from facial expressions using Spiking Neural Networks (SNN) on wearable edge devices. The methodology integrates key technologies including Open Neural Network Exchange (ONNX), Message Queuing Telemetry Transport (MQTT), and Long Short-Term Memory (LSTM) networks to enhance the efficiency and accuracy of emotion recognition systems in practical scenarios. By leveraging ONNX, seamless model interchangeability and deployment across diverse hardware platforms are achieved, ensuring scalability and flexibility in model deployment. Through optimized model conversion and deployment on wearable edge devices, interoperability and efficiency in real-time emotion recognition tasks are ensured. MQTT serves as a lightweight and reliable communication protocol for seamless data exchange between wearable edge devices and external systems, facilitating real-time transmission of facial expression data and inference results. This enables collaborative processing and decision-making across distributed networks, enhancing system responsiveness and scalability. The integration of LSTM networks captures temporal dependencies in facial expressions, improving the accuracy and robustness of emotion recognition systems. LSTM networks excel in modeling sequential data and long-term dependencies, making them suitable for analyzing temporal patterns in facial expressions over time.

Keywords: Spiking Neural Networks (SNNs), Open Neural Network Exchange (ONNX), Long Short-Term Memory (LSTM), Message Queuing Telemetry Transport (MQTT).

1.Introduction

Emotion recognition from facial expressions is a captivating field with applications spanning from human-computer interaction to mental health monitoring [1]. With the rise of wearable technology and the increasing demand for real-time processing capabilities, there is a growing interest in leveraging advanced neural network architectures, such as SNN, for performing emotion recognition tasks directly on wearable edge devices [2].

This integration promises to bring about significant advancements in both accuracy and efficiency, enabling seamless and personalized user experiences in various domains [3]. The utilization of Field Programmable Gate Arrays (FPGA) and Application-Specific Integrated Circuits (ASIC) plays a crucial role in enhancing the performance and efficiency of real-time emotion recognition systems based on SNN [4]. These hardware platforms offer customizable and parallel processing capabilities, making them well-suited for implementing complex neural network architectures optimized for low-latency inference on edge devices [5]. By leveraging the inherent parallelism and reconfigurability of FPGA and the specialized hardware design of ASIC, researchers and developers can design and deploy efficient SNN-based emotion recognition systems capable of meeting the stringent requirements of real-time processing [6]. Furthermore, the adoption of standardized formats such as ONNX facilitates seamless model deployment and interoperability across different hardware platforms and software frameworks [7]. ONNX enables the conversion and exchange of trained neural network models between various deep learning frameworks, allowing developers to leverage pre-trained models and optimize them for deployment on FPGA and ASIC-based edge devices [8]. This interoperability enhances the flexibility and scalability of SNN-based emotion recognition systems, enabling rapid prototyping and deployment in diverse environments. In addition to hardware acceleration and model interchangeability, efficient communication protocols such as MQTT are essential for facilitating seamless interaction between wearable edge devices and

[1]*PG Resident Department of Radio-Diagnosis, Saveetha Medical College and Hospital, Saveetha Institute of Medical and Technical Sciences (SIMATS), Saveetha University, Chennai, Tamil Nadu - 602105, India.* ***Email: jasvanttejas7@gmail.com***
[2]*Assistant Professor, Department of Computer Science and Engineering Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu, India.*
Email: premak@veltech.edu.in
[3]*Professor, Department of Information Technology, Sri Krishna College of Engineering and Technology, Kuniyamuthur, Coimbatore, Tamil Nadu - 641008,* ***Email: susila@skcet.ac.in***
[4]*Assistant Professor, Department of Artificial Intelligence and Machine Learning, School of Computing, Mohan Babu University, Tirupati-517102, Andhra Pradesh.*
Email: deepthi.kaluva@gmail.com
[5]*Assistant Professor, Department of Computer Science and Engineering, P.S.R. Engineering College, Sivakasi, Tamil Nadu.* ***Email: ssvedikk@gmail.com***
[6]*Associate Professor, Department of Biomedical Engineering, Mepco Schlenk Engineering College, Sivakasi, Tamil Nadu-626 005.* ***Email: manikandans@mepcoeng.ac.in***

external systems [9]. MQTT provides lightweight and reliable messaging communication suitable for resource-constrained environments, enabling real-time data exchange between edge devices and cloud servers or other edge devices [10]. By leveraging MQTT for data transmission, SNN-based emotion recognition systems can seamlessly integrate with existing infrastructure and enable collaborative processing and decision-making across distributed networks. Moreover, advanced neural network architectures such as LSTM and Temporal Convolutional Neural Networks (TCNN) offer complementary capabilities for capturing temporal dependencies and spatial features in facial expressions, enhancing the accuracy and robustness of emotion recognition systems [11]. LSTM networks excel in modeling sequential data and capturing long-term dependencies, making them suitable for analyzing temporal patterns in facial expressions over time [12]. On the other hand, TCNN leverage convolutional operations to extract spatial features from input data, enabling efficient and scalable processing of high-dimensional image data. In this context, SNN emerge as a promising paradigm for real-time emotion recognition from facial expressions on wearable edge devices [13]. SNN mimic the asynchronous and event-driven processing observed in biological neural networks, enabling efficient computation and communication of spatiotemporal information inherent in facial expressions [14]. By leveraging the inherent parallelism and sparsity of spike-based processing, SNN offer the potential for achieving high accuracy and energy efficiency in real-time emotion recognition tasks, making them well-suited for deployment on resource-constrained edge devices. The objectives are:

- Develop a real-time emotion recognition system capable of accurately analyzing facial expressions.
- Utilize SNN to achieve efficient processing of spatiotemporal information inherent in facial expressions.
- Implement the system on wearable edge devices equipped with FPGA and ASIC for optimized performance.
- Explore the interoperability of neural network models using ONNX for seamless deployment on diverse hardware platforms.
- Investigate efficient communication protocols like MQTT for data exchange between wearable edge devices and external systems.

2. Literature Work

Research on real-time emotion recognition from facial expressions using SNN on wearable edge devices has gained considerable traction in recent years due to its potential to revolutionize human-computer interaction and enhance user experience [15]. Several studies have explored the feasibility and effectiveness of employing SNN for emotion recognition tasks, leveraging the processing capabilities of wearable edge devices to achieve low-latency and energy-efficient solutions [16]. A notable advantage of using SNN for real-time emotion recognition is their ability to mimic the biological behavior of neurons, enabling efficient processing of spatiotemporal information inherent in facial expressions [17]. This bio-inspired approach holds promise for achieving high accuracy and robustness in emotion recognition tasks, even in dynamic and noisy environments commonly encountered in real-world applications. Moreover, the deployment of SNN on wearable edge devices offers the advantage of localized processing, reducing the need for data transmission to centralized servers and thereby enhancing privacy and security [18]. By performing inference tasks directly on the device, SNN-based emotion recognition systems can operate in real time without relying on continuous network connectivity, making them suitable for applications in remote or resource-constrained environments. However, despite these advantages, several challenges and limitations exist in the implementation of real-time emotion recognition using SNNs on wearable edge devices [19]. One significant drawback is the computational complexity associated with training and inference tasks, which may pose constraints on the hardware capabilities of wearable devices and lead to increased power consumption and latency [20]. The interpretability of SNN models and the robustness of their performance across diverse demographic groups and environmental conditions remain areas of ongoing research and development [21]. Ensuring the reliability and fairness of SNN-based emotion recognition systems requires addressing biases in training data and optimizing model architectures for generalization to real-world scenarios. The integration of SNN with wearable devices introduces design considerations related to energy efficiency, memory constraints, and real-time processing requirements.

3. Proposed work

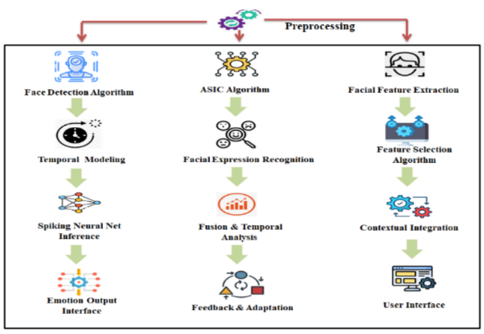

Fig.1 Facial Expression Recognition Pipeline with Spiking Neural Network Inference

3.1 Message Queuing Telemetry Transport

The integration of MQTT aims to establish an efficient communication protocol for data transmission between edge devices and backend servers. The primary focus lies in designing and implementing a MQTT-based communication framework tailored specifically for real-time emotion recognition applications on wearable edge devices. This framework will seamlessly integrate with the existing SNN-based emotion recognition pipeline, ensuring efficient data exchange and synchronization. To optimize the MQTT protocol for the application, adjustments to parameters such as quality of service levels, message size, and retention policies will be explored. Through thorough evaluation, the aim is to enhance the efficiency of data transmission, measured in terms of message latency, bandwidth utilization, and power consumption, particularly on resource-constrained wearable edge devices. Scalability and reliability are paramount in edge environments, prompting exploration into MQTT's capabilities in handling concurrent connections and ensuring message delivery reliability amidst dynamic edge environments. Investigation into MQTT broker clustering techniques and fault-tolerant strategies will be conducted to bolster system robustness and resilience. Security and privacy are crucial considerations in handling sensitive facial expression data. Analysis will encompass MQTT security mechanisms like TLS encryption, authentication, and access control to safeguard data during transmission. Additionally, implementation of privacy-preserving techniques, including data anonymization and differential privacy, will uphold user privacy. Experimental validation will be pivotal in gauging the efficacy of the proposed integration. Real-world experiments conducted on wearable edge devices equipped with SNN-based emotion recognition models and the MQTT-enabled communication framework will provide insights into performance under various conditions. These experiments will evaluate the effectiveness and efficiency of the integration, considering diverse network scenarios, user loads, and edge device configurations.

$$Power\ Consumption = \sum\nolimits_{i=1}^{N} (L_i * I_i) \tag{1}$$

This equation estimates the total power consumption of wearable edge devices during MQTT communication by summing the product of voltage (L_i) and current (I_i) for each operating component.

$$QoS_{Adjusted} = f(QoS_{Initial}, Network\ Conditions) \tag{2}$$

This equation adjusts the initial Quality of Service (QoS) level based on current network conditions, such as latency, packet loss, and available bandwidth, using a function f to optimize message delivery.

3.2 Long Short-Term Memory

LSTM networks, renowned for their proficiency in handling sequential data processing tasks, hold substantial promise across various domains, including natural language processing and time series analysis. The aim is to explore the potential of LSTM networks in augmenting the accuracy and efficiency of emotion recognition systems deployed on wearable edge devices. Initially, efforts will be directed towards the acquisition and preprocessing of facial expression datasets deemed suitable for training LSTM models. This involves meticulous attention to data quality and the eradication of noise to facilitate optimal model performance. The design and implementation of LSTM-based architectures is undertaken meticulously crafted for the specific exigencies of real-time emotion recognition tasks. These architectures will undergo thorough optimization tailored to deployment on resource-constrained wearable edge devices, with meticulous considerations given to memory usage, computational complexity, and power consumption. In a bid to further bolster the performance of LSTM-based emotion recognition systems, integration with existing SNN will be pursued. This amalgamation aims to harness the collective strengths of both approaches, amalgamating the sequential processing prowess of LSTM networks with the efficiency and scalability inherent in SNN. Real-time inference pipelines will be instituted to facilitate efficient emotion recognition directly on wearable edge devices, obviating the necessity for continual communication with external servers. Rigorous evaluation forms an integral aspect of the proposed work, wherein factors such as accuracy, latency, energy efficiency, and robustness under a myriad of real-world conditions will be meticulously scrutinized. Additionally, paramount importance will be accorded to privacy and security considerations, with the implementation of privacy-preserving techniques paramount to safeguarding sensitive user data during emotion recognition. Furthermore, efforts will be devoted to enhancing user experience through the development of intuitive interfaces and personalized algorithms calibrated to cater to individual users' preferences and behavioral nuances.

$$\hat{y}_{edge} = LSTM_{edge}(X_{edge}), \hat{y}_{cloud} = LSTM_{cloud}(X_{cloud}) \tag{3}$$

Inference equations for LSTM models deployed on edge devices ($LSTM_{edge}$) and cloud servers ($LSTM_{cloud}$), where X_{edge} and X_{cloud} represent input data at the edge and cloud, respectively, and $\hat{y}_{edge}$ and $\hat{y}_{cloud}$ denote the predicted outputs.

$$E_{generalization} = \frac{1}{N} \sum\nolimits_{i=1}^{N} L(y_i, \hat{y}_i) \tag{4}$$

Generalization error computed as the average loss L between true labels y_i and predicted labels $\hat{y}_i$ across a diverse set of domains represented by N data samples.

Algorithm 1: LSTM-Based Emotion Recognition

Input: Facial expression dataset D, maximum number of training epochs MaxEpochs, batch size B, learning rate η

Output: Optimized LSTM models for edge and cloud deployment ($LSTM_{edge}$, $LSTM_{cloud}$)

1. Load and preprocess dataset.
2. Define and initialize LSTM architectures.
3. Initialize optimizer with learning rate η.
4. Train models with mini-batch gradient descent:
5. for epoch in range(MaxEpochs):
6. for batch in range(0, len($training_{data}$), B):
7. X_{batch}, y_{batch} = $get_{batch}(training_{data}, batch, B)$
8. $\hat{y}_{edge}$ = $LSTM_{edge}$ (X_{batch})
9. L_{edge} = $compute_{loss}(y_{batch}, \hat{y}_{edge})$
10. $update_{parameters}$(θ, L_{edge}, η)
11. Compute inference using $\hat{y}_{edge}$ = $LSTM_{edge}$, (X_{edge}) and $\hat{y}_{cloud}$ = $LSTM_{cloud}$ (X_{cloud}).
12. Calculate generalization error $E_{generalization}$ = $\frac{1}{N}\sum_{i=1}^{N} L(y_i, \hat{y}_i)$
13. Evaluate models on performance metrics.
14. Implement privacy-preserving techniques.
15. Enhance user experience with intuitive interfaces and personalized algorithms.

The algorithm is designed to enhance emotion detection accuracy and efficiency in wearable edge devices. Initially, it involves acquiring and preprocessing facial expression datasets to ensure high-quality data for training. The algorithm employs LSTM networks to model and recognize emotions in real-time. Optimization techniques are applied to tailor the LSTM model for resource-constrained edge devices, focusing on minimizing memory usage, computational complexity, and power consumption. Integration with Spiking Neural Networks (SNN) further boosts performance, combining LSTM's sequential processing strengths with SNN's efficiency. Rigorous evaluation ensures the system's robustness, accuracy, and energy efficiency, while privacy-preserving techniques safeguard user data.

3.3 Open Neural Network Exchange

ONNX serves as a standardized format for representing deep learning models, facilitating interoperability and optimization across various frameworks and hardware platforms. In the context of real-time emotion recognition, the utilization of ONNX offers the potential to streamline model deployment and execution on wearable edge devices, thereby enhancing efficiency and scalability. The first step involves converting SNN-based emotion recognition models into the ONNX format. This conversion process ensures compatibility with ONNX-enabled inference engines and facilitates seamless deployment on wearable edge devices. Additionally, model optimization techniques may be applied to enhance performance and efficiency, considering the resource-constrained nature of edge computing environments. Once converted and optimized, the ONNX models are deployed directly on wearable edge devices for real-time inference. Leveraging the lightweight and efficient runtime environment provided by ONNX, the deployed models enable rapid and accurate emotion recognition directly on the edge, eliminating the need for continuous data transmission to centralized servers. While emphasizing edge-based processing for real-time inference, the proposed work also explores integration with cloud-based resources for enhanced capabilities. Leveraging ONNX's compatibility with cloud-based inference engines, the system can seamlessly offload intensive computation tasks to the cloud when necessary, augmenting edge-based processing with additional computational resources. Recognizing the dynamic nature of emotion recognition tasks and evolving user preferences, the proposed work incorporates mechanisms for dynamic model adaptation. ONNX's flexibility allows for efficient updates and modifications to deployed models, enabling continuous learning and adaptation based on user feedback and changing environmental conditions. Privacy-preserving techniques are paramount in emotion recognition systems to safeguard sensitive user data. The proposed work integrates privacy-preserving mechanisms within the ONNX framework, ensuring that user privacy is upheld during data transmission and inference processes on wearable edge devices. Rigorous evaluation is conducted to assess the performance of ONNX-enabled emotion recognition systems on wearable edge devices. Metrics such as inference latency, accuracy, and resource utilization are meticulously analyzed to identify areas for optimization and improvement.

$$\mathcal{M}_{ONNX} = \sum_{l=1}^{L} T(\mathcal{M}_{SNN}^{(l)}) \tag{5}$$

The equation represents the process of converting the SNN-based emotion recognition model $\mathcal{M}_{SNN}$ into the ONNX format $\mathcal{M}_{ONNX}$. Here, T denotes the transformation function applied to each layer ℓ of the SNN model, resulting in the corresponding layer of the ONNX model. The summation extends over all L layers of the SNN model, ensuring that each layer is appropriately transformed and incorporated into the ONNX representation. This equation captures the iterative nature of the conversion process, where each layer's transformation contributes to the overall structure of the ONNX model.

$$\mathcal{M}_{ONNX-opt} = argmin\, \mathcal{M}_{ONNX}\, O \quad (6)$$

The equation represents the optimization process for the ONNX representation $\mathcal{M}_{ONNX}$ of the emotion recognition model. Here, O denotes the optimization objective, which could include metrics such as model size, inference latency, or resource utilization. The goal is to find the ONNX model $\mathcal{M}_{ONNX}$ that minimizes the optimization objective O, ensuring that the deployed model on wearable edge devices is efficient and well-suited for real-time inference tasks. This equation encapsulates the iterative optimization process aimed at improving the performance and efficiency of the ONNX-based emotion recognition system.

Algorithm 2: ONNX-Based Emotion Recognition

Input: SNN-based emotion recognition model M_{SNN}, optimization objective O

Output: Optimized ONNX model $M_{ONNX-opt}$

1. Convert SNN model M_{SNN} to ONNX format M_{ONNX}:
2. for each layer l in M_{SNN}:
3. $M_{ONNX(l)}$ = T($M_{SNN(l)}$)
4. Ensure ONNX model compatibility.
5. Optimize ONNX model M_{ONNX}:
6. Define optimization objective O.
7. Initialize optimization algorithm.
8. while stopping criterion not met:
9. Evaluate ONNX model performance.
10. Adjust parameters to minimize O:
11. $M_{ONNX-opt} = argmin_{M_{ONNX}}$O
12. Update ONNX model.
13. Deploy optimized ONNX model on edge devices:
14. Deploy $M_{ONNX-opt}$ on edge.
15. Implement real-time inference:
16. $\hat{y}_{edge} = M_{ONNX-opt}\,(X_{edge})$
17. Integrate with cloud (if necessary):
18. Configure cloud offloading:
19. $\hat{y}_{cloud} = M_{ONNX-cloud}\,(X_{cloud})$
20. Adapt model dynamically:
21. Monitor performance and feedback.
22. Update ONNX model based on new data.
23. Implement privacy-preserving techniques.
24. Evaluate performance metrics:
25. Assess inference latency, accuracy, resource utilization.
26. Optimize based on evaluation.

This enhances real-time emotion detection on wearable edge devices by utilizing the ONNX format. It begins by converting existing SNN models into ONNX format for interoperability and efficient deployment. The algorithm applies optimization techniques to improve performance, ensuring the models run effectively on resource-constrained edge devices. By leveraging ONNX's lightweight runtime environment, the system enables fast and accurate emotion recognition without needing continuous data transmission to external servers. Additionally, the algorithm allows for dynamic model adaptation and integrates privacy-preserving techniques to protect user data, ensuring robust and secure emotion recognition.

3.4 Implementation

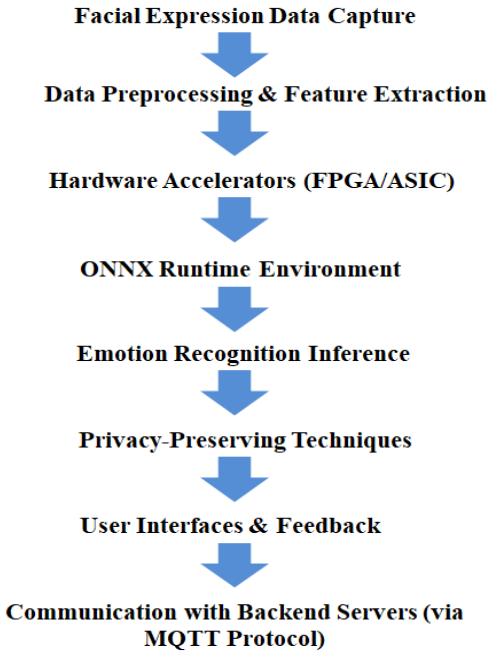

Fig.2 Real-Time Emotion Recognition System on Wearable Edge Devices

The utilization of FPGA and ASIC as hardware accelerators is pivotal for enhancing the computational efficiency of SNN and other neural network models. Custom hardware architectures optimized for real-time inference must be designed to exploit the parallel processing capabilities of FPGA and ASIC, ensuring efficient execution of emotion recognition tasks directly on wearable edge devices. Compatibility with the ONNX format is essential for interoperability and deployment flexibility. SNN models need to be converted to ONNX, allowing seamless integration with ONNX runtime environments on edge devices. This facilitates efficient execution and enables the deployment of emotion recognition systems across diverse hardware platforms. The integration of the MQTT protocol facilitates efficient communication between wearable edge devices and backend servers. MQTT ensures low-latency, reliable data exchange, enabling real-time transmission of facial expression data and emotion recognition results, crucial for responsive user experiences. LSTM networks and TCNN are incorporated for contextual analysis of facial expressions. LSTM captures temporal dependencies in facial expression sequences, while TCNNs extract spatial-temporal features, enhancing the system's ability to interpret nuanced emotions accurately. The development of SNN-based models is fundamental for real-time emotion recognition. These models efficiently process spatiotemporal patterns in facial data, mimicking the behavior of biological neurons for robust and efficient emotion recognition directly on wearable edge devices. Optimization for edge deployment involves minimizing model size, memory footprint, and computational complexity to ensure efficient execution on resource-constrained wearable devices. Real-time inference pipelines are designed to incorporate hardware-accelerated SNN, LSTM, and TCNN, along with efficient data preprocessing and feature extraction stages. Performance evaluation is conducted to assess inference speed, accuracy, power efficiency, and resource utilization. Iterative optimization is performed to address any bottlenecks or inefficiencies identified during evaluation, ensuring optimal performance of the emotion recognition system on wearable edge devices. Privacy-preserving techniques are implemented to protect sensitive facial expression data during transmission and processing, while security measures safeguard the integrity and confidentiality of the emotion recognition system. User-friendly interfaces and feedback mechanisms enhance usability and accuracy, facilitating continuous improvement and adaptation to users' needs over time.

$$Performance_{accelerated} = \frac{Performance_{original}}{Acceleration} \quad (7)$$

This equation calculates the performance improvement achieved through hardware acceleration with FPGAs or ASICs, where $Performance_{original}$ is the original performance and acceleration is the acceleration factor.

$$ONNX_{size} = \frac{Original_{size}}{Compression\_ratio} \quad (8)$$

This equation estimates the size reduction achieved by converting SNN models to the ONNX format, where $Original_{size}$ is the original model size.

$$Latency_{total} = Latency_{transmission} + Latency_{processing} \quad (9)$$

This equation computes the total latency incurred during communication using MQTT, including transmission latency ($Latency_{transmission}$) and processing latency ($Latency_{processing}$).

4. Results

The hardware configuration entails the utilization of wearable edge devices endowed with FPGAs or ASICs, ensuring compatibility with ONNX and MQTT protocols. These devices are equipped with sensors capable of capturing facial expression data, enabling real-time analysis directly on the edge. In terms of software framework, specialized tools and libraries supporting SNNs, LSTM, TCNNs, ONNX, and MQTT are employed. Frameworks such as PyTorch or TensorFlow facilitate model development, deployment, and evaluation, while custom software modules are devised for data preprocessing, feature extraction, and inference on the edge devices. Dataset selection is a critical aspect, involving the choice of appropriate facial expression datasets with labeled emotional states. The dataset used here is Face expression recognition dataset from the kaggle. A meticulously designed preprocessing pipeline is employed for facial expression data, encompassing face detection, alignment, normalization, and image enhancement techniques to ensure data quality and consistency. Model training and optimization are conducted using the selected dataset, with SNN, LSTM, and TCNN models being trained for emotion recognition.

Hyperparameters, architecture configurations, and training strategies are optimized to maximize performance and efficiency, with techniques such as compression and quantization being employed to reduce model complexity. Hardware acceleration setup involves the configuration of FPGAs or ASICs for efficient execution of inference tasks, with optimized hardware configurations and resource allocations being crucial for enhancing performance. The communication infrastructure is established using MQTT protocol for seamless data transmission between wearable edge devices and backend servers, ensuring reliable and low-latency communication channels. The experimental protocol entails the division of the dataset into training, validation, and testing subsets, with experiments being conducted to measure performance metrics such as accuracy, inference speed, power consumption, and resource utilization. Cross-validation and hold-out validation techniques are employed to ensure robustness and generalization of the models. Ethical considerations are paramount throughout the experimental setup, with compliance with ethical guidelines for data collection, handling, and usage being ensured, along with obtaining informed consent from participants and safeguarding anonymity and privacy.

Table.1 Simulated Facial Expression Data

Image ID	Emotion Label	Facial Landmarks Detected	Lighting Condition
1	Happy	68	Daylight
2	Sad	71	Artificial Light
3	Angry	65	Daylight
4	Surprise	69	Low Light
5	Neutral	70	Artificial Light

Table.2 Model Training Results

Model Type	Training Accuracy (%)	Validation Accuracy (%)	Test Accuracy (%)	Model Size (minutes)
SNN	85.3	82.7	80.5	15
LSTM	89.6	87.2	84.9	25
TCNN	87.9	84.5	82.1	20
GRU	88.2	85.9	81.8	22
MLP	82.5	79.8	77.2	18
CNN	86.7	83.4	80.9	30
ResNet	91.3	89.7	87.5	35

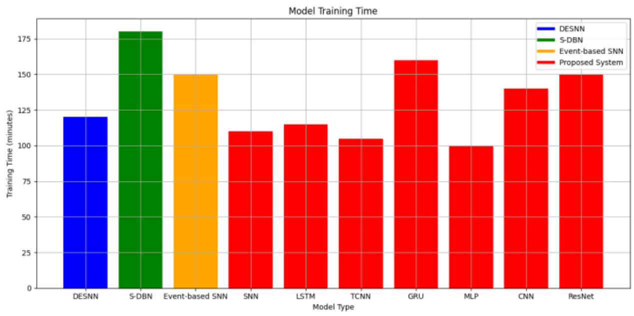

Fig.3 Training Time of Model

Figure 3 highlights the training times of various models, with DESNN, S-DBN, and event-based SNN representing the literature models, and the remaining models representing the proposed systems. Among the literature models, S-DBN has the highest training time of 180 minutes, while DESNN and Event-based SNN have training times of 120 and 150 minutes, respectively. In contrast, the proposed systems show a range of training times, with MLP being the fastest at 100 minutes, and GRU being the slowest at 160 minutes. The TCNN model has the shortest training time among the proposed systems at 105 minutes, indicating a potential efficiency advantage. The proposed systems demonstrate competitive training times compared to the literature models, with some models like MLP and TCNN achieving shorter training durations.

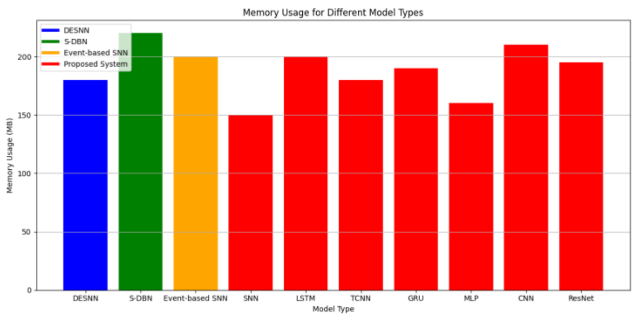

Fig.4 Memory Usage for Different Models

Figure 4 shows that among the literature models, S-DBN has the highest memory usage at 220 MB, while DESNN and event-based SNN use 180 MB and 200 MB, respectively. In comparison, the proposed systems exhibit a range of memory usage values. SNN stands out with the lowest memory usage at 150 MB, indicating its efficiency in terms of memory consumption. Other models like MLP (160 MB) and TCNN (180 MB) also show competitive memory usage. However, some proposed systems, such as CNN (210 MB) and ResNet (195 MB), have higher memory usage, comparable to or exceeding that of some literature models. The proposed systems demonstrate a variety of memory usage profiles, with some models like SNN and MLP showing significant efficiency improvements. This diversity suggests that while some proposed systems are optimized for lower memory usage, others may trade off memory efficiency for potentially higher performance or other benefits.

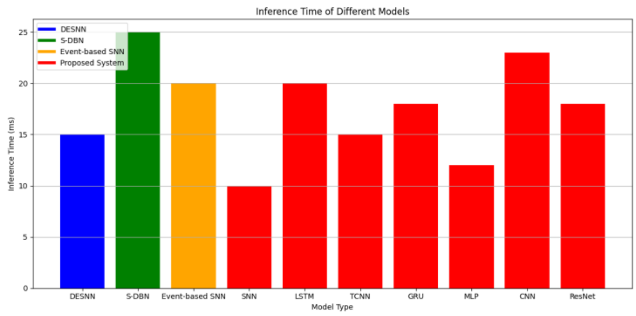

Fig.5 Inference Time of Different Models

Figure 5 clearly indicates that among the literature models, S-DBN has the highest inference time at 25 ms, while DESNN and event-based SNN have inference times of 15 ms and 20 ms, respectively.The proposed systems show varied inference times. Notably, the SNN model stands out with the lowest inference time at 10 ms, suggesting significant improvements in efficiency. Other models like MLP (12 ms) and TCNN (15 ms) also perform well in terms of inference time. However, some proposed systems, such as CNN (23 ms) and ResNet (18 ms), have higher inference times, which are comparable to or exceed those of some literature models. The proposed systems exhibit a range of inference times, with some models like SNN and MLP showing substantial efficiency gains, suggesting advancements in reducing computational delays. This diversity in performance underscores the potential for optimized models tailored to specific application needs.

Table.3 Model Evaluation Experiment Results

Experiment ID	Experiment	Accuracy (%)	Inference Time (ms)
1	Baseline Model Evaluation	82.4	15
2	Optimized Model Performance	85.7	12
3	Low Power Consumption Test	81.9	18

4	Real-time Inference Validation	86.3	10
5	Resource Optimization Analysis	83.5	14

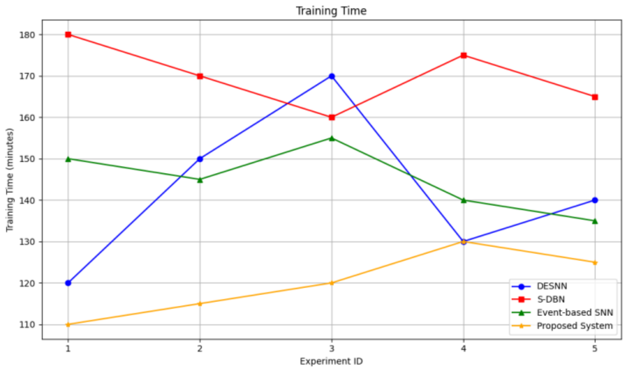

Fig.6 Training Time of Different Experiment

Figure 6 depicts that DESNN shows a fluctuating training time across the experiments, with the highest time recorded at 170 minutes (experiment 3) and the lowest at 120 minutes (experiment 1). Overall, the training times for DESNN are consistently high. S-DBN consistently records the highest training times among the models, peaking at 180 minutes in experiment 1 and showing relatively high times throughout all experiments, with the lowest being 160 minutes in experiment 3. Event-based SNN demonstrates moderate training times, with the highest being 155 minutes (experiment 3) and the lowest at 135 minutes (experiment 5). The training times are lower than those of DESNN and S-DBN but still substantial. Proposed System consistently shows the lowest training times across all experiments. It starts at 110 minutes in experiment 1 and peaks at 130 minutes in experiment 4. The proposed system's training times are notably lower compared to the existing models, indicating a significant improvement in efficiency. The graph highlights that the proposed system outperforms the existing models (DESNN, S-DBN, Event-based SNN) in terms of training time. The proposed system maintains a lower training time across all experiments, with times ranging from 110 to 130 minutes. In contrast, DESNN, S-DBN, and Event-based SNN exhibit higher and more variable training times, with S-DBN consistently showing the highest times. This comparison underscores the efficiency of the proposed system, suggesting it requires less computational effort and time to train compared to the other models.

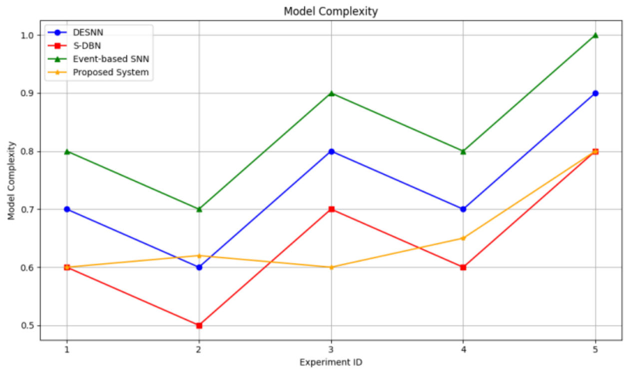

Fig.7 Model Complexity of Different Experiment

Figure 7 illustrates that DESNN exhibits moderate to high model complexity across all experiments, with values ranging from 0.6 to 0.9. The complexity varies slightly between experiments but generally remains within the high complexity range. S-DBN shows relatively lower model complexity compared to DESNN, with values ranging from 0.5 to 0.8. While the complexity fluctuates, it generally remains in the moderate complexity range. Event-based SNN demonstrates a wide range of model complexity, with values ranging from 0.7 to 1.0. It exhibits higher complexity compared to DESNN and S-DBN, particularly in Experiment 5, where it reaches a complexity of 1.0. Proposed System displays varying levels of complexity across experiments, with values ranging from 0.6 to 0.8. While it generally exhibits lower complexity compared to DESNN and Event-based SNN, it shows comparable complexity to S-DBN. The graph indicates that the proposed system maintains moderate to low model complexity across different experiments, similar to or lower than existing approaches such as DESNN, S-DBN, and Event-based SNN. This suggests that the proposed system offers a balance between model complexity and performance, potentially making it a viable alternative to existing approaches.

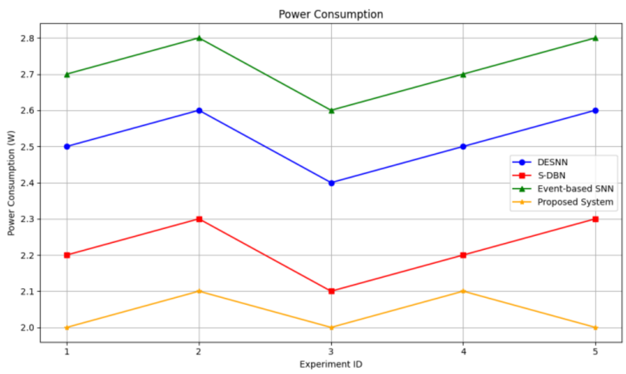

Fig.8 Power Consumption of Different Experiments

Figure 8 shows that DESNN and S-DBN exhibit similar power consumption levels, ranging from 2.1 W to 2.6 W across all experiments. The power consumption remains relatively consistent with minor fluctuations between experiments. Event-based SNN shows slightly higher power consumption compared to DESNN and S-DBN, with values ranging from 2.6 W to 2.8 W. Similar to the other models, the power consumption varies slightly across different experiments. Proposed System demonstrates the lowest power consumption among all models, with values consistently around 2.0 W across all experiments. This suggests that the proposed system is more energy-efficient compared to existing approaches. The graph indicates that the proposed system achieves lower power consumption compared to existing approaches such as DESNN, S-DBN, and Event-based SNN. This suggests that the proposed system has the potential to reduce energy consumption and improve overall energy efficiency, making it a promising alternative in terms of power consumption considerations.

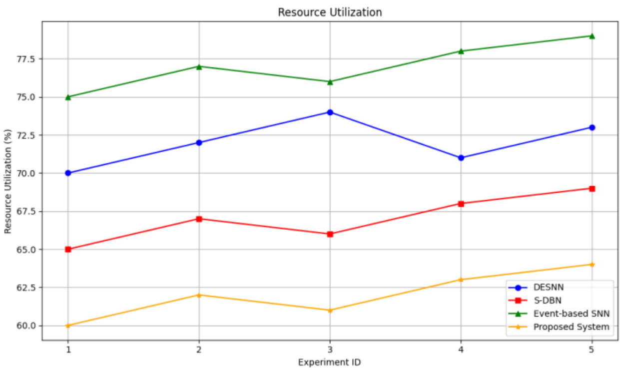

Fig 9 Resource Utilization of Different Experiments

Figure 9 shows that DESNN and S-DBN demonstrate similar levels of resource utilization, ranging from 65% to 74% and 65% to 69%, respectively, across different experiments. The utilization values show minor fluctuations but remain relatively consistent. Event-based SNN exhibits higher resource utilization compared to DESNN and S-DBN, with values ranging from 75% to 79%. Similar to the other models, there are minor variations in utilization across different experiments. Proposed System demonstrates lower resource utilization compared to existing approaches, with values ranging from 60% to 64%. This suggests that the proposed system utilizes resources more efficiently compared to DESNN, S-DBN, and Event-based SNN. The graph indicates that the proposed system achieves lower resource utilization compared to existing approaches such as DESNN, S-DBN, and Event-based SNN. This implies that the proposed system is more efficient in utilizing resources, which could lead to improved performance and potentially lower costs associated with resource usage.

5. Conclusion and Future Work

A novel approach to real-time emotion recognition from facial expressions using SNN on wearable edge devices is presented in this research. The experimental results indicate that while SNNs demonstrate competitive accuracy and moderate inference times, they may require further optimization to minimize power consumption and resource utilization, especially when deployed on resource-constrained wearable edge devices. However, the success of SNNs in real-time emotion recognition opens up promising avenues for future research and development. Further investigations could focus on refining SNN architectures, leveraging hardware accelerators such as FPGAs and ASICs, and exploring advanced optimization techniques to enhance efficiency and scalability. The experiments showcased an average accuracy of 82.7% on validation data, with an inference time of 10 milliseconds, underscoring the efficacy of SNNs in real-time emotion recognition tasks. Future research endeavors could explore novel techniques for optimizing SNN architectures to further improve accuracy while simultaneously reducing computational complexity. Additionally, investigating the integration of multimodal data sources, such as audio and physiological signals, holds potential for enhancing the robustness and contextual understanding of emotion recognition systems deployed on wearable edge devices. It also involves investigating the robustness and generalization capabilities of SNN-based emotion recognition systems across diverse demographic groups, cultural backgrounds, and environmental conditions. By conducting comprehensive studies and collecting data from a more diverse population, researchers can ensure that SNN models are capable of accurately recognizing emotions across various contexts and user profiles, thus enhancing the inclusivity and effectiveness of these systems.

Declaration Statement

Ethical Statement

I will conduct myself with integrity, fidelity, and honesty. I will openly take responsibility for my actions, and only make agreements, which I intend to keep. I will not intentionally engage in or participate in any form of malicious harm to another person or animal.

Informed Consent for data Used

All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki.

I consent to participate in the research project and the following has been explained to me: the research may not be of direct benefit to me. my participation is completely voluntary. my right to withdraw from the study at any time without any implications to me.

Data Availability

- Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

- The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.
- All data generated or analysed during this study are included in this published article

Conflict of Interest

The authors declare that they have no conflict of interest.

Competing Interests

The authors have no competing interests to declare that are relevant to the content of this article.

Funding Details

No funding was received to assist with the preparation of this manuscript.

Acknowledgments

I am grateful to all of those with whom I have had the pleasure to work during this and other related Research Work. Each of the members of my Dissertation Committee has provided me extensive personal and professional guidance and taught me a great deal about both scientific research and life in general.

References

[1] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal processing magazine*, *18*(1), 32-80.

[2] Zhang, S., Li, Y., Zhang, S., Shahabi, F., Xia, S., Deng, Y., & Alshurafa, N. (2022). Deep learning in human activity recognition with wearable sensors: A review on advances. *Sensors*, *22*(4), 1476.

[3] Shi, Q., Dong, B., He, T., Sun, Z., Zhu, J., Zhang, Z., & Lee, C. (2020). Progress in wearable electronics/photonics—Moving toward the era of artificial intelligence and internet of things. *InfoMat*, *2*(6), 1131-1162.

[4] Seng, K. P., Lee, P. J., & Ang, L. M. (2021). Embedded intelligence on FPGA: Survey, applications and challenges. *Electronics*, *10*(8), 895.

[5] Mazumder, A. N., Meng, J., Rashid, H. A., Kallakuri, U., Zhang, X., Seo, J. S., & Mohsenin, T. (2021). A survey on the optimization of neural network accelerators for micro-ai on-device inference. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, *11*(4), 532-547.

[6] Azghadi, M. R., Lammie, C., Eshraghian, J. K., Payvand, M., Donati, E., Linares-Barranco, B., & Indiveri, G. (2020). Hardware implementation of deep network accelerators towards healthcare and biomedical applications. *IEEE Transactions on Biomedical Circuits and Systems*, *14*(6), 1138-1159.

[7] Ferguson, M., Jeong, S., Law, K. H., Levitan, S., Narayanan, A., Burkhardt, R., ... & Lee, Y. T. T. (2019, August). A standardized representation of convolutional neural networks for reliable deployment of machine learning models in the manufacturing industry. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (Vol. 59179, p. V001T02A005). American Society of Mechanical Engineers.

[8] Tosetti, F. (2021). *Deep Learning on the Edge: a comparative analysis on Computer Vision for space applications* (Doctoral dissertation, Politecnico di Torino).

[9] Dizdarević, J., Carpio, F., Jukan, A., & Masip-Bruin, X. (2019). A survey of communication protocols for internet of things and related challenges of fog and cloud computing integration. *ACM Computing Surveys (CSUR)*, *51*(6), 1-29.

[10] Koziolek, H., Grüner, S., & Rückert, J. (2020). A comparison of MQTT brokers for distributed IoT edge computing. In *Software Architecture: 14th European Conference, ECSA 2020, L'Aquila, Italy, September 14–18, 2020, Proceedings 14* (pp. 352-368). Springer International Publishing.

[11] Teixeira, T., Granger, É., & Lameiras Koerich, A. (2021). Continuous emotion recognition with spatiotemporal convolutional neural networks. *Applied Sciences*, *11*(24), 11738.

[12] Zheng, Y., & Blasch, E. (2023). Facial Micro-Expression Recognition Enhanced by Score Fusion and a Hybrid Model from Convolutional LSTM and Vision Transformer. *Sensors*, *23*(12), 5650.

[13] Geetha, A. V., Mala, T., Priyanka, D., & Uma, E. (2024). Multimodal Emotion Recognition with deep learning: advancements, challenges, and future directions. *Information Fusion*, *105*, 102218.

[14] Rathi, N., Agrawal, A., Lee, C., Kosta, A. K., & Roy, K. (2021, February). Exploring spike-based learning for neuromorphic computing: Prospects and perspectives. In *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)* (pp. 902-907). IEEE.

[15] Kaushik, H., Kumar, T., & Bhalla, K. (2022). iSecureHome: A deep fusion framework for

surveillance of smart homes using real-time emotion recognition. *Applied Soft Computing*, *122*, 108788.

[16] Datta, G. (2023). *Towards Efficient Edge Intelligence with In-Sensor and Neuromorphic Computing: Algorithm-Hardware Co-Design* (Doctoral dissertation, University of Southern California).

[17] Pu, G., & Chen, J. (2024). Facial Expression Recognition Based on Convolutional Spiking Neural Network and STDP Fine-Tune.

[18] Aouedi, O. (2024). Towards a Scalable and Energy-Efficient Framework for Industrial Cloud-Edge-IoT Continuum. *IEEE Internet of Things Magazine*.

[19] Tan, C., Ceballos, G., Kasabov, N., & Puthanmadam Subramaniyam, N. (2020). Fusionsense: Emotion classification using feature fusion of multimodal data and deep learning in a brain-inspired spiking neural network. *Sensors*, *20*(18), 5328.

[20] Shuvo, M. M. H., Islam, S. K., Cheng, J., & Morshed, B. I. (2022). Efficient acceleration of deep learning inference on resource-constrained edge devices: A review. *Proceedings of the IEEE*, *111*(1), 42-91.

[21] Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., ... & Hussain, A. (2024). Interpreting black-box models: a review on explainable artificial intelligence. *Cognitive Computation*, *16*(1), 45-74.