

An Interpretable KAN-Guided SNN Architecture with Channel-Wise Contribution Tracking for EEG-Based Emotion Recognition

by

Abdul Kariem Bin Kamal

22201192

S.M. Hasnat Fardin

22201199

Sumaiya Khan Maisha

22201206

Saeed Rahman Samin

22301289

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
June 2026.

© 2026. Brac University
All rights reserved.

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Abdul Kariem Bin Kamal

22201192

S.M. Hasnat Fardin

22201199

Sumaiya Khan Maisha

22201206

Saeed Rahman Samin

22301289

Approval

The thesis/project titled “An Interpretable KAN-Guided SNN Architecture with Channel-Wise Contribution Tracking for EEG-Based Emotion Recognition” submitted by

1. Abdul Kariem Bin Kamal (22201192)
2. S.M. Hasnat Fardin (22201199)
3. Sumaiya Khan Maisha (22201206)
4. Saeed Rahman Samin (22301289)

of Spring, 2026 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering in June, 2026.

Examining Committee:

Supervisor:
(Member)

Dr. Md. Ashraful Alam

Associate Professor
Department of Computer Science and Engineering
BRAC University

Co-Supervisor:
(Member)

Dr. Md. Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
BRAC University

Thesis Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi

Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

EEG-based emotion recognition is challenging due to high signal-to-noise ratio, subtle difference in emotional states and mainly inter-participant variability. Previous works have shown that Spiking Neural Networks (SNN) are suitable for capturing temporal EEG activity, however, their learned representations are difficult to interpret at a channel level. This study proposes a channel-wise KAN (Kolmogorov-Arnold Networks) guided SNN for interpretable emotion recognition using EEG data from the DER-VREEG dataset. The dataset classifies emotion from four distinct emotion classes: happy, bored, calm and scared using four channels that capture EEG data from TP9, TP10, AF7 and AF8 which are four distinct regions of the human brain. Unlike traditional SNN classifiers, our proposed model can estimate the contribution factor of each of the four EEG channels towards predicting a certain emotion class, which is crucial for neuroscientists to study how a specific region of the brain reacts to a specific emotion. The framework thus provides explainable EEG emotion recognition along with classification accuracy that is on par with existing related architectures.

Keywords: Emotion recognition, EEG, SNN, KAN

Table of Contents

Declaration	i
Approval	ii
Abstract	iv
Table of Contents	v
List of Figures	viii
List of Tables	ix
Nomenclature	ix
1 Introduction	1
1.1 Background	1
1.2 Rational of the Study or Motivation	2
1.3 Problem Statement	3
1.4 Objective	4
1.5 Methodology in Brief	4
1.6 Scopes and Challenges	5
2 Literature Review	7
2.1 Preliminaries	7
2.1.1 EEG-Based Emotion Recognition	7
2.1.2 Kolmogorov–Arnold Representation Theorem	7
2.1.3 Kolmogorov–Arnold Networks	8
2.1.4 Spline-Based Function Approximation in KAN	8
2.1.5 Spiking Neural Networks	9
2.1.6 Interpretability in EEG Modelling	10
2.2 Review of Existing Research	10
2.2.1 DER-VREEG and Low-Cost VR-EEG Emotion Recognition	10
2.2.2 Deep-Learning Approaches for EEG Emotion Recognition	11
2.2.3 Channel-Aware and Graph-Based EEG Emotion Recognition	12
2.2.4 Spiking Neural Networks for EEG and Emotion-Related Recognition	13
2.2.5 Kolmogorov–Arnold Networks and Their Relevance to EEG Modelling	14
2.2.6 Research Gap	15
2.3 Summary of Key Findings	16

3	Requirements, Impacts and Constraints	18
3.1	Final Specifications and Requirements	18
3.2	Societal Impact	19
3.3	Environmental Impact	20
3.4	Ethical Issues	20
3.5	Standards	21
3.6	Project Management Plan	22
3.7	Risk Management	23
3.8	Economic Analysis	23
4	Proposed Methodology	25
4.1	Design Process or Methodology Overview	25
4.2	Preliminary Design or Design (Model) Specification	26
4.2.1	Alternative Design 1: Conventional Machine Learning	26
4.2.2	Alternative Design 2: Conventional Deep Neural Network	27
4.2.3	Alternative Design 3: Plain SNN with Linear Readout	27
4.2.4	Alternative Design 4: KAN Encoder with SNN	27
4.2.5	Selected Design: SNN Backbone with KAN Readout	27
4.2.6	Input Representation	28
4.2.7	Exponential Latency-Based Spike Encoding	28
4.2.8	SNN Backbone	29
4.2.9	Surrogate Gradient Learning	31
4.2.10	KAN Readout	31
4.2.11	Spline Parameterization of KAN Functions	33
4.2.12	Loss Function	34
4.2.13	Final Proposed Architecture	34
4.3	Data Collection	35
4.3.1	Data Cleaning	36
4.3.2	Data Transformation	36
4.3.3	Data Integration	38
4.3.4	Data Reduction	38
4.3.5	Summary of Preprocessed Data	38
4.4	Implementation of Selected Design	39
4.4.1	Step 1: Dataset Loading	39
4.4.2	Step 2: Cleaning and Label Encoding	39
4.4.3	Step 3: Normalization	39
4.4.4	Step 4: Window Generation	40
4.4.5	Step 5: Spike Encoding	40
5	Result Analysis	41
5.1	Performance Evaluation	41
5.2	Analysis of Design Solutions	45
5.3	Final Design Adjustments	45
5.4	Statistical Analysis	48
5.5	Comparisons and Relationships	51
5.6	Discussions	54

6	Conclusion	58
6.1	Summary of Findings	58
6.2	Contributions to the Field	58
6.3	Recommendations for Future Work	59
	Bibliography	62

List of Figures

1.1	High-level methodology of the proposed EEG emotion-recognition framework. The workflow begins with the DER-VREEG EEG dataset, followed by preprocessing, spike encoding, SNN-based temporal feature extraction, KAN readout, emotion prediction, and final evaluation with interpretability analysis.	5
4.1	Final proposed SNN + KAN readout architecture for interpretable EEG emotion recognition. The architecture converts normalized EEG windows into spike-based representations, processes them through an SNN backbone, and uses a KAN readout to produce channel-wise class contribution vectors before final emotion prediction.	35
5.1	Training and validation accuracy of the selected SNN + KAN readout model on DER-VREEG.	43
5.2	Training and validation loss of the selected SNN + KAN readout model on DER-VREEG.	43
5.3	Per-class precision, recall, and F1-score of the selected SNN + KAN readout model.	44
5.4	Overall performance statistics of the final SNN + KAN readout model on the DER-VREEG dataset.	49
5.5	Confusion matrix of the final SNN + KAN readout model for four-class DER-VREEG emotion classification.	50
5.6	Channel-wise KAN contribution curves toward the Scared class. The curves illustrate how TP9, AF7, AF8, and TP10 can contribute differently to the Scared prediction through the KAN readout.	53

List of Tables

3.1	Project Management Plan	22
3.2	Risk Management Plan	23
3.3	Estimated Cost Components	24
5.1	Performance comparison of evaluated model variants	42
5.2	Analysis of design solutions	46
5.3	Statistical summary of the final SNN + KAN readout model	50
5.4	Internal comparison of evaluated model variants	52
5.5	Comparison of the proposed model with related modelling approaches	55

Chapter 1

Introduction

1.1 Background

Human emotion is one of the crucial components of behaviour, communication, cognition, as well as decision-making. In HCI, mental health monitoring, affective computing, adaptive learning systems and brain-computer interface applications, the ability to recognize emotional states of the person involved can help machines to respond in a more personalized, context-aware and human-like manner. In this regard, physiological signals that are generated by the internal body and brain activity provide a way of studying emotions. Among such signals, electroencephalography (EEG) is widely used for emotion recognition tasks as it records electrical activity in the brain directly from the scalp in a non-invasive manner [14], [16]. EEG-based emotion recognition has been studied extensively using traditional machine learning methods as well as deep learning approaches like graph-based models, recurrent networks, multi-layered architectures, convolutional neural networks and more recently, biologically inspired structures. However, challenges remain in EEG emotion recognition that include EEG signals being noisy, non-stationary, heavily subject-dependent, sensitive to electrode placement, device quality, as well as preprocessing choices [16], [18].

Most conventional deep learning approaches process EEG data using continuous-valued activations and a massive number of floating point computations. While it is true that such models can achieve strong classification performance, they do not reflect the event-based nature of biological neural communication. Spiking Neural Networks (SNN), often referred to as the third generation of neural networks, offer a different computational approach by representing information through discrete spike events and membrane potential dynamics [3], [10]. This makes SNNs a strong candidate for EEG analysis because EEG is generated by brain electrical activity, and SNNs provide a biologically inspired way of modelling spatio-temporal signal processing where both the location and timing of an electrical event equally matters.

In addition to biological plausibility, SNNs are closely associated with neuromorphic computing. Neuromorphic computing involves designing brain-inspired computational systems that analyze information efficiently through event-driven mechanisms [9], [24]. Although this research has been tested out on a software level and not on neuromorphic hardware, because of biological plausibility, the use of spike-based

architecture keeps the possibilities towards future low-power neuromorphic EEG systems open. This is relevant due to the fact that practical EEG emotion recognition may eventually need to run on real-time, compact and energy efficient devices.

Despite all these advantages, SNN-based emotion recognition has the limitation of loss of information owing to the conversion of continuous EEG values into spike trains if the appropriate encoding technique is not applied. In addition, SNNs are more difficult to train than ANNs because the spikes, which are basically instantaneous changes with respect to time, are non-differentiable by nature. Surrogate gradient descent has become a common tool in order to deal with this issue, where the gradients are approximated via a smoother function during backpropagation while maintaining spike-based forward propagation [8], [11]. Another crucial limitation is interpretability of EEG channels. Since each EEG channel captures brain activity from a distinct region of the brain, channel-level interpretability would give neuroscientists the opportunity to study the interaction among different parts of brain as well as their contribution towards a specific emotion. Which is why, when a model predicts an emotional state, it is important to understand not only whether the prediction is correct, but also which EEG channels learned signal patterns that contributed to that decision. Prior work on interpretable EEG has shown that neural networks that act as a black box limit neuroscientific usefulness as they may classify EEG windows without actually revealing the neurophysiological evidence behind their decisions [4]. Therefore, interpretability has been treated in this study as a core design requirement rather than an optional step.

To address this interpretability issue, this study integrates a Kolmogorov-Arnold Network (KAN)-based readout module after the SNN layer. KANs replace fixed scalar weights or fixed node activations as seen in MLP with learnable univariate functions on edges represented using spline functions [19]. This structure makes KANs useful for modelling nonlinear relationships while offering a more explainable view of learned transformations. In this study, the KAN module does not replace the SNN, but stand as an interpretable readout mechanism that maps channel-wise SNN features into class-wise contribution scores. In other words, the proposed model combines spike-based temporal processing from SNNs with nonlinear interpretable classification from KANs.

1.2 Rational of the Study or Motivation

The main rationale of this study is based on the need for an EEG emotion recognition model that is not only computationally efficient but also biologically motivated and interpretable. Since EEG signals are temporal activity generated from the brain, using a neural firing dynamics inspired model is suitable for this task. SNNs provide this biological motivation by converting the EEG input signals into spikes and processing them through membrane potential updates, allowing the model to treat the problem as a temporal neural-processing task rather than a static feature-classification problem. Another motivation is the connection between SNN and neuromorphic computing. Traditional deep neural networks involve dense numerical operations, whereas SNNs can represent information through sparse spike events.

As this event-driven style of computation could potentially support low-power processing as opposed to currently used wearable EEG systems that are constrained by device size and power consumption, even though the current study is limited to software level experimentation, one of the motivations of this study is to look towards a meaningful future direction to make the shift to power efficient neuromorphic hardware possible. Finally, the study is also motivated by the interpretability problem in emotion recognition. DER-VREEG, the dataset used in this work, has four emotion labels: happy, calm, scared and bored. The KAN readout is designed to produce a four-dimensional class contribution vector with learnable coefficients that can adapt while training to easily plot how each EEG channel pushes the model towards predicting a certain emotion.

1.3 Problem Statement

EEG-based emotion recognition remains a difficult problem because EEG signals are non-stationary, noisy, low-amplitude and highly variable across participants. Contemporary deep-learning architectures can learn complex patterns from EEG signals, but they often provide limited explanation of how individual EEG channels contribute to emotion classification. This lack of explainability is problematic in affective computing because a prediction alone is not sufficient, neuroscientists also have to understand how different regions of the brain contribute to specific emotion classes. Meanwhile, although SNNs are biologically inspired and suitable for temporal modelling, their use in EEG emotion recognition poses additional challenges. Continuous signals must first be converted into spike trains, and this conversion might lose valuable signal information. Besides, after spike encoding and SNN processing, the final output might become even harder to interpret. A simple linear readout after the SNN can classify the extracted features, but it may not capture nonlinear relationships between channel-specific spike features and emotion classes well enough.

Therefore, the central problem addressed in this study is:

How can an EEG emotion recognition model be designed using a biologically inspired SNN backbone while also providing interpretable channel-wise class contributions through nonlinear KAN-based readout mechanism?

This problem leads to three connected research needs. First, the model should preserve the temporal nature of EEG windows through spike-based processing. Second, the spike-encoding method should be logically explainable in relation to signal saliency and spike latency. Third, the final classifier should provide channel-wise contribution information so that the prediction can be treated beyond a completely black-box decision.

1.4 Objective

The primary objective of this study is to develop an interpretable EEG-based emotion recognition framework using a Spiking Neural Network backbone and a Kolmogorov-Arnold Network readout for channel-wise class contribution analysis. The specific objectives are:

1. To preprocess the DER-VREEG dataset by handling missing values, organizing participant-wise recordings, normalizing EEG signals, and creating fixed-length temporal windows from the four EEG channels used in the dataset.
2. To design an exponentially weighted latency-based spike encoding method that converts continuous EEG windows into spike representations suitable for SNN processing.
3. To implement an SNN backbone that processes the encoded spike trains using membrane-potential dynamics and extracts temporal spike-based features.
4. To integrate a KAN-based readout module after the SNN layer, where each EEG channel produces a four-dimensional contribution vector corresponding to the four emotion classes.
5. To evaluate the proposed model on the four-class emotion recognition task: Happy, Calm, Scared, and Bored.
6. To analyze the interpretability of the model by examining how the channel-wise KAN readout contributes to class-level predictions.

1.5 Methodology in Brief

This study follows a structured workflow for EEG-based emotion recognition using an SNN backbone with a KAN-based interpretable readout. The overall methodology is shown in Figure 1.1. The workflow begins with the DER-VREEG EEG dataset, which contains four-channel wearable EEG recordings from TP9, AF7, AF8, and TP10 for four emotional classes: Calm, Scared, Happy, and Bored.

The first stage of the methodology is preprocessing. This includes handling missing or invalid EEG values, applying participant-wise channel normalization, and dividing the continuous EEG recordings into fixed-length temporal windows. These steps prepare the raw EEG signals for model training while preserving the temporal structure and channel identity of the data.

The second stage is spike encoding. Each preprocessed EEG window is converted into a spike-based representation using threshold-based spike generation followed by exponential latency weighting. This allows the continuous EEG window to be represented in a form suitable for SNN processing, while maintaining a simple relationship between EEG activity, spike occurrence, and temporal weighting.

The third stage is temporal feature extraction using the SNN backbone. The encoded spike-based input is processed through membrane-potential dynamics to extract channel-aware temporal features. These features are then passed to the KAN readout. The KAN readout performs nonlinear classification and produces channel-wise class contribution information, allowing the model to show how each EEG channel contributes toward the four possible emotion classes.

The final stage is emotion prediction and evaluation. The model predicts one of four classes: Calm, Scared, Happy, and Bored. Its performance is evaluated using accuracy, F1-score, Cohen’s kappa, Matthews correlation coefficient, and confusion matrix analysis. In addition, channel-wise contribution analysis is used to interpret the decision behaviour of the KAN readout. A detailed explanation of the pre-processing, spike encoding, SNN backbone, KAN readout, training procedure, and evaluation process is provided in the methodology chapter.

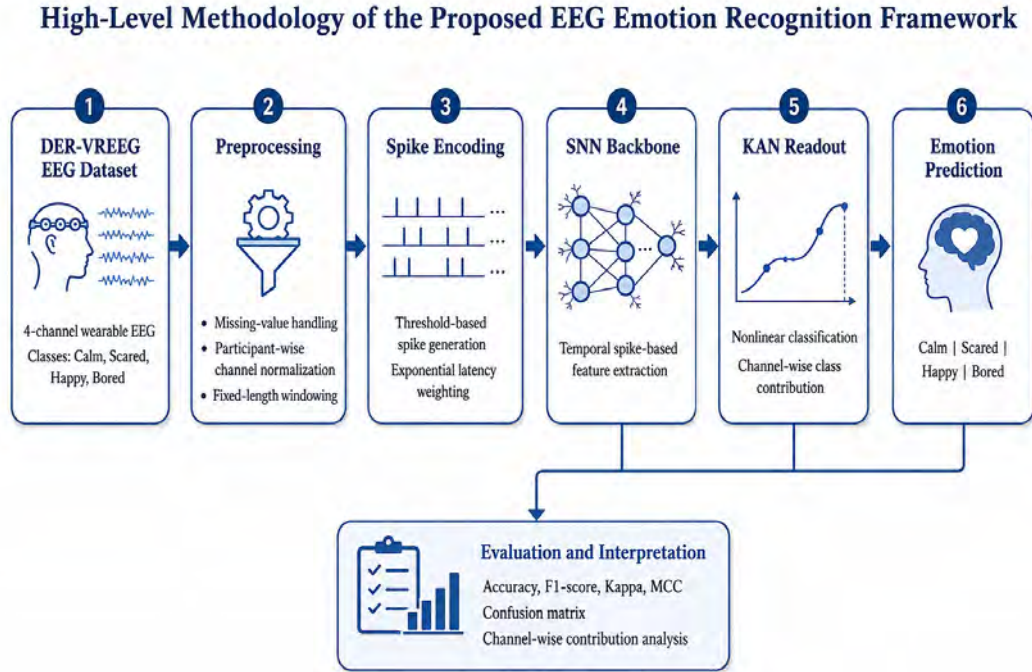


Figure 1.1: High-level methodology of the proposed EEG emotion-recognition framework. The workflow begins with the DER-VREEG EEG dataset, followed by pre-processing, spike encoding, SNN-based temporal feature extraction, KAN readout, emotion prediction, and final evaluation with interpretability analysis.

1.6 Scopes and Challenges

The scope of this study is limited to EEG-based four-class emotion recognition using the DER-VREEG dataset. The work focuses on the design and evaluation of an SNN-based architecture with a KAN readout for interpretable channel-wise classification. The four emotional states considered in this study are Calm, Scared, Happy,

and Bored.

The study does not include real-time EEG acquisition, hardware-level neuromorphic deployment, or emotion intervention. The neuromorphic computing discussion is used as a theoretical and future-oriented motivation for choosing SNNs, not as a claim that the model has been implemented on neuromorphic hardware. Similarly, the biological plausibility of the model is limited to its spike-based computational inspiration; the model does not attempt to reproduce the full biological complexity of the human brain.

Several challenges are associated with this study. Firstly, EEG data is prone to drastic participant variability. EEG signals differ across individuals due to physiological differences, electrode contact quality, head shape, emotional response differences, and recording conditions. A model trained on some participants may not generalize well to unseen participants. To reduce this issue, participant-wise normalization and participant-aware train-validation-test splitting are important parts of the methodology. It also has to be kept in check so that the model does not fine tune too much on the specific participants it was trained on, which would result in overfitting.

The second challenge is spike encoding. Since EEG signals are continuous-valued but SNNs operate on spike events, the encoding stage strongly affects model performance. A poorly designed encoding method may remove useful information before the data reach the SNN. The exponentially weighted latency-based encoding method is used to make this step more explainable, but encoding remains a sensitive design choice.

The third challenge is SNN training. Spike functions are non-differentiable, making ordinary backpropagation difficult. Surrogate-gradient methods help solve this problem, but SNN training can still be unstable and sensitive to hyperparameters such as threshold, decay factor, learning rate, and time-window length [8], [11].

The fourth challenge is interpretability. Although the KAN readout is introduced to improve channel-wise explanation, interpretation must be done carefully. A higher contribution from one EEG channel does not automatically mean that the corresponding brain region alone caused the emotional state. Instead, the contribution score should be understood as the model’s learned evidence from that channel for a particular class.

Finally, the dataset size and channel count are limited. DER-VREEG uses only four EEG channels, which makes the setup practical and wearable-oriented but restricts spatial brain coverage. This affects the complexity of emotional patterns that the model can learn. For this reason, the results should be interpreted as an exploratory study on interpretable SNN-KAN modelling for low-channel EEG emotion recognition rather than as a final generalized solution for all EEG-based affect recognition tasks.

Chapter 2

Literature Review

2.1 Preliminaries

This section introduces the theoretical and technical concepts required to understand the proposed EEG emotion-recognition framework. The main concepts include EEG-based emotion recognition, the Kolmogorov–Arnold representation theorem, Kolmogorov–Arnold Networks, spline-based function approximation, Spiking Neural Networks, and interpretability in EEG modelling.

2.1.1 EEG-Based Emotion Recognition

Electroencephalography (EEG) is a non-invasive technique used to record electrical activity from the scalp. Since emotion is closely associated with neural and physiological processes, EEG has become an important signal modality for affective computing and brain–computer interface research. Compared with external behavioural signals such as facial expression, speech, or gesture, EEG has the advantage of measuring internal brain activity more directly. This makes EEG useful for emotion recognition, especially when external emotional expressions are weak, intentionally controlled, or affected by environmental conditions.

In EEG-based emotion recognition, the objective is to classify a person’s emotional state from recorded EEG signals. Depending on the dataset and experimental design, emotions may be represented using discrete labels such as happy, calm, scared, and bored, or using dimensional models such as valence and arousal.

2.1.2 Kolmogorov–Arnold Representation Theorem

The Kolmogorov–Arnold representation theorem provides an important mathematical foundation for representing multivariate continuous functions. In simplified form, the theorem states that any continuous multivariate function can be represented as a finite composition of continuous univariate functions and addition operations [1], [2]. For a continuous function

$$f : [0, 1]^n \rightarrow R, \tag{2.1}$$

the theorem states that it can be represented as

$$f(x_1, x_2, \dots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^n \phi_{q,p}(x_p) \right), \quad (2.2)$$

where $x_1, x_2, \dots, x_n$ are input variables, $\phi_{q,p}$ are inner univariate functions, and Φ_q are outer univariate functions. The important idea is that a complex multivariate function can be decomposed into combinations of one-dimensional functions.

This theorem is significant because it suggests that high-dimensional nonlinear mappings can be represented through sums and compositions of simpler one-dimensional transformations. In machine learning, this idea motivates architectures that learn nonlinear functions directly instead of relying only on fixed activation functions and scalar weights.

2.1.3 Kolmogorov–Arnold Networks

Kolmogorov–Arnold Networks (KANs) are neural network architectures inspired by the Kolmogorov–Arnold representation theorem. Unlike a conventional Multilayer Perceptron (MLP), where scalar weights are placed on edges and fixed activation functions are applied at nodes, KANs place learnable activation functions on the edges of the network [19]. Therefore, each connection in a KAN does not simply multiply an input by a scalar weight. Instead, it applies a learnable nonlinear function to the input.

For a standard linear layer followed by activation, an MLP layer can be written as

$$\mathbf{h} = \sigma(W\mathbf{x} + \mathbf{b}), \quad (2.3)$$

where W contains scalar weights, $\mathbf{b}$ is the bias vector, and σ is a fixed activation function. In contrast, a KAN layer can be written as

$$y_j = \sum_{i=1}^d \phi_{j,i}(x_i), \quad (2.4)$$

where $\phi_{j,i}$ is a learnable univariate function associated with the connection between input x_i and output y_j . This means that each edge learns how to transform its input before contributing to the output.

The major advantage of this structure is interpretability. Since the edge functions are explicit one-dimensional functions, they can be visualized and inspected. This is useful in scientific machine learning because it allows researchers to examine how individual input variables are transformed by the model. In this study, KAN is not used as the main temporal feature extractor. Instead, it is used as a nonlinear and interpretable readout after the SNN backbone.

2.1.4 Spline-Based Function Approximation in KAN

In practical KAN implementations, the learnable edge functions are often represented using splines. A spline is a piecewise polynomial function defined over a set of intervals. Instead of learning an arbitrary continuous function directly, the model

learns a weighted combination of spline basis functions.

A learnable univariate function $\phi(x)$ can be approximated as

$$\phi(x) = \sum_{m=1}^M c_m B_m(x), \quad (2.5)$$

where $B_m(x)$ represents the m -th spline basis function, c_m is a learnable coefficient, and M is the number of basis functions. The basis functions determine the shape of the local polynomial pieces, while the coefficients control how strongly each basis function contributes to the final learned curve.

For cubic B-splines, each basis function is a piecewise polynomial of degree three. Cubic splines are commonly used because they provide smooth function approximation while remaining flexible enough to model nonlinear relationships. The learnable coefficients are updated during training through gradient-based optimization. As a result, the shape of $\phi(x)$ changes gradually so that the function becomes more suitable for the target classification task.

In the context of this research, the spline-based KAN idea is useful because EEG emotion recognition involves nonlinear relationships between signal features and emotional classes. However, the detailed channel-wise use of KAN in the proposed architecture is not discussed in this chapter. That implementation-level explanation is provided in the methodology chapter.

2.1.5 Spiking Neural Networks

Spiking Neural Networks (SNNs) are biologically inspired neural networks that process information using discrete spike events. Unlike conventional artificial neural networks, which pass continuous activations between layers, SNNs use membrane-potential dynamics and spike-generation mechanisms. A common neuron model used in SNNs is the Leaky Integrate-and-Fire (LIF) neuron.

A simplified membrane-potential update can be written as

$$V_t = \alpha V_{t-1} + I_t, \quad (2.6)$$

where V_t is the membrane potential at time t , α is the decay factor, and I_t is the input current at time t . When the membrane potential crosses a threshold θ , the neuron emits a spike:

$$S_t = \begin{cases} 1, & \text{if } V_t \geq \theta, \\ 0, & \text{otherwise.} \end{cases} \quad (2.7)$$

After a spike is generated, the membrane potential is usually reset or reduced. This dynamic process allows SNNs to model temporal information through spike timing and membrane accumulation.

SNNs are relevant to EEG emotion recognition for two reasons. First, EEG is a temporal signal, and SNNs are naturally designed to process time-dependent spike

sequences. Second, SNNs are biologically inspired, which makes them conceptually suitable for modelling signals generated by brain activity. However, training SNNs is challenging because the spike-generation function is non-differentiable. Surrogate-gradient learning is commonly used to overcome this issue by approximating gradients during backpropagation [8], [11].

2.1.6 Interpretability in EEG Modelling

Interpretability is especially important in EEG-based research because EEG signals are physiological measurements, not ordinary numerical features. A model that predicts an emotional state should ideally provide some information about which EEG channels or signal patterns contributed to that prediction. In neuroscience-oriented EEG analysis, high classification accuracy alone may be insufficient if the model behaves as a black box.

Prior work has shown that interpretability can make EEG deep-learning models more useful for understanding brain-signal patterns. Sturm et al. proposed interpretable deep neural networks for single-trial EEG classification and showed that relevance-based explanations can reveal signal regions used by the model [4]. Schirrmeyer et al. also demonstrated that deep ConvNets can be combined with visualization techniques to support EEG decoding and brain mapping [5]. In EEG emotion recognition specifically, channel-wise attention and relevant-channel analysis have been used to study how different EEG electrodes contribute to emotional-state prediction [15], [27].

Therefore, interpretability is treated in this study as a central design requirement. The proposed model is not intended only to classify EEG windows. It is also designed to provide channel-wise class contribution information through the KAN readout.

2.2 Review of Existing Research

This section reviews existing research related to the proposed study. The review is organized into five groups: DER-VREEG and low-cost VR-EEG emotion recognition, deep-learning approaches for EEG emotion recognition, channel-aware and graph-based EEG emotion recognition, SNN-based EEG modelling, and KAN-based or KAN-related neural architectures.

2.2.1 DER-VREEG and Low-Cost VR-EEG Emotion Recognition

Suhaimi et al. introduced the DER-VREEG dataset for emotion recognition using virtual reality and EEG [12]. Their study used VR stimuli to induce four emotional states: happy, scared, calm, and bored. EEG signals were collected using a low-cost wearable Muse headset with four electrodes: TP9, AF7, AF8, and TP10. The study explored the feasibility of using a compact EEG headset together with immersive VR stimuli for emotion recognition.

The DER-VREEG study is directly relevant to this research because it provides the dataset used in the proposed work. The original study evaluated several conventional machine-learning classifiers, including KNN, SVM, FANN, DRF, GBM, and Naive Bayes. Their results showed that emotion classification is possible even with a low-channel wearable EEG device. The reported performance was strong, especially for inter-subject classification, where an SVM-based model achieved 85.01% accuracy [12].

However, the original DER-VREEG work mainly focused on dataset creation, VR-based emotion induction, feature-based classification, and comparison of traditional machine-learning algorithms. It did not investigate biologically inspired spike-based modelling, neuromorphic relevance, or KAN-based interpretability. Moreover, the original work did not focus on channel-wise class contribution analysis. Therefore, the DER-VREEG dataset provides a suitable foundation for this study, while the proposed SNN-KAN approach explores a different modelling direction.

Another important consideration is the nature of the dataset. DER-VREEG was collected using a low-cost wearable EEG device, which makes it suitable for practical and wearable-oriented emotion-recognition research. However, the dataset has limited spatial coverage because it contains only four EEG channels. This means that any model developed on this dataset should be interpreted carefully. The plus side of this dataset, and the very reason why this dataset was chosen for our research, is that emotions were elicited during EEG collection using VR technology, which is more reliable than EEG elicited via videos displayed on a screen. The goal of this study is therefore not to make broad claims about full-brain emotional dynamics, but to explore interpretable low-channel EEG emotion recognition using a biologically inspired SNN backbone and a KAN readout.

2.2.2 Deep-Learning Approaches for EEG Emotion Recognition

Deep learning has been widely applied to EEG emotion recognition because of its ability to learn nonlinear representations from complex EEG signals. Yang et al. proposed a hierarchical network with subnetwork nodes for EEG-based emotion recognition [7]. Their model used a hierarchical structure to classify emotional states and showed that deep architectures can learn useful EEG representations. This work is important because it moved beyond shallow classifiers and demonstrated the potential of neural-network-based EEG emotion recognition.

However, hierarchical deep networks can still behave as black-box models. Although they may improve classification performance, they do not necessarily provide clear channel-wise explanations. In addition, many deep EEG models are designed mainly for accuracy improvement rather than interpretability. This creates a gap for models that can combine nonlinear learning with explicit channel-level contribution analysis.

Schirrneister et al. proposed deep ConvNets for EEG decoding and visualization [5]. Their work is not limited to emotion recognition, but it is important for EEG deep learning because it showed that deep neural networks can be used not only for

decoding but also for visualizing learned EEG patterns. This supports the argument that EEG models should be evaluated not only by prediction accuracy but also by their ability to reveal meaningful signal patterns.

Lawhern et al. introduced EEGNet, a compact convolutional neural network for EEG-based brain-computer interfaces [6]. EEGNet is important because it uses an EEG-specific architecture and demonstrates that compact models can perform well across BCI tasks. Although EEGNet is not the main model used in this study, it supports the broader idea that EEG architectures should be designed with signal structure and practical constraints in mind.

Overall, deep-learning studies show that neural networks can improve EEG decoding, but many of them still require additional mechanisms for interpretability. The proposed study addresses this issue by placing a KAN-based readout after the SNN backbone so that the model can produce channel-wise class contribution information.

2.2.3 Channel-Aware and Graph-Based EEG Emotion Recognition

EEG signals are spatially structured because each channel corresponds to a specific electrode location on the scalp. For this reason, many recent EEG emotion-recognition models attempt to capture channel relationships, spatial dependencies, and electrode-level importance.

Tao et al. proposed an EEG-based emotion-recognition model using channel-wise attention and self-attention [15]. Their work is highly relevant to this study because it explicitly treats EEG channels as having different importance for emotion recognition. The channel-wise attention mechanism allows the model to assign different weights to different electrodes, while self-attention helps capture relationships within the learned representation. This supports the idea that channel-level information is meaningful in EEG emotion recognition.

However, attention weights are not always straightforward to interpret as direct causal explanations. They can indicate relative importance within the model, but they do not necessarily show an explicit nonlinear function from channel-specific features to class-specific evidence. In contrast, the proposed KAN readout is designed to generate channel-wise class contribution vectors, making the contribution structure more direct at the output level.

Jin et al. proposed PGCN, a Pyramidal Graph Convolutional Network for EEG emotion recognition [17]. Their method models EEG electrode relationships at multiple levels: local, mesoscopic, and global. This is important because EEG channels are not independent; emotional processing may involve relationships across different brain regions. PGCN addresses the limitation of shallow graph convolutional networks, which may capture only local electrode relationships, and the over-smoothing problem that can occur when too many graph convolutional layers are stacked.

PGCN is a strong example of spatially aware EEG emotion recognition. However, it is mainly focused on graph-based spatial relationship modelling. The proposed study has a different focus. Since DER-VREEG contains only four EEG channels, constructing a complex multi-level graph structure may not be as meaningful as in datasets with many electrodes. Instead, this study focuses on temporal spike-based processing and channel-wise contribution analysis using a KAN readout.

Valderrama and Sheoran investigated relevant EEG channels for subject-independent emotion recognition using attention network layers [27]. Their work further supports the importance of identifying which EEG channels contribute most to emotion recognition. This is closely aligned with the interpretability goal of the present study. However, there is a subtle difference that their work does not explicitly show how a specific channel reacts to a specific emotion.

Taken together, channel-aware and graph-based studies show that EEG emotion recognition benefits from modelling electrode-level information. These works motivate the channel-wise interpretability component of the proposed model. However, most existing models do not combine channel-wise output contribution analysis with biologically inspired spike-based temporal processing.

2.2.4 Spiking Neural Networks for EEG and Emotion-Related Recognition

Spiking Neural Networks have gained attention because they process information using spike events and membrane-potential dynamics. Their biological inspiration makes them conceptually suitable for EEG modelling, where the input signal itself reflects brain electrical activity. SNNs are also relevant to neuromorphic computing, where event-driven processing can potentially support low-power and real-time applications.

Li et al. proposed a fractal spiking neural network scheme for EEG-based emotion recognition [13]. This work is directly related to the present study because it uses SNNs for EEG emotion recognition. The fractal SNN approach shows that spike-based models can be applied to affective EEG tasks and supports the suitability of SNNs for emotion-recognition research.

However, the fractal SNN study mainly focuses on improving emotion-recognition performance through a spiking architecture. It does not address KAN-based readout interpretability or explicit channel-wise class contribution analysis. Therefore, while it supports the use of SNNs in EEG emotion recognition, it does not solve the interpretability problem targeted in the present study.

Joshi et al. studied EEG-based stress detection using Spiking Neural Networks and Convolutional Spiking Neural Networks [22]. Although this work focuses on stress detection rather than four-class emotion recognition, it is relevant because stress detection is also an affective-state recognition problem using EEG. Their work shows that SNN and CSNN models can be effective for EEG signal classification. It also supports the idea that spike-based architectures are becoming increasingly relevant

in EEG-based mental-state analysis.

More broadly, SNN research has shown that spike-based neural networks can process temporal information through biologically inspired mechanisms [3], [10]. However, training SNNs remains difficult because spike functions are non-differentiable. Surrogate-gradient methods have become a common solution for training SNNs using backpropagation-like optimization [8], [11]. These methods are important for the proposed study because the SNN backbone requires trainable spike-based processing.

The key limitation in existing SNN-based EEG research is that interpretability is often secondary to classification performance. Existing SNN studies demonstrate biological motivation and temporal processing ability, but they do not fully explain how each EEG channel contributes to each predicted emotional class. This motivates the integration of a KAN readout in the present study.

2.2.5 Kolmogorov–Arnold Networks and Their Relevance to EEG Modelling

Kolmogorov–Arnold Networks (KANs) were introduced by Liu et al. as an alternative to conventional multilayer perceptrons [25]. Unlike MLPs, where scalar weights are applied between neurons and fixed activation functions are placed at nodes, KANs use learnable univariate functions on network edges. These functions are commonly parameterized using splines. Liu et al. reported that this design can improve both function approximation and interpretability, since the learned edge functions can be visualized and inspected directly [25]. This property makes KANs especially relevant for scientific and biomedical machine learning, where understanding the learned relationship between input features and output decisions is often as important as prediction accuracy.

Recent studies have started applying KANs to EEG and biosignal modelling. Contreras et al. proposed KAN–EEG for epileptic seizure detection, where KAN was used as a replacement for the conventional MLP backbone in an EEG classification system [21]. Their results suggest that KANs can be effective for EEG classification and may provide a useful alternative to MLP-based classifiers. However, this work focused on seizure detection rather than emotion recognition. It also did not investigate spike-based temporal processing or SNN-based biological plausibility.

KANs have also been explored specifically for EEG emotion recognition. Ping et al. proposed KAN-GCNN, a Kolmogorov–Arnold Network-enhanced Graph Convolutional Neural Network for EEG-based emotion recognition [26]. Their method integrates KAN with graph convolution to improve nonlinear feature modelling in EEG emotion recognition. This paper is important because it shows that KAN is already being applied to the same general task area as the present study: EEG-based emotion recognition. However, KAN-GCNN follows a graph-convolutional modelling direction that focuses mostly on spatial dynamics. It does not use an SNN backbone, spike encoding, membrane-potential dynamics, temporal dynamics or neuromorphic motivation. Therefore, it supports the relevance of KAN for EEG emotion recognition, while also showing that the proposed SNN-KAN direction re-

mains distinct.

A small number of recent works have also explored combinations of KANs and SNNs outside EEG emotion recognition. Kruttsylo et al. presented hybrid Spiking–Kolmogorov–Arnold Networks, including a KAN-readout SNN that replaces the standard linear readout with a spline-based KAN readout, and a KAN-integrated SNN that inserts spline transformations inside spiking layers [23]. This is closely related to the design logic of the present study because it shows that using KAN as a readout after an SNN is an emerging architectural direction. However, their work was not focused on EEG or emotion recognition. Similarly, Ghadekar et al. proposed a multimodal learning framework combining SNNs and KANs with modified SHAP visualization [20]. This supports the broader idea that SNN–KAN hybrid models can be combined with interpretability methods, but the study is not specifically designed for EEG emotion recognition. Another recent preprint, E-SKAN, introduced an event-driven spiking KAN architecture intended to bridge SNN efficiency and KAN expressiveness [28]. Since it is a preprint, it should be treated as emerging evidence rather than established literature.

These studies clarify the research position of the present work. It would not be accurate to claim that KAN has never been used for EEG, because KAN–EEG and KAN–GCNN already demonstrate KAN-based EEG applications. It would also not be accurate to claim that KAN and SNN have never been combined, because recent hybrid SNN–KAN studies exist. However, the reviewed literature suggests that the specific combination targeted in this study remains underexplored: EEG emotion recognition using spike-based temporal processing through an SNN backbone together with a KAN-based readout for interpretable channel-wise class contribution analysis. Therefore, the novelty of this research is not simply the use of KAN or SNN alone, but the way KAN is used as an interpretable readout after biologically inspired SNN processing for low-channel EEG emotion recognition.

2.2.6 Research Gap

The reviewed studies show that EEG emotion recognition has been approached from several directions. Traditional machine-learning studies, such as the original DER-VREEG work, show that low-cost wearable EEG can be used for four-class emotion recognition. Deep-learning studies show that neural networks can learn complex EEG representations. Channel-wise attention and graph-based models show that electrode-level relationships and channel importance are important for EEG emotion recognition. SNN-based studies show that spike-based models are suitable for EEG and affective-state recognition. KAN-based studies show that learnable non-linear edge functions can support flexible and interpretable modelling.

However, several gaps remain. First, many EEG emotion-recognition models focus mainly on classification performance and provide limited explanation of channel-wise class evidence. Second, SNN-based EEG models are biologically inspired but are often not designed with explicit output-level interpretability. Third, KAN-based EEG models are emerging, but their integration with SNNs for interpretable EEG emotion recognition remains limited. Fourth, the DER-VREEG dataset has been

studied using traditional machine-learning classifiers, but not sufficiently explored using a biologically inspired SNN architecture with KAN-based channel-wise interpretability.

Therefore, this study addresses the following gap: existing EEG emotion-recognition studies have explored deep learning, graph modelling, attention mechanisms, SNNs, and KANs separately, but there is limited work on combining SNN-based temporal spike processing with a KAN readout that provides channel-wise class contribution information for low-channel EEG emotion recognition.

2.3 Summary of Key Findings

The literature review leads to several key findings that guide the proposed research.

First, the DER-VREEG dataset is appropriate for this study because it provides low-cost wearable EEG recordings for four emotional states: Calm, Scared, Happy, and Bored recorded under a VR environment. The original DER-VREEG study demonstrated that four-class emotion recognition is possible using traditional machine-learning methods. However, it did not investigate SNN-based temporal modelling or KAN-based interpretability.

Second, deep-learning models such as hierarchical networks, CNN-based EEG decoders, attention models, and graph neural networks have improved EEG emotion recognition by learning complex spatial, temporal, or channel-aware representations. However, many of these models remain difficult to interpret, especially when trying to understand how individual EEG channels support a particular class prediction.

Third, channel-aware studies show that different EEG electrodes contribute differently to emotion recognition. This supports the design decision of treating channel-wise contribution as an important part of the proposed model. Since the DER-VREEG dataset contains only four EEG channels, each channel has high importance, and the model should preserve channel-level interpretability.

Fourth, SNNs are suitable for this research because they provide biologically inspired temporal processing through spike events and membrane-potential dynamics. This aligns conceptually with EEG as a brain-generated electrical signal and with neuro-morphic computing principles. However, existing SNN-based EEG studies generally emphasize classification performance more than interpretable channel-wise class contribution.

Fifth, KANs provide a useful readout mechanism because they use learnable nonlinear functions that can be inspected more directly than ordinary scalar weights. This makes them suitable for modelling nonlinear relationships between SNN-extracted EEG features and emotion classes.

Based on these findings, the proposed study uses an SNN backbone for spike-based temporal EEG processing and a KAN readout for nonlinear classification and channel-wise interpretability. It explores a direction complementary to the key find-

ings: biologically inspired, low-channel EEG emotion recognition with interpretable channel-wise class contribution analysis.

Chapter 3

Requirements, Impacts and Constraints

3.1 Final Specifications and Requirements

This project focuses on EEG-based emotion recognition using a biologically inspired Spiking Neural Network (SNN) backbone and an interpretable Kolmogorov–Arnold Network (KAN) readout. The final system is designed as a research prototype rather than a deployable clinical or commercial product. Therefore, the main requirements are methodological, computational, and experimental.

The first requirement is the availability of a suitable EEG dataset. This study uses the DER-VREEG dataset, which contains EEG recordings collected using a Muse headset with four channels: TP9, AF7, AF8, and TP10. The dataset includes four emotional classes: Calm, Scared, Happy, and Bored [12]. Since this work uses an existing dataset, no new EEG data collection or participant recruitment is required.

The second requirement is proper preprocessing of EEG data. The raw EEG recordings must be cleaned by handling missing values, preserving temporal order, normalizing signal values, and generating fixed-length EEG windows. These steps are necessary because EEG signals are noisy, participant-dependent, and sensitive to scale differences across channels and individuals.

The third requirement is the design of a spike-encoding method. Since SNNs process spike trains rather than ordinary continuous values, the EEG windows must be converted into spike-based representations. In this project, an exponentially weighted latency-based encoding strategy is used so that the model can represent EEG windows in a temporally meaningful spike format.

The fourth requirement is the implementation of the proposed learning architecture. The model must include an SNN backbone for temporal spike-based processing and a KAN readout for nonlinear classification and channel-wise interpretability. The KAN readout should support the analysis of how each EEG channel contributes to the four emotion classes.

The fifth requirement is an appropriate evaluation strategy. The model is evaluated

using classification metrics such as accuracy, macro F1-score, weighted F1-score, Cohen’s kappa, Matthews correlation coefficient, and confusion matrix analysis. Since EEG emotion recognition is sensitive to participant variability, the evaluation should avoid data leakage and should be reported honestly, including both strengths and limitations.

The software requirements of the project include Python, common scientific computing libraries such as NumPy and Pandas, machine learning libraries such as Scikit-learn, and deep learning frameworks such as PyTorch and SnnTorch. Additional libraries may be used for visualization, metric calculation, and result analysis. The project does not require paid software licenses, as it can be implemented using open-source tools.

3.2 Societal Impact

The project contributes to the broader field of affective computing, where the goal is to develop systems that can recognize or respond to human emotional states. Emotion-aware systems may eventually support applications in education, virtual reality, human–computer interaction, mental-health research, driver monitoring, and assistive technologies. For example, adaptive learning systems may respond to frustration, and immersive environments may adjust their content according to user engagement or discomfort.

The societal contribution of this project is mainly research-oriented. It does not produce a commercial emotion-recognition system, but it investigates whether an SNN-based model with a KAN readout can support interpretable EEG emotion recognition. This may help future researchers design models that are not only accurate but also more transparent in terms of channel-wise contribution.

The project may also have long-term relevance for health-related technologies. EEG-based emotion recognition could contribute to future mental-health monitoring, stress analysis, and affective brain–computer interface systems. However, this study does not diagnose mental illness, measure clinical conditions, or provide medical recommendations. Its contribution is limited to emotion-classification research using an existing EEG dataset.

There are also possible negative societal implications. Emotion recognition technology can be misused for emotional surveillance, workplace monitoring, targeted advertising, or behavioural profiling. Such applications may reduce personal autonomy and privacy. Therefore, any future use of EEG-based emotion recognition should be carefully regulated and should involve informed consent, transparency, and human oversight.

Another societal concern is over-reliance on automated emotional assessment. Human emotions are complex, and no machine learning model can fully understand a person’s emotional state. A model may classify EEG patterns associated with a label, but this should not be treated as a definitive judgment about the person’s inner experience. Therefore, emotion-recognition outputs should be interpreted as

probabilistic model predictions rather than absolute emotional truth.

3.3 Environmental Impact

The direct environmental impact of this project is relatively small because it is a software-based academic study using an existing dataset. No new EEG headset, VR equipment, laboratory setup, or participant experiment is required for this thesis. Reusing the DER-VREEG dataset reduces the need for additional experimental data collection and supports sustainable research practice by extracting further value from already collected data.

The main environmental cost of the project comes from computational energy consumption. Model development requires repeated preprocessing, training, testing, hyperparameter adjustment, and result analysis. These computational processes consume electricity. Although the scale of this thesis is small compared with large industrial AI systems, the environmental cost of computation should still be acknowledged.

The use of SNNs provides a relevant future sustainability direction. SNNs are often associated with event-driven processing and neuromorphic computing, which may support energy-efficient AI implementations in the future [9]. However, since this project is implemented on conventional computing hardware rather than neuromorphic hardware, it would be inappropriate to claim direct energy savings. But the fact remains that the model design is conceptually aligned with future low-power EEG processing systems.

The broader field of EEG and VR research also depends on electronic devices such as headsets, sensors, computers, and display equipment. The manufacturing, maintenance, and disposal of such devices can contribute to environmental burdens and electronic waste. This project does not directly produce such hardware waste, but responsible use of existing equipment and proper electronic waste management remain important considerations for the wider field.

3.4 Ethical Issues

The most important ethical issue in this project is the sensitivity of EEG and emotional data. EEG signals are generated from brain activity, and although they cannot read thoughts, they may still reveal information related to cognitive or emotional states. Since this study aims to classify emotions such as Calm, Scared, Happy, and Bored, the data should be treated as sensitive human-subject information.

Informed consent is another important ethical consideration. The DER-VREEG dataset was collected by the original researchers using VR-based emotional stimulation. Participants in such studies should be informed about the purpose of the experiment, the type of EEG data collected, the emotional nature of the stimuli, and their right to withdraw. Although this thesis does not collect new participant

data, it depends on ethically collected existing data.

The use of VR to induce emotions also raises psychological safety concerns. Some emotional stimuli, especially those designed to induce fear or discomfort, may cause temporary stress, anxiety, dizziness, or motion sickness. These concerns are more directly relevant to the original data collection process than to the present secondary analysis, but they remain important when discussing the ethical context of VR-based emotion recognition.

Another ethical issue is the risk of misinterpretation. A machine learning model may classify a signal as “Scared” or “Bored”, but this does not mean that the model fully understands the participant’s subjective emotional experience. Therefore, results should be communicated carefully and should not be exaggerated as mind-reading or clinical emotional diagnosis.

Bias and fairness are also relevant. EEG responses may vary according to participant characteristics, cultural background, emotional interpretation, neurological differences, and recording conditions. A model trained on one dataset may not generalize equally well to all populations. Therefore, this study avoids universal claims and treats its findings as dataset-specific.

Professional responsibility requires honest reporting of results, proper citation of data and prior work, transparent documentation of preprocessing and model design, and clear communication of limitations. Therefore, this study reports both successful and unsuccessful findings and avoids presenting the model as more reliable or generalizable than the evidence supports.

3.5 Standards

This project does not implement a medical device, clinical diagnostic system, or commercial EEG product. Therefore, no hardware safety or clinical certification standard is directly applied. However, several professional and ethical standards are relevant to the research process.

The first relevant standard is IEEE 7000-2021, which provides a model process for addressing ethical concerns during system design. This standard is relevant because the project involves emotion-related AI and sensitive EEG data. Its emphasis on ethical values, stakeholder concerns, transparency, and traceability supports the responsible design of human-centered AI systems.

The second relevant framework is the ACM Code of Ethics and Professional Conduct. This study follows the general principles of honesty, privacy protection, avoidance of harm, proper credit to original work, and responsible communication of system limitations. These principles are important because the work uses human-related data and machine learning predictions.

From a technical reporting perspective, the project also follows standard machine-learning research practices: clear dataset description, preprocessing documentation,

proper train-validation-test separation, use of appropriate evaluation metrics, and reproducible reporting of experimental settings.

3.6 Project Management Plan

The project management plan is organized around the main stages of a research-based machine learning thesis. Since the project uses an existing dataset and open-source software, the main resources are time, computing capacity, literature support, and careful documentation.

Table 3.1: Project Management Plan

Phase	Main Activities	Deliverables
Literature Review	Study EEG emotion recognition, SNNs, KANs, interpretability, DER-VREEG, and related models.	Chapter 2 draft and reference list
Dataset Preparation	Load dataset, inspect structure, handle missing values, normalize signals, and create EEG windows.	Cleaned dataset and preprocessing pipeline
Model Design	Design spike encoding, SNN backbone, and KAN readout.	Final architecture description
Implementation	Implement preprocessing, training, evaluation, and visualization code.	Working model code
Experimentation	Train baseline and proposed models, collect performance metrics, and analyze results.	Accuracy, F1-score, MCC, kappa, confusion matrix
Interpretability Analysis	Analyze channel-wise class contribution from KAN readout.	Interpretability figures and discussion
Report Writing	Prepare introduction, literature review, methodology, results, limitations, and conclusion.	Final thesis report
Final Review	Check formatting, references, grammar, figures, and consistency.	Submission-ready report

The estimated schedule follows a compact thesis workflow. Literature review and dataset understanding are completed first, followed by preprocessing and model development. Experimental evaluation and report writing are performed in parallel after the core model is implemented. Final revision focuses on consistency, formatting, and clear presentation of limitations.

The resource plan includes open-source Python libraries, the DER-VREEG dataset, LaTeX-based report preparation, and reference-management tools. Since no new data collection is performed, the project does not require participant compensation, laboratory booking, or EEG/VR equipment purchase.

3.7 Risk Management

Several risks may affect the project. These risks are mainly related to dataset limitations, model performance, time constraints, interpretability, and reproducibility.

Table 3.2: Risk Management Plan

Risk	Possible Impact	Mitigation Strategy
Small number of EEG channels	Limits spatial brain coverage and channel relationship modelling.	Focus on low-channel wearable EEG and channel-wise contribution analysis.
Overfitting	Model may learn dataset-specific patterns instead of generalizable features.	Use proper train-validation-test split, regularization, and careful evaluation.
Spike encoding information loss	Useful EEG information may be lost during conversion to spikes.	Use explainable encoding and discuss encoding limitations.
Interpretability overclaim	Channel contribution may be misread as causal brain explanation.	Describe contribution as model-level evidence, not biological causality.
Reproducibility issues	Results may be difficult to verify without clear documentation.	Document preprocessing, model settings, metrics, and experimental configuration.

The most important mitigation strategy is honest scope control. This project should not claim to solve EEG emotion recognition universally. Instead, it should be presented as a research exploration of interpretable SNN-KAN modelling for low-channel EEG emotion recognition.

3.8 Economic Analysis

The economic analysis of this project should be understood in the context of an academic research prototype rather than a commercial product. The project does not generate direct financial revenue. Its value lies in knowledge creation, methodological exploration, and contribution to EEG-based affective computing research.

A major economic advantage is the use of an existing dataset. Collecting a new EEG-VR emotion dataset would require EEG headsets, VR equipment, laboratory space, participant recruitment, ethical approval procedures, researcher time, and possibly participant compensation. By using the DER-VREEG dataset, the project avoids these costs and becomes financially feasible within a student research setting.

The main cost of the project is computational. Training and evaluating machine learning models require electricity and computing resources. However, because the dataset is relatively small and the project is conducted at thesis scale, the cost is modest compared with large-scale AI research. The use of open-source software further reduces financial burden because no paid licenses are required.

Table 3.3: Estimated Cost Components

Cost Component	Estimated Cost Consideration
Dataset	No direct cost because an existing dataset is used.
EEG and VR equipment	No purchase required for this thesis because no new data collection is performed.
Software	No direct licensing cost if open-source Python and LaTeX tools are used.
Computation	Low to moderate electricity/computing cost depending on training repetitions.
Researcher time	Main investment; includes literature review, coding, analysis, and report writing.
Documentation and publication preparation	Minimal direct cost; mainly time and formatting effort.

The benefit of the project is mainly indirect. It may help future researchers understand the strengths and limitations of combining SNNs and KANs for EEG emotion recognition. It may also provide insight into channel-wise interpretability for low-channel EEG systems. These findings can reduce future research effort by identifying useful modelling directions and limitations.

In the long term, emotion-recognition technologies may have economic value in education, healthcare research, virtual reality, user-experience design, and human-computer interaction. However, such applications are beyond the scope of this thesis and should be treated as future possibilities rather than immediate outcomes.

Overall, the project has a favourable cost-benefit balance for an academic study. The financial cost is low because it uses existing data and open-source tools, while the potential research value is meaningful because it explores an interpretable and biologically inspired approach to EEG emotion recognition.

Chapter 4

Proposed Methodology

4.1 Design Process or Methodology Overview

This chapter presents the design and implementation methodology of the proposed EEG-based emotion-recognition framework. The main objective of the design is to classify four emotional states, namely Calm, Scared, Happy, and Bored, from low-channel EEG recordings while maintaining biological motivation and channel-wise interpretability. The proposed framework combines three major components: pre-processing of raw EEG signals, spike-based temporal processing using a Spiking Neural Network (SNN), and nonlinear interpretable classification using a Kolmogorov–Arnold Network (KAN) readout.

The overall design process is guided by the objectives and constraints identified in the previous chapters. Since EEG signals are temporal, noisy, participant-dependent, and low-dimensional in the DER-VREEG dataset, the methodology must preserve temporal order, reduce scale variation, avoid data leakage between participants, and provide a meaningful explanation of the model’s decision. The use of an SNN is motivated by the temporal and biologically inspired nature of spike-based computation, while the KAN readout is used to improve interpretability by estimating channel-wise class contributions.

The proposed methodology follows the sequence below:

1. Load the DER-VREEG EEG dataset containing raw EEG recordings from TP9, AF7, AF8, and TP10 channels.
2. Clean the dataset by handling missing and invalid EEG values.
3. Preserve participant-wise and class-wise temporal order.
4. Normalize EEG signals to reduce channel-scale and participant-scale variation.
5. Segment the continuous EEG recordings into fixed-length temporal windows.
6. Convert each EEG window into spike trains using exponentially weighted latency-based spike encoding.
7. Process the encoded spike trains using an SNN backbone.

8. Extract channel-aware SNN features from the final membrane or spike-based responses.
9. Pass the channel-wise SNN features into a KAN readout.
10. Produce channel-wise four-dimensional class contribution vectors.
11. Combine channel-wise contributions to generate final emotion logits.
12. Apply softmax classification and evaluate the predicted emotion class.

The high-level flow of the proposed methodology can be summarized as:

$$\begin{aligned} \text{Raw EEG} &\rightarrow \text{Preprocessing} \rightarrow \text{Windowing} \rightarrow \text{Spike Encoding} \\ &\rightarrow \text{SNN Backbone} \rightarrow \text{KAN Readout} \rightarrow \text{Emotion Prediction.} \end{aligned} \quad (4.1)$$

The main design idea is that the SNN learns temporal spike-based representations from EEG windows, while the KAN readout learns nonlinear mappings from channel-wise SNN features to emotion-class contributions. This design allows the final prediction to be decomposed according to EEG channels, which supports interpretability.

4.2 Preliminary Design or Design (Model) Specification

Before finalizing the proposed architecture, multiple alternative modelling directions were considered. The alternatives were selected based on three major criteria: suitability for EEG temporal modelling, ability to work with a low-channel dataset, and capacity to provide interpretable output.

4.2.1 Alternative Design 1: Conventional Machine Learning

The first possible design was a conventional machine-learning pipeline. In this design, statistical or handcrafted features would be extracted from EEG windows and classified using models such as Support Vector Machine, Random Forest, K-Nearest Neighbours, or Gradient Boosting. This direction is consistent with the original DER-VREEG study, where several traditional classifiers were evaluated [12].

The advantage of this design is simplicity. It is computationally inexpensive and easier to implement. However, it has two major limitations for the present research. First, handcrafted features may fail to preserve the temporal dynamics of EEG windows. Second, conventional classifiers do not naturally provide biologically inspired spike-based modelling or channel-wise nonlinear contribution analysis. Therefore, this design was not selected as the final architecture.

4.2.2 Alternative Design 2: Conventional Deep Neural Network

The second possible design was a conventional deep-learning model such as a Multi-layer Perceptron (MLP), Convolutional Neural Network (CNN), or recurrent model. These models can learn nonlinear EEG representations and have been widely used in EEG emotion-recognition research. However, they usually process information through continuous activations rather than spike events. As a result, they do not directly support the biological plausibility and neuromorphic motivation of this study.

Another limitation is interpretability. Although attention or visualization methods can be added to deep models, a standard dense classifier does not automatically provide channel-wise class contribution vectors. Since interpretability is a central requirement of this research, a conventional deep network alone was not selected.

4.2.3 Alternative Design 3: Plain SNN with Linear Readout

The third design was a plain SNN model followed by a standard linear classifier. This design supports biologically inspired temporal processing and can handle spike-encoded EEG windows. The SNN converts spike sequences into temporal hidden states or membrane features, and the linear readout maps those features to four emotion classes.

This design is useful as a baseline because it tests whether spike-based temporal processing alone is sufficient for the task. However, the linear readout has limited nonlinear modelling capacity and provides weak interpretability. It produces final class logits directly, without showing how individual EEG channels contribute to each class. Therefore, although the plain SNN is important for comparison, it does not fully satisfy the interpretability objective.

4.2.4 Alternative Design 4: KAN Encoder with SNN

Another possible design was to place a KAN module before spike encoding. In this structure, the KAN would transform raw or normalized EEG channel values before converting them into spikes. This design could improve adaptive channel calibration before spike generation. However, it also increases architectural complexity without adding much value, and given the relatively small size of the dataset, it had lower validation accuracy. In other words, the model overfits as the KAN encoder fits the input EEG values too well, resulting in poor generalization.

Furthermore, the final chosen direction emphasizes interpretability at the decision level. For this reason, the KAN component is placed after the SNN, where it can directly map channel-wise SNN features into class-wise contribution scores.

4.2.5 Selected Design: SNN Backbone with KAN Readout

The selected design uses an SNN backbone followed by a KAN-based readout. This design was selected because it satisfies the two main goals of the project:

1. It uses spike-based temporal processing through an SNN, which supports the biological and neuromorphic motivation of the study.
2. It uses a KAN readout to generate nonlinear channel-wise class contribution scores, supporting interpretability.

The selected model can be written conceptually as:

$$\hat{y} = f_{\text{KAN}}(f_{\text{SNN}}(E(X))), \quad (4.2)$$

where X is the preprocessed EEG window, $E(\cdot)$ is the spike-encoding function, $f_{\text{SNN}}(\cdot)$ is the SNN temporal feature extractor, and $f_{\text{KAN}}(\cdot)$ is the KAN readout classifier.

4.2.6 Input Representation

Let one EEG window be represented as:

$$X \in R^{T \times C}, \quad (4.3)$$

where T is the number of time steps in the window and C is the number of EEG channels. In this thesis,

$$C = 4, \quad (4.4)$$

corresponding to the four DER-VREEG EEG channels:

$$\{TP9, AF7, AF8, TP10\}. \quad (4.5)$$

Each window can therefore be written as:

$$X = \begin{bmatrix} x_{1,TP9} & x_{1,AF7} & x_{1,AF8} & x_{1,TP10} \\ x_{2,TP9} & x_{2,AF7} & x_{2,AF8} & x_{2,TP10} \\ \vdots & \vdots & \vdots & \vdots \\ x_{T,TP9} & x_{T,AF7} & x_{T,AF8} & x_{T,TP10} \end{bmatrix}. \quad (4.6)$$

Here, $x_{t,c}$ represents the normalized EEG value at time step t for channel c .

4.2.7 Exponential Latency-Based Spike Encoding

Since SNNs process spike events rather than ordinary continuous values, each normalized EEG window must be converted into a spike-based representation. In this study, an exponentially weighted latency-based encoding scheme is used. The purpose of this design is to convert EEG activity into binary spike events while giving higher importance to earlier spike occurrences within a temporal window.

Let $\tilde{x}_{t,c}$ represent the normalized EEG value at time step t for channel c , where $t = 0, 1, \dots, T - 1$. A binary spike is generated by comparing the normalized EEG value with an encoding threshold θ_e :

$$s_{t,c} = \begin{cases} 1, & \text{if } \tilde{x}_{t,c} \geq \theta_e, \\ 0, & \text{otherwise.} \end{cases} \quad (4.7)$$

Here, $s_{t,c} = 1$ indicates that the EEG activity at time step t and channel c is strong enough to produce a spike, while $s_{t,c} = 0$ indicates no spike. The complete spike-encoded window is therefore represented as:

$$S \in \{0, 1\}^{T \times C}, \quad (4.8)$$

where T is the window length and C is the number of EEG channels.

To introduce latency-sensitive behaviour, an exponential temporal weighting term is applied to the generated spikes:

$$w_t = e^{-\lambda t}, \quad (4.9)$$

where λ is the decay parameter. Since w_t decreases as t increases, spikes occurring earlier in the window receive higher importance than spikes occurring later in the same window. This produces a simple latency-inspired interpretation: earlier threshold-crossing events are treated as stronger evidence for the SNN input.

For each EEG channel, the exponentially weighted spike activity is summarized as:

$$I_c = \sum_{t=0}^{T-1} e^{-\lambda t} s_{t,c}. \quad (4.10)$$

The resulting encoded current vector for one EEG window is:

$$\mathbf{I} = [I_{TP9}, I_{AF7}, I_{AF8}, I_{TP10}]. \quad (4.11)$$

This vector is then used as the input current to the SNN backbone. The reasoning behind this design is that the EEG signal is first normalized to remove scale effects, then converted into discrete spike events through thresholding, and finally weighted according to spike timing. Therefore, the encoder preserves a clear and explainable relationship between signal intensity, spike occurrence, and latency-sensitive SNN input. Compared with using raw continuous EEG values directly, this representation is more consistent with the spike-based processing principle of SNNs.

4.2.8 SNN Backbone

The encoded EEG spike train is passed into the SNN backbone. In the proposed architecture, the SNN backbone is designed to preserve channel-wise interpretability while also allowing cross-channel interaction learning. Therefore, the model contains four channel-specific SNN branches, one for each EEG channel, and one additional interaction branch that receives all channels jointly. Let the encoded spike vector at time t be denoted as:

$$\mathbf{S}_t = [S_{t,TP9}, S_{t,AF7}, S_{t,AF8}, S_{t,TP10}]^T, \quad (4.12)$$

where $S_{t,c}$ denotes the encoded spike value of channel c at time t . The four channel-wise SNN branches process the spike sequence of each EEG channel separately. This preserves a direct relationship between each channel and its learned temporal representation. In addition, the interaction branch processes the full four-channel spike vector $\mathbf{S}_t$ in order to capture spatial or inter-channel dependencies that may

be relevant for emotion recognition. For each channel-specific branch, the input current is computed as:

$$\mathbf{u}_{t,c} = W_c \mathbf{I}_{t,c} + \mathbf{b}_c, \quad (4.13)$$

where $c \in \{TP9, AF7, AF8, TP10\}$, $S_{t,c}$ is the spike value of channel c at time t , W_c is the trainable input projection for the corresponding channel branch, $\mathbf{b}_c$ is the bias term, and $e^{-\lambda t}$ is the exponential latency-weighting factor. This term gives stronger influence to earlier spikes within the EEG window. The interaction branch receives the complete four-channel spike vector:

$$\mathbf{u}_{t,int} = W_{int} \mathbf{I}_t + \mathbf{b}_{int}, \quad (4.14)$$

where W_{int} is the trainable projection matrix of the interaction branch and $\mathbf{b}_{int}$ is its bias term. Unlike the channel-specific branches, this branch intentionally mixes information from all four EEG channels. It is included to capture inter-channel relationships such as asymmetry or joint activation patterns between electrodes. Each SNN branch is composed of Leaky Integrate-and-Fire neurons. For a branch receiving input current $\mathbf{I}_t$, the membrane potential is updated as:

$$\mathbf{V}_t = \alpha \mathbf{V}_{t-1} + \mathbf{u}_t, \quad (4.15)$$

where $\mathbf{V}_t$ is the membrane potential vector at time t , α is the membrane decay factor, and $\mathbf{I}_t$ is the input current vector of the corresponding branch. A spike is generated when the membrane potential crosses the spiking threshold:

$$\mathbf{O}_t = H(\mathbf{V}_t - \boldsymbol{\theta}_s), \quad (4.16)$$

where $\mathbf{O}_t$ is the output spike vector, $\boldsymbol{\theta}_s$ is the spiking threshold vector, and $H(\cdot)$ is the Heaviside step function:

$$H(z) = \begin{cases} 1, & \text{if } z \geq 0, \\ 0, & \text{otherwise.} \end{cases} \quad (4.17)$$

After spike generation, a soft reset is applied to the membrane potential:

$$\mathbf{V}_t \leftarrow \mathbf{V}_t - \mathbf{O}_t \odot \boldsymbol{\theta}_s, \quad (4.18)$$

where $\odot$ denotes element-wise multiplication. Since the Heaviside function is non-differentiable, surrogate gradients are used during training to enable gradient-based optimization. The hidden spiking state of each branch evolves across the 20-second EEG window as:

$$\mathbf{O}_1, \mathbf{O}_2, \dots, \mathbf{O}_T. \quad (4.19)$$

Since the final task is classification, the temporal SNN response must be summarized into a fixed-length representation. In this study, membrane-potential aggregation is used to obtain the SNN feature representation:

$$\mathbf{h}_r = \frac{1}{T} \sum_{t=1}^T \mathbf{V}_{t,r}, \quad (4.20)$$

where r denotes a branch. For the four channel-specific branches, this gives:

$$\mathbf{h}_{TP9}, \quad \mathbf{h}_{AF7}, \quad \mathbf{h}_{AF8}, \quad \mathbf{h}_{TP10}. \quad (4.21)$$

The interaction branch produces an additional representation:

$$\mathbf{h}_{int}. \quad (4.22)$$

Therefore, the complete SNN output representation is:

$$H_{\text{SNN}} = [\mathbf{h}_{TP9}, \mathbf{h}_{AF7}, \mathbf{h}_{AF8}, \mathbf{h}_{TP10}, \mathbf{h}_{int}]. \quad (4.23)$$

The four channel-specific branches provide interpretable channel-wise temporal features, while the interaction branch captures relationships across channels. This design avoids assigning channel meaning to mixed neurons after the fact. Instead, channel-specific features are produced by branches that only receive the corresponding channel input, making the later channel-wise KAN contribution analysis more logically valid.

4.2.9 Surrogate Gradient Learning

Training SNNs is difficult because the Heaviside spike function is non-differentiable. During backpropagation, the derivative of the spike function is zero almost everywhere and undefined at the threshold. To solve this problem, surrogate-gradient learning is used [8].

In the forward pass, the actual spike function is used:

$$O_t = H(V_t - \theta_s). \quad (4.24)$$

During the backward pass, the derivative of $H(\cdot)$ is replaced by a smooth sigmoid-based surrogate derivative:

$$\frac{\partial O_t}{\partial V_t} \approx \gamma \sigma(\gamma(V_t - \theta_s)) (1 - \sigma(\gamma(V_t - \theta_s))), \quad (4.25)$$

where γ controls the steepness of the surrogate function. This allows gradients to flow through the spike-generation process during training while preserving binary spike behaviour during the forward pass.

4.2.10 KAN Readout

The KAN readout is placed after the SNN backbone. Its purpose is to map the SNN-derived features into class-wise contribution vectors. Unlike a conventional linear classifier, the KAN readout uses learnable univariate functions to model nonlinear relationships between SNN features and emotion classes. For a standard linear classifier, the final logit for class j can be written as:

$$z_j = \sum_{i=1}^d w_{j,i} h_i + b_j, \quad (4.26)$$

where z_j is the logit for class j , h_i is the hidden feature, $w_{j,i}$ is a scalar weight, and b_j is the bias. This structure is simple, but each feature contributes only through

scalar multiplication. In a KAN layer, the scalar weight is replaced by a learnable univariate function:

$$z_j = \sum_{i=1}^d \phi_{j,i}(h_i) + b_j. \quad (4.27)$$

Here, $\phi_{j,i}(\cdot)$ is a learnable nonlinear function. In this project, these functions are represented using spline basis functions, allowing the readout to model nonlinear feature-class relationships while still retaining an interpretable additive structure. The proposed architecture uses separate KAN readouts for the four EEG channels and an additional KAN readout for the interaction branch. For the channel-specific branches, the KAN readouts are defined as:

$$\mathbf{c}_{TP9} = KAN_{TP9}(\mathbf{h}_{TP9}), \quad (4.28)$$

$$\mathbf{c}_{AF7} = KAN_{AF7}(\mathbf{h}_{AF7}), \quad (4.29)$$

$$\mathbf{c}_{AF8} = KAN_{AF8}(\mathbf{h}_{AF8}), \quad (4.30)$$

$$\mathbf{c}_{TP10} = KAN_{TP10}(\mathbf{h}_{TP10}). \quad (4.31)$$

The interaction branch is mapped using a separate KAN readout:

$$\mathbf{c}_{int} = KAN_{int}(\mathbf{h}_{int}). \quad (4.32)$$

Each channel-specific KAN produces a four-dimensional class contribution vector:

$$\mathbf{c}_c = [c_{c,Calm}, c_{c,Scared}, c_{c,Happy}, c_{c,Bored}], \quad (4.33)$$

where $c \in \{TP9, AF7, AF8, TP10\}$. This means that each EEG channel contributes separately to the four emotion classes. Similarly, the interaction KAN produces:

$$\mathbf{c}_{int} = [c_{int,Calm}, c_{int,Scared}, c_{int,Happy}, c_{int,Bored}], \quad (4.34)$$

which represents class-wise evidence arising from cross-channel interactions. For a branch feature vector $\mathbf{h}_r$, the contribution to class j can be written as:

$$c_{r,j} = \sum_{i=1}^{d_r} \phi_{r,j,i}(h_{r,i}), \quad (4.35)$$

where $r \in \{TP9, AF7, AF8, TP10, int\}$, d_r is the feature dimension of branch r , and $\phi_{r,j,i}(\cdot)$ is the learnable KAN function associated with feature i and class j . Using spline basis functions, each learnable function can be expressed as:

$$\phi_{r,j,i}(h_{r,i}) = \sum_{m=1}^M a_{r,j,i,m} B_m(h_{r,i}), \quad (4.36)$$

where $B_m(\cdot)$ denotes the m -th spline basis function and $a_{r,j,i,m}$ is a trainable coefficient. During training, these coefficients are updated through backpropagation along with the trainable SNN parameters. The final class logit vector is obtained by adding the contribution vectors from the four channel-specific KAN readouts and the interaction KAN readout:

$$\mathbf{z} = \mathbf{c}_{TP9} + \mathbf{c}_{AF7} + \mathbf{c}_{AF8} + \mathbf{c}_{TP10} + \mathbf{c}_{int} + \mathbf{b}, \quad (4.37)$$

where $\mathbf{b}$ is the class bias vector. The contribution vectors are added element-wise, so each component of $\mathbf{z}$ corresponds to one emotion class:

$$\mathbf{z} = [z_{Calm}, z_{Scared}, z_{Happy}, z_{Bored}]. \quad (4.38)$$

The predicted class probabilities are computed using the softmax function:

$$\hat{y}_j = \frac{e^{z_j}}{\sum_{k=1}^4 e^{z_k}}, \quad (4.39)$$

where j represents one of the four emotion classes. The final predicted class is then obtained as:

$$\hat{y} = \arg \max_j \hat{y}_j. \quad (4.40)$$

This readout structure provides two levels of interpretability. First, the four channel-specific KAN outputs indicate how each EEG channel contributes to each emotion class. Second, the interaction KAN output shows how much additional class evidence is produced by combined multi-channel activity. Therefore, the model does not rely only on isolated channel effects, but also captures cross-channel EEG patterns that may be useful for emotion recognition.

4.2.11 Spline Parameterization of KAN Functions

Each KAN function $\phi(x)$ is represented using spline basis functions. A learnable function can be written as:

$$\phi(x) = \sum_{m=1}^M a_m B_m(x), \quad (4.41)$$

where $B_m(x)$ is the m -th spline basis function, a_m is a learnable coefficient, and M is the number of spline basis functions.

For a class-specific channel contribution, this becomes:

$$c_{c,j} = \phi_{c,j}(h_c) = \sum_{m=1}^M a_{c,j,m} B_m(h_c), \quad (4.42)$$

where c is the EEG channel and j is the emotion class. The coefficient $a_{c,j,m}$ determines how much the m -th spline basis contributes to the mapping from channel feature h_c to class j .

The coefficients are updated during training through gradient descent. If the total loss is $\mathcal{L}$, then each coefficient is updated as:

$$a_{c,j,m}^{new} = a_{c,j,m}^{old} - \eta \frac{\partial \mathcal{L}}{\partial a_{c,j,m}}, \quad (4.43)$$

where η is the learning rate. Since

$$c_{c,j} = \sum_{m=1}^M a_{c,j,m} B_m(h_c), \quad (4.44)$$

the derivative with respect to a coefficient is:

$$\frac{\partial c_{c,j}}{\partial a_{c,j,m}} = B_m(h_c). \quad (4.45)$$

Using the chain rule:

$$\frac{\partial \mathcal{L}}{\partial a_{c,j,m}} = \frac{\partial \mathcal{L}}{\partial c_{c,j}} \cdot B_m(h_c). \quad (4.46)$$

Thus, if a particular spline basis is active for a channel feature value, its coefficient receives a stronger update. Over training, the KAN readout learns nonlinear class-specific functions for each EEG channel.

4.2.12 Loss Function

The model is trained using categorical cross-entropy loss. For a sample with true class label y and predicted class probabilities $\hat{y}_j$, the loss is:

$$\mathcal{L} = - \sum_{j=1}^4 y_j \log(\hat{y}_j), \quad (4.47)$$

where y_j is the one-hot encoded true label. If the correct class is r , then the loss simplifies to:

$$\mathcal{L} = -\log(\hat{y}_r). \quad (4.48)$$

The trainable parameters of the full model include the SNN weights, SNN biases, KAN spline coefficients, and output bias terms. These parameters are optimized using gradient-based learning.

4.2.13 Final Proposed Architecture

Based on the design alternatives, methodological constraints, and interpretability requirements discussed earlier, the final selected architecture is an SNN backbone with a KAN-based readout for channel-wise class contribution analysis. The overall architecture is shown in Figure 4.1. The model follows a sequential pipeline beginning with preprocessed DER-VREEG EEG windows and ending with four-class emotion prediction.

The input to the model is a fixed-length EEG window containing four channels: TP9, AF7, AF8, and TP10. After participant-wise channel normalization, each window is converted into a spike-based representation using threshold-based spike generation and exponential latency weighting. The resulting latency-weighted current vector is passed into the SNN backbone, where membrane-potential dynamics are used to extract temporal spike-based features.

The SNN output is preserved in a channel-aware form so that the contribution of each EEG channel can be analyzed separately. Instead of using a conventional linear classifier, the final architecture uses a KAN readout. For each EEG channel, the KAN readout maps the corresponding SNN feature into a four-dimensional class

contribution vector. These vectors represent the channel's learned contribution toward Calm, Scared, Happy, and Bored. This final design was selected because it satisfies the two main objectives of the study. First, the SNN backbone provides biologically inspired temporal spike processing through membrane-potential dynamics. Second, the KAN readout provides interpretable nonlinear class contribution scores for each EEG channel. As a result, the architecture is able to perform four-class EEG emotion recognition while also allowing the final decision to be examined in terms of channel-wise evidence.

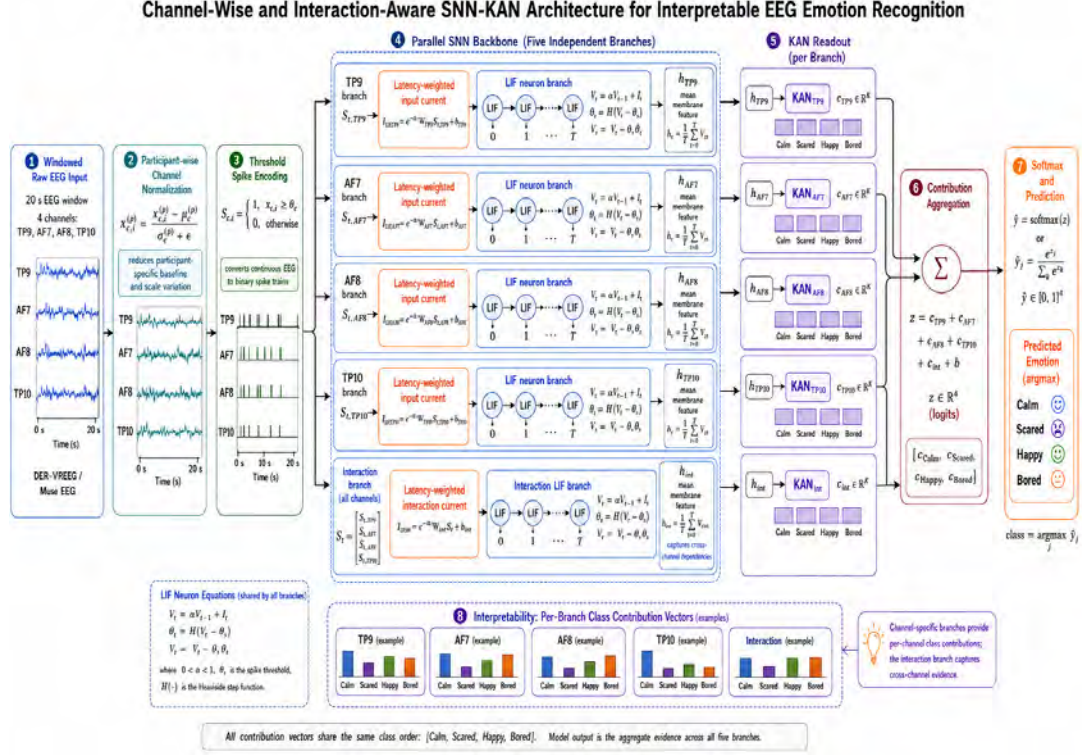


Figure 4.1: Final proposed SNN + KAN readout architecture for interpretable EEG emotion recognition. The architecture converts normalized EEG windows into spike-based representations, processes them through an SNN backbone, and uses a KAN readout to produce channel-wise class contribution vectors before final emotion prediction.

It should be noted that the channel-wise contribution values produced by the KAN readout represent model-level evidence rather than direct biological causality. Therefore, the architecture supports interpretability in the machine-learning sense, but it does not claim to identify causal neural mechanisms behind emotional states.

4.3 Data Collection

This study does not involve new data collection. Instead, it uses the publicly available DER-VREEG dataset introduced by Suhaimi et al. [12]. The dataset was created for emotion recognition using EEG signals collected during virtual-reality-based emotional stimulation. This dataset was selected despite its low channel count because it evoked emotional stimuli using a realistic VR-powered environment rather

than videos displayed on a screen that is commonly used in other datasets.

The EEG signals were recorded using a Muse 2016 headset through the Mind Monitor application. The headset contains four EEG electrodes placed at TP9, AF7, AF8, and TP10, with reference at Fpz. The dataset includes four emotion classes: Calm, Scared, Happy, and Bored. The data were recorded in CSV format, and each row contains participant information, emotion classification label, and raw EEG values from the four channels.

The use of an existing dataset is suitable for this project because the thesis focuses on model design, spike encoding, SNN-based temporal processing, and KAN-based interpretability rather than experimental EEG acquisition.

4.3.1 Data Cleaning

The raw DER-VREEG dataset contains missing EEG values marked as invalid entries. Missing values may occur due to poor electrode contact, temporary signal loss, or recording errors.

Let the original dataset be represented as:

$$D = \{(x_i, y_i, p_i)\}_{i=1}^N, \quad (4.49)$$

where x_i is the EEG sample, y_i is the emotion label, and p_i is the participant identifier. If any EEG channel value in x_i is missing or invalid, that sample is excluded:

$$D_{clean} = \{(x_i, y_i, p_i) \in D \mid x_i \text{ contains no invalid EEG value}\}. \quad (4.50)$$

This choice is reasonable because the number of invalid values is very small. Dropping such rows avoids introducing artificial EEG values through interpolation and keeps the first experimental pipeline simple and reproducible.

After cleaning, all EEG channel values are converted into numerical format. The class labels are also encoded into numerical class IDs:

$$Calm \rightarrow 0, \quad Scared \rightarrow 1, \quad Happy \rightarrow 2, \quad Bored \rightarrow 3. \quad (4.51)$$

4.3.2 Data Transformation

Data transformation includes normalization, temporal windowing, label encoding, and spike encoding.

Normalization

Raw EEG values in the DER-VREEG dataset contain large absolute values and strong channel-level offsets. Feeding raw EEG values directly into the model may cause unstable training and may allow amplitude differences between participants or channels to dominate the learning process. Therefore, normalization is applied

before windowing and spike encoding.

In this study, participant-wise channel normalization is used. This means that each EEG channel is normalized separately for each participant. For participant p , channel c , and time step t , the normalized EEG value is computed as:

$$\tilde{x}_{t,c}^{(p)} = \frac{x_{t,c}^{(p)} - \mu_c^{(p)}}{\sigma_c^{(p)} + \epsilon}, \quad (4.52)$$

where $x_{t,c}^{(p)}$ is the raw EEG value, $\mu_c^{(p)}$ is the mean value of channel c for participant p , $\sigma_c^{(p)}$ is the standard deviation of channel c for participant p , and ϵ is a small constant used to avoid division by zero.

This normalization strategy reduces participant-specific baseline differences caused by factors such as electrode contact, headset placement, scalp characteristics, and individual physiological variation. It also preserves channel identity because TP9, AF7, AF8, and TP10 are normalized separately rather than being mixed together. As a result, the model learns from relative deviations within each participant and channel instead of relying on raw amplitude offsets.

Participant-wise channel normalization does not force all values into the exact same fixed range, such as $[0, 1]$ or $[-1, 1]$. Instead, it centers each participant-channel distribution around zero and scales it by its own standard deviation. This is suitable for the proposed model because the following spike-encoding stage only needs normalized relative activity values for threshold-based spike generation, not fixed-range probability values.

Temporal Windowing

After normalization, the continuous samples are segmented into fixed-length windows. Let L be the window length. A window beginning at time index r is defined as:

$$X_r = [\tilde{x}_r, \tilde{x}_{r+1}, \dots, \tilde{x}_{r+L-1}], \quad (4.53)$$

where each $\tilde{x}_t \in R^4$. Therefore,

$$X_r \in R^{L \times 4}. \quad (4.54)$$

In this thesis, the EEG recordings are grouped by participant and emotion class before windowing. This ensures that windows do not accidentally cross from one participant to another or from one emotion label to another. This step is important because mixing samples across participant or class boundaries would create invalid temporal windows.

Spike Encoding

Each normalized EEG window is converted into spike form using the exponentially weighted latency-based encoding method described earlier. The final spike-encoded representation has the form:

$$S_r \in \{0, 1\}^{L \times 4}. \quad (4.55)$$

This transformation makes the data suitable for the SNN backbone.

4.3.3 Data Integration

This project uses only one dataset: DER-VREEG. Therefore, no multi-source data integration is required. The dataset already contains participant identifiers, class labels, and EEG channel values in a single CSV file.

However, internal integration is still performed during preprocessing. Participant IDs, emotion labels, and four EEG channels are aligned so that each window has the correct label and participant information. The final processed sample can be represented as:

$$(X_r, y_r, p_r), \quad (4.56)$$

where X_r is the EEG window, y_r is the class label, and p_r is the participant ID.

4.3.4 Data Reduction

No aggressive dimensionality reduction method is applied because the dataset already contains only four EEG channels. Applying dimensionality reduction could remove channel identity, which would conflict with the interpretability goal of the project.

Instead, the main form of data reduction is temporal windowing and feature summarization inside the SNN. Raw sequential samples are converted into fixed-length windows, and the SNN later summarizes the temporal spike response into channel-aware features. This keeps the data compact while preserving channel-level information.

The decision not to reduce the four EEG channels is intentional. Since the proposed KAN readout explains class contributions at the channel level, each original channel must remain identifiable throughout the modelling pipeline.

4.3.5 Summary of Preprocessed Data

After preprocessing, the final dataset consists of cleaned, normalized, fixed-length EEG windows. Each window contains EEG values from four channels and is assigned one emotion label. The final input shape for one sample is:

$$X \in R^{L \times 4}, \quad (4.57)$$

and after spike encoding:

$$S \in \{0, 1\}^{L \times 4}. \quad (4.58)$$

The target label belongs to one of four classes:

$$y \in \{Calm, Scared, Happy, Bored\}. \quad (4.59)$$

For the implemented experiments, the cleaned dataset contains 4,408,320 valid EEG samples after removing rows with missing values. Using fixed-length windowing, the final windowed dataset contains 1,722 EEG windows, each with 2560 time steps at 128 Hz for 20 seconds and 4 EEG channels:

$$X_{final} \in R^{1722 \times 2560 \times 4}. \quad (4.60)$$

The corresponding label vector is:

$$y_{final} \in R^{1722}. \quad (4.61)$$

A participant-wise split is used to reduce data leakage. Participants in the training set, validation set, and test set are kept separate so that the model is evaluated on participants not seen during training.

4.4 Implementation of Selected Design

The selected design is implemented as a complete EEG emotion-recognition pipeline. The implementation consists of preprocessing, spike encoding, SNN feature extraction, KAN readout classification, training, and evaluation.

4.4.1 Step 1: Dataset Loading

The raw CSV file is loaded using Python data-processing tools. The required columns are:

$$Participant, Classification, RAW_TP9, RAW_AF7, RAW_AF8, RAW_TP10. \quad (4.62)$$

The participant column is used for grouping and participant-wise splitting. The classification column is used as the emotion label. The four raw EEG columns are used as input channels.

4.4.2 Step 2: Cleaning and Label Encoding

Invalid EEG values are removed. The four class names are mapped into numerical labels. The cleaned EEG matrix is then separated from the label vector and participant identifiers.

4.4.3 Step 3: Normalization

The EEG channels are normalized to reduce scale variation. Participant-wise normalization is preferred because EEG values can differ significantly across participants. This step produces normalized EEG values that are more suitable for spike encoding and model training.

4.4.4 Step 4: Window Generation

The normalized EEG samples are grouped according to participant and emotion class. Fixed-length windows are created within each group. This ensures that a window contains temporally consistent samples from the same participant and the same emotional label.

If we want windows of a duration of 20 seconds, at 128 Hz, the window length is $L = 20 \times 128 = 2560$

4.4.5 Step 5: Spike Encoding

Each normalized EEG window is converted into binary spike form using the exponentially weighted latency-based encoder. This step produces a spike tensor that can be passed into the SNN backbone.

The encoder is designed to keep the relationship between signal magnitude and spike timing interpretable. Stronger normalized responses are more likely to generate earlier or stronger spike activity, while weaker responses generate fewer or later spikes.

Chapter 5

Result Analysis

5.1 Performance Evaluation

This section evaluates the performance of the proposed EEG emotion-recognition framework. The evaluation focuses on classification performance, model feasibility, reliability, and interpretability. Since the task is a four-class classification problem, the model predicts one of four emotional states: Calm, Scared, Happy, and Bored.

The main evaluation criteria used in this study are accuracy, macro F1-score, weighted F1-score, Cohen's kappa, Matthews correlation coefficient, confusion matrix analysis, and interpretability of channel-wise class contribution. Accuracy measures the overall proportion of correct predictions. However, accuracy alone is not sufficient for EEG emotion recognition because class-wise performance and chance-level agreement also need to be considered. Therefore, macro F1-score, weighted F1-score, Cohen's kappa, and Matthews correlation coefficient are also used.

Accuracy is defined as:

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}. \quad (5.1)$$

Precision, recall, and F1-score are computed as:

$$Precision = \frac{TP}{TP + FP}, \quad (5.2)$$

$$Recall = \frac{TP}{TP + FN}, \quad (5.3)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \quad (5.4)$$

Macro F1-score gives equal importance to each class, while weighted F1-score considers the number of samples in each class. Cohen's kappa is used to measure agreement between predicted and true labels while accounting for chance agreement. Matthews correlation coefficient is also included because it provides a balanced evaluation of classification quality, especially when class imbalance or uneven prediction patterns are present.

The testing method follows a participant-wise evaluation strategy. Participants used for training, validation, and testing are kept separate. This is important because EEG signals are strongly participant-dependent. If windows from the same participant appear in both training and testing sets, the model may learn participant-specific patterns rather than general emotion-related patterns. Therefore, participant-wise splitting provides a more realistic evaluation of generalization.

The selected architecture was compared with alternative design solutions. The comparison is shown in Table 5.1.

Table 5.1: Performance comparison of evaluated model variants

Model Variant	Validation Accuracy	Interpretability	Observation
Plain SNN with Linear Readout	0.8142	Low	Uses spike-based temporal processing, but provides weak output-level explanation.
KAN Encoder + SNN	0.7534	Moderate	Adds nonlinear pre-processing before SNN, but does not improve classification performance.
Selected SNN + KAN Readout	0.8892	High	Achieves the best observed accuracy and provides channel-wise class contribution.

The selected SNN + KAN readout model achieved the highest observed accuracy among the tested variants. It is also important because the selected model also satisfies the interpretability objective of the study. Unlike the plain SNN model, the selected architecture can decompose the final prediction into channel-wise contribution vectors for TP9, AF7, AF8, and TP10. Therefore, we will be focusing on analyzing the performance of this model in this section.

The learning behaviour of the selected SNN + KAN readout model is shown in Figure 5.1 and Figure 5.2. The training and validation accuracy curves show that the model learns progressively over the training epochs. At the beginning of training, both training and validation accuracy are low, indicating that the model has not yet learned discriminative EEG patterns. After the first few epochs, both curves increase sharply, showing that the SNN backbone and KAN readout start learning useful representations from the spike-encoded EEG windows.

From the accuracy curve, the validation accuracy improves steadily and reaches its best value of approximately 0.8892. The training accuracy is slightly higher than the validation accuracy near the final epochs, which is expected because the model is optimized directly on the training data. However, the validation accuracy follows a similar upward trend instead of collapsing, suggesting that the model does not simply memorize the training set. This indicates that the selected architecture is

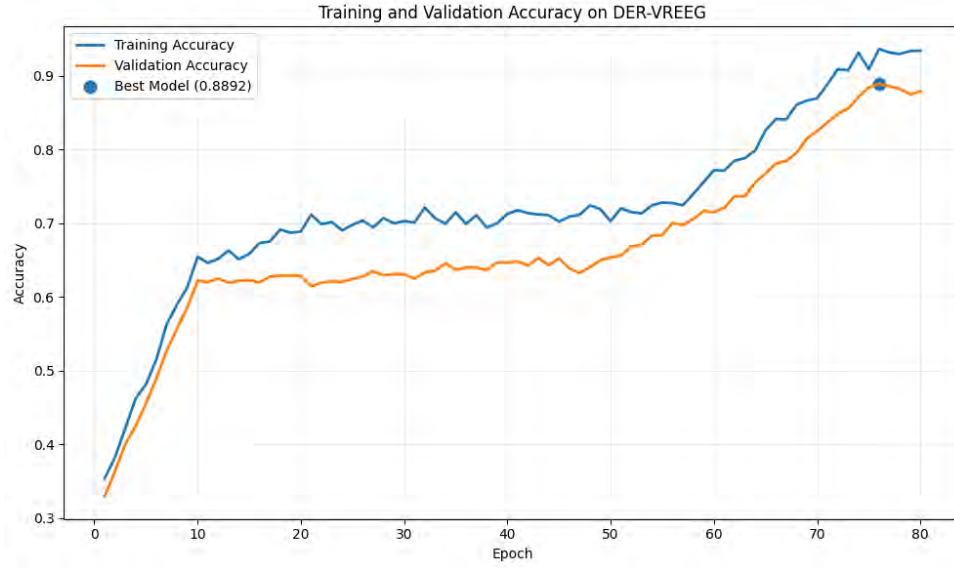


Figure 5.1: Training and validation accuracy of the selected SNN + KAN readout model on DER-VREEG.

able to learn emotion-related patterns that generalize reasonably well to the validation set.

The loss curve in Figure 5.2 provides further evidence of the model’s learning behaviour. Both training and validation loss decrease over time. The training loss decreases more smoothly, while the validation loss shows small fluctuations during the middle epochs. These fluctuations are expected in EEG classification because EEG signals are noisy, participant-dependent, and difficult to separate perfectly. Nevertheless, the overall downward trend of validation loss indicates that the model continues to improve during training.

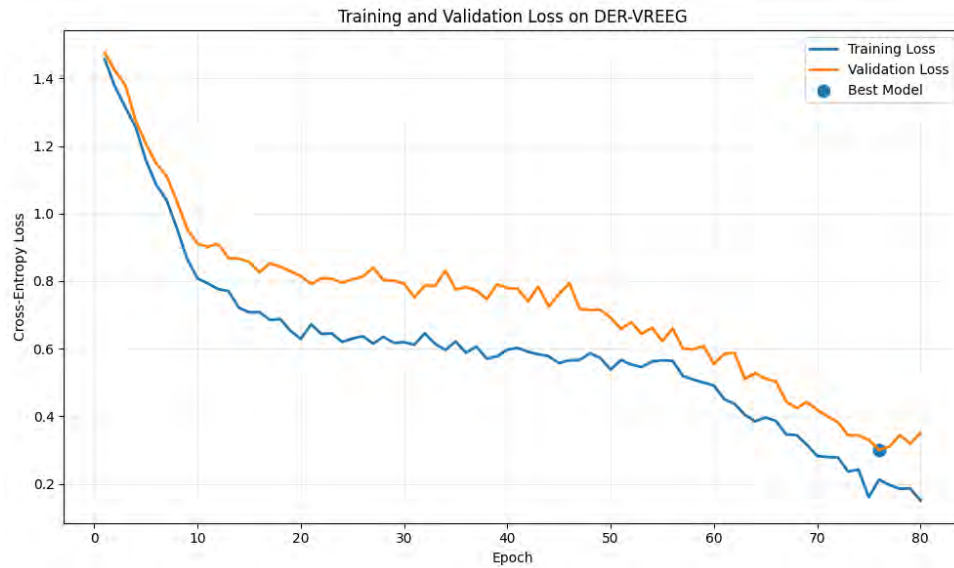


Figure 5.2: Training and validation loss of the selected SNN + KAN readout model on DER-VREEG.

The best model is selected at the epoch where the validation performance is highest. After this point, further training may improve the training accuracy but does not necessarily improve validation performance. This justifies using validation-based model selection rather than simply using the final training epoch. The small gap between training and validation curves suggests that the model has some degree of generalization ability, although the presence of a gap also indicates that overfitting should still be monitored.

The per-class precision, recall, and F1-score of the selected model are shown in Figure 5.3. This figure provides a more detailed view of the classification performance across the four emotional classes. Accuracy alone cannot show whether the model performs equally well for all classes. Therefore, per-class precision, recall, and F1-score are important for understanding class-wise behaviour.

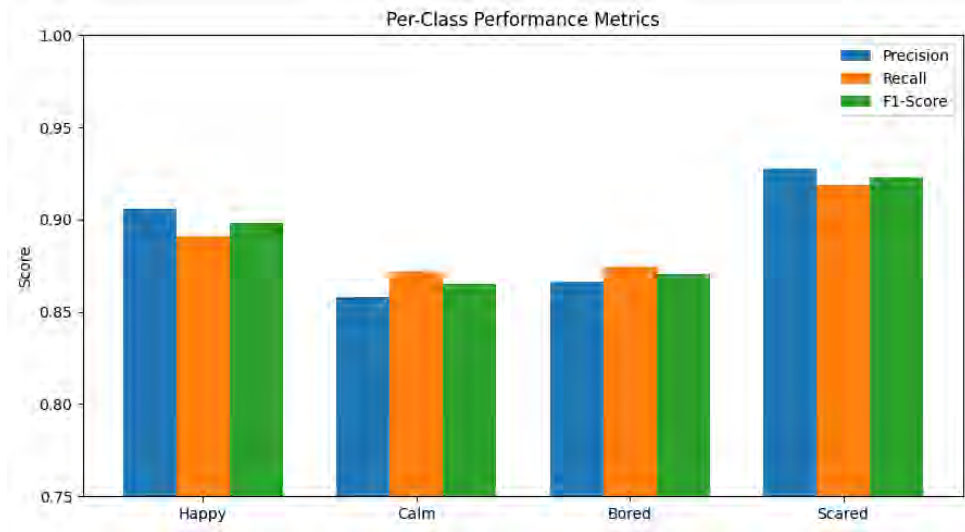


Figure 5.3: Per-class precision, recall, and F1-score of the selected SNN + KAN readout model.

The results show that the Scared class achieves the strongest performance among the four classes, with precision, recall, and F1-score all above 0.90. This suggests that the EEG patterns associated with the Scared class are more distinguishable for the proposed model. The Happy class also achieves strong performance, with all three metrics close to 0.90. This indicates that the model can identify Happy EEG windows with relatively high reliability.

The Calm and Bored classes show slightly lower but still stable performance. Their F1-scores are close to 0.86–0.87, indicating that these classes are more difficult to distinguish than Scared and Happy. This may be because Calm and Bored are both relatively low-arousal emotional states, and their EEG patterns may be more similar than high-arousal states such as Scared. The Bored class shows balanced precision and recall, suggesting that the model does not heavily overpredict or underpredict this class.

Overall, the per-class results indicate that the selected SNN + KAN readout model performs consistently across all four classes. The model does not rely on only one

dominant class, and all classes achieve relatively high F1-scores. This supports the use of macro-level evaluation metrics and confirms that the selected architecture is not only accurate overall but also reasonably balanced across emotional categories.

The performance trend can be linked to the architectural design. The threshold-based spike encoder converts normalized EEG activity into discrete spike events, while exponential latency weighting emphasizes earlier threshold-crossing activity within each window. The SNN backbone then processes these spike-based temporal representations through membrane-potential dynamics. Finally, the KAN readout learns nonlinear mappings from channel-aware SNN features to class-wise contribution scores. The strong validation accuracy and balanced per-class performance suggest that this combination is effective for the DER-VREEG four-class emotion-recognition task.

5.2 Analysis of Design Solutions

The first alternative was a plain SNN with a linear readout. This design was useful because it directly tested whether spike-based temporal processing could classify EEG emotion windows. Its advantage is simplicity and biological motivation. However, its linear readout provides limited nonlinear modelling capacity and does not clearly explain how each EEG channel contributes to the final class prediction.

The second alternative was a KAN encoder followed by an SNN. This design placed the KAN before spike encoding or before SNN processing. The motivation was to allow nonlinear transformation of EEG features before spike-based temporal learning. However, this design did not improve performance in the observed experiments. It also made the interpretation less direct because the KAN transformation occurred before spike processing rather than at the final decision stage.

The third and selected design was the SNN backbone with a KAN readout. This architecture uses the SNN for temporal spike-based feature extraction and uses the KAN readout for nonlinear classification and interpretability. The selected model is more suitable for the research objective because it combines biologically inspired temporal processing with channel-wise class contribution analysis.

The strengths and weaknesses of the evaluated designs are summarized in Table 5.2. Based on this comparison, the SNN + KAN readout architecture was selected as the final design. The selection was not based only on accuracy. It was chosen because it provides the best balance between performance, feasibility, and interpretability.

5.3 Final Design Adjustments

The final design was adjusted based on the observed performance and methodological clarity. The first adjustment was made in the spike-encoding stage. An earlier version considered a sigmoid-based spike probability formulation. However, this introduced unnecessary complexity and made the encoding harder to justify. The final

Table 5.2: Analysis of design solutions

Design Solution	Strengths	Weaknesses	Feasibility
Plain SNN	Simple, biologically inspired, supports spike-based temporal processing.	Limited interpretability and weak nonlinear readout.	High
KAN Encoder + SNN	Allows nonlinear EEG transformation before SNN processing.	Did not improve performance and makes interpretation less direct.	Moderate
SNN + KAN Readout	Best observed accuracy, biologically inspired processing, channel-wise contribution analysis.	Performance improvement is moderate; interpretation is model-level, not causal.	High

version uses a simpler threshold-based spike encoder followed by exponential latency weighting.

The final spike generation rule is:

$$s_{t,c} = \begin{cases} 1, & \text{if } \tilde{x}_{t,c} \geq \theta_e, \\ 0, & \text{otherwise.} \end{cases} \quad (5.5)$$

The exponentially weighted channel current is:

$$I_c = \sum_{t=0}^{T-1} e^{-\lambda t} s_{t,c}. \quad (5.6)$$

This design is more explainable because it directly connects normalized EEG activity, threshold crossing, spike generation, and latency-sensitive current formation.

The second adjustment was made in the SNN membrane equation. A separate reset term was removed because it was redundant with the soft reset operation. The final membrane dynamics are written as:

$$V_t = \alpha V_{t-1} + I_t, \quad (5.7)$$

$$O_t = H(V_t - \theta_s), \quad (5.8)$$

$$V_t \leftarrow V_t - O_t \theta_s. \quad (5.9)$$

This makes the model formulation cleaner and avoids double-counting the reset operation.

The third adjustment was the placement of the KAN module. Instead of placing KAN before spike encoding, the final design places KAN after the SNN as a readout module. This decision was made because the final objective is not only to transform EEG signals, but also to explain the final decision. By placing KAN after the SNN,

each channel-specific SNN feature can be mapped into a class-wise contribution vector.

Thus, for each channel c :

$$\mathbf{c}_c = KAN_c(h_c), \quad (5.10)$$

where:

$$\mathbf{c}_c = [c_{c,Calm}, c_{c,Scared}, c_{c,Happy}, c_{c,Bored}]. \quad (5.11)$$

The final logits are obtained as:

$$\mathbf{z} = \mathbf{c}_{TP9} + \mathbf{c}_{AF7} + \mathbf{c}_{AF8} + \mathbf{c}_{TP10} + \mathbf{b}. \quad (5.12)$$

The fourth adjustment was the use of adaptive gradient-based optimization to improve convergence and model accuracy. Since the proposed architecture contains both SNN parameters and KAN spline coefficients, stable optimization is important. A fixed learning-rate gradient descent method may update all parameters with the same step size, which can be inefficient when different parts of the model learn at different rates. Therefore, an adaptive optimizer was used to update the trainable parameters more effectively during training.

Let Θ represent all trainable parameters of the model, including the SNN weights, SNN biases, KAN spline coefficients, and class-bias terms. The training objective is to minimize the cross-entropy loss:

$$\mathcal{L}(\Theta) = - \sum_{j=1}^4 y_j \log(\hat{y}_j), \quad (5.13)$$

where y_j is the true one-hot encoded label and $\hat{y}_j$ is the predicted probability for class j .

At each training step, the gradient of the loss with respect to the model parameters is computed as:

$$g_t = \nabla_{\Theta} \mathcal{L}_t(\Theta). \quad (5.14)$$

To improve convergence, the optimizer maintains an exponentially decaying average of the gradients and squared gradients. These are computed as:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \quad (5.15)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \quad (5.16)$$

where m_t is the first-moment estimate, v_t is the second-moment estimate, and β_1 and β_2 are decay parameters. Bias-corrected estimates are then computed as:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad (5.17)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}. \quad (5.18)$$

The parameters are updated using:

$$\Theta_{t+1} = \Theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}, \quad (5.19)$$

where η is the learning rate and ϵ is a small constant used for numerical stability.

This adaptive update is useful for the proposed architecture because the SNN and KAN readout have different parameter behaviours. The SNN parameters are affected by surrogate-gradient learning through spike-generation functions, while the KAN readout updates spline coefficients that control nonlinear edge functions. Adaptive optimization allows each parameter to receive an update magnitude based on its own gradient history, which helps improve training stability and convergence.

In addition to adaptive optimization, validation-based model selection was used. During training, the model was evaluated on the validation set after each epoch. The final saved model was selected from the epoch with the best validation accuracy rather than simply using the last epoch. This can be written as:

$$\Theta^* = \arg \max_{\Theta_t} Accuracy_{val}(\Theta_t), \quad (5.20)$$

where Θ^* represents the parameters of the selected best model. This adjustment prevents the final model from being chosen after unnecessary additional training if the validation performance stops improving.

Together, adaptive optimization and validation-based model selection improved the convergence behaviour of the selected SNN + KAN readout model. As shown in the training curves, the validation loss decreased over time and the validation accuracy reached its best value before the final epoch. Therefore, the final reported model corresponds to the best validation checkpoint rather than an arbitrary final training state.

These adjustments made the final model more consistent with the research aim: interpretable low-channel EEG emotion recognition using an SNN backbone and KAN readout.

5.4 Statistical Analysis

The statistical analysis in this study focuses on evaluating the final SNN + KAN readout model using classification-based statistical measures. Since the task is a four-class emotion-recognition problem, accuracy alone is not sufficient to judge model performance. Therefore, the model is evaluated using accuracy, macro F1-score, Cohen’s kappa, Matthews correlation coefficient, and confusion matrix analysis.

The overall statistical performance of the final model is shown in Figure 5.4. The model achieved an accuracy of 0.8891 and a macro F1-score of 0.8892. The similarity between accuracy and macro F1-score indicates that the model performs consistently across the four emotion classes rather than performing well only on one or two

dominant classes. This is important because the DER-VREEG task involves four emotional states, and a reliable model should maintain balanced performance across Calm, Scared, Happy, and Bored.

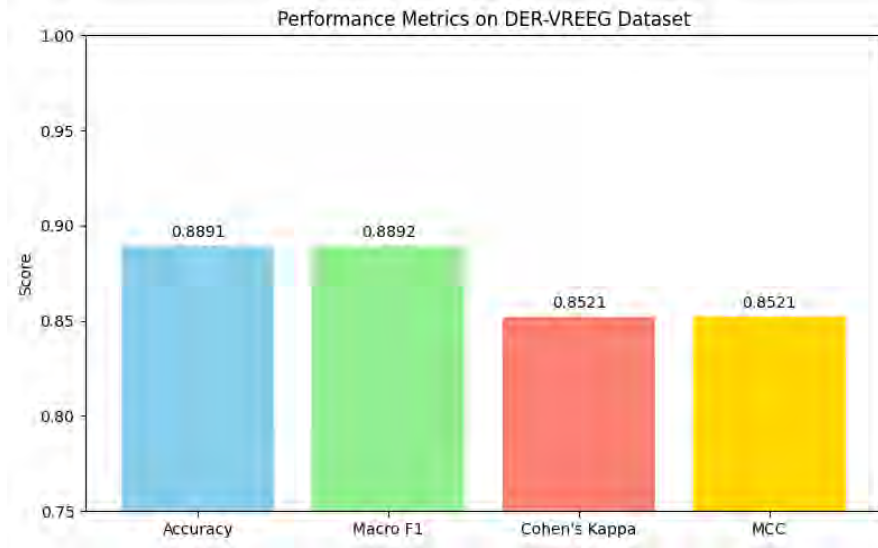


Figure 5.4: Overall performance statistics of the final SNN + KAN readout model on the DER-VREEG dataset.

Cohen’s kappa and Matthews correlation coefficient were also used to evaluate the model beyond simple accuracy. Cohen’s kappa measures the agreement between the predicted and true labels while accounting for agreement that may occur by chance. It is defined as:

$$\kappa = \frac{p_o - p_e}{1 - p_e}, \quad (5.21)$$

where p_o is the observed agreement and p_e is the expected agreement by chance. The final model achieved a Cohen’s kappa value of 0.8521. Since the task has four classes, the chance-level accuracy is approximately 25%. A high kappa value therefore indicates that the model’s predictions are substantially better than chance-level classification.

Matthews correlation coefficient (MCC) was also computed because it is a balanced measure that considers the full confusion matrix. The final model achieved an MCC value of 0.8521. This value supports the result obtained from Cohen’s kappa and indicates strong agreement between predicted and actual labels across the multi-class classification task.

The confusion matrix of the final model is shown in Figure 5.5. The diagonal elements represent correct predictions, while the off-diagonal elements represent misclassifications. The model correctly classified 384 Happy samples, 375 Calm samples, 376 Bored samples, and 396 Scared samples. These values show that the model performs strongly across all four classes, with the highest number of correct predictions observed for the Scared class.

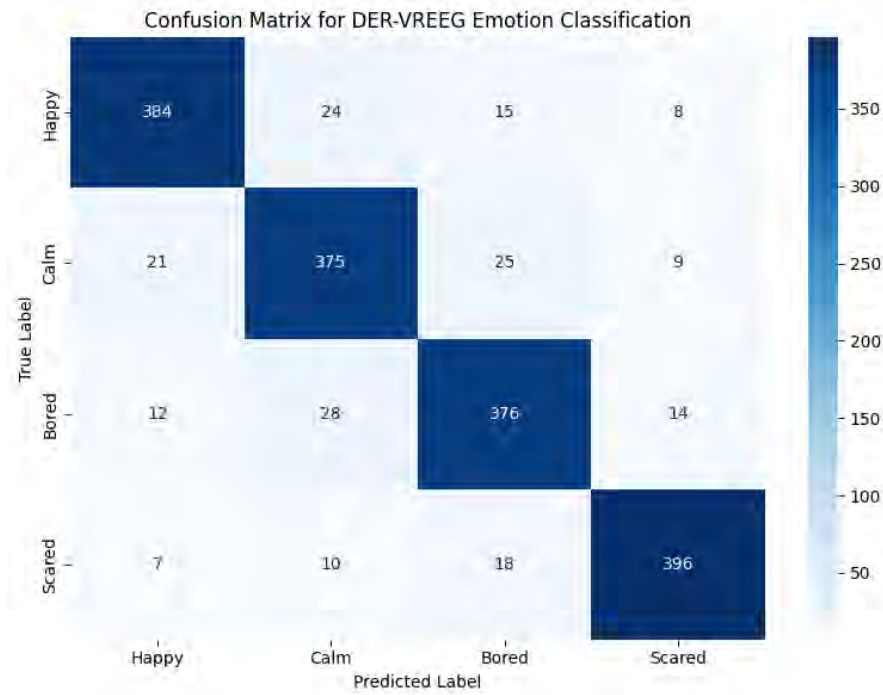


Figure 5.5: Confusion matrix of the final SNN + KAN readout model for four-class DER-VREEG emotion classification.

The confusion matrix also shows the main misclassification patterns. Happy is most frequently confused with Calm, with 24 Happy samples predicted as Calm. Calm is most frequently confused with Bored, with 25 Calm samples predicted as Bored. Bored is most frequently confused with Calm, with 28 Bored samples predicted as Calm. These patterns suggest that Calm and Bored are relatively more difficult to separate, which is reasonable because both can be considered lower-arousal emotional states. In contrast, the Scared class shows the strongest separability, with 396 correct predictions and relatively few misclassifications into other classes.

The statistical results are summarized in Table 5.3.

Table 5.3: Statistical summary of the final SNN + KAN readout model

Metric	Value	Interpretation
Accuracy	0.8891	High overall proportion of correct predictions.
Macro F1-score	0.8892	Balanced class-wise performance across four emotions.
Cohen's kappa	0.8521	Strong agreement beyond chance-level prediction.
Matthews correlation coefficient	0.8521	Strong balanced multi-class classification performance.

Overall, the statistical analysis shows that the final SNN + KAN readout model performs reliably on the DER-VREEG emotion-recognition task. The high accuracy and macro F1-score demonstrate strong predictive performance, while the high

Cohen’s kappa and MCC values confirm that the model is not merely benefiting from chance agreement or class imbalance. The confusion matrix further supports this conclusion by showing strong diagonal dominance and limited misclassification across classes. At the same time, the remaining confusion between Calm and Bored indicates that emotionally similar or low-arousal states remain more challenging to distinguish.

5.5 Comparisons and Relationships

The proposed SNN + KAN readout model was compared from two perspectives. First, it was compared with internal design alternatives developed during this study. Second, it was compared conceptually with existing EEG emotion-recognition approaches from the literature. This comparison is important because the purpose of the proposed model is not only to improve classification performance, but also to provide biologically inspired temporal processing and channel-wise interpretability.

Among the internal model variants, the selected SNN + KAN readout achieved the strongest validation performance. As shown earlier, the selected model obtained an accuracy of 0.8891 and a macro F1-score of 0.8892. The closeness between these two values indicates that the model performs consistently across the four emotional classes rather than relying on one dominant class. In addition, the model achieved Cohen’s kappa and Matthews correlation coefficient values of 0.8521, indicating strong agreement beyond chance and balanced multi-class classification performance.

The internal comparison is summarized in Table 5.4. The plain SNN with a linear readout achieved a validation accuracy of 0.8142. This shows that spike-based temporal processing alone is useful for the DER-VREEG task. However, the plain SNN has weak output-level interpretability because its linear readout does not explicitly decompose the prediction into channel-wise class contributions. The KAN encoder + SNN variant achieved a lower validation accuracy of 0.7534, suggesting that placing KAN before SNN processing did not improve the model in this setting. In contrast, the selected SNN + KAN readout achieved the highest validation accuracy of 0.8892 while also providing the most direct interpretability structure.

The relationship between the final model and the confusion matrix further supports the selection of the SNN + KAN readout architecture. The confusion matrix shows strong diagonal dominance, with 384 correctly classified Happy samples, 375 correctly classified Calm samples, 376 correctly classified Bored samples, and 396 correctly classified Scared samples. This indicates that the model is not only accurate overall, but also capable of separating all four emotional classes reasonably well. The remaining errors mainly occur between Calm and Bored, which may be expected because both classes can represent relatively lower-arousal emotional states.

The per-class precision, recall, and F1-score results also support the effectiveness of the selected model. The Scared class shows the strongest performance, with all three metrics above 0.90. Happy also performs strongly, with precision, recall, and F1-score close to 0.90. Calm and Bored show slightly lower but still stable performance, with F1-scores around 0.86–0.87. This suggests that the selected architecture does

Table 5.4: Internal comparison of evaluated model variants

Model Variant	Validation Accuracy	Interpretability	Main Observation
Plain SNN with Linear Readout	0.8142	Low	Learns spike-based temporal features, but does not provide explicit channel-wise class evidence.
KAN Encoder + SNN	0.7534	Moderate	Adds nonlinear transformation before SNN processing, but reduces validation performance in this experiment.
Selected SNN + KAN Readout	0.8892	High	Provides the best validation performance and produces channel-wise class contribution vectors.

not overfit to a single class and maintains balanced class-wise behaviour.

Compared with the original DER-VREEG study, the proposed model follows a different methodological direction. The original work mainly evaluated conventional machine-learning classifiers such as SVM, KNN, FANN, DRF, GBM, and Naive Bayes on the DER-VREEG dataset [12]. Those models are useful for establishing the feasibility of emotion recognition from low-channel wearable EEG. However, they do not use spike-based temporal processing or KAN-based channel-wise contribution analysis. Therefore, the present work should not be treated simply as a direct leaderboard comparison unless the same preprocessing, split strategy, and evaluation setting are used. Instead, the contribution of this study is the design of an interpretable SNN-KAN architecture for the same low-channel EEG emotion-recognition problem.

Compared with attention-based EEG emotion-recognition models, the proposed architecture shares the objective of identifying meaningful EEG-channel information. Attention-based methods usually assign weights to channels or features to indicate their relative importance. However, attention weights do not always directly represent class-specific evidence. In contrast, the KAN readout in this study produces a separate four-dimensional contribution vector for each EEG channel. This allows TP9, AF7, AF8, and TP10 to contribute differently to Calm, Scared, Happy, and Bored, as illustrated for the Scared class in the following contribution-curve visualization.

To illustrate the channel-wise interpretability provided by the KAN readout, Figure 5.6 presents representative KAN contribution curves toward the Scared class for the four EEG channels. The horizontal axis represents the channel-specific SNN feature, while the vertical axis represents the KAN contribution toward the Scared

class. A positive contribution indicates that the channel feature supports the Scared prediction, whereas a lower or negative contribution indicates weaker support for that class.

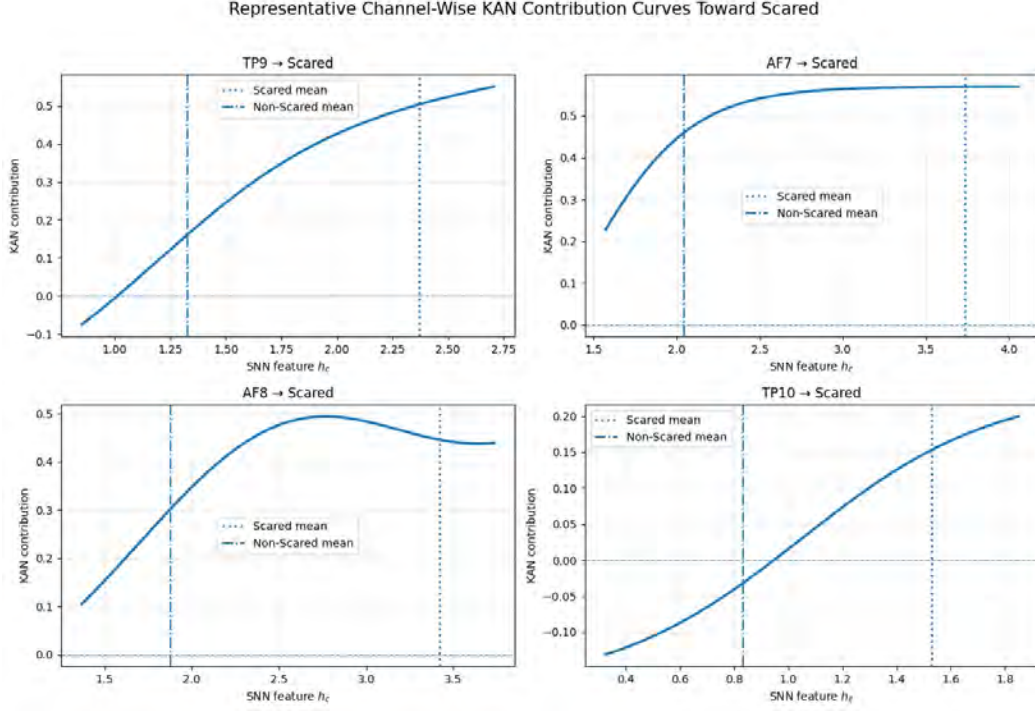


Figure 5.6: Channel-wise KAN contribution curves toward the Scared class. The curves illustrate how TP9, AF7, AF8, and TP10 can contribute differently to the Scared prediction through the KAN readout.

The curves show that the four EEG channels do not contribute identically toward the Scared class. AF7 shows the steepest positive contribution pattern, indicating that higher AF7 SNN feature values strongly push the model toward Scared. AF8 also shows a strong positive contribution, but its curve rises more gradually, suggesting a smoother and more distributed contribution. TP9 shows a moderate increasing pattern, meaning that it supports the Scared class when its feature value increases, but its influence is less dominant than AF7 and AF8. TP10 shows the weakest and more saturating contribution pattern, indicating that it provides supportive evidence only at higher feature values.

This comparison demonstrates why the KAN readout is more informative than a conventional linear classifier. Instead of producing only a final class score, the KAN readout allows each EEG channel to express a distinct nonlinear relationship with a target class. For the Scared class, the visualization suggests that frontal channels AF7 and AF8 provide stronger class-supporting evidence, while temporal channels TP9 and TP10 provide more moderate or supportive contributions. This is consistent with the strong per-class performance observed for Scared, where the final model achieved the highest class-wise precision, recall, and F1-score.

It should be noted that this visualization is used to demonstrate the interpretability mechanism of the final SNN + KAN readout design. The contribution values should

be interpreted as model-level evidence rather than biological causality. Therefore, a stronger AF7 or AF8 contribution does not prove that those scalp locations alone caused the Scared emotional state; it only indicates that, within the model’s learned representation, those channel features provide stronger evidence for the Scared class.

Compared with graph-based EEG emotion-recognition models, the proposed model is simpler and more suitable for the DER-VREEG setting. Graph neural networks are powerful when datasets contain many EEG electrodes because they can model spatial relationships between brain regions. However, DER-VREEG contains only four EEG channels. In this low-channel setting, preserving the individual contribution of each channel is more practical than constructing a complex graph structure. Therefore, the proposed model emphasizes temporal spike-based processing and channel-wise contribution analysis instead of multi-level electrode graph modelling.

Compared with existing SNN-based EEG models, the proposed model adds a more explicit decision-level interpretability mechanism. SNNs provide biologically inspired temporal processing through spike events and membrane-potential dynamics. However, a standard SNN classifier does not automatically explain how each EEG channel supports each emotion class. The KAN readout addresses this limitation by transforming channel-specific SNN features into class-wise contribution scores.

The overall relationship between different modelling approaches is summarized in Table 5.5.

Overall, the comparisons show that the final SNN + KAN readout model provides the best balance among the evaluated solutions. It improves validation accuracy over the internal alternatives, achieves strong macro F1-score, Cohen’s kappa, and MCC values, and provides a clearer interpretability mechanism than a conventional readout. Therefore, the selected model is more suitable for the research objective of interpretable low-channel EEG emotion recognition.

5.6 Discussions

The results show that the selected SNN + KAN readout architecture provides the best performance among the evaluated internal variants. The final model achieved an accuracy of 0.8891, a macro F1-score of 0.8892, Cohen’s kappa of 0.8521, and Matthews correlation coefficient of 0.8521. These results indicate that the proposed architecture performs strongly on the DER-VREEG four-class emotion-recognition task. The high macro F1-score is especially important because it shows that the model does not achieve high accuracy by favouring only one or two classes; instead, it maintains balanced performance across Calm, Scared, Happy, and Bored.

The comparison with internal alternatives shows that the placement of the KAN module is important. The plain SNN with a linear readout achieved a validation accuracy of 0.8142, while the KAN encoder + SNN variant achieved 0.7534. In contrast, the selected SNN + KAN readout achieved 0.8892 validation accuracy. This suggests that applying KAN after SNN feature extraction is more effective than

Table 5.5: Comparison of the proposed model with related modelling approaches

Approach	Main Strength	Main Limitation	Relation to This Study
Traditional Machine Learning	Simple, efficient, and useful for baseline EEG classification.	Often depends on handcrafted features and limited temporal modelling.	Used as background from the original DER-VREEG study.
CNN/Deep Learning Models	Learn nonlinear EEG representations automatically.	Often behave as black-box models and may require larger datasets.	Supports the need for learned EEG representations.
Attention-Based EEG Models	Highlight channel or feature importance.	Attention scores may not directly represent class-wise evidence.	Motivates channel-level interpretability.
Graph-Based EEG Models	Capture spatial relationships between EEG electrodes.	More suitable for datasets with many electrodes than four-channel EEG.	Related but less appropriate for DER-VREEG’s low-channel setting.
Plain SNN	Provides biologically inspired temporal spike processing.	Linear readout gives weak interpretability.	Internal baseline with 0.8142 validation accuracy.
KAN Encoder + SNN	Adds nonlinear transformation before SNN processing.	Reduced validation performance in this experiment.	Internal alternative with 0.7534 validation accuracy.
Selected SNN + KAN Readout	Combines spike-based processing, strong validation performance, and channel-wise class contribution.	Interpretability is model-level evidence, not biological causality.	Final selected architecture with 0.8892 validation accuracy.

applying KAN before spike-based processing for this dataset. One possible reason is that the SNN first extracts temporal spike-based features from EEG windows, and the KAN readout then learns nonlinear class-specific mappings from these features. In comparison, using KAN before SNN processing may transform the EEG signals before spike encoding in a way that does not benefit temporal spike learning.

The training and validation curves also support the effectiveness of the selected model. The validation accuracy increased steadily during training and reached its best value near the later epochs, while the validation loss generally decreased over time. This indicates that the model was able to learn discriminative EEG representations rather than remaining near chance-level performance. The gap between training and validation performance remained acceptable, suggesting that the model generalized reasonably well within the selected validation setting.

The confusion matrix provides further insight into the model’s behaviour. The diagonal values are dominant across all four classes, showing that most samples were classified correctly. The Scared class achieved the highest number of correct predictions, indicating that the EEG patterns associated with this class were more separable for the model. Happy also showed strong classification performance. The main remaining confusion occurred between Calm and Bored, which may be expected because both can be interpreted as relatively lower-arousal emotional states. This suggests that the model is better at distinguishing high-arousal or more distinct emotional responses than separating emotionally similar states.

The per-class precision, recall, and F1-score results confirm that the model performs consistently across the four classes. Scared achieved the strongest performance, with all three metrics above 0.90. Happy also performed close to 0.90 across the three metrics. Calm and Bored achieved slightly lower but still strong F1-scores, around 0.86–0.87. These results show that the model does not depend on a single dominant class and can recognize all four emotional states with relatively balanced effectiveness.

The most important contribution of the selected design is not only its classification performance but also its interpretability. The KAN readout allows the final prediction to be decomposed into channel-wise class contribution vectors. This means that the model can show how TP9, AF7, AF8, and TP10 contribute toward Calm, Scared, Happy, and Bored before the final logits are combined. This is more informative than a conventional classifier that only produces a final class label or probability distribution. Therefore, the selected architecture supports the objective of interpretable low-channel EEG emotion recognition.

However, the interpretation must still be handled carefully. A high contribution from a particular EEG channel does not prove that the corresponding brain region caused the emotional state. The contribution values represent model-level evidence learned from the available EEG data. Therefore, the interpretability provided by the KAN readout should be understood as computational explanation, not causal neuroscience evidence or clinical diagnosis.

The biological motivation of the model is meaningful but also limited. The SNN backbone processes information through spike events and membrane-potential dynamics, which provides a biologically inspired modelling direction. The threshold-based spike encoding and exponential latency weighting also give the input representation a spike-timing interpretation. However, the DER-VREEG dataset has a low sampling rate, and the model is implemented on conventional computing hardware rather than neuromorphic hardware. Therefore, this study does not claim to reproduce real biological neural firing or achieve hardware-level neuromorphic efficiency. The SNN is used as a biologically inspired temporal classifier.

The results also show that adding model complexity does not automatically improve performance. The KAN encoder + SNN model performed worse than the plain SNN, even though it introduced an additional nonlinear transformation. This suggests that the position and role of KAN are important. In this study, KAN is more effective as a readout after the SNN than as a preprocessing or encoding component before spike-based learning. This supports the final design decision of using the SNN as the temporal feature extractor and KAN as the interpretable nonlinear classifier.

Overall, the selected SNN + KAN readout architecture achieves the best balance among the tested models. It provides strong classification performance, maintains biological motivation through spike-based SNN processing, and supports interpretability through KAN-based channel-wise class contribution analysis. The model should still be understood as a research prototype for interpretable low-channel EEG emotion recognition rather than a clinically deployable emotion-recognition system. Future work should test the architecture on additional EEG emotion datasets, evaluate subject-independent generalization more extensively, compare against stronger deep-learning baselines, and investigate whether the spike-based design can be implemented efficiently on neuromorphic hardware.

Chapter 6

Conclusion

6.1 Summary of Findings

This study presented an interpretable EEG-based emotion-recognition framework using a Spiking Neural Network backbone and a Kolmogorov–Arnold Network readout. The proposed model was evaluated on the DER-VREEG dataset for four emotional classes: Calm, Scared, Happy, and Bored. The final architecture used participant-wise channel normalization, threshold-based spike encoding, exponentially weighted latency current formation, SNN-based temporal processing, and KAN-based channel-wise class contribution analysis.

The experimental results showed that the selected SNN + KAN readout model achieved strong classification performance, with an accuracy of 0.8891, macro F1-score of 0.8892, Cohen’s kappa of 0.8521, and Matthews correlation coefficient of 0.8521. The confusion matrix showed strong diagonal dominance across all four classes, indicating that most samples were classified correctly. The Scared and Happy classes showed the strongest performance, while Calm and Bored showed slightly higher confusion, which may be due to their relatively similar lower-arousal characteristics.

The comparison with internal alternatives showed that the final SNN + KAN readout model performed better than the plain SNN with linear readout and the KAN encoder + SNN variant. This indicates that placing the KAN module after the SNN is more effective for this task than using KAN as a preprocessing or encoding component. The results also showed that the proposed architecture can provide both strong predictive performance and interpretable channel-wise class contribution information.

6.2 Contributions to the Field

The main contribution of this study is the design of an interpretable SNN-KAN framework for low-channel EEG emotion recognition. Instead of using a conventional black-box classifier, the proposed model produces channel-wise class contribution vectors, allowing the final prediction to be examined in terms of how TP9, AF7, AF8, and TP10 contribute to Calm, Scared, Happy, and Bored.

A second contribution is the integration of spike-based temporal processing with KAN-based nonlinear readout. The SNN backbone provides biologically inspired processing through spike events and membrane-potential dynamics, while the KAN readout provides a more interpretable nonlinear mapping from channel-aware SNN features to class-level evidence.

A third contribution is the use of a simple and explainable spike-encoding strategy. The final encoder avoids unnecessary probability-based transformations and instead uses threshold-based spike generation followed by exponential latency weighting. This makes the encoding process easier to interpret and more consistent with the biological motivation of SNNs.

Finally, this study contributes an experimental analysis of different SNN-KAN design choices on the DER-VREEG dataset. The findings suggest that KAN is more effective as a post-SNN readout than as a pre-SNN encoder for this specific low-channel EEG emotion-recognition task.

6.3 Recommendations for Future Work

Future work should evaluate the proposed SNN + KAN readout framework on additional EEG emotion-recognition datasets to test its generalizability beyond DER-VREEG. Since EEG signals are highly subject-dependent, future studies should also perform more extensive subject-independent validation and cross-dataset testing.

The model can also be extended by exploring more advanced spike-encoding strategies, adaptive thresholds, and different SNN neuron models. Further work may investigate whether the threshold-based latency encoder can be improved without sacrificing interpretability.

Another important direction is deeper interpretability analysis. The channel-wise KAN contribution vectors provide model-level explanations, but future work should compare these explanations with established EEG neuroscience findings or expert interpretation. This would help determine whether the learned channel contributions are neurophysiologically meaningful or mainly model-specific.

Future research may also compare the proposed model with stronger baselines, including CNNs, attention-based models, graph neural networks, and transformer-based EEG models under the same preprocessing and evaluation protocol. This would provide a clearer understanding of the relative strengths and limitations of the SNN-KAN approach.

Finally, since SNNs are conceptually related to neuromorphic computing, future work could explore implementation on neuromorphic or edge-computing hardware. This would help determine whether the proposed architecture can support low-power, real-time EEG emotion recognition in wearable or embedded systems.

Bibliography

- [1] V. I. Arnold, “On Functions of Three Variables,” *Doklady Akademii Nauk SSSR*, vol. 114, pp. 679–681, 1957.
- [2] A. N. Kolmogorov, “On the Representation of Continuous Functions of Many Variables by Superposition of Continuous Functions of One Variable and Addition,” *Doklady Akademii Nauk SSSR*, vol. 114, pp. 953–956, 1957.
- [3] W. Maass, “Networks of Spiking Neurons: The Third Generation of Neural Network Models,” *Neural Networks*, vol. 10, no. 9, pp. 1659–1671, 1997. DOI: 10.1016/S0893-6080(97)00011-7 [Online]. Available: [https://doi.org/10.1016/S0893-6080\(97\)00011-7](https://doi.org/10.1016/S0893-6080(97)00011-7)
- [4] I. Sturm, S. Lapuschkin, W. Samek, and K.-R. Müller, “Interpretable Deep Neural Networks for Single-Trial EEG Classification,” *Journal of Neuroscience Methods*, vol. 274, pp. 141–145, 2016. DOI: 10.1016/j.jneumeth.2016.10.008 [Online]. Available: <https://doi.org/10.1016/j.jneumeth.2016.10.008>
- [5] R. T. Schirrmeister et al., “Deep Learning with Convolutional Neural Networks for EEG Decoding and Visualization,” *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017. DOI: 10.1002/hbm.23730 [Online]. Available: <https://doi.org/10.1002/hbm.23730>
- [6] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, “EEGNet: A Compact Convolutional Neural Network for EEG-Based Brain–Computer Interfaces,” *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, 2018. DOI: 10.1088/1741-2552/aace8c [Online]. Available: <https://doi.org/10.1088/1741-2552/aace8c>
- [7] Y. Yang, Q. M. J. Wu, W.-L. Zheng, and B.-L. Lu, “EEG-Based Emotion Recognition Using Hierarchical Network with Subnetwork Nodes,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 2, pp. 408–419, 2018. DOI: 10.1109/TCDS.2017.2685338 [Online]. Available: <https://doi.org/10.1109/TCDS.2017.2685338>
- [8] 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. DOI: 10.1109/MSP.2019.2931595 [Online]. Available: <https://doi.org/10.1109/MSP.2019.2931595>
- [9] K. Roy, A. Jaiswal, and P. Panda, “Towards Spike-Based Machine Intelligence with Neuromorphic Computing,” *Nature*, vol. 575, no. 7784, pp. 607–617, 2019. DOI: 10.1038/s41586-019-1677-2 [Online]. Available: <https://doi.org/10.1038/s41586-019-1677-2>

- [10] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, and A. Maida, “Deep Learning in Spiking Neural Networks,” *Neural Networks*, vol. 111, pp. 47–63, 2019. DOI: 10.1016/j.neunet.2018.12.002 [Online]. Available: <https://doi.org/10.1016/j.neunet.2018.12.002>
- [11] F. Zenke and T. P. Vogels, “The Remarkable Robustness of Surrogate Gradient Learning for Instilling Complex Function in Spiking Neural Networks,” *Neural Computation*, vol. 33, no. 4, pp. 899–925, 2021. DOI: 10.1162/neco_a_01367 [Online]. Available: https://doi.org/10.1162/neco_a_01367
- [12] N. S. Suhaimi, J. Mountstephens, and J. Teo, “A Dataset for Emotion Recognition Using Virtual Reality and EEG (DER-VREEG): Emotional State Classification Using Low-Cost Wearable VR-EEG Headsets,” *Big Data and Cognitive Computing*, vol. 6, no. 1, p. 16, 2022. DOI: 10.3390/bdcc6010016 [Online]. Available: <https://doi.org/10.3390/bdcc6010016>
- [13] W. Li, C. Fang, Z. Zhu, C. Chen, and A. Song, “Fractal Spiking Neural Network Scheme for EEG-Based Emotion Recognition,” *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 12, pp. 106–118, 2023. DOI: 10.1109/JTEHM.2023.3320132 [Online]. Available: <https://doi.org/10.1109/JTEHM.2023.3320132>
- [14] X. Li et al., “EEG Based Emotion Recognition: A Tutorial and Review,” *ACM Computing Surveys*, vol. 55, no. 4, p. 79, 2023. DOI: 10.1145/3524499 [Online]. Available: <https://doi.org/10.1145/3524499>
- [15] W. Tao et al., “EEG-Based Emotion Recognition via Channel-Wise Attention and Self-Attention,” *IEEE Transactions on Affective Computing*, vol. 14, no. 1, pp. 382–393, 2023. DOI: 10.1109/TAFFC.2020.3025777 [Online]. Available: <https://doi.org/10.1109/TAFFC.2020.3025777>
- [16] X. Wang, T. Zhang, X. Xu, L. Chen, X. Xing, and C. L. P. Chen, “Deep Learning-Based EEG Emotion Recognition: A Survey,” *Frontiers in Psychology*, vol. 14, p. 1126994, 2023. DOI: 10.3389/fpsyg.2023.1126994 [Online]. Available: <https://doi.org/10.3389/fpsyg.2023.1126994>
- [17] M. Jin, C. Du, H. He, T. Cai, and J. Li, “PGCN: Pyramidal Graph Convolutional Network for EEG Emotion Recognition,” *IEEE Transactions on Multimedia*, vol. 26, pp. 9070–9082, 2024. DOI: 10.1109/TMM.2024.3385676 [Online]. Available: <https://doi.org/10.1109/TMM.2024.3385676>
- [18] C. Liu et al., *A Comprehensive Survey on EEG-Based Emotion Recognition: A Graph-Based Perspective*, 2024. arXiv: 2408.06027. [Online]. Available: <https://arxiv.org/abs/2408.06027>
- [19] Z. Liu et al., *KAN: Kolmogorov-Arnold Networks*, 2024. arXiv: 2404.19756. [Online]. Available: <https://arxiv.org/abs/2404.19756>
- [20] A. Ghadekar et al., *Multimodal Learning with Spiking Neural Networks and Kolmogorov-Arnold Networks Enhanced through Modified SHAP Visualization*, Emerging conference/preprint reference. Verify final publication details before submission if cited., 2025.

- [21] L. F. Herbozo Contreras et al., “KAN-EEG: Towards Replacing Backbone-MLP for an Effective and Efficient EEG Seizure Detection System,” *Royal Society Open Science*, vol. 12, no. 3, p. 240 999, 2025. DOI: 10.1098/rsos.240999 [Online]. Available: <https://doi.org/10.1098/rsos.240999>
- [22] A. Joshi, P. S. Matharu, L. Malviya, M. Kumar, and A. Jadhav, “Advancing EEG Based Stress Detection Using Spiking Neural Networks and Convolutional Spiking Neural Networks,” *Scientific Reports*, vol. 15, p. 26 267, 2025. DOI: 10.1038/s41598-025-10270-0 [Online]. Available: <https://doi.org/10.1038/s41598-025-10270-0>
- [23] D. Kruttsylo et al., *Hybrid Spiking-Kolmogorov-Arnold Networks*, Workshop/preprint reference. Verify final publication details before submission if cited., 2025.
- [24] D. Kudithipudi et al., “Neuromorphic Computing at Scale,” *Nature*, vol. 637, pp. 801–812, 2025. DOI: 10.1038/s41586-024-08253-8 [Online]. Available: <https://doi.org/10.1038/s41586-024-08253-8>
- [25] Z. Liu et al., “KAN: Kolmogorov-Arnold Networks,” in *International Conference on Learning Representations*, 2025. [Online]. Available: <https://openreview.net/forum?id=Ozo7qJ5vZi>
- [26] J. Ping, B. Xu, X. Wang, W. Zhang, Z. Gao, and A. Song, “KAN-GCNN: EEG-Based Emotion Recognition with a Kolmogorov-Arnold Network-Enhanced Graph Convolutional Neural Network,” in *Proceedings of the 2025 5th International Conference on Robotics and Control Engineering*, ser. RobCE '25, Linköping, Sweden: Association for Computing Machinery, 2025, pp. 44–49. DOI: 10.1145/3733774.3735333 [Online]. Available: <https://doi.org/10.1145/3733774.3735333>
- [27] C. E. Valderrama and A. Sheoran, “Identifying Relevant EEG Channels for Subject-Independent Emotion Recognition Using Attention Network Layers,” *Frontiers in Psychiatry*, vol. 16, p. 1 494 369, 2025. DOI: 10.3389/fpsyt.2025.1494369 [Online]. Available: <https://doi.org/10.3389/fpsyt.2025.1494369>
- [28] E-SKAN Authors, *E-SKAN: Breaking the Efficiency-Accuracy Frontier in Event-Driven Spiking Kolmogorov-Arnold Networks*, engrXiv preprint, 2026. [Online]. Available: <https://engrxiv.org/preprint/view/6365/version/8302>