EEG SIGNAL CLASSIFICATION USING FEW-SHOT LEARNING

A Thesis Submitted
in Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

in

Computer Science

Submitted by

Chirag Ahuja

2K20/PhDCO/13

Under the Supervision of

Dr Divyashikha Sethia

Associate Professor
Department of Software Engineering

To the

Department of Computer Science and Engineering

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi 110042, India.

February, 2026

Acknowledgement

I would like to express my heartfelt gratitude to my supervisor, Dr Divyashikha Sethia, Department of Software Engineering, for her invaluable guidance, steadfast motivation and continued support. Her insights, thoughtful feedback and academic rigour have significantly strengthened the depth and clarity of this work. The intellectual freedom and confidence she instilled in me have been instrumental in shaping both this thesis and my broader research perspective.

I extend my sincere thanks to Prof. Anil Singh Parihar, Head, Department of Computer Science and Engineering, for his support and for facilitating the necessary resources for this work. His encouragement and administrative guidance ensured a smooth academic progression throughout this research. I am also grateful to Prof. Prateek Sharma, Hon'ble Vice-Chancellor, Delhi Technological University, for providing an excellent research environment and opportunities to pursue my PhD, fostering a culture of enquiry and academic excellence.

I am deeply indebted to my wife, Mrs Jigeesha Gulati, whose unwavering support, patience and belief in my abilities have been a constant source of strength throughout this journey. Her understanding during demanding phases of research has been invaluable. I am profoundly grateful to my parents, Mr Sushil Kumar Ahuja and Mrs Komal Ahuja, for their unconditional love, encouragement and blessings, which have been the foundation of all my achievements. I also wish to acknowledge my sister, Ms Nishtha Ahuja, for her constant encouragement, thoughtful conversations and steadfast belief in my aspirations. I also wish to acknowledge my daughter, Sahaana Ahuja, whose innocence, joy and boundless curiosity have been a continual source of inspiration and motivation during the course of this work.

I wish to express my deep appreciation to my friends for their constant encouragement, patience and unwavering support throughout this endeavour. I also extend my sincere gratitude to my colleagues across the various organisations I have had

the privilege to work with, whose professional collaboration, stimulating discussions and encouragement have contributed positively to my academic and professional growth during this journey. I would like to especially thank my fellow PhD colleague, Shikha, for her encouragement and for helping me remain motivated during challenging times.

Lastly, I extend my sincere appreciation to my lab members for their co-operation, consistent support and valuable suggestions, which contributed meaningfully to the completion of this research. The collaborative environment and exchange of ideas within the lab greatly enriched this work.

Chirag Ahuja
2K20/PhDCO/13

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE'S DECLARATION

I Chirag Ahuja hereby certify that the work which is being presented in the synopsis entitled **"EEG Signal Classification using Few-Shot Learning"** in partial fulfillment of the requirements for the award of the Degree of Doctor of Philosophy, submitted in the Department of Computer Science and Engineering, Delhi Technological University is an authentic record of my own work carried out under the supervision of Dr. Divyashikha Sethia, Associate Professor, Department of Software Engineering, Delhi Technological University. I further undertake that the work embodied in this synopsis has not been submitted for the award of any other degree elsewhere.

Chirag Ahuja
2K20/PHDCO/13

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis and the statement made by the candidate is correct to the best of our knowledge.

Signature of Supervisor **Signature of External Examiner**

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE BY THE SUPERVISOR

This is to certify that **Chirag Ahuja** (Roll no. 2K20/PHDCO/13) has carried out his research work presented in this synopsis entitled **"EEG Signal Classification using Few-Shot Learning"** for the award of Doctor of Philosophy from Department of Computer Science and Engineering, Delhi Technological University, Delhi, under my supervision. The synopsis embodies results of original work, and studies are carried out by the student himself and the contents of the synopsis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

Supervisor
Dr. Divyashikha Sethia
Associate Professor
Department of Software Engineering
Delhi Technological University

Date:

Indian Examiner

ABSTRACT

Electroencephalogram (EEG) signals are crucial in various applications, including Motor Imagery, Emotion Recognition, Visual Evoked Potentials, and Mental Workload assessment. However, EEG classification remains challenging due to limited labelled data, high noise levels, and substantial inter- and intra-subject variability. This thesis addresses these challenges by leveraging Few-Shot Learning (FSL) techniques to enable effective learning from minimal data for EEG signal classification.

To overcome key limitations, this research integrates Data Augmentation, Transfer Learning, and Self-Supervised Learning (SSL) within the FSL framework. Specifically, it focuses on (1) developing EEG-specific data augmentation strategies to mitigate data scarcity, (2) designing transfer learning methodology to facilitate efficient knowledge transfer across subjects, and (3) formulating SSL methods to enhance FSL with minimal labelled data.

Firstly, the thesis presents a comprehensive literature review of FSL techniques in EEG classification, detailing data augmentation, transfer learning, and SSL methodologies. It establishes best practices for FSL for EEG classification and provides standardized guidelines for reporting results in future studies.

Secondly, it explores data augmentation techniques to reduce dependence on limited EEG datasets by generating realistic augmented samples. It introduces Auto-Augmentation for Emotion Recognition in EEG - A Class and Subject Invariant Approach (ADAPTER) framework, which, when integrated with the cross-subject model Self-Organizing Graph Neural Network (SOGNN), achieves around 2% F1-score gain over vanilla SOGNN achieving 88.54% of cross-subject accuracy on SEED.

Thirdly, recognizing the need for improved subject adaptation, the thesis proposes a novel framework called Transfer and Robust Adaptation of New Subjects in EEG

Technology (TRANSIT-EEG). It combines a subject-specific data-augmentation - Individualised Denoising Probabilistic Model (IDPM) with Low-Rank Adaptation (LoRA) based transfer learning on an enhanced SOGNN model called Self-Organizing Graph Attention Transformer (SOGAT). Experimental evaluations on SEED and Phyaat datasets demonstrate superior cross-subject F1 scores of 91.53% and 87.78%, respectively.

Finally, the work addresses cross-device generalization in EEG classification through two Self-Supervised Learning frameworks: (i) Self-Supervised Enhancement for Multidimensional Emotion Recognition using GNNs for EEG (SS-EMERGE) and (ii) Unified Framework for Yielding EEG-based Emotion Recognition Model with Self-Supervised Learning (UNIFY-ESSL). SS-EMERGE employs a multidimensional architecture to capture temporal, spectral, and spatial features. A meiosis-based data-augmentation pretext task drives cross-subject generalization. The model delivers Macro-F1 scores of 92.35% and 81.51% on SEED and SEED-IV, respectively. When fine-tuned with only half of the labels, it still achieves 86.13% and 76.75% on SEED and SEED-IV, respectively. UNIFY-ESSL evaluates Contrastive Learning (SimCLR) and Contrastive Predictive Coding (CPC) based pretext tasks alongside a proposed data sampling strategy. The experimental results show that SimCLR attains F1-scores of 82.62%, 87.83%, and 89.05% on SEED, DEAP, and DREAMER datasets, respectively, while CPC achieves 81.35%, 82.27%, and 91.23%. It improves cross-dataset generalization, with a 1-2% performance gain on DREAMER and maintained performance on DEAP despite channel reduction, although SEED experiences a 3% F1-score drop due to significant channel reduction.

These contributions enable realistic data augmentation, rapid adaptation to new subjects for personalization, and unified modeling across datasets—advancing robust, adaptable, and generalizable EEG classification for diverse real-world applications.

List of Publications

Papers Published in International Journals

Journal 1: Chirag Ahuja, and Divyashikha Sethia. "Harnessing Few-Shot Learning for EEG Signal Classification: A Survey of State-of-the-Art Techniques and Future Directions." **Frontiers in Human Neuroscience**, 18: 1421922 (2024). (**SCIE, Publisher: Frontiers Media SA, Impact Factor: 2.4**). Doi: https://doi.org/10.3389/fnhum.2024.1421922

Journal 2: Chirag Ahuja, and Divyashikha Sethia. "TRANSIT-EEG - A Framework for Cross-Subject Classification with Subject-Specific Adaptation." **IEEE Transactions on Cognitive and Developmental Systems** (2024). (**SCIE, Publisher: IEEE, Impact Factor: 5.0**). Doi: https://doi.org/10.1109/TCDS.2025.3529669

Journal 3: Chirag Ahuja, and Divyashikha Sethia. "SS-EMERGE: Self-Supervised Enhancement for Multidomain Emotion Recognition using GNNs for EEG" **Scientific Reports** (2025). (**SCIE, Publisher: Springer Nature, Impact Factor: 3.8**). Doi: https://doi.org/10.1038/s41598-025-98623-7

Papers Published in International Conferences

Conference 1: Chirag Ahuja, and Divyashikha Sethia. "Measuring Human Auditory Attention with EEG." **In Proc. International Conference on COMmunication Systems & NETworkS (COMSNETS), Bangalore, India pp. 774-778, 2022**. (**Publisher: IEEE**). Doi: https://10.1109/COMSNETS53615.2022.9668363

Conference 2: **Chirag Ahuja**, and Divyashikha Sethia. "ADAPTER: Auto-Augmentation for Emotion Recognition in EEG—A Class and Subject Invariant Approach." **In Proc. International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, pp. 1-6, 2023**. (**Publisher: IEEE**). Doi: https://10.1109/ICCCNT56998.2023.10308153

Conference 3: **Chirag Ahuja**, and Divyashikha Sethia. "UNiFY-ESSL: UNIfied Framework for Yielding EEG-based Emotion Recognition Model with Self-Supervised Learning." **In Proc. International Conference on Brain Computer Interface & Healthcare Technologies (iCon-BCIHT), Thiruvananthapuram, India, pp. 249-256, 2024**. (**Publisher: IEEE**). Doi: https://10.1109/iCon-BCIHT63907.2024.10882427

Contents

List of Tables

List of Figures

List of Abbreviations

ADAPTER Automatic Data Augmentation and PerTubation for Emotion Recognition 14

BCI Brain-Computer Interfaces 1

BENDR BErt-inspired Neural Data Representations 12

CBAM Convolutional Block Attention Module 73

CDEER Cross-Domain EEG Emotion Recognition 94

CNN Convolution Neural Network 5

CPC Contrastive Predictive Coding 15

CSP Common Spatial Pattern 41

CV Computer Vision 2

DADA Differentiable Auto Data Augmentation 100

DCGAN Deep Convolutional Generative Adversarial Network 65

DDPM Denoising Diffusion Probabilistic Model 68

DE Differential Entropy 35

DEAP Dataset for Emotion Analysis using Physiological Signals 67

DL Deep Learning 57

DMMR Domain-aware Multi-decoder Mutual Reconstruction 13

EEG Electroencephalogram x

EMD Empirical Mode Decomposition 32

ER Emotion Recognition 20

ERP Event-Related Potential 20

ETL Explicit Transfer Learning 76

Chapter 1

Introduction

Electroencephalography (EEG) captures real-time brain activity using non-invasive and portable devices. Its high temporal resolution and ease of use make it suitable for clinical diagnostics, cognitive state monitoring, and Brain-Computer Interfacess (BCIs) [19, 20, 21, 22]. These advantages enable researchers to deploy EEG systems in practical, real-world environments. Despite these strengths, building reliable machine-learning models for EEG data presents several challenges. Researchers face high costs and effort to collect large, labelled datasets. EEG signals vary across individuals due to differences in brain structure, mental state, and electrode placement. These signals also suffer from a low signal-to-noise ratio and are non-stationary over time. Environmental noise, hardware-specific distortions, and inconsistent recording setups further reduce model robustness. To improve generalisation, researchers often combine multiple datasets. However, datasets collected from different devices introduce new problems. Each device uses a different number of electrodes, sampling rate, and preprocessing method. The preprocessing pipeline must resample signals to a fixed frequency, align them to a shared set of electrodes, and normalise amplitudes to remove device-specific biases. Without these steps, models fail to generalise across datasets and perform inconsistently on new inputs.

Few-Shot Learning (FSL) addresses many of these issues. FSL enables models to learn from only a few labelled examples by transferring knowledge from related subjects or tasks [23, 24, 25, 26]. In EEG settings, FSL reduces the burden of data collection and supports fast adaptation to new users or sessions. FSL also helps detect

rare events, such as neurological anomalies or emotional outliers, that are difficult to capture in bulk. However, standard FSL methods often struggle with EEG data. Many approaches assume stable and clean inputs, which EEG does not provide. Models that reuse features from NLP [27] or Computer Vision (CV) [28] domains perform poorly on subject-specific, noisy EEG signals. To make FSL effective for EEG, models must learn robust, transferable representations that handle device and subject variability. These methods must also adapt quickly using limited labelled data. This thesis addresses these challenges through proposed frameworks that combine data generation, subject-encoding, and label-efficient modelling to improve task-specific classification. These techniques make EEG classification more robust and practical in low-data environments.

1.1 Electroencephalogram (EEG) Overview

EEG [19] is a widely used neuroimaging technique that records electrical activity generated by neuronal populations in the brain. EEG is non-invasive, cost-effective, and provides high temporal resolution, making it an essential tool for BCIs, medical diagnostics, cognitive research, and affective computing [29, 30]. Unlike other neuroimaging modalities such as Functional Near-Infrared Spectroscopy (fNIRS) [31] and Magnetoencephalography (MEG) [32], EEG is portable and allows real-time neural signal acquisition, enabling applications in clinical and non-clinical environments. It is acquired using special sensors from manufacturers such as Emotiv [21], Biosemic [33], Neuroscan [34] and OpenBCI [4]. Figure 1.1 shows a typical setup of EEG signal acquisition, where a human subject is wearing an EEG cap consisting of various electrodes placed around different brain regions and data feeding into the computer.

1.1.1 Applications of EEG

EEG has a wide range of applications across medical, technological, and cognitive domains due to its ability to capture real-time brain activity non-invasively. The high temporal resolution of EEG makes it particularly useful for studying dynamic neural processes, enabling its integration into various practical fields. EEG is a foundational tool in neuroscience and applied research, from diagnosing neurological disorders to

Figure 1.1: EEG cap for brainwave monitoring and neural signal analysis [4].

enhancing human-computer interaction. The following are key applications of EEG:

1. **Clinical Applications**: EEG plays an important role in diagnosing and treating neurological disorders. Applications such as epilepsy monitoring to detect abnormal brain activity associated with seizures [29] use EEG signal. EEG is a fundamental component of polysomnography, helping to classify sleep stages and diagnose sleep disorders such as insomnia, narcolepsy, and sleep apnea [30] for sleep medicine. Also, EEG is employed in neurorehabilitation, assisting in the recovery of motor function following stroke or traumatic brain injury. Moreover, the recent advancements in EEG-based biomarkers have further enabled early detection of neurodegenerative diseases such as Alzheimer's and Parkinson's disease, improving patient outcomes through early intervention [22].

2. **Brain-Computer Interfacess (BCIs)**: EEG is pivotal for BCIs [35], facilitating direct communication between the brain and external devices without reliance on muscular activity. EEG-based BCIs are instrumental in assistive technologies, enabling individuals with severe motor impairments to control prosthetic limbs, wheelchairs, and communication devices [36]. EEG-based BCIs also find applications in cognitive workload monitoring, gaming, and virtual reality, where they enhance the user experience by adapting system responses to mental states.

3. **Cognitive and Affective Computing**: EEG is widely used to study cognitive

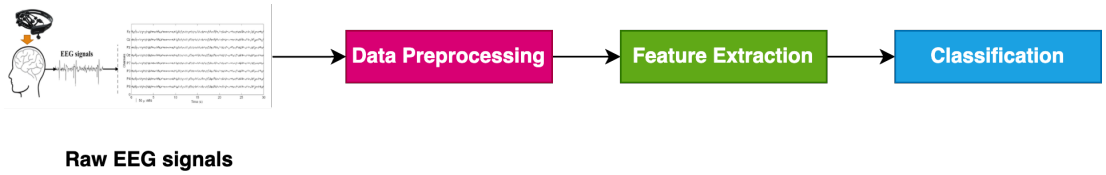


Figure 1.2: Automated EEG-based classification workflow.

functions such as attention, memory, and decision-making. By analysing EEG spectral components, researchers can infer mental states, assess cognitive workload, and track real-time attention levels [36]. EEG-based systems help personalise learning experiences in education by monitoring student engagement and cognitive load. Similarly, affective computing leverages EEG signals to detect emotions such as stress, frustration, and relaxation, contributing to mental health monitoring and adaptive human-computer interaction.

1.1.1.1 Automated EEG-based Classification Workflow

EEG signals capture brain activity as high-dimensional, multichannel time-series data recorded via scalp electrodes. The workflow typically processes these signals using a three-stage pipeline: **Preprocessing**, **Feature Extraction**, and **Classification** (Figure 1.2). Preprocessing enhances signal quality by applying baseline correction, filtering, re-referencing, epoching, and artifact removal [37, 20]. Feature extraction derives meaningful representations in time, frequency, or spatial domain. Many studies rely on frequency-domain features such as differential entropy because it effectively distinguishes cognitive and emotional states [38]. In the final stage, classification models—ranging from classical Machine Learning (ML) to deep learning—predict target outcomes such as mental workload or emotional state [1, 5, 2, 39, 11].

Evolution of EEG Classification Models

The field of EEG classification evolves through several architectural paradigms, each addressing the limitations of earlier approaches [40, 41, 42, 43]. Early studies rely on **Classical Machine Learning models** such as Support Vector Machines (SVMs), k-Nearest Neighbors (k-NN), and Random Forests. These models perform well on small-scale datasets and offer interpretability, but they critically depend on handcrafted

feature extraction. Researchers typically compute statistical and spectral features such as Differential Entropy (DE), Power Spectral Density (PSD), Wavelet Transform coefficients, and Hjorth parameters to represent EEG signals [44, 45]. While these features capture valuable frequency-domain and time-domain characteristics, selecting and engineering them is manual, subjective, and prone to suboptimal generalization.

The introduction of deep learning shifts the focus toward automated feature learning. **Convolution Neural Networks (CNNs)** mark the first major breakthrough by learning hierarchical representations directly from raw or minimally preprocessed EEG signals [46]. This development removes the need for handcrafted features and allows the model to discover informative patterns across time and channels. However, CNNs treat EEG data as a grid-like structure, which limits their ability to model long-range temporal dependencies and to account for spatial relationships among non-adjacent brain regions.

To address temporal modelling limitations, researchers adopt **Recurrent Neural Networks (RNNs)** and **Long Short-Term Memory (LSTM) networks**. These models explicitly capture sequential dependencies and perform well on short- to medium-term temporal dynamics in EEG data [47, 7]. Despite their strengths, they face challenges in learning long-range dependencies due to vanishing gradients and suffer from high computational costs because of their sequential processing nature.

To incorporate the brain's structural and functional topology, the field transitions to **Graph Neural Networkss (GNNs)**. Architectures such as Graph Convolution Networks (GCNs) [9] and Graph Attention Networks (GATs) [10] model EEG electrodes as nodes in a graph and use edge connections to encode spatial or functional relationships [48, 8]. This graph-based representation allows GNNs to capture inter-regional dependencies more effectively than CNNs or GNNs. However, their performance heavily depends on the choice of graph construction, and they lack a native mechanism for modelling long-range temporal patterns.

The current frontier in EEG classification lies in Transformer-based models. These architectures use self-attention to capture complex and global dependencies across time without relying on recurrence [49, 39, 11]. Transformers outperform previous methods in temporal modelling and offer scalability through parallel computation. Nevertheless, they are data- and resource-intensive, often requiring large datasets and

powerful hardware—constraints that hinder their use in many real-world or clinical scenarios.

1.1.2 Challenges in EEG processing

EEG signals are instrumental in decoding brain dynamics across various applications. However, practical challenges—such as data collection, model robustness, inter-subject and intra-subject variability, and dataset adaptation—have hindered the development of effective affective processing systems for real-world use.

1. **Data collection:**

   (a) *Volume and quality of data*: Advance signal processing methods require extensive, high-quality EEG datasets to function effectively. Low Signal-to-Noise-Ratio (SNR), significant inter-subject and session variability of EEG signals poses challenges during data collection, limiting less open datasets [50].

   (b) *Methodological constraints*: Acquiring high-fidelity EEG data requires sophisticated equipment, precise sensor placement, and skilled supervision. These requirements can be resource-intensive and may limit widespread data collection efforts [50].

   (c) *Labelling difficulties*: Labelling EEG data, especially for complex cognitive or affective states, demands domain expertise and is labour-intensive. This complexity makes it challenging to generate robust training datasets [50].

2. **Model robustness:**

   (a) *Adaptation to variability*: EEG signals exhibit significant variability across individuals and sessions due to physiological and psychological factors, such as fatigue or emotional states. This variability complicates model generalisation and necessitates strategies to enhance robustness [50, 51, 8].

   (b) *Real-World Conditions*: Laboratory settings often fail to replicate real-world scenarios, leading to a disconnect between experimental results

and practical applications. For instance, controlled environments may not account for movement artifacts present in naturalistic settings [51, 8, 52].

(c) *Noise Resilience*: EEG signals are susceptible to noise from muscle movements, eye blinks, and external electrical interference. While preprocessing techniques like filtering and artifact removal are essential, they are not foolproof and may inadvertently remove relevant signal components [37, 53].

3. **Inter-subject and intra-subject variability:**

(a) *Inter-Subject Variability*: Differences in brain anatomy, age, gender, and lifestyle result in unique EEG patterns for each individual, posing challenges for cross-subject generalisation [51, 8, 54, 55, 56].

(b) *Intra-subject variability*: Temporal changes within an individual, such as fluctuations in mood or alertness, lead to shifts in EEG signal characteristics over time, affecting model consistency [57, 57].

(c) *Dataset shift*: These variabilities contribute to covariate shifts, where the statistical properties of the training and test data differ, leading to degraded model performance on unseen subjects or sessions [58, 59].

4. **Device variability:**

(a) *Lack of standardisation*: The diversity in EEG recording devices, ranging from high-density systems with numerous electrodes to consumer-grade devices with fewer sensors, introduces data quality and format inconsistencies.

(b) *Electrode types*: Variations between wet and dry electrodes further complicate standardisation efforts. Wet electrodes typically offer better signal quality but are less practical for everyday use than dry electrodes.

5. **Class imbalance:**

(a) *Skewed datasets*: Many EEG datasets underrepresent certain states or conditions and suffer from class imbalance, such as DEAP [5], SEED-IV

[3], DREAMER [15]. The imbalance leads to biased models that perform well on majority classes but poorly on minority classes [60, 61, 19, 62, 63].

(b) *Class Boundaries*: Capturing the full spectrum of human cognitive or emotional states is inherently challenging due to their fluid and context-dependent nature, making clear-cut class definitions difficult [60].

6. **Temporal variability and model adaptation**: The non-stationary nature of EEG signals, with factors like ageing or emotional fluctuations influencing them, challenges static models. With this, models need continuous adaptation mechanisms to maintain performance over time.

Similar challenges exist across various domains, including computer vision, natural language processing, and audio processing, leading to Few-Shot Learning (FSL) as a dedicated research area. FSL has evolved as a transformative paradigm designed to mitigate the limitations of scarce labelled data, improve task generalisation, and reduce the risk of overfitting [25, 12]. This approach enables machine learning models to learn effectively from minimal training samples, making it particularly valuable when data collection is costly, time-consuming, or impractical.

1.2 Few-Shot Learning (FSL) Overview

FSL enables models to learn and generalise from a few examples, significantly advancing the field of machine learning [23, 24]. In light of the challenges associated with EEG classification — such as limited labelled data, inter-subject variability, and the high data acquisition costs FSL emerges as a promising solution. FSL enables models to generalise from minimal data by leveraging prior knowledge and experience, making it particularly relevant for scenarios where data acquisition is expensive, time-consuming, or impractical. It is especially beneficial in EEG-based applications, where collecting extensive labelled datasets is often challenging [64].

1.2.1 Advantages of FSL

FSL offers advantages over conventional approaches, particularly in data-scarce environments where acquiring large-scale labelled datasets is impractical [64, 23, 25,

[24, 12]. The key advantages include:

1. **Learning from Rare Cases:** A well-designed model can handle diverse real-world scenarios, including those that are difficult to collect, simulate, or anticipate. Traditional models often struggle with rare events due to insufficient training samples. FSL mitigates this limitation by enabling models to adapt to infrequent occurrences using minimal supervision, making it particularly valuable in anomaly detection, medical diagnosis, and conducting studies of complex systems such as the brain.

2. **Reducing Data Collection Effort and Computational Cost:** Acquiring and labelling large datasets is often resource-intensive, particularly in biomedical imaging, robotics, and remote sensing. FSL reduces the dependency on extensive data collection efforts by enabling models to transfer knowledge from previously learned tasks. This capability decreases the computational cost associated with model retraining and facilitates efficient adaptation to new datasets without requiring vast amounts of labelled data.

1.2.2 Challenges of Few-Shot Learning (FSL)

Despite its advantages, FSL presents several challenges stemming from limitations in data availability, model generalisation, and algorithmic constraints. Song et al. [12] identifies key difficulties, including the inability to assess data distributions due to limited training samples accurately, the sensitivity of feature reuse mechanisms, the restricted generalisability of learnt representations to future tasks, and the inherent limitations posed by single-modality learning frameworks. Addressing these issues requires advancements in data augmentation, meta-learning, domain adaptation, and multi-modal learning techniques to ensure that FSL models maintain high performance across diverse applications [64]. These challenges can be broadly categorised into four primary areas as follows:

1. **Inaccurate data distribution:** The limited number of training samples restricts the ability of machine learning models to accurately infer the underlying data distribution, resulting in suboptimal generalisation.

2. **Feature reuse sensitivity:** The reliance on prior knowledge from auxiliary datasets introduces sensitivities in feature reuse mechanisms, potentially leading to biased representations not well-adapted to novel tasks.

3. **Generality of future tasks:** Ensuring that models learn representations that generalise beyond the current task remains a fundamental challenge. Overfitting to a specific task often hampers the model's ability to adapt to new, unseen tasks.

4. **Deficiency of single-modal information:** The limited availability of information from a single modality makes feature learning less effective. While integrating multiple modalities can enhance representation learning, achieving seamless multi-modal integration remains complex and computationally demanding.

1.3 Research Gaps in Few-Shot Learning (FSL) for EEG

This thesis explores FSL for EEG based signal classification and has identified the following gaps in the current literature, laying the foundation for the rest of the thesis. Deep learning models have taken over traditional machine learning in the past decade due to their efficiency and effectiveness in all fields, including EEG processing [42, 65]. Therefore, deep learning is the foundation for most of the research gaps as follows:

1. **RG1 - Effective ordering of existing data augmentation**: The ordering of geometric transformations for EEG data augmentation (including time shifts, frequency shifts, signal flipping, and channel shuffling) creates a combinatorial challenge in determining the most effective sequence of transformations for a given dataset. While automatic data augmentation techniques using reinforcement learning have shown promise in computer vision [66], the direct application to EEG signals remains computationally intractable due to EEG data's high-dimensional, multivariate nature. Current approaches rely on manual selection of transformations or computationally expensive search methods. The field lacks efficient algorithms for automatically discovering

optimal transformation sequences that preserve class-specific characteristics while maximising data diversity, mainly when working with limited samples.

2. **RG2 - Subject-invariance guarantees of data augmentation**: The inherent entanglement of subject-specific characteristics with task-relevant features in EEG signals presents a fundamental challenge for data augmentation. Current approaches address this through a two-stage process: applying traditional signal processing techniques (like filtering and artifact removal) to remove subject-specific components, followed by data augmentation [12, 8, 48, 67, 68, 69]. However, this sequential approach has challenges: (1) there is a risk of removing task-relevant information during preprocessing, (2) it cannot adapt to individual subject characteristics, and (3) it fails to capture the complex, non-linear relationships between subject-specific and task-relevant features.

3. **RG3 - Inadequate Mechanisms for Personalized Subject Adaptation**: Personalising EEG models across subjects presents a critical challenge in balancing generalisable representations and individual-specific nuances [51, 8]. Current transfer learning methods typically demand complete model retraining for new subjects, which is computationally inefficient and prone to catastrophic forgetting [70]. Moreover, they lack mechanisms for selective parameter adaptation, making them ill-suited for low-resource, real-time applications [69]. There is a pressing need for architectures that support rapid, few-shot personalisation—enabling efficient adaptation to new users while retaining transferable knowledge from prior EEG experiences.

4. **RG4 – Inefficiencies in Few-Shot Transfer Learning for EEG Applications**: Current transfer learning techniques in EEG signal processing often rely on full-model finetuning, which is computationally intensive and time-consuming. It makes them impractical for real-time or embedded systems where rapid adaptation is crucial. Moreover, these methods typically require substantial labelled data, which is scarce in many EEG applications. FSL techniques, designed to operate with minimal data, have shown promise in addressing these challenges. For instance, Mammone et al. [71] introduced a few-shot transfer learning approach for motion intention decoding from EEG signals,

demonstrating the potential of FSL in adapting to new tasks with limited data. However, generalising such methods beyond specific tasks like emotion recognition, metal workflow detection, and many more remains under explored. There is a need to develop efficient FSL frameworks that can generalise across diverse EEG tasks and subject variations, enabling rapid and reliable adaptation in low-data setups.

5. **RG5 - Unified foundation model for EEG signal processing**: Despite rapid advancements in EEG signal processing techniques, it faces significant limitations due to heterogeneity across various EEG devices, resulting in heterogeneous datasets. This results in a lack of a foundation model that can generalise across diverse EEG tasks, datasets, and devices. The current landscape relies on task-specific approaches, requiring separate models for distinct applications like emotion recognition, sleep staging, and seizure detection from different devices [1, 8, 14, 67, 72]. Similar to the revolution in computer vision, natural language processing, and audio processing [28, 73, 74], a robust foundation model trained across various tasks, devices, and subjects can enable many real-world applications using EEG devices.

6. **RG6 – Lack of multidimensional representation learning in SSL**: Current SSL approaches, such as BErt-inspired Neural Data Representations (BENDR) [39], Generative Adversarial Network-based Self-supervised EEG Representation learning (GANSER) [75], and Self-supervised Group Meiosis Contrastive Learning (SGMC) [14], capture only subsets of EEG signal dimensions—temporal, spectral, or spatial—but rarely integrate all three. SGMC [14] used temporal meiosis-based augmentations but lacks frequency or spatial modelling. Multi-View Spectral-Spatial-Temporal Masked Autoencoder (MV-SSTMA) [76] integrated all signal dimensions but relied on generative training and lacked cross-subject evaluation. No current framework jointly optimises pretext tasks and architectures across time, frequency, and space. A unified multidimensional strategy remains missing.

7. **RG7 – Challenges in Cross-Subject Generalization within Self-Supervised Learning for EEG**: Self-supervised learning (SSL) has emerged as a powerful

tool for learning representations from unlabelled EEG data. However, existing SSL methods often struggle with cross-subject generalisation due to the high inter-subject variability inherent in EEG signals. This variability leads to models that overfit subject-specific patterns, reducing their effectiveness when applied to new individuals. Recent studies have attempted to address this by incorporating contrastive learning frameworks and attention mechanisms to enhance feature extraction [77]. Additionally, approaches like the Domain-aware Multi-decoder Mutual Reconstruction (DMMR) framework utilise multi-decoder autoencoders to promote subject-invariant feature learning [78]. Despite these advancements, developing robust SSL strategies that can effectively generalise across subjects remains an open research area. Innovations in pretext task design, data augmentation techniques, and hybrid learning objectives are needed to improve the cross-subject applicability of SSL models in EEG analysis.

1.4 Problem Statement

This thesis addresses the critical challenges of developing robust EEG-based applications with limited labelled data. The research focuses on three fundamental barriers in EEG signal processing: data scarcity, low Signal-to-Noise-Ratio (SNR), and adaptation challenges. The research proposes a comprehensive Few-Shot Learning (FSL) framework that incorporates data augmentation to mitigate data scarcity, efficient cross-subject knowledge transfer to acknowledge low Signal-to-Noise-Ratio (SNR) across subjects, and using Self-Supervised Learning (SSL) to leverage multiple datasets for improved adaptation from auxiliary sources to overcome these challenges.

1.5 Research Objectives and Contribution of the Thesis

The main objective of the thesis is to enhance the cross-subject transfer with few-shot adaptation. The thesis achieves the following objectives:

1. **Research Objective I: Literature review - To perform a systematic literature review on few-shot learning techniques for EEG signal classification.**

Contribution: This thesis conducts a comprehensive literature review on few-shot learning techniques for EEG signal classification. It covers the characteristics and techniques for strategies to achieve few-shot learning via Data Augmentation, Transfer Learning and Self-Supervised Learning (SSL). This literature survey identifies research gaps and proposes the best practices for conducting the FSL research. (Completed by **Journal 1** and **Conference 1**).

2. **Research Objective II: Data augmentation - To enhance EEG signal classification through an advanced few-shot learning framework incorporating automated data augmentation.**

   Contribution: This thesis proposes two novel data augmentation techniques: (1) **Automatic Data Augmentation and PerTubation for Emotion Recognition (ADAPTER)** formulates augmentation as an optimal sequence search problem of geometric augmentation. ADAPTER is a class-invariant augmentation technique which evaluates the classification performance of a subject-invariant model - Self Organizing Graph Neural Network (SOGNN) [8]. (2) *Individualised Denoising Probabilistic Model (**IDPM**)* is another augmentation technique that learns a subject-specific generative model, which learns to generate synthetic EEG samples for each subject and class. ADAPTER uses simple geometric transformations that are not subject-specific, which IDPM addresses, and finally yields a subject and class-invariant augmentation technique. (Completed by **Conference 2** and **Journal 2** respectively).

3. **Research Objective III: Transfer learning - To develop an adapter-based transfer learning framework for efficient few-shot EEG classification.**

   Contribution: This thesis introduces a framework called *Transfer and Robust Adaptation of New Subjects in EEG Technology (**TRANSIT-EEG**)*, which proposes and end-to-end cross-subject modelling framework. It first enhances the cross-subject model, i.e. SOGNN [8] by replacing standard GNN with GAT [10], referred to as *Self Organizing Graph Attention Transformer (**SOGAT**)* in TRANSIT-EEG. It also uses Objective II's proposed model - *Individualised Denoising Probabilistic Model (**IDPM**)* for subject-specific data augmentation. Finally, it uses Low Rank Adaptation (LoRA) to enable

efficient subject-specific adaptation while preserving the SOGAT performance of other subjects. TRANSIT-EEG enhances the cross-subject performance while enhancing subject-specific components in every modelling stage. (Completed by **Journal 2**).

4. **Research Objective IV: Self-Supervised Learning – To enhance EEG classification through a self-supervised few-shot learning framework.**

   Contribution: This objective addresses key limitations in existing self-supervised EEG classification approaches by proposing a more expressive and generalisable framework. It builds upon the Self-supervised Group Meiosis Contrastive (SGMC) framework [14], which uses a ResNet-based architecture focused on temporal modelling and applies grouped meiosis data augmentation. While effective in capturing temporal patterns, SGMC does not exploit spatial and spectral information critical to EEG signal diversity. To overcome this, the thesis proposes, *Self-Supervised Enhancement for Multidimension Emotion Recognition using Graph Neural Networks (**SS-EMERGE**)* framework, which introduces a multidimensional self-supervised architecture that integrates EEG signal temporal, spatial, and frequency representations. Inspired by multivariate attention-based models such as MV-SSTMA [76], SS-EMERGE extends SGMC by incorporating graph-based spatial learning and spectral encoding within a contrastive learning framework. This work evaluates the proposed SS-EMERGE framework extensively under subject-independent and cross-subject settings, showing notable improvements in generalisation, especially under low-data conditions. The thesis also proposes *Unify-Emotion Self Supervised Learning (**Unify-ESSL**)* framework, a unified pretraining strategy aimed at cross-device and cross-dataset EEG learning. By combining heterogeneous datasets and leveraging established pretext tasks such as SimCLR [79] and Contrastive Predictive Coding (CPC) [80], the framework sets the foundation for robust few-shot learning across devices. Experiments demonstrate that the proposed sampling strategy and joint pretraining significantly improve performance in diverse EEG environments. (Completed by **Journal 3, Conference 3**)

1.6 Significance of the Thesis

This thesis advances EEG signal processing by enhancing Few-Shot Learning (FSL) methodologies to address key challenges in EEG signal classification, including data scarcity, subject variability, and adaptation limitations. EEG-based BCIss hold great promise for assistive communication, neurorehabilitation, and cognitive monitoring applications. However, the need for large labelled datasets and model robustness across diverse users and environments limits their real-world applications.

The research contributes novel techniques to enhance EEG classification efficiency and adaptability. It tackles data scarcity by leveraging FSL and data augmentation strategies that reduce dependence on large labelled datasets while preserving neural signal integrity. A transfer learning framework improves cross-subject generalisation, minimising the need for subject-specific retraining to address inter-subject variability. Furthermore, SSL integration enables EEG models to extract meaningful representations from unlabelled data, improving performance in low-data settings. Ultimately, this thesis lays the foundation for more generalisable and data-efficient EEG classification models, bridging the gap between research and real-world applications using EEG.

1.7 Thesis Overview

This section outlines the structure of the thesis, providing an overview of each chapter's content and purpose.

Chapter 2: Background for EEG Signal Classification provides background on EEG-based signal classification, detailing fundamental concepts of EEG signals, including different paradigms and acquisition methods. It discusses the advantages of classical machine learning techniques, feature-based approaches, and deep learning methods for EEG classification. Additionally, it introduces widely used open datasets relevant to EEG signal classification in this thesis.

Chapter 3: Literature Review presents a systematic review of FSL techniques for EEG signal processing, covering data augmentation, transfer learning, and SSL approaches. It analyses existing methods, identifies research gaps, and establishes the

foundation for the proposed solutions.

Chapter 4: ADAPTER — Auto-Augmentation and PerTubation for Emotion Recognition presents a data augmentation framework, ADAPTER, designed to enhance cross-subject generalisation in EEG-based emotion recognition. The chapter outlines the ADAPTER methodology, including its augmentation formulation, class-invariant constraints, and optimization strategy. It specifically highlights the integration of ADAPTER with a subject-invariant classification model - SOGNN, and conducts extensive cross-subject validation on the SEED [1]. The chapter further includes experimental evaluations that assess the contribution of different augmentation components, compares performance against baseline augmentation methods, and provides ablation studies to validate the effectiveness of the proposed approach in improving generalisation across unseen subjects.

Chapter 5: TRANSIT-EEG: A Framework for Cross-Subject Classification with Subject Specific Adaptation introduces Transfer and Robust Adaptation of New Subjects in EEG Technology (TRANSIT-EEG), a novel end-to-end cross-subject adaptation framework designed to rapidly adapt to new subjects using only a few labelled samples. This chapter systematically outlines the methodology underpinning the TRANSIT-EEG, which integrates subject-agnostic augmentation, a robust classification model, and an efficient transfer learning strategy. The discussion begins with the proposed data augmentation technique, Individualised Denoising Probabilistic Model (IDPM), which enhances generalization across subjects by modelling intra-class variability. It then presents the subject-oriented graph-based model, Self Organizing Graph Attention Transformer (SOGAT), which leverages spatial and temporal dependencies in EEG signals. Finally, the framework incorporates lightweight subject-specific adaptation through Low Rank Adaptation (LoRA)-based fine-tuning, enabling efficient personalization with minimal overhead. The chapter concludes with an extensive ablation analysis that isolates the contribution of each component in the TRANSIT-EEG framework. Empirical evaluations on the SEED [1] and PhyAAT [2] datasets demonstrate the effectiveness of the framework in both emotion recognition and cognitive activity classification tasks.

Chapter 6: Generalisable Self-Supervised Learning for Few Shot EEG Signal Classification presents two self-supervised learning frameworks for EEG signal

classification: Self-Supervised Enhancement for Multidimension Emotion Recognition using Graph Neural Networks (SS-EMERGE) and Unify-Emotion Self Supervised Learning (Unify-ESSL). The chapter is organised in two main parts. It first details the methodology, training strategy, and architectural components of SS-EMERGE, followed by a comprehensive set of experiments on homogeneous EEG datasets, including SEED [1] and SEED-IV [3]. This section includes ablation studies, evaluation under few-shot settings, and comparisons with existing baselines. The second part of the chapter focuses on Unify-ESSL, outlining its design choices for handling heterogeneous EEG datasets. It discusses the unified pretraining strategy and balanced sampling procedures across diverse datasets such as SEED [1], DEAP [5], and DREAMER [15]. The chapter concludes with detailed experimental results that assess Unify-ESSL's cross-dataset and cross-device generalisation capabilities, supported by ablation analyses and benchmark comparisons.

Chapter 7: Conclusion, Future Work & Social Impact summarises the key contributions and findings of the research work, discussing limitations, potential future research directions, and broader social applications in areas such as healthcare and BCI. It also reflects the need for few-shot learning in other machine learning paradigms, such as federated learning for real-world applications.

Chapter 2

Background for EEG Signal Classification

This chapter dives deeper into the fundamental concepts of EEG signal processing. It first introduces the acquisition methodology of EEG signals and various EEG paradigms and applications. Gradually, it goes deeper into EEG montages and different electrodes and then introduces various EEG frequency bands and their significance. After introducing the fundamental concepts, it moves into the EEG processing workflow while introducing data processing, feature extraction and modelling, and, specifically, classification for the scope of this thesis. It sets a background of the evolution of various deep learning models for EEG signal classification while highlighting their strengths and limitations. This chapter also introduces some open EEG datasets and their intricacies that are widely used to study EEG signal classification.

2.1 Fundamentals of EEG Signals

Electroencephalography (EEG) signals are complex, multivariate time series data from multiple electrodes placed on the scalp at distinct brain regions. However, EEG recordings are highly susceptible to noise artifacts from various sources, including muscular activity such as eye blinks, limb movements, and environmental interferences [81, 82, 83]. Several key parameters characterise EEG signals, including their spatial location, amplitude, frequency, morphology, continuity, synchronisation, symmetry,

and reactivity.

2.1.1 EEG Signal Acquisition Methods

The acquisition of EEG signals involves measuring the brain's electrical activity via electrodes positioned on the scalp. It provides valuable insights into cognitive processes, neurological disorders, and machine-learning applications, such as few-shot learning-based classification.

2.1.2 EEG Paradigms and Applications

EEG-based BCIs employ various experimental paradigms to extract meaningful neural information [84]. Several key paradigms are central to this field:

1. **Motor Imagery (MI) [85]:** This paradigm involves the modulation of neural activity within sensory-motor regions when an individual imagines executing specific movements. MI-based BCIs enable direct control by interpreting the user's intended actions through EEG signals without requiring actual physical movement.

2. **Event-Related Potential (ERP) [86]:** ERPs represent specific neural responses elicited by external stimuli. A particularly significant component in BCI applications is the P300 potential, which emerges approximately 300 milliseconds after an unexpected stimulus [87]. The reliability and consistency of ERPs make them a widely adopted paradigm in BCI research.

3. **Passive paradigms:** These paradigms analyse neural activity during spontaneous brain states without requiring active user engagement. Two significant passive paradigms include:

   (a) *Emotion Recognition (ER):* This approach utilises EEG signals to classify affective states based on the Valence-Arousal-Dominance (VAD) model [5]. Researchers can infer the subject's emotional state by analysing neural responses to emotionally salient stimuli, such as audiovisual content.

(b) *Human Auditory Attention (HAA):* This paradigm quantifies attention levels based on EEG patterns [2]. It has important applications in education, workplace monitoring, and human-computer interaction, providing insights into sustained and transient attention fluctuations.

2.1.3 EEG Montages and Electrode Placement

EEG montages define the spatial arrangement of electrodes during signal acquisition, influencing the quality and interpretability of recorded brain activity. The choice of montage depends on the intended application, whether for clinical diagnostics, neurophysiological research, or BCI systems. The two primary montage techniques are:

1. **Bipolar Montage:** This configuration records the voltage difference between two adjacent electrodes, enhancing spatial resolution and reducing common-mode noise. Bipolar montages are particularly useful in detecting localised brain activity and minimising external interference [88].

2. **Referential Montage:** In this approach, each electrode records signals relative to a standard reference electrode, typically placed at a neutral site such as the mastoid (A1, A2) or ear lobes. This technique enables broader comparisons of cortical activity across different brain regions but may be more susceptible to reference noise [89, 90].

2.1.3.1 10-20 System and Electrode Placement

The 10-20 system is the most widely adopted standard for EEG electrode placement, ensuring consistency and comparability across clinical and research applications [91]. EEG manufacturers, including Biosemi [33], ESI NeuroScan [34], and Emotiv Epoc [21], extensively utilise this system. Electrodes are positioned based on proportional distances between anatomical landmarks on the scalp, standardising signal acquisition across different studies [92, 93, 94]. Figure 2.1 illustrates the electrodes corresponding to their cortical regions as follows :

- **F - Frontal Lobe:** Associated with higher-order cognitive functions, including decision-making, executive control, and voluntary motor activity.
- **T - Temporal Lobe:** Involved in auditory processing, language comprehension, and memory encoding.
- **C - Central Region:** Plays a key role in sensorimotor integration, governing movement coordination and proprioception.
- **P - Parietal Lobe:** Engaged in spatial awareness, sensory processing, and multimodal integration.
- **O - Occipital Lobe:** Primarily responsible for visual processing and interpretation of visual stimuli.

Numerical designations further specify electrode positioning relative to cerebral hemispheres: **odd numbers** indicate electrode placement on the left hemisphere, while **even numbers** indicate placement on the right hemisphere. Additionally, midline electrodes such as Fz, Cz, and Pz provide critical reference points in EEG signal interpretation [95, 96]. Reference electrodes positioned on neutral sites such as A1, A2 are commonly employed to improve signal stability and reduce noise artifacts in referential montages. The 10-20 system provides a foundational framework for EEG recording, facilitating interoperability across different EEG devices and ensuring compatibility with advanced methodologies, including high-density electrode arrays used in modern BCI research.

2.1.4 EEG Frequency Bands

Electroencephalography (EEG) signals decompose into distinct frequency bands, each corresponding to different neural states and cognitive processes. These frequency components play a crucial role in understanding brain function and aid in diagnosing neurological disorders, conducting cognitive research, and developing brain-computer interfaces.

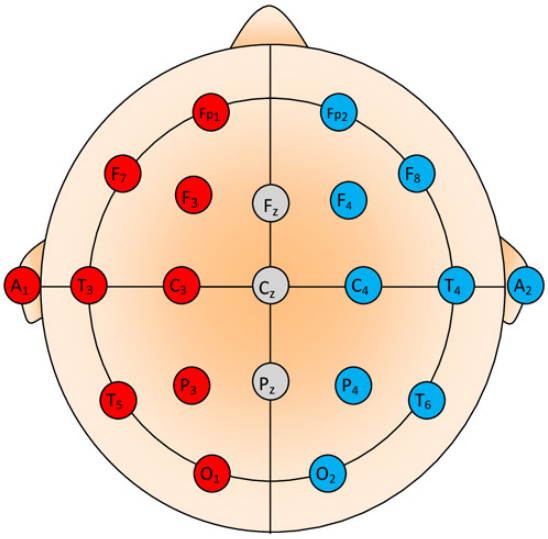


Figure 2.1: Illustration of the 10-20 EEG electrode placement system, widely used in clinical and research applications.

Waves	Frequency bands (Hz)	Behavioral Trait	Signal Waveform
Delta	0.3 – 4	Deep sleep	
Theta	4 – 8	Deep Meditation	
Alpha	8 – 13	Eyes closed, awake	
Beta	13 – 30	Eyes opened, thinking	
Gamma	30 and above	Unifying consciousness	

Table 2.1: Frequency Bands of EEG Signals.

2.1.4.1 Characteristics of EEG Frequency Bands

Researchers categorise EEG signals into five primary frequency bands: Delta, Theta, Alpha, Beta, and Gamma. Each band exhibits distinct characteristics and associates with specific behavioural and cognitive states, as summarised in Table 2.1 [97, 98, 99, 88, 100, 1, 5].

2.1.4.2 Functional Roles of EEG Frequency Bands

1. **Delta Activity ($<$ 4 Hz):** Delta waves predominantly appear in deep sleep and contribute to restorative physiological processes [101]. However, their presence in awake individuals may indicate pathological conditions such as brain injury or encephalopathy [99]. The spatial distribution of abnormal delta activity depends

on the underlying neurological impairment.

2. **Theta Activity (4–7 Hz):** Theta waves commonly emerge in adults during drowsiness or transitioning sleep states. However, excessive theta activity in alert individuals can indicate neurological dysfunction. In pediatric EEG recordings, theta waves often dominate occipital and central regions as part of standard developmental patterns [97, 98].

3. **Alpha Activity (8–13 Hz):** Alpha waves dominate the rhythm in awake adults with closed eyes, primarily localising in the occipital region. This activity typically appears symmetrical, with an amplitude range of 40–100 µV. Alpha wave attenuation upon eye-opening serves as a hallmark feature of typical EEG reactivity [93].

4. **Beta Activity ($>$ 13 Hz):** Beta waves correlate with active cognitive processing, alertness, and attentional focus. Central and frontal regions predominantly record beta waves. Excessive beta activity may also link to anxiety states or cortical excitability [88].

5. **Gamma Activity ($>$ 30 Hz):** Gamma waves contribute to higher-order cognitive functions, such as perception, problem-solving, and consciousness. Increased gamma activity occurs during tasks requiring complex neural coordination and links to memory consolidation and sensory processing [100].

Understanding EEG frequency bands is essential for clinical diagnostics and computational applications, such as feature extraction in machine-learning-based classification systems. The interactions between these frequency components provide valuable insights into neural dynamics, cognitive states, and pathological conditions.

2.2 Open EEG Datasets

EEG research relies heavily on publicly available datasets to develop and benchmark algorithms for applications such as emotion recognition, sleep staging, and seizure detection. These datasets provide standardized EEG recordings that support fair comparisons across methods and enable reproducibility. Table 2.2 summarizes

Table 2.2: Characteristics of benchmark EEG datasets.

Attribute	**DEAP [5]**	**SEED [1]**	**SEED-IV [1]**	**DREAMER [15]**	**PhyAat [2]**
Paradigm / Task	ER	ER	ER	ER	HA
#Subjects	32	15	15	23	25
#Session	32	3	3	23	25
Sampling Frequency (Hz)	512	1000	1000	128	128
Down sampled Frequency (Hz)	128	200	200	128	128
#Electrodes	32	62	62	14	14
Stimuli	Music videos	Movie clips	Movie clips	Movie clips	Listening / Writing / Resting
#Stimuli	40	15	15	18	144
Baseline Duration (seconds)	3 and 5	5	60	60	1
Stimuli Duration (seconds)	60	240	120	65	4–6
Classes	Valence, Arousal	Positive, Neutral and Negative	Happy, Sad, Fear, Neutral	Valence, Arousal, Dominance	Listening, Writing and Resting

ER: Emotion Recognition; HA: Human Attention

commonly used open-access EEG datasets, most of which focus on emotion recognition and auditory attention tasks, including SEED [1], SEED-IV [3], DREAMER [15], DEAP [5], and PhyAat [2]. This thesis uses all five datasets to benchmark few-shot learning (FSL) techniques under varying subject counts, recording frequencies, and paradigm complexities.

2.2.1 SEED Dataset

The SJTU Emotion EEG Dataset (SEED) [1] dataset served as a benchmark for EEG-based emotion recognition. Developed at Shanghai Jiao Tong University, it enabled early evaluation of deep learning methods for affective computing. Figure 2.2 shows

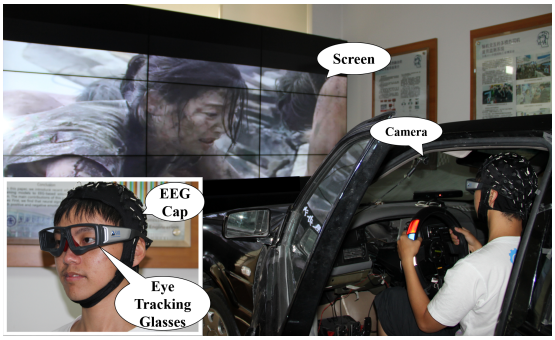


Figure 2.2: SEED data collection setup [1].

the setup where 15 participants viewed emotionally curated Chinese film clips and reported their responses via questionnaires. Researchers recorded EEG signals using a 62-channel ESI NeuroScan System [102] at 1000 Hz, then downsampled them to 200 Hz. Preprocessing included bandpass filtering (0.1–75 Hz) and denoising via a linear dynamic model. This thesis uses SEED [1] for subject adaptation and augmentation experiments, leveraging its multiple-session structure and balanced emotion classes.

2.2.2 SEED-IV Dataset

The SEED-IV [3] dataset extended SEED [1] by introducing finer-grained emotion classes and multimodal data. It retained the original 15 participants and recorded EEG using the same 62-channel ESI NeuroScan system. Participants viewed stimuli labeled as happy, sad, fear, or neutral, while wearing SMI eye-tracking glasses. EEG was sampled at 1000 Hz and downsampled to 200 Hz. This multimodal setup jointly captured eye movement and neural signals, enabling richer emotion recognition models. This thesis includes SEED-IV [3] in few-shot learning evaluations, particularly for cross-task generalization and adapter-based emotion recognition.

2.2.3 DREAMER Dataset

The DREAMER dataset [15] enabled multimodal emotion research by combining EEG and ECG recordings. It included data from 23 participants using a 14-channel Emotiv EPOC headset [21] and a Shimmer2 wireless ECG sensor. Participants watched 18 emotionally evocative film clips and rated their emotional responses using the Self-Assessment Manikin (SAM) scale [103], quantifying Valence, Arousal, and

Figure 2.3: DEAP data collection setup [5].

Dominance. EEG signals were sampled at 128 Hz, and ECG was recorded at 256 Hz. Although DREAMER provides multimodal physiological recordings, this thesis considers only the EEG modality for all experiments. DREAMER [15] is used to evaluate model generalization under constraints such as low electrode count and reduced temporal resolution.

2.2.4 DEAP Dataset

The **DEAP dataset** [5] is one of the most widely used benchmarks for EEG-based emotion recognition. Figure 2.3 illustrates the experimental setup, where researchers recorded EEG and physiological signals while participants watched emotionally evocative one-minute music video excerpts. The dataset contains recordings from 32 participants, collected using a 32-channel Biosemi ActiveTwo EEG system [33]. In addition to EEG, the recording protocol included Electrocardiography (ECG), Electrooculography (EOG), and Galvanic Skin Response (GSR). Participants were exposed to 40 video stimuli, and their emotional responses were self-reported along four dimensions: Valence, Arousal, Dominance, and Liking—each rated on a scale from 1 to 9. EEG signals were recorded at 512 Hz and then downsampled to 128 Hz for further analysis. The DEAP [5] dataset is a foundational benchmark in affective computing and has supported extensive research on multimodal and EEG-only emotion recognition. This thesis uses only the EEG modality from DEAP [5] to develop and evaluate the

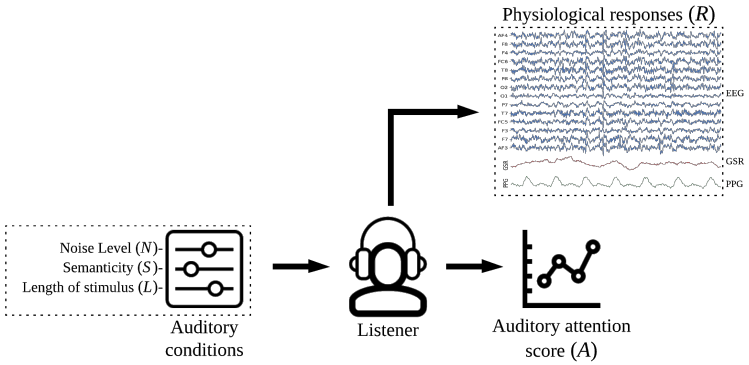


Figure 2.4: PhyAat data collection setup [2].

foundation model for emotion classification under diverse emotional stimuli.

2.2.5 PhyAat Dataset

The **PhyAat dataset** [2] was designed to study auditory attention and classify cognitive tasks such as listening, writing, and resting using EEG and physiological signals. Figure 2.4 shows the experimental setup, where participants were exposed to auditory stimuli that varied in noise level, semantic content, and duration. The dataset contains recordings from 25 university students (21 males, 4 females), acquired using a 14-channel Emotiv EPOC EEG headset [21] to monitor brain activity. In addition to EEG, physiological measurements such as Galvanic Skin Response (GSR), Photoplethysmography (PPG), and heart rate in Beats Per Minute (BPM) were recorded. Participants listened to 144 audio samples comprising semantic and non-semantic speech under six different noise conditions, simulating realistic auditory attention scenarios. Each trial was annotated with cognitive task labels, enabling task-specific classification. EEG signals were sampled at 128 Hz, making the dataset suitable for low-resolution, wearable EEG system research. The dataset also included a regression target for estimating auditory attention using Word Error Rate (WER) from a speech-to-text task. This thesis uses only the EEG modality from PhyAat [2] and focuses on classifying cognitive tasks—listening, writing, and reading—under varying auditory conditions.

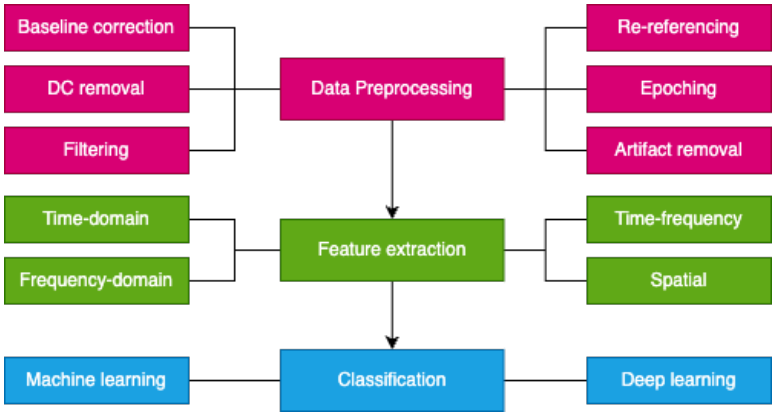


Figure 2.5: EEG signal processing workflow.

2.3 EEG Classification Workflow

EEG data undergoes a structured processing pipeline to extract meaningful information while reducing noise and artifacts. Figure 2.5 illustrates a typical flow of EEG processing. It primarily has three stages: data processing, feature extraction, and classification. The following details of each of these three stages:

2.3.1 Data Preprocessing

The initial stage of the workflow entails preprocessing raw EEG signals to eliminate unwanted noise and interference [20, 104]. This preprocessing typically comprises six sequential steps as follows:

1. **Baseline correction:** Adjusts EEG signals to account for pre-stimulus neural activity and external noise, helping isolate the brain's true response to a stimulus. Traditionally, this is achieved by subtracting the mean voltage recorded during a pre-defined baseline period from the entire epoch. However, this method assumes the baseline period is free from task-related neural activity, which may not always be true. Alternative approaches include high-pass filtering to attenuate low-frequency drifts and regression-based correction, which incorporates the baseline interval as a predictor in a generalised linear model, allowing the data to determine the extent of correction needed [20, 104].

2. **DC removal:** Eliminates low-frequency drift in EEG signals caused by direct current (DC) components, which can otherwise distort subsequent analyses. It

is essential to remove the DC component, ensuring that only relevant brain signal fluctuations remain. The high-pass filter attenuates frequencies below a specified cutoff, eliminating the DC offset [105]. A high-pass filter [106] allows frequencies higher than its cutoff frequency to pass through while attenuating lower frequencies, including the DC component (0 Hz). Applying a high-pass filter with an appropriate cutoff frequency in EEG signal processing removes slow drifts and DC offsets, enhancing the clarity of the neural signals. The choice of cutoff frequency is crucial; setting it too high may distort low-frequency neural activities, while setting it too low might not effectively remove the unwanted DC component.

Among various filter designs, the Butterworth filter [105] is known for its flat frequency response in the passband, ensuring minimal signal distortion within the desired frequency range. Its smooth response makes it a popular choice in EEG preprocessing. Implementing a Butterworth high-pass filter involves selecting the filter order and cutoff frequency. A higher-order filter provides a steeper roll-off, attenuating unwanted low frequencies more effectively. However, it is essential to balance the filter order to avoid excessive phase distortion or increased computational complexity [105, 1, 2].

3. **Filtering:** Enhances EEG signal quality by isolating frequency bands related to brain activity and removing unwanted noise. To extract particular EEG frequency bands — such as alpha, beta, and gamma rhythms—band-pass filters are employed [105, 1, 20]. Band-pass filters allow signals within a specified frequency range to pass through while attenuating frequencies outside this range. Researchers can analyse brain activities associated with these rhythms by designing filters targeting specific EEG bands' frequency ranges. Moreover, downsampling (reducing the data's sampling rate) is often conducted after filtering to make data handling more efficient. This step should follow filtering to prevent aliasing artifacts, ensuring that high-frequency components are adequately attenuated before the reduction in sampling rate. Before downsampling, an anti-aliasing filter is applied to suppress frequencies above the new Nyquist limit [104].

4. **Re-referencing:** Redefines the reference point for all EEG channels to improve signal interpretability [20, 1, 5, 107]. Since EEG measures the difference in electrical potential between electrodes, the choice of reference significantly influences the recorded signals' amplitude and spatial distribution. The common re-referencing methods include:

   (a) **Average Reference (AR):** Calculates the mean voltage across all electrodes and uses this average as the reference for each channel. This method reduces the influence of any single electrode and is effective in high-density electrode arrays. However, it may introduce bias if certain electrodes are noisy or the electrode coverage is not uniform [107].

   (b) **Linked Mastoids (LM):** Uses the average of signals from electrodes placed on both mastoid processes (bony areas behind the ears) as the reference. This approach is common in auditory studies and provides a stable reference point. However, it may not be suitable for all experimental setups, especially if mastoid areas are involved in the study [107].

   (c) **Reference Electrode Standardisation Technique (REST):** Estimates a reference at infinity by modelling the potential distribution over the scalp to approximate a neutral reference point. This method provides a more accurate representation of brain activity by minimising reference-related biases but requires complex computations and accurate head models [107].

   The choice of reference affects both the amplitude and polarity of EEG signals. Re-referencing can alter the topographical distribution of recorded potentials, potentially influencing the interpretation of neural activity patterns. Therefore, selecting an appropriate reference is crucial for accurate data analysis.

5. **Epoching:** EEG data preprocessing includes epoching, a crucial step that segments continuous recordings into discrete time-locked intervals or epochs centred around specific events or stimuli [20, 1, 5, 15, 2]. This process facilitates the analysis of brain responses to particular events by aligning data segments to a common temporal reference point. Firstly, it determines the events of interest within the continuous EEG data, typically marked by triggers or markers

corresponding to specific stimuli or responses. Secondly, the time window for each epoch relative to the event marker will be tuned, including pre-stimulus and post-stimulus durations. For example, an epoch may range from 100 ms before to 500 ms after the event onset. Finally, the defined time windows are extracted from the continuous data to create epochs for each event occurrence.

6. **Artifact removal:** Artifacts in EEG recordings originate from non-neural sources such as eye blinks, muscle movements, and environmental noise. Several techniques are employed to remove these unwanted components:

   (a) **Independent Component Analysis (ICA)** is a computational method that separates multivariate signals into additive, independent components [81]. In EEG processing, ICA decomposes recorded signals to identify and remove components associated with artifacts like eye movements or muscle activity. This technique assumes that underlying sources are statistically independent and non-Gaussian, allowing for reconstructing EEG signals with reduced artifact interference. Studies have demonstrated ICAs effectiveness in eliminating artifacts from EEG data.

   (b) **Empirical Mode Decomposition (EMD)** is a data-driven technique that decomposes nonlinear and non-stationary signals into Intrinsic Mode Functions (IMFs), each representing a simple oscillatory mode within the signal [108]. In EEG applications, EMD can separate neural activity from artifacts by extracting IMFs corresponding to different frequency components, facilitating the attenuation of artifact-related IMFs and resulting in cleaner EEG signals. The foundational work on EMD and the Hilbert spectrum has been pivotal in advancing nonlinear and non-stationary time series analysis.

   (c) **Adaptive filtering** involves designing filters that adjust their parameters in real time based on the statistical properties of the input signal. In EEG applications, adaptive filters can dynamically identify and suppress artifacts by distinguishing them from neural signals. For instance, when reference channels detect specific artifacts (e.g., eye movements), the filter adapts to remove these components from the EEG data. This method

Table 2.3: Time-domain features in EEG signal analysis.

Feature	**Description**	**Relevance to EEG Analysis**
Mean	Average amplitude of the signal.	Indicates overall signal level [109].
Variance	Measure of signal amplitude variability.	Reflects signal variability [109].
Standard Deviation	Square root of variance.	Measures signal amplitude dispersion [109].
Skewness	Asymmetry of the amplitude distribution.	Indicates signal shape and potential abnormalities [110].
Kurtosis	Peakedness of the amplitude distribution.	Reflects signal's outlier characteristics [110].
Root Mean Square (RMS)	Square root of the mean of squared values.	Represents signal power [111].
Zero-Crossing Rate	Number of times the signal crosses zero amplitude.	Indicates frequency content [112].
Hjorth Parameters	Activity, Mobility, and Complexity metrics.	Describe signal's time-domain properties [113].

is beneficial when artifacts overlap frequency spectra with neural signals, making traditional filtering techniques less effective [82].

2.3.2 Feature Extraction

After preprocessing, transforming EEG signals into meaningful features is crucial for effective analysis. The feature extraction step enhances signal interpretability and reduces dimensionality. The following are the common and effective features:

1. **Time-Domain Features:** Time-Domain features involve statistical properties such as mean, variance, skewness, and kurtosis, which summarise the signal's amplitude characteristics over time. Time-domain features are essential for capturing variations in EEG signals, providing insights into fluctuations in neural activity. Table 2.3 lists all the widely used time-domain features for EEG signal processing.

2. **Frequency-Domain Features:** Analysing the signal in the frequency domain

Table 2.4: Frequency-domain features in EEG analysis.

Feature	**Description**	**Relevance to EEG Analysis**
Power Spectral Density (PSD)	Power distribution across frequencies	Identifies dominant rhythms and oscillations [40, 114]
Fast Fourier Transform (FFT)	Converts time-domain to frequency-domain	Reveals spectral content of the signal [40, 114]
Spectral Edge Frequency (SEF)	Frequency below which a certain percentage of total power lies	Indicates overall frequency distribution [40]
Intensity Weighted Mean Frequency (IWMF)	Average frequency weighted by signal intensity	Represents central tendency of frequency content [40]
Band Power	Power within specific frequency bands	Analyzes specific brain rhythms [115]
Relative Band Power	Ratio of power in a specific band to total power	Compares different frequency bands [114]
Peak Frequency	Frequency with highest amplitude in a given band	Identifies dominant frequencies [40]
Spectral Entropy	Regularity or complexity of the power spectrum	Measures signal's frequency domain complexity [40]
Coherence	Correlation between two signals in frequency domain	Analyzes functional connectivity between brain regions [115]
Phase Synchronization Index (PSI)	Phase relationship between different EEG channels	Measures synchronization between brain areas [40]

provides insights into its spectral content. Standard techniques include Power Spectral Density (PSD) estimation and Fourier Transform (FFT) [116, 117]. PSD measures the power distribution over frequency components, aiding in identifying dominant rhythms and oscillations in EEG signals. Various frequency bands such as Delta (0.5–4 Hz), Theta (4–8 Hz), Alpha (8–13 Hz), Beta (13–30 Hz), and Gamma (above 30 Hz) have been analysed for emotion recognition and cognitive state estimation [1]. Table 2.4 presents key frequency-domain features used in EEG analysis along with their descriptions and relevance.

3. **Time-Frequency Features:** Unlike traditional time-domain or frequency-

domain approaches, time-frequency features capture transient signal changes by simultaneously analysing time and frequency information. Methods such as Wavelet Transform [118] and Short Time Fourier Transform (STFT) [118] enable the decomposition of EEG signals into different frequency components while preserving their temporal evolution. These techniques are extensively used in EEG-based emotion recognition [1, 5].

4. **Spatial Features:** Spatial features capture the distribution and relationships between EEG signals recorded from different electrodes on the scalp. These features are essential for understanding brain activity across different regions, as EEG signals vary spatially due to functional specialisation in the brain. Common Spatial Patterns (CSP) is a widely used technique for enhancing discriminative EEG features by optimising spatial filters, commonly applied in motor imagery and BCI applications [119, 120, 121]. CSP functions by identifying spatial filters that maximise variance for one class of signals while minimising it for another, enhancing discriminative features across different brain regions.

5. **Differential Entropy (DE):** As a measure of signal complexity, DE quantifies the randomness or unpredictability within EEG signals. Unlike traditional entropy measures, typically applied to discrete signals, DE is an extension designed for continuous distributions. It is defined by Equation 2.1:

$$H(X) = -\int p(x) \log p(x) dx \tag{2.1}$$

where $p(x)$ represents the probability density function of the continuous EEG signal X. In practice, DE is often computed by assuming a Gaussian distribution of EEG signals, leading to the simplified form as in Equation 2.2

$$H(X) = \frac{1}{2} \log(2\pi e \sigma^2) \tag{2.2}$$

where σ^2 is the variance of the signal. This formulation efficiently computes DE features across EEG frequency bands.

DE is a widely used feature in EEG-based affective computing due to its

correlation with brain activity variations. Research utilising the SEED [1] dataset has demonstrated that DE features can effectively capture emotional components of EEG signals, reflecting activity differences across brain regions [122, 1]. Additionally, DE is a discriminative feature for classifying emotional states across different frequency bands, making it a robust feature for emotion recognition tasks [1, 3].

2.3.3 Deep Learning

In the model-building phase of EEG classification, researchers apply various machine-learning algorithms to classify and interpret EEG signals. For instance, in detecting epileptic seizures, researchers have utilised logistic regression, demonstrating its applicability in binary classification tasks within EEG studies. Similarly, practitioners employ k-Nearest Neighbors (kNN) in EEG-based machine-learning applications, where its instance-based learning approach aids in classifying EEG patterns. Support Vector Machines (SVM) is also a prominent model in EEG research, offering robust performance in high-dimensional feature spaces, which benefits the analysis of complex EEG data. Tree-based methods, such as random forests, are applied to EEG data, providing advantages in handling nonlinear relationships and interactions between features. Furthermore, ensemble-based methods, including stacking ensemble classifiers, enhance the accuracy of EEG-based emotion classification, demonstrating the effectiveness of combining multiple models to improve predictive performance [1, 2, 1, 122].

In recent years, Deep Learning (DL) models have significantly advanced the analysis of EEG signals, enabling more accurate and efficient classification models. Unlike classical EEG classification techniques, DL-based techniques are straightforward and do not require extensive feature extraction; instead, they directly work on the time series EEG signals. This subsection delves into the evolution of DL models for EEG processing. It begins with Convolution Neural Network (CNN) [123, 124] based methods that learn to extract features and then moves to Recurrent Neural Network (RNN) [125, 126] to learn sequential dependencies in the temporal domain. Further, it elaborates Graph Neural Networks (GNN) [127, 8] that leverages the spatial

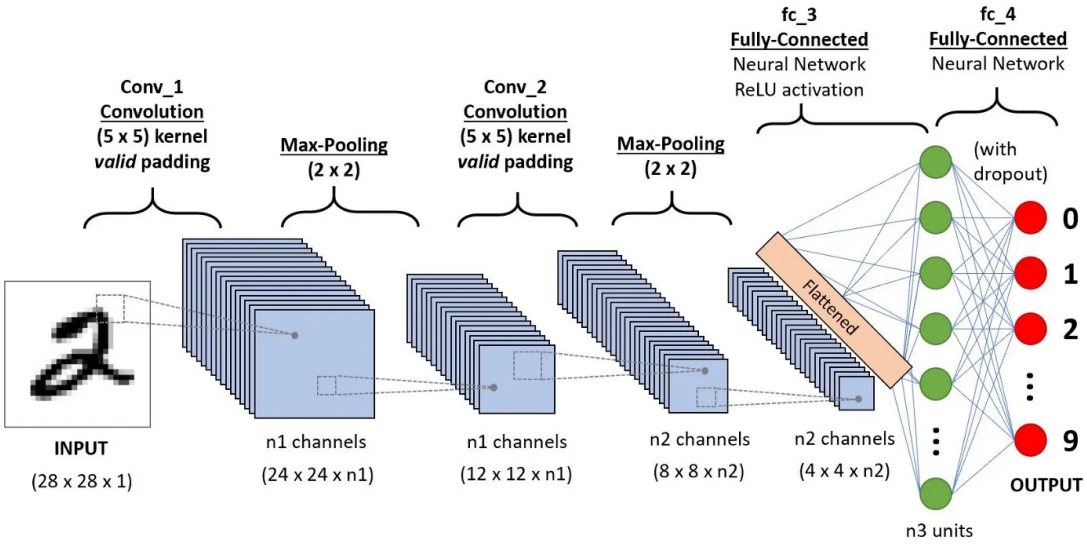


Figure 2.6: CNN architecture [6].

arrangement of EEG signals along with time. Finally, Transformers [49, 11, 39] can learn features in space and frequency domain features while exploiting long-range time dependencies of EEG signals.

2.3.3.1 Convolutional Neural Networks for EEG

CNNs process EEG signals by learning spatial and temporal patterns directly from raw or transformed representations. These deep learning models use convolutional layers to extract features from grid-like data, such as time-series EEG. In EEG analysis, CNNs operates either on raw signals or transformed inputs like spectrograms. The convolutional layers serve as feature extractors by identifying relevant spatial and temporal patterns while pooling layers reduce the feature dimensionality and preserve key information. Final classification or regression occurs through fully connected layers integrating the extracted features. Figure 2.7 illustrates this structure. Early studies adopted CNNs to extract spectral-spatial features from preprocessed EEG data. Lun et al. [128] designed a five-layer CNN with separate temporal and spatial filters and achieved 97.28% accuracy on MI tasks. This design operated directly on raw EEG signals and eliminated manual preprocessing, marking an early shift toward end-to-end learning. EEGNet [6] further advanced this direction by introducing depthwise separable convolutions to reduce computational complexity

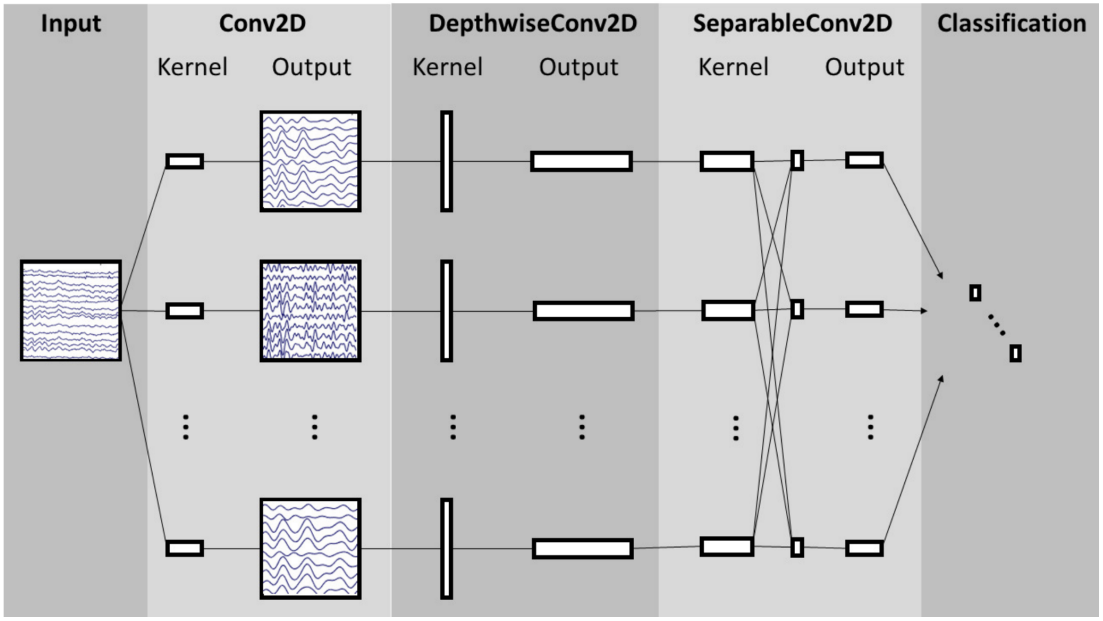


Figure 2.7: Neural network architecture of EEGNet [6].

while maintaining classification performance. Figure 2.6 presents the EEGNet architecture, which includes:

1. **Depthwise convolution:** It applies a separate filter to each input channel, learning spatial features within individual channels.

2. **Pointwise convolution:** It performs a 1×1 convolution across channels, integrating spatial information from different channels.

This architecture extracts spatiotemporal EEG features efficiently and supports real-time BCIs due to its lightweight structure. Table 2.5 summarizes common CNN-based architectures proposed for EEG classification.

2.3.3.2 Recurrent Neural Networks (RNNs) for EEG

RNNs [125] model sequential data by maintaining a hidden state that evolves, making them suitable for capturing the temporal dynamics of EEG signals. EEG signals exhibit non-stationary behavior and temporal dependencies, which recurrent connections in RNNs can learn by updating the hidden state at each time step. Figure 2.8 illustrates the unfolded structure of an RNN across multiple timesteps, where each block corresponds to a specific time step and shares the same set of weights. This unrolled view shows

Table 2.5: Summary of CNN-based EEG processing studies.

Reference	Model	Dataset(s)	Performance	# Classes	DA	TL	SSL
Lawhern et al. [6]	**EEGNet**: Depthwise and separable convolutions; designed for efficiency.	BCI paradigms: P300 [129], ERN [130], SMR [129]	Demonstrated robust performance across multiple BCI paradigms, especially with limited training data.	Varies	No	No	No
Sors et al. [131]	**1D-CNN**: Consists of convolutional and pooling layers followed by fully connected layers.	Sleep-EDF [132]	Achieved an accuracy of 74.8% in sleep stage classification.	5	No	No	No
Zhang and Wu [133]	**Multi-Scale CNN**: Incorporates multi-scale convolutional layers to capture diverse temporal features.	Clinical sleep data	Achieved 96% accuracy in sleep stage classification.	5	No	No	No
Zhu et al. [134]	**Attention-Based CNN**: Combines convolutional layers with attention mechanisms to enhance feature extraction.	Sleep-EDF [132]	Achieved 93.7% accuracy in sleep stage classification.	5	No	No	No
Zhao et al. [135]	**CNN-LSTM Hybrid**: Integrates convolutional layers with LSTM units to capture spatial and temporal features.	Sleep-EDF [132]	Achieved 93.47% accuracy in sleep stage classification.	5	No	No	No
Li et al. [136]	**CAttSleepNet**: Integrates convolutional layers, BiLSTM units, and attention mechanisms for comprehensive feature extraction.	Sleep-EDF [132]	Achieved superior performance compared to other models on Sleep-EDF datasets.	5	No	No	No
Li et al. [137]	**Simplified CNN**: 5-layer CNN with temporal and spatial convolutions, max pooling, and a fully connected layer.	BCI Competition IV-2a [129]	Achieved competitive performance in motor imagery classification tasks.	4	No	No	No
Zhang et al. [138]	**SaleNet**: Combines CNN layers optimized for low-power FPGA implementation.	Custom EEG dataset	Achieved a state-of-the-art subject-independent sustained attention level classification accuracy of 84.2%.	2	No	No	No
Alghanim et al. [139]	**Hybrid Ensemble CNN**: Ensemble of CNNs trained on decomposed EEG components, integrated using two ensemble modes.	Custom EEG dataset	Achieved an average classification accuracy of 83.48% in cross-subject driver fatigue recognition tasks.	2	No	No	No
Lu et al. [140]	**CIT-EmotionNet**: Combines CNN for local feature extraction and Transformer for global feature extraction, with an interactive module for feature fusion.	SEED [1], SEED-IV [3]	Achieved an average recognition accuracy of 98.57% on SEED and 92.09% on SEED-IV datasets.	Varies	No	No	No

the information flow from input x to hidden state h and output $\hat{y}$ at each time step, effectively capturing sequential dependencies in the data. Traditional RNNs often face difficulties in learning long-range dependencies due to vanishing or exploding gradients during training. To overcome these issues, researchers proposed advanced variants such as Long-Short-Term Memory (LSTM) networks [126]. LSTM networks introduced gating mechanisms that regulate information flow and preserve long-term dependencies across sequences [126, 141]. Building on this, Bidirectional LSTM

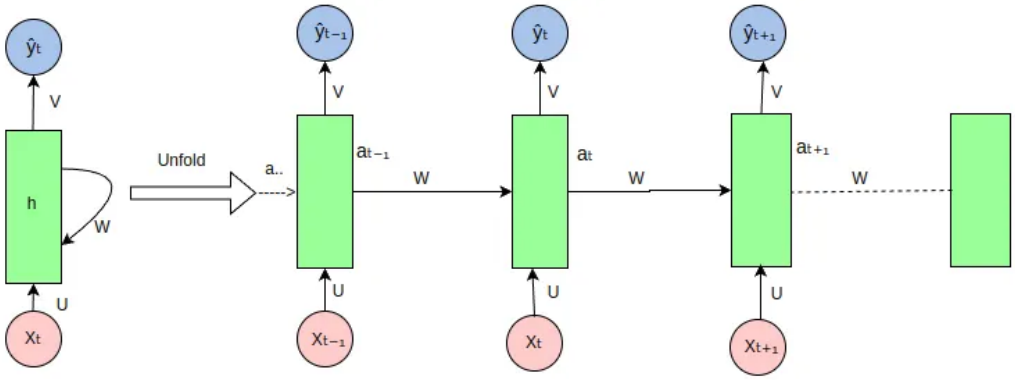


Figure 2.8: RNN architecture [7].

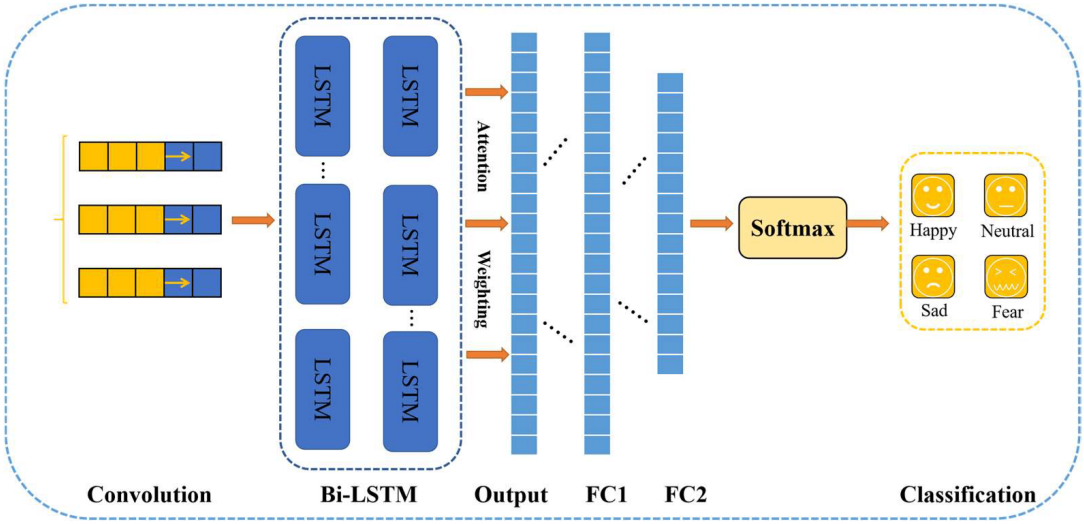


Figure 2.9: Neural network architecture of CNN-LSTM based EEG signal classification [7].

(BiLSTM) [141] networks extended LSTMs by processing input sequences in both forward and backward directions. This bidirectional design enabled the model to access past and future context at each time step, improving its ability to model complex temporal patterns in EEG classification tasks.

To effectively capture both spatial and temporal features in EEG signals, hybrid architectures that integrate CNNs and RNNs have been developed. In these models, CNNs are typically employed to extract spatial features from EEG data, which are then fed into RNNs (such as LSTMs or BiLSTMs) to model temporal dependencies as summarized by various architectures in Table 2.6. For example, Omar et al. [7] proposed a hybrid CNN and Bi-LSTM model for EEG signal classification as shown in Figure 2.9. In this architecture, initial convolutional layers extract spatial features

Table 2.6: Summary of RNN-based EEG processing studies.

Reference	Model	Dataset(s)	Performance	# Classes	DA	TL	SSL
Supratak et al. [142]	**DeepSleepNet**: A model combining CNN and bidirectional LSTM. CNN extracts time-invariant features, while BiLSTM captures temporal dependencies.	Sleep-EDF [132]	Achieved 82% accuracy on Sleep-EDF dataset.	5	No	No	No
Sun et al. [143]	**1D Convolutional LSTM**: Integrates 1D convolutional layers with LSTM to learn spatial and temporal features for user identification.	Private dataset	Reported 99.58% accuracy on a 109-subject database.	109	No	No	No
Xu et al. [144]	**CNN-LSTM**: CNN extracts spatial features; LSTM captures temporal sequence learning in EEG data.	Sleep-EDF [132]	Achieved 93.47% accuracy using Fpz-Cz channel.	5	No	No	No
Wilaiprasitporn et al. [145]	**CNN-GRU**: Combines CNN for spatial feature extraction and GRU for temporal modeling in EEG signals.	DEAP [5]	Achieved up to 100% mean Correct Recognition Rate (CRR).	32	No	No	No
Saha and Fels [146]	**CNN-RNN Autoencoder**: Uses parallel CNN and RNN, followed by autoencoders to capture spatio-temporal features in EEG.	Various public EEG datasets	Demonstrated a 23.45% improvement over baseline methods.	Varies	No	No	No
Narin [147]	**Deep CRNN**: Combines CNN for spatial feature extraction and RNN for temporal dynamics in EEG classification.	Custom EEG dataset	Demonstrated robustness and high accuracy in EEG classification tasks.	Varies	No	No	No
Spampinato et al. [148]	**RNN for Visual Classification**: RNN learns discriminative brain activity patterns corresponding to visual object categories.	Custom EEG dataset	Achieved approximately 40% accuracy in object classification using EEG data.	40	No	No	No
Li et al. [149]	**BiHDM**: A Bidirectional Hierarchical Dynamic Model using RNNs to capture short-term and long-term dependencies in EEG signals.	SEED [1], DEAP [5]	Achieved 90.1% accuracy on SEED and 85.2% on DEAP datasets.	Varies	No	No	No
Chowdary et al. [150]	**Emotion-Specific RNN**: Uses emotion-specific representations in RNN to enhance EEG-based emotion recognition.	SEED [1]	Achieved 86.65% accuracy on SEED dataset.	3	No	No	No
Tripathi et al. [151]	**Deep RNN**: Deep RNN architecture models temporal dependencies in EEG for emotion classification.	DEAP [5]	Achieved 87.58% accuracy in binary classification of valence and arousal.	2	No	No	No

from EEG signals, which are then processed by Bi-LSTM layers to capture temporal dependencies. An attention mechanism further enhances the model's ability to focus on the most relevant features. The extracted representations are processed through fully connected layers before classification via a softmax function, distinguishing between emotional states such as happy, neutral, sad, and fearful.

Another such study proposed OPTICAL framework by integrating LSTM networks with Common Spatial Pattern (CSP) [152] filtering and sliding window processing, achieving significant improvements in classifying motor imagery tasks [153]. This approach combines LSTM's regression outputs with traditional CSP features, enhancing feature discriminability through support vector machines. In epilepsy detection,

Table 2.7: Summary of GNN-based EEG processing studies.

Reference	Model	Dataset(s)	Performance	# Classes	DA	TL	SSL
Demir et al. [156]	**EEG-GNN**: Projects EEG electrodes onto graph nodes using neuroscientific policies; GNN captures functional neural connectivity.	ErrP [157], RSVP [158]	Outperformed standard CNN classifiers across ErrP and RSVP datasets.	Varies	No	No	No
Zhong et al. [159]	**RGNN**: Captures local/global EEG channel relations using biologically-inspired topologies; includes regularizers for robustness.	SEED [1], SEED-IV [3]	Demonstrated superior performance over state-of-the-art models.	Varies	No	No	No
Zhang et al. [160]	**GNN4EEG**: Benchmark and toolkit with implementations of various GNN models for EEG analysis.	Multiple EEG Datasets	Facilitated comprehensive GNN evaluation across EEG tasks.	Varies	No	No	No
Hajisafi et al. [161]	**NeuroGNN**: Dynamically constructs graphs capturing spatial, temporal, semantic, and taxonomic EEG correlations.	Real-world EEG Data	Significantly outperformed existing models in seizure detection and classification.	2	No	No	No
Jin et al. [162]	**PGCN**: Pyramidal GCN for emotion recognition, capturing dependencies at local, mesoscopic, and global scalp levels.	SEED [1], DEAP [5], DREAMER [15]	Achieved state-of-the-art results in subject-dependent and subject-independent scenarios.	Varies	No	No	No
Klepl et al. [163]	**GNN**: Applied for Alzheimer's detection and seizure type classification using EEG graphs.	Custom Dataset	Effective in classifying Alzheimer's and seizure types.	3	No	No	No
Hou et al. [165]	**GCNs-Net**: Constructs graphs using Pearson correlation, applies GCN layers for motor imagery decoding.	PhysioNet [166], High Gamma	93.06% accuracy (subject), 88.57% (group) on PhysioNet dataset [166].	Varies	No	No	No
Chen et al. [167]	**DAMGCN**: Dual attention GCN assigns weights to electrodes and frequency bands for emotion recognition.	DEAP [5], SEED [1], SEED-IV [1]	99.42% (subject-dependent) and 73.21% (subject-independent) accuracy on SEED.	Varies	No	No	No
Kumar et al. [168]	**ST-GCN**: Spatio-Temporal GCN with GRU for motor imagery classification.	BCI Competition IV [129]	Achieved 85.3% accuracy, outperforming CNN and RNN models.	4	No	No	No
Demir et al. [169]	**EEG-GAT**: Applies Graph Attention Networks with multi-head attention to learn EEG graph structures.	RSVP [158], ErrP [157]	Improved classification accuracy over traditional CNNs.	Varies	No	No	No
Li et al. [169]	**SOGNN**: Self-Organized GNN dynamically constructs EEG graphs for cross-subject emotion recognition.	SEED [1], SEED-IV [3]	Achieved 86.81% on SEED and 75.27% on SEED-IV	3 (SEED), 4 (SEED-IV)	No	No	No

contemporary architectures combine convolutional and recurrent components. The Conv-1D-BiLSTM [154] model captures spatial patterns through one-dimensional convolutions and long-term temporal dependencies via bidirectional LSTM layers, demonstrating high accuracy in seizure detection tasks. However, standard RNN architectures face challenges modelling long-range dependencies in prolonged EEG recordings, prompting the integration of attention mechanisms [155].

2.3.3.3 Graph Neural Networks (GNNs) for EEG

Graph Neural Networks (GNNs) [127] model non-Euclidean data by explicitly representing relationships as graphs. Each data point becomes a node, and edges

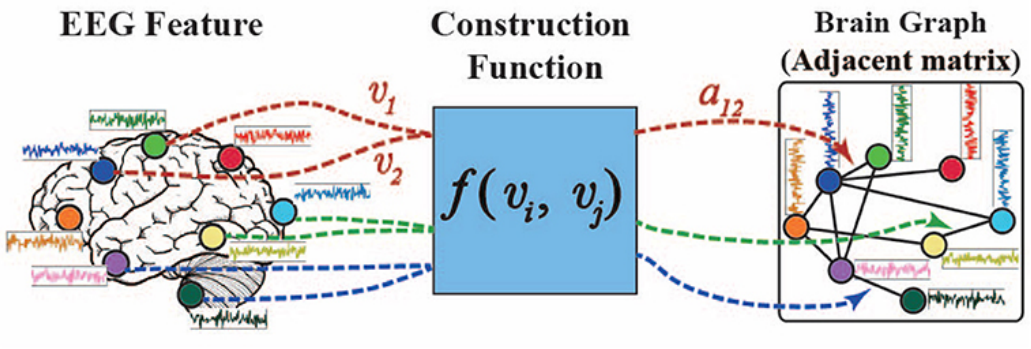


Figure 2.10: Graph construction from EEG features [8].

capture structural or spatial dependencies. GNNs learn by exchanging information between connected nodes, enabling deep learning on structured data. GNNs represent data using nodes and edges and learn by exchanging information between connected nodes. The ability to model spatial and structural dependencies through message passing makes GNNs well-suited for tasks involving complex relational structures. EEG signals naturally benefit from graph-based modelling. EEG electrodes are distributed irregularly across the scalp, and their signals often exhibit spatial or functional relationships. GNNs treat each electrode as a node and define edges based on spatial distance or signal similarity. Table 2.7 summarizes famous GNN-based models used for EEG signal processing. GNN models EEG signals by computing pairwise similarity between electrode feature vectors v_i and v_j using a learnable function to construct the graph. Equation 2.3: defines the similarity score a_{ij} between electrodes:

$$a_{ij} = \frac{\exp(\theta(v_i W)\theta(v_j W)^\top)}{\sum_{i=1}^{N} \exp(\theta(v_i W)\theta(v_j W)^\top)} \tag{2.3}$$

Equation 2.4 uses these scores to compute the adjacency matrix A, where $G = \tanh(VW)$ denotes transformed electrode features:

$$A = \text{Softmax}(GG^\top) \tag{2.4}$$

2.3.3.3.1 Graph Convolutional Networks (GCNs) Graph Convolution Network (GCN) [9] generalize convolution to graph domains. Instead of applying filters over grid-based neighbourhoods (as in CNNs), GCNs aggregate information from

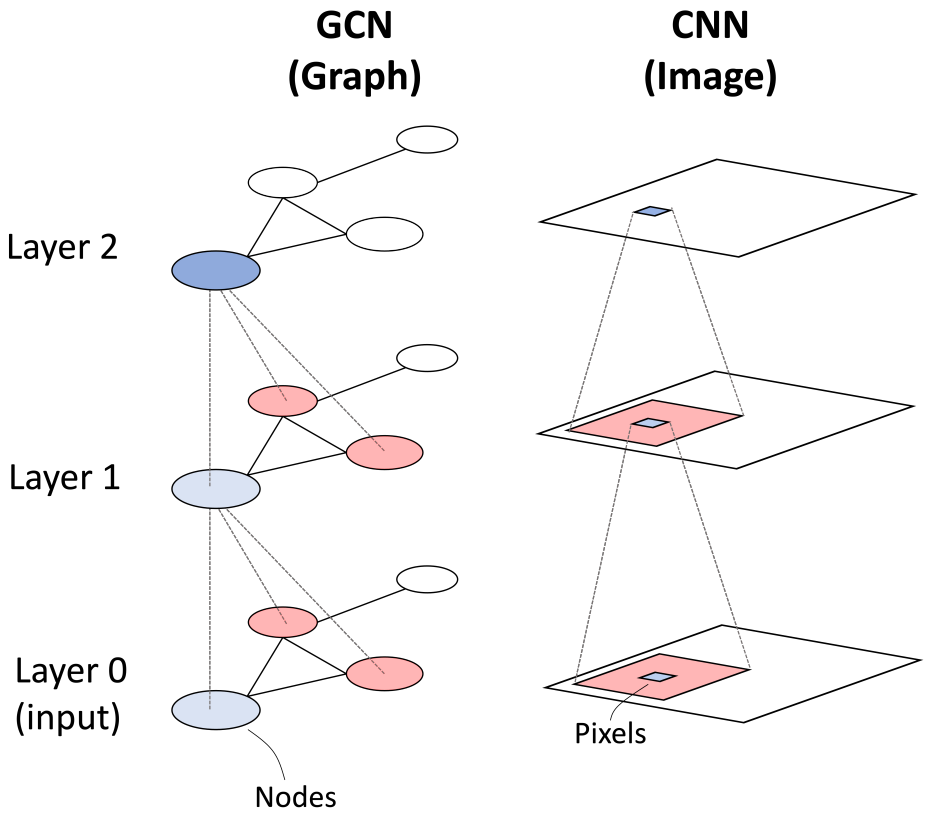


Figure 2.11: Comparison between CNN and GCN. GCNs aggregate information from neighbouring nodes, while CNNs apply fixed kernels over image grids. [9]

neighbours in a graph. Equation 2.5: defines the update rule for GCN layers:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H^{(l)}W^{(l)}\right) \tag{2.5}$$

where $\tilde{A} = A + I$ adds self-loops, $\tilde{D}$ is the degree matrix, $H^{(l)}$ is the node feature matrix, and $W^{(l)}$ and $\sigma(\cdot)$ are the trainable weight matrix and non-linear activation, respectively. EEG-GNN [156] introduced a graph-based framework for EEG classification that addressed limitations of traditional CNNs in capturing the complex spatial dependencies inherent in EEG data. Unlike CNNs, which assumed equidistant relationships among inputs, EEG-GNN [156] represented each electrode as a node in a graph, with edges defined based on spatial distances or functional connectivity measures. This graph structure allowed the model to reflect the spatial and functional relationships between EEG channels accurately. In this framework, EEG signals were processed through GCN [9] layers that performed message passing among connected nodes. This mechanism enabled the extraction of spatial patterns

by aggregating information from neighbouring electrodes, effectively capturing the underlying neural connectivity. The model's architecture facilitated the learning of subject-invariant features, enhancing its ability to generalize across individuals. Empirical evaluations demonstrated that EEG-GNN [156] outperformed standard CNN classifiers on various EEG classification tasks, including emotion recognition and motor imagery, by leveraging the graph-based representation to better model the spatial dynamics of EEG signals. Additionally, the framework supported neuroscientific interpretability and could assist in EEG channel selection, which was beneficial for reducing computational costs and designing portable EEG headsets.

2.3.3.3.2 Graph Attention Networks (GATs) GATs [10] enhance GCNs by introducing a learnable attention mechanism that assigns different importance to neighboring nodes. Unlike GCNs, which aggregates neighbour features using fixed or normalized weights such as mean or degree-based, GATs compute attention scores that allow the model to focus on the most informative neighbours. Equation 2.6 expresses the attention coefficient α_{ij} between a target node i and its neighbour j. The model first transforms input features $\mathbf{h}_i$ and $\mathbf{h}_j$ using a shared weight matrix $\mathbf{W}$, concatenates the transformed features, and applies a learnable attention vector $\mathbf{a}$ followed by a LeakyReLU activation. The final attention score is normalized using a softmax over all neighbours:

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\mathbf{a}^\top [\mathbf{W}\mathbf{h}_i \| \mathbf{W}\mathbf{h}_j]\right)\right)}{\sum_{k \in \mathcal{N}(i)} \exp\left(\text{LeakyReLU}\left(\mathbf{a}^\top [\mathbf{W}\mathbf{h}_i \| \mathbf{W}\mathbf{h}_k]\right)\right)} \tag{2.6}$$

Figure 2.12 visualizes this mechanism. The left panel illustrates how the attention module computes α_{ij} for each neighbour, and the right panel shows the aggregation of neighbours' features using these learned weights. In contrast to GCNs [9] that treat all edges uniformly or based on static adjacency, GAT [10] allows dynamic, context-dependent edge weighting. *DenseGATConv* further extends the GAT [10] formulation by enabling fully connected attention across all nodes in the graph, not just the sparse set defined by physical or functional adjacency. This approach treats the input graph as a dense graph and allows each node to attend to every other node, enabling global context modelling. Multi-head attention improves learning stability and expressiveness

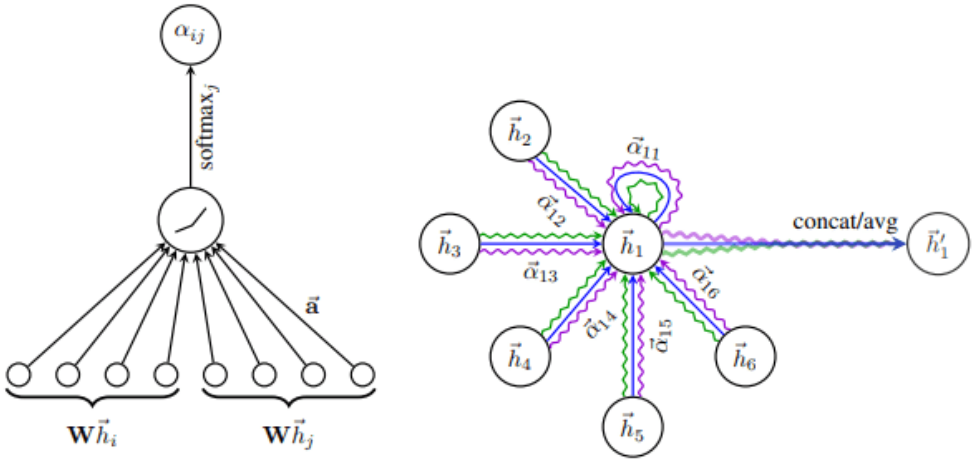


Figure 2.12: Graph Attention Network (GAT) architecture. Left: attention score computation between node pairs. Right: attention-based aggregation of neighbours using learned coefficients [10].

by aggregating multiple independent attention mechanisms before concatenation or averaging.

EEG-GAT [169] extended EEG-GNN [156] by replacing uniform neighbourhood aggregation with DenseGATConv layers and multi-head attention. This design allowed each EEG electrode to attend to all other electrodes dynamically, capturing complex spatial relationships across the scalp. The model achieved improved classification accuracy on EEG datasets such as SEED-IV [3] and DEAP [5] over EEG-GNN [156] demonstrating the benefit of attention-based graph modelling in EEG signal analysis.

2.3.3.3.3 Self Organizing Graph Neural Network (SOGNN) EEG-GNN [156]and EEG-GAT [169] rely on fixed graphs that remain unchanged across subjects or input samples. This limitation restricts their ability to model subject-specific variability in EEG signals. To address this, SOGNN [8] constructs dynamic graphs that adapt to each EEG trial, enabling individualized graph representations. The model begins by computing a dense affinity matrix A from pairwise similarity scores between EEG channel embeddings. The model derives these similarities from feature transformations of the input EEG signals. SOGNN applies a top-k pruning operation to adjacency matrix A to enforce sparsity and retain meaningful structure. Equation 2.7 defines the pruning step, which retains only the top-k highest affinity scores per row to obtain the sparse matrix $\tilde{A}$:

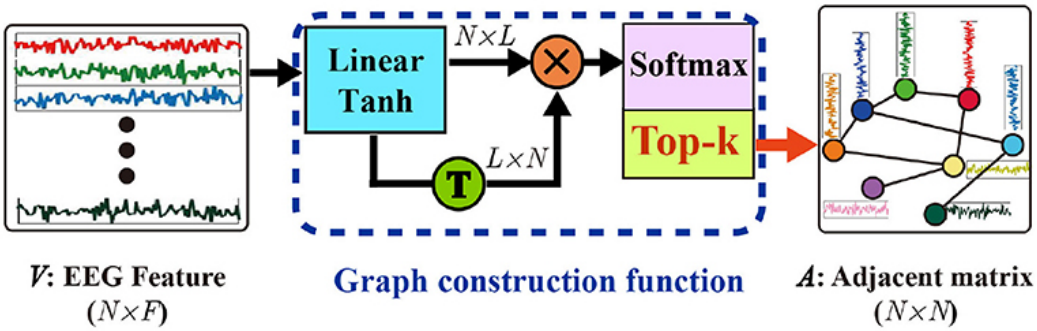


Figure 2.13: Dynamic graph construction in SOGNN [8].

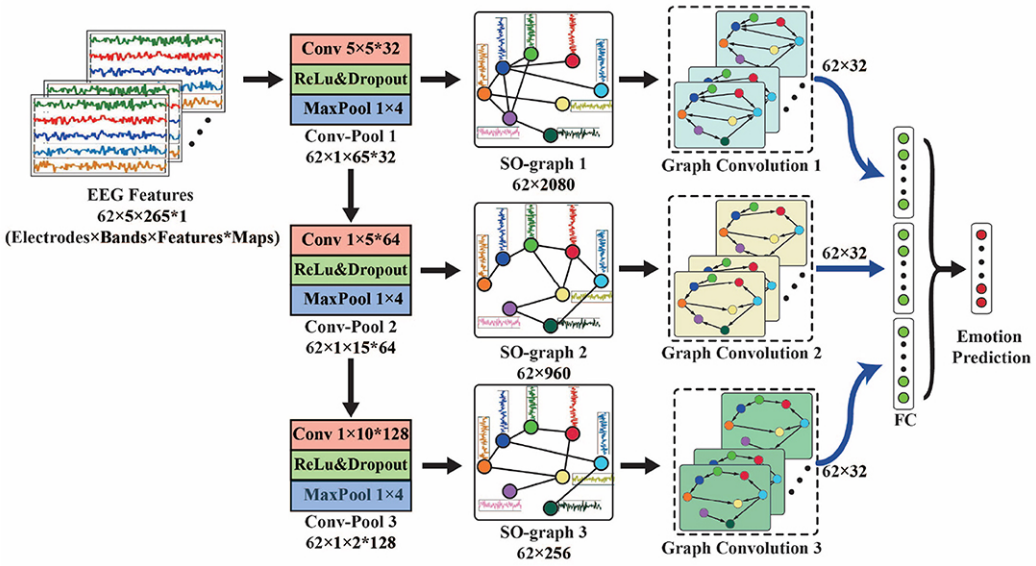


Figure 2.14: Overall architecture of SOGNN. EEG features are transformed, used to construct graphs, and passed through GCN layers and classifiers [8].

$$\tilde{A}_{ij} = \begin{cases} A_{ij}, & \text{if } j \in \text{Top}_k(A[i,:]) \\ 0, & \text{otherwise} \end{cases} \tag{2.7}$$

Figure 2.13 illustrates this process, where the model removes weak connections and preserves the strongest and most relevant neighbours for each node. This adaptive pruning step results in a sample-specific sparse graph $\tilde{A}$, which enables localized but flexible spatial reasoning. After constructing the dynamic graph, SOGNN applies GCN layers over the pruned adjacency matrix. The model stacks multiple GCN layers across different abstraction levels and fuses their outputs through a feature integration module. This architecture enables hierarchical representation learning while preserving input-specific structural information. Figure 2.14 shows the whole pipeline

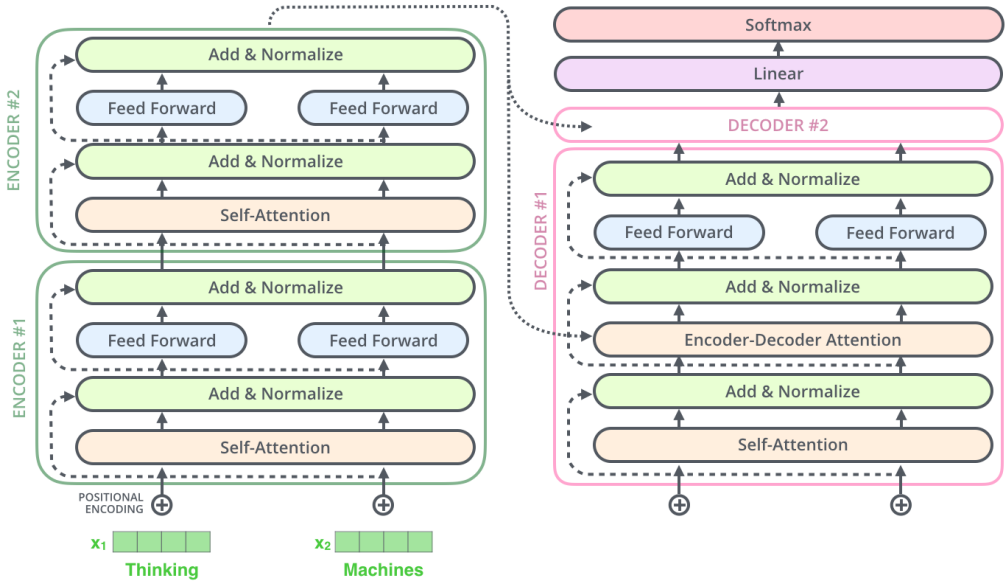
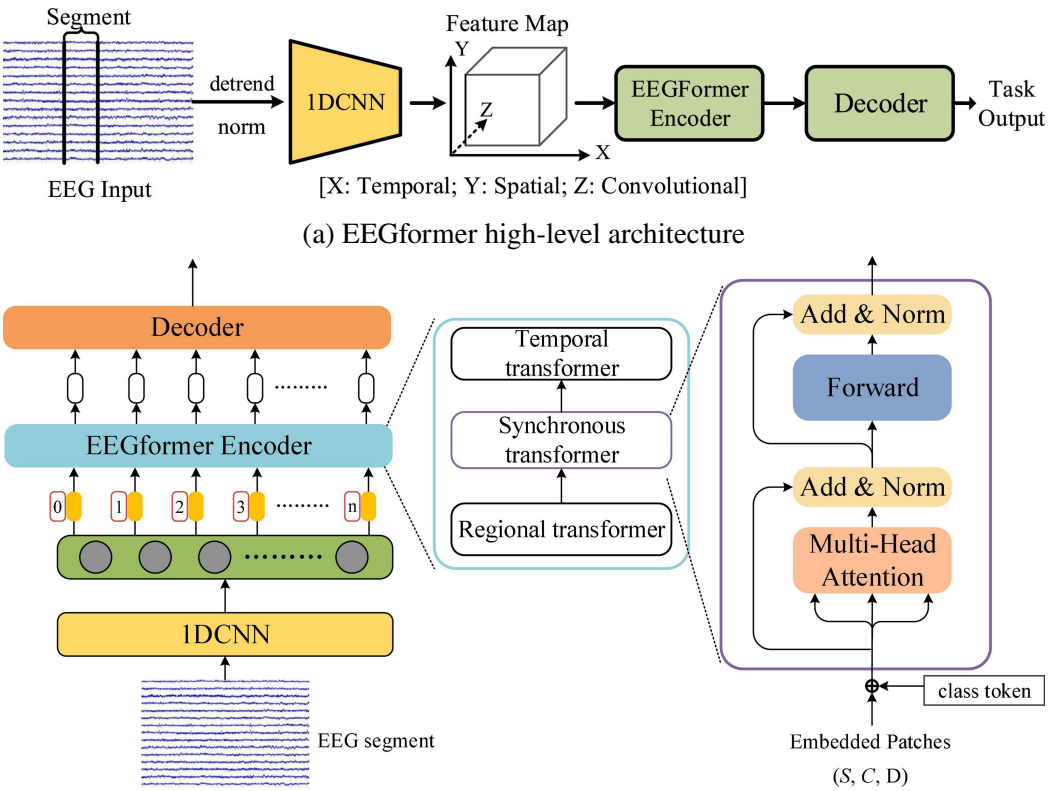


Figure 2.15: Standard Transformer architecture comprising encoder and decoder stacks.

of SOGNN [8]. The model first transforms EEG features using temporal convolution or feedforward layers, constructs a graph based on the transformed features, and applies GCNs for spatial encoding. The final graph-level representations are passed through a classification head to predict cognitive or emotional states. By learning sample-specific topologies and combining multi-scale graph reasoning, SOGNN improves subject generalization and classification accuracy. This dynamic graph formulation enables the model to adaptively encode individual differences in EEG patterns, especially in cross-subject scenarios where static graphs fail to capture signal variability.

2.3.3.4 Transformers for EEG

Transformers [49] serve as foundational models in deep learning by capturing complex dependencies in sequential data through attention mechanisms. Researchers initially developed transformers for natural language processing and later adapted them to domains such as computer vision, audio processing, and, more recently, EEG signal analysis. The key innovation in transformers, the self-attention mechanism, enables models to dynamically focus on relevant parts of the input sequence when computing representations. Unlike RNNs, which process inputs sequentially and often struggle with long-range dependencies, transformers operate on entire sequences in parallel. This design improves computational efficiency and captures global patterns, crucial for



(a) EEGformer high-level architecture

(b) Low-level architecture of EEGFormer

Figure 2.16: Neural network architecture of EEGFormer [11]

modelling high-dimensional, multi-channel EEG data.

In the self-attention mechanism, the model projects each input element into query (Q), key (K), and value (V) vectors. It computes attention scores by taking the scaled dot product between the query and key vectors, normalizes these scores using a softmax function, and uses them to weight the corresponding value vectors. Equation 2.8 shows how the model first calculates the similarity between elements using $QK^\top$, then scales it by the dimension of the key vectors d_k, and finally applies the softmax to generate a probability distribution over all positions in the sequence. This formulation allows the model to assign higher importance to relevant elements and aggregate contextual information effectively across the sequence.

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \tag{2.8}$$

Transformers typically employ *multi-head attention*, where multiple attention heads operate in parallel to capture diverse relational patterns, ranging from short-term local interactions to long-range global dependencies. To encode temporal order—since

transformers lack inherent recurrence—*positional encodings* is added to the input embeddings using fixed sinusoidal patterns or learnable vectors. Figure 2.15 illustrates the standard transformer architecture, which includes stacked encoder and decoder blocks. Each encoder block contains a self-attention layer followed by a feed-forward network, with residual connections and layer normalization applied after each sub-layer. Decoder blocks introduce an additional encoder-decoder attention mechanism, which allows the model to condition outputs on the encoded input sequence. While the full architecture suits sequence-to-sequence tasks, EEG models utilize only the encoder to extract spatiotemporal features from multichannel signals.

In EEG modelling, transformer encoders are the primary component for learning subject-invariant or task-specific patterns. Therefore, most transformer-based architectures utilize only the encoder stack [39, 171, 11], focusing on extracting robust spatiotemporal features from multichannel EEG inputs. EEG signals are multivariate time series, recorded from spatially distributed electrodes placed across scalp. Hence, this requires to model temporal dynamics and spatial dependencies jointly. The global attention mechanism of transformers makes them well-suited for capturing complex temporal and spatial dependencies in EEG signals. It models complex interdependencies across time and channels without relying on handcrafted filters or fixed spatial assumptions. Researchers have proposed specialized transformer variants to better align with EEG data characteristics. Figure 2.16 illustrates the architecture of EEGFormer [11], a hierarchical model that decomposed EEG representation learning into modular components. The authors first applied a 1D convolutional layer to extract low-level temporal features. The resulting feature map was fed into the EEGFormer encoder, which included three specialized transformer branches. Each branch captured a distinct structural property of the EEG signal: the temporal Transformer modeled sequential dependencies, the synchronous Transformer learned inter-channel phase relationships and the regional Transformer encoded spatial groupings of electrodes. The model fused these representations and passed them to a lightweight decoder for classification. Other models, such as EEG-Deformer [172], extended this design by introducing a coarse-to-fine temporal attention mechanism. This approach progressively refined the focus from global temporal patterns to localized fine-grained dynamics. Such hierarchical attention enabled multi-resolution analysis and improved

Table 2.8: Summary of Transformer-based EEG processing studies

Reference	Model	Dataset(s)	Performance	# Classes	DA	TL	SSL
Song et al. [173]	**Transformer-based Spatial-Temporal Feature Learning**: Applies attention mechanisms to enhance spatial features; slices time for attention-based transformation.	BCI Competition IV 2a [129]	Achieved competitive accuracy with fewer parameters than state-of-the-art.	4	No	No	No
Pan et al. [174]	**DuA Transformer**: Dual attentive modules for spatial-spectral and temporal modeling of long-term EEG signals.	Proprietary emotion EEG dataset	Outperformed existing methods with 5.28% average improvement.	Varies	No	No	No
Zhao and Iramina [175]	**CwA-T**: Combines channelwise CNN-based autoencoder with Transformer classifier to preserve independence and model dependencies.	TUH [176]	85.0% accuracy, 76.2% sensitivity, 91.2% specificity.	2	No	No	No
Yang and Modesitt [177]	**ViT2EEG**: Uses pretrained Vision Transformer (ViT) from ImageNet for EEG regression tasks.	Proprietary EEG Regression Task Dataset	Showed clear performance gains over non-pretrained ViT models.	Varies	No	**Yes**	No
Kostas et al. [39]	**BENDR**: Bidirectional Encoder Representations from Transformers for EEG with contrastive SSL.	Various large EEG datasets	Transferable to multiple EEG tasks, outperforming prior methods.	Varies	**Yes**	**Yes**	**Yes**
Zhang et al. [11]	**EEGFormer**: Foundation Transformer model pretrained on large-scale EEG data for generalization.	Compound EEG data	Adaptable to various downstream EEG tasks with interpretability.	Varies	No	**Yes**	**Yes**
Zhu et al. [171]	**Eeg2vec**: Combines contrastive and reconstruction SSL losses; CNN + Transformer for downstream tasks.	ICASSP2023 Auditory EEG Challenge [178]	Improved EEG match-mismatch and regression performance.	Varies	**Yes**	No	**Yes**
Zhang et al. [75]	**GANSER**: Uses GANs and SSL to generate synthetic EEG for data augmentation in emotion recognition.	SEED [1], SEED-IV [3]	Achieved state-of-the-art accuracy.	SEED (3), SEED-IV (4)	**Yes**	No	**Yes**
Zheng et al. [14]	**SGMC**: Genetics-inspired augmentation via Meiosis and group-level contrastive learning.	DEAP [5], SEED [1]	94.72% (valence), 95.68% (arousal) on DEAP; 94.04% on SEED.	2 or 3	**Yes**	No	**Yes**
Li et al. [76]	**MV-SSTMA**: Multi-view CNN-Transformer hybrid reconstructs masked EEG from spectral, spatial, temporal views.	DEAP [5], SEED [1]	Outperformed prior state-of-the-art models.	Varies	No	No	**Yes**
Cui et al. [179]	**Neuro-GPT**: Combines EEG encoder (CNN + Transformer) with GPT trained to predict masked EEG.	TUH [176], BCI IV 2a [129]	Improved motor imagery classification in low-data settings.	2	No	**Yes**	**Yes**
Wang et al. [180]	**EEGPT**: 10M-parameter pretrained Transformer for universal EEG representation learning.	Multi-task EEG dataset	SOTA performance across downstream tasks with linear probing.	Varies	No	**Yes**	**Yes**

sensitivity to transient and sustained EEG phenomena. Table 2.8 summarizes several transformer-based models developed for EEG signal processing. These architectures demonstrated how researchers adapted the Transformer's core strengths—scalability, parallelism, and flexible dependency modeling—using EEG-specific inductive biases to capture the complex spatiotemporal structure of brain activity effectively.

2.4 Summary

This chapter provides a foundational understanding of EEG signals and their acquisition methodologies. It explores various EEG paradigms, montages, electrodes, and frequency bands. The chapter also presents an overview of diverse open EEG datasets, including specific statistics about subjects, channels, and sampling frequency. It delves into a typical EEG classification workflow, detailing data preprocessing, feature extraction, and various deep learning models. These open EEG datasets highlight the scarcity of EEG data and serve as evaluation datasets in subsequent chapters, demonstrating the need for few-shot learning. Simultaneously, the evolution of deep learning models establishes a foundation of various processing techniques for EEG signal classification.

Chapter 3

Literature Review

This chapter reviews key Few-Shot Learning (FSL) techniques for EEG signal processing, including data augmentation, transfer learning, and Self-Supervised Learning (SSL). It presents a taxonomy of data augmentation strategies, techniques for domain adaptation through transfer learning, and SSL techniques for unsupervised feature learning. The chapter critically analyses how these techniques address challenges like data scarcity, subject adaptation, and generalisation. It also identifies current research gaps and suggests future directions.

3.1 Few-Shot Learning (FSL)

Few-Shot Learning (FSL) is a sub-area in Machine Learning (ML), the formal definition of FSL, originates from machine learning. Definition 3.1.1 defines machine learning before defining FSL.

Definition 3.1.1 (**Machine Learning** [12, 181, 182])**.** A computer program is said to learn from experience (E) with respect to some classes of task T and performance measure (P) if its performance can improve with (E) on (T) measured by (P).

A suitable example can be an EEG-based emotion recognition task (T); a machine learning algorithm can improve its classification accuracy (P) through E obtained by training on a large number of labelled EEG samples [1, 5]. Typical machine learning applications requires many examples with supervised information. However, collecting such extensive labelled data in EEG is often difficult or infeasible due to high labelling

costs, subject fatigue, and variability in cognitive states across individuals and sessions. FSL is a special case of machine learning, which targets obtaining good learning performance given limited supervised information provided in the training set D_{train}, which consists of examples of inputs $x_i's$ along with their corresponding output $y_i's$. Formally, Definition 3.1.2 formulates FSL as machine learning for a limited number of samples.

Definition 3.1.2. Few-Shot Learning (FSL) [12] is a type of machine learning problem (specified by E, T and P), where E contains only a limited number of examples with supervised information for the target T.

Concretely, few-shot classification learns classifiers given only a few labelled examples of each class. For example, in an emotion recognition task, the model receives only a few EEG samples for each emotional category—such as happiness, sadness, or fear—rather than the full dataset. This approach also applies to other EEG-based tasks such as mental workload detection, where only limited examples of high or low cognitive load are available; motor imagery classification, where the system must recognise imagined movements from a few training trials; and sleep stage classification, where labelling long overnight recordings is costly and time-consuming. In all these cases, few-shot learning enables models to learn effectively from limited supervision, reducing the dependency on large-scale labelled datasets while improving adaptability across subjects and recording conditions.

3.1.1 FSL Techniques

FSL [12] addresses fundamental challenges in machine learning by enabling models to perform well with limited labelled data. Research on FSL directly tackles the challenges outlined in Chapter 1 and highlighted in Figure 3.1. Specifically, FSL techniques aim to achieve the following goals:

1. Augment data using prior knowledge to learn the approximate distribution of the dataset.

2. Constrain the embedding space with prior knowledge to improve task generalisation.

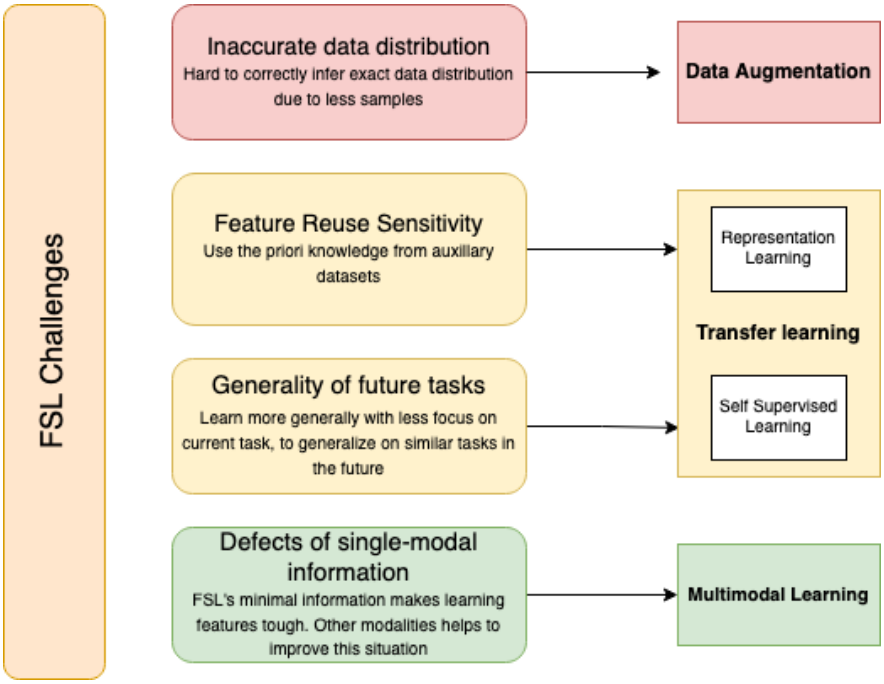


Figure 3.1: Challenges and techniques of FSL [12].

3. Refine pre-trained models on new data by altering the search strategy in the hypothesis space based on prior knowledge.

The techniques for achieving these goals are referred to as Data augmentation [50, 183], transfer learning [184, 185] and Self-Supervised Learning (SSL) [186]. Subsequent sections of this chapter detail the existing work of the following techniques and help identify the research gaps and future directions for each technique.

1. **Data augmentation:** In real-world FSL tasks, the number of samples in the support and query sets is limited due to privacy, collection, and labelling costs. Data augmentation is the most direct way to increase the sample richness. Nevertheless, the core risk of the FSL data augmentation is how likely the augmented dataset can evaluate the distribution behind the real data.

2. **Transfer learning:** Transfer learning is a classical learning paradigm that aims to solve the challenging problem that there are only a few or even no labelled samples in the FSL [184, 185]. The essential operation is to pre-train the model on an extensive dataset and to finetune it on the limited support set. Feature reuse is the core idea of transfer learning to solve the absence of data sets.

3. **Self-Supervised Learning (SSL):** SSL provides an alternative to supervised pretraining by using unlabelled data to learn general-purpose feature representations. SSL avoids task-specific labels and instead creates surrogate pretext tasks — such as predicting signal segments, solving contrastive pairs, or reconstructing masked inputs—that encourage the model to capture intrinsic structure in the data [186, 187, 188]. By learning from large volumes of unlabelled EEG data, SSL produces representations that generalise across subjects, sessions, and even datasets. Through finetuning, these features transfer efficiently to downstream few-shot classification tasks, such as emotion recognition and seizure detection. This capability makes SSL a powerful strategy for addressing label scarcity in EEG-based few-shot learning.

3.2 Data Augmentation

Data augmentation creates new training examples by modifying existing data. This helps address common challenges in EEG signal classification, such as limited data, class imbalance, and high variability across subjects and sessions. Data augmentation improves model robustness by generating realistic variations of EEG signals [189, 190]. Several factors make robust model training difficult in EEG:

1. EEG datasets usually contain fewer examples overall, and the number of samples is often uneven across different classes. For example, in seizure detection, there are far fewer recordings from subjects with seizures compared to recordings from healthy subjects [191].

2. Collecting diverse EEG datasets is difficult because it needs costly EEG devices and controlled conditions to record good-quality brain signals.

3. Due to limited and less diverse data, models may overfit and yield poor performance on unseen data, especially on new subjects or sessions.

4. The model trained can also be unstable, such as small fluctuations in the input can yield very different predictions.

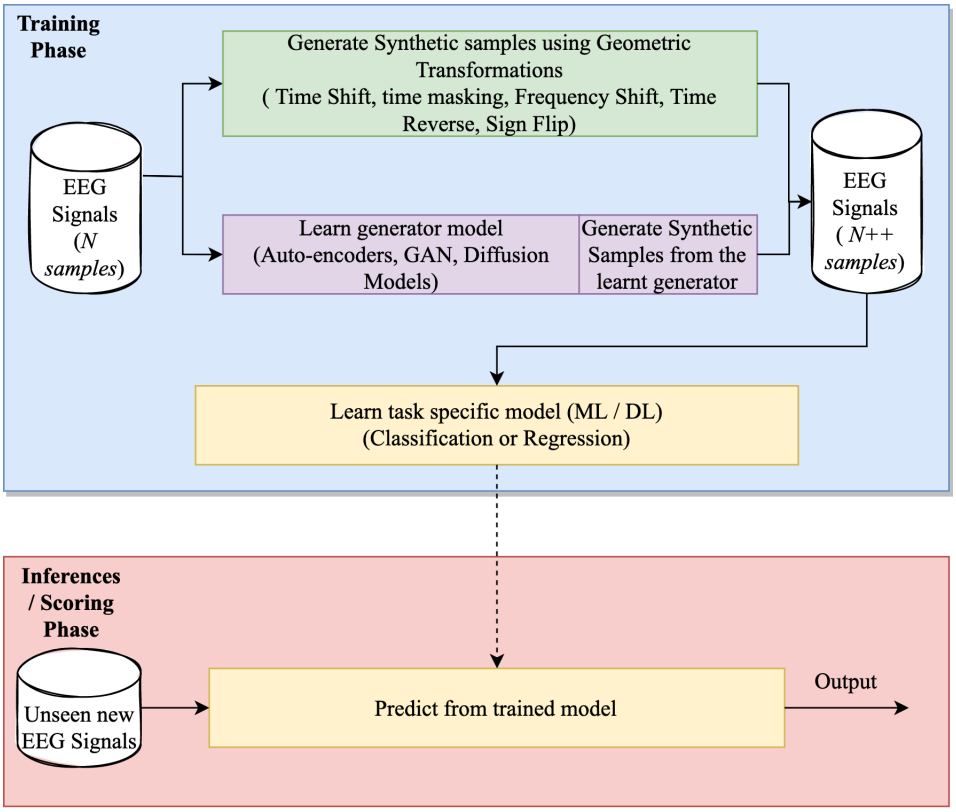


Figure 3.2: Schematic representation of an algorithm for generating synthetic EEG signals.

This section analyses recent advancements in data augmentation for EEG signal processing and organises them based on the augmentation characteristics and corresponding techniques. Figure 3.2 illustrates a typical workflow for synthetic EEG signal generation and its usage during model training. The process begins with collecting N real EEG samples. Transformations such as geometric modifications such as time shift, frequency shift, time masking, time reversal and noise injection techniques generate additional samples by perturbing the original signals. Also, generative models such as Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), and diffusion models synthesise new samples by learning the underlying data distribution. These augmentation strategies expand the training dataset to $N+$ samples. The augmented dataset subsequently trains a task-specific model using either ML or Deep Learning (DL) techniques for classification or regression tasks. The trained model predicts outputs from previously unseen EEG signals during inference.

3.2.1 Characteristics of Data Augmentation

Effective data augmentation adds variety to training data while preserving the key patterns in EEG signals. This thesis identifies three important characteristics from a systematic review of existing studies: subject invariance, session invariance, and class invariance. These characteristics show how different augmentation methods handle changes across subjects, sessions, and class distributions.

Studies that train and test models within the same subject—often across different sessions—focus on *session invariance*. These methods evaluate whether augmentations help the model perform well on new recordings from the same individual. In contrast, studies that train on some subjects and test on unseen subjects aim for *subject invariance*. These methods test whether augmentation supports generalisation across individuals with varying brain signals. *Class invariance* applies when studies handle class imbalance by either balancing the number of samples per class or applying augmentation to each class separately. This approach prevents the model from relying on classes with more examples and encourages learning from rare or minority classes.

Researchers state these characteristics explicitly in their methods or demonstrate them through their design of experiments and report results. For example, studies that report results separately for each subject typically test session invariance, while those that test on unseen subjects demonstrate subject invariance. This classification supports a fair comparison of augmentation methods and helps evaluate their effectiveness in real-world EEG tasks involving new subjects, sessions, or unbalanced class distributions.

1. *Subject Invariance:* EEG signals often encapsulate features indicative of individual subjects. Consequently, a generator trained on such datasets becomes sensitive to these subject-specific features. An optimal generator, Subject Invariant Generator, should adeptly filter out these features. Both Aznan et al. [192] and P. Sharaj and J. Tzyy-Ping [193] proposed a Wasserstein Generative Adversarial Network (WGAN) for Rapid Serial Visual Presentation (RSVP) using gradient penalty to synthesise EEG data.

   Aznan et al. [192] proposed the Subject Invariant Generative Adversarial Network (SIS-GAN) to generate synthetic EEG signals for Steady-State Visual Evoked Potential (SSVEP) tasks. SIS-GAN aimed to remove subject-specific

information while preserving SSVEP frequencies. The architecture included a generator, a discriminator, an auxiliary classifier, and a pre-trained subject-biometric classifier to enforce subject invariance. Their method improved SSVEP classification performance compared to other augmentation techniques.

2. *Session Invariance:* Achieving session invariance requires a generator capable of filtering out session-specific features from a signal. This generator retains only the essential information within the feature space while disregarding session-related variations.

   ”Özdenizci et al. [194] proposed a method based on invariant representation learning for person identification using EEG signals. Their approach applied adversarial inference to learn both session-invariant and subject-invariant representations with long-term applicability. In within-session evaluations, their model achieved accuracies of $98.7\% \pm 0.005$, $99.3\% \pm 0.003$, and $98.6\% \pm 0.006$ for Sessions 1, 2, and 3, respectively. These results came from analysing 2,760 half-second EEG epochs using a 20% test split. The evaluation focused on measuring model performance within each session. For cross-session person identification, the authors tested their model on an unseen session and reported accuracies of up to 72% for 10-class person identification. This evaluation used 13,800 half-second epochs. They attributed improvements of up to 6% to the use of adversarial learning, which enhanced session invariance.

3. *Class Invariance:* In addition to filtering out subject and session-specific characteristics, achieving class invariance in classification problems requires a generator to eliminate class-related features from the signal effectively. Rommel et al. [60] introduced an advanced automated differentiable data augmentation approach for EEG data, comparing class-wise augmentation to class-agnostic augmentation.

3.2.2 Data Augmentation Techniques

This section provides a comprehensive overview of various data augmentation techniques for EEG signals from an extensive review of the previous research.

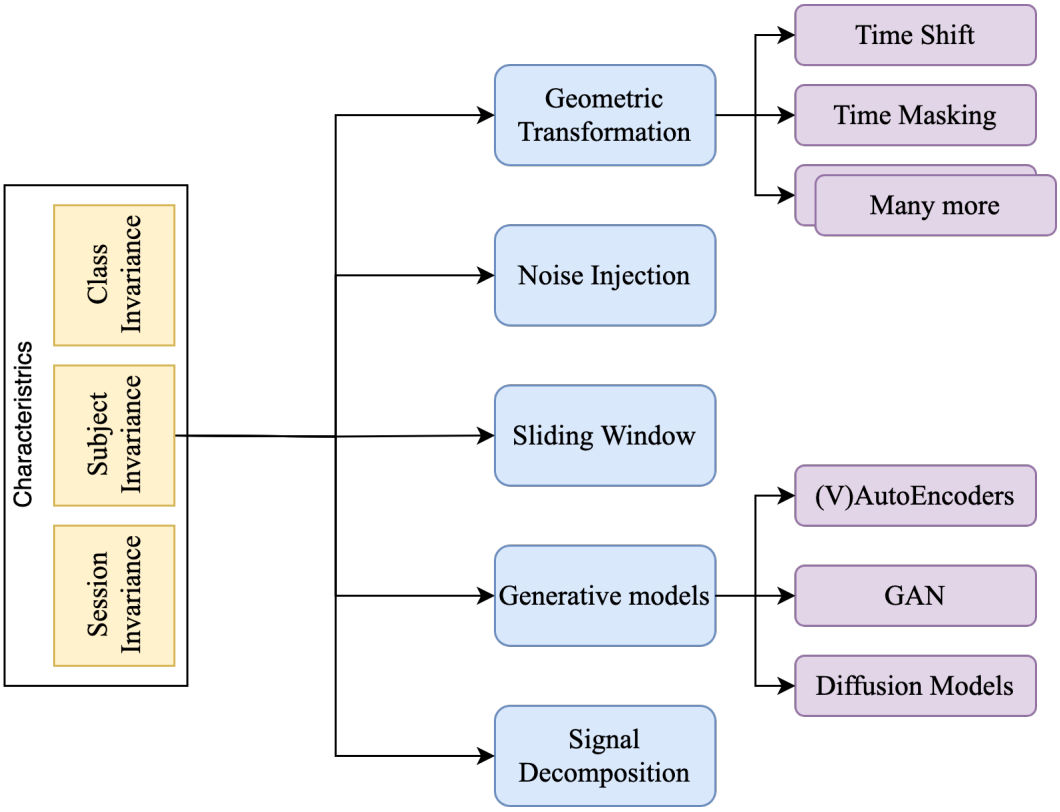


Figure 3.3: Taxonomy of data augmentation techniques for EEG signal processing.

Data augmentation enhances the robustness and generalisation of machine learning models by artificially expanding the training dataset. Based on the analysis, a proposed taxonomy outlines different data augmentation techniques tailored explicitly for EEG signal processing. These techniques encompass various approaches, including temporal augmentation, spatial augmentation, frequency-domain augmentation, and hybrid methods that combine multiple augmentation strategies. Each technique offers unique benefits and addresses specific challenges associated with EEG data, ultimately contributing to improved performance and adaptability of EEG-based machine learning models.

Figure 3.3 shows the taxonomy of data augmentation techniques explored in the literature. Research in EEG focuses on improving the robustness of the model on unseen data while showcasing the improvement with and without augmentation. This subsection discusses various data augmentation techniques for EEG classification.

3.2.2.1 Geometric Transformation (GT)

Geometric transformations, including flipping, rotation, and cropping, modify the spatial structure of input signals. These operations capture direction, contour information, and support tasks in speech signal processing [195] and computer vision [196]. K., Mario Michael and K. Su Kyoung [197] proposed rotational data augmentation for EEG signals by simulating spatial distortions through rotations around the y- and z-axes. Specifically, rotation around the y-axis tilts the EEG sensor layout side-to-side (left–right), while rotation around the z-axis simulates angular shifts across the scalp, such as twisting the cap clockwise or counterclockwise. These transformations aimed to mimic realistic electrode placement variations, including cap misalignments and subject movement, commonly occurring in practice. Evaluations on P300 and Movement-Related Cortical Potential (MRCP) paradigms using the BCI Competition IV Datasets 2a and 2b [129] showed that applying rotations of approximately $\pm 18^{\circ}$ improved classification accuracy by 1% to 6%. The study primarily focused on within-subject evaluations and did not explicitly address whether the proposed augmentation preserved session, subject, or class invariance.

Table 3.1 summarises common transformations used to augment EEG signals across time, frequency, and spatial domains. Table 3.2 reviews selected studies that apply these techniques for EEG classification tasks. Several studies report improvements in model generalisation across sessions and subjects. However, only a few works explicitly assess whether these augmentations preserve class-wise statistical properties of the signals, highlighting a key limitation in current evaluation practices.

Several studies evaluated the effectiveness of geometric transformations for EEG signal classification: Shovon et al. [203] designed a multi-input CNN for Motor Imagery (MI) classification. They converted EEG signals into spectrograms using Short-Time Fourier Transform (STFT) and applied rotation, flipping, and zooming as augmentation. Their method achieved 97% and 89.19% accuracy on the BCI III [204] and BCI IV 2b [129] datasets, respectively. However, the scalability of their approach to higher-dimensional EEG datasets remained uncertain. Freer and Yang [205] introduced a data augmentation framework for MI classification using a Convolution-LSTM model combined with Filter Bank Common Spatial Patterns

Table 3.1: Common geometric transformations used for EEG data augmentation.

Augmentation	Domain	Description
FTSurrogate [61]	Frequency	Randomize Fourier phases of all channels.
FrequencyShift [60]	Frequency	Randomly shift power spectral density (PSD) of all channels.
SignFlip [60]	Time	Randomly flip the sign of all channels.
TimeReverse [60]	Time	Reverse time axis of all channels.
ChannelsSymmetry [198]	Spatial	Swap signals between hemispheres to simulate brain symmetry.
ChannelsDropout [199]	Spatial	Randomly set selected channels to zero.
ChannelsShuffle [199]	Spatial	Randomly permute the order of EEG channels.
SensorsRotation [200]	Spatial	Interpolate signals on rotated virtual sensor positions.
TimeShift [201]	Time	Translate signal within a time window by -0.5 to 0.5 seconds.
TimeMasking [202]	Time	Zero out random time segments.

(FBCSP) features. The authors applied noise addition, constant multiplication, sign flip, and frequency shift as augmentations. Their method improved classification accuracy by approximately 5.3% on the BCIC IV 2a dataset [129]. Table 3.2 summarises these studies, listing the EEG paradigm, dataset, augmentation target, feature space, model architecture, and reported performance improvements.

3.2.2.2 Noise Injection (NI)

Generally, NI adds gaussian white noise to the original signal or the generated features as a common data augmentation technique. Since EEG signals exhibit a low signal-to-noise ratio, excessive noise can degrade the original signal. Therefore, applying NI requires careful consideration. Research works listed in Table 3.3 employ Noise Injection in training protocols and demonstrate performance improvements compared to models without noise injection across various EEG paradigms such as ER, Mental Workload Detection (MWD), and MI. Despite the effectiveness of noise injection in data augmentation, important questions remain regarding the optimal noise level and the methods for evaluating the augmented signal quality.

Table 3.2: Data augmentation using geometric transformations for EEG signal classification.

References	EEG Paradigm	Datasets	DA Chararacter istics	Feature Space	Geometric Transformation	Improvements
Shovon et al. [203]	MI	BCIC III (II), IV (2b)	Subject Invariance	Time-Frequency (STFT)	Rotation, Flipping, Zooming	NA to 89.9%, 97%
Freer and Yang [205]	MI	BCIC IV (2a)	Session Invariance	Time Series	Signal Flipping, Frequency Shifting	∼14% Avg Gain
Mokatren et al. [206]	ER	DEAP	Session Invariance	Wavelet Domain	Horizontal and Vertical Shifting	2.2%–5% Avg Gain
Rommel et al. [60]	MI, SS	Sleep-EDF	Class Invariance	Time-Frequency	Time Reversal, Sign Flip, Frequency Shift	Not Reported

3.2.2.3 Sliding Window (SW)

Models typically utilise the complete signal for classification problems. When applied to small datasets, the sliding window mechanism creates multiple instances from a single sample by setting the *window size* to a value smaller than the original signal length. Table 3.4 lists studies that generate multiple instances from a single sample using SW for EEG signal classification and demonstrate performance improvements with the sliding window approach across various EEG paradigms such as Seizure Detection (SD), Sleep Staging (SS), and MI. A sliding window is not an explicit augmentation technique as it does not significantly change the original signal. However, when looking at EEG data as a series of time-based signals, it is important to recognise that each segment may show some differences over time. Despite these variations, the key features used for classification remain somewhat similar across all time segments. Choosing the best window size becomes subjective depending on the dataset's characteristics, making it a hyperparameter that researchers carefully adjust.

3.2.2.4 Generative Models

Unlike deterministic transformation techniques, generative models learn the underlying data distribution and generate samples from this distribution, offering a more robust

Table 3.3: Data Augmentation using noise injection for EEG signal classification.

Reference	Model	EEG Paradigm	Dataset	DA Characteristic	Feature Space	Improvement
Wang et al. [207]	SVM	ER	SEED, MAHNOB-HCI	Session Invariant	DE	40.8% to 45.4%
Salama et al. [208]	3D-CNN	ER	DEAP	–	Time Series	79.11% to 89.4%
Li et al. [209]	Channel-Projection Mixed-Net	MI	BCIC IV (2a), HGD	–	Spectral Image	Avg of 1.1%
Yin and Zhang [55]	RNN	MWD	Proprietary	Session Invariant	Stacked Autoencoder (SAE)	34.2% to 75%
Kuanar et al. [210]	CNN	MWD	Proprietary	–	Spectral Image	NA to 93%
Zhang et al. [211]	Inception Net	MI	BCIC IV (2a & 2b)	Subject Invariant	Time Series	90% to 95%

approach that facilitates automatic end-to-end modelling. Determining whether the generated samples remain subject- or session-invariant poses a challenge. Nonetheless, Table 3.5 highlights the effectiveness of these models in EEG signal classification and their corresponding improvements.

1. *Generative Adversarial Network (GAN):* GANs generate artificial data through adversarial learning. In this process, two sub-networks, the Discriminator (D) and the Generator (G), strive to match the statistical distribution of the target data [232]. The discriminator differentiates between genuine and artificial input samples, whereas the generator produces realistic artificial sample distributions to deceive the discriminator. The output from the discriminator provides the likelihood of a sample being genuine. Probabilities near 0 or 1 denote distinct distributions, while values around 0.5 suggest challenges in discrimination. GANs effectively produce synthetic data resembling real distributions, as illustrated in Equation 3.1:

Table 3.4: Data augmentation using sliding window for EEG signal classification.

Reference	Model	EEG Paradigm	Dataset(s)	DA Character istic	Feature Space	Improvement
Majidov and Whangbo [212]	Regularized Multinomial Logistic Regression	MI	BCIC IV (2a & 2b)	NA	Time Series	NA to 80.4–82.39%
O'Shea et al. [213]	CNN	SD	Proprietary Clinic Dataset	NA	Spatial-Temporal	NA to 97%
Mousavi et al. [214]	CNN	SS	SleepEDF	Subject & Session Invariant	Time Series	82.9 to 85%
Avcu et al. [215]	CNN	SD	Proprietary Clinic Dataset	NA	Time Series	NA to 93%
Tayeb et al. [216]	CNN	MI	BCIC IV (2b)	NA	Spectral Image	NA to 84%
Tsiouris et al. [217]	LSTM	SD	CHB-MIT	Subject Invariant	Spatial-Temporal	70 to 80%

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{3.1}$$

Here, x represents a sample from the real data distribution $p_{\text{data}}(x)$, and z is a latent noise vector drawn from a prior distribution $p_z(z)$, typically gaussian or uniform. The generator G maps z to the data space to produce synthetic samples, while the discriminator D assigns a probability to indicate whether the input is real or generated.

Below are prominent GAN architectures adapted for EEG-specific data generation:

(a) *Deep Convolutional Generative Adversarial Network (DCGAN):* DCGAN replaces pooling layers with fractional-strided convolutions in the generator and employs stride convolutions in the discriminator [223]. Integrating this architectural design with adversarial training ensures adherence to

Table 3.5: Data augmentation using generative models for EEG signal classification.

Reference	Model	Paradigm	Dataset	DA Character istic	Technique	Feature Space	Improvement
Wei et al. [218]	GCNN	SD	CHB-MI	NA	WGAN	Time Series	72.11 to 95.89%
Chang and Jun [219]	DNN	ER	Proprietary	Subject & Session Invariant	GAN	Time Series	NA to 98.4%
Luo et al. [220]	SVM	ER	SEED, DEAP	Subject Invariant	cWGAN + sVAE	PSD + DE	NA to 90.8%
Zhu et al. [221]	CNN	ER	JAFFE	Subject & Session Invariant	CycleGAN	Time Series	Avg 3.7–8%
Kaper et al. [222]	CNN	MI	BCIC II (3)	NA	cDCGAN	Time-Frequency	78–83%
Zhang et al. [223]	CNN	MI	BCIC IV (1, 2b)	NA	DCGAN	Spectral Image	74.5–83.2%; 80.6–93.2%
Fahimi et al. [224]	EEGNet	MI	Proprietary	–	cDCGAN	Spectral Image	Avg 3.22–5.45%
Piplani et al. [225]	Boosting	MWD	Proprietary	NA	GAN	Spectral Image	90–95%
Hartmann et al. [226]	k-NN	MI	NA	NA	GAN	Spectral Image	NA
Luo and Lu [227]	CWGAN	ER	SEED	Subject Invariant	GAN	Spectral Image	Avg 3–20%
Zhang et al. [228]	CNN	MI	BCIC IV (2b)	–	GAN	Spectral Image	77–79%
Chang and Jun [219]	GAN	ER	CHB-MIT	Subject & Session Invariant	GAN	Time Series	97–98%
Zhang et al. [229]	CNN + LSTM	MI	BCIC IV (2b)	Subject Invariant	GAN	Time Series	NA to 76%
P. et al. [230]	CNN	RSVP	BCIT X2 RSVEP	NA	GAN	Time Series	0.7–2%
Arỳ et al. [231]	CNN	ER	MAHNOB-HCI	NA	AE	Wavelet	Avg ∼22%
Aznan et al. [68]	CNN	SSVEP	Proprietary	Subject Invariant	VAE	Time Series	Avg ∼35%
Zhang et al. [75]	CNN	ER	DEAP, DREAMER	Subject, Session, Class Invariant	GAN	Time Series	NA

feature distribution principles. Studies have demonstrated that DCGAN produces samples that are both diverse and closely resemble real data. For instance, in epilepsy seizure detection using EEG signals from the CHB-MIT dataset [166], a DCGAN-based augmentation was employed, followed by classification using ResNet50. This approach achieved a 3% performance improvement over non-augmented datasets [190].

(b) *Wasserstein Generative Adversarial Network (WGAN):* WGAN addressed the challenge of discontinuous GAN divergence for generator parameters, which could result in training instability or convergence issues [233]. Equation 3.2 defines Earth-Mover (Wasserstein-1) distance for stable training while replacing standard Jensen-Shannon divergence:

$$W(\mathbb{P}_r, \mathbb{P}_g) = \inf_{\gamma \in \Pi(\mathbb{P}_r, \mathbb{P}_g)} \mathbb{E}_{(x,y)\sim\gamma}[\mathsf{x} - \mathsf{y}] \tag{3.2}$$

where $\mathbb{P}_r$ and $\mathbb{P}_g$ denote the real and generated data distributions. WGAN enforces 1-Lipschitz continuity on the discriminator, stabilising training, as shown in Equation 3.3:

$$\min_G \max_{D \in \mathscr{D}} \mathbb{E}_{x\sim\mathbb{P}_r}[D(x)] - \mathbb{E}_{\tilde{x}\sim\mathbb{P}_g}[D(\tilde{x})] \tag{3.3}$$

(c) *Conditional Wasserstein Generative Adversarial Network (cWGAN):* In emotion recognition tasks using the SEED [1] and Dataset for Emotion Analysis using Physiological Signals (DEAP) [5] datasets, combining cWGAN with manifold sampling has enhanced classifier performance by approximately 10% [220]. cWGAN, a variant of WGAN, incorporates auxiliary label information by introducing the real label Y_r to both the discriminator and generator. The generator combines the latent input X_z with Y_r, while the discriminator merges the real input X_r or generated input X_g with Y_r. Equation 3.4 expresses the objective for cWGAN:

$$\begin{aligned}\min_{\theta_G}\max_{\theta_D} L(X_r, X_g, Y_r) =& \mathbb{E}_{x_r\sim X_r, y_r\sim Y_r}[D(x_r \mid y_r)] \\ &- \mathbb{E}_{x_g\sim X_g, y_r\sim Y_r}[D(x_g \mid y_r)] \\ &- \lambda \mathbb{E}_{\hat{x}\sim \hat{X}, y_r\sim Y_r}\left[\left(\|\nabla_{\hat{x}|y_r} D(\hat{x} \mid y_r)\|_2 - 1\right)^2\right]\end{aligned} \tag{3.4}$$

In this formulation, $\hat{x}$ denotes an interpolated sample between real and generated data points used for gradient penalty enforcement. The hyperparameter λ controls the strength of this penalty, encouraging smooth gradients and stable training.

2. *Autoencoders (AE):* Autoencoders [234] refer to feed-forward neural networks that encode raw data into low-dimensional vector representations through an encoder and then reconstruct these vectors back into artificial data using a decoder. Instead of outputting vectors in the latent space, the encoder of a VAE [235] produces parameters of a predefined distribution in the latent space for each input. The VAE enforces constraints on this latent distribution, ensuring its normality. Compared with standard Autoencoders, VAEs ensure that generated data adheres to a specific probability distribution. Komolovaitė et al. [236] introduced a synthetic data generator using VAE for stimuli classification employing EEG signals from healthy individuals and those with Alzheimer's disease. This technique exhibited a 2% improvement over models without augmentation.

3. *Diffusion Models:* Diffusion models introduce a fundamentally different generative approach by modelling data generation as a gradual denoising of noise over multiple time steps. Denoising Diffusion Probabilistic Model (DDPM) [28] formulate a two-part process: forward diffusion and reverse denoising processes. The forward diffusion process defines a Markov chain that progressively adds gaussian noise to the input data $\mathbf{x}_0$ over T steps, as shown in Equation 3.5:

$$q(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t;\ \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I}) \tag{3.5}$$

where β_t denotes a slight positive variance that gradually increases across

diffusion steps. The reverse denoising process aims to reconstruct the original signal by learning the conditional distribution $p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$. Instead of directly modeling p_θ, DDPMs reparameterize the objective by training a neural network ε_θ to predict the added noise, as shown in Equation 3.6. In most DDPM implementations adapted for EEG signals, the denoising model ε_θ adopts a U-Net backbone architecture [237].

$$\mathscr{L}_{\text{simple}} = \mathbb{E}_{\mathbf{x}_0,t,\varepsilon}\left[\|\varepsilon - \varepsilon_\theta(\mathbf{x}_t,t)\|_2^2\right] \tag{3.6}$$

In EEG signal processing, Torma and Szegletes [238] introduced Brain Activity-guided Unsupervised Domain Adaptation (BAUDA) to align EEG feature spaces across subjects. However, domain adaptation methods still heavily rely on real EEG samples. Diffusion models overcome this limitation by generating synthetic EEG data. Pascual and Gramfort [239] proposed a DDPM-based framework capturing temporal and statistical properties of EEG signals. Lu et al. [240] introduced EEG-Diffusion, enhancing subject-specific feature generation. Ye et al. [241] proposed Time Series Diffusion (TSDiffusion) for general time series, demonstrating strong performance on EEG datasets. These methods highlight the potential of diffusion models, especially U-Net-based architectures, to generate diverse, temporally coherent EEG signals that improve data augmentation and model robustness.

The results in Table 3.5 demonstrate that generative models, including GANs, VAEs, and diffusion models, consistently improve classification performance across various EEG paradigms. Techniques such as conditional GANs (cGAN) and Wasserstein GANs (WGAN) particularly show substantial gains in subject-invariant and session-invariant settings. While most studies report significant improvements in accuracy, the effectiveness often depends on the choice of generative model and the target feature space, highlighting the importance of model selection and training stability in generative data augmentation for EEG signals. However, these methods cannot guarantee that the subject-specific nuances are preserved when augmenting from a single subject's data, which may affect model generalisation to unseen subjects.

3.2.2.5 Signal Decomposition

Unlike other data augmentation techniques, signal decomposition in feature space is another technique that has shown some promising results. Generally, these techniques decompose a signal using methods such as Empirical Model Composition (EMD) or Fourier Transform (FT), transform the signal in the decomposed space, and then reconstruct the time domain signal. All the studies utilising Signal Decomposition have augmented signals on a per-class basis, thereby establishing class invariance.

Kalaganis et al. [183] proposed a DA method based on Graph-Empirical Mode Decomposition (Graph-EMD) to generate EEG data, which combined the advantages of the multiplex network model and the graph variant of classical empirical mode decomposition. They used a graph CNN to implement the automated identification of human states while designing a continuous attention-driving activity in a virtual reality environment. The experimental results demonstrated that investigating the EEG signal's graph structure could reflect the signal's spatial characteristics and that merging graph CNN with data augmentation produced more reliable performance. [229] suggested a new way to classify EEG data by combining DL and DA. The classifier consisted of Morlet wavelet features as input and a two-layered CNN followed by a pooling layer NN architecture. The author used EMD on the EEG record, mixing their Intrinsic Mode Functions (IMF) to create a new synthetic EEG record.

Huang et al. [242] proposed three different augmentation techniques: Segmentation, Time domain exchange, and Frequency domain exchange to generate more training samples. Combined with a CNN for training the classification model, these techniques yielded a gain of 5 to 10 % accuracy compared to pre-augmentation Schwabedal et al. [61] proposed a novel augmentation technique called FT Surrogates. FT Surrogates generated new samples by changing the signal phase by decomposing the signal using the Fourier Transform (FT) and then reconstructing it into the time domain using the Inverse Fourier Transform (IFT).

Data augmentation techniques aim to increase data diversity and improve model generalisation by creating synthetic samples. However, despite their effectiveness, these methods do not directly address domain shifts due to individual subject variability, different sessions, or cross-dataset settings. Transfer learning provides

a complementary strategy to further enhance model adaptability across such variations. Transfer learning leverages prior knowledge from a related domain to improve performance in a target domain with limited labelled data. The following section explores transfer learning approaches developed for EEG signal classification.

3.3 Transfer learning for EEG Signal Classification

Transfer learning is a method in which a model trained on a source domain assists learning in a target domain with limited data availability. Continuous variations in human physiology and psychology require frequent model updates.

1. Human physiology and psychology exhibit inherent dynamics as individuals undergo constant physical and mental fluctuations.

2. The profound inter-subject variability in human behaviour and characteristics accentuates the need for continuous model updates.

3. The acquisition of EEG signals, essential for decoding brain activity, introduces further complexity due to sensor variations from diverse manufacturers.

All these reasons require an FSL technique that can frequently update the model rather than retraining from scratch. Transfer Learning is common in FSL, where prior information is transferred from the source task to the few-shot task through transfer learning methods [243, 244, 245]. Figure 3.4 illustrates the entire transfer learning procedure, involving pre-training using similar datasets and finetuning the pre-trained model for the downstream data. The following subsections present various characteristics that can be transferred, followed by the techniques explored in the literature for transfer learning.

3.3.1 Transfer of Characteristics

Transfer Learning enables the transfer of diverse data characteristics across datasets. This section reviews the characteristics transferred in the context of EEG signals and highlights the corresponding research efforts.

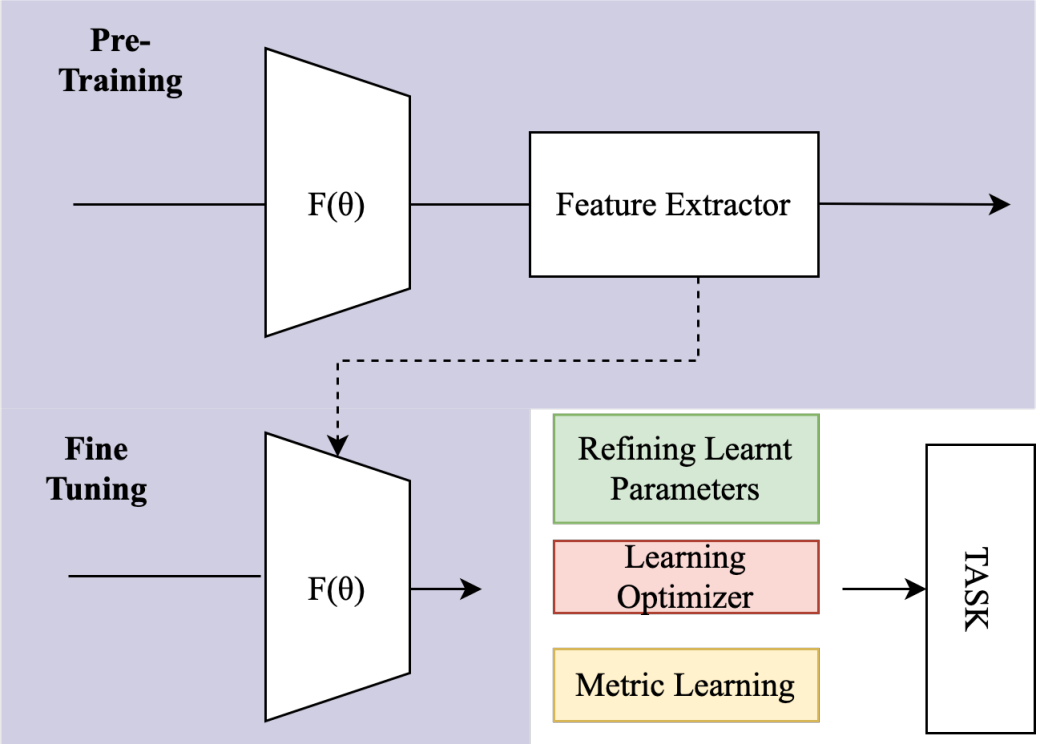


Figure 3.4: Process flow of transfer learning.

1. *Cross-Subject and Cross-Session Calibration:* Cross-subject and cross-session calibration actively combines information from source domains (other subjects or previous sessions) to calibrate a new target domain (new subject or session). Models transfer cross-subject knowledge by incorporating EEG data from multiple subjects performing the same task, directly enhancing generalisation. Consistency across sessions is maintained using historical session data to calibrate new session recordings, capture temporal dependencies and reduce session-specific biases. This combined approach reduces reliance on large quantities of subject-specific or session-specific labelled examples, improving few-shot learning performance in EEG classification tasks.

   Yang et al. [246] examined cross-subject emotion classification on the DEAP [5] and SEED [1] datasets. The study extracts ten linear and nonlinear features from each EEG channel, including standard deviation, power spectral density (alpha, beta, gamma, and theta bands), sample entropy, wavelet entropy, and Hjorth coefficients (activity, mobility, and complexity). Significance testing and sequential backward feature selection identify discriminative features. A Support Vector Machine (SVM) with a radial basis function (RBF) kernel performs the

final classification.

Fahimi et al. [56] proposed a CNN-based framework that first trains on EEG signals from source subjects, followed by finetuning using a small amount of calibration data from the target subject. The method filters EEG signals through a bandpass filter and extracts frequency bands, including delta, theta, alpha, beta, and gamma. This approach demonstrates strong generalisation ability across subjects by leveraging shared frequency band characteristics.

Li et al. [247] introduced a neural network model for cross-subject and cross-session EEG emotion recognition that eliminates the need for target domain labels. The network minimises source domain classification error and aligns latent representations of the source and target domains. The early network layers perform marginal distribution alignment through adversarial training, while the later layers align conditional distributions using association reinforcement mechanisms. This joint adaptation significantly improves classification performance across unseen domains.

Song et al. [48] proposed a Dynamical Graph Convolutional Neural Network (DGCNN) for cross-subject emotion recognition. The model uses differential entropy features extracted from five frequency bands. Each EEG channel forms a node in a dynamically constructed graph, with graph convolutional filtering capturing spatial relationships. A 1×1 convolutional layer learns channel-specific differences across frequency bands, followed by ReLU activation to enforce positive activations. Evaluation on the DREAMER [15] dataset reported recognition accuracies of 86.23% for valence, 84.54% for arousal, and 85.02% for dominance prediction.

Ning et al. [248] proposed the Subject-Domain Adaptation for Few-Shot Learning (SDA-FSL) framework for cross-subject EEG emotion recognition. The method integrates a Convolutional Block Attention Module (CBAM) -based feature mapping module, a domain adaptation module, and Prototypical Networks enhanced with an instance-attention mechanism. Evaluation of the DEAP [5] and SEED [1] datasets demonstrated improved cross-dataset performance. However, limitations include dependence on sufficiently large calibration sets

and interpretability challenges associated with attention-based adaptations.

2. *Cross-Device:* Cross-device transfer uses data from a source EEG device (source domain) to calibrate another EEG device (target domain), ensuring both devices perform the same task using common electrodes.

   Lan et al. [249] utilised various EEG devices, each with different numbers of electrodes, to record data from the DEAP [5] and SEED [1] datasets. The authors collected samples from different sets of individuals. The study focused on the 32 channels shared by both datasets, each containing three trials corresponding to positive, neutral, and negative outcomes. The final feature set for each channel involved extracting and combining five distinct frequency bands. The study demonstrated that applying domain adaptation techniques [250], particularly Transfer Component Analysis (TCA) [251] and maximal independence domain adaptation, significantly augmented classification accuracy. Deep learning approaches greatly enhanced cross-device transfer learning in BCI applications. In particular, transforming EEG data into images before feeding into deep models standardised the outputs across different devices.

   Similarly, Siddharth et al. [252] engaged in cross-dataset emotion categorisation, employing various modalities such as EEG, electrocardiography (ECG), and facial expression recognition (FAR). The authors discussed their EEG-based deep learning technique for emotion classification, where the model trained on DEAP [5] and tested on the MAHNOB-HCI dataset [253]. This approach remained effective despite differences in the number and placement of electrodes and variations in sampling rates across devices. The study extracted PSD of theta, alpha, and beta bands from EEG signals for each trial. It generated topographic PSD images for each trial by blending information from multiple EEG devices using an alpha blending ratio to construct colour images. VGG-16 [254] extracted 4,096 features from these images, which were then reduced to 30 using Principal Component Analysis (PCA). After pre-training, the authors employed an extreme learning machine as the final classifier.

   Cimtay and Ekmekcioglu [255] utilised an advanced CNN model, Inception-ResNet-v2 [256], to transfer information across subjects and datasets. The study

added gaussian random noise to expand the number of channels from 75 to 80 to match the input size requirements of Inception-ResNet-v2, structured as $(N_1, N, 3)$, where N_1 represents the number of EEG channels and $N \geq 75$ denotes the number of time-domain samples. An $80 \times 300 \times 3$ matrix was constructed and fed into Inception-ResNet-v2 for each trial. The output was then passed through a global average pooling layer and five dense layers for classification purposes.

3. *Cross-Task:* : Cross-task calibration leverages labelled data from prior activities (source domains) to improve calibration for new but related tasks (target domain). In most scenarios, the subject and EEG recording device remain the same. For instance, left- and right-hand motor imagery data can calibrate foot and tongue MI tasks, enabling broader generalisation across task types. The key challenge lies in aligning feature distributions between tasks that differ in cognitive or motor context. Several feature alignment strategies address the domain shift between EEG datasets by aligning either feature distributions or their statistical properties. Label Alignment (LA) aligns class-level distributions by re-centring the source class features to match the target domain using class-wise covariance matrices. This approach handles cases where source and target domains may not share identical label sets, enabling class-specific transfer. Riemannian Alignment (RA) leverages the geometry of symmetric positive definite matrices, which naturally represent EEG covariance structures. RA performs alignment on the Riemannian manifold by computing geodesic distances, preserving the intrinsic structure of the data. In contrast, Euclidean Alignment (EA) simplifies the process by treating the feature space as Euclidean and directly aligning the means and covariances of source and target feature distributions. He and Wu [257] compared LA, RA, and EA in a cross-task calibration using EEG-based MI. Subjects performed left- and right-hand MI as source tasks, while foot and tongue MI were the target tasks. LA effectively adapted using only one target label per class by estimating and aligning class-wise statistics. RA preserved the manifold structure of EEG signals, whereas EA prioritised computational efficiency. Each method showed distinct advantages depending on the representational structure and task similarity

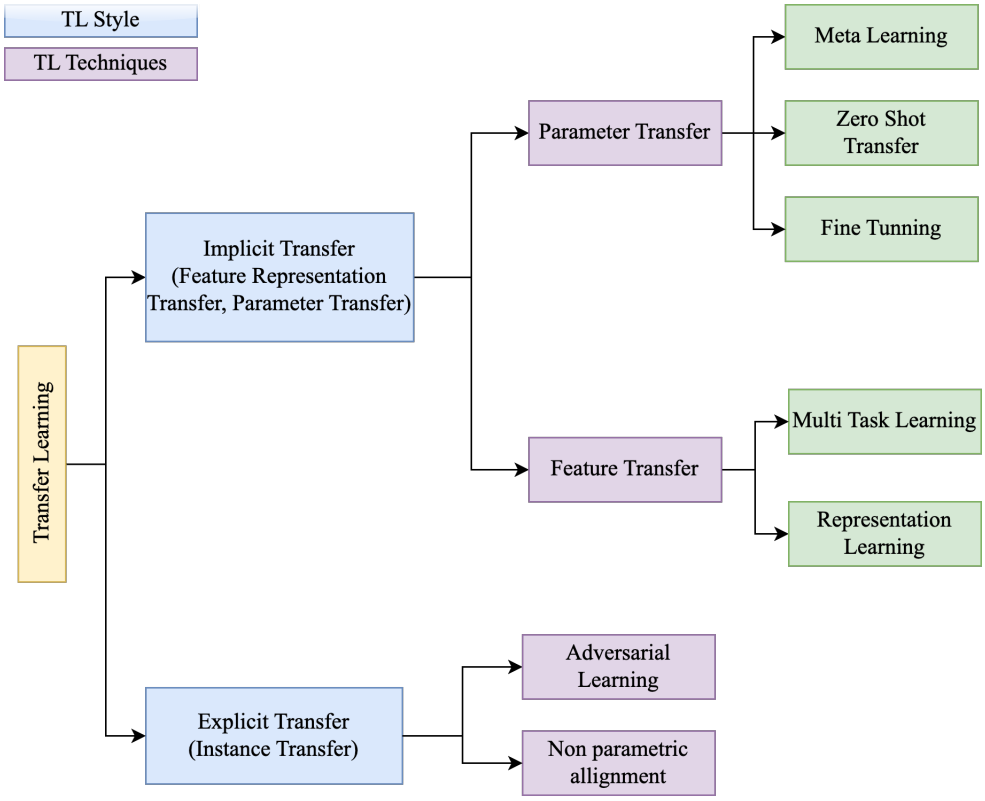


Figure 3.5: Taxonomy of transfer learning.

[258].

Zheng et al. [259] proposed a task-transfer method to enhance usability in MI-BCI applications. The study enlarged the MI command set by combining traditional MI tasks, expanding the range of BCI commands available to users. A transfer learning-based feature extraction algorithm supported this framework, significantly reducing calibration time while improving classification accuracy, particularly on low-quality datasets. The proposed approach demonstrated superior performance over traditional methods, especially under limited training samples. However, the study focused exclusively on MI tasks without exploring broader categories.

3.3.2 Techniques for Transfer learning

The categorisation of transfer learning by [260] defines two key categories: *Implicit Transfer Learning (ITL)* and *Explicit Transfer Learning (ETL)*. ITL transfers knowledge without directly aligning domains and encompasses methodologies such

as Representation Learning [261], Finetuning [262, 56, 263, 264], and Meta-Learning [265], as illustrated in Figure 3.5. In contrast, ETL explicitly addresses distinctions between domains, such as subjects or training sessions, by mitigating disparities through feature space alignment during training. Similarly, ITL includes strategies where domain shifts are handled implicitly, while ETL includes techniques such as Nonparametric Alignment [266] and Adversarial Learning [267]. Further categorisation by [268] classifies transfer learning into instance, parameter, and feature transfer. Domain adaptation through feature and parameter transfer, occurring without explicit instance alignment, falls under ITL, while instance transfer, which explicitly aligns instances across domains, corresponds to ETL.

3.3.2.1 Implicit Learning

Implicit learning refers to transfer strategies that adapt through internal representations, shared parameters, or latent embeddings rather than explicit domain alignment. These approaches assume that a model trained on a source domain can learn generalisable patterns that extend to a new domain with minimal adaptation. Implicit learning encompasses parameter transfer, meta-learning, and feature transfer, each focusing on internalising transferable knowledge rather than transforming the data. It makes implicit learning especially appealing for EEG-based systems where obtaining large-scale labelled data for every subject or condition is impractical. Reusing model parameters or embedding spaces learnt from previous domains enhances generalisation, reduces calibration effort, and improves performance in cross-subject and cross-session EEG classification. The following list outlines the key paradigms of implicit learning: parameter transfer, feature transfer, and representative EEG-based implementations.

1. *Parameter Transfer:* Implicit learning forms a core approach within transfer learning, encompassing the transfer of pre-trained parameters, latent spaces, or task-agnostic representations. Parameter transfer enhances performance and enables efficient adaptation across diverse ML applications, particularly in EEG signal processing. Among various transfer learning strategies, finetuning remains a widely adopted technique, where a pre-trained model from a source domain adapts to a target domain with limited labelled data. Finetuning refines model

parameters using a small dataset, improving the model's ability to extract meaningful representations from EEG signals. In contrast, zero-shot transfer leverages semantic representations to infer target tasks without explicit retraining, enabling knowledge transfer across domains. Meta-learning further enhances adaptability by training meta-learners to rapidly adjust to novel tasks, even with minimal labelled samples, thus promoting efficient few-shot adaptation in EEG-based systems.

(a) *Finetuning:* Finetuning first pre-trains a model on a source dataset and then adapts it on a smaller target dataset to capture domain-specific characteristics [269]. Finetuning strategies vary from updating the whole network end-to-end to updating only selected layers (e.g., classification heads or intermediate layers). Table 3.6 summarises the developments across different finetuning mechanisms for ITL.

(b) *Meta Transfer:* Meta-learning improves conventional transfer mechanisms by optimising the learning process, enabling efficient adaptation to unseen tasks with minimal data. A meta-learner acquires task-agnostic adaptation strategies, while a base-learner rapidly adjusts to task-specific nuances. In DL-based BCI applications, meta-transfer learning demonstrates strong potential for handling subject-specific and session-specific variability [23, 270, 271, 272, 273].

Duan et al. [270] employed a Model-Agnostic Meta-Learning (MAML) [265] framework to seek optimal initialisations that adapt rapidly through a few gradient steps. Equation 3.7 defines the MAML objective, which minimises the expected loss across tasks after inner-loop adaptation:

$$\min_{\theta} \sum_{\mathscr{T}_i \sim p(\mathscr{T})} \mathscr{L}_{\mathscr{T}_i}(U_\theta(\theta)), \tag{3.7}$$

where θ denotes the initial model parameters, $\mathscr{T}_i$ represents a task sampled from $p(\mathscr{T})$, $\mathscr{L}_{\mathscr{T}_i}$ indicates the task-specific loss, and $U_\theta(\theta)$ refers to the adapted parameters after few inner-loop updates. Model Agnostic Meta Learning (MAML) optimises θ such that the model can rapidly adapt to

new tasks with minimal calibration data, improving generalisation for EEG classification.

Pati et al. [271] proposed an optimisation-based meta-learning framework that pre-trained a model on motor imagery data across multiple subjects and Finetuned it on unseen subjects with few-shot data. Their approach used EEGNet [6] and demonstrated superior classification accuracy compared to traditional finetuning methods. Li et al. [272] applied MAML for cross-subject motor imagery decoding, reporting improved adaptation to new subjects and sessions through few-shot training. Han et al. [274] introduced META-EEG, a meta-learning framework integrating intermittent freezing, and demonstrated robust generalisation across subjects in a Leave One Subject Out (LOSO) setting.

Despite these advancements, meta-learning approaches face challenges such as computational overhead, risks of overfitting, and difficulty capturing complex class-discriminative features. Simplified algorithms such as Reptile [275] attempted to reduce meta-optimisation complexity, although Xiaoli and Rosa [276] found limited benefits for low signal-to-noise EEG tasks compared to vision applications.

2. *Feature Transfer:* Feature transfer forms another pillar of ITL, enabling models to generalise by transferring features learnt from source domains without direct reliance on target task labels. Embedding Learning (EL) supervises the mapping of samples into a latent space where intra-class samples cluster together and inter-class samples separate [277, 278]. This embedding structure reduces hypothesis complexity and improves few-shot adaptability.

   (a) *Representation Learning:* Representation learning (or embedding learning) transforms each input sample $x_i \in \mathscr{X} \subseteq \mathbb{R}^d$ into an embedding $z_i \in \mathscr{Z} \subseteq \mathbb{R}^m$, ensuring that similar samples remain close in $\mathscr{Z}$, while dissimilar ones separate. This transformation shrinks the hypothesis space $\tilde{\mathscr{H}}$ and facilitates generalisation [277, 278]. Embedding strategies vary between task-specific, task-invariant, and hybrid forms. Suh and Kim [279] proposed Riemannian manifold learning for EEG feature representation,

improving spatial structure modelling across subjects. Yu et al. [53] proposed DeepSeparator, a sequence-to-sequence model that learns task-specific latent representations for EEG artifact removal. The model captures multiscale temporal features by employing inception modules within the encoder, enabling the latent space to disentangle clean EEG signal characteristics from noise. The decoder reconstructs artifact-free EEG signals from these learnt representations, optimising the model through a reconstruction loss.

(b) *Multi-Task Learning (MTL):* Multi-task Learning jointly optimises multiple related tasks to exploit shared structures and improve generalisation [280, 281]. In EEG applications, Autthasan et al. [282] demonstrated MTL by simultaneously learning signal reconstruction and classification tasks on motor imagery datasets, showing that MTL acts as an effective regulariser for small datasets and strengthens few-shot learning performance.

3.3.2.2 Explicit Learning

Explicit learning strategies aim to reduce domain shifts by directly aligning feature spaces between source and target EEG distributions. These methods operate without relying on labelled data from the target domain, making them particularly suitable for real-world applications where labelled EEG data is scarce or costly. Unlike implicit approaches that rely on downstream model adaptation, explicit learning focuses on preprocessing or intermediate transformation to harmonise distributions. Two dominant strategies in this category include nonparametric alignment and adversarial training. Both techniques promote the development of domain-invariant representations and enhance generalisation across subjects or sessions. Table 3.7 summarises the research studies that utilise these strategies for cross-subject and cross-session EEG classification, demonstrating their versatility and effectiveness.

(a) *Nonparametric alignment:* In the context of EEG signal processing, addressing distribution shifts across subjects or sessions remains critical.

Table 3.6: Implicit Transfer Learning (ITL) via feature transfer and parameter transfer.

Reference	Model	Paradigm	TL Characteristics	IL Type	Datasets	Results	Main Contribution
Jeon et al. [283]	CNN	MI	CS	RepL	BCIC-IV-2a	–	PSD-based domain adaptation
Pal et al. [284]	Linear SVM	MI	CS	MultiT	BCIC-III-IVa	75.8%	Multi-objective optimization
Hossain et al. [285]	LDA	MI	CS	RepL	BCIC-IV-2a	67.7%	Informative TL with adaptation layer
Gaur et al. [286]	LDA	MI	CS	RepL	BCIC-IV-2a	–	Tangent space-based TL
Yair et al. [287]	Linear SVM	MI	CT	ZT	BCIC-IV-2a	–	Zero-shot TL using SPD cone manifold
Chiang et al. [288]	–	SSVEP	CS	PT	BCIC-IV-2a	82.1%	LS transformation-based finetuning
Xu et al. [289]	CNN	MI	CSS	FT	BCIC-IV-2b	74.2%	VGG-based finetuning
Behncke et al. [290]	CNN	ERP	CT	FT	Epilepsy patients	81.5%	Pretrained CNN finetuning
Pati et al. [271]	EEGNet + Meta-opt	MI	CS	MetaT	BCIC-IV-2a	55.6–63.1%	Meta-learnt init. for rapid subj. adaptation
Li et al. [272]	MAML + CNN	MI	CSS	MetaT	Physionet MI	60–80%	MAML for few-shot subj. adaptation
Duan et al. [270]	MUPS-EEG	MI, ER	CS	MetaT	BCI-IV-2a, DEAP	76.3%, 67.2%	Meta-updates for fast generalisation
Jaiswal et al. [291]	MeLaDA	ER	CS	MetaT	SEED	86.4%	Meta + adversarial DA with shift control
Ng and Guan [273]	Meta framework	MI, Inner Speech	CS	MetaT	Multiple datasets	88.7%	Fast adaptation with few target samples
An et al. [23]	CNN + Attention	MI	CS	MetaT	BCIC-IV-2b	74.6%	Relation scoring in meta-TL
Han et al. [274]	META-EEG	MI	CS	MetaT	Multiple datasets	+12%	Grad. meta-learning + feature freezing

CS: Cross-Subject, CT: Cross-Task, CD: Cross-Device, CSS: Cross Session & Subject, RepL: Representation Learning, FT: Fine Tuning, PT: Parameter Transfer, ZT: Zero-Shot Transfer, MetaT: Meta Training, MultiT: Multi-Task Transfer

Table 3.7: Explicit Transfer Learning (ETL) using nonparametric alignment and adversarial learning.

Reference	Model	Paradigm	TL Characteristics	EL Type	Datasets	Results	Main Contribution
Azab et al. [292]	LR+CSP	MI	CS	NPA	19 subj., BCIC-IV-2a, III-4a	70.3%, 75%, 75%	KL divergence for inter-subject similarity
Giles et al. [293]	CSP+LDA	MI	CS	NPA	BCIC-IV-2a	77%	Jensen-Shannon ratio + rule-based TL
Adair et al. [294]	Bayesian LDA	P300	CS	NPA	8 participants	62.5%	Ensemble of generic subjects
Wei et al. [295]	Clustering	SSVEP	CS	NPA	8 subjects	–	Fisher ratio for variability assessment
Hossain et al. [296]	LDA	MI	CS	NPA	BCIC-IV-2a	76%	Selective boundary-driven TL
Zhang et al. [297]	CNN	MI	CS	NPA	BCIC-IV-2b	–	p-hash-based instance-level TL
”Özdenizci et al. [298]	DANN cVAE	MI	CSS	AL	BCI2000	69.8%	CNN-based feature extractor in DANN
Zhao et al. [299]	DANN CNN	MI	CT	AL	BCIC-IV	83.98%	Centre loss for intra-class compactness
Tang and Zhang [300]	DANN CNN	MI	CT	AL	BCIC-IV	74.55%	Classifier output to domain discriminator
Jeon et al. [283]	DANN CNN	MI	CT	AL	HGD	92.5%	Source selection via resting-state EEG
Wei et al. [301]	DANN CNN	RSVP	CSS	AL	11 subjects	–	Subject ranking for source selection
Wang et al. [302]	SPD + CNN	ER	CSS	AL	DREAMER, DEAP	75.44%	Centroid alignment + source ranking

Nonparametric alignment provides a flexible technique to mitigate such differences. Unlike parametric methods, nonparametric alignment does not assume specific distributional forms or fixed models. Instead, it directly aligns the distributions of EEG signals between source and target domains by exploiting inherent data characteristics. This flexibility enables nonparametric alignment to handle the complex and diverse structures present in EEG datasets, improving its applicability in transfer learning scenarios. Nonparametric alignment reduces distribution shifts and promotes effective knowledge transfer even under heterogeneous domain conditions.

(b) *Adversarial training:* Adversarial training introduces a dual mechanism to achieve domain alignment. A discriminator network distinguishes between source and target domain samples, while a feature extractor network simultaneously learns to produce domain-invariant representations. The training objective minimises the discriminator's ability to separate domains while maximising the feature extractor's capacity to encode shared characteristics across domains. This adversarial process promotes cohesive feature alignment, enabling robust transfer learning in heterogeneous settings.

3.3.2.3 Hybrid Learning

Hybrid learning methods combine multiple alignment and adaptation strategies to overcome the limitations of standalone techniques in EEG domain adaptation. These approaches integrate explicit alignment with transfer learning or domain reweighting to create a more robust cross-subject and cross-session generalisation pipeline. By leveraging statistical alignment and learnt representations, hybrid methods aim to reduce domain shifts and preserve task-relevant features simultaneously.

Recent hybrid frameworks such as such as Double Stage Transfer Learning (DSTL) [303] and the Multi-layer Transfer Learning Algorithm based on Improved Common Spatial Patterns (MTICSP) [304] attracted significant attention due to their innovative methodologies and promising results. DSTL [303] addressed the challenge

of limited EEG data availability by employing a double-stage transfer learning strategy. The method first performed Euclidean alignment across EEG trials from different subjects and reweighted the aligned trials based on the distance between the covariance matrices of the source and target domains. After feature extraction using Common Spatial Patterns (CSP) [305], Transfer Learning-Principal Component Analysis (TL-PCA) [306] reduced domain discrepancies further. Experimental evaluations on two public datasets, BCIC IV-1 [47] and BCIC IV-2a [204], under Multi-Source to Single-Target (MTS) and Single-Source to Single-Target (STS) paradigms, reported superior classification accuracies of 84.64% and 77.16% for MTS, and 73.38% and 68.58% for STS. These results highlighted the effectiveness of DSTL in mitigating domain shifts and outperforming existing state-of-the-art methods for EEG classification.

Similarly, MTICSP [304] proposed a multi-stage adaptation pipeline to address challenges in decoding motor imagery tasks for BCIs. The method begins with Target Alignment (TA), standardising feature distributions between source and target domains by applying domain-wise covariance normalisation. This step reduces inter-subject variability by transforming the EEG signals from the target domain to better match the statistical properties of the source domain. Following TA, the method reweights the mean covariance matrices based on trial-level distances between source and target trials, allowing the model to emphasise more transferable examples. An improved CSP variant with a regularisation coefficient extracts spatial features while minimising domain-specific distortions. Finally, Joint Distribution Adaptation (JDA) aligns feature blocks between domains by minimising differences in both marginal and conditional distributions. Experimental results on proprietary five-subject and nine-subject datasets demonstrate the effectiveness of MTICSP, with classification accuracies of 80.21% and 77.58% under MTS settings and 80.10% and 73.91% under STS settings. These findings confirm MTICSP's strong performance in combining transfer learning with motor imagery decoding. However, DSTL and MTICSP depend on labelled calibration samples and task-specific alignment steps, limiting their generalisability beyond motor imagery. Recent research advocates for meta-learning, self-supervised learning, and uncertainty-aware techniques to overcome these limitations and enable robust generalisation across broader EEG paradigms.

Transfer learning techniques, encompassing both implicit and explicit strategies, play

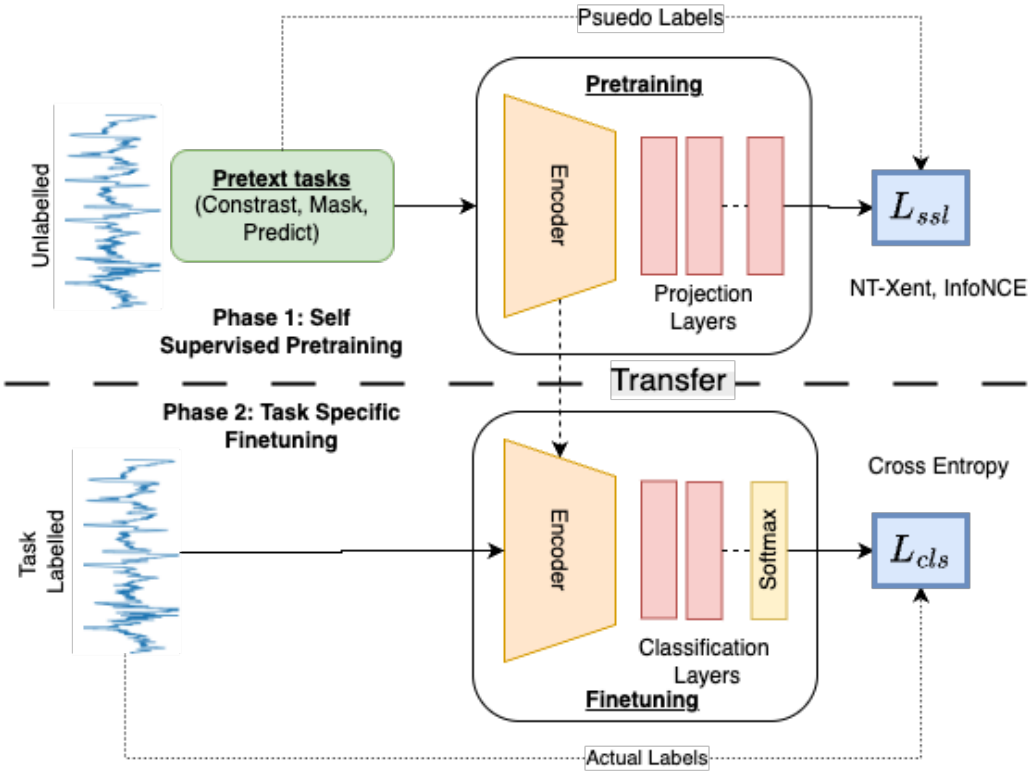


Figure 3.6: Overview of a SSL pipeline for EEG classification.

a critical role in improving generalisation across subjects, sessions, devices, and tasks in EEG-based classification. However, transfer learning methods still face challenges when adapting to highly personalised brain signal patterns or scenarios with severely limited labelled data. Motivated to address these limitations, researchers have increasingly focused on SSL and FSL paradigms. These approaches aim to further reduce dependency on extensive labelled datasets by leveraging intrinsic signal structures or rapid adaptation mechanisms. The following sections explore these emerging directions in EEG signal classification.

3.4 Self-Supervised Learning (SSL)

EEG classification faces persistent challenges such as low signal-to-noise ratio, high inter-subject variability, and limited access to labelled data. Data augmentation increases sample diversity but fails to fully address subject-specific variations and temporal inconsistencies. Transfer learning techniques, including finetuning and meta-learning, improve adaptability but depend on labelled datasets for pretraining. These

limitations create the need for learning meaningful EEG representations directly from unlabelled data. Self-Supervised Learning (SSL) addresses this need by enabling models to extract generalisable features without manual labels.

Figure 3.6 illustrates a typical self-supervised learning pipeline for EEG classification, consisting of two stages: *pretraining* and *transfer*. In the `pretraining stage`, the model receives raw EEG signals and performs one or more pretext tasks such as contrastive learning, temporal masking, or prediction. These tasks use pseudo-labels derived from the structure of the signal rather than external annotations. An encoder processes the input and extracts latent representations, which pass through projection layers designed to suit the pretext objective. The model optimizes a self-supervised loss $\mathscr{L}_{ssl}$ to learn invariant and meaningful features. In the `transfer stage`, the pretrained encoder transfers to a new model setup for classification. The EEG signal again passes through the encoder, followed by classification layers including a softmax output. This stage uses actual class labels and optimizes a classification loss $\mathscr{L}_{cls}$. The finetuning step adapts the representations learned from unlabelled data to improve performance on the labelled target task. Pretext tasks typically follow either contrastive or generative formulations [307, 39].

1. **Contrastive learning** trains the model to identify whether two samples originate from the same EEG trial. The system applies augmentations such as time shifting or frequency perturbation to generate two views of the same trial. It treats these as a positive pair and treats views from different trials as negative pairs. For example, two augmented versions of a 5-second EEG segment form a positive pair whose embeddings move closer in the latent space, while embeddings of unrelated segments move farther apart. The model optimises this contrastive objective using loss functions such as Normalized Temperature-scaled Cross Entropy Loss (NT-Xent) or Information Noise-Contrastive Estimation Loss (InfoNCE) [79, 80].

2. **Generative learning** trains the model to reconstruct or predict missing parts of the EEG signal. Masking-based tasks provide the model with an input that omits certain time steps, channels, or frequency bands. The model reconstructs the missing content from the surrounding context. Prediction tasks require the model to forecast future segments using past observations. These objectives force

the encoder and decoder to capture spatial, temporal, and spectral dependencies in the signal.

During pretraining, the model minimises a self-supervised loss $\mathscr{L}_{\text{ssl}}$ that arises from the selected pretext task. The model uses only unlabelled data in this stage. In the second stage, the pipeline attaches a task-specific classification head and finetunes the full model using a small labelled dataset. The training process minimises a supervised loss $\mathscr{L}_{\text{cls}}$ such as cross-entropy [258]. This two-stage process reduces reliance on large labelled corpora, improves robustness across subjects, and provides a scalable foundation for EEG classification.

3.4.1 Contrastive Learning Techniques

Contrastive learning forms a core component of self-supervised learning by training models to distinguish between similar and dissimilar examples in the embedding space. The objective encourages the model to pull together representations of positive pairs—typically augmented views of the same EEG trial—and push apart representations of negative pairs drawn from different trials. This mechanism helps the model learn a discriminative and structured latent space without using task labels. The most widely used loss functions in contrastive learning are the Normalized Temperature-scaled Cross Entropy Loss (NT-Xent) [79] and Information Noise-Contrastive Estimation Loss (InfoNCE) [80] losses, both of which operationalise this objective by comparing distances between positive and negative samples in a batch.

1. *Simple Contrastive Learning (SimCLR):* Chen et al. [308] proposed SimCLR, which learns representations by maximising the similarity between two augmented views of the same input. As illustrated in Figure 3.7, SimCLR applies random augmentations t_1 and t_2 to an input EEG segment x, resulting in two views x_i and x_j. The model then maximises agreement between their embeddings z_i and z_j.

   Equation 3.8 defines the Normalized Temperature-scaled Cross-Entropy Loss (NT-Xent Loss) used to train the SimCLR framework.

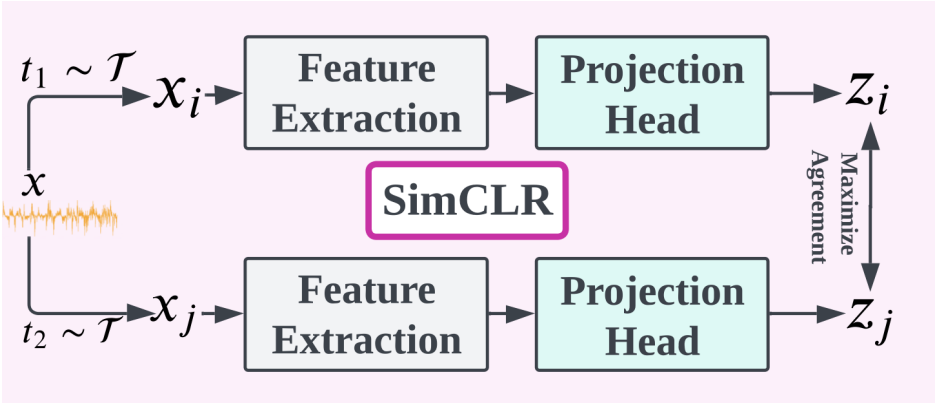


Figure 3.7: SimCLR framework: Two augmented views from the same EEG segment are encoded and projected to a latent space to maximise the similarity [13].

$$\mathscr{L}_{i,j} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\text{sim}(z_i, z_k)/\tau)} \tag{3.8}$$

where $\text{sim}(\cdot)$ denotes cosine similarity, τ is the temperature parameter, and $2N$ represents the number of examples in the batch.

Although SimCLR performs strongly in computer vision, its direct application to electroencephalography (EEG) remains challenging due to high subject variability and low signal-to-noise ratio. Several studies have adapted SimCLR to better handle EEG signals. Mohsenvand et al. [201] proposed a contrastive learning framework specifically designed for EEG signal classification. Their method employed time-domain augmentations, including noise injection and time shifting, to generate positive pairs for self-supervised pretraining. Each original EEG segment underwent random perturbations through additive gaussian noise or temporal shifts to create two correlated views in the proposed framework. These augmented views are passed through a shared encoder, and the model was trained to maximise agreement between the representations using a contrastive loss objective. The study demonstrated that this contrastive approach improved robustness in downstream sleep stage classification tasks, even when only limited labelled data were available. The findings highlighted that simple augmentation strategies combined with contrastive learning could significantly enhance feature representations in EEG signals and support more effective few-shot learning performance.

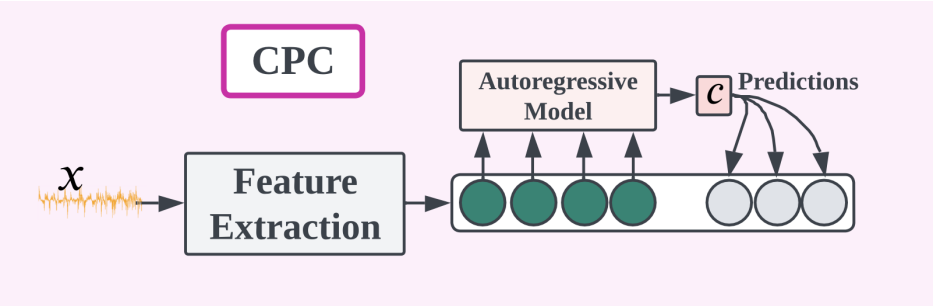


Figure 3.8: CPC framework: The model learns to predict future EEG feature representations based on the past context [13].

2. *Contrastive Predictive Coding (CPC):* van den Oord et al. [80] proposed CPC to model temporal dependencies in sequential data. Figure 3.8 shows CPC segments an EEG signal into temporal windows, encodes each segment into a latent representation, and aggregates past segments into a context vector using an autoregressive model. The model predicts future representations based on this context.

   Equation 3.9 expresses the InfoNCE Loss (Information Noise Contrastive Estimation Loss) optimised the CPC model.

$$\mathscr{L}_{\text{CPC}} = -\mathbb{E}\left[\log \frac{\exp(f_k(c_t, x_{t+k}))}{\sum_{x_j \in X} \exp(f_k(c_t, x_j))}\right] \tag{3.9}$$

   where c_t denotes the context vector at time t, x_{t+k} is the future sample to predict, x_j are negative samples, and f_k is a scoring function at step k.

   While CPC captures meaningful temporal structures, it does not explicitly address inter-subject variability, which is critical for robust EEG emotion recognition across datasets.

3. *Self-supervised Group Meiosis Contrastive Learning (SGMC):* Zheng et al. [14] introduced SGMC, a contrastive self-supervised learning framework tailored for EEG signals. SGMC addressed inter-subject variability by proposing a novel augmentation strategy called *meiosis-based grouping*. Instead of generating positive pairs through standard time or frequency augmentations, SGMC partitioned each EEG trial into semantically similar but non-overlapping temporal windows, referred to as *meiosis segments*. These segments served as

diverse yet coherent augmentations of the same signal. The model grouped these segments to form positive pairs and used segments from different trials as negative samples. This stimuli-aware grouping reduced reliance on handcrafted augmentations and captured intra-trial variability more effectively. SGMC demonstrated improved few-shot emotion recognition performance and achieved better generalisation across EEG datasets compared to traditional contrastive baselines such as SimCLR [79].

3.4.2 Generative SSL Techniques

Generative SSL techniques train models to reconstruct or generate EEG signals from corrupted or partial inputs. These methods encourage the encoder to capture meaningful spatial, temporal, and spectral patterns necessary for signal recovery. By solving such reconstruction-based tasks without supervision, the model learns representations that can transfer to downstream tasks such as classification or regression. Common generative tasks include input masking, temporal prediction, and autoencoding. In masking-based approaches, the model receives an EEG signal with selected time segments, channels, or frequency bands removed and learns to reconstruct the missing portions using the surrounding context. Temporal prediction tasks require the model to forecast future segments based on past observations, while autoencoder-based methods compress the EEG signal into a latent representation and reconstruct the full signal from it. During pretraining, the model typically minimises a reconstruction loss, such as mean squared error, that quantifies the difference between the original and recovered signal.

Several frameworks have explored generative self-supervised learning for EEG. Kostas et al. [39] introduced BENDR, a transformer-based model trained using a masked reconstruction objective inspired by BERT-style masking in NLP [27, 49]. The model learnt to predict masked temporal regions of EEG signals using surrounding context and was pretrained on the large-scale Temple University Hospital EEG Corpus (TUH) corpus [176]. finetuning on tasks such as P300 [204], BCI Competition IV [129], and Sleep-EDF [132] showed strong gains for P300 classification, but more modest improvements on motor imagery and sleep staging tasks. These results suggested that

the effectiveness of generative pretext tasks may depend on the downstream domain. Similarly, Zhang et al. [75] proposed GANSER, which also adopted a masked signal modeling approach. The model used an encoder–decoder architecture to reconstruct missing segments in unlabelled EEG inputs and was evaluated on emotion recognition tasks. GANSER outperformed contrastive baselines such as SimCLR and CPC, highlighting the potential of generative pretext tasks to capture rich temporal dynamics for affective state decoding.

Wang et al. [309] introduced BrainBERT, which extended masked modeling to naturalistic EEG to learn task-agnostic representations. The model improved neural decoding performance while maintaining interpretability, making it suitable for brain-machine interface applications. In a different approach, You et al. [25] combined generative adversarial training with few-shot learning in SleepGAN, which synthesised EEG signals for data-scarce sleep stage classification tasks and achieved gains on the Sleep-EDF dataset.

Li et al. [76] proposed the MV-SSTMA framework, which reconstructed masked EEG signals across spatial, spectral, and temporal views. MV-SSTMA achieved high supervised accuracy on the SEED dataset (95.32% ± 3.05), and maintained reasonable performance in semi-supervised settings with few labelled samples (81.04% ± 6.57). However, the evaluation remained limited to subject-dependent settings, and the model was not assessed in cross-subject configurations. As a generative model, MV-SSTMA may overfit to subject-specific signal characteristics such as baseline rhythms and artefacts, which hinders generalisation to unseen subjects.

While generative self-supervised methods consistently show strong performance in subject-mixed or subject-dependent settings, most frameworks have not been thoroughly evaluated for cross-subject generalisation. This lack of evaluation may stem from a tendency of generative models to overfit to individual subject characteristics during signal reconstruction. Since these models are optimised to recover detailed signal structure, they can inadvertently learn subject-specific patterns that do not transfer across individuals. As a result, despite their success in controlled or intra-subject scenarios, generative pretext tasks require further investigation to assess their robustness in generalisable EEG classification pipelines.

SSL is an effective solution for improving few-shot learning in EEG classification [307,

39, 309, 75]. Contrastive learning techniques, such as SimCLR [308] and CPC [80], enable models to learn structural and temporal relationships directly from unlabelled EEG signals. Generative tasks, including BENDR [39], BrainBERT [309], SleepGAN [25], MV-SSTMA [76], and GANSER [75], enrich latent feature representations by reconstructing masked or corrupted inputs. Each method addresses critical challenges inherent to EEG learning, including low SNR, high inter-subject variability, and the scarcity of labelled data. Despite these advances, limitations such as cross-dataset generalisation and robustness under real-world recording conditions persist, motivating continued research in this domain.

3.5 Research Gaps and Future Directions

Research in EEG signal classification using FSL has significantly progressed in recent years. However, some open challenges and research gaps still need to be addressed to improve the accuracy and robustness of the models. This section lays out the research gaps in each FSL technique for EEG signal classification—data augmentation, transfer learning, and self-supervised learning and provides the rationale for the top research gaps outlined in Section 1.3 of Chapter 1.

3.5.1 Data Augmentation Techniques in EEG

EEG-specific data augmentation remains a critical but under explored area. Existing approaches often rely on heuristics, lack theoretical guarantees, and fail to ensure consistency across subjects or transformations.

1. **Effective Ordering of Existing Data Augmentation:** EEG augmentations such as time shifting, frequency perturbation, signal flipping, and channel shuffling are applied in fixed or manually designed orders. Unlike vision, EEG's multivariate, nonstationary structure makes the ordering of transformations crucial for preserving class semantics. Exhaustive search is computationally infeasible, and current methods lack efficient algorithms to discover valid augmentation sequences [60, 66]. The thesis refers this research gap as RG1 in chapter 1.

2. **Subject-Invariance Guarantees of Data Augmentation:** Augmentations that disrupt subject-specific features can degrade performance on personalised or cross-subject tasks. Most literature applies augmentation after filtering and artifact removal [48, 68], which cannot adapt to nonlinear entanglement between subject identity and task-relevant features. No existing method jointly optimises augmentation while preserving subject-specific traits. The thesis refers this research gap as RG2 in chapter 1.

3.5.2 Transfer Learning for Subject and Device Adaptation

Transfer learning for EEG presents unique challenges due to high inter-subject variability and heterogeneous sensor configurations across datasets.

1. **Inadequate Mechanisms for Personalized Subject Adaptation:** Standard fine-tuning requires large labelled sets from new subjects and leads to catastrophic forgetting of previous knowledge. Although meta-learning [271] improves adaptation efficiency, it remains sensitive to the number and quality of adaptation samples. There is no lightweight architecture that enables rapid, few-shot personalisation without retraining the entire model. The thesis refers this research gap as RG3 in chapter 1.

2. **Inefficiencies in Few-Shot Transfer Learning for EEG Applications:** Existing transfer techniques perform poorly under low-resource conditions. Most frameworks are not optimised for real-time or embedded systems, as they require full-model finetuning and large annotated datasets [71]. A generalizable few-shot transfer pipeline that can adapt across EEG tasks remains an open challenge. The thesis refers this research gap as RG4 in chapter 1.

3.5.3 Self-Supervised Learning for Unlabelled EEG

SSL offers a promising direction for EEG representation learning without labelled data. However, existing models suffer from limited modality coverage, poor generalisation across datasets, and lack of invariance to subject variation.

1. **Lack of Multidimensional Representation Learning in SSL:** Most SSL models encode only one of the three core EEG dimensions—time, frequency, or space. BENDR [39] focuses on temporal encodings; GANSER [75] captures spatial information; SGMC [14] uses temporal contrastive learning but lacks spectral-spatial modeling. MV-SSTMA [76] proposes a multidimensional encoder but is limited by generative training and lacks subject-invariance. No existing model jointly optimises all three signal modalities using contrastive learning. The thesis refers this research gap as RG5 in chapter 1.

2. **Challenges in Cross-Subject Generalization within SSL:** SSL models often overfit to subject-specific patterns. While methods like DMMR [78] attempt to enforce invariance using multi-decoder architectures, subject generalisation remains limited. The field lacks subject-aware pretext task design and batch sampling strategies that mitigate inter-subject variability. The thesis refers this research gap as RG6 in chapter 1.

3. **Unified Foundation Model for EEG Signal Processing:** Unlike vision and NLP [28, 74], EEG lacks a foundation model that supports task-agnostic and device-agnostic transfer. Current models are often limited to specific datasets such as SEED [1], DEAP [5] and cannot generalise across domains [310, 1]. Also, the lack of cross-dataset pretraining limits real-world applicability of SSL in EEG. Models such as Cross-Domain EEG Emotion Recognition (CDEER) [311] and Multi-Source Domain Generalisation (MSDG) [312] explore multi-dataset setups, but no framework robustly handles device heterogeneity, subject imbalance, and task diversity in pretraining. Developing a unified foundational model, pre-trained on a diverse range of tasks, devices, and subjects, can significantly accelerate progress across various EEG applications. Such a model would facilitate seamless transferability across different tasks, datasets, and devices. The thesis refers this research gap as RG7 in chapter 1.

3.6 Summary

The literature review for FSL includes strategies such as data augmentation, transfer learning, and SSL for FSL in EEG signal classification. It presents a taxonomy-driven study for data augmentation techniques, including geometric transformations, noise injection, and generative models such as GANs, VAEs, and diffusion models, to help improve model robustness with small datasets. Similarly, it categorises transfer learning methods into explicit and implicit, addressing domain shifts and reducing calibration time in EEG powered applications. This literature review details the recent work towards few-shot learning using SSL for EEG signal classification, leveraging methods such as contrastive learning and generative models to enhance feature representation. It deep dives into a few recent foundation models for EEG trained using SSL such as BENDR [39] and BrainBERT [309] paving the future of EEG signal processing. Finally, it discusses the research gaps for these strategies, such as the need for automatic and subject-specific augmentations and cross-subject transfer while keeping the model robust to other subjects and leveraging heterogeneous datasets for few-shot adaptation.

This chapter is based on the following work:

Journal 1: Chirag Ahuja, and Divyashikha Sethia. "Harnessing Few-Shot Learning for EEG Signal Classification: A Survey of State-of-the-Art Techniques and Future Directions." **Frontiers in Human Neuroscience**, 18: 1421922 (2024). (**SCIE, Publisher: Frontiers Media SA**). Doi: https://doi.org/10.3389/fnhum.2024.1421922

Chapter 4

ADAPTER — Auto-Augmentation and PerTubation for Emotion Recognition

Deep learning models often rely on data augmentation to improve generalisation, especially when training data is limited or noisy. In image classification, simple geometric transformations such as cropping, flipping, and masking preserve semantic labels while enhancing robustness. These operations are easy to apply, compose sequentially, and visually verify. EEG signals, however, present a fundamentally different challenge. They exhibit low signal-to-noise ratios, encode discriminative features across spectral and temporal dimensions, and vary significantly across subjects. Even minor perturbations in amplitude, timing, or spatial topology can distort the underlying neural information. Unlike images, EEG lacks visual interpretability, making it challenging to ensure label consistency after transformation. Transformations that disrupt informative components often degrade classification performance.

Most EEG pipelines either avoid augmentation entirely or apply static transformation rules based on domain heuristics. These manually designed policies fail to account for class-specific signal characteristics—such as emotion-specific frequency patterns—and limit model generalisation across subjects. To address this limitation, this chapter describes **Automatic Data Augmentation and PerTubation for Emotion Recognition (ADAPTER)**, a class-specific augmentation framework that learns label-preserving transformations tailored to each class. ADAPTER applies task-relevant operations such as *Frequency Shift*, *Time Masking*, and *Channel Shuffle*, and optimises

their type, order, and intensity using a differentiable policy search. It samples class-conditional sub-policies and integrates directly into the training loop through bilevel optimisation. This design enables the model to discover augmentation strategies that improve generalisation while respecting class semantics in EEG signals.

4.1 Background and Motivation

Geometric transformations play a foundational role in data augmentation for image-based tasks. In these domains, simple transformations such as flipping, cropping, and colour jitter preserve class semantics and enhance model robustness against overfitting [313]. These transformations are safe to apply sequentially, and their visual nature allows easy inspection of augmentation effects. EEG signals differ fundamentally from images. They represent non-stationary, multi-channel time series with low signal-to-noise ratio and strong inter-subject variability [44]. Geometric transformations for EEG signals fall into three main domains: frequency, temporal, and spatial. Each transformation introduces structured variability while aiming to preserve class-relevant information and improve generalisation [314, 313, 205].

1. **Frequency-Domain Transformations**: Frequency-domain transformations modify the spectral structure of EEG signals while preserving their physiological characteristics. These are particularly relevant in emotion recognition and cognitive workload tasks, where power in specific frequency bands (e.g., alpha, beta, gamma) plays a key role [315].

   (a) *Frequency Shift* simulates spectral shifts caused by inter-subject variability or hardware drift. Equation 4.1 presents the shifting of the Fourier spectrum of the signal as:

   $$X'(f) = X(f - f_s) \tag{4.1}$$

   where, $X(f)$ is the Fourier Transform of the input signal $x(t)$, and f_s is the frequency shift value. The inverse Fourier transform reconstructs the perturbed signal $x'(t)$.

   (b) *Bandpass Perturbation* selectively amplifies or attenuates energy in specific

frequency bands. Equation 4.2 shows how the model applies a bandpass filter in the frequency domain and reconstructs the perturbed signal in the time domain:

$$x'(t) = \mathscr{F}^{-1}\left(H(f)X(f)\right) \tag{4.2}$$

where, $X(f) = \mathscr{F}[x(t)]$ is the Fourier Transform of the original signal, and $H(f)$ is a bandpass filter that perturbs target frequency bands. The inverse transform returns the modified time-domain signal $x'(t)$ [236].

(c) *Fourier Surrogate* introduces temporal variability by randomising phase while preserving the power spectral density of the original signal. Equation 4.3 expresses the representation of the EEG signal in the frequency domain:

$$X(f) = A(f)e^{j\phi(f)} \tag{4.3}$$

where $A(f)$ denotes the amplitude spectrum and $\phi(f)$ denotes the phase spectrum of the signal.

The method perturbs the phase to generate a surrogate signal while keeping the amplitude unchanged. Equation 4.4 explains the symmetrical phases shuffling:

$$\phi'(f) = \phi(\pi - f) \tag{4.4}$$

The Inverse Fourier Transform (IFT) reconstructs the time-domain signal $x'(t)$ using the original amplitude $A(f)$ and the perturbed phase $\phi'(f)$, producing a temporally varied but spectrally consistent surrogate [66].

2. **Temporal-Domain Transformations**: Temporal augmentations alter the signal along the time axis, introducing variations that mimic delays or missing segments while maintaining spectral structure.

   (a) *Time Shift* introduces a fixed delay or advancement in the signal. Equation 4.5 defines this operation:

$$x'(t) = x(t - \tau) \tag{4.5}$$

where τ is the shift offset sampled from a predefined range [316].

(b) *Time Warping* applies nonlinear distortions by stretching or compressing time segments. Equation 4.6 presents this transformation:

$$x'(t) = x(f_w(t)) \tag{4.6}$$

where $f_w(t)$ is a smooth time-warping function that modifies the input time index [317].

(c) *Time Masking* simulates dropout or transient signal loss. Equation 4.7 expresses this operation:

$$x'(t) = x(t) \cdot M(t) \tag{4.7}$$

where $M(t)$ is a binary mask that zeros out randomly selected time intervals [66].

3. **Spatial-Domain Transformations**: Spatial augmentations modify the channel-level structure, introducing variation across the electrode space. These help reduce sensitivity to sensor-specific biases or placement.

(a) *Channel Dropout* [316] removes a subset of channels to simulate sensor failure or reduce channel dependence. Equation 4.8 defines this operation:

$$X' = X \cdot M_c \tag{4.8}$$

where M_c is a binary channel mask with dropout probability p.

(b) *Channel Shuffle* [316] randomly reorders EEG channels to improve model robustness against electrode position variability. Equation 4.9 shows this transformation:

$$X' = \sigma(X) \tag{4.9}$$

where σ is a random permutation function applied across channels.

(c) *Sign Flip* inverts the polarity of selected channels to account for inconsistent

hardware configurations. Equation 4.10 shows this augmentation:

$$X' = X \cdot (-1)^{M_f} \tag{4.10}$$

where M_f is a binary mask indicating which channels undergo sign inversion.

These transformations simulate realistic noise and inter-subject differences, but uncontrolled augmentation can easily distort the fragile characteristics of EEG signals. Class-relevant information often resides in specific frequency bands or temporally localised patterns [314]. Even small perturbations in amplitude, timing, or spatial arrangement can degrade the signal structure and alter the class label. Unlike images, EEG signals provide no visual feedback to verify the integrity of augmented samples, making aggressive augmentation risky [205]. Despite this, geometric transformations remain among the most interpretable and practical techniques to improve EEG model generalisation under real-world variability [6].

AutoAugment [316] formulated policy learning as a discrete optimisation problem, using reinforcement learning to search over transformation sequences that improve validation performance. This approach discovered augmentation strategies empirically, but it incurs a high computational cost, as each policy evaluation requires training a complete model. Moreover, its discrete sampling procedure is non-differentiable and unsuitable for end-to-end training.

Differentiable Auto Data Augmentation (DADA) [66] addressed these issues by introducing a continuous relaxation framework. DADA replaces reinforcement learning with a differentiable architecture based on Gumbel-Softmax [318] for sub-policy sampling and Binary Concrete distributions for transformation application. It employed bilevel optimisation, where the classifier and augmentation policy co-trained in a nested loop, allowing direct gradient updates based on validation performance. However, DADA [66] assumes a single global policy for all classes and applies a fixed number of transformations, possibly leading to over-augmentation in EEG. These limitations are critical in the context of EEG signal classification. Different classes rely on distinct frequency bands, topographies, or temporal signatures [1, 3, 314]. A transformation that improves one class may degrade another. Moreover, EEG signals are fragile;

uncontrolled application of multiple transformations can distort the signal beyond recovery.

This thesis proposes Automatic Data Augmentation and PerTubation for Emotion Recognition (ADAPTER), a class-specific and differentiable augmentation framework for EEG signal classification. Existing approaches, such as AutoAugment [316] and DADA [66], treat the augmentation policy as a global process, applying the same transformation strategies across all classes. This assumption fails in EEG settings, where each class may exhibit distinct spectral, temporal, or spatial patterns. Applying irrelevant or excessive transformations can distort signal semantics, reduce label fidelity, and harm generalisation. Moreover, EEG signals offer no visual feedback to verify the quality of augmented data, making controlled and targeted augmentation critical. ADAPTER addresses these challenges by learning augmentation strategies conditioned on class labels, allowing the model to adapt the type, intensity, and number of transformations to the structure of each class. This class-aware formulation improves robustness in cross-subject EEG classification while enabling end-to-end training through differentiable sampling mechanisms.

This chapter discusses the proposed ADAPTER, a class-specific and differentiable EEG augmentation framework and its experimental validation on an open SEED [1]. ADAPTER learns separate transformation distributions conditioned on class labels. It selects the type, order, and intensity of augmentations using Gumbel-Softmax [318] and Binary Concrete [319] relaxations, enabling end-to-end training. By aligning augmentation strategies with class-dependent variability, ADAPTER improves generalisation in cross-subject EEG emotion classification. The main contributions of this chapter are:

1. **A class-specific augmentation framework for EEG**: ADAPTER introduces class-specific augmentation strategies that adapt to the characteristics of each signal label, unlike global strategies used in prior work such as AutoAugment [316] and Differentiable Auto Data Augmentation (DADA) [66].

2. **A differentiable policy learning formulation**: ADAPTER optimises transformation type and strength using Gumbel-Softmax [318] for sub-policy selection and Binary Concrete [319] relaxations for transformation application while

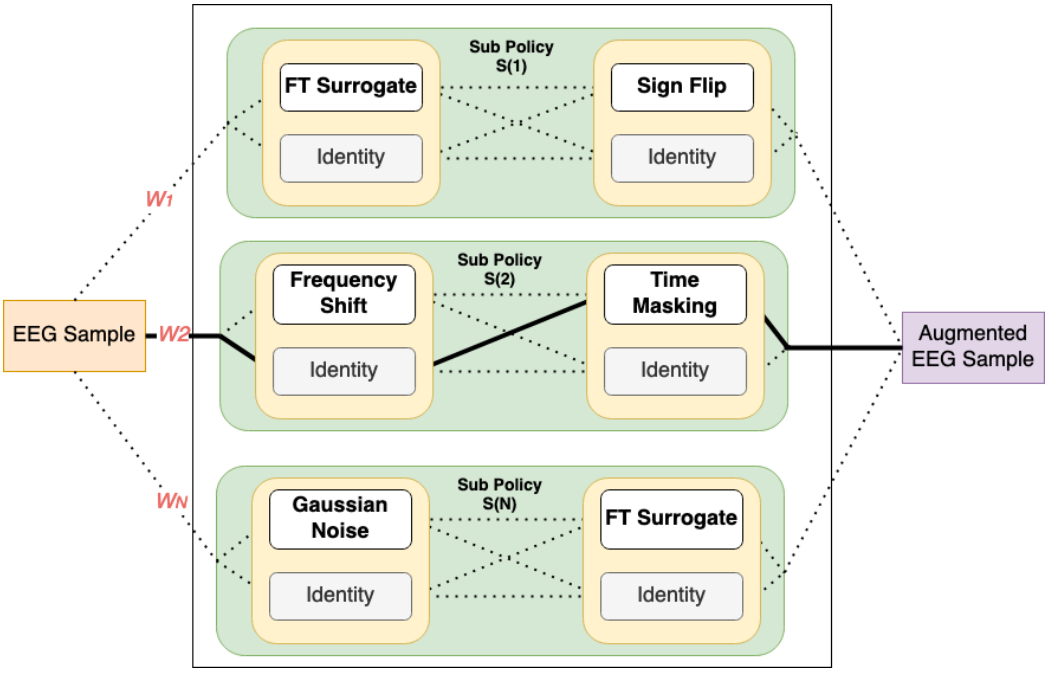


Figure 4.1: ADAPTER framework: class-specific sub-policy selection and stochastic transformation blocks.

controlling the number of applied augmentations.

3. **A class-specific mechanism**: The augmentation module conditions sampling on class labels during training, enabling the model to learn class-specific transformations specific to each category.

4. **Cross-subject evaluation on SEED**: Experiments on the SEED dataset [1] show that ADAPTER improves cross-subject generalisation by applying class-specific augmentation strategies that preserve signal semantics.

4.2 Methodology

This section presents the ADAPTER framework, which automatically learns class-specific data augmentation strategies for EEG-based emotion recognition. EEG signals are complex, non-stationary, and sensitive to transformations. A fixed or manually designed augmentation pipeline often fails to generalise across subjects or classes. Applying an inappropriate transformation—or applying transformations in the wrong order—can distort class-relevant information of the signal and harm model performance.

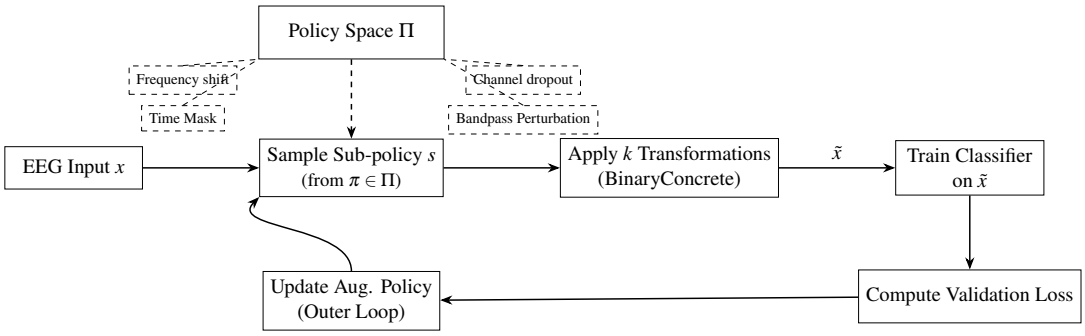


Figure 4.2: Workflow of the Automatic Data Augmentation and PerTubation for Emotion Recognition framework for automatic EEG data augmentation.

ADAPTER addresses this by framing augmentation as a sequence learning problem. The framework learns not only which transformations to apply but also how to order them and how to adapt them to each class. To achieve this, ADAPTER defines an augmentation *policy* as a distribution over a set of candidate transformation sequences, called *sub-policies*. Each sub-policy represents a fixed, ordered pair of transformations—such as Time Masking followed by Channel Dropout—that the model can apply to the input EEG signal. The policy learning process involves two key components: (i) selecting the most suitable sub-policy for a given input and class, and (ii) determining whether to apply each transformation within the selected sub-policy. Instead of learning these transformations directly, ADAPTER learns to sample sub-policies and control their application using class-specific probabilities. This modular structure allows the model to learn effective, class-aware augmentation strategies in a differentiable and interpretable way. Figure 4.1 illustrates the sub-policy structure in the ADAPTER framework. Given an EEG input $x \in \mathscr{X}$ with class label $y \in \mathscr{Y}$, the model selects one sub-policy $s \in \mathscr{S} = \{s_1, s_2, \ldots, s_N\}$ using a class-specific categorical distribution defined by weights $\mathbf{w}^y \in \mathbb{R}^N$. Each sub-policy s consists of a fixed sequence of k transformation blocks. Each block includes a stochastic gate that selects between a transformation (e.g., FT Surrogate, Time Masking) and the Identity function using BinaryConcrete relaxations.

4.2.1 Mathematical Formulation

The ADAPTER framework formulates this process as a class-conditional, differentiable augmentation policy. Each transformation block $\bar{T}_i^s$ is defined by two parameters: an application probability $\hat{p}_i^s \in [0,1]$ and a magnitude m_i^s.

Equation 4.11 shows the stochastic decision rule that governs whether to apply a transformation or retain the input:

$$\bar{T}_i^s(x \mid \hat{p}_i^s, m_i^s) = \begin{cases} T_i^s(x \mid m_i^s), & \text{with probability } \hat{p}_i^s \\ x, & \text{with probability } 1 - \hat{p}_i^s \end{cases} \tag{4.11}$$

A sub-policy s is defined as a composition of k such transformations, each with its application probability and magnitude parameters:

$$\begin{aligned} s(x) &= \bar{T}_k^s(x \mid \hat{p}_k^s, m_k^s) \circ \bar{T}_{k-1}^s(x \mid \hat{p}_{k-1}^s, m_{k-1}^s) \circ \cdots \circ \bar{T}_1^s(x \mid \hat{p}_1^s, m_1^s) \\ &= \bar{T}_k^s\left(\bar{T}_{k-1}^s(\ldots \bar{T}_1^s(x)\ldots)\right) \end{aligned} \tag{4.12}$$

Equation 4.13 expresses this stochasticity in terms of a Bernoulli trial:

$$\begin{aligned} \bar{T}(x \mid \hat{p}, m) &= b \cdot T(x \mid m) + (1-b) \cdot x \\ \text{where} \quad b &\sim \text{Bernoulli}(\hat{p}) \end{aligned} \tag{4.13}$$

The framework samples a sub-policy by drawing from a softmax distribution over logits $\alpha \in \mathbb{R}^N$. To enable class-aware learning, the framework extends these logits to class-specific forms, denoted as $\alpha^{(y)}$. Equation 4.14 shows this sampling process:

$$\begin{aligned} \bar{s}(x) &= \sum_{s \in \mathscr{S}} c_s \cdot s(x), \\ c &\sim \text{Categorical}(\pi), \\ \pi_s &= \frac{\exp(\alpha_s)}{\sum_{s' \in \mathscr{S}} \exp(\alpha_{s'})}, \quad \forall s \in \{1, \ldots, N\} \end{aligned} \tag{4.14}$$

Figure 4.2 shows the complete training loop. Given an input x, the model samples a sub-policy $s \sim \pi_y$, applies the stochastic transformations and generates an augmented sample $\tilde{x}$. The classifier parameters θ are trained on $\tilde{x}$ in the inner loop. The

augmentation policy parameters (e.g., α, $\hat{p}_i^s$) are updated in the outer loop using validation loss.

Equation 4.15 expresses this bilevel optimisation formally:

$$\begin{aligned} \min_{\pi} \quad & \mathscr{L}_{\text{val}}(\theta^* \mid D_{\text{val}}) \\ \text{s.t.} \quad & \theta^* = \arg\min_{\theta} \mathscr{L}_{\text{train}}(\theta \mid \pi(D_{\text{train}})) \end{aligned} \tag{4.15}$$

Here, θ represents the classifier parameters, and D_{train}, D_{val}, and D_{test} denote the training, validation, and test sets respectively.

To prevent over-augmentation and support class-aware learning, ADAPTER extends the DADA [66] formulation by conditioning policies on class labels. Unlike DADA [66], ADAPTER samples transformations within sub-policies and learns both sub-policy selection and block-level gates using Gumbel-Softmax and BinaryConcrete relaxations. The whole system optimises the bilevel objective in Equation 4.15 using gradient descent.

4.2.1.1 Glossary of Notations

The ADAPTER framework integrates several components, such as class-specific sub-policy selection, stochastic transformation application, and bilevel optimisation. These components draw upon mathematical tools including learnable parameters, probabilistic relaxations, and differentiable gates. To enhance clarity and maintain consistency in the subsequent formulations, Table 4.1 defines the key symbols used throughout the methodology. This glossary outlines the type and role of each symbol, serving as a quick reference for interpreting equations and algorithms discussed in later sections.

4.2.1.2 Sub-Policy Sampling and Application

Given an EEG input signal $x \in \mathbb{R}^{C \times T}$, the augmentation process begins by selecting one sub-policy $s \in \mathscr{S}$ from a set of N candidate sub-policies. Each sub-policy defines a fixed sequence of transformations that the model may apply to the input. The model assigns each sub-policy a learnable logit score α_s, which is converted into a probability

Table 4.1: Glossary of symbols used for the ADAPTER framework.

Symbol	Description
Input and Output Signals	
$x \in \mathbb{R}^{C \times T}$	Original EEG input with C channels and T time steps
$\tilde{x} \in \mathbb{R}^{C \times T}$	Augmented EEG signal after applying transformations
Augmentation Policy and Sub-Policies	
$\mathscr{S} = \{s_1, \ldots, s_N\}$	Set of N sub-policies
$s \in \mathscr{S}$	A selected sub-policy
$k \in \mathbb{N}$	Number of transformations in sub-policy s
T_i^s	Transformation at step i of sub-policy s
$\hat{p}_i^s \in [0,1]$	Probability of applying transformation T_i^s
$m_i^s \in \mathbb{R}_+$	Magnitude parameter of transformation T_i^s
$\bar{T}_i^s$	Stochastic transformation operator for step i
$\mathscr{A}_s(x)$	Augmented signal from applying sub-policy s to input x
Stochastic Relaxation Variables	
$\alpha_s \in \mathbb{R}$	Logit score for selecting sub-policy s
$\pi_s \in [0,1]$	Softmax-normalised probability of sub-policy s
$\tilde{c}^{(y)} \in \Delta^{N-1}$	Relaxed sub-policy selection vector for class y (Gumbel-Softmax)
$\tilde{b}_i^s \in (0,1)$	Relaxed gate for transformation T_i^s (Binary Concrete)
$\beta_i^s \in (0,1)$	Bernoulli parameter for transformation gate T_i^s
$\tau > 0$	Temperature for smoothing discrete relaxations
Model and Loss	
θ	Parameters of the classifier (e.g., SOGNN)
$P = \{\alpha, \hat{\mathbf{p}}, \mathbf{m}\}$	Augmentation policy parameters: logits, probabilities, and magnitudes
$\mathscr{L}(\cdot)$	Loss function (e.g., cross-entropy)
Datasets	
$D_{\text{train}}, D_{\text{valid}}$	Training and validation datasets

distribution using the softmax function, as defined in Equation 4.16:

$$\pi_s = \frac{\exp(\alpha_s)}{\sum_{s'} \exp(\alpha_{s'})} \tag{4.16}$$

This distribution enables the model to favour certain sub-policies over others, depending on the class or training dynamics. After sampling a sub-policy s, the framework applies its k transformations in a fixed order. Each transformation T_i^s is executed stochastically, based on a learnable probability $\hat{p}_i^s$ and a magnitude parameter m_i^s. Equation 4.11 presents the rule for applying transformation i in sub-policy s. This stochastic switch allows the model to skip unnecessary transformations and control the intensity of augmentation during training. Finally, the augmented signal $\tilde{x}$ is produced by composing all k transformation layers in sequence, as defined in Equation 4.17:

$$\mathscr{A}_s(x) = \bar{T}_k^s(x \mid \hat{p}_k^s, m_k^s) \circ \bar{T}_{k-1}^s(x \mid \hat{p}_{k-1}^s, m_{k-1}^s) \circ \cdots \circ \bar{T}_1^s(x \mid \hat{p}_1^s, m_1^s) \tag{4.17}$$

Each sub-policy introduces structured variability while maintaining class semantics. This design enables adaptive selection and tuning of transformations, allowing the framework to learn class-aware augmentation strategies that improve generalisation.

4.2.1.3 Differentiability Relaxations

ADAPTER performs two discrete operations that are non-differentiable in their standard form: (i) sampling a sub-policy from a categorical distribution and (ii) deciding whether to apply each transformation within the selected sub-policy. These operations prevent gradient flow through the augmentation pipeline, making joint optimisation with the classifier infeasible under standard backpropagation. To enable end-to-end training, ADAPTER replaces these discrete choices with continuous, differentiable relaxations. For transformation gating, the framework uses the Binary Concrete distribution [319]. This relaxation allows the model to interpolate between applying a transformation and using the identity function. Equation 4.18 defines the relaxed gate variable $\tilde{b}_i^s \in (0,1)$ for transformation i in sub-policy s:

$$\tilde{b}_i^s = \sigma\left(\frac{1}{\tau}\left(\log\frac{\beta_i^s}{1-\beta_i^s} + \log\frac{v_i}{1-v_i}\right)\right), \quad v_i \sim \mathscr{U}(0,1) \tag{4.18}$$

Here, $\beta_i^s \in (0,1)$ parameterises the probability of applying the transformation, and $\tau > 0$ is a temperature hyperparameter that controls the sharpness of the decision. As

$\tau \to 0$, the gating becomes increasingly binary. For sub-policy selection, ADAPTER employs the Gumbel-Softmax trick [318] to approximate categorical sampling in a differentiable manner. Instead of drawing a one-hot vector from a categorical distribution over sub-policies, the framework produces a soft probability vector $\tilde{c}^{(y)} \in \Delta^{N-1}$ using the following relaxation:

$$\tilde{c}^{(y)} = \text{softmax}\left(\frac{\alpha^{(y)} + \mathbf{g}}{\tau}\right), \quad g_i = -\log(-\log v_i), \quad v_i \sim \mathscr{U}(0,1) \tag{4.19}$$

In Equation 4.19, $\alpha^{(y)} \in \mathbb{R}^N$ are the class-specific logits that define the sub-policy distribution for class y, and τ controls the smoothness of the sampling. Lower values of τ lead to sharper, near-discrete selections. The model may anneal τ during training to transition from exploration to exploitation. These relaxations make the entire augmentation process differentiable. They allow ADAPTER to jointly optimise both the classifier and the augmentation policy using standard gradient-based methods.

4.2.1.4 Class-specific Sub-Policy Selection

To support class-specific augmentation strategies, ADAPTER conditions the sub-policy selection on the class label of the input EEG sample. Instead of using a single global distribution over sub-policies, the framework maintains a separate set of logits $\alpha^{(c)}$ for each class $c \in \{1, \ldots, C\}$. These logits parameterise a class-specific categorical distribution over sub-policies. During training, given a labelled input (x, y), the framework uses the class label y to select the corresponding logits $\alpha^{(y)}$. It then applies the Gumbel-Softmax [318] trick to sample a sub-policy from this distribution in a differentiable way. Equation 4.19 shows this class-conditioned relaxation. This class-specific sampling allows the model to learn distinct augmentation patterns for each class. For example, one class may benefit more from frequency-domain transformations, while another may benefit from time masking. ADAPTER captures such preferences automatically through this design. At inference time, the model applies the learnt class-specific policy using predicted labels, ensuring that augmentation remains aligned with the training strategy.

4.2.2 Data Preprocessing

The SEED dataset [1] provides 62-channel EEG signals recorded from 15 subjects at a sampling rate of 1000 Hz. The preprocessing pipeline begins by applying a bandpass filter between 0.5 Hz and 75 Hz to retain EEG activity in the delta to gamma frequency range while removing low-frequency drift and high-frequency noise. To remove non-neural artifacts such as eye blinks and muscle movements, the framework uses a linear dynamic system-based denoising method provided by the dataset authors [1, 3]. This artifact suppression improves the signal quality for downstream emotion classification tasks. After denoising, the signals are downsampled to 200 kHz to reduce computational cost during training and augmentation.

The dataset provides precomputed Differential Entropy (DE) features that serve as the input representation for the model. The preprocessing pipeline extracts DE features using a 512-point Short-Time Fourier Transform (STFT) with a non-overlapping 1-second Hanning window. It averages the STFT outputs across five standard EEG frequency bands: delta (1–3 Hz), theta (4–7 Hz), alpha (8–13 Hz), beta (14–30 Hz), and gamma (31–50 Hz). This process produces a feature matrix of shape $5 \times T$ for each trial, where T denotes the number of time windows. Since the trial lengths vary depending on the associated film clip, the time dimension T ranges from 185 to 265 across samples. To ensure uniform input dimensions, the framework zero-pads all trials to a fixed temporal length of 265 and forms a standard DE feature tensor of size 5×265 per trial. This pipeline aligns all inputs temporally, preserves spectral information, and suppresses artifacts, enabling consistent training and augmentation across subjects and classes.

4.2.3 Implementation Details

The ADAPTER framework implements automatic data augmentation using a class-conditional sub-policy learning strategy. Each sub-policy consists of a fixed sequence of $k = 2$ transformation blocks, where each block contains a stochastic gate for applying a geometric transformation or the identity operation. The framework selects transformations from a predefined set of EEG-specific augmentations, implemented using the Braindecode library [317]. These include operations such as Fourier

Algorithm 1 Training ADAPTER with bilevel optimization.

Require: Training set D_{train}, validation set D_{valid}
Require: Initial classifier parameters θ, augmentation policy parameters P
while not converged **do**
  Sample mini-batch $x \sim D_{\text{train}}$
  Sample sub-policy $s \sim \text{GumbelSoftmax}(\alpha)$
  for $i = 1$ to k **do**
    Sample gate $b_i \sim \text{BinaryConcrete}(\hat{p}_i^s)$
    Apply: $x \leftarrow b_i \cdot T_i^s(x \mid m_i^s) + (1 - b_i) \cdot x$
  end for
  Compute training loss $\mathscr{L}_{\text{train}}$
  Update classifier parameters θ using gradient descent
  Compute validation loss $\mathscr{L}_{\text{valid}}$
  Update policy parameters P via backpropagation through relaxed stochastic nodes
end while
return θ, P

surrogate, time masking, frequency scaling, temporal warping, and channel dropout. Each transformation operates on DE feature tensors of shape 5×265 and preserves the label semantics of the original trial.

During training, the model samples sub-policies using the Gumbel-Softmax relaxation over class-specific logits $\alpha^{(y)} \in \mathbb{R}^N$, where N denotes the number of candidate sub-policies. Within each sub-policy, transformation blocks apply with relaxed binary gates $\tilde{b}_i^s \in (0,1)$, derived from the Binary Concrete distribution. The temperature τ in both relaxations is annealed from 1.0 to 0.1 across training epochs to encourage sharper decisions. The classifier uses a Self Organizing Graph Neural Network (SOGNN) backbone [8] trained with a cross-entropy loss. The inner loop minimises training loss using the Adam optimiser with an initial learning rate of 0.001, a batch size of 64, and early stopping based on validation performance. The outer loop updates the augmentation policy parameters $\{\alpha, \hat{\mathbf{p}}, \mathbf{m}\}$ using gradient signals from the validation loss, propagated through the relaxed stochastic nodes. Training proceeds for 100 epochs, with temperature decay applied every 10 epochs.

4.2.4 Bilevel Optimization

ADAPTER trains the classifier and the augmentation policy jointly using a bilevel optimisation framework. Algorithm 1 describes the whole training procedure for the

ADAPTER framework. The training process uses a bilevel optimisation strategy. The inner loop updates the classifier using augmented training data, while the outer loop updates the augmentation policy based on validation loss. The algorithm uses Gumbel-Softmax and Binary Concrete relaxations to enable gradient flow through sub-policy and transformation selection.

The model begins each iteration by sampling a mini-batch x from the training set. It selects a sub-policy s using the Gumbel-Softmax distribution [318] over the learnable logits α. Each transformation T_i^s in the sub-policy applies stochastically. A relaxed binary gate $b_i \sim \text{BinaryConcrete}(\hat{p}_i^s)$ controls whether the model applies the transformation or retains the original signal. After applying all k transformations, the model computes training loss and updates the classifier parameters θ. It then computes validation loss and updates the augmentation policy parameters P using gradients propagated through the Gumbel-Softmax [318] and Binary Concrete relaxations [319]. This two-level optimisation process continues until convergence. Equation 4.20 defines the bilevel optimisation objective:

$$\min_{P} \mathscr{L}(\theta^* \mid D_{\text{valid}}), \quad \text{subject to} \quad \theta^* \in \arg\min_{\theta} \mathscr{L}(\theta \mid P(D_{\text{train}})) \tag{4.20}$$

where θ denotes the classifier parameters, P represents the augmentation policy parameters (i.e., $\alpha, \hat{p}_i^s$), $D_{\text{train}}, D_{\text{valid}}$ denote training and validation datasets, $P(D_{\text{train}})$ refers to training data augmented using policy P, and θ^* indicates the optimal classifier trained under the current policy.

This bilevel structure explicitly separates the optimisation goals: the inner loop finds classifier parameters that minimise training loss under a given augmentation policy, while the outer loop updates the augmentation policy to minimise validation loss. This formulation allows ADAPTER to align augmentation with generalisation objectives rather than overfitting to the training set. By using differentiable relaxations—Gumbel-Softmax for sub-policy selection and Binary Concrete for transformation gates [318]—ADAPTER enables end-to-end training of both the classifier and the augmentation policy using standard gradient-based methods. This structure allows the model to discover class-specific augmentation strategies that improve generalisation across subjects and emotion classes—an essential requirement

in EEG-based emotion recognition.

4.3 Experiments and Results

4.3.1 Experimental Setup

This section details the experimental setup that evaluates the effectiveness of the proposed ADAPTER framework on the SEED dataset [1] for EEG-based emotion recognition. It utilises SOGNN [8] as the predictive model and integrates ADAPTER into its training pipeline. ADAPTER aims to improve classification performance by learning class-specific augmentation strategies. By combining with a subject-invariant model like SOGNN [8], ADAPTER enhances robustness by learning class-specific augmentation strategies, which contributes to improved generalisation. Table 4.2 presents the performance gains over the baseline SOGNN [8] using LOSO cross-validation. In each fold, it randomly selects five subjects as the test set and uses data from the remaining subjects for training. The experiment includes all EEG frequency bands and relies on the official SOGNN implementation.[1] It wrap geometric augmentations implemented in Braindecode [317] with the proposed ADAPTER module. Model training runs on a Tesla A100 GPU and completes in approximately 37 GPU hours using PyTorch.

The ADAPTER training setup follows the optimisation procedure for SOGNN [8]. The training process minimises the model loss using the *Adam optimiser* with a *learning rate of 0.00001*, *weight decay of 0.0001*, and a *mini-batch size of 16*. During training, it applies *dropout with a rate of 0.1* to prevent overfitting. It trains both the classifier and the augmentation policy using the *bilevel optimisation algorithm* defined in Algorithm 1. This algorithm consists of two nested optimisation loops, which makes ADAPTER computationally demanding. To balance augmentation diversity with training efficiency, it restricts each sub-policy to contain *exactly three transformations*. It constructs *165 unique sub-policies* per emotion class by enumerating all three transformation combinations from the augmentation pool. During training, the *Gumbel-Softmax distribution* [318] operates over these sub-policy candidates to enable

[1] https://github.com/tailofcat/SOGNN

differentiable selection.

4.3.2 Results

Figure 4.6 shows the grouped probability distribution of all geometric transformations across all possible sub-policies for each class in the SEED dataset [1]. During training, the model tracks the average Area Under the Curve (AUC) from the receiver operating characteristic curve across all emotion categories. Once the average AUC reaches 0.99, the training process terminates. The framework then applies the ADAPTER's learnt strategy to the SOGNN [1], which includes 15 subjects. The evaluation conducts a LOSO experiment across all subjects. The validation process calculates the mean accuracy to evaluate the model's performance and facilitates comparison with other EEG-based emotion recognition models. This dual optimisation process allows the model to perform better on the target task and adapt more effectively to new data. The experimental results highlight that the proposed automatic data augmentation strategy ADAPTER significantly improves emotion recognition performance on the SEED dataset [1].

Table 4.2 presents the classification accuracies for different emotion categories under the LOSO cross-validation protocol. The results demonstrate that when integrated with SOGNN [8], ADAPTER consistently improves across all emotion classes, yielding an overall accuracy gain of approximately 1.85% compared to the baseline SOGNN [8] model. The improvement highlights the robustness of the proposed augmentation strategy in enhancing EEG-based emotion recognition. The learnt geometric transformations introduce meaningful variability in the training data, improving generalisation to unseen subjects. To further analyse model behaviour, this experiment computes confusion matrices that reflect classification performance across all emotion categories. The SEED [1] dataset contains EEG recordings from 15 subjects, each with 15 sessions. This work follows a Leave One Subject Out (LOSO) cross-validation protocol, where the model is trained on data from 14 subjects and tested on the remaining subjects. The process repeats 15 times, once for each subject. For every fold, the evaluation constructs a confusion matrix where each entry $M_{i,j}$ counts the instances with accurate label i that the model predicts as class j. The evaluation then

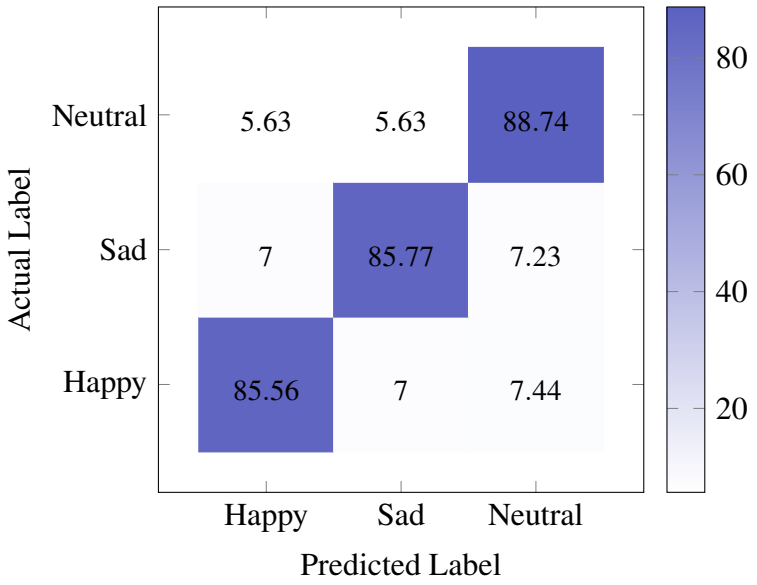


Figure 4.3: ADAPTER's normalised confusion matrix for the baseline SOGNN model on the SEED [1] dataset using LOSO cross-validation.

aggregates all confusion matrices from the LOSO folds to produce an overall confusion matrix. To enable fair comparison, the process normalises each matrix row-wise to ensure that every row sums to 100%. This normalisation converts raw prediction counts into percentages, which indicate the likelihood of each predicted class given a specific ground truth label.

Figures 4.3 and 4.4 present the final normalised confusion matrices for the baseline SOGNN [8] model and the ADAPTER-enhanced SOGNN [8] model, respectively. Each matrix displays the average percentage of predictions per emotion category. ADAPTER demonstrates a noticeable reduction in cross-class confusion and a consistent boost in per-class accuracy, further supporting its effectiveness for subject-independent EEG-based emotion classification.

Figure 4.5 illustrates the training and validation curves for the ADPTER enhanced SOGNN [8] model across the first three folds of the LOSO cross-validation setup, corresponding to the first three subjects in the SEED dataset [1]. Each row represents one fold and visualises the model's learning behaviour over training epochs regarding loss, accuracy, and AUC. The training loss decreases steadily in all folds while the validation loss begins to plateau—indicating convergence. Training and validation accuracy curves show rapid early improvements, with fold-specific gaps emerging later in training, suggesting varying degrees of generalisation and mild overfitting. Similarly,

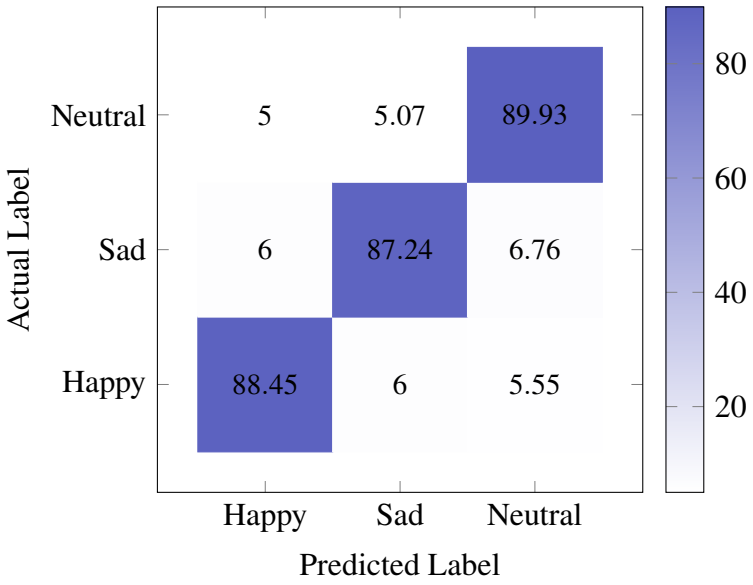


Figure 4.4: ADAPTER's normalised confusion matrix for the ADAPTER-enhanced SOGNN model on the SEED [1] dataset using LOSO cross-validation.

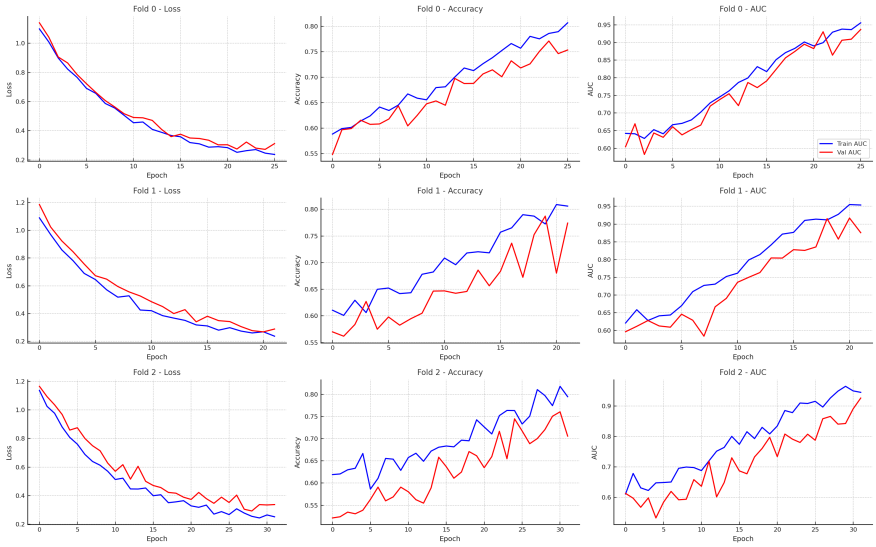


Figure 4.5: Training curves of ADAPTER enhanced SOGNN.

the Area Under Curve (AUC) curves climb sharply in the initial epochs and stabilise close to optimal values. The increasing discrepancy between training and validation performance across subjects highlights inter-subject variability, a well-known challenge in EEG-based emotion recognition. These trends confirm that the model is learning meaningful temporal-spatial features while demonstrating the importance of cross-subject robustness in evaluating EEG classification systems.

Moreover, the probability distribution of learnt transformations, shown in Figure 4.6, reveals meaningful patterns in how ADAPTER selects augmentations for

Table 4.2: Comparison of base model and ADAPTER on SEED [1].

Method	Average Accuracy	Happy (%)	Sad (%)	Neutral (%)
SOGNN [8]	86.69	85.56	85.77	88.74
ADAPTER + SOGNN	**88.54**	**88.45**	**87.24**	**89.93**

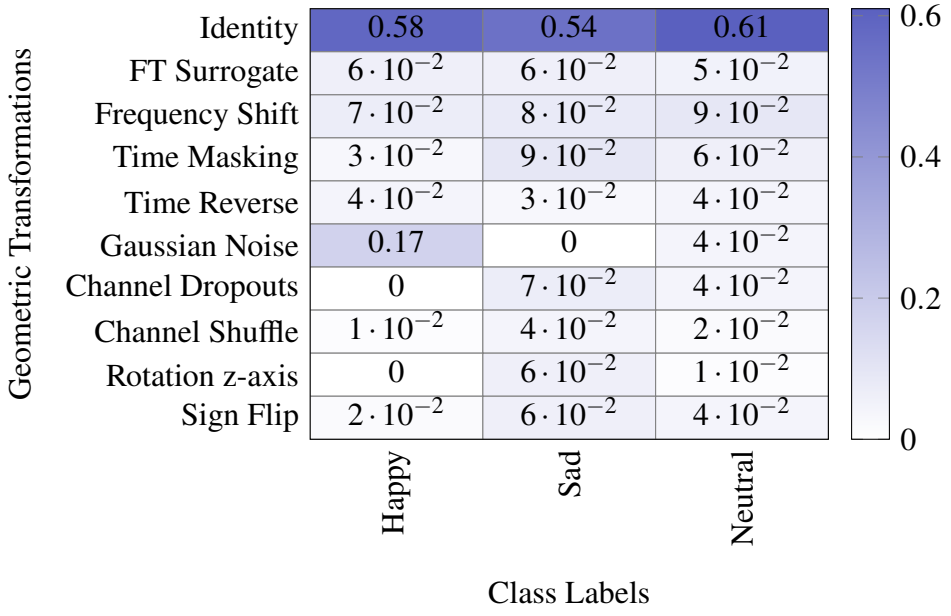


Figure 4.6: Estimated probability of geometric transformations learnt using ADAPTER.

different emotion classes. The heatmap illustrates that certain transformations occur more frequently for specific classes, indicating that the learnt augmentation policy aligns with class-dependent signal characteristics. The model effectively learns which transformations most improve class separability. For example, *FT surrogate* [61] and *frequency shift* appear frequently in the **happy** class. These transformations operate in the frequency domain and introduce mild temporal distortions, suggesting that frequency-related features play a key role in detecting **happy** states. In contrast, *Gaussian noise* and *channel dropout* are more commonly applied to the **sad** class, implying that enhancing robustness to noisy or spatially diffuse signals improves recognition of **sad** emotions. For the **neutral** class, the model often retains the original signal using the *Identity* transformation, suggesting that minimal perturbation benefits recognition of neutral states. Transformations such as `rotation`, `channel shuffle`, and `sign flip` appear infrequently, indicating that they contribute less to performance or may introduce distortions detrimental to certain emotion classes. These results

demonstrate that ADAPTER not only improves classification accuracy but also provides interpretable insights into the relevance of specific transformations for each emotion class. This understanding can inform the design of more effective preprocessing strategies and guide the development of class-aware augmentation pipelines for EEG-based emotion recognition tasks.

4.4 Summary

This chapter introduces a novel auto-augmentation ADAPTER framework tailored for EEG signal processing tasks, aiming to enhance the performance of state-of-the-art models by incorporating a data-driven augmentation approach during training. Previously, most auto-augmentation research focused on computer vision, the proposed framework is effective in EEG-based tasks. The learnt augmentation strategy is class invariant, and when combined with a subject-invariant model like SOGNN [8], yields a robust model invariant to both class and subject. The framework forms a solid foundation for robust and generalisable EEG classification models. A key advantage of the ADAPTER framework lies in its intuitive and controllable set of geometric augmentations, which offers better understanding and finetuning of the augmentation process compared to black-box generative models, all while maintaining interpretability. The proposed ADAPTER framework demonstrates significant promise for improving models in various EEG-based tasks, including emotion recognition. However, a notable limitation arises from its lack of subject-invariant transformations, which can hinder its effectiveness for cross-subject transfer in real-world adaptation scenarios.

This chapter is based on the following work:

Conference 2: **Chirag Ahuja**, and Divyashikha Sethia. "ADAPTER: Auto-Augmentation for Emotion Recognition in EEG—A Class and Subject Invariant Approach." **In Proc. International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, pp. 1-6, 2023**. (**Publisher: IEEE**). Doi: https://10.1109/ICCCNT56998.2023.10308153

Chapter 5

TRANSIT-EEG: A Framework for Cross-Subject Classification with Subject Specific Adaptation

The previous chapter introduced Automatic Data Augmentation and PerTubation for Emotion Recognition (ADAPTER), which uses simple geometric augmentations optimally to enhance training diversity but does not explicitly adapt the model to a new subject. In EEG signal classification, the low signal-to-noise ratio makes it necessary for models to adapt specifically to each new subject to achieve good performance. Hence, this chapter describes a novel unified new subject transfer framework called Transfer and Robust Adaptation of New Subjects in EEG Technology (TRANSIT-EEG). At TRANSIT-EEG's core, there are primarily three innovations that allow subject adaptation. First, it proposes a subject-specific data augmentation technique - Individual-specific Diffusion Probabilistic Model (IDPM). Second, it improves the existing cross-subject model SOGNN [8] using a Graph Attention Transformer (GAT) and proposes a Self-Organizing Graph Attention Network (SOGAT). Third, it enhances the model adaptation through finetuning of IDPM for generating synthetic data for new subjects and uses it for finetuning the cross-subject task classifier - SOGAT, using LoRA [320] with only a limited number of trials of the new subject.

5.1 Background and Motivation

Chapter 1 highlights that building robust EEG classification models across different subjects is challenging due to data scarcity, low signal-to-noise ratio of EEG signals, and high subject variability. Standard supervised deep learning models require extensive subject-specific data to generalise effectively. However, collecting sufficient EEG data for each new individual remains impractical because of cost, time, and ethical constraints.

Data augmentation often serves as a solution to develop generalised models that handle variability in EEG signals. Techniques based on Generative Adversarial Networks (GANs)[75] and Variational Autoencoders (VAEs)[68, 321] offer early examples of synthetic EEG generation. Recent advances using Denoising Diffusion Probabilistic Model (DDPM) [322, 238] further suggest that diffusion-based approaches can generate higher-quality EEG signals. While these methods improve model robustness by creating synthetic signals, they do not generalise well to new subjects because they lack explicit mechanisms for subject adaptation. Unlike images, EEG signals have a low-signal-to-noise ratio where noise is subject-specific and often induced through subject-specific artifacts such as eye blinks, muscle movements, or other cognitive states. Hence, standard DDPM requires learning what is noise and clean signal during the generation process.

Cross-subject EEG classification models, such as EEG-GNN [156], EEG-GAT [169], and SOGNN [8], promote subject generalization through dynamic spatio-temporal learning. However, these models primarily focus on general robustness and do not explicitly adapt to unseen individuals. There remains significant room for improvement, particularly in models like SOGNN. Instead of assuming a static graph structure across all subjects, SOGNN [8] dynamically prunes the graph for each subject and models interactions among EEG channels using Graph Convolution Networks (GCNs) [9]. Although GCNs effectively capture local neighbourhood relationships, they often fail to model global interactions across distant nodes. In contrast, Graph Attention Networks (GATs) [10] can model both local and global dependencies by attending to all nodes in the graph. Replacing the GCN modules in SOGNN with GAT layers can, therefore, potentially enhance model performance,

similar to how EEG-GAT [169] demonstrated improvements over EEG-GNN [156].

Domain adaptation techniques such as Meta Update Strategy (MUPS-EEG) [323] aim to adapt models to new subjects with minimal data. However, these techniques often assume the availability of subject-specific data during finetuning, which might not be feasible in real-world scenarios. Frameworks such as META-EEG [274] and Model Agnostic Meta Learning (MAML)-based methods [324] attempt to improve cross-subject generalisation but typically operate under the assumption that EEG signals used during finetuning are preprocessed and free from artifacts. This assumption overlooks the practical challenges of dealing with noisy real-world EEG recordings, potentially limiting the robustness and applicability of these models.

This thesis proposes **Transfer and Robust Adaptation of New Subjects in EEG Technology (TRANSIT-EEG)**, a unified framework that directly addresses the limitations of prior work by enabling subject-specific adaptation under low-data and noisy conditions. TRANSIT-EEG separates clean signal from subject-specific noise using a generative model conditioned on subject embeddings, improving the quality of synthetic data. It enhances cross-subject modelling by replacing GCNs [9] with GAT [10] to capture both local and global dependencies in EEG signals. It further supports efficient adaptation to new subjects through meta-learning and parameter-efficient finetuning strategies. This design eliminates the need for clean, labelled target-subject data and improves generalisation across individuals in realistic EEG scenarios.

This chapter discusses the proposed TRANSIT-EEG framework and its experimental validation on two open datasets SEED [1] and PhyAAt [2], which systematically addresses the challenge of subject-specific adaptation in EEG classification under limited data and high variability. TRANSIT-EEG includes three core components: a subject-specific generative model that separates clean signal from subject-induced noise, a graph-based classifier that captures both local and global EEG dynamics, and efficient adaptation strategies that require minimal subject-specific data. Together, these components enable robust and generalisable EEG classification across unseen individuals. The main contributions of this chapter are:

1. **Unified Framework for Subject Adaptation:** Proposal of a novel unified

framework for cross-subject adaptation called TRANSIT-EEG, which integrates subject-specific data augmentation, dynamic GNN-based modelling, and efficient finetuning strategies.

(a) **Improved Graph Architecture:** Extension of Self-Organizing Graph Neural Networks (SOGNN) [8] by replacing Graph Convolutional Networks (GCN) [325] with Graph Attention Transformers (GAT) [326], resulting in a more effective model called Self-Organizing Graph Attention Transformer (SOGAT).

(b) **Subject-Specific Data Generation:** Proposal of a subject-specific generative model called Individual-specific Diffusion Probabilistic Model (IDPM). IDPM uses diffusion-based generation constrained by subject embeddings to disentangle subject-specific noise from clean EEG signals and generates synthetic samples for SOGAT finetuning.

(c) **Finetuning Strategies for New Subjects:** Evaluation of full-model, head-only, and adapter-based finetuning strategies (including LoRA) for adapting SOGAT to new subjects under few-shot conditions.

2. **Empirical Validation:** Demonstration of consistent performance gains on two benchmark datasets—SEED [1] and PhyAAt [2]—using the proposed TRANSIT-EEG framework.

3. **Model Interpretability:** Analysis of learnt EEG topographic maps and graph structures, showing meaningful spatial patterns across emotional states.

5.2 Methodology

This section presents the methodology of the proposed TRANSIT-EEG framework for adapting EEG models to new subjects.

5.2.1 Proposed TRANSIT-EEG Framework

The proposed novel TRANSIT-EEG framework addresses the challenge of cross-subject classification in EEG paradigms by combining data augmentation, meta-learning, and

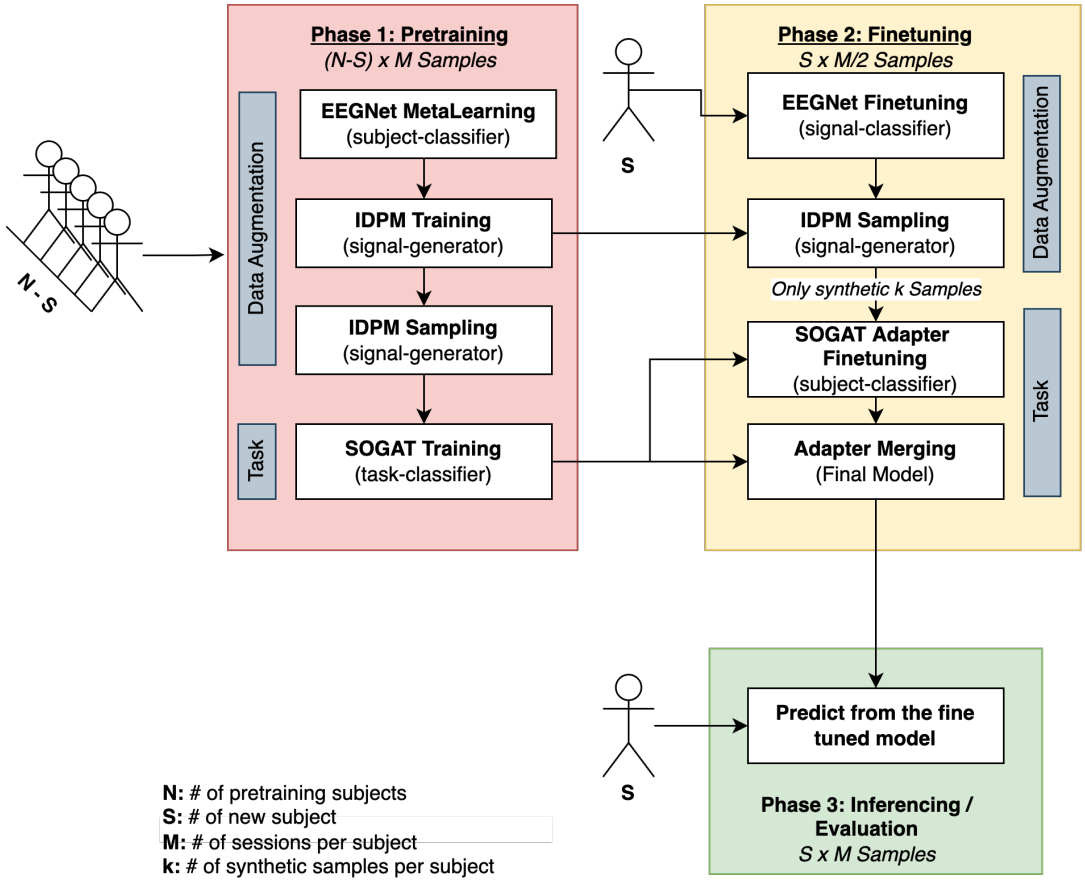


Figure 5.1: Schematic representation of the proposed TRANSIT-EEG framework.

efficient finetuning using LoRA [320]. It enables the model to adapt to a new subject with limited sessions. Figure 5.1 provides a schematic overview of the proposed TRANSIT-EEG framework. The overall system integrates three key components:

1. **Subject Classifier**: The subject-classifier uses EEGNet [6] to learn compact, subject-specific embeddings from EEG signals.

2. **Signal Generator**: The Individualised Denoising Probabilistic Model (IDPM) builds upon DDPM [28] and incorporates an ArcMargin loss [327] to constrain the generation process and preserve subject identity while separating subject-specific noise from the clean, informative signal.

3. **Task Classifier**: The Self Organizing Graph Attention Transformer (SOGAT) extends the Self Organizing Graph Neural Network (SOGNN) [8] by replacing GCN layers with Graph Attention layers [10] to better capture both local and global spatio-temporal interactions.

The proposed TRANSIT-EEG framework trains and deploys its components sequentially across three phases: pretraining, finetuning, and inferencing.

1. Pretraining: TRANSIT-EEG meta-learns the subject-classifier EEGNet [6] using data from $N-1$ training subjects. EEGNet [6] uses ArcMargin loss introduced in a face recognition model called ArcFace [327] to learn discriminative subject embeddings that capture subject-specific EEG characteristics that condition the proposed signal generator model - IDPM. IDPM modifies the standard U-Net architecture used in Denoising Diffusion Probabilistic Models (DDPM) [28] by introducing two separate output branches: one branch reconstructs the clean EEG signal, while the other models subject-specific noise. Through cross-attention mechanisms [49], IDPM integrates subject embeddings into the denoising process at each timestep, constraining the generation toward the characteristics of the target subject. This conditioning enables IDPM to generate realistic, subject-specific synthetic EEG samples without explicit labels. After subject-specific data augmentation, the task-classifier SOGAT trains using original EEG data and synthetic samples generated by IDPM. SOGAT models spatial and temporal dependencies across EEG channels using attention mechanisms, improving robustness to subject variability.

2. Finetuning: TRANSIT-EEG adapts the models to a new subject using a limited number of real EEG samples. The subject-classifier EEGNet [6] finetunes on a small subset of the new subject's data using ArcMargin loss [327], allowing the subject embedding space to adjust rapidly to new EEG patterns. The updated embeddings condition the IDPM model, which remains frozen during finetuning. IDPM uses the updated embeddings to generate subject-specific synthetic EEG samples without modifying internal parameters. The task-classifier SOGAT finetunes using original and synthetic data from the new subject. Low-Rank Adaptation (LoRA) [320] enables efficient task-specific personalisation without full model retraining.

3. Inferencing: TRANSIT-EEG applies the finetuned task classifier SOGAT to unseen EEG data from the new subject. The model predicts target labels such as emotional states or cognitive conditions. The evaluation uses classification accuracy and F1-score metrics across all available sessions from the new subject to comprehensively assess generalisation performance after adaptation.

5.2.2 Subject-Specific Data Augmentation Using IDPM

The proposed IDPM builds upon recent advances in diffusion modelling with Denoising Diffusion Probabilistic Model (DDPM) [28] and metric learning techniques [327] to enable subject-specific EEG signal generation. IDPM modifies the standard U-Net architecture used in DDPM to serve as the backbone network for feature extraction and reconstruction. The U-Net architecture [237], initially developed for biomedical image segmentation, effectively captures detailed localisation and reconstruction patterns, making it well-suited for modelling the complex spatiotemporal structure of EEG signals. The encoder of U-Net progressively downsamples the noisy EEG input through stacked residual blocks consisting of convolution layers. At the same time, the decoder restores the original spatial and temporal dimensions through interpolation and deconvolution operations.

The training of IDPM follows the standard DDPM [28] methodology, which learns to predict and remove progressively added Gaussian noise across a sequence of diffusion steps. However, unlike standard DDPM [28], which primarily focuses on denoising the entire input, IDPM explicitly separates the reconstruction of the clean neural signal and the generation of subject-specific noise through two parallel decoding branches. This separation enables better modelling of subject-specific variations inherent in EEG recordings. IDPM leverages cross-attention mechanisms that condition intermediate representations on subject embeddings learnt using EEGNet [6] to inject subject-specific information into the denoising process. An ArcMargin loss [327] ensures that extracted embeddings remain discriminative amongst different subjects. These components enable IDPM to synthesise high-quality, subject-specific EEG samples by disentangling clean neural activity from subject-specific variations during diffusion-driven reconstruction.

5.2.2.1 Architecture Overview of IDPM

Figure 5.2 shows the proposed IDPM architecture, which extends the standard U-Net used in diffusion models such as DDPM [28]. Unlike the original U-Net, which employs a single encoder–decoder path, IDPM introduces two parallel decoders to model clean neural activity and subject-specific noise components within EEG signals.

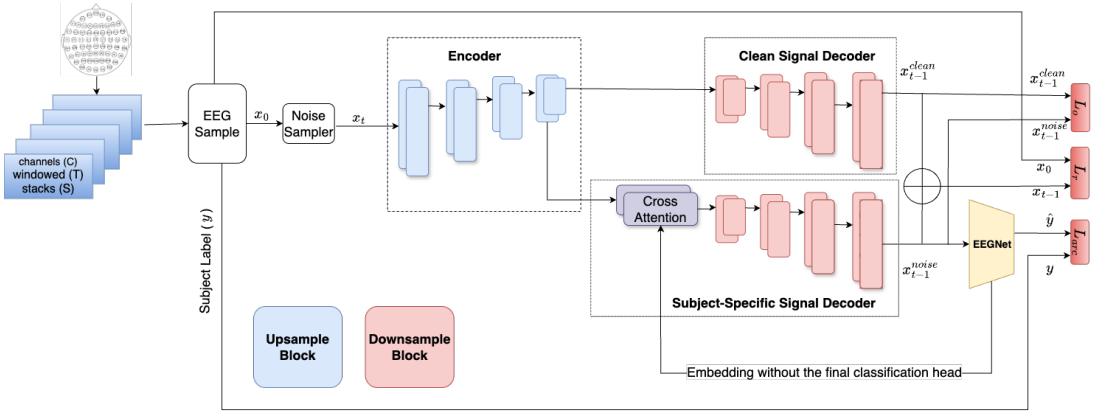


Figure 5.2: Schematic overview of the IDPM model architecture, highlighting the shared encoder in TRANSIT-EEG framework.

The encoder compresses the input into a latent representation, which both decoders process independently. The clean signal decoder reconstructs task-relevant information, while the subject-specific decoder uses cross-attention with subject embeddings to inject individual variability into the reconstruction. The standard U-Net is a symmetric architecture with a downsampling encoder and an upsampling decoder, originally designed for image segmentation tasks [237]. It uses skip connections to pass feature maps from encoder layers to their corresponding decoder layers, helping preserve spatial structure and fine details. This architecture supports the effective reconstruction of structured signals and forms the foundation for IDPM's dual-decoder design. By integrating this structure, IDPM enables personalized yet consistent EEG signal generation, improving data augmentation for subject-adaptive classification. At each training iteration, IDPM processes three inputs:

1. **Noisy EEG Signal** (x_t): IDPM generates x_t by adding Gaussian noise to a clean EEG sample x_0 using a dedicated **Noise Sampler** module that performs the forward diffusion step.

2. **Diffusion Time Step** (t): IDPM embeds t using sinusoidal positional encodings to provide a time-dependent feature vector that guides the denoising process.

3. **Subject Label** (y): IDPM uses y to supervise identity preservation, ensuring that the subject-specific noise component retains discriminative features linked to subject identity.

IDPM passes the noisy input x_t through a shared encoder composed of stacked downsampling blocks, where each block contains a residual block [28] followed by spatial pooling. The encoder progressively reduces spatial and temporal dimensions while extracting hierarchical features. The latent representation from the encoder branches into two separate decoding paths:

1. **Clean Signal Decoder**: This branch processes the encoded features through upsampling operations, concatenation with corresponding skip connections, and residual blocks. It reconstructs the clean EEG signal $\hat{x}_{\text{clean}}$ and models subject-independent brain activity without subject-specific conditioning.

2. **Subject-Specific Noise Decoder**: This branch first applies a cross-attention mechanism, allowing the encoded features to interact with subject embeddings extracted from a meta-learnt EEGNet [6]. After conditioning, the decoder processes the modified features through a separate upsampling path to reconstruct the subject-specific noise component $\hat{x}_{\text{noise}}$. This path captures individual-specific variations in EEG recordings, such as muscle artifacts, electrode placement effects, and other personal characteristics.

IDPM enables efficient adaptation to new subjects by updating only the subject embeddings while separating subject-specific noise from the clean, informative signal. This separation occurs without retraining the diffusion model, allowing the system to preserve invariant subject dynamics and generalize effectively to unseen individuals.

5.2.2.2 Loss Functions in IDPM

The training of IDPM optimises a combination of three primary loss functions — the **Reconstruction Loss** ($\mathscr{L}_{\text{recon}}$), the **Identity Preservation Loss** ($\mathscr{L}_{\text{arc}}$), and the **Orthogonality Loss** ($\mathscr{L}_{\text{orth}}$) — along with one regularisation term, the **Temporal Difference Regularisation** ($\mathscr{R}_{\text{temp}}$). These objectives enable accurate signal reconstruction, subject-specific identity preservation, effective noise separation, and realistic temporal dynamics without uneven changepoints in the generated EEG signals.

To formally define these objectives, let x_0 denote the original clean EEG signal, x_t denote the noisy EEG signal at diffusion step (t), $\hat{x}_{\text{clean}}$ denotes the predicted clean EEG

signal, $\hat{x}_{\text{noise}}$ denotes the predicted subject-specific noise, $\hat{x}_{\text{t}} = \hat{x}_{\text{clean}} + \hat{x}_{\text{noise}}$ denotes the reconstructed signal, y denotes the subject label, and T denotes the number of temporal steps in the EEG trial. Equation 5.1 shows the total training objective ($\mathscr{L}_{\text{total}}$) combining three losses and a regularization term:

$$\mathscr{L}_{\text{total}} = \lambda_1 \mathscr{L}_{\text{recon}} + \lambda_2 \mathscr{L}_{\text{arc}} + \lambda_3 \mathscr{L}_{\text{orth}} + \lambda_4 \mathscr{R}_{\text{temp}} \tag{5.1}$$

where $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are weighting hyperparameters for each loss and regularisation terms and are explained below:

1. **Reconstruction Loss ($\mathscr{L}_{\text{recon}}$)**: This loss ensures that the model accurately reconstructs both the clean EEG signal and the subject-specific noise component. Equation 5.2 shows the reconstruction loss:

$$\mathscr{L}_{\text{recon}} = \|\hat{x}_{\text{clean}} - x_0\|^2 + \|\hat{x}_{\text{noise}} - (x_t - x_0)\|^2 \tag{5.2}$$

   The first term ($\|\hat{x}_{\text{clean}} - x_0\|^2$) penalises differences between the predicted clean signal ($\hat{x}_{\text{clean}}$) and the ground-truth clean EEG (x_0). The second term ($\|\hat{x}_{\text{noise}} - (x_t - x_0)\|^2$) penalises differences between the predicted noise ($\hat{x}_{\text{noise}}$) and the true residual noise ($x_t - x_0$).

2. **Identity Preservation Loss ($\mathscr{L}_{\text{arc}}$)**: This loss preserves subject identity in the noise branch by applying an ArcMargin loss [327] on the extracted subject embeddings. Equation 5.3 shows the identity preservation loss:

$$\mathscr{L}_{\text{arc}} = \text{ArcMarginLoss}\left(\text{EEGNet}\left(\hat{x}_{\text{noise}}\right), y\right) \tag{5.3}$$

   Here, $\hat{x}_{\text{noise}}$ passes through a meta-learnt EEGNet [6] to produce a subject embedding. ArcMargin loss encourages embeddings of the same subject to cluster together while maintaining separation between different subjects.

3. **Orthogonality Loss ($\mathscr{L}_{\text{orth}}$)**: This loss enforces that the clean and noise representations remain decorrelated, promoting effective signal separation.

Equation 5.4 shows the orthogonality loss:

$$\mathscr{L}_{\text{orth}} = \left\| f_{\text{clean}}^{\top} f_{\text{noise}} \right\|_F^2 \tag{5.4}$$

Here, f_{clean} and f_{noise} denote intermediate feature maps of the clean and noise decoders, respectivelMinMinimising their Frobenius inner product ensures the two components capture independent information.

4. **Temporal Difference Regularisation ($\mathscr{R}_{\text{temp}}$)**: This regularisation encourages temporal smoothness in the reconstructed EEG signals. Equation 5.5 shows the temporal regularisation:

$$\mathscr{R}_{\text{temp}} = \sum_{i=1}^{T-1} \left\| (\hat{x}_t^{(i+1)} - \hat{x}_t^{(i)}) - (x_0^{(i+1)} - x_0^{(i)}) \right\|_2^2 \tag{5.5}$$

The first term models frame-to-frame differences in the reconstructed signal $\hat{x}_t$, and the second term models actual frame differences in the clean signal x_0, minimising this difference preserves the temporal dynamics of real brain activity in the generated EEG.

5.2.2.3 IDPM workflow

Table 5.1 summarises the overall workflow of the proposed IDPM across the training, finetuning, and inference phases. During training, the model jointly optimises a modified U-Net architecture as explained in Section 5.2.2.1, cross-attention modules, and a meta-learnt EEGNet [6] to learn disentangled representations of clean EEG signals and subject-specific noise. During finetuning, the model adapts only the EEGNet [6] component to a small set of new subject data, enabling flexible subject conditioning without retraining the entire model. IDPM generates subject-specific EEG samples during inference by progressively denoising random noise through reverse diffusion steps conditioned on the adapted subject embedding. This modular training and inference design enables efficient adaptation and scalable generalisation across unseen subjects. While IDPM generates subject-specific data, SOGAT performs robust task classification using real and synthetic EEG signals.

Table 5.1: Workflow overview of IDPM across Phase 1 (Pretraining), Phase 2 (Finetuning), and Phase 3 (Inference) in TRANSIT-EEG framework.

Phase	**Step**	**IDPM Module**	**Description**
Phase 1: Pretraining	1. Pretrain EEGNet	EEGNet (Meta-learnt Subject Classifier)	Meta-learn EEGNet [6] using ArcMargin loss [327] across multiple subjects to extract identity-discriminative embeddings.
	2. Generate Noisy EEG	Forward Diffusion Process	Add Gaussian noise to clean EEG signals at randomly selected diffusion steps using a Noise Sampler.
	3. Encode Noisy EEG	U-Net Encoder	Pass noisy signal and time embedding through a modified U-Net encoder with Conv2D residual blocks.
	4. Dual-Path Decoding	Clean Decoder, Noise Decoder, Cross-Attention, EEGNet, Loss Heads	• Clean Branch: Apply vanilla deconvolution to predict clean EEG. • Noise Branch: Apply cross-attention with subject embeddings and deconvolution to predict subject-specific noise. Supervise with MSE, ArcMargin [327], Orthogonality, and Temporal Difference losses.
Phase 2: Finetuning	1. Few-Shot EEGNet Finetuning	EEGNet (Finetuned Subject Classifier)	Finetune EEGNet [6] on a small subset of new subject's data using ArcMargin loss; freeze U-Net and cross-attention modules.
	2. Extract New Subject Embedding	EEGNet	Use the finetuned EEGNet to compute the subject embedding for conditioning IDPM.
	3. Prepare for Inference	All Modules Frozen	Freeze all modules after subject embedding extraction; no further parameter updates.
Phase 3: Inference	1. Noise Initialisation	Random Initialisation	Sample random Gaussian noise tensor x_T for starting point of the reverse diffusion.
	2. Reverse Diffusion	U-Net Encoder, Cross-Attention, Clean/Noise Decoders	Iteratively denoise from x_T to x_0 conditioned on time and the extracted subject embedding; sum outputs from clean and noise branches at each step.
	3. Output Generation	Output Synthesiser	Obtain the final subject-specific EEG sample at step $t = 0$.

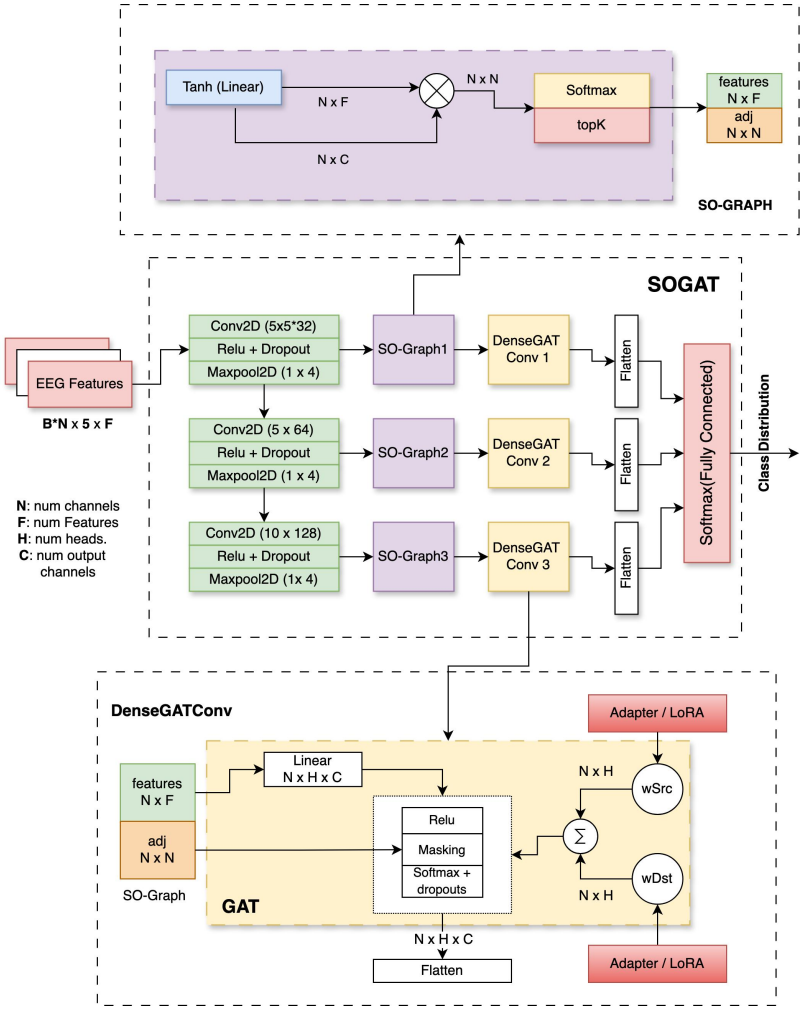


Figure 5.3: Architecture of the proposed SOGAT model for EEG signal classification in TRANSIT-EEG framework.

5.2.3 Task Classification Using SOGAT

The proposed TRANSIT-EEG framework employs a dedicated task classification model — Self Organizing Graph Attention Transformer (SOGAT) — to classify EEG signals based on spatial and temporal patterns. SOGAT builds upon principles from Graph Neural Networkss (GNNs) and models the dynamic relationships between EEG electrodes using self-organised graphs and attention-based aggregation. Section 2.3.3.3 in Chapter 2 provides an overview of GNNs, including a detailed discussion of Self Organizing Graph Neural Network (SOGNN) [8]. SOGAT extends this foundation by

replacing GCN layers with GAT layers [10] in SOGNN, enabling dynamic capture of both local and global dependencies based on input signals.

The main components of SOGAT are as follows:

1. **Feature Engineering Using Differential Entropy (DE):**

$$h(X) = -\int_{-\infty}^{\infty} f(x) \log f(x)\, dx \tag{5.6}$$

SOGAT extracts Differential Entropy (DE) features from EEG signals across five standard frequency bands. DE measures the uncertainty of a continuous random variable and captures the complexity of EEG signal distributions. Equation 5.6 shows the formulation of DE where $f(x)$ denotes the probability density function of X. By using DE features, SOGAT captures nuanced spatial and temporal patterns critical for EEG-based classification [8].

2. **Dynamic Graph Construction:**

$$a_{ij} = \frac{\exp(\theta(v_i W)\theta(v_j W)^\top)}{\sum_{i=1}^{N} \exp(\theta(v_i W)\theta(v_j W)^\top)} \tag{5.7}$$

SOGAT constructs dynamic graphs where nodes represent EEG electrodes and edges represent functional relationships. Traditional static graphs based on spatial arrangement of electrodes as per EEG devices fail to capture subject-specific spatial dynamics [8, 328, 329, 330, 331]. Inspired by SOGNN [8], SOGAT learns adjacency matrices where each cell a_{ij} is calculated using Equation 5.7 from DE features using a self-organised graph module where, v_i represents the feature vector of node i, W is a learnable weight matrix, and θ applies a tanh activation.

$$A = \text{Softmax}(GG^\top) \tag{5.8}$$

$$A'_{ij} = \begin{cases} A_{ij}, & \text{if } j \in \text{Top}_k(A[i,:]) \\ \\ 0, & \text{otherwise} \end{cases} \tag{5.9}$$

Equation 5.8 shows the normalization that generates the dense adjacency matrix A, where $G = \tanh(\mathcal{Y}W)$ denotes the transformed node embeddings. Equation 5.9 enforces sparsity by retaining only the top-k strongest connections per node, resulting in a sparse adjacency matrix A'. In this formulation, a_{ij} denotes the raw affinity score, A represents the normalized dense graph, and A' defines the final sparse graph used for graph convolution operations.

3. **Signal Classification Using GAT:** SOGAT applies Graph Attention Networks (GATs) [10] to dynamically constructed graphs. The GAT layers learn attention coefficients that weight neighbouring nodes based on their importance, enabling the model to focus on the most informative electrode interactions. As shown in Figure 5.3, DE features first pass through convolutional layers to extract local spatial patterns. SOGAT constructs three self-organised graphs at different abstraction levels (SO-Graph1, SO-Graph2, SO-Graph3), each retaining the top-$k = 10$ edges. GAT layers independently process these graphs, and the outputs are concatenated and passed through a fully connected layer for classification. SOGAT optionally incorporates an Adapter or Low-Rank Adaptation (LoRA) [320] module to enable efficient parameter adaptation on new subjects after the GAT layers. SOGAT contains approximately 386,991 trainable parameters, balancing expressiveness with training efficiency.

5.2.4 New Subject Adaptation

The proposed TRANSIT-EEG framework enables efficient adaptation to new subjects using selective finetuning strategies that avoid catastrophic forgetting. Conventional full-model finetuning often suffers from catastrophic forgetting, where the model overwrites important features learnt from earlier subjects [332]. To mitigate this issue, TRANSIT-EEG selectively finetunes only necessary components while freezing others.

Before describing the full adaptation pipeline, it is important to review the different finetuning strategies explored for adapting models to new subjects:

1. **Full Finetuning:** This approach updates all network parameters using new data. Although full finetuning enables complete adaptation, it risks overfitting and catastrophic forgetting.

2. **Classification Head Finetuning:** This approach updates only the final classification layers, preserving deeper pre-trained features. It reduces computational load and mitigates catastrophic forgetting. However, it limits the model's ability to capture complex subject-specific features.

3. **Adapter Finetuning:** This strategy inserts small learnable modules, called adapters, within the pre-trained network. Only these adapters are updated during finetuning, leaving the rest of the model frozen. Adapter finetuning balances adaptability and efficiency, making it suitable for preserving shared knowledge while capturing subject-specific variations. In the proposed SOGAT model, adapter finetuning supports subject adaptation without full model retraining.

5.2.4.1 Low-Rank Decomposition

Adapter finetuning achieves parameter efficiency by leveraging low-rank decomposition. One widely used method, **LoRA** [320], inserts low-rank matrices to approximate weight updates. Consider a linear layer with weight matrix $W \in \mathbb{R}^{d\times k}$. Instead of learning a full update $\Delta W \in \mathbb{R}^{d\times k}$, LoRA learns two smaller matrices $A \in \mathbb{R}^{d\times r}$ and $B \in \mathbb{R}^{r\times k}$, where $r \ll \min(d,k)$. Equation 5.10 shows this decomposition:

$$\Delta W = AB \tag{5.10}$$

LoRA then updates the original weight matrix using the low-rank approximation as shown in Equation 5.11:

$$W' = W + \Delta W = W + AB \tag{5.11}$$

The variable r represents the **rank** of the adapter. The rank of a matrix refers to the number of linearly independent rows or columns. In low-rank adaptation, a

weight matrix of size $d \times k$ is decomposed into the product of two smaller matrices of dimensions $d \times r$ and $r \times k$. This factorization reduces the number of trainable parameters from $d \times k$ to $r \times (d + k)$, which is significantly smaller when $r \ll d$ and $r \ll k$

Comparison with Full-Rank Adapters Full-rank adapters directly learn an update matrix $\Delta W \in \mathbb{R}^{d \times k}$ without any decomposition. While this method offers higher expressive capacity, it increases the number of trainable parameters and risks overfitting, especially in low-data regimes. Equation 5.10 and Equation 5.11 illustrate how LoRA achieves efficient finetuning by constraining updates to a low-dimensional subspace. This design is particularly suitable for EEG applications, where the limited subject-specific data makes full-rank adaptation less effective.

5.2.4.2 Adaptation of IDPM

The adaptation strategy for IDPM draws inspiration from face recognition research, particularly ArcFace [327], where meta-learnt embedding networks enable rapid subject-specific adaptation using few-shot data. By applying a similar principle to EEG signals, the framework trains the subject-classifier EEGNet [6] to extract robust subject embeddings that generalise across individuals while allowing efficient finetuning.

TRANSIT-EEG adapts IDPM by finetuning only the subject-classifier - EEGNet [6] while freezing the U-Net encoder, clean decoder, noise decoder, and cross-attention modules. Specifically, the adaptation performs **full finetuning** of all parameters within the EEGNet [6] on the new subject's few-shot EEG data. This complete parameter update enables flexible early and deep feature representation adjustment. During the pretraining phase, EEGNet [6] undergoes meta-training using a Model Agnostic Meta Learning (MAML)-based optimisation strategy [324]. This meta-learning process exposes EEGNet [6] to multiple simulated few-shot tasks, optimising its parameters to become sensitive to new subjects while requiring minimal updates. Meta-learning prepares the model to quickly minimise loss when presented with new data rather than merely fitting to the pretraining distribution. As a result, after meta-training, EEGNet [6] reaches a favourable state where full finetuning during adaptation behaves like an efficient second training phase, requiring fewer iterations and data points for

convergence. After adaptation, the updated subject embeddings condition IDPM's generation process, enabling the model to synthesise high-quality, subject-specific EEG samples without retraining the full diffusion model.

5.2.4.3 Adaptation of SOGAT

The proposed TRANSIT-EEG framework adapts the task-classifier SOGAT using real and synthetic EEG samples generated by IDPM. Figure 5.3 shows the DenseGATConv layers that model spatial interactions across EEG electrodes within the SOGAT architecture. For faster adaptation, TRANSIT-EEG applies Low-Rank Adaptation (LoRA) [320] directly to the DenseGATConv layers, allowing fast subject-specific updates without retraining the entire network.

The framework finetunes DenseGATConv in four steps:

1. **Low-Rank Decomposition:** It decomposes the original attention matrices (*att_src* and *att_dst*) in DenseGATConv into two smaller matrices A and B, approximating the full matrix $W \approx AB$. This decomposition reduces model complexity and minimises the number of trainable parameters.

2. **Integration of Low-Rank Matrices:** It then modifies the forward pass to compute attention features using low-rank projections. Equation 5.12 shows the updated attention computation:

$$att\_src = A_{\text{src}} \times B_{\text{src}}, \quad att\_dst = A_{\text{dst}} \times B_{\text{dst}} \tag{5.12}$$

3. **Training with Low-Rank Parameters:** During adaptation, the framework updates only the low-rank matrices A and B while keeping the backbone DenseGATConv weights frozen. This selective optimisation captures subject-specific patterns efficiently with minimal computation.

4. **Subject-Specific SOGAT Adaptation:** The adapted SOGAT model learns to specialise in the new subject's EEG distribution using synthetic EEG samples generated by IDPM and few-shot real EEG data. This adaptation improves task classification performance across unseen subjects.

Table 5.2: Dataset statistics for various experiments in TRANSIT-EEG framework.

SEED [1]

Experimental Setup	**# subjects × # sessions**	**# features**	**# classes**	**Augmentation factor**	**# sessions**
Original Dataset	15 × 45	5 × 265	3	1x	675
Single Fold of LOSO training w/o IDPM	14 × 45	5 × 265	3	1x	630
Single Fold of LOSO training with IDPM	14 × 45	5 × 265	3	1.5x	945
Single Fold of LOSO finetuning with IDPM	1 × 21	5 × 265	3	5x	105

PhyAAt [2]

Experimental Setup	**# subjects × # sessions**	**# features**	**# classes**	**Augmentation factor**	**# sessions**
Original Dataset	25 × 432	5 × 384	3	1x	10,800
Single Fold of LOSO training w/o IDPM	24 × 432	5 × 384	3	1x	10,368
Single Fold of LOSO training with IDPM	24 × 432	5 × 384	3	1.5x	15,552
Single Fold of LOSO finetuning with IDPM	1 × 216	5 × 384	3	5x	1,080

5.3 Experiments and Results

This section presents the experimental setup and results for the various components of the proposed TRANSIT-EEG framework. The evaluation compares the proposed model against state-of-the-art graph-based architectures and analyses the effectiveness of different finetuning strategies and augmentation ratios. The framework implementation is publicly available on GitHub.[1]. All experiments run on a platform equipped with an Nvidia T4 GPU, Ubuntu 22.04, PyTorch 2.1, and PyTorch-Geometric 1.5.0.

5.3.1 Experimental Setup

The evaluation uses two widely studied EEG datasets, SEED [1] and PhyAAt [2]. SEED [1] supports emotion recognition tasks, while PhyAAt [2] targets cognitive task classification. Both datasets provide balanced class labels, enabling fair cross-subject evaluation. SEED [1] contains EEG recordings from 15 participants collected

[1] https://github.com/cahuja1992/transit-eeg

with a 62-channel ESI NeuroScan system. Each session labels three emotional states: happiness (0), sadness (1), and neutrality (2). PhyAAt [2] collects EEG, GSR, and PPG signals from 25 subjects using a 14-channel Emotiv Epoc device, with tasks labelled as listening (0), writing (1), and resting (2). The evaluation follows a LOSO strategy. The evaluation follows a LOSO strategy. The framework first pretrains the model on $N-1$ subjects while leaving out the i-th subject. After pretraining, it finetunes using half of the available sessions from the held-out subject, augmented with synthetic samples generated by IDPM. Finally, it evaluates performance on all sessions from the held-out subject using the updated SOGAT classifier.

Table 5.2 summarises the dataset statistics for different experimental setups on the SEED [1] and PhyAAt [2] datasets. It outlines four configurations: original data, LOSO training without augmentation, LOSO training with IDPM, and LOSO finetuning with IDPM. The table shows the number of subjects and sessions for each case, feature dimensions, classes, augmentation factors, and total sessions. In the original datasets, SEED [1] includes 15 subjects with 45 sessions each, while PhyAAt [2] contains 25 subjects and 432 sessions each. In LOSO settings, one subject is removed from training. IDPM-based augmentation increases the session counts by 1.5x during training and 5x during finetuning. During finetuning, only a subset of sessions — 21 for SEED [1] and 216 for PhyAAt [2] — is used per subject. This overview helps clarify the scale and structure of the data used across different phases of the experiments.

5.3.2 Experiment 1: Choice of the network architecture

This experiment evaluates the cross-subject generalisation capabilities of the proposed SOGAT model against existing graph-based models: EEG-GNN [156], EEG-GAT [169], and SOGNN [8]. The evaluation follows a robust cross-validation strategy, LOSO, which measures performance without data leakage and provides a reliable estimate of cross-subject transfer effectiveness. LOSO validation trains the model on data from $N-1$ subjects while holding out one subject for validation. For SEED [1], each of the 15 subjects is left out once while training on the remaining 14 subjects. Similarly, for PhyAAt [2], LOSO validates each of the 25 subjects iteratively while training only 24 subjects at a time. This setup ensures fair evaluation across all subjects

Table 5.3: LOSO performance of different graph-based models across SEED [1] and PhyAAt [2] for choosing the best model architecture in TRANSIT-EEG framework.

SEED [1]

Model	**Macro-F1 (%)**	**Happy F1 (%)**	**Sad F1 (%)**	**Neutral F1 (%)**
EEG-GNN [156]	77.35	79.38	75.83	76.83
EEG-GAT [169]	78.16	81.38	75.75	77.34
SOGNN [8]	86.69	85.56	85.77	88.74
SOGAT	**87.75**	**86.82**	**86.52**	**89.92**

PhyAAt [2]

Model	**Macro-F1 (%)**	**Listening F1 (%)**	**Writing F1 (%)**	**Resting F1 (%)**
EEG-GNN [156]	69.35	72.38	66.83	68.83
EEG-GAT [169]	73.16	78.38	68.75	72.34
SOGNN [8]	80.05	80.81	77.44	81.92
SOGAT	**81.64**	**82.72**	**79.78**	**82.43**

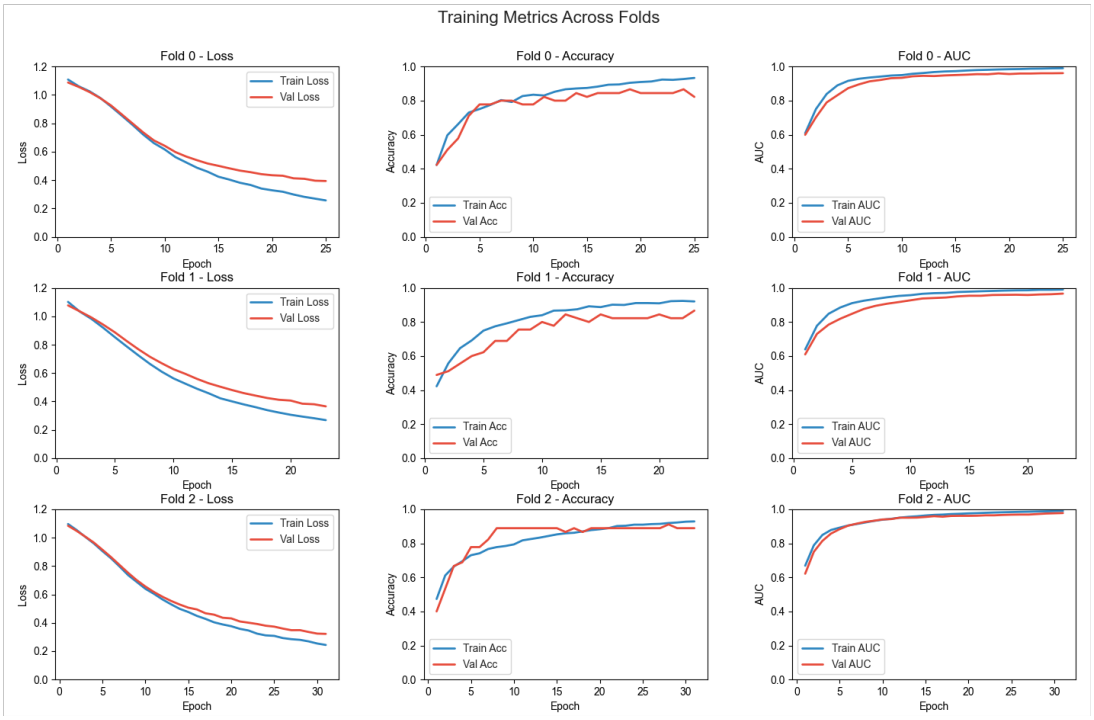


Figure 5.4: Training curves of SOGAT for 3 folds of LOSO as part of TRANSIT-EEG learning process.

without overlap.

This study selects the Macro-averaged F1 score (or Macro-F1 score) as the evaluation metric because both datasets exhibit an approximately balanced class distribution. During each LOSO split, the framework divides the data according to the statistics reported in Table 5.2. Specifically, the SEED [1] uses 630 sessions (each with 1,325 features) for training and 45 sessions for validation. PhyAAt [2] uses 10,368 sessions (each with 1,920 features) for training and 144 sessions for validation. The framework ensures that training and validation sessions maintain the same feature dimensionality. The model trains for 200 epochs using the Adam optimiser, with a learning rate of 1×10^{-3}. Training terminates early if the Area Under Curve (AUC) metric reaches 0.99 on the training set, ensuring efficient convergence without overfitting.

Table 5.3 reports the LOSO Macro-F1 scores achieved by different models across SEED [1] and PhyAAt [2]. Analysis of these results shows that SOGAT consistently outperforms previous models. On the SEED [1], SOGAT achieves a Macro-F1 score of 87.75%, improving by 1.06 percentage points compared to SOGNN's 86.69%. Similarly, on the PhyAAt [2], SOGAT attains a Macro-F1 score of 81.64%, surpassing SOGNN's 80.05% by 1.59 percentage points. Although the improvements appear modest, they highlight the enhanced robustness and generalisation capabilities of SOGAT across subjects. Figure 5.4 shows the training and validation loss curves for SOGAT across three LOSO folds. The training curves demonstrate that the training loss decreases as the number of epochs increases, whereas the validation loss saturates early. The training stops according to the AUC threshold criterion, ensuring convergence to an optimal model without significant overfitting.

5.3.3 Experiment 2: Effect of Subject-Specific Data Augmentation Using IDPM

This experiment builds upon Experiment 1 by introducing subject-specific data augmentation using the proposed Individualised Diffusion Probabilistic Model (IDPM). While Experiment 1 establishes a baseline without any augmentation, this experiment evaluates the impact of IDPM-augmented training on model performance. The

framework implements Phase 1 of the methodology, where IDPM trains on data from all subjects and their trials and generates new synthetic samples to augment the training datasets. The evaluation follows the same LOSO validation strategy as in Experiment 1, enabling a direct comparison of model performance with and without IDPM augmentation. During each fold of LOSO cross-validation, the framework augments the training data by 50%. For SEED [1], this augmentation results in 924 training sessions and 66 validation sessions. Similarly, for PhyAAt [2], the training dataset expands to 5,182 sessions with 216 validation sessions.

To generate synthetic samples, IDPM performs linear noise scheduling over 50 iterations. This sampling strategy increases the adequate dataset size by a factor of five, thereby improving the diversity and subject coverage of the training data. The model without IDPM augmentation acts as the baseline for comparison. Table 5.4 reports the LOSO Macro-F1 scores for both datasets, where bold values represent the best performance achieved in each category.

The results show that IDPM-based data augmentation consistently enhances model performance across both datasets. For SEED [1], SOGAT combined with IDPM achieves a Macro-F1 score of 88.62%, improving by 0.87 percentage points over the baseline model's 87.75%. Similarly, for PhyAAt [2], the model attains a Macro-F1 score of 82.50%, surpassing the baseline by 0.86 percentage points. In addition to overall improvements, each class F1 score improves when IDPM augmentation is applied. These findings suggest that the proposed IDPM-based augmentation improves the model's ability to generalise across subjects, as evidenced by the consistent gains in LOSO cross-validation metrics. The augmentation technique proves particularly effective for mitigating class imbalance and modelling inter-subject variability, thereby leading to more robust EEG signal classification.

5.3.4 Experiment 3: Effectiveness of Adapter Finetuning

5.3.4.1 Overall Effect of Adapter Finetuning with IDPM-Based Data Augmentation

This experiment builds upon Experiment 2 by evaluating the impact of adapter-based finetuning after IDPM-aided pretraining. The framework increases the finetuning data

Table 5.4: Proposed SOGAT performance with IDPM augmentation in TRANSIT-EEG framework.

SEED [1]

Technique	**Macro-F1 (%)**	**Happy F1 (%)**	**Sad F1 (%)**	**Neutral F1 (%)**
SOGAT without IDPM	87.75	86.82	86.52	89.92
SOGAT with IDPM (Phase 1)	**88.62**	**87.36**	**88.25**	**90.27**

PhyAAt [2]

Technique	**Macro-F1 (%)**	**Listening F1 (%)**	**Writing F1 (%)**	**Resting F1 (%)**
SOGAT without IDPM	81.64	82.72	79.78	82.43
SOGAT with IDPM (Phase 1)	**82.50**	**82.72**	**81.62**	**83.18**

by five times by generating synthetic sessions for new subjects using the IDPM sampler.

For each LOSO fold, the framework uses 50% of the new subject's sessions for finetuning and reserves the remaining 50% for evaluation. In the SEED [1], this approach retains 21 sessions for finetuning and 24 sessions for evaluation, evenly distributed across classes. The PhyAAt [2] allocates 216 sessions per class for both finetuning and evaluation.

During finetuning, the framework augments the few-shot sessions by a factor of five, resulting in 105 sessions for SEED [1] and 1,080 sessions for PhyAAt [2]. The model finetunes over 5 epochs using a reduced learning rate of 1×10^{-6}. The framework explores two finetuning strategies:

- **TRANSIT-EEG.A**: This variant uses a full-rank adapter for task-specific adaptation.
- **TRANSIT-EEG.B**: This variant uses a low-rank adapter based on LoRA for efficient parameter updates.

Table 5.5 presents both datasets' LOSO Macro-F1 scores and finetuning times. Bold values indicate the best-performing configurations.

Table 5.5: TRANSIT-EEG performance across different phases on SEED [1] and PhyAAt [2].

SEED [1]

Technique	**Happy F1 (%)**	**Sad F1 (%)**	**Neutral F1 (%)**	**Macro-F1 (%)**	**Fine-tuning time (sec)**
SOGAT with IDPM (Phase 1)	87.36	88.25	90.27	88.62	-
SOGAT + Phase 1 + Phase 2 (Only Full Rank Adapter)	91.00	90.20	91.56	90.92	2273
SOGAT + Phase 1 + Phase 2 (Only LoRA)	89.80	90.02	89.82	89.88	1186
TRANSIT-EEG.A = SOGAT + Phase 1 + Phase 2 (IDPM + Full Rank Adapter)	**90.89**	**90.81**	**92.89**	**91.53**	3929
TRANSIT-EEG.B = SOGAT + Phase 1 + Phase 2 (IDPM + LoRA)	90.02	89.64	91.82	90.49	**2082**

PhyAAt [2]

Technique	**Listening F1 (%)**	**Writing F1 (%)**	**Resting F1 (%)**	**Macro-F1 (%)**	**Fine-tuning time (sec)**
SOGAT with IDPM (Phase 1)	82.72	81.62	83.18	82.50	-
SOGAT + Phase 1 + Phase 2 (Full Rank Adapter)	83.51	85.50	85.42	84.81	1927
SOGAT + Phase 1 + Phase 2 (LoRA)	82.50	84.00	85.00	83.81	983
TRANSIT-EEG.A = SOGAT + Phase 1 + Phase 2 (IDPM + Full Rank Adapter)	**84.70**	**88.82**	**89.92**	**87.78**	2089
TRANSIT-EEG.B = SOGAT + Phase 1 + Phase 2 (IDPM + LoRA)	85.72	87.98	87.43	87.04	**1184**

For SEED [1], "TRANSIT-EEG.A" improves the LOSO Macro-F1 score from 90.92% (finetuning without IDPM) to 91.53%, representing a gain of 0.61 percentage points. Similarly, "TRANSIT-EEG.B" improves from 89.88% to 90.49%. Regarding computational efficiency, "TRANSIT-EEG.B" achieves 90.49% F1 in 2,082 seconds, nearly halving the finetuning time compared to "TRANSIT-EEG.A" 3,929 seconds. Both methods significantly outperform Phase 1 training alone, where SOGAT with only IDPM augmentation achieves a maximum average F1 score of 88.62%.

For PhyAAt [2], "TRANSIT-EEG.A" demonstrates a substantial improvement, increasing the Macro-F1 from 84.81% (finetuning without IDPM) to 87.78%, a gain of 2.97 percentage points. "TRANSIT-EEG.B" similarly improves from 83.81% to 87.04%. Compared to Phase 1 results (82.50% Macro-F1), both TRANSIT-EEG variants show marked improvements. "TRANSIT-EEG.B" accomplishes this improvement in only 1,186 seconds, significantly faster than "TRANSIT-EEG.A" 2,089 seconds for PhyAAt [2]. These results confirm that adapter-based finetuning, especially when combined with IDPM-based data augmentation, enhances adaptation accuracy and computational efficiency in the proposed TRANSIT-EEG framework.

5.3.4.2 Comparison of Various Finetuning Methodologies Without IDPM-Based Data Augmentation

This subsection presents a comparative analysis of different finetuning methodologies for EEG signal classification, focusing on their performance using LOSO evaluation [333, 70, 334]. The evaluation uses the PhyAAt [2], considering only 50% of the available samples per subject without applying any data augmentation. The framework tests multiple finetuning strategies, including full finetuning, classification head finetuning, and adapter-based finetuning with different ranks (full rank, 32, 16, 8). The results show that lower-rank adapters correspond to decreased LOSO Macro-F1 scores. However, LoRA, with a rank of 32, provides an optimal trade-off between latency and accuracy, making it suitable for applications with constraints such as on-device finetuning. Table 5.6 summarises the LOSO results, showing that higher adapter ranks correlate with better performance but at the cost of increased computational time. The full-rank adapter achieves the highest Macro-F1 score of 84.81%, while LoRA with rank 32 achieves a Macro-F1 of 83.81% with the fastest finetuning time of 983 seconds. In comparison, full finetuning achieves 81.98% Macro-F1, and classification head finetuning achieves 83.26% Macro-F1.

To statistically validate the performance differences, this study conducts paired t-tests at a significance level of $\alpha = 0.05$. The test between LoRA (rank 32) and the classification head finetuning yields a t-statistic of 7.0356 and a p-value of 0.0196. Similarly, comparing the full-rank adapter and full finetuning produces a t-statistic of

Table 5.6: Finetuning performance of SOGAT on PhyAAt [2] for TRANSIT-EEG's subject adaptation.

Finetuning Technique	**Rank**	**Macro-F1 (%)**	**Listening F1 (%)**	**Writing F1 (%)**	**Resting F1 (%)**	**Fine-tuning Time (sec)**
Full finetuning	-	81.98	80.90	80.90	84.14	2572
Classification head	-	83.26	81.40	83.92	84.46	1583
Full rank adapter	full	**84.81**	**83.51**	**85.50**	**85.42**	1927
LoRA	32	83.81	82.50	84.00	85.00	**983**
LoRA	16	**84.88**	83.90	83.90	86.85	690
LoRA	8	84.45	83.50	83.50	86.35	430

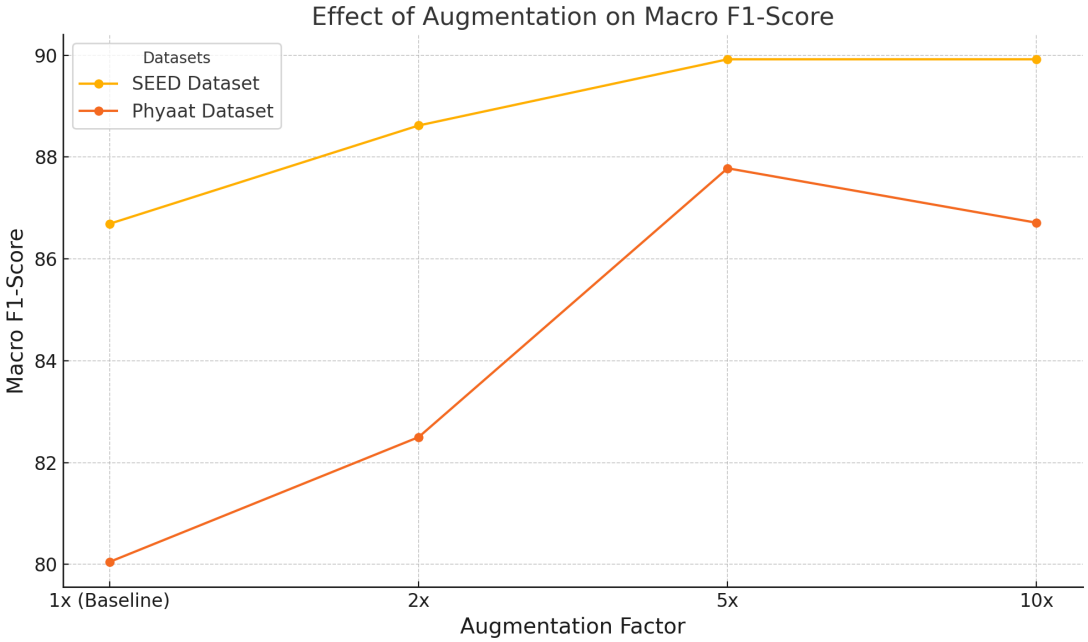


Figure 5.5: Impact of TRANSIT-EEG's proposed IDPM data augmentation on macro F1-score for SEED [1] and PhyAAt [2].

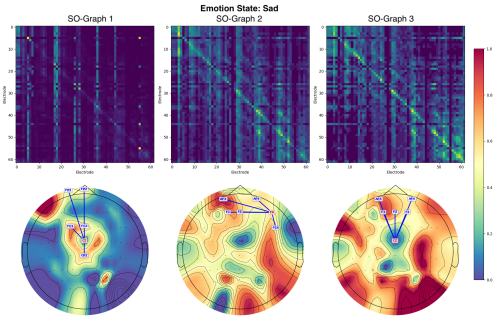
6.0584 with a p-value of 0.0262. These results statistically substantiate the effectiveness of adapter-based finetuning methods over traditional finetuning approaches for adapting to new subjects.

5.3.4.3 Impact of IDPM-Based Data Augmentation During Finetuning

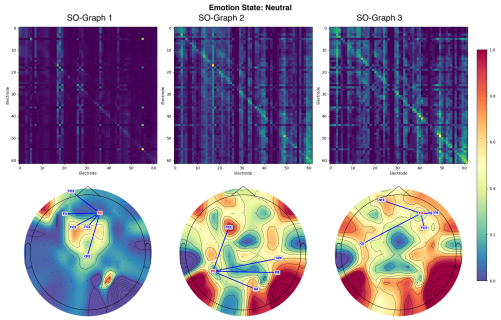
This experiment examines the effect of session-based data augmentation on EEG signal classification performance using the SEED [1]. This experiment selects SEED [1] due to its balanced class distribution, which helps mitigate potential class bias in the results. The experimental procedure randomly selects 21 sessions per subject for finetuning. It then applies session augmentation with factors of $1\times$ (no augmentation), $2\times$, $5\times$, and $10\times$. For each augmentation factor, the model finetunes on the augmented data.

This experiment adopts a Leave One Subject's Session Out (LOSSO) cross-validation strategy, analogous to the LOSO used for cross-subject generalisation. LOSSO holds out one session as the test set for each subject while using the remaining sessions and synthetically generated signals for finetuning. This process repeats for all subjects, and the experiment averages the results across all folds. This experiment serves two primary purposes. First, it assesses the model's ability to generalise across different sessions within subjects. Second, it tests the model's robustness against variations introduced by session-based training data. By systematically increasing the available finetuning data through augmentation, it evaluates how data quantity impacts model generalisation to unseen samples.

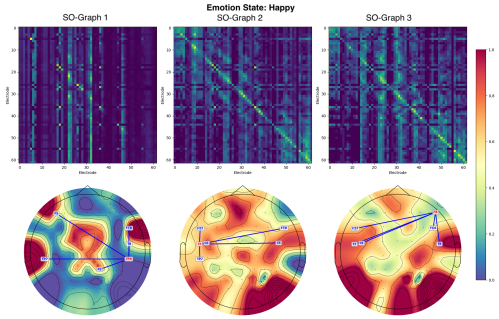
Figure 5.5 summarises the classification performance across different augmentation scales. The experiment computes the F1-score and classification accuracy for each subject and reports the average across the dataset. The results show that increasing the number of training sessions through augmentation enhances the model's generalisation performance. For SEED [1], accuracy improves from 86.81% at baseline ($1\times$) to 91.53% at $5\times$ augmentation, with similar gains observed in Macro-F1 score. However, further augmentation to $10\times$ does not yield additional improvements, indicating a saturation point where more samples no longer enhance performance. Similarly, for PhyAAt [2], this experiment observes an accuracy increase from 84.74% at baseline to 87.78% at $5\times$ augmentation, followed by a slight decrease at $10\times$ augmentation. This decline likely results from overfitting or noise introduced by excessive synthetic data. These findings suggest that IDPM-based session augmentation significantly improves the model's ability to generalise during subject adaptation but highlights the importance of tuning augmentation factors to avoid overfitting.



(a) Sad emotional state



(b) Neutral emotional state



(c) Happy emotional state

Figure 5.6: Channel adjacency matrices and topographic maps of each graph learnt using proposed SOGAT as part of TRANSIT-EEG framework for SEED [1] subject 0 with different emotional states [1].

5.3.5 Understanding the Relationship Between EEG Channels from the Learnt Model

Graph Neural Networks (GNNs) provide powerful tools for visualising inter-channel relationships in EEG data. This analysis explores the inter-channel connectivity patterns captured by the model for Subject 0 from the SEED [1]. The experiment averages and normalises the adjacency matrices derived from the Self-Organised graphs (SO-graphs 1, 2, and 3) for each emotional state: sadness, neutrality, and happiness. Based on these matrices, it identifies the key channels and visualises their influence using topographic brain maps.

Figures 5.6a, 5.6b, and 5.6c display the corresponding adjacency matrices and topographic representations for each emotional state of Subject 0 in the SEED [1]. SO-graph 1 exhibits diagonal dominance, indicating that a small subset of EEG channels exhibits high discriminative power during the early stages of graph convolution. In contrast, SO-graphs 2 and 3 reveal richer inter-channel connectivity patterns, capturing more complex brain dynamics necessary for differentiating emotional states. For

sadness and neutrality, prefrontal, centro-parietal, and parietal electrodes (such as F6, CPZ, and P5) demonstrate the largest connection weights. For happiness, the model highlights prefrontal, temporal, and temporal-parietal electrodes (such as TP8, T7, and F8) as the most discriminative regions. These findings align with previous neuroscience studies [335, 336], which associate prefrontal and centro-parietal activations with emotional processing. The analysis thus confirms the model's ability to capture meaningful inter-channel relationships critical for robust EEG-based emotion recognition

5.4 Summary

This chapter proposes the novel TRANSIT-EEG framework to address the challenge of EEG signal classification variability across subjects. The proposed framework systematically adapts models to new subjects through a two-stage process: generating subject-specific synthetic data for augmentation during finetuning, and adapting the task classifier using efficient low-rank finetuning. The proposed TRANSIT-EEG integrates three key innovations: the proposed Self-Organising Graph Attention Transformer (SOGAT) for robust EEG feature extraction, the proposed Individualised Diffusion Probabilistic Model (IDPM) for subject-specific data generation, and Low-Rank Adaptation (LoRA) modules for fast, memory-efficient model adaptation.

TRANSIT-EEG first employs meta-learning to train the IDPM-based synthetic data generator, enabling it to synthesise high-quality, subject-specific EEG signals conditioned on few-shot real samples. These synthetic samples augment the training dataset, allowing the model to maintain strong generalisation capabilities even with limited real data. Subsequently, the framework adapts the task classifier by finetuning only low-rank adapter modules within SOGAT, thus preserving the pre-trained knowledge while specialising the model for new subjects without catastrophic forgetting. Through this structured approach, TRANSIT-EEG significantly improves the accuracy and applicability of cross-subject EEG classification, particularly for neuroscience and healthcare applications. By addressing the critical bottlenecks of subject variability and data scarcity through innovative augmentation and adaptation techniques, the proposed framework provides a practical and scalable solution for

real-world EEG-based systems.

This chapter is based on the following work:

Journal 2: **Chirag Ahuja**, and Divyashikha Sethia. "TRANSIT-EEG - A Framework for Cross-Subject Classification with Subject-Specific Adaptation." **IEEE Transactions on Cognitive and Developmental Systems** (2024). (**SCIE, Publisher: IEEE**). Doi: https://doi.org/10.1109/TCDS.2025.3529669

Chapter 6

Generalisable Self-Supervised Learning for Few Shot EEG Signal Classification

Previous chapters present frameworks such as ADAPTER and TRANSIT-EEG to enhance cross-subject generalisation in few-shot EEG classification. However, these methods rely on labelled data or task-specific tuning and remain confined to single datasets. This dependence limits generalisation because high-quality EEG labels require costly, controlled recording setups. Self-supervised learning (SSL) eliminates the need for manual annotations during pretraining, enabling models to learn subject-invariant representations across diverse datasets. This chapter describes two SSL-based frameworks - **Self-Supervised Enhancement for Multidimension Emotion Recognition using Graph Neural Networks (SS-EMERGE)** and **Unify-Emotion Self Supervised Learning (Unify-ESSL)** - to address three core challenges in EEG classification: subject variability, label scarcity, and device heterogeneity. Using a unified encoder, the proposed `SS-EMERGE` framework learns representations across time, space, and frequency domains. It applies stimulus-aware contrastive learning to enable few-shot classification on homogenous datasets such as SEED [1] and SEED-IV [3]. The proposed encoder architecture integrates differential entropy features (spectral), GNN-based spatial modelling, and causal temporal convolution. The `Unify-ESSL` framework extends pretraining to heterogeneous datasets that differ in

sampling rates, number of channels, and stimuli durations. It proposes an intelligent dataset-alignment strategy that balances contributions across datasets, subjects, and sessions.

6.1 SS-EMERGE: Self-Supervised Multidimensional Representation Learning

6.1.1 Background and Motivation

EEG-based signal classification faces major challenges due to high inter-subject variability and the scarcity of labelled data. Variations in individual brain dynamics, coupled with the high annotation cost, make it difficult to train subject-invariant models using fully supervised learning. Sleep Staging (SS) offers a promising alternative by enabling representation learning from unlabelled EEG data. However, many existing SSL approaches ignore EEG's structured nature or fail to capture generalisable, subject-invariant features. SimCLR [79] and CPC [80] are widely used self-supervised methods. SimCLR [79] learns by maximising agreement between two augmented views of the same instance using a contrastive loss, while CPC [80] learns to predict future latent representations in a sequence. Although effective in vision and speech domains, these instance-level methods do not account for shared semantic structure in EEG trials collected from subjects exposed to the same stimulus. Moreover, their standard augmentations or temporal continuity assumptions may break the underlying semantic consistency of EEG signals.

Self-supervised Group Meiosis Contrastive Learning (SGMC) [14] introduced a contrastive approach tailored to EEG by grouping trials from subjects who viewed the same stimulus. Within each group, SGMC [14] applies a meiosis-inspired augmentation recombining segments from different subjects to generate homologous views. It then uses a group-level contrastive loss to align representations from the same stimulus group while separating those from unrelated groups. This formulation enables stimulus-aware and subject-invariant learning. However, SGMC [14] only models temporal features using a ResNet backbone and does not exploit EEG's rich spectral or spatial structure.

Multi-View Spectral-Spatial-Temporal Masked Autoencoder (MV-SSTMA) [76] addressed the need for multidimensional EEG processing by introducing a masked autoencoder that jointly models spectral, spatial, and temporal dimensions. It used a reconstruction-based objective to recover missing data across views. While this generative strategy improves within-dataset performance, it inherently preserves subject-specific details, which hinders cross-subject generalisation. Generative models are optimised to reconstruct all aspects of the input, including noise and identity-specific artifacts, making them less effective for learning subject-agnostic representations.

This thesis proposes **Self-Supervised Enhancement for Multidimension Emotion Recognition using Graph Neural Networks (SS-EMERGE)**, a contrastive representation learning framework for EEG that addresses the limitations of both SGMC [14] and MV-SSTMA [76]. SS-EMERGE combines stimulus-aware grouping and meiosis-inspired augmentation from SGMC with a multidimensional encoder structure adapted from MV-SSTMA. The encoder consists of: (i) a spectral embedding layer that maps Differential Entropy (DE) features from five canonical EEG frequency bands into learnable representations [337], (ii) a Graph Attention Network (GAT)[10] to model spatial relationships among electrodes, and (iii) a Temporal Convolutional Network (TCN)[338] to extract long-range temporal dependencies using causal convolutions.

This chapter discusses the proposed SS-EMERGE framework in detail. SS-EMERGE avoids generative reconstruction and instead learns subject-invariant homologous representations through contrastive training within stimulus groups. The learnt representations generalise effectively across subjects and support downstream classification tasks, such as emotion recognition, under few-shot learning settings. SS-EMERGE provides a scalable and label-efficient framework for EEG representation learning in homogeneous dataset scenarios. The main contributions of SS-EMERGE are as follows:

1. **Group-Based Contrastive Learning:** Application of a meiosis-inspired signal recombination strategy for group-level contrastive augmentation, enabling better cross-subject generalisation.

2. **Unified Multi-Domain Encoder:** Design of an encoder that integrates temporal causal convolutions, spectral embeddings from differential entropy features, and

spatial modelling using graph neural networks.

3. **Few-Shot Evaluation Across Protocols:** Empirical evaluation under few-shot learning protocols using both supervised and self-supervised training modes on SEED [1] and SEED-IV [3] datasets.

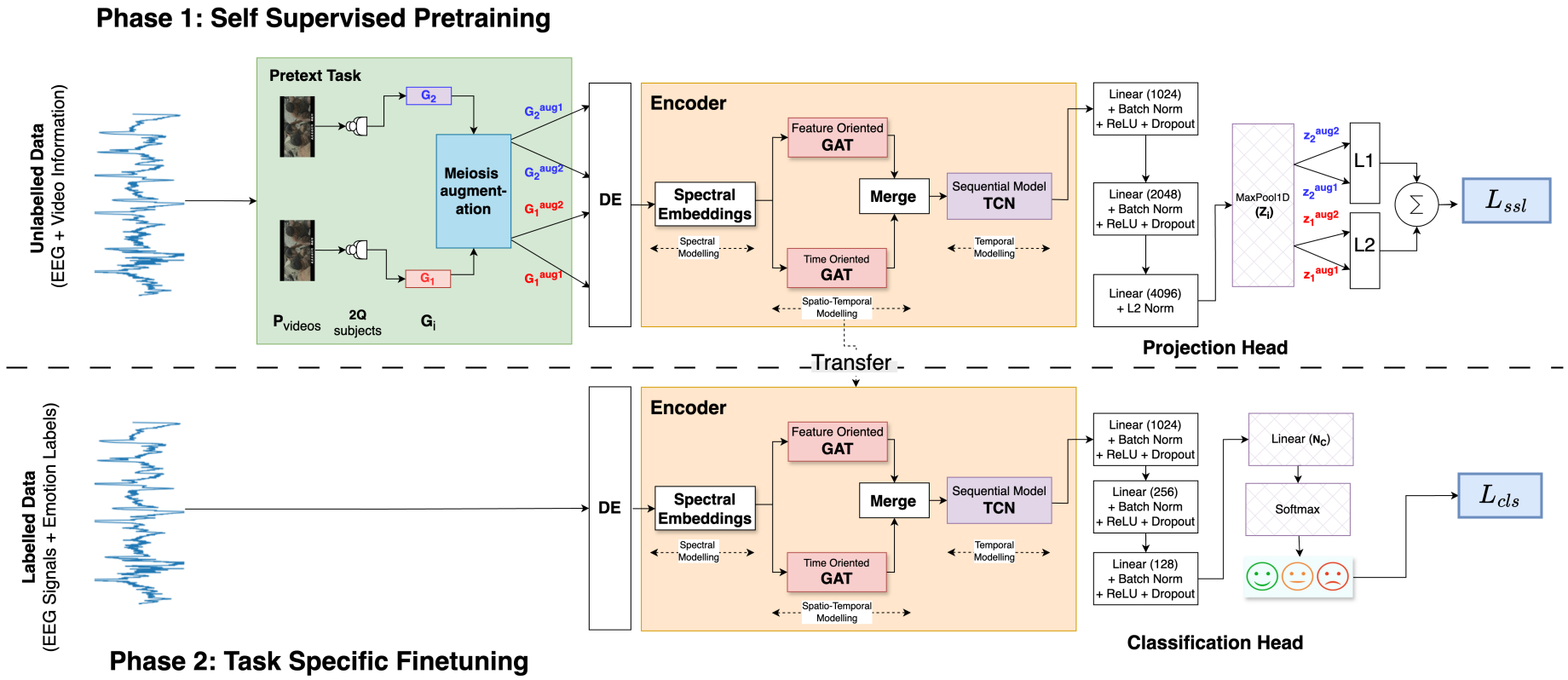
4. **Foundation Model Pretraining :** Cross-dataset pretraining using SS-EMERGE across SEED and SEED-IV, establishing a reusable foundation model for downstream EEG classification tasks.

6.1.2 Methodology

This section describes the proposed **Self-Supervised Enhancement for Multidimension Emotion Recognition using Graph Neural Networks (SS-EMERGE)** framework, designed to learn robust and subject-invariant representations from EEG signals. The methodology is structured into three parts as illustrated in Figure 6.1: The Encoder Architecture, Phase-1 (Self-Supervised pretraining), and Phase-2 (Task Specific finetuning).

6.1.2.1 Data Preprocessing and Input

SS-EMERGE processes EEG signals segmented into non-overlapping 1-second intervals. For each interval, it extracts the Differential Entropy (DE) features across five standard EEG frequency bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (14–31 Hz), and gamma (31–50 Hz). The spectral transformation captures the logarithmic energy of band-limited EEG activity and is robust for emotion recognition [1, 3]. In both the SEED [1] and SEED-IV [3], each trial is recorded from 62 EEG channels using a consistent electrode montage. The extracted DE features are organised as a tensor of shape $[F \times C \times T]$, where $F = 5$ denotes the number of frequency bands, $C = 62$ the number of channels, and T the number of temporal steps per trial (determined by the duration and sampling window). For SEED [1], all trials are of fixed duration, and T remains constant across the dataset. However, for SEED-IV [3], trial durations can vary across subjects and stimuli. Consequently, T is variable and requires padding to ensure uniformity for batch processing. The channel order is



P: #videos; **Q:** #subjects; $\mathbf{N_C}$: #classes; **G:** Group of subject for i^{th} video; $\mathbf{z_{sk}}$: Latent Feature for sk^{th} subject for vi^{th} video

Figure 6.1: SS-EMERGE framework showing Phase-1 pretraining with meiosis-based contrastive learning on unlabelled EEG and Phase-2 finetuning with supervised emotion classification using labelled data.

rearranged based on functional brain regions such as frontal, parietal, and temporal, following the canonical SEED [1] channel layout. This reordering facilitates structured attention masking in subsequent GAT layers [10], allowing for biologically grounded spatial reasoning over the EEG sensor space.

6.1.2.2 Encoder Architecture

The SS-EMERGE encoder architecture learns meaningful representations from *unlabelled* EEG data during a self-supervised pretraining phase, subsequently adapting these representations to downstream *labelled* tasks such as emotion classification. This two-stage pipeline enables the model to generalise across subject variability and recording conditions without relying on extensive labelled datasets upfront. The SS-EMERGE encoder architecture aims to extract rich and generalisable representations of emotional states from raw EEG signals. EEG data is inherently high-dimensional, noisy, and varies significantly across subjects. To address these challenges, the SS-EMERGE model integrates frequency-specific information (spectral modelling), spatial dependencies across brain regions (via graph attention), and temporal evolution of neural patterns (using a TCN [338]). The design is modular, enabling separate but robust learning across these three axes of signal processing. At a high level, the encoder processes frequency-domain DE features, embeds them using spectral projections, captures spatial-temporal relationships through GATs, and finally aggregates temporal dynamics with a TCN [338] to yield a compact and informative representation suitable for self-supervised and supervised learning.

The SS-EMERGE encoder integrates spectral, spatial, and temporal components to model EEG dynamics effectively. The architecture begins with a spectral embedding layer that projects Differential Entropy (DE) features from F frequency bands into a D-dimensional space using a learnable linear transformation. To capture spatial dependencies, the encoder employs a Feature-GAT module that applies graph attention across EEG channels at each time step, enabling the model to learn inter-regional relationships. The model captures temporal dependencies in two stages. First, the Time-GAT module applies graph attention across time steps within each channel to model the local temporal context. Then, the Temporal Convolutional Network (TCN)

uses dilated causal convolutions to capture long-range temporal dependencies across the EEG sequence. These modules jointly enable the encoder to represent EEG signals in a domain-aware and subject-invariant manner.

1. **Spectral Modelling:** The encoder begins by encoding the spectral structure of EEG signals using a learnable projection layer. The input is a tensor of shape $[F \times C \times T]$, where $F = 5$ represents the number of frequency bands (delta, theta, alpha, beta, gamma), $C = 62$ denotes the EEG channels, and T is the number of time steps, which remains fixed in SEED [1] and varies in SEED-IV [3]. For each time step t and channel c, the model extracts a feature vector $x_{c,t} \in \mathbb{R}^F$, which contains the Differential Entropy (DE) values across all five frequency bands. Equation 6.1 applies a shared linear transformation to project the band-specific input vector into a D-dimensional latent space:

$$z_{c,t} = W^{(s)} x_{c,t} + b^{(s)} \tag{6.1}$$

where, $x_{c,t} \in \mathbb{R}^F$ is the DE vector for channel c at time t, $W^{(s)} \in \mathbb{R}^{D \times F}$ is the learnable weight matrix, $b^{(s)} \in \mathbb{R}^D$ is the bias term, and $z_{c,t} \in \mathbb{R}^D$ is the resulting spectral embedding. The model applies this projection across all channels and time steps to obtain the tensor $Z \in \mathbb{R}^{C \times T \times D}$, which encodes frequency-aware embeddings aligned across the spatial and temporal axes. The resulting representation captures both spectral content and temporal position, forming the basis for downstream spatial and temporal reasoning.

2. **Spatio-Temporal Modelling:** After extracting frequency-aware embeddings, SS-EMERGE models the spatial and temporal relationships in the EEG signal using two stacked GAT [10]. These modules operate on graph representations constructed across channels (spatial) or time steps (temporal). Combining both captures the brain's dynamic behaviour across regions and over time.

   (a) **Feature-GAT:** The first GAT models spatial relationships between EEG channels. At each time step t, SS-EMERGE defines a graph $\mathcal{G}_t = (\mathcal{V}, \mathcal{E})$, where each node $v_i \in \mathcal{V}$ corresponds to an EEG channel and edges $\mathcal{E}$ are defined based on anatomical or functional proximity. The node

feature $h_{i,t} \in \mathbb{R}^D$ encodes the spectral embedding for channel i at time t. Equation 6.2 computes the attention score $\alpha_{ij}^{(t)}$ between node i and its neighbour j, allowing the model to weigh contributions from spatially adjacent regions:

$$\alpha_{ij}^{(t)} = \frac{\exp\left(\text{LeakyReLU}\left(\mathbf{a}^\top [W h_{i,t} \,\|\, W h_{j,t}]\right)\right)}{\sum_{k \in \mathcal{N}(i)} \exp\left(\text{LeakyReLU}\left(\mathbf{a}^\top [W h_{i,t} \,\|\, W h_{k,t}]\right)\right)} \tag{6.2}$$

where, $W \in \mathbb{R}^{D' \times D}$ is a learnable weight matrix, $\mathbf{a} \in \mathbb{R}^{2D'}$ is the attention vector, $\|$ denotes concatenation, and $\mathcal{N}(i)$ is the neighbourhood of node i. The resulting embedding for node i is a weighted sum of its neighbours, capturing spatial interactions at each time step.

(b) **Time-GAT:** The second GAT captures temporal dependencies across the EEG sequence. For each channel c, the model constructs a graph $\mathcal{G}_c = (\mathcal{V}, \mathcal{E})$, where each node corresponds to a time step $t \in \{1, \ldots, T\}$. Since GATs are inherently permutation-invariant, SS-EMERGE adds sinusoidal positional encodings to each time step to preserve order. Equation 6.3 computes the attention weight $\beta_{ts}^{(c)}$, which models how strongly time step s influences time step t for a given channel:

$$\beta_{ts}^{(c)} = \frac{\exp\left(\text{LeakyReLU}\left(\mathbf{a}^\top [W h_{c,t}^{(f)} \,\|\, W h_{c,s}^{(f)}]\right)\right)}{\sum_{k=1}^{T} \exp\left(\text{LeakyReLU}\left(\mathbf{a}^\top [W h_{c,t}^{(f)} \,\|\, W h_{c,k}^{(f)}]\right)\right)} \tag{6.3}$$

where, $h_{c,t}^{(f)} \in \mathbb{R}^{D'}$ is the output of the *Feature-GAT* for channel c at time t, and $\beta_{ts}^{(c)}$ denotes the normalized attention from step s to t. The output $h_{c,t}^{(g)}$ captures context-aware temporal representations enriched with both feature and time dynamics.

(c) **Merge:** A merge operation fuses the outputs from the `Feature-GAT` and `Time-GAT` branches before passing the representation to the TCN [338] module. This results in a unified tensor of shape $C \times T \times D'$ that captures both spatial and temporal context. The fused representation serves as the input to the temporal modelling stage.

3. **Temporal Modelling:** Following the spatial encoding from the dual GAT blocks, SS-EMERGE models the temporal progression of EEG signals using a TCN [338]. Although the Time-GAT captures non-local dependencies across time steps by attending to salient moments in the signal, it lacks an explicit inductive bias for sequential continuity or causality. To address this, SS-EMERGE introduces a TCN [338] module that complements the Time-GAT by enforcing temporal ordering and learning hierarchical temporal dynamics. The input to the TCN [338] is the fused tensor $H \in \mathbb{R}^{C \times T \times D'}$, where C is the number of EEG channels, T is the number of time steps, and D' is the feature dimension after GAT. The model first reshapes this tensor into $X \in \mathbb{R}^{T \times (C \cdot D')}$, treating each time step as a multivariate feature vector concatenated across channels. SS-EMERGE applies stacked 1D causal convolutional layers with increasing dilation factors and residual connections. These layers ensure that each output at time t is computed only from inputs at time steps $\leq t$, thereby preserving causal structure. Equation 6.4 expresses the output of a single dilated convolutional layer in the TCN [338]:

$$y_t = \sum_{i=0}^{k-1} w_i \cdot x_{t-d \cdot i} \tag{6.4}$$

where, $y_t \in \mathbb{R}^{D_{\text{tcn}}}$ denotes the output at time step t, $x_{t-d \cdot i} \in \mathbb{R}^{C \cdot D'}$ is the input from an earlier time step with dilation factor d, k is the kernel size, and w_i are learnable convolutional weights. The stacked dilated convolutions expand the receptive field exponentially, enabling the model to capture both local and long-range dependencies in the EEG sequence. After the final TCN [338] layer, SS-EMERGE applies a temporal pooling operation—such as average or max pooling—across time steps to obtain a fixed-length feature vector $z \in \mathbb{R}^{512}$. This vector aggregates frequency, spatial, and temporal information into a compact representation used by the projection and classification heads. Combining Time-GAT with TCN [338], the model learns salient temporal attention and smooth sequential transitions, yielding a temporally grounded and context-aware embedding.

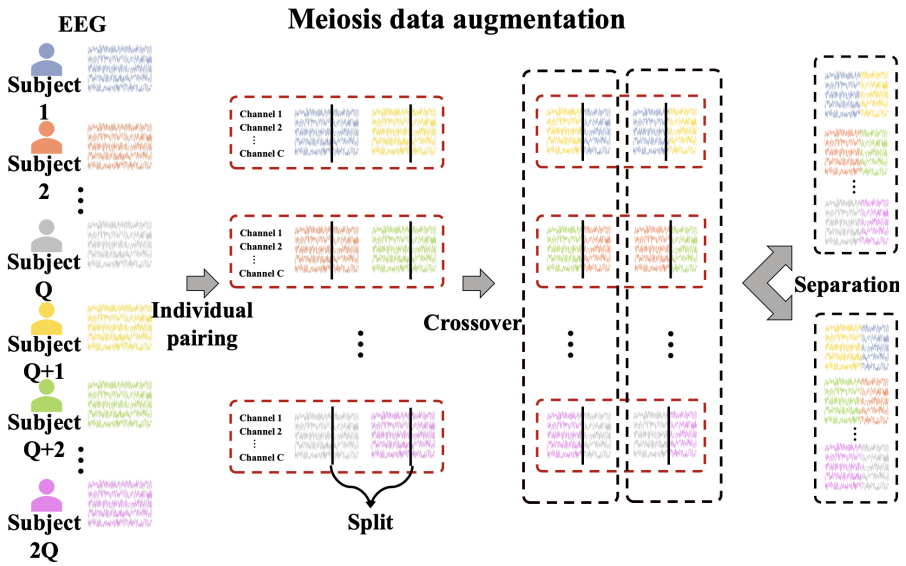


Figure 6.2: Meiosis-based data augmentation [14] used in SS-EMERGE framework.

6.1.2.3 Phase-1: Self Supervised Pretraining

Phase-1 (pretraining) improves the discriminative ability of the encoder by learning latent representations through Meiosis-based contrastive learning, using augmented pairs created via group-level crossover. In the context of EEG data, contrastive learning learns distinct differences in brain responses that can signify various cognitive or emotional states in response to identical stimuli. Meiosis augmentation enhances EEG-based emotion recognition by cross-exchanging signal segments from different subjects within the same group. It creates augmented pairs, where positive samples come from subjects who watched the exact video (same stimuli), and negative samples come from different video groups (different stimuli). Meiosis-based contrastive learning helps the model to distinguish between similar and dissimilar stimuli, improving its ability to recognise emotions. Algorithm 2 outlines the end-to-end Phase-1 pretraining using contrastive learning with meiosis augmentation. It begins by extracting group-level features and projecting them into a latent space. This step derives compact group representations for contrastive learning. The algorithm then minimises the distance between positive pairs and maximises amongst the negative pairs [79, 80, 14].

1. **Stimulus-Driven Contrastive Learning in EEG:** Learning subject-invariant representations remains a core challenge in EEG-based emotion recognition due to high inter-subject variability and limited labelled data. Contrastive

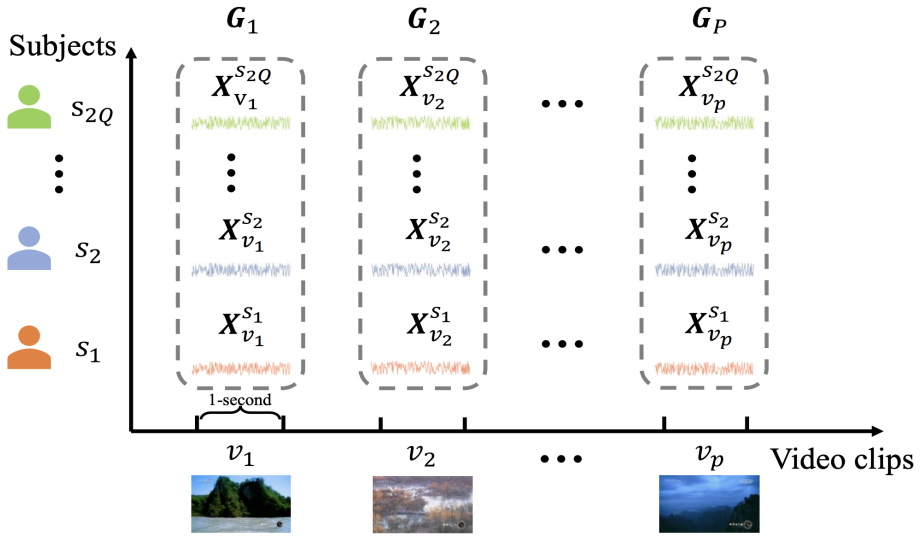


Figure 6.3: Group formation in meiosis-based data augmentation [14] used in SS-EMERGE framework.

learning has emerged as a powerful strategy for learning robust representations by distinguishing between semantically similar and dissimilar examples. Traditional contrastive methods such as Simple Contrastive Learning (SimCLR) [79] and Contrastive Predictive Coding (CPC) [80] define positive pairs using augmented views of the same trial or temporal context while treating all other samples as negatives. Although effective in other modalities, these approaches fail to capitalise on the unique structure of EEG signals, where responses to the same stimulus across subjects exhibit shared emotional patterns despite individual variability.

To address the limitations of instance-based contrastive learning for EEG signals, Zheng et al. [14] proposed the Self-supervised Group Meiosis Contrastive Learning (SGMC) framework—*a stimulus-aware self-supervised learning method that leverages cross-subject consistency in emotional responses*. The key insight behind SGMC is that EEG signals recorded from different individuals, when exposed to the same stimulus (e.g., a video clip), share underlying emotional structure despite individual variability. Instead of treating each EEG trial independently, SGMC constructs *stimulus-specific groups* G_i, where each group contains EEG segments from $2Q$ different subjects who viewed the same video clip v_i (see Figure 6.3). Within each group, SGMC applies a

meiosis-inspired data augmentation strategy to create two homologous views, G_i^A and G_i^B (illustrated in Figure 6.2). Each EEG trial is split at the temporal midpoint, and crossover recombination is performed between randomly paired signals to simulate structured but diverse variants while preserving stimulus semantics. The model splits each EEG trial at the temporal midpoint and performs crossover recombination between randomly paired signals to simulate structured yet diverse variants while preserving stimulus semantics. It then passes the augmented groups through a shared encoder and applies a group-level projector that aggregates individual trial embeddings into a single group-level representation using a symmetric function, specifically MaxPool1D.

SGMC formulates the contrastive learning objective at the group level. Given a minibatch of P stimulus groups, the framework computes a temperature-scaled cross-entropy loss between the group-level embeddings of homologous views. Equation 6.5 defines the loss for an anchor view $\mathbf{z}_i^A$, encouraging similarity with its corresponding positive view $\mathbf{z}_i^B$ while contrasting against all other groups in the minibatch:

$$\ell_i^A = -\log \frac{\exp\left(s(\mathbf{z}_i^A, \mathbf{z}_i^B)/\tau\right)}{\sum_{j=1}^{P} \mathbb{1}_{[j\neq i]} \exp\left(s(\mathbf{z}_i^A, \mathbf{z}_j^A)/\tau\right) + \sum_{j=1}^{P} \exp\left(s(\mathbf{z}_i^A, \mathbf{z}_j^B)/\tau\right)} \tag{6.5}$$

where, $s(\cdot,\cdot)$ denotes cosine similarity, and τ is a temperature hyperparameter. The indicator function $\mathbb{1}_{[j\neq i]}$ masks out the positive pair in the denominator. Equation 6.6 computes the total contrastive objective by averaging over both augmented views of all P groups in the batch:

$$\mathscr{L} = \frac{1}{2P} \sum_{i=1}^{P} \left(\ell_i^A + \ell_i^B\right) \tag{6.6}$$

This stimulus-grouped contrastive formulation increases the number of informative negatives while maintaining high semantic consistency among positives. As a result, it yields more *discriminative and subject-agnostic representations*. SGMC has demonstrated superior performance over instance-based contrastive

Algorithm 2 Pretraining with meiosis-based contrastive learning used in SS-EMERGE framework.

Initialization:
Set the number of epochs E_1
Initialize the model parameters
for each epoch $t = 1, 2, \ldots, E_1$ **do**
    Video Clip Sampling:
    Sample P video clips per iteration until all clips are enumerated
    EEG Sample Extraction:
    Extract $2P \times Q$ EEG samples: $D = \{X_{sk}^{vi} \mid i = 1, \ldots, P;\ k = 1, \ldots, 2Q\}$
    Pack these into groups: $G = \{G_i \mid i = 1, \ldots, P\}$
    Meiosis Data Augmentation:
    Randomly generate split position c such that $1 < c < M - 1$
    Apply Meiosis crossover on each group to get:
        $\tilde{G} = \{\tilde{G}_i^t \mid i = 1, \ldots, P;\ t \in \{X, Y\}\}$
    Feature Extraction and Projection:
    Extract group-level features
    Project features into latent space to get representations
    Loss Calculation:
    Compute contrastive loss L using Equation 6.5
    Model Update:
    Perform backpropagation to compute gradients
    Update model parameters using the optimiser
end for

learning methods in within-dataset EEG emotion recognition. However, it remains limited to single-dataset settings. It does not explicitly address the challenges of *cross-dataset generalisation*, where variability in recording setups, hardware, and stimulus labelling can degrade performance.

SS-EMERGE integrates this group-based contrastive learning strategy into its multidimensional encoder. Phase-1 in Figure 6.1 shows the setup, where each stimulus-specific group G_i undergoes meiosis augmentation to produce two homologous views $G_{i,\text{aug1}}$ and $G_{i,\text{aug2}}$. Each view passes through the encoder, and the resulting group representations $\mathbf{z}_i$ and $\mathbf{z}_j$ are optimised using the contrastive loss in Equation 6.5. Algorithm 2 outlines the pretraining process with meiosis-based augmentation. The algorithm samples stimulus groups, generates augmented pairs, extracts group-level features, computes the contrastive loss, and updates the encoder parameters through gradient descent.

2. **Projection Head and Group-Level Representation:** During Phase-1, SS-

EMERGE projects the encoder output through a multi-layer projection head to generate high-dimensional embeddings for contrastive learning. As shown in Phase-1 of Figure 6.1, the projection head consists of three sequential linear layers. The first two layers apply ReLU activation, dropout, and batch normalisation, followed by a final linear layer that outputs a 4096-dimensional vector, which undergoes ℓ_2-normalisation. Given a stimulus-specific group G_i, the model encodes individual subject trials into feature vectors using the shared encoder and projection head. SS-EMERGE applies a **1D max pooling** operation across the subject dimension to obtain a single group-level representation. It produces a pooled embedding $\mathbf{z}_i \in \mathbb{R}^{4096}$, which captures the most prominent activations across subjects while remaining invariant to individual variability. This pooling ensures robustness and reduces sensitivity to outlier signals. The model computes the contrastive loss over these group-level embeddings from homologous augmentations G_i^{aug1} and G_i^{aug2}. It uses the NT-Xent loss to bring augmented views of the same stimulus group closer in embedding space while pushing apart representations of different groups. This setup enables the encoder to learn subject-invariant and stimulus-consistent EEG features without supervision.

6.1.2.4 Phase-2: Task Specific Finetuning

In Phase-2 (finetuning), the pretrained encoder generates the latent representation and further finetunes using the classification head seen in Phase-2 of Figure 6.1 during task-specific finetuning. It updates only the classification head while keeping the encoder parameters frozen. SS-EMERGE appends the classification head over the learnt model.

1. **Classification Head**: The classifier includes a sequence of fully connected layers that transform the latent representation of dimension D. These layers culminate in a softmax layer that produces the final emotion class probabilities (N_C).

   (a) *Fully Connected Layers*: The layers with 1024, 256, and 128 neurons progressively transform and reduce the dimensionality of the latent representation.

(b) *Class Logits (FC (N_C))*: This layer outputs N_C neurons, where N_C corresponds to the number of emotion classes.

(c) *Softmax Layer*: Converts the logits into class probabilities using Softmax.

2. **Training Procedure**:

(a) It finetunes the classification head over the pretrained model using labelled EEG data.

(b) The training process involves minimising a cross-entropy loss [258], which learns to discriminate different predicted class probabilities and the true emotion labels.

(c) It optimises the network using Stochastic Gradient Descent (SGD), with appropriate learning rate scheduling to ensure convergence.

6.1.3 Experiments and Results

6.1.3.1 Experimental Setup

This section evaluates the proposed SS-EMERGE model on two publicly available EEG datasets: SEED [1] and SEED-IV [3]. SEED [1] contains recordings from 15 subjects across three emotion categories: positive, neutral, and negative. SEED-IV [3] expands the emotion set to four classes: happy, sad, fear, and neutral, with 72 trials per subject. Both datasets utilise a 62-channel ESI NeuroScan [102] acquisition system and adhere to a uniform preprocessing pipeline. The model segments each EEG trial into 1-second non-overlapping windows. It extracts Differential Entropy (DE) features across five standard frequency bands and passes them into the multidimensional encoder. During pretraining, the model groups unlabelled EEG segments by stimulus and applies meiosis augmentation to form contrastive pairs. In the finetuning phase, the model adds a classification head and optimises using a small number of labelled samples. The evaluation considers two experimental setups: a *subject-independent* validation using LOSO and a *subject-mixed* validation using an 80-20 train-test split across all subjects. Each setup measures classification accuracy under varying levels of label availability (10%, 50%, 100%) to simulate few-shot learning conditions.

Table 6.1: Hyperparameters used to tune SS-EMERGE framework.

Phase	Parameter	Value
SSL Phase	Learning Rate	1×10^{-3}
	Momentum	0.1
	Batch Size	64
	Epochs	3288
	Meiosis Augmentation - P	16 (video clips per iteration) (SGMC [14])
	Meiosis Augmentation - Q	2 (samples per group) (SGMC [14])
Classification Phase	Learning Rate	1×10^{-3}
	Batch Size	256
	Epochs	100
	Optimization	Adam with learning rate scheduling
	Loss Function	Cross-entropy
	Dropout	0.5

Phase-1 (pretraining) and Phase-2 (finetuning) select the hyperparameters used by original SGMC [14] along with the default settings of the optimiser. These settings yield good subject-independent and subject-mixed performance without the rigorous need for hyperparameter tuning. Table 6.1 lists all the hyperparameters, along with their corresponding default values, during the pretraining and finetuning of the Emotion Classification model.

6.1.3.2 Experiment 1: Baseline Model Comparison

This experiment compares the effectiveness of two model architectures for EEG-based emotion classification under subject-independent and subject-mixed settings. The baseline model utilises a ResNet encoder, as adopted in SGMC [14]. The proposed SS-EMERGE model integrates temporal, spectral, and spatial processing through a multidimensional encoder. The evaluation uses a LOSO validation for subject-independent/cross-subject testing. In each fold, the model trains on EEG trials from $N-1$ subjects and tests on the remaining subjects. This setup provides a rigorous measure of cross-subject generalisation. A subject-mixed setup also complements the evaluation, where the model randomly splits the data into 80% training and 20% testing sets across subjects. Table 6.2 reports per-class and average classification accuracies for both models on the SEED [1] to test the robustness of the proposed architecture. The results highlight the performance gap between architectures under both evaluation schemes.

Table 6.2: SS-EMERGE's performance on SEED [1] dataset under subject-independent and cross-subject setups.

Model	**Setup**	**Positive Accuracy (%)**	**Neutral Accuracy (%)**	**Negative Accuracy (%)**	**Average Accuracy (%)**
ResNet (used in SGMC [14])	Subject-mixed	89.7	90.0	89.8	89.83
	Cross-Subject	68.5	69.8	65.2	67.83
SS-EMERGE	Subject-mixed	96.2	92.5	94.1	94.27
	Cross-Subject	87.4	83.1	81.9	84.13

Table 6.2 shows that ResNet achieves 67.83% accuracy under the cross-subject setup, indicating limited generalisation to unseen subjects. SS-EMERGE significantly outperforms ResNet, reaching 84.13% accuracy in the same setting. This improvement demonstrates the benefit of multidimensional representation learning for subject-invariant emotion classification. Both models report higher accuracy in the subject-mixed setting, where the training and test sets share subject-specific patterns.

6.1.3.3 Experiment 2: Self-Supervised vs Fully Supervised Training

This experiment evaluates the proposed SS-EMERGE model under two training paradigms: fully supervised and self-supervised learning. The model undergoes pretraining using a stimuli-aware contrastive learning task in the self-supervised setup. The model groups EEG trials from subjects exposed to the same video stimulus and applies meiosis-based recombination to generate homologous views. It then applies a contrastive loss to maximise the similarity between augmented views from the same group while separating them from views of unrelated stimuli. This procedure enables the encoder to learn stimulus-consistent and subject-invariant features without supervision.

Following pretraining, the model freezes the encoder and finetunes only a classification head. The finetuning process gradually introduces labelled data from the same training split used during pretraining, simulating a few-shot setup. The finetuning stages use 10%, 50%, and 100% of the available labelled data to evaluate performance under varying levels of supervision. In contrast, the fully supervised

Table 6.3: Comparison of self-supervised and fully supervised performance of SS-EMERGE framework on SEED [1] and SEED-IV [3].

Training Type	**Label Proportion**	**SEED [1] Accuracy (%)**	**SEED-IV [3] Accuracy (%)**
Self-Supervised **SGMC** [14]	10%	73.29	52.30
Self-Supervised **SS-EMERGE**	10%	75.50	58.10
Self-Supervised **SGMC** [14]	50%	82.91	68.05
Self-Supervised **SS-EMERGE**	50%	86.13	76.75
Self-Supervised **SGMC** [14]	100% (Finetuned)	90.75	75.20
Self-Supervised **SS-EMERGE**	100% (Finetuned)	92.35	81.51
Fully Supervised **SS-EMERGE**	100% (LOSO)	**94.27**	**84.25**

model trains end-to-end using only labelled data and optimises a cross-entropy loss without pretraining. Both setups train and evaluate on SEED [1] and SEED-IV [3] using a subject-mixed 80-20 train-test split, where both training and testing sets may include data from the same subjects.

Table 6.3 summarises the performance of SS-EMERGE and SGMC under both self-supervised and fully supervised training settings across the SEED and SEED-IV datasets. SS-EMERGE consistently outperforms SGMC across all proportions of labelled data. When finetuned using only 10% of labelled samples, SS-EMERGE achieves 75.50% on SEED and 58.10% on SEED-IV, surpassing SGMC by 2.21% and 5.8%, respectively. This margin widens further as label availability increases. At 50% supervision, SS-EMERGE improves to 86.13% on SEED and 76.75% on SEED-IV, outperforming SGMC by over 3% and 8.7% respectively. With full finetuning (100%), SS-EMERGE reaches 92.35% and 81.51%, establishing a strong performance ceiling under the self-supervised setup. Notably, even with 50% labelled data, SS-EMERGE outperforms SGMC trained on the full dataset, demonstrating better label efficiency. The fully supervised SS-EMERGE model trained on 100% data under a subject-mixed setup achieves the highest scores: 94.27% on SEED and 84.25% on SEED-IV.

However, the self-supervised SS-EMERGE variant closes the gap substantially and offers a compelling alternative in low-label conditions. These results confirm that the multidimensional encoder design and stimulus-aware contrastive pretraining strategy in SS-EMERGE support robust generalisation under both few-shot and fully supervised scenarios.

6.1.3.4 Experiment 3: Foundational Pretraining with Combined Datasets

This experiment evaluates the impact of foundational pretraining using a merged dataset composed of SEED [1] and SEED-IV [3]. Both datasets share the same 62-channel electrode configuration, allowing direct trial integration. To disambiguate stimuli across datasets, the model uses composite identifiers formed by concatenating the dataset and video ID during group formation. This evaluation includes the full SS-EMERGE model, referred to as SS-EMERGE.FTS, which integrates spectral (F), temporal (T), and spatial (S) encoders. Two ablation variants isolate subsets of these components: SS-EMERGE.FT retains only spectral and temporal branches, while SS-EMERGE.TS includes temporal and spatial modelling but excludes spectral input. The baseline SGMC.T [14] model uses a ResNet-based encoder that operates solely in the temporal dimensions. Due to computational constraints, all models use a subject-mixed setup when combining SEED [1] and SEED-IV [3].

Table 6.4 compares the performance of SS-EMERGE under different modelling configurations. SS-EMERGE.FTS integrates all three dimensions—**F**requency (spectral), **T**ime (temporal), and **S**pace (spatial)—and achieves the highest accuracy across both datasets and all label availability settings. SS-EMERGE.FT removes spatial modelling to retain only frequency and temporal branches, while SS-EMERGE.TS excludes frequency information and retains temporal and spatial modelling. SGMC.T refers to the baseline model from SGMC [14], which operates only in the temporal dimension using a ResNet backbone. The results confirm that SS-EMERGE.FTS provides the most discriminative representation while removing any branch leads to a consistent drop in performance. The additive benefit of each modelling dimension highlights the importance of multidimensional learning for generalisable EEG representation.

Table 6.4: SS-EMERGE's foundational pretraining performance on SEED [1] and SEED-IV [3].

#Samples	SGMC [14] T	SS-EMERGE.FT FT(Proposed)	SS-EMERGE.TS TS(Proposed)	SS-EMERGE FTS(Proposed)
SEED [1]				
10%	73.29%	73.82%	73.51%	**74.11%**
50%	83.71%	84.02%	83.84%	**84.42%**
100%	83.04%	85.56%	85.12%	**88.87%**
SEED-IV [3]				
10%	58.33%	56.81%	55.53%	**57.92%**
50%	66.44%	62.32%	61.02%	**63.54%**
100%	77.64%	76.31%	75.08%	**82.75%**

6.1.3.5 Visualisation of underlying embedding

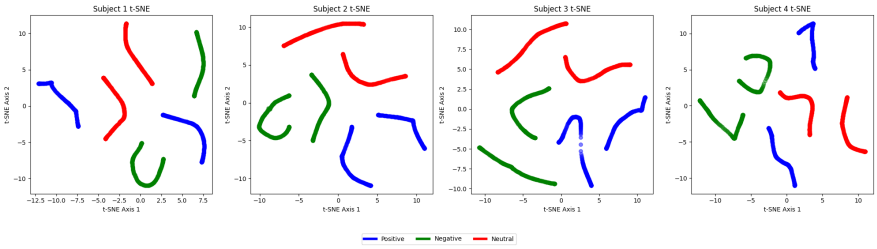


Figure 6.4: t-SNE visualisation of learnt embeddings by SS-EMERGE framework on SEED [1].

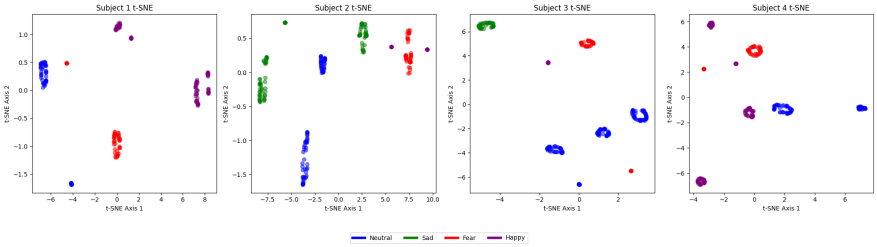


Figure 6.5: t-SNE visualisation of learnt embeddings by SS-EMERGE framework on SEED-IV [3].

Figures 6.4 and 6.5 visualise the learnt 4096-dimensional embeddings from SS-EMERGE using t-distributed Stochastic Neighbour Embedding (t-SNE). The projections reduce high-dimensional representations into two dimensions for interpretability. The visualisations show well-separated clusters corresponding to different emotion categories in SEED [1] and SEED-IV [3]. SS-EMERGE produces compact and discriminative feature groupings, confirming that the model effectively captures emotion-specific neural representations across subjects.

These experiments demonstrate that SS-EMERGE provides strong cross-subject

and subject-mixed generalisation with robust performance under a few-shot setup. The model consistently outperforms the baseline SGMC [14] across multiple evaluation settings. Foundational pretraining on merged datasets and ablation results further validate that SS-EMERGE outperforms SGMC and affects each modelling dimension's contribution, namely F, T, and S.

6.2 Unify-ESSL: A Unified Self-Supervised Learning Framework for Cross-Dataset EEG Emotion Recognition

6.2.1 Background and Motivation

Existing self-supervised learning frameworks for EEG, such as SS-EMERGE, assume homogeneous data distributions across subjects and sessions. SS-EMERGE achieves strong generalisation within datasets like SEED [1] and SEED-IV [3], which maintain consistent recording setups, including electrode montage and sampling frequency. However, real-world EEG data originate from heterogeneous sources that differ in acquisition setups, hardware specifications, and subject populations. These discrepancies introduce significant domain shifts that reduce model transferability. To address this, this thesis proposes Unify-ESSL, a framework for self-supervised pretraining across multiple heterogeneous EEG datasets, eliminating the reliance on homogeneous conditions.

Public EEG datasets such as SEED [1], DEAP [5], and DREAMER [15] differ in the number of EEG channels, sampling frequencies, trial durations, and stimulus setups. These structural mismatches cause severe shifts in the spatial and temporal signal distributions, rendering direct transfer between datasets unreliable. Models trained or pretrained on one dataset frequently fail when evaluated on another with different sensor layouts or annotation schemes. A unified and label-efficient pretraining strategy is therefore essential for robust cross-dataset generalisation.

Earlier cross-dataset methods mainly rely on supervised domain adaptation. Li et al. [170] align label distributions between SEED [1] and DEAP [5] through emotion

mapping, while Zhou et al. [339] use deep transfer learning with shared latent spaces. Although these approaches improve performance across datasets, they depend on labels from both source and target domains and remain sensitive to differences in protocol and device configuration. Recent techniques such as CDEER [311] and MSDG [312] adopt adversarial and meta-learning strategies to align dataset-specific distributions. While these methods reduce domain gaps, they continue to require supervision and perform poorly under mismatched electrode configurations or trial structures. Self-supervised methods such as BENDR [39] and NeuroGPT [340] remove the need for labels by pretraining on large EEG corpora. BENDR applies contrastive objectives to the TUH EEG dataset [176], and NeuroGPT uses masked segment prediction with transformer decoders. These studies show that large-scale pretraining benefits transfer learning. However, they require substantial computation and do not explicitly address structural mismatches in channel layout or stimulus design—factors critical for emotion recognition tasks.

This thesis proposes **Unify-Emotion Self Supervised Learning (Unify-ESSL)**, a unified and scalable self-supervised framework designed for cross-dataset EEG representation learning. Unify-ESSL targets heterogeneity in channel layouts, trial designs, and sampling frequencies by introducing structured alignment and balanced training. It aligns spatial dimensions using a 14-channel intersection across SEED, DEAP, and DREAMER [15], ensuring electrode consistency. Prior studies validate that this subset retains discriminative power for affective decoding. To handle dataset imbalance, Unify-ESSL introduces a stratified sampling scheme with a fixed 1:2:3 ratio across DREAMER, DEAP, and SEED, maintaining proportionate representation during pretraining. The encoder extends EEGNet [6] with Multi-Head Self-Attention (MHSA)[49] for enhanced temporal modelling. Unify-ESSL adopts self-supervised objectives such as SimCLR[79] and CPC [80] during pretraining. During finetuning, it freezes the encoder and updates a lightweight classifier on a few labelled samples, enabling generalisation with minimal supervision.

This chapter discusses the proposes Unify-ESSL framework, which enables label-efficient and device-agnostic representation learning for EEG emotion recognition across heterogeneous datasets. The unified spatial layout, stratified sampling, and multi-objective contrastive training support robust cross-dataset transfer without requiring

target-domain labels. The main contributions of Unify-ESSL are as follows:

1. **Cross-Dataset Alignment:** Temporal and spatial alignment of SEED [1], DEAP [5], and DREAMER [15] datasets into a unified representation with 14 common EEG channels and a fixed sampling rate.

2. **Stratified Pretraining Sampler:** Development of a stratified data sampler that ensures balanced coverage across datasets, subjects, sessions, and time windows during contrastive pretraining.

3. **Contrastive Learning on Heterogeneous EEG:** Evaluation of representation learning using contrastive methods, including SimCLR [79] and CPC [80], applied to diverse EEG datasets under a unified pretraining setup.

6.2.2 Methodology

Figure 6.6 presents the Unify-ESSL framework for cross-dataset EEG representation learning using self-supervised objectives. The proposed method addresses a central challenge in EEG-based emotion recognition: the lack of generalisation across datasets collected using different acquisition setups, hardware configurations, and subject populations. These heterogeneities introduce severe distributional shifts in the spatial and temporal characteristics of signals, which degrade model transferability. Unify-ESSL mitigates these challenges by standardising input representations, leveraging contrastive self-supervision, and enabling adaptation with minimal labelled data. The model comprises three components: a backbone encoder for spatiotemporal feature extraction, a pretraining stage for self-supervised learning using unlabelled EEG data from multiple datasets, and a finetuning stage for supervised emotion classification using limited labels from a single dataset. During pretraining, the model learns subject-invariant and device-agnostic EEG representations using objectives such as SimCLR and CPC. The framework uses a shared encoder across all pretext tasks and reuses it for downstream classification. In Phase-2, the pretrained encoder transfers to a target emotion recognition task, where only the classification head is updated using the available labelled data. To enable learning across heterogeneous datasets, Unify-ESSL includes two critical components. First, it applies a 14-channel

alignment strategy to ensure consistent spatial input across datasets. Second, it employs a stratified batch sampler that balances contributions from SEED [1], DEAP [5], and DREAMER [15] in a fixed ratio, accounting for differences in dataset size, trial durations, and sampling frequency. These design choices preserve representational uniformity during training and prevent overfitting to dominant sources. The remainder of this section describes the full Unify-ESSL pipeline in detail.

6.2.2.1 Encoder Architecture

Unify-ESSL uses a modified **EEGNet** [6] as the base encoder to extract representations from raw EEG signals. The framework selects EEGNet due to its lightweight architecture, parameter efficiency, and compatibility with large-scale self-supervised pretraining. The encoder processes input tensors with shape $[B \times C \times T \times 1]$, where B represents the batch size, C indicates the number of aligned EEG channels (14), and T specifies the number of temporal steps per trial.

The EEGNet encoder comprises three convolutional blocks:

- A **Conv2D** layer extracts initial temporal features across the T dimension.
- A **DepthwiseConv2D** layer independently learns spatial patterns across EEG channels for each temporal filter.
- A **SeparableConv2D** layer captures complex spatiotemporal interactions while reducing the number of parameters.

Batch Normalisation, ELU activation, Average Pooling, and Dropout follow each convolutional block. The final output from EEGNet is reshaped to a 3D tensor of shape $[B \times T' \times D]$, where T' is the reduced temporal dimension and D is the feature dimension. To enhance temporal modelling, Unify-ESSL augments the encoder with a **Temporal Encoder** block that includes sinusoidal positional encodings and a **Multi-Head Self-Attention (MHSA)** layer [49]. The MHSA layer attends all T' time steps in parallel, allowing the encoder to capture long-range dependencies. A Global Average Pooling layer then compresses the temporal dimension, yielding a fixed-length latent representation $z \in \mathbb{R}^D$.

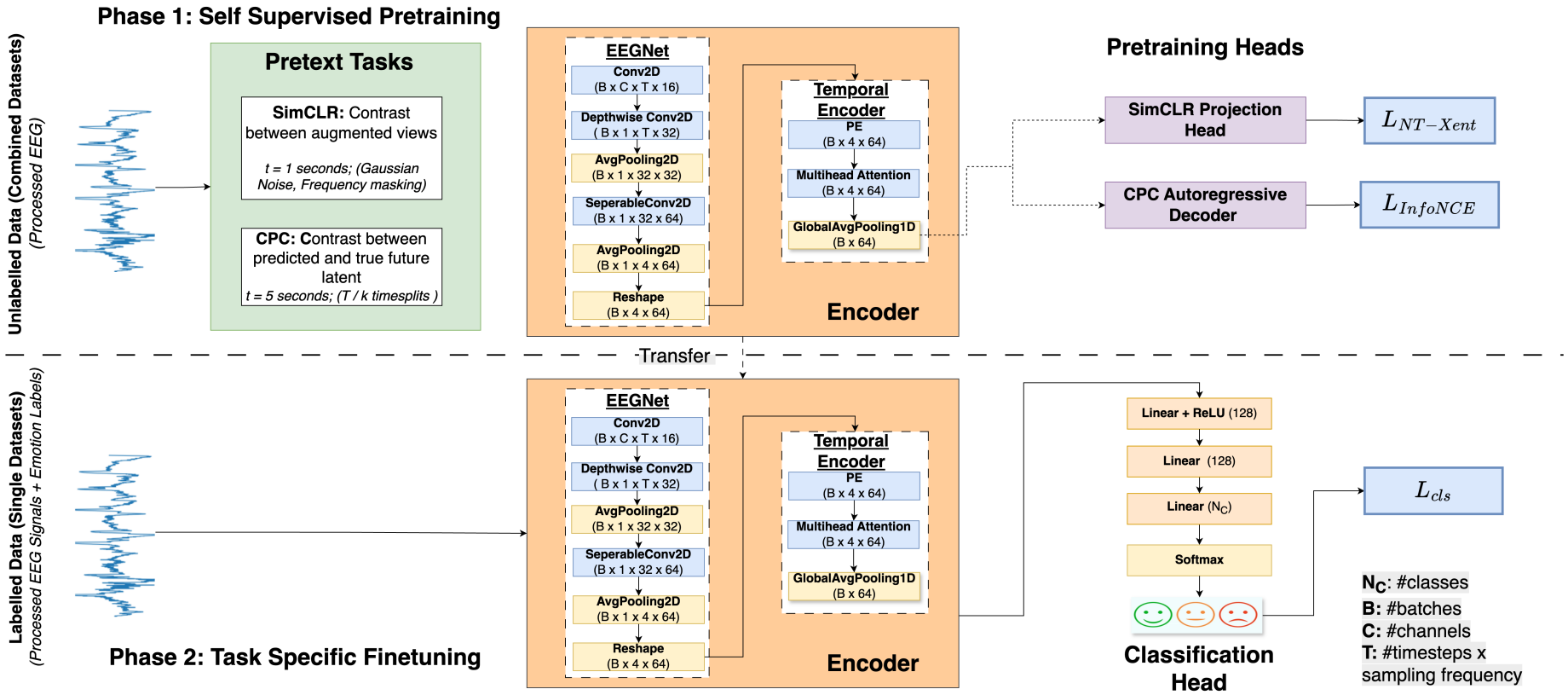


Figure 6.6: Unify-ESSL framework for cross-dataset EEG learning. Phase-1 pretrains a shared encoder using SimCLR and CPC on unlabelled EEG. Phase-2 finetunes a classification head using labelled data from a target dataset.

Table 6.5: Statistics of public EEG emotion recognition datasets used for evaluating Unify-ESSL framework.

Characteristic	**DEAP [5]**	**SEED [1]**	**DREAMER [15]**
Subjects	32	15	23
Trials per Subject	40	15	18
Channels	32	62	14
Stimuli Type	Music Videos	Film Clips	Film Clips
Emotion Labels	Valence, Arousal	3-Class Discrete	Valence, Arousal, Dominance
Average Trial Length (s)	60	240	65
Segments Generated (1s windows)	76,800	54,000	26,910

Algorithm 3 Stratified batch sampling in Unify-ESSL framework.

Define dataset weights: $w_{\text{DEAP}} = 1$, $w_{\text{SEED}} = 2$, $w_{\text{DREAMER}} = 3$
Compute sampling probabilities p_i using Equation 6.7
Initialise empty batch B
while B $<$ batch_size **do**
    Sample dataset $D_i \sim p_i$
    Sample subject s, trial t, segment w from D_i
    Add segment $(x_{s,t,w}, D_i)$ to batch B
end while
return B

6.2.2.2 Phase-1: Self-Supervised Pretraining

Phase-1 pretrains the Unify-ESSL encoder on a combined corpus of unlabelled EEG trials from SEED [1], DEAP [5], and DREAMER [15]. This phase enables the encoder to learn generalisable and subject-invariant EEG representations through contrastive self-supervision. It supports two self-supervised tasks: SimCLR [79] and CPC [80], each operating on different temporal input windows and training objectives.

1. **Data Integration and Preprocessing:** Table 6.5 compares the three EEG emotion recognition datasets used for pretraining. These datasets differ significantly in spatial layout (number of channels), temporal resolution (sampling frequency), trial duration, and label structure. For example, SEED contains long trials with 62 channels, while DREAMER [15] uses only 14 channels with shorter durations. These variations introduce inconsistencies in

spatial and temporal feature distributions, which can degrade unified model performance if left unaddressed. To standardise input structure across datasets, Unify-ESSL applies the following preprocessing steps:

(a) Downsample all EEG signals to 128 Hz to normalise sampling frequency.

(b) Apply a bandpass filter between 1–50 Hz to remove noise and baseline drift.

(c) Use ICA [341] to remove artefacts such as eye blinks and muscle activity.

(d) Retain only 14 electrodes (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4) that are common across all three datasets, as shown in Table 6.5.

(e) Segment EEG trials into fixed-length windows:

i. *1-second segments* (128 samples) for SimCLR to increase view diversity.

ii. *5-second segments* (640 samples) for CPC to ensure sufficient context for temporal prediction.

(f) Apply per-segment z-score normalisation independently per subject.

The framework selects different segment lengths based on the average trial durations reported in Table 6.5. All three datasets include sufficiently long trials (60–240 seconds) to support the 5-second segments required for CPC. In contrast, 1-second segments enable SimCLR to generate more augmented views per trial, which improves the quality of contrastive learning.

2. **Cross-Dataset Sampling Strategy:**Table 6.5 shows that the datasets vary significantly in both total segment count (ranging from 26,910 to 76,800) and trial duration. Naive random sampling builds batches in a way that lets larger datasets such as DEAP [5] dominate training and push the encoder toward those distributions. To address this imbalance, Unify-ESSL uses a stratified batch sampling strategy that explicitly controls dataset representation within each mini-batch. Algorithm 3 outlines the batch construction process. This sampler enforces balanced representation and allows the encoder to learn domain-agnostic features. By doing so, it prevents overfitting to any single dataset—a limitation observed in earlier approaches such as SS-EMERGE.

(a) **Static dataset weights:** Sampling weights are fixed as 1:2:3 for DEAP [5], SEED [1], and DREAMER [15], respectively. These weights compensate for differences in segment count and subject coverage.

(b) **Effective sampling probability:** Equation 6.7 defines the dataset selection probability p_i:

$$p_i = \frac{w_i N_i}{\sum_j w_j N_j} \tag{6.7}$$

(c) **Multi-level stratification:** Samples are drawn hierarchically—first by dataset, then by subject, trial, and time window—ensuring temporal and subject-wise diversity.

6.2.2.3 Pretraining Heads and Objectives

Unify-ESSL supports two contrastive self-supervised learning tasks during Phase-1: Simple Contrastive Learning (SimCLR) [79] and Contrastive Predictive Coding (CPC) [80]. Both tasks share the same backbone encoder but use distinct projection heads, input granularity, and training objectives. Figure 6.7 illustrates the architecture of the SimCLR projection head and the CPC autoregressive decoder.

1. **SimCLR Projection Head:** The SimCLR head maps encoder outputs to a 256-dimensional contrastive embedding. It applies a Multi-Layer Perceptron (MLP) with three operations:

   (a) A fully connected layer with 128 hidden units followed by Batch Normalisation.

   (b) A ReLU activation with dropout for regularisation.

   (c) A second fully connected layer that expands the embedding to 256 dimensions, followed by L2-normalisation.

   The model trains using the Normalised Temperature-scaled Cross Entropy (NT-Xent) loss [79]. Equation 6.8 defines this objective:

$$\mathscr{L}_{\text{NT-Xent}} = -\log \frac{\exp\left(s(\mathbf{z}_i, \mathbf{z}_j)/\tau\right)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp\left(s(\mathbf{z}_i, \mathbf{z}_k)/\tau\right)} \tag{6.8}$$

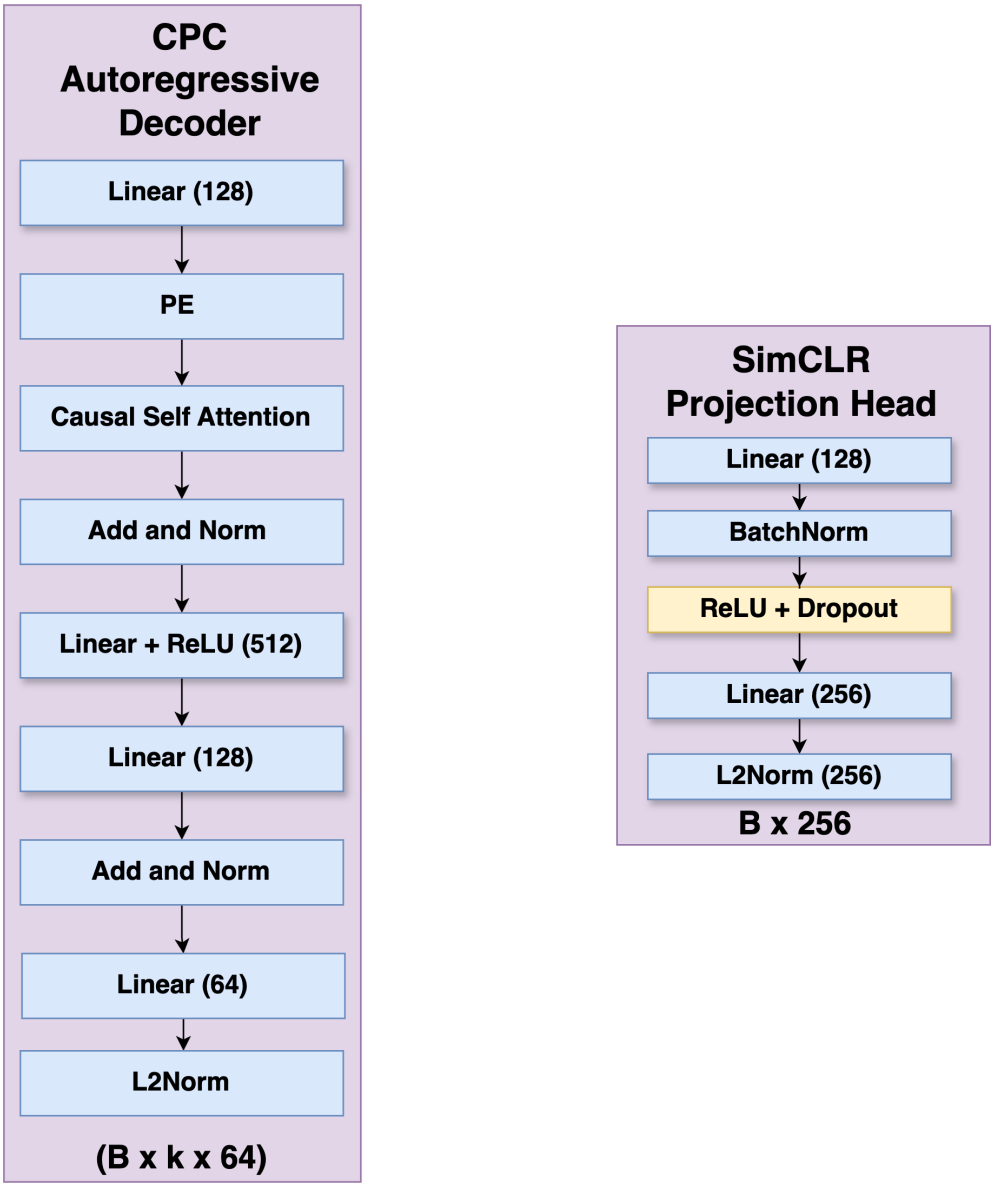


Figure 6.7: Pretraining heads used in the model architecture used Unify-ESSL framework. CPC autoregressive decoder (left) and SimCLR projection head (right).

where $s(\cdot,\cdot)$ denotes cosine similarity, τ is the temperature hyperparameter, and $2N$ is the total number of samples in the batch.

2. **CPC Autoregressive Decoder:** The CPC decoder predicts future latent embeddings from the past context using a transformer-style architecture. The decoder includes the following stages:

   (a) A linear layer projects the encoder output to 128 dimensions.

   (b) Sinusoidal positional encodings inject temporal order into the sequence.

(c) A causal self-attention block models autoregressive dependencies across time.

(d) Two residual blocks (Add and Norm) and a feed-forward network with ReLU activation (512 units) refine the context representation.

(e) The decoder projects the context to 64-dimensional vectors for future timestep prediction, followed by L2-normalisation.

The decoder optimises the InfoNCE loss [80], defined in Equation 6.9:

$$\mathscr{L}_{\text{InfoNCE}} = -\sum_{t=1}^{T-K} \log \frac{\exp\left(s(\mathbf{z}_{t+K}, \hat{\mathbf{z}}_{t+K})/\tau\right)}{\sum_{j=1}^{B} \exp\left(s(\mathbf{z}_{t+K}, \hat{\mathbf{z}}_j)/\tau\right)} \tag{6.9}$$

where $\hat{\mathbf{z}}_{t+K}$ is the predicted latent at time $t+K$, $\mathbf{z}_{t+K}$ is the ground-truth latent, and the denominator includes negatives sampled from the batch.

6.2.2.4 Phase-2: Task-Specific Finetuning

Phase-2 finetunes the learnt pretrained model by transferring the encoder to a downstream EEG-based emotion recognition task. It uses labelled samples from a single target dataset—typically SEED [1] or DEAP [5] or DREAMER [15] — to train a lightweight classification head while keeping the encoder weights frozen. This head consists of:

1. A fully connected layer with 128 hidden units and ReLU activation,
2. A second linear layer mapping to the number of emotion classes N_C,
3. A softmax layer to output class probabilities.

The model minimises the categorical cross-entropy loss $\mathscr{L}_{\text{cls}}$ between predicted and actual emotion labels. The model minimises the categorical cross-entropy loss [258] between predicted and actual emotion labels. The classification head finetuning uses the Adam optimiser with a learning rate of 1×10^{-5} and a batch size of 64 with early stopping based on the validation set accuracy. The encoder remains frozen throughout finetuning to prevent gradient updates, ensuring that performance reflects the quality of the representation learnt during pre-training. Unify-ESSL uses 1-second segments for classification, consistent with SimCLR's pretraining view granularity.

6.2.3 Experiments and Results

This section presents the results of experiments designed to optimise the Unify-ESSL configuration. The experiments systematically compare various pretraining and finetuning strategies across three public emotion recognition datasets collected with different setups, thus making them heterogeneous (SEED [1], DEAP [5], and DREAMER [15]) to determine the most effective approach.

6.2.3.1 Experimental Setup

The experimental setup evaluates model performance under both subject-mixed and cross-subject conditions. It prioritises cross-subject evaluation where feasible, using a Leave-One-Subject-Out (LOSO) to measure generalisation rigorously. However, the setup defaults to a standard subject-mixed train-test split due to computational constraints when full LOSO evaluation becomes impractical. This alternative ensures fast yet reliable model comparisons while maintaining subject-wise stratification. The setup consists of three stages: pretraining, finetuning, and testing. During pretraining, the model receives unlabelled EEG segments from the combined training split of the SEED [1], DEAP [5], and DREAMER [15] datasets. This phase trains the encoder using contrastive objectives to learn domain-invariant EEG representations. In the finetuning stage, the model freezes the encoder and attaches a classification head configured for the number of emotion classes in the target dataset. This head adapts to the specific label space while leveraging the pretrained features. The testing stage evaluates model performance on a held-out test split. This stage uses the **Macro-average F1 score (Macro-F1)** as the evaluation metric. Macro-F1 calculates the unweighted mean of F1 scores across all classes, making it robust to class imbalance and suitable for datasets with diverse emotion labels. This metric ensures fair and consistent evaluation across different datasets and learning strategies.

6.2.3.2 Experiment 1: Evaluating Self-Supervised Learning Strategies

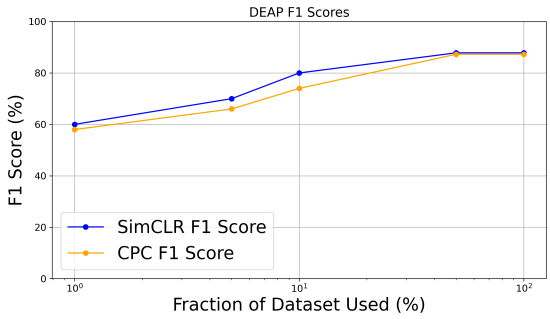
Experiment 1 compares the effectiveness of two SSL techniques, SimCLR [79] and CPC [80], in pretraining the model. Table 6.6 summarises the results obtained from this evaluation. Stratified sampling addresses the data imbalance in the number of

Table 6.6: Test F1-scores of Unify-ESSL framework on various datasets post finetuning on an individual dataset with 100% labelled data.

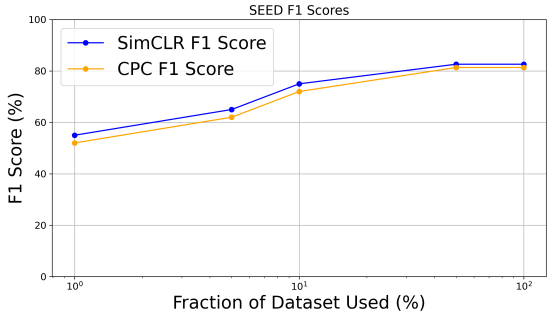
Learning Paradigm	**SSL Technique**	**DEAP [5] Macro-F1 (%)**	**SEED [1] Macro-F1 (%)**	**DREAMER [15] Macro-F1 (%)**
Fully-Supervised (32, 62, 14 channels respectively)	-	87.75	**85.28**	89.28
Self Supervised (14 channels)	SimCLR (1:1:1)	81.04	76.01	85.52
	CPC (1:1:1)	79.89	73.75	83.80
	SimCLR (1:2:3)	**87.83**	82.62	89.05
	CPC (1:2:3)	87.27	81.35	**91.23**

EEG samples and the length of each sample across the datasets. As shown in Table 6.5, SEED [1] has fewer EEG samples, but each sample is approximately 4 minutes long. When segmented with 1-second non-overlapping windows, this created 54,000 records. DEAP [5], on the other hand, has each EEG sample of 3-minute duration, but due to the inclusion of more subjects and trials per subject, it has a total of 76,800 EEG samples. DREAMER [15]contains 26,910 EEG samples, where each sample is around 1 minute only. Based on the total number of segments, the calculated weights for the sampling scheme are 1:1.45:2.85 for DEAP:SEED:DREAMER, which gets rounded to 1:2:3 for simplicity.

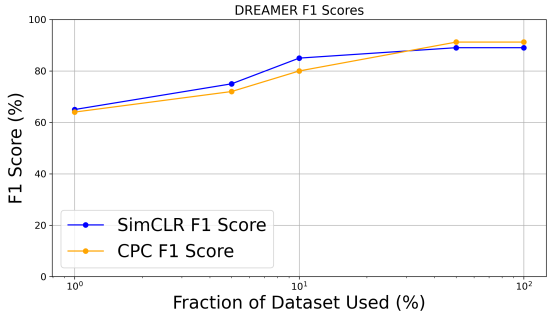
The results reveal that SimCLR slightly outperforms CPC across all datasets. However, the sampling ratio exerts a more significant impact on performance than the choice of SSL technique. With an optimal sampling ratio of 1:2:3, both SSL techniques approach or exceed the fully supervised performance. Figure 6.8 further illustrates the impact of SSL techniques on model performance with varying amounts of finetuning data.



(a) DEAP [5]



(b) SEED [1]



(c) DREAMER [15]

Figure 6.8: Unify-ESSL framework performance under different finetuning data sizes for SimCLR and CPC on SEED [1], DEAP [5], and DREAMER [15] datasets.

6.2.3.3 Experiment 2: Investigating Sampling Strategies Across Datasets

Experiment 2 investigates the impact of varying sampling ratios on model performance across different datasets. It tests various sampling ratios (0.5:1:2, 0.5:2:3, 1:1:2, 1:2:2, 1:2:3, and 1:1:1) for the DEAP:SEED:DREAMER datasets. Figure 6.9 shows the 1:2:3 sampling ratio achieves the best F1-scores of 87.82, 83.36, and 91.23 using CPC, compared to 84.82, 81.25, and 86.05 with the 1:1:1 ratio. These findings indicate that variations in sampling ratios can significantly impact model efficacy, particularly in addressing imbalances in the number of samples and their lengths. The oversampling strategy with a ratio of 1:2:3 for DEAP:SEED:DREAMER emerges as the

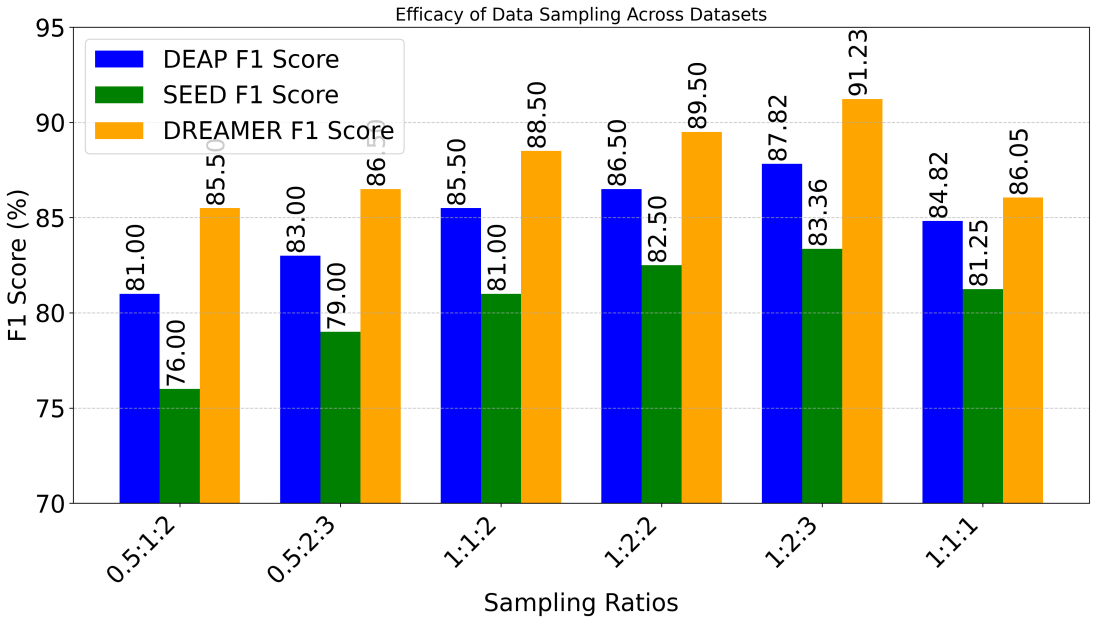


Figure 6.9: Effectiveness of Unify-ESSL proposed data sampling strategies across datasets.

most effective, yielding the highest F1 scores across all datasets. Figure. 6.9 illustrates the effectiveness of various sampling ratios, with 1:1:1 being the least effective and 1:2:3 the most effective. The results consistently show that the 1:2:3 sampling ratio produces the highest F1 scores across all datasets, underscoring its efficacy in enhancing model performance. Unify-ESSL enables scalable cross-dataset EEG learning using lightweight architecture and stratified contrastive pretraining.

6.3 Summary

This chapter presents two complementary self-supervised learning frameworks — SS-EMERGE and Unify-ESSL — that address the challenges of few-shot learning and cross-dataset generalisation in EEG-based emotion recognition.

SS-EMERGE framework, proposes a multidimensional encoder architecture that integrates causal convolutions to model temporal dynamics, differential entropy embeddings for spectral features, and graph attention networks to capture spatial relationships among electrodes. The framework employs a group-level contrastive learning objective based on meiosis augmentation, building on the stimulus-aware contrastive strategy introduced in SGMC [14]. This objective aims to align

representations for EEG trials from different subjects responding to the same stimulus, thereby enhancing subject-invariant encoding. Experimental results on SEED [1] and SEED-IV [3] demonstrate that SS-EMERGE outperforms supervised and self-supervised baselines under cross-subject and low-label conditions. Ablation studies validate the additive benefit of the multidimensional architecture, and t-SNE visualisations confirm that the learnt embeddings form well-separated clusters across emotion classes.

The Unify-ESSL framework aims to address SS-EMERGE's shortcomings in cross-dataset scenarios, facilitating generalizable pretraining across heterogeneous EEG corpora. While SS-EMERGE operates on datasets recorded with the same device and channel configuration, Unify-ESSL integrates SEED [1], DEAP [5], and DREAMER [15] using a unified 14-channel layout and standardised preprocessing. The model retains the group-level contrastive objective of SS-EMERGE but addresses the dataset imbalance and acquisition variability through a stratified batch sampling strategy. This sampler proportionally samples trials across datasets based on their segment counts and temporal granularity, preventing bias during pretraining. The encoder employs a lightweight EEGNet [6] backbone enhanced with self-attention, favouring scalability over the architectural complexity proposed in earlier chapters. Empirical results demonstrate that Unify-ESSL achieves strong generalisation performance across datasets, with performance on DREAMER [15] surpassing fully supervised baselines. Notably, the sampling ratio 1:2:3 for DEAP [5], SEED [1], and DREAMER [15] yields consistent improvements across SSL variants, confirming the importance of balanced dataset representation during pretraining. Together, SS-EMERGE and Unify-ESSL advance self-supervised learning in EEG by enabling robust, stimulus-aware representation learning across subjects and datasets. These models provide a foundation for building scalable, label-efficient EEG models that can adapt to real-world recording conditions and population diversity variability.

This chapter is based on the following work:

Conference 3: Chirag Ahuja, and Divyashikha Sethia. "UNiFY-ESSL: UNIfied Framework for Yielding EEG-based Emotion Recognition Model with Self-Supervised Learning." **2024 International Conference onBrain-Computer Interface & Healthcare Technologies (iCon-BCIHT), Thiruvananthapuram, India, 2024, pp. 249-256**. (**Publisher: IEEE**) https://10.1109/iCon-BCIHT63907.2024.10882427

Journal 3: Chirag Ahuja, and Divyashikha Sethia. "SS-EMERGE: Self-Supervised Enhancement for Multidimensional Emotion Recognition using GNNs for EEG" **Scientific Reports** (2025). (**SCIE, Publisher: Springer Nature, Impact Factor: 3.8**) https://doi.org/10.1038/s41598-025-98623-7

Chapter 7

Conclusion, Future Work & Social Impact

This thesis investigates the problem of few-shot EEG signal classification by addressing three fundamental challenges in EEG signal processing: subject generalisation, scarcity of labelled data, and cross-device variability. The overarching objective is to develop a comprehensive Few-Shot Learning (FSL) framework that overcomes these challenges through principled data augmentation, efficient cross-subject knowledge transfer, and self-supervised learning from auxiliary datasets. To address these challenges, the thesis proposes four core frameworks: (1) a principled taxonomy of EEG-specific few-shot learning strategies, (2) an automated augmentation method (Automatic Data Augmentation and PerTubation for Emotion Recognition (ADAPTER)) that learns class-conditional policies over geometric transformations, (3) a subject-adaptive classification framework (Transfer and Robust Adaptation of New Subjects in EEG Technology (TRANSIT-EEG)) that integrates meta-learned subject embeddings, generative augmentation using Individualised Denoising Probabilistic Model (IDPM), and parameter-efficient finetuning with Low Rank Adaptation (LoRA), and (4) two self-supervised learning frameworks (Self-Supervised Enhancement for Multidimension Emotion Recognition using Graph Neural Networks (SS-EMERGE) and Unify-Emotion Self Supervised Learning (Unify-ESSL)) that enable robust EEG representation learning across heterogeneous subjects and devices.

Collectively, these frameworks form a unified system that improves generalisation,

reduces reliance on labelled data, and supports scalable EEG learning across both subjects and acquisition settings. By combining principled augmentation, subject-conditioned adaptation, and multidimensional representation learning, the thesis advances the field toward real-world deployment of EEG systems in applications such as brain-computer interfaces, mental health monitoring, and neuroadaptive technologies. The proposed frameworks demonstrate significant improvements across multiple EEG datasets and settings.

Automatic Data Augmentation and PerTubation for Emotion Recognition (ADAPTER) enhances SEED LOSO accuracy from 86.69% to 88.54% by learning class-specific transformation sequences through differentiable policy search. In addition to boosting performance, it reveals interpretable augmentation patterns—such as frequency-based transformations tailored to specific emotion classes—thereby improving both the accuracy and transparency of the model.

Transfer and Robust Adaptation of New Subjects in EEG Technology (TRANSIT-EEG) integrates meta-learned subject embeddings with synthetic signal generation using a denoising diffusion probabilistic model (IDPM) and applies LoRA-based finetuning for efficient adaptation. This approach improves SOGNN accuracy by 4.6% on SEED and achieves a 2% performance gain over the non-adaptive SOGAT baseline on PhyAAt, demonstrating its effectiveness in subject-specific adaptation.

Self-Supervised Enhancement for Multidimension Emotion Recognition using Graph Neural Networks (SS-EMERGE) employs a multidimensional encoder that explicitly models temporal, spectral, and spatial features using a stimulus-aware contrastive objective. The framework achieves 84.13% cross-subject accuracy on SEED and 82.75% on SEED-IV, surpassing the prior state-of-the-art SGMC model in both supervised and few-shot configurations. To support cross-dataset generalisation, **Unify-Emotion Self Supervised Learning (Unify-ESSL)** standardises EEG signals using a unified 14-channel layout and applies stratified batch sampling during contrastive pretraining. This approach increases macro-F1 on the DREAMER dataset from 86.05% to 91.23%, demonstrating its ability to learn robust and transferable EEG representations under heterogeneous recording conditions.

These findings provide a foundation for scalable EEG classification under real-world conditions. The proposed frameworks reduce dependency on labelled data, adapt to

unseen subjects, and generalise across devices and corpora. It enables the practical deployment of EEG-based systems in applications such as brain-computer interfaces, mental health diagnostics, and neuroadaptive computing.

7.1 Lessons Learnt

Based on the literature review, experimental results, and key findings, the following are the main insights gained from this research on few-shot EEG signal classification:

1. **Integrating multiple strategies enhances Few-Shot Learning for EEG:** Combining data augmentation, transfer learning, and self-supervised learning is necessary to address the challenges of low signal-to-noise ratio, high inter-subject variability, and limited labelled data. No single method suffices; their integration leads to more generalisable models.

2. **Optimised simple augmentations improve generalisation:** Basic transformations such as Gaussian noise injection, temporal jittering, and frequency perturbation significantly improve model robustness and generalisation when applied in a class- and data-aware manner.

3. **Subject- and Class-Invariant augmentation enhances transferability:** Augmentations that maintain consistency across subjects and emotion classes reduce subject-specific biases and support better generalisation to unseen individuals.

4. **Adapter-based finetuning prevents catastrophic forgetting in transfer learning:** While subject-specific data augmentation on EEG datasets improves cross-subject performance, full model finetuning can still lead to catastrophic forgetting of prelearned representations. Adapter-based methods such as LoRA enable efficient adaptation to new subjects while preserving generalisable knowledge from the source domain. This selective updating reduces the risk of overfitting and enhances transfer stability in few-shot settings.

5. **Cross-dataset training improves real-world adaptability** Training on multiple heterogeneous EEG datasets exposes models to varied acquisition protocols and

subject demographics. This diversity increases the model's robustness to new recording conditions, devices, and user populations.

6. **Multidimensional modelling boosts synthetic EEG quality** Capturing temporal, spectral, and spatial information during self-supervised pretraining leads to more realistic and informative synthetic signals. These multidomain features enhance physiological fidelity and improve downstream classification performance.

7.2 Ethical Implications and Responsible Use

Integrating few-shot learning techniques into EEG signal classification introduces significant opportunities for advancing clinical diagnostics, brain-computer interfaces, and cognitive monitoring. These techniques aim to build effective models from limited labelled data, making them highly attractive in neuroscience and healthcare settings, where data collection can be resource-intensive. However, their application in sensitive domains also raises important ethical questions. The inherent characteristics of EEG data, combined with the generalisation behaviour of few-shot models, demand careful attention to ethical implications. This section outlines key concerns, including bias, privacy, risk, and responsible use.

7.2.1 Bias

Few-shot learning models rely heavily on the support sets provided during training. Because these sets contain only a small number of examples, any imbalance or lack of diversity strongly influences the model's learned representations. EEG signals vary across individuals due to factors such as age, gender, neurological conditions, and socio-environmental background. When the support set fails to capture this diversity, the model learns biased patterns that do not generalise well to the broader population. In applications such as medical diagnosis or assistive technologies, this bias leads to unequal performance, potentially misdiagnosing some individuals or misinterpreting their neural activity. These issues raise serious concerns in high-stakes domains where fairness, accuracy, and reliability are critical.

7.2.2 Privacy Concerns

EEG data carries inherently sensitive information. It can reflect brain states related to specific tasks and underlying emotional responses, cognitive traits, or health conditions. Techniques commonly used in few-shot learning pipelines—such as Self-Supervised Learning (SSL), Transfer Learning (TL), and Data Augmentation (DA) raise additional privacy concerns when working with EEG signals. In SSL, models are often trained on large volumes of unlabeled EEG data to extract general-purpose features. These features may unintentionally capture subject-specific neural patterns, allowing for the inference of private information or the re-identification of individuals even when labels are absent. TL compounds this issue by transferring these learnt features to new tasks, where individualised information may persist in downstream applications, regardless of context or consent. DA further introduces privacy risks by creating multiple variations of EEG signals that still retain the core identity-related structure. It can lead to models that overfit specific subjects or become more sensitive to personal traits embedded in the data. Reusing publicly available EEG datasets—often collected under limited consent—intensifies these risks, as models trained across subjects may generalise patterns not initially intended for broad reuse.

7.2.3 Risk and Accountability

Real-world applications require reliable few-shot EEG classifiers, particularly when decisions have a significant impact on critical domains such as healthcare, education, and human-computer interaction. Errors in classification—such as false positives or false negatives—can lead to significant consequences. Because these systems train on limited datasets and operate under controlled settings, they often fail when encountering out-of-distribution or complex real-world inputs. Additionally, many models operate as black boxes with limited interpretability, raising concerns about their accountability and transparency. When these systems generate harmful or incorrect predictions, stakeholders often struggle to identify whether the responsibility lies with model developers, dataset curators, or deployment teams. This ambiguity obstructs practical impact evaluation and reduces user trust and adoption.

7.2.4 Responsible Use

Ethical concerns arise alongside technical challenges when deploying EEG-based few-shot learning systems. Organisations may deploy these systems in settings where individuals lack awareness or understanding of how the systems interpret or apply their neural data. Actors may use these models for surveillance, manipulation, or decision-making without obtaining meaningful consent or providing adequate oversight. Unequal access to EEG technologies and machine learning expertise creates disparities, allowing specific populations to benefit while excluding others. Without careful ethical guidance, the rapid expansion of neurotechnology can reinforce existing inequalities or enable its misuse for unintended purposes.

7.3 Limitation

Few-shot learning (FSL) has emerged as a robust framework to tackle data scarcity in EEG signal classification. By leveraging Data Augmentation (DA), Transfer learning (TL), and Self-Supervised Learning (SSL), these methods aim to learn generalised representations with minimal labelled data. FSL applications in the EEG domain face several critical limitations, outlined below:

1. **Lack of Evaluation Frameworks for Data Augmentation:** EEG-specific data augmentation techniques are often chosen heuristically, with no principled mechanism for evaluating their effectiveness beyond extrinsic metrics such as classification accuracy. It makes it difficult to determine whether improvements in model performance are due to meaningful representation learning or artifacts induced by overfitting from augmentation. The absence of intrinsic or domain-aware metrics limits the interpretability and trustworthiness of augmented EEG data.

2. **Limited Control Over Transferred Representations in Transfer Learning:** In transfer learning, finetunes models are pre-trained on large source datasets for downstream EEG tasks. However, this process offers limited control over the type of features transferred. EEG signals encode overlapping dimensions, such as sleep stages, emotional states, and cognitive workload, and current TL methods

cannot selectively transfer only the task-relevant characteristics. It can result in the unintended retention of irrelevant or confounding information, reducing model interpretability and performance in the target domain.

3. **Challenges in Dataset Unification Due to Channel Variability:** EEG datasets vary significantly in terms of electrode configurations, sampling rates, and hardware specifications. Methods like SSL or TL require consistent input structures to generalise across datasets or subjects. It necessitates preprocessing steps such as channel selection or interpolation to standardise electrode layouts. These steps may discard useful information or introduce distortions, especially when datasets do not share a common channel subset. As a result, model generalizability and reproducibility suffer due to inherent inconsistencies in EEG acquisition setups.

7.4 Future Work

Advancements in Few-Shot Learning (FSL) for EEG signal classification open several directions for further research. The following areas play a critical role in improving the robustness, generalizability, and real-world applicability of FSL models:

1. **Towards foundation models for EEG signal classification:** Building a foundation model for EEG requires unifying diverse datasets collected under varying conditions, such as differences in channel configurations, sampling frequencies, and experimental protocols. Future research should focus on creating scalable pretraining frameworks that can ingest heterogeneous EEG data and learn general-purpose representations transferable across tasks and subjects. Key challenges include harmonising electrode layouts, aligning temporal dynamics, and designing architectures that capture cross-domain invariances. Such foundation models would enable robust performance across datasets and support a wide range of downstream EEG applications with minimal supervision.

2. **Establishment of comprehensive evaluation suites for Self-Supervised Learning (SSL):** Current evaluations of SSL models in EEG applications

predominantly rely on classification accuracy, which may not fully capture the quality of learnt representations. Future work should aim to develop robust evaluation suites that include tasks such as classification, signal reconstruction, retrieval, and clustering. These additional metrics would provide a more holistic assessment of SSL models, ensuring they capture meaningful and generalisable features from EEG data.

3. **Enhancement of model robustness to channel placement Variability:** In practical scenarios, EEG channel placements can be inconsistent due to factors such as user movement or sensor misalignment, resulting in increased noise and reduced signal quality. Future research should focus on developing models that can detect and compensate for such variability. It includes designing architectures that can identify misaligned or noisy channels and adjust their processing accordingly, thereby enhancing the robustness and reliability of EEG-based FSL systems in real-world applications.

7.5 Social Impact and Applications

FSL has emerged as a pivotal advancement in EEG signal classification, offering profound social implications and a spectrum of real-world applications. By enabling models to learn effectively from minimal labelled data, FSL addresses critical challenges in EEG analysis, leading to significant societal benefits and practical implementations.

7.5.1 Social Impact

FSL significantly enhances the social impact of EEG applications by improving accessibility to neurological diagnostics, enabling personalised medicine, and advancing mental health monitoring. These advancements contribute to more equitable and effective healthcare solutions worldwide as follows:

1. **Privacy-preserving learning:** EEG signals encode personal information, including emotions, cognitive states, and potential health conditions. Transmitting

raw EEG data to centralised servers for training increases risks of identity inference, data leakage, and misuse. FSL minimises this risk by requiring only a few local examples for finetuning. With federated learning, FSL enables privacy-preserving training where data remains on-device, supporting decentralised mental health monitoring, neuroadaptive interfaces, and emotion-aware systems without compromising user privacy.

2. **Scalability with limited resources:** Many low- and middle-income regions lack access to EEG labs or large annotated datasets. FSL reduces dependence on centralised datasets by enabling local model adaptation with limited examples. It democratises EEG applications by lowering the entry barrier for clinics, schools, and researchers in under-resourced settings, promoting global equity in brain health technologies.

3. **Reduced environmental footprint:** Training deep neural networks from scratch requires substantial computational resources and energy, especially for large-scale EEG datasets. FSL follows a ”train once, adapt many” approach, allowing pre-trained models to adapt with minimal computational overhead. It leads to significantly fewer GPU cycles and lower carbon emissions. In wearable or edge-device scenarios, this energy-efficient learning is essential for sustainable deployment in healthcare, education, and consumer technologies.

4. **Enabling real-time on-device learning:** Wearable EEG devices in smart classrooms, neurofeedback systems, or assistive technologies must adapt to users in real time without relying on cloud dependency. FSL supports lightweight, fast personalisation on-device with minimal calibration. It enhances responsiveness and usability, reducing reliance on external infrastructure and facilitating deployment in remote areas with limited connectivity.

5. **Accelerating regulatory and clinical adoption:** The ability to finetune models with minimal data facilitates safer and more explainable adaptations in clinical trials or medical deployments. Regulators and clinicians can more easily validate models personalised for individual patients, making FSL-based systems more suitable for high-stakes medical use.

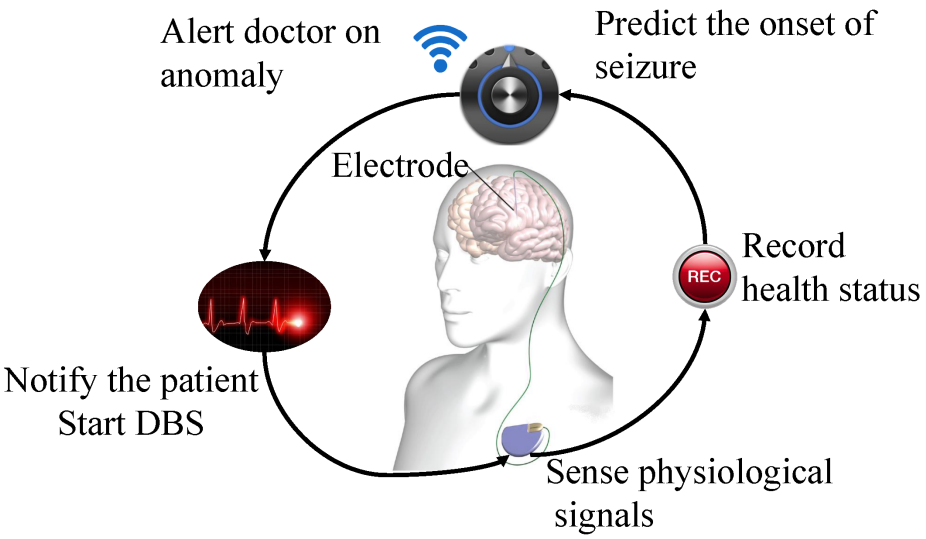


Figure 7.1: EEG-based seizure detection.

7.5.2 Applications

EEG signal classification has various applications, ranging from clinical diagnosis to brain-computer interface (BCI) applications. Specifically, FSL enables many real-world applications as follows:

1. **Automatic seizure detection and prediction:**

   Epileptic seizures often occur unpredictably, posing significant health risks. EEG is a non-invasive tool for detecting abnormal neural discharges preceding a seizure. However, conventional seizure detection systems require long-term monitoring and large quantities of labelled EEG data, which are impractical in clinical or wearable settings. Few-shot learning (FSL) provides a compelling solution by enabling rapid adaptation of seizure classifiers to new patients using only a few labelled examples. For instance, models trained on public datasets such as the CHB-MIT Scalp EEG corpus [219] or the TUH EEG Seizure Corpus [176] can be finetuned on individual-specific data using meta-learning or embedding-based approaches [342]. These methods significantly reduce patient burden while maintaining high prediction accuracy.

   Figure 7.1 illustrates a closed-loop seizure detection and intervention system. EEG electrodes placed on the scalp continuously sense physiological signals. These signals are recorded and analysed in real-time to predict the onset of

Figure 7.2: Adaptive Learning environment.

Figure 7.3: EEG-based adaptive learning with CogniFit [16].

seizures. Upon anomaly detection, the system can notify the patient and initiate therapeutic actions, such as Deep Brain Stimulation (DBS) [343], while simultaneously alerting medical professionals. Few-shot learning enables the personalisation of the detection module, allowing rapid model updates without requiring exhaustive data collection for each new patient. This FSL-powered system integrates seamlessly into wearable or implantable devices, enabling continuous, real-time seizure monitoring and prevention. It improves clinical responsiveness and patient quality of life.

2. **Adaptive Learning:**

   Students experience a wide range of cognitive and emotional states, such as boredom, excitement, and stress, during the learning process, which influences their engagement and retention. EEG-based classifiers can detect these states by

analysing brain activity related to attention, workload, and emotional arousal. However, large-scale, labelled EEG datasets are not feasible for every learner to collect due to concerns about privacy, fatigue, and variability. Few-shot learning enables personalised classification models to adapt quickly to new learners using only a few calibration samples. Once adopted, these models can monitor engagement in real time, enabling digital learning platforms to adjust content complexity or pacing accordingly. Figure 7.2 illustrates how EEG-based systems capture student states such as boredom, excitement, and stress during online education. These adaptive platforms enhance student outcomes by minimising disengagement and cognitive overload, especially in remote or self-paced environments.

In addition to classroom or online education settings, commercial platforms such as CogniFit have started integrating EEG-based adaptive systems to personalise learning experiences [16, 344]. Figure 7.3 shows a student interacting with such a system, where EEG sensors continuously monitor cognitive load and attention levels to adjust task difficulty and instructional timing in real time. These systems aim to enhance retention and engagement by matching learning demands to the learner's mental state. Despite their promise, personalised neuroadaptive platforms face practical barriers—chief among them the difficulty of collecting extensive labelled EEG data from each user. Variability across learners due to age, cognitive capacity, fatigue, or environmental factors further complicates model generalisation. Few-shot learning addresses these challenges by enabling classifiers to adapt using only a few user-specific examples. By finetuning emotion or attention detection models with minimal calibration data, few-shot learning makes EEG-based adaptive learning scalable and deployable in real-world educational environments. Future applications may embed these models into wearable EEG devices for seamless, real-time personalisation of mobile or at-home learning systems. It would enable students to benefit from continuously adjusted instructional pacing, automatic disengagement detection, and proactive content scaffolding without requiring expert configuration or model retraining.

3. **Personalized Music Recommendation:**

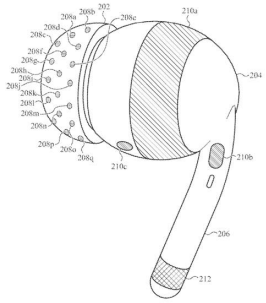


Figure 7.4: Apple's patent sketch of an EEG-enabled AirPod with integrated surface electrodes for real-time biosignal sensing [17].

Music has a profound influence on and reflects emotional and cognitive states. EEG-based classifiers can detect emotional responses such as stress, relaxation, excitement, or boredom, enabling systems to personalise music playback based on the user's real-time brain activity. However, emotion recognition from EEG data faces two challenges: subject-specific variability and the limited availability of labelled emotional EEG data for each user. Few-shot learning addresses these issues by enabling emotion classifiers to adapt to individual EEG patterns using only a few labelled examples. It can support the lightweight personalisation of music recommendation systems without requiring exhaustive calibration.

Recent developments highlight the commercial potential of this approach. Apple's patent (US20230225659A1) [17] proposes a biosignal-sensing AirPods device as shown in Figure 7.4 that captures EEG, EMG, and EOG signals through dynamic electrode selection on the ear surface. The system can infer a user's affective state in real time and adjust music playback accordingly, thereby enhancing emotional resonance and listener satisfaction. This integration of EEG-based emotion detection into consumer-grade audio hardware represents a significant step toward commercialising neuroadaptive music recommendation. By combining few-shot learning with such wearable biosignal sensors, future music platforms could deliver hyper-personalised listening experiences—automatically selecting or adapting songs to match or modulate the listener's mood. It enhances user satisfaction and opens up new avenues in affective computing and wellness-focused entertainment technology.

4. **Personalized Mental Health Monitoring:**

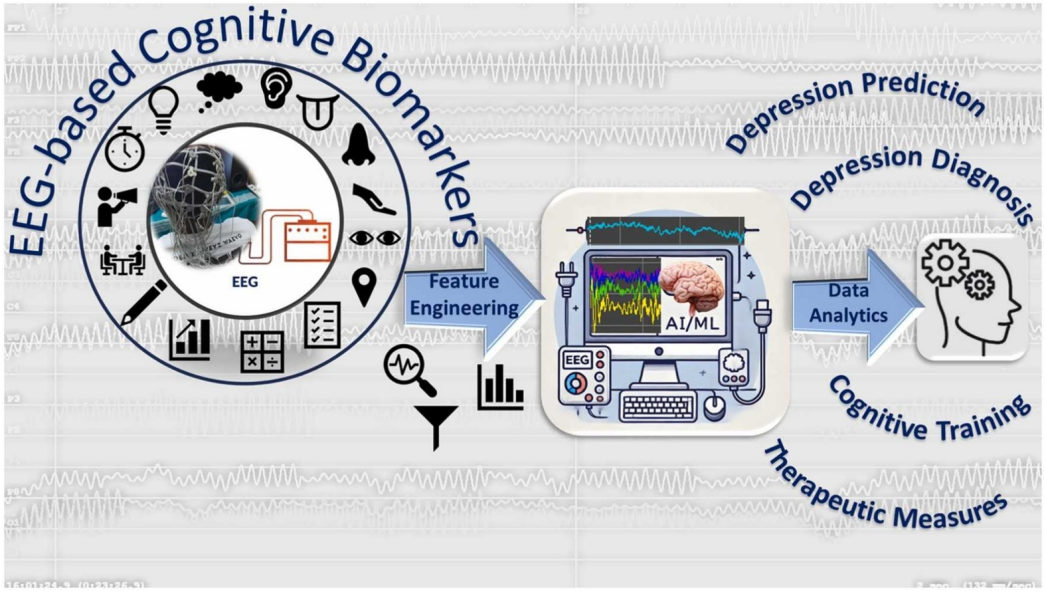


Figure 7.5: EEG-based mental health monitoring using cognitive biomarkers and AI/ML for depression diagnosis [18].

EEG-based mental health monitoring systems aim to detect cognitive and affective disorders such as depression, anxiety, and stress through the classification of brainwave patterns [345]. These disorders manifest with subtle, subject-specific neural signatures that vary across individuals and contexts. Traditional machine learning models struggle to generalise across such variability, especially when large-scale labelled EEG data is unavailable due to privacy and diagnostic limitations. Few-Shot Learning (FSL) enables rapid personalisation of emotion and disorder classifiers by finetuning on a limited set of labelled EEG samples per individual. It is particularly beneficial in mental health applications where patients may only provide minimal data during the initial assessment. By adapting pre-trained models to a user's baseline EEG profile, FSL enhances classification accuracy without the need for extensive per-subject data collection. Figure 7.5 illustrates an EEG-based mental health monitoring pipeline [18]. The system extracts cognitive biomarkers from EEG signals using feature engineering, followed by AI/ML-based classification. These insights inform diagnostic and therapeutic pipelines, enabling the prediction of depression, personalised cognitive training, and targeted interventions. FSL strengthens this loop by facilitating patient-specific calibration at minimal cost, supporting real-time, scalable mental health solutions for clinics, telehealth platforms, or wearable

neurotech.

5. **Cognitive State Monitoring:**

   Cognitive state monitoring is vital in education, workplace efficiency, and adaptive human-computer interaction. EEG signals are non-invasive in classifying task-specific cognitive states such as focused attention, cognitive workload, and mental fatigue. These states are crucial for understanding how individuals interact with activities such as reading, writing, or listening. However, EEG patterns linked to these cognitive processes vary significantly across users and recording sessions, making it difficult for traditional models to generalise effectively. Few-Shot Learning (FSL) addresses this challenge by enabling EEG classifiers to adapt to individual neural signatures using only a few labelled samples. This approach enables the rapid personalisation of models for cognitive state decoding without requiring extensive, subject-specific datasets. For example, in auditory attention tasks, EEG can reveal the focus of auditory attention [346], with datasets such as the KUL Dataset [347] supporting model development. In reading and writing contexts, frontal and temporal EEG activity captures mental effort and linguistic processing, as shown in the ZuCo dataset [348], which combines EEG with eye-tracking during natural reading tasks.

   FSL enables models trained on such datasets to adapt quickly to new users, supporting real-world applications that depend on fast, subject-specific tuning. Adaptive learning platforms can use these models to adjust content difficulty or pacing in response to real-time measures of mental workload. Similarly, smart notetaking systems can identify cognitive overload during writing tasks and suggest breaks or summaries to prevent fatigue. Attention-aware audiobook players can modulate playback speed or highlight important sections based on inferred attention levels from EEG signals in auditory learning or entertainment contexts. By supporting these applications with minimal calibration effort, FSL transforms EEG-based cognitive monitoring into a scalable and practical tool for personalised interaction across educational and cognitive support technologies.

References

[1] W.-L. Zheng and B.-L. Lu. "Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks". *IEEE Transactions on Autonomous Mental Development*, vol. 7, pp. 162–175, 2015.

[2] N. Bajaj, J. Requena Carrión, and F. Bellotti. "PhyAAt: Physiology of Auditory Attention to Speech Dataset". *arXiv Preprint arXiv:2004.00000*, 2020.

[3] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu. "EmotionMeter: A Multimodal Framework for Recognizing Human Emotions". *IEEE Transactions on Cybernetics*, vol. "49", pp. "1110–1122", "2018".

[4] OpenBCI. "EEG Electrode Cap Kit". URL https://shop.openbci.com/products/openbci-eeg-electrocap. Accessed: 2025-02-23.

[5] S. Koelstra, C. Muhl, and M. Soleymani. "DEAP: A Database for Emotion Analysis; Using Physiological Signals". *IEEE Transactions on Affective Computing*, vol. 3, pp. 18–31, 2011.

[6] V. J. Lawhern, A. J. Solon, and N. R. Waytowich. "EEGNet: A Compact Convolutional Neural Network for EEG-Based Brain – Computer Interfaces". *Journal of Neural Engineering*, vol. 15, pp. 056013, 2018.

[7] S. M. Omar et al. "Enhancing EEG Signals Classification Using LSTM-CNN Architecture". *Engineering Reports*, vol. 6, pp. e12827, 2024.

[8] J. Li, S. Li, and J. Pan. "Cross-Subject EEG Emotion Recognition with Self-Organized Graph Neural Network". *Frontiers in Neuroscience*, pp. 689, 2021.

[9] T. N. Kipf and M. Welling. "Semi-Supervised Classification with Graph Convolutional Networks". *arXiv Preprint arXiv:1609.02907*, 2017.

[10] P. Velić"ković et al. "Graph Attention Networks". In *Proc. International Conference on Learning Representations (ICLR)*, 2018.

[11] Y. Zhang, J. Chen, and J. H. Tan. "EEGformer: A Transformer – Based Brain Activity Classification Method Using EEG Signal". *Frontiers in Neuroscience*, vol. 17, pp. 1148855, 2023.

[12] Y. Song, T. Wang, and S. K. Mondal. "A Comprehensive Survey of Few-Shot Learning: Evolution, Applications, Challenges, and Opportunities". *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, pp. 10427–10449, 2023.

[13] E. Eldele et al. "Self-Supervised Learning for Label-Efficient Sleep Stage Classification: A Comprehensive Evaluation". *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 1333–1342, 2023.

[14] W.-L. Zheng, J.-C. Duan, and B.-L. Lu. "Self-Supervised Group Meiosis Contrastive Learning for EEG-Based Emotion Recognition". In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1335–1339, 2021.

[15] S. Katsigiannis and N. Ramzan. "DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals From Wireless Low-Cost Off-the-Shelf Devices". *IEEE Journal of Biomedical and Health Informatics*, vol. 22, pp. 98–107, 2018.

[16] CogniFit. "CogniFit's Cognitive General Assessment to Help Adaptive Learning Using EEG". https://www.newswire.com/news/cognifits-cognitive-general-assessment-to-help-daptive-learning-using-208 2019. Accessed on 11 May 2025.

[17] Apple Inclusive. "Biosignal Sensing Device Using Dynamic Selection of Electrodes", July 2023. https://patents.google.com/patent/US20230225659A1/en.

[18] K. Boby and S. Veerasingam. “Depression Diagnosis: EEG-based Cognitive Biomarkers and Machine Larning”. *Behavioural Brain Research*, vol. 478, pp. 115325, 2025. ISSN 0166-4328.

[19] M. P. Hosseini et al. “A Review on Machine Learning for EEG Signal Processing in Bioengineering”. *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 204–218, 2021.

[20] B. Shamlo et al. “The PREP Pipeline: Standardized Preprocessing for Large-Scale EEG Analysis”. *Frontiers in Neuroinformatics*, vol. 9, 2015.

[21] “Emotiv Epoc Neuroheadset”, 2022. Accessed: 2024-05-30.

[22] S. M. Imran, M. Saad, and M. Adeel. “Electroencephalography: Clinical Applications During the Perioperative Period”. *Frontiers in Medicine*, vol. 7, pp. 251, 2020.

[23] S. An, S. Kim, and P. Chikontwe. “Few-Shot Relation Learning with Attention for EEG-Based Motor Imagery Classification”. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 10933–10938, 2020.

[24] OpenMMLab. “MMFewShot: OpenMMLab FewShot Learning Toolbox and Benchmark”. https://github.com/open-mmlab/mmfewshot", 2021. Accessed: 2024-05-30.

[25] Y. You, X. Guo, and X. Zhong. “A Few-Shot Learning-Based EEG and Stage Transition Sequence Generator for Improving Sleep Staging Performance”. *Biomedicines*, vol. 10, pp. 3006, 2022.

[26] S. Lu. “Few-Shot Learning for EEG-Based Image Tasks”. https://github.com/SaMueL1u/EEG_HELPS_FEW_SHOT_LEARNING", 2024. Accessed: 2024-05-30.

[27] J. Devlin et al. “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding”. In *Proc. Conference of the North American Chapter of the Association for Computational Linguistics*, volume 1, pp. 4171–4186, 2019.

[28] J. Ho, A. Jain, and P. Abbeel. "Denoising Diffusion Probabilistic Models". In *Proc. Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.

[29] J. H. Medicine. "Electroencephalogram (EEG)". URL https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/electroencephalogram-eeg. Accessed: 2025-02-21.

[30] N. H. Service. "Electroencephalogram (EEG)". URL https://www.nhs.uk/conditions/electroencephalogram/. Accessed: 2025-02-21.

[31] J. Li, Y. Li, and M. Huang. "The Most Fundamental and Popular Literature on Functional Near-Infrared Spectroscopy: A Bibliometric Analysis of the Top 100 Most Cited Articles". *Frontiers in Neurology*, vol. 15, pp. 1388306, 2024.

[32] M. Hämäläinen et al. "Magnetoencephalography — Theory, Instrumentation, and Applications to Noninvasive Studies of the Working Human Brain". *Reviews of Modern Physics*, vol. 65, pp. 413–497, 1993.

[33] BioSemi. "ActiveTwo System", 2025. URL https://www.biosemi.com/Products_ActiveTwo.htm. Accessed: 2025-02-17.

[34] "ESI NeuroScan System". https://compumedicsneuroscan.com/applications/eeg/". Accessed: 2022-08-07.

[35] J. R. Wolpaw, N. Birbaumer, and D. J. McFarland. "Brain-Computer Interfaces for Communication and Control". *Clinical Neurophysiology*, vol. 113, pp. 767–791, 2002.

[36] BrainSigns. "Electroencephalography (EEG)", n.d. URL https://www.brainsigns.com/en/science/s2/technologies/eeg. Accessed: 2025-02-21.

[37] J. A. Urigüen and B. Garcia-Zapirain. "EEG Artifact Removal State-of-the-Art and Guidelines". *Journal of Neural Engineering*, vol. 12, pp. 031001, 2015.

[38] W. A. W Azlan and Y. F. Low. "Feature Extraction of Electroencephalogram (EEG) Signal-A Review". In *Proc. IEEE Conference on Biomedical Engineering and Sciences (IECBES)*, pp. 801–806, 2014.

[39] D. Kostas, S. Aroca-Ouellette, and F. Rudzicz. "BENDR: Using Transformers and a Contrastive Self-Supervised Learning Task to Learn From Massive Amounts of EEG Data". *Frontiers in Human Neuroscience*, vol. 15, pp. 653659, 2021.

[40] A. K. Singh and S. Krishnan. "Trends in EEG Signal Feature Extraction Applications". *Frontiers in Artificial Intelligence*, vol. 5, pp. 1072801, 2023.

[41] X. Chen, C. Li, and A. Liu. "Toward Open-World Electroencephalogram Decoding Via Deep Learning: A Comprehensive Survey". *IEEE Signal Processing Magazine*, vol. 39, pp. 117–134, 2021.

[42] Y. Roy et al. "Deep Learning-Based Electroencephalography Analysis: A Systematic Review". *Journal of Neural Engineering*, vol. 16, pp. 051001, 2019.

[43] K. M. Hossain et al. "Status of Deep Learning for EEG-Based Brain – Computer Interface Applications". *Frontiers in Computational Neuroscience*, vol. 16, 2023.

[44] F. Lotte et al. "A Review of Classification Algorithms for EEG-Based Brain – Computer Interfaces: A 10 Year Update". *Journal of Neural Engineering*, vol. 15, pp. 031005, 2018.

[45] B. Blankertz et al. "Optimizing Spatial Filters for Robust EEG Single-Trial Analysis". *IEEE Signal Processing Magazine*, vol. 25, pp. 41–56, 2008.

[46] H. Cecotti et al. "Convolutional Neural Networks for P300 Detection with Application to Brain-Computer Interfaces". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 433–445, 2010.

[47] D. Blankertz, Benjamin et al. "The Non-Invasive Berlin Brain – Computer Interface: Fast Acquisition of Effective Performance in Untrained Subjects". *NeuroImage*, vol. 37, pp. 539–550, 2007.

[48] T. Song, W. Zheng, and P. Song. "EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks". *IEEE Transactions on Affective Computing*, vol. 11, pp. 532–541, 2018.

[49] A. Vaswani et al. "Attention Is All You Need". In *Proc. ACM International Conference on Neural Information Processing Systems (NIPS)*, 2017.

[50] C. He et al. "Data Augmentation for Deep Neural Networks Model in EEG Classification Task: A Review". *Frontiers in Human Neuroscience*, pp. 747, 2021.

[51] J. Ferreira. "Cross-Subject Emotion Recognition From EEG Using Convolutional Neural Networks". https://github.com/javiferfer/cross-subject-eeg-emotion-recognition-through-nn". Accessed: 2025-04-23.

[52] R. Thiyagarajan, C. Curro, and S. Keene. "A Learned Embedding Space for EEG Signal Clustering". In *Proc. Signal Processing in Medicine and Biology Symposium (SPMB)*, pp. 1–4. IEEE, 2017.

[53] J. Yu, C. Li, and K. Lou. "Embedding Decomposition for Artifacts Removal in EEG Signals". *Journal of Neural Engineering*, 2022.

[54] Z. Yin, Y. Wang, and L. Liu. "Cross-Subject EEG Feature Selection for Emotion Recognition Using Transfer Recursive Feature Elimination". *Frontiers in Neurorobotics*, vol. 11, pp. 19, 2017.

[55] Z. Yin and J. Zhang. "Cross-Subject Recognition of Operator Functional States Via EEG and Switching Deep Belief Networks with Adaptive Weights". *Neurocomputing*, vol. 260, pp. 349–366, 2017.

[56] F. Fahimi, Z. Zhang, and W. B. Goh. "Inter-Subject Transfer Learning with an End-to-End Deep Convolutional Neural Network for EEG-Based BCI". *Journal of Neural Engineering*, vol. 16, pp. 026007, 2019.

[57] Z. Yin and J. Zhang. "Cross-Session Classification of Mental Workload Levels

Using EEG and an Adaptive Deep Learning Model". *Biomedical Signal Processing and Control*, vol. 33, pp. 30–47, 2017.

[58] G. Rafiei, Mohammad H. et al. "Self-Supervised Learning for Electroencephalography". *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, pp. 1–15, 2022.

[59] AliasVishnu. "EEG-SSL: Self-Supervised Learning for EEG". https://github.com/aliasvishnu/EEG-SSL". Accessed: 2025-04-23.

[60] C. Rommel et al. "Class-Wise Automatic Differentiable Data Augmentation for EEG Signals". *arXiv Preprint arXiv:2106.13695*, 2021.

[61] J. T. C. Schwabedal, J. C. Snyder, and A. Cakmak. "Addressing Class Imbalance in Classification Problems of Noisy Signals by Using Fourier Transform Surrogates". *arXiv Preprint arXiv:1806.08675*, 2019.

[62] C. Rommel. "EEG-AUGMENTATION-BENCHMARK-2022: Benchmark of Data Augmentations for EEG". https://github.com/eeg-augmentation-benchmark/eeg-augmentation-benchmark-2022". Accessed: 2023-03-23.

[63] A. D. Jacobsen. "EEG Data Augmentation: Project Work on TUH EEG Artifact Classification". https://github.com/AronDJacobsen/EEG_Data_Augmentation". Accessed: 2025-04-23.

[64] Y. Wang and Q. Yao. "Few-Shot Learning: A Survey". *arXiv Preprint arXiv:1904.05046*, 2019.

[65] G. Li, C. H. Lee, and J. J. Jung. "Deep Learning for EEG Data Analytics: A Survey". *Concurrency and Computation: Practice and Experience*, vol. 32, pp. e5199, 2020.

[66] Y. Li, G. Hu, and Y. Wang. "DADA: Differentiable Automatic Data Augmentation". In *Proc. Springer European Conference on Computer Vision*, pp. 580–595, 2020.

[67] W. Bomela, S. Wang, and C.-A. Chou. "Real-Time Inference and Detection of Disruptive EEG Networks for Epileptic Seizures". *Scientific Reports*, vol. 10, pp. 1–10, 2020.

[68] N. K. N. Aznan, A. Atapour-Abarghouei, and S. Bonner. "Simulating Brain Signals: Creating Synthetic EEG Data Via Neural-Based Generative Models for Improved Ssvep Classification". In *Proc. IEEE International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2019.

[69] IMICS-Lab. "EEG-Transfer-Learning: Source Code for Self-Supervised EEG Data Transfer Learning". https://github.com/imics-lab/eeg-transfer-learning". Accessed: 2025-04-23.

[70] S. V. Eeckt and H. Van Hamme. "Using Adapters to Overcome Catastrophic Forgetting in End-to-End Automatic Speech Recognition". In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2023.

[71] N. Mammone et al. "A Few-Shot Transfer Learning Approach for Motion Intention Decoding From Electroencephalographic Signals". *International Journal of Neural Systems*, vol. 34, pp. 2350068, 2024.

[72] A. Guillot, F. Sauvet, and E. H. During. "Dreem Open Datasets: Multi-Scored Sleep Datasets to Compare Human and Automated Sleep Staging". *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, pp. 1955–1965, 2020.

[73] S. Raschka. "Practical Tips for Finetuning LLMs Using LoRA (Low-Rank Adaptation)". https://magazine.sebastianraschka.com/p/practical-tips-for-finetuning-llms". Accessed: 2023-08-19.

[74] Z. Liu, Y. Guo, and K. Yu. "DiffVoice: Text-to-Speech with Latent Diffusion". In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

[75] Z. Zhang, Y. Liu, and S.-h. Zhong. "GANSER: A Self-Supervised Data Augmentation Framework for EEG-Based Emotion Recognition". *IEEE Transactions on Affective Computing*, vol. 14, pp. 2048–2063, 2023.

[76] R. Li, Y. Wang, and W.-L. Zheng. "A Multi-View Spectral-Spatial-Temporal Masked Autoencoder for Decoding Emotions with Self-Supervised Learning". In *Proc. ACM International Conference on Multimedia*, pp. 4002–4010, 2022.

[77] W. Li et al. "Self-Supervised Contrastive Learning for EEG-Based Cross-Subject Motor Imagery Recognition". *Journal of Neural Engineering*, vol. 21, 2024.

[78] Y. Zhang et al. "DMMR: Cross-Subject Domain Generalization for EEG-Based Emotion Recognition Via Denoising Mixed Mutual Reconstruction". In *Proc. AAAI Conference on Artificial Intelligence*, volume 37, pp. 1234–1242, 2023.

[79] T. Chen, S. Kornblith, and M. Norouzi. "A Simple Framework for Contrastive Learning of Visual Representations". In *Proc. International Conference on Machine Learning*, pp. 1597–1607, 2020.

[80] A. van den Oord, Y. Li, and O. Vinyals. "Representation Learning with Contrastive Predictive Coding". *arXiv Preprint arXiv:1807.03748*, 2019.

[81] R. Vigario. "Extraction of Ocular Artifacts From EEG Using Independent Component Analysis". *Electroencephalography and Clinical Neurophysiology*, vol. 103, pp. 395–404, 1997.

[82] P. He, G. Wilson, and C. Russell. "Removal of Ocular Artifacts From Electro-Encephalogram by Adaptive Filtering". *Medical and Biological Engineering and Computing*, vol. 42, pp. 407–412, 2004.

[83] A. Villena et al. "Preprocessing for Lessening the Unfluence of Eye Artifacts in EEG Analysis". *Applied Sciences*, vol. 9, pp. 1757, 2024.

[84] R. Abiri et al. "A Comprehensive Review of EEG-Based Brain – Computer Interface Paradigms". *Journal of Neural Engineering*, vol. 16, pp. 011001, 2019.

[85] G. Pfurtscheller and C. Neuper. “Motor Imagery and Direct Brain-Computer Communication”. In *Proc. IEEE*, volume 89, pp. 1123–1134, 2001.

[86] S. J. Luck. “Event-Related Potentials”. In *APA Handbook of Research Methods in Psychology, Vol. 1: Foundations, Planning, Measures, and Psychometrics*, pp. 523–546. 2012.

[87] S. Sutton, M. Braren, and J. Zubin. “Evoked-Potential Correlates of Stimulus Uncertainty”. *Science*, vol. 150, pp. 1187–1188, 1965.

[88] J. N. Acharya et al. “American Clinical Neurophysiology Society Guideline 3: A Proposal for Standard Montages to Be Used in Clinical EEG”. *Journal of Clinical Neurophysiology*, vol. 33, pp. 312–316, 2016.

[89] D. Goldman. “The Clinical Use of the ” Average ” Reference Electrode in Monopolar Recording”. *Electroencephalography and Clinical Neurophysiology*, vol. 2, pp. 209–212, 1950.

[90] F. F. Offner. “The EEG as Potential Mapping: the Value of the Average Monopolar Reference”. *Electroencephalography and Clinical Neurophysiology*, vol. 2, pp. 213–214, 1950.

[91] G. H. Klem et al. “The Ten Twenty Electrode System: International Federation of Societies for Electroencephalography and Clinical Neurophysiology”. *American Journal of EEG Technology*, vol. 1, pp. 13–19, 1961.

[92] J. N. Acharya et al. “American Clinical Neurophysiology Society Guideline 2: Guidelines for Standard Electrode Position Nomenclature”. *Journal of Clinical Neurophysiology*, vol. 33, pp. 4, 2016.

[93] M. R. Nuwer et al. “IFCN Standards for Digital Recordings of Clinical EEG”. *Electroencephalography and Clinical Neurophysiology*, vol. 106, pp. 259–261, 1998.

[94] A. C. N. Society. “Guideline 5: Guidelines for Standard Electrode Position Nomenclature”. *Journal of Clinical Neurophysiology*, vol. 23, pp. 107–110, 2006.

[95] M. Seeck et al. “The Standardized EEG Electrode Array of the IFCN”. *Clinical Neurophysiology*, vol. 128, pp. 2070–2077, 2017.

[96] R. Oostenveld and P. Praamstra. “The Five Percent Electrode System for High-Resolution EEG and ERP Measurements”. *Clinical Neurophysiology*, vol. 112, pp. 713–719, 2001.

[97] B. Hjorth. “An on-Line Transformation of EEG Scalp Potentials Into Orthogonal Source Derivations”. *Electroencephalography and Clinical Neurophysiology*, vol. 39, pp. 526–530, 1975.

[98] “Pediatric EEG”. https://www.learningeeg.com/pediatric". Accessed on 04 May 2025.

[99] A. B. Justesen et al. “Added Clinical Value of the Inferior Temporal EEG Electrode Chain”. *Clinical Neurophysiology*, pp. 291–295, 2018.

[100] C. A. Bosman et al. “Attentional Stimulus Selection Through Selective Synchronization Between Monkey Visual Areas”. *Neuron*, vol. 75, pp. 875–888, 2012.

[101] P. D. Emmady and A. C. Anilkumar. “EEG Abnormal Waveforms”. In *StatPearls [Internet]*. Treasure Island (FL), 2023. Updated 2023 May 1. Available from: https://www.ncbi.nlm.nih.gov/books/NBK557655/.

[102] “ESI NeuroScan System”. https://compumedicsneuroscan.com/applications/eeg/", 2025. Accessed: August 7, 2022.

[103] M. M. Bradley and P. J. Lang. “Measuring Emotion: the Self-Assessment Manikin and the Semantic Differential”. *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, pp. 49–59, 1994.

[104] A. Delorme and S. Makeig. “EEGLAB: an Open Source Toolbox for Analysis of Single-Trial EEG Dynamics Including Independent Component Analysis”. *Journal of Neuroscience Methods*, vol. 134, pp. 9–21, 2004.

[105] A. Widmann, E. Schröger, and B. Maess. "Digital Filter Design for Electrophysiological Data – A Practical Approach". *Journal of Neuroscience Methods*, vol. 250, pp. 34–46, 2015.

[106] A. de Cheveigné and I. Nelken. "Filters: When, Why, and How (Not) to Use Them". *Neuron*, vol. 102, pp. 280–293, 2019.

[107] P. Yang et al. "A Comparative Study of Average, Linked Mastoid, and REST References for ERP Components Acquired During fMRI". *Frontiers in Neuroscience*, vol. 11, 2017.

[108] N. E. Huang et al. "The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis". In *Proc. Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, volume 454, pp. 903–995, 1998.

[109] "Time Frequency Analysis of Bio Signals". https://bpa-au.vlabs.ac.in/exp/analysis-biosignals/theory.html". Accessed on 04 May 2025.

[110] K. Menon. "The Complete Guide to Skewness and Kurtosis". https://www.simplilearn.com/tutorials/statistics-tutorial/skewness-and-kurtosis". Accessed: 2025-02-21.

[111] Y. Wang et al. "A New Statistical Features Based Approach for Bearing Fault Diagnosis Using Vibration Signals". *Sensors*, vol. 22, pp. 2012, 2022.

[112] Y. Li et al. "An EEG-Based Attention Recognition Method: Fusion of Time Domain, Frequency Domain, and Non-Linear Dynamics Features". *Frontiers in Neuroscience*, 2023.

[113] A. Rajalakshmi and S. S. Sridhar. "A Study of Time Domain Features of EEG Signal Analysis". In *Proc. International Conference on Recent Trends in Data Science and Its Applications (ICRTDA)*, 2020.

[114] S. Liu et al. "EEG Power Spectral Analysis of Abnormal Cortical Activations During REM / NREM Sleep in Obstructive Sleep Apnea". *Frontiers in Neurology*, vol. 12, pp. 643855, 2021.

[115] R. Wang et al. "Power Spectral Density and Coherence Analysis of Alzheimer 's EEG". *Cognitive Neurodynamics*, vol. 9, pp. 291–304, 2015.

[116] E. Niedermeyer, D. Schomer, and F. da Silva. *"Event-Related Potentials: Methodology and Quantification"*. Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. Wolters Kluwer/Lippincott Williams & Wilkins Health, 2011. ISBN 9780781789424.

[117] M. Cohen. *"Analyzing Neural Time Series Data: Theory and Practice"*. Issues in Clinical and Cognitive Neuropsychology. MIT Press, 2014. ISBN 9780262319560.

[118] A. Subasi. "EEG Signal Classification Using Wavelet Feature Extraction and a Mixture of Expert Model". *Expert Systems with Applications*, vol. 32, pp. 1084–1093, 2007.

[119] Y. Wang, S. Gao, and X. Gao. "Common Spatial Pattern Method for Channel Selection in Motor Imagery Based Brain-Computer Interface". In *Proc. IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp. 5392–5395, 2005.

[120] J. Park and W. Chung. "Common Spatial Patterns Based on Generalized Norms". In *Proc. International Winter Workshop on Brain-Computer Interface (BCI)*, pp. 39–42, 2013.

[121] X. Yong, R. K. Ward, and G. E. Birch. "Robust Common Spatial Patterns for EEG Signal Preprocessing". In *Proc. IEEE International Conference of Engineering in Medicine and Biology Society*, pp. 2087–2090, 2008.

[122] Y. Zhang et al. "An Investigation of Deep Learning Models for EEG-Based Emotion Recognition". *Frontiers in Neuroscience*, vol. 14, pp. 622759, 2020.

[123] Y. LeCun et al. "Gradient-Based Learning Applied to Document Recognition". In *Proc. IEEE*, volume 86, pp. 2278–2324, 1998.

[124] Y. LeCun et al. "Backpropagation Applied to Handwritten Zip Code Recognition". In *Proc. Neural Computation*, volume 1, pp. 541–551, 1989.

[125] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. "Learning Internal Representations by Error Propagation". Technical Report ICS 8504, Institute for Cognitive Science, University of California, San Diego, California, 1985.

[126] S. Hochreiter and J. Schmidhuber. "Long Short-Term Memory". *Neural Computation*, vol. 9, pp. 1735–1780, 1997.

[127] F. Scarselli et al. "The Graph Neural Network Model". *IEEE Transactions on Neural Networks*, vol. 20, pp. 61–80, 2009.

[128] X. Lun et al. "A Simplified CNN Classification Method for MI-EEG Via the Electrode Pairs Signals". *Frontiers in Human Neuroscience*, vol. 14, 2020.

[129] C. Brunner et al. "BCI Competition IV – Dataset 2b". http://www.bbci.de/competition/iv/". Accessed: 2024-02-23.

[130] B. Blankertz, G. Curio, and K.-R. Mueller. "The BCI-NER Dataset", 2004. URL https://www.bbci.de/competition/ii/albany_desc/albany_desc.pdf. Accessed: May 3, 2025.

[131] A. Sors et al. "A Convolutional Neural Network for Sleep Stage Scoring From Raw Single-Channel EEG". *Biomedical Signal Processing and Control*, vol. 42, pp. 107–114, 2018.

[132] B. Kemp et al. "Analysis of a Sleep-Dependent Neuronal Feedback Loop: the Slow-Wave Microcontinuity of the EEG". *IEEE Transactions on Biomedical Engineering*, vol. 47, pp. 1185–1194, 2000.

[133] X. Zhang and Y. Wu. "Automated Multi-Model Deep Neural Network for Sleep Stage Scoring with Unfiltered Clinical Data". *Sleep and Breathing*, vol. 24, pp. 687–698, 2020.

[134] T. Zhu, W. Luo, and F. Yu. "Convolution and Attention-Based Neural Network for Automated Sleep Stage Classification". *International Journal of Environmental Research and Public Health*, vol. 17, pp. 4152, 2020.

[135] H. Zhao et al. “A Novel Sleep Staging Network Based on Multi-Scale Dual Attention”. *Biomedical Signal Processing and Control*, vol. 68, pp. 102581, 2021.

[136] C. Li et al. “A Deep Learning Method Approach for Sleep Stage Classification with EEG Spectrogram”. *International Journal of Environmental Research and Public Health*, vol. 19, pp. 6322, 2022.

[137] Y. Li et al. “A Simplified CNN Classification Method for MI-EEG Via the Electrode Pairs Signals”. *Frontiers in Human Neuroscience*, vol. 14, pp. 338, 2020.

[138] C. Zhang et al. “SaleNet: A Low-Power End-to-End CNN Accelerator for Sustained Attention Level Evaluation Using EEG”. In *Proc. IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 2304–2308, 2022.

[139] M. Alghanim et al. “A Hybrid Deep Neural Network Approach to Recognize Driving Fatigue Based on EEG Signals”. *Interntional Journal of Intelligent Systems*, vol. 2024, 2024. ISSN 0884-8173.

[140] W. Lu et al. “CIT-EmotionNet: Convolution Interactive Transformer Network for EEG Emotion Recognition”. *PeerJ Computer Science*, vol. 10, pp. e2610, 2024.

[141] A. Graves and J. Schmidhuber. “Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures”. In *Proc. Neural Networks*, volume 18, pp. 602–610, 2005.

[142] A. Supratak et al. “DeepSleepNet: A Model for Automatic Sleep Stage Scoring Based on Raw Single-Channel EEG”. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, pp. 1998–2008, 2017.

[143] Y. Sun, F. P.-W. Lo, and B. Lo. “EEG-Based User Identification System Using 1D-Convolutional Long Short-Term Memory Neural Networks”. *Expert Systems with Applications*, vol. 125, pp. 259–267, 2019.

[144] G. Xu et al. “A One-Dimensional CNN-LSTM Model for Epileptic Seizure Recognition Using EEG Signal Analysis”. *Frontiers in Neuroscience*, vol. 14, pp. 578126, 2020.

[145] T. Wilaiprasitporn et al. “Affective EEG-Based Person Identification Using the Deep Learning Approach”. In *Proc. International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2018.

[146] S. Saha and S. Fels. “A Multi-Modal Deep Learning Approach for Subject-Independent Emotion Recognition Using EEG and Eye Tracking Data”. *arXiv Preprint arXiv:1905.10988*, 2019.

[147] A. Narin. “Detection of Focal and Non-Focal Epileptic Seizure Using Continuous Wavelet Transform-Based Scalogram Images and Pre-Trained Deep Neural Networks”. *Innovation and Research in BioMedical Engineering*, vol. 43, pp. 22–31, 2022.

[148] C. Spampinato et al. “Deep Learning Human Mind for Automated Visual Classification”. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4503–4511. IEEE, 2017.

[149] Y. Li et al. “A Novel Bi-Hemispheric Discrepancy Model for EEG Emotion Recognition”. *IEEE Transactions on Cognitive and Developmental Systems*, vol. 13, pp. 354–367, 2021.

[150] M. K. Chowdary, J. Anitha, and D. J. Hemanth. “Emotion Recognition From EEG Signals Using Recurrent Neural Networks”. *Electronics*, vol. 11, 2022.

[151] S. Tripathi et al. “Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset”. *arXiv Preprint arXiv:1707.02011*, 2017.

[152] P. Gaur et al. “A Sliding Window Common Spatial Pattern for Enhancing Motor Imagery Classification in EEG-BCI”. *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021.

[153] S. Kumar, A. Sharma, and T. Tsunoda. "Brain Wave Classification Using Long Short-Term Memory Network Based OPTICAL Predictor". *Scientific Reports*, vol. 9, pp. 9153, 2019.

[154] Chetana R, A Shubha Rao, and Mahantesh K. "Application of Conv-1D and Bi-LSTM to Classify and Detect Epilepsy in EEG Data". *International Journal of Advanced Computer Science and Applications*, vol. 14, pp. 228–235, 2023.

[155] Y. Huang et al. "An Improved Model Using Convolutional Sliding Window-Attention Network for Motor Imagery EEG Classification". *Frontiers in Neuroscience*, vol. 17, pp. 1204385, 2023.

[156] A. Demir et al. "EEG-GNN: Graph Neural Networks for Classification of Electroencephalogram (EEG) Signals". In *Proc. IEEE International Conference of Engineering in Medicine & Biology Society (EMBS)*, pp. 1061–1067, 2021.

[157] M. Agarwal, E. Gupta, and S. Raghupathy. "EEG Error Related Potential Signal Using an Atari-Based Maze Game-BCI ErrP Dataset", 2024. URL https://dx.doi.org/10.21227/7rwy-9q83.

[158] K. Won et al. "EEG Dataset for RSVP and P300 Speller Brain-Computer Interfaces". *Scientific Data*, vol. 9, pp. 388, 2022.

[159] P. Zhong, D. Wang, and C. Miao. "EEG-Based Emotion Recognition Using Regularized Graph Neural Networks". *IEEE Transactions on Affective Computing*, vol. 13, pp. 1290–1301, 2022.

[160] Y. Zhang et al. "GNN4EEG: A Benchmark for Graph Neural Networks in EEG Classification". *arXiv Preprint arXiv:2301.12345*, 2023.

[161] A. Hajisafi et al. "NeuroGNN: Dynamic Graph Neural Networks for Precise Seizure Detection and Classification From EEG Data". *arXiv Preprint arXiv:2405.09568*, 2024.

[162] M. Jin et al. "PGCN: Pyramidal Graph Convolutional Network for EEG Emotion Recognition". *IEEE Transactions on Multimedia*, vol. 26, pp. 9070–9082, 2024.

[163] D. Klepl et al. “EEG-Based Graph Neural Network Classification of Alzheimer ’s Disease: an Empirical Evaluation of Functional Connectivity Methods”. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 2651–2660, 2022.

[164] Y. Hou et al. “GCNs-Net: A Graph Convolutional Neural Network Approach for Decoding Time-Resolved EEG Motor Imagery Signals”. *arXiv Preprint arXiv:2006.08924*, 2020.

[165] S. Tang et al. “Self-Supervised Graph Neural Networks for Improved Electroencephalographic Seizure Analysis”. *arXiv Preprint arXiv:2104.08336*, 2021.

[166] A. L. Goldberger et al. “PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals”. *Circulation*, vol. 101, pp. e215–e220, 2000.

[167] X. Chen et al. “DAMGCN: Dual Attention Mechanism Graph Convolutional Network for EEG-Based Emotion Recognition”. *IEEE Transactions on Affective Computing*, 2024.

[168] A. Kumar et al. “ST-GCN: Spatio-Temporal Graph Convolutional Network for Motor Imagery Classification”. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 1234–1243, 2023.

[169] A. Demir et al. “EEG-GAT: A Graph Attention Networks for Classification of Electroencephalogram (EEG) Signals”. In *Proc. IEEE International Conference of Engineering in Medicine & Biology Society (EMBS)*, pp. 30–35, 2022.

[170] H. Li et al. “Cross-Subject Emotion Recognition Using Deep Adaptation Networks”. In *Proc. Neural Information Processing*, pp. 403–413, 2018.

[171] X. Zhu et al. “Learning Generalizable EEG Representations with Hybrid Contrastive and Reconstruction Losses”. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, 2023.

[172] Y. Ding et al. “EEG-Deformer: A Dense Convolutional Transformer for Brain-Computer Interfaces”. *arXiv Preprint arXiv:2405.00719*, 2024.

[173] Y. Song et al. “Transformer-Based Spatial-Temporal Feature Learning for EEG Decoding”. *arXiv Preprint arXiv:2106.11170*, 2021.

[174] Y. Pan et al. “Dual Attentive Transformer in Long-Term Continuous EEG Emotion Analysis”. *arXiv Preprint arXiv:2401.01234*, 2024.

[175] M. Zhao and K. Iramina. “Channel-Wise Autoencoder Transformer for EEG Signal Classification”. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, pp. 123–134, 2024.

[176] I. Obeid and J. Picone. “The Temple University Hospital EEG Data Corpus”. *Frontiers in Neuroscience*, vol. 10, pp. 196, 2016.

[177] W. Yang and S. Modesitt. “Transfer Learning From Vision Transformers to EEG Classification Using Attention Pooling”. *IEEE Transactions on Biomedical Engineering*, vol. 70, pp. 45–56, 2023.

[178] M. J. Monesi et al. “ICASSP 2023 Auditory EEG Decoding Challenge”. https://exporl.github.io/auditory-eeg-challenge-2023/", 2023. Accessed: 2025-05-03.

[179] Y. Cui et al. “A Generative Pre-Trained Transformer for EEG Signal Generation and Classification”. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

[180] H. Wang et al. “Universal EEG Representation Learning Via Transformer Pretraining”. *arXiv Preprint arXiv:2403.01234*, 2024.

[181] T. Mitchell. *“Machine Learning”*. McGraw-Hill International Editions - Computer Science Series. McGraw-Hill Education, 1997. ISBN 9780070428072.

[182] M. Mohri, A. Rostamizadeh, and A. Talwalkar. *“Foundations of Machine Learning”*. Adaptive Computation and Machine Learning series. MIT Press, 2012. ISBN 9780262018258.

[183] F. P. Kalaganis, N. A. Laskaris, and E. Chatzilari. “A Data Augmentation Scheme for Geometric Deep Learning in Personalized Brain – Computer Interfaces”. *IEEE Access*, vol. 8, pp. 162218–162229, 2020.

[184] F. Zhuang, Z. Qi, and K. Duan. “A Comprehensive Survey on Transfer Learning”. In *Proc. IEEE*, volume 109, pp. 43–76, 2021.

[185] I. Kevin, K. Wang, and X. Zhou. “Federated Transfer Learning Based Cross-Domain Prediction for Smart Manufacturing”. *IEEE Transactions on Industrial Informatics*, vol. 18, pp. 4088–4096, 2021.

[186] S. Gidaris et al. “Boosting Few-Shot Visual Learning with Self-Supervision”. In *Proc. IEEE / CVF International Conference on Computer Vision*, pp. 8059–8068, 2019.

[187] Z. Chen, J. Ge, and H. Zhan. “Pareto Self-Supervised Training for Few-Shot Learning”. In *Proc. IEEE / CVF Conference on Computer Vision and Pattern Recognition*, pp. 13663–13672, 2021.

[188] X. Li, Q. Sun, and Y. Liu. “Learning to Self-Train for Semi-Supervised Few-Shot Classification”. *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[189] K. Simonyan and A. Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition”. *arXiv Preprint arXiv:1409.1556*, 2014.

[190] K. He, X. Zhang, and S. Ren. “Deep Residual Learning for Image Recognition”. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

[191] Andrzejak et al. “Indications of Nonlinear Deterministic and Finite-Dimensional Structures in Time Series of Brain Electrical Activity: Dependence on Recording Region and Brain State”. *Physical Review E*, vol. 64, pp. 061907, 2001.

[192] N. K. N. Aznan et al. “Leveraging Synthetic Subject Invariant EEG Signals for Zero Calibration BCI”. In *Proc. International Conference on Pattern Recognition (ICPR)*, pp. 10418–10425, 2021.

[193] R. P. Sharaj and J. Tzyy-Ping. "Modeling EEG Data Distribution with a Wasserstein Generative Adversarial Network to Predict RSVP Events". *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, pp. 1720–1730, 2020.

[194] O. "Özdenizci et al. "Adversarial Deep Learning in EEG Biometrics". *IEEE Signal Processing Letters*, vol. 26, pp. 710–714, 2019.

[195] X. Cui, V. Goel, and B. Kingsbury. "Data Augmentation for Deep Neural Network Acoustic Modeling". *IEEE / ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, pp. 1469–1477, 2015.

[196] M. Paschali et al. "Manifold Exploring Data Augmentation with Geometric Transformations for Increased Performance and Robustness". In A. C. S. Chung, J. C. Gee, P. A. Yushkevich, and S. Bao, editors, *Proc. Information Processing in Medical Imaging*, pp. 517–529, Cham, 2019.

[197] K., Mario Michael and K. Su Kyoung. "Rotational Data Augmentation for Electroencephalographic Data". In *Proc. IEEE International Conference of Engineering in Medicine and Biology Society (EMBS)*, pp. 471–474, 2017.

[198] O. Deiss, S. Biswal, and J. Jin. "HAMLET: Interpretable Human and Machine Co-LEarning Technique". *arXiv Preprint arXiv:1803.09702*, 2018.

[199] A. Saeed, D. Grangier, and O. Pietquin. "Learning From Heterogeneous EEG Signals with Differentiable Channel Reordering". In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1255–1259, 2021.

[200] M. M. Krell and S. K. Kim. "Rotational Data Augmentation for Electroencephalographic Data". In *Proc. IEEE International Conference of Engineering in Medicine & Biology Society (EMBS)*, volume 2017, pp. 471–474, Jul 2017.

[201] M. N. Mohsenvand, M. R. Izadi, and P. Maes. "Contrastive Representation Learning for Electroencephalogram Classification". In *Proc. Machine Learning for Health*, pp. 238–253, 2020.

[202] J. Y. Cheng et al. “Subject-Aware Contrastive Learning for Biosignals”. *Clinical Orthopaedics and Related Research*, vol. abs/2007.04871, 2020.

[203] T. H. Shovon, Z. Al Nazi, and S. Dash. “Classification of Motor Imagery EEG Signals with Multi-Input Convolutional Neural Network by Augmenting STFT”. In *Proc. International Conference on Advances in Electrical Engineering (ICAEE)*, pp. 398–403. IEEE, 2019.

[204] C. Brunner et al. “BCI Competition 2008–Graz Data Set A”. *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology*, vol. 16, pp. 1–6, 2008.

[205] D. Freer and G.-Z. Yang. “Data Augmentation for Self-Paced Motor Imagery Classification with C-LSTM”. *Journal of Neural Engineering*, vol. 17, pp. 016041, 2020.

[206] L. S. Mokatren, R. Ansari, and A. E. Cetin. “EEG Classification by Factoring in Sensor Configuration”. *arXiv Preprint arXiv:2001.00000*, 2020.

[207] F. Wang, S.-h. Zhong, and J. Peng. “Data Augmentation for EEG-Based Emotion Recognition with Deep Convolutional Neural Networks”. In *Proc. International Conference on Multimedia Modeling*, pp. 82–93. Springer, 2018.

[208] E. S. Salama, R. A. El-Khoribi, and M. E. Shoman. “EEG-Based Emotion Recognition Using 3D Convolutional Neural Networks”. *International Journal of Advanced Computer Science and Applications*, vol. 9, 2018.

[209] Y. Li, X.-R. Zhang, and B. Zhang. “A Channel-Projection Mixed-Scale Convolutional Neural Network for Motor Imagery EEG Decoding”. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, pp. 1170–1180, 2019.

[210] S. Kuanar, V. Athitsos, and N. Pradhan. “Cognitive Analysis of Working Memory Load From EEG, by a Deep Recurrent Neural Network”. In *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2576–2580. IEEE, 2018.

[211] C. Zhang, Y.-K. Kim, and A. Eskandarian. "EEG-Inception: an Accurate and Robust End-to-End Neural Network for EEG-Based Motor Imagery Classification". *arXiv Preprint arXiv:2101.00000*, 2021.

[212] I. Majidov and T. Whangbo. "Efficient Classification of Motor Imagery Electroencephalography Signals Using Deep Learning Methods". *Sensors*, vol. 19, pp. 1736, 2019.

[213] A. O'Shea, G. Lightbody, and G. Boylan. "Neonatal Seizure Detection Using Convolutional Neural Networks". In *Proc. International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6. IEEE, 2017.

[214] Z. Mousavi, T. Y. Rezaii, and S. Sheykhivand. "Deep Convolutional Neural Network for Classification of Sleep Stages From Single-Channel EEG Signals". *Journal of Neuroscience Methods*, vol. 324, pp. 108312, 2019.

[215] M. T. Avcu, Z. Zhang, and D. W. S. Chan. "Seizure Detection Using Least EEG Channels by Deep Convolutional Neural Network". In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1120–1124. IEEE, 2019.

[216] Z. Tayeb et al. "Validating Deep Neural Networks for Online Decoding of Motor Imagery Movements From EEG Signals". *Sensors*, vol. 19, pp. 210, 2019.

[217] K. M. Tsiouris, V. C. Pezoulas, and M. Zervakis. "A Long Short-Term Memory Deep Learning Network for the Prediction of Epileptic Seizures Using EEG Signals". *Computers in Biology and Medicine*, vol. 99, pp. 24–37, 2018.

[218] Z. Wei, J. Zou, and J. Zhang. "Automatic Epileptic EEG Detection Using Convolutional Neural Network with Improvements in Time-Domain". *Biomedical Signal Processing and Control*, vol. 53, pp. 101551, 2019.

[219] S. Chang and H. Jun. "Hybrid Deep-Learning Model to Recognise Emotional Responses of Users Towards Architectural Design Alternatives". *Journal of Asian Architecture and Building Engineering*, vol. 18, pp. 381–391, 2019.

[220] Y. Luo, L.-Z. Zhu, and Z.-Y. Wan. "Data Augmentation for Enhancing EEG-Based Emotion Recognition with Deep Generative Models". *Journal of Neural Engineering*, vol. 17, pp. 056021, 2020.

[221] X. Zhu, Y. Liu, and J. Li. "Emotion Classification with Data Augmentation Using Generative Adversarial Networks". In *Proc. Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 349–360. Springer, 2018.

[222] M. Kaper et al. "BCI Competition 2003-Data Set IIb: Support Vector Machines for the P300 Speller Paradigm". *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 1073–1076, 2004.

[223] K. Zhang, G. Xu, and Z. Han. "Data Augmentation for Motor Imagery Signal Classification Based on a Hybrid Neural Network". *Sensors*, vol. 20, pp. 4485, 2020.

[224] F. Fahimi, S. Dosen, and K. K. Ang. "Generative Adversarial Networks-Based Data Augmentation for Brain – Computer Interface". *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[225] T. Piplani, N. Merill, and J. Chuang. "Faking It, Making It: Fooling and Improving Brain-Based Authentication with Generative Adversarial Networks". In *Proc. IEEE International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pp. 1–7. IEEE, 2018.

[226] K. G. Hartmann, R. T. Schirrmeister, and T. Ball. "EEG-GAN: Generative Adversarial Networks for Electroencephalographic (EEG) Brain Signals". *arXiv Preprint arXiv:1806.01875*, 2018.

[227] Y. Luo and B.-L. Lu. "EEG Data Augmentation for Emotion Recognition Using a Conditional Wasserstein GAN". In *Proc. IEEE International Conference of Engineering in Medicine and Biology Society (EMBS)*, pp. 2535–2538. IEEE, 2018.

[228] D. Zhang, L. Yao, and K. Chen. "Ready for Use: Subject-Independent Movement Intention Recognition Via a Convolutional Attention Model". In *Proc. ACM*

International Conference on Information and Knowledge Management, pp. 1763–1766, 2018.

[229] Z. Zhang, F. Duan, and J. Sole-Casals. "A Novel Deep Learning Approach with Data Augmentation to Classify Motor Imagery Signals". *IEEE Access*, vol. 7, pp. 15945–15954, 2019.

[230] S. P., P. Rad, and J. Quarles. "Generating EEG Signals of an RSVP Experiment by a Class Conditioned Wasserstein Generative Adversarial Network". In *Proc. IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 1304–1310. IEEE, 2019.

[231] B. Arỳ et al. "Wavelet ELM-AE Based Data Augmentation and Deep Learning for Efficient Emotion Recognition Using EEG Recordings". *IEEE Access*, 2022.

[232] I. Goodfellow, J. Pouget-Abadie, and M. Mirza. "Generative Adversarial Nets". *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[233] M. Arjovsky, S. Chintala, and L. Bottou. "Wasserstein Generative Adversarial Networks". In *Proc. International Conference on Machine Learning*, pp. 214–223. PMLR, 2017.

[234] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. "Learning Internal Representations by Error Propagation". Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

[235] D. P. Kingma and M. Welling. "Auto-Encoding Variational Bayes". *arXiv Preprint arXiv:1312.6114*, 2013.

[236] D. Komolovaitė, R. Maskeliūnas, and R. Damaš"evič"ius. "Deep Convolutional Neural Network-Based Visual Stimuli Classification Using Electroencephalography Signals of Healthy and Alzheimer 's Disease Subjects". *Life*, vol. 12, pp. 374, 2022.

[237] O. Ronneberger, P. Fischer, and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In *Proc. International Conference on Medical*

Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234–241. Springer, 2015.

[238] S. Torma and L. Szegletes. "Brain Signal Generation and Data Augmentation with a Single-Step Diffusion Probabilistic Model". *arXiv Preprint arXiv:2302.03549*, 2023.

[239] J. Pascual and A. Gramfort. "Generating Realistic Neurophysiological Time Series with Denoising Diffusion Probabilistic Models". *Patterns*, vol. 5, pp. 100919, 2024.

[240] X. Lu et al. "EEG-Diffusion: A Diffusion Model for Electroencephalogram Data Generation". *arXiv Preprint arXiv:2307.10757*, 2023.

[241] J. Ye, Y. Zhang, and J. Wang. "TSDiffusion: Time-Series Diffusion Models with Dynamic Threshold Guidance". *arXiv Preprint arXiv:2306.01948*, 2023.

[242] W. Huang, L. Wang, and Z. Yan. "Classify Motor Imagery by a Novel CNN with Data Augmentation". In *Proc. IEEE International Conference on Engineering in Medicine & Biology Society (EMBS)*, pp. 192–195, 2020.

[243] S. Azadi, M. Fisher, and V. G. Kim. "Multi-Content GAN for Few-Shot Font Style Transfer". In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7564–7573, 2018.

[244] B. Liu, X. Wang, and M. Dixit. "Feature Space Transfer for Data Augmentation". In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9090–9098, 2018.

[245] Z. Luo, Y. Zou, and J. Hoffman. "Label Efficient Learning of Transferable Representations Across Domains and Tasks". *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[246] F. Yang, X. Zhao, and W. Jiang. "Multi-Method Fusion of Cross-Subject Emotion Recognition Based on High-Dimensional EEG Features". *Frontiers in Computational Neuroscience*, pp. 53, 2019.

[247] J. Li, S. Qiu, and C. Du. "Domain Adaptation for EEG Emotion Recognition Based on Latent Representation Similarity". *IEEE Transactions on Cognitive and Developmental Systems*, vol. 12, pp. 344–353, 2019.

[248] P. C. Ning, Run et al. "Cross-Subject EEG Emotion Recognition Using Domain Adaptive Few-Shot Learning Networks". In *Proc. IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1468–1472, 2021.

[249] Z. Lan et al. "Domain Adaptation Techniques for EEG-Based Emotion Recognition: a Comparative Study on Two Public Datasets". *IEEE Transactions on Cognitive and Developmental Systems*, vol. 11, pp. 85–94, 2018.

[250] K. Yan, L. Kou, and D. Zhang. "Learning Domain-Invariant Subspace Using Domain Features and Independence Maximization". *IEEE Transactions on Cybernetics*, vol. 48, pp. 288–299, 2017.

[251] S. J. Pan, I. W. Tsang, and J. T. Kwok. "Domain Adaptation Via Transfer Component Analysis". *IEEE Transactions on Neural Networks*, vol. 22, pp. 199–210, 2010.

[252] S. Siddharth, T.-P. Jung, and T. J. Sejnowski. "Utilizing Deep Learning Towards Multi-Modal Bio-Sensing and Vision-Based Affective Computing". *IEEE Transactions on Affective Computing*, 2019.

[253] M. Soleymani, J. Lichtenauer, and T. Pun. "A Multimodal Database for Affect Recognition and Implicit Tagging". *IEEE Transactions on Affective Computing*, vol. 3, pp. 42–55, 2011.

[254] S. Tammina. "Transfer Learning Using VGG-16 with Deep Convolutional Neural Network for Classifying Images". *International Journal of Scientific and Research Publications (IJSRP)*, vol. 9, pp. 143–150, 2019.

[255] Y. Cimtay and E. Ekmekcioglu. "Investigating the Use of Pretrained Convolutional Neural Network on Cross-Subject and Cross-Dataset EEG Emotion Recognition". *Sensors*, vol. 20, pp. 2034, 2020.

[256] C. Szegedy, S. Ioffe, and V. Vanhoucke. “Inception-V4, Inception-ResNet and the Impact of Residual Connections on Learning”. *Clinical Orthopaedics and Related Research*, vol. abs/1602.07261, 2016.

[257] H. He and D. Wu. “Different Set Domain Adaptation for Brain-Computer Interfaces: a Label Alignment Approach”. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, pp. 1091–1108, 2020.

[258] A. Mao, M. Mohri, and Y. Zhong. “Cross-Entropy Loss Functions: Theoretical Analysis and Applications”. In *Proc. International Conference on Machine Learning*, 2023.

[259] X. Zheng et al. “Task Transfer Learning for EEG Classification in Motor Imagery-Based BCI System”. *Computational and Mathematical Methods in Medicine*, vol. 2020, pp. 6056383, 2020.

[260] W. Ko, E. Jeon, and S. Jeong. “A Survey on Deep Learning-Based Short / Zero-Calibration Approaches for EEG-Based Brain-Computer Interfaces”. *Frontiers in Human Neuroscience*, vol. 15, pp. 643386, 2021.

[261] R. T. Schirrmeister et al. “Deep Learning with Convolutional Neural Networks for EEG Decoding and Visualization”. *Human Brain Mapping*, vol. 38, pp. 5391–5420, 2017.

[262] F. Andreotti, H. Phan, and N. Cooray. “Multichannel Sleep Stage Classification and Transfer Learning Using Convolutional Neural Networks”. In *Proc. IEEE International Conference of Engineering in Medicine and Biology Society (EMBS)*, pp. 171–174, 2018.

[263] D. Zhao, F. Tang, and B. Si. “Learning Joint Space – Time – Frequency Features for EEG Decoding on Small Labeled Data”. *Neural Networks*, vol. 114, pp. 67–77, 2019.

[264] S. Raghu, N. Sriraam, and Y. Temel. “EEG Based Multi-Class Seizure Type Classification Using Convolutional Neural Network and Transfer Learning”. *Neural Networks*, vol. 124, pp. 202–212, 2020.

[265] A. Finn, Chelsea et al. “Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks”. In *Proc. International Conference on Machine Learning*, pp. 1126–1135, 2017.

[266] A. Gretton et al. “A Kernel Two-Sample Test”. *The Journal of Machine Learning Research*, vol. 13, pp. 723–773, 2012.

[267] H. He and D. Wu. “Transfer Learning for Brain – Computer Interfaces: A Euclidean Space Data Alignment Approach”. *IEEE Transactions on Biomedical Engineering*, vol. 67, pp. 399–410, 2019.

[268] K. Zhang et al. “Application of Transfer Learning in EEG Decoding Based on Brain-Computer Interfaces: A Review”. *Sensors*, vol. 20, 2020.

[269] Y. Pan, Sinno Jialin et al. “A Survey on Transfer Learning”. *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, pp. 1345–1359, 2009.

[270] T. Duan et al. “Ultra Efficient Transfer Learning with Meta Update for Cross Subject EEG Classification”. *Clinical Orthopaedics and Related Research*, vol. abs/2003.06113, 2020.

[271] A. Pati, D. Mewada, and D. Samanta. “Meta-Learning for Subject Adaptation in Low-Data Environments for EEG-Based Motor Imagery Brain-Computer Interfaces”. *arXiv Preprint arXiv:2301.10058*, 2023.

[272] O. Li, Denghao et al. “Model-Agnostic Meta-Learning for EEG Motor Imagery Decoding in Brain-Computer-Interfacing”. In *Proc. IEEE / EMBS Neural Engineering (NER) Conference*, pp. 527–530, 2021.

[273] H. W. Ng and C. Guan. “Subject-Independent Meta-Learning Framework Towards Optimal Training of EEG-Based Classifiers”. *Neural Networks*, 2024.

[274] J.-W. Han et al. “META-EEG: Meta-Learning-Based Class-Relevant EEG Representation Learning for Zero-Calibration Brain – Computer Interfaces”. *Expert Systems with Applications*, vol. 238, pp. 121986, 2024.

[275] A. Nichol, J. Achiam, and J. Schulman. “On First-Order Meta-Learning Algorithms”. *arXiv Preprint arXiv:1803.02999*, 2018.

[276] W. Xiaoli and H. M. C. Rosa. "Does Meta-Learning Improve EEG Motor Imagery Classification?". In *Proc. IEEE International Conference of Engineering in Medicine & Biology Society (EMBS)*, pp. 4048–4051, 2022.

[277] T. Willmore. *"An Introduction to Differential Geometry"*. Oxford University Press, 1982. URL https://books.google.co.in/books?id=sfdGNAEACAAJ.

[278] Y. Jia, E. Shelhamer, and J. Donahue. "Caffe: Convolutional Architecture for Fast Feature Embedding". In *Proc. ACM International Conference on Multimedia*, pp. 675–678, 2014.

[279] Y.-J. Suh and B. H. Kim. "Riemannian Procrustes Analysis: Transfer Learning for Brain – Computer Interfaces". In *Proc. AAAI Conference on Artificial Intelligence*, pp. 854–862, 2021.

[280] R. Caruana. "Multitask Learning". *Machine Learning*, vol. 28, pp. 41–75, 1997.

[281] Y. Zhang and Q. Yang. "A Survey on Multi-Task Learning". *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, pp. 5586–5609, 2022.

[282] P. Autthasan, R. Chaisaen, and T. Sudhawiyangkul. "MIN2Net: End-to-End Multi-Task Learning for Subject-Independent Motor Imagery EEG Classification". *IEEE Transactions on Biomedical Engineering*, pp. 1–1, 2021.

[283] E. Jeon, W. Ko, and H.-I. Suk. "Domain Adaptation with Source Selection for Motor-Imagery Based BCI". In *Proc. International Winter Conference on Brain-Computer Interface (BCI)*, pp. 1–4, 2019.

[284] M. Pal, S. Bandyopadhyay, and S. Bhattacharyya. "A Many Objective Optimization Approach for Transfer Learning in EEG Classification". *arXiv Preprint arXiv:1904.04156*, 2019.

[285] I. Hossain, A. Khosravi, and I. Hettiarachchi. "Calibration Time Reduction Using Subjective Features Selection Based Transfer Learning for Multiclass BCI". In *Proc. IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 491–498, 2018.

[286] P. Gaur et al. “Tangent Space Features-Based Transfer Learning Classification Model for Two-Class Motor Imagery Brain – Computer Interface”. *International Journal of Neural Systems*, vol. 29, pp. 1950025, 2019.

[287] O. Yair, M. Ben-Chen, and R. Talmon. “Parallel Transport on the Cone Manifold of SPD Matrices for Domain Adaptation”. *IEEE Transactions on Signal Processing*, vol. 67, pp. 1797–1811, 2019.

[288] K.-J. Chiang, C.-S. Wei, and M. Nakanishi. “Cross-Subject Transfer Learning Improves the Practicality of Real-World Applications of Brain-Computer Interfaces”. In *Proc. IEEE International Conference on Neural Engineering (NER)*, pp. 424–427, 2019.

[289] G. Xu, X. Shen, and S. Chen. “A Deep Transfer Convolutional Neural Network Framework for EEG Signal Classification”. *IEEE Access*, vol. 7, pp. 112767–112776, 2019.

[290] J. Behncke, R. T. Schirrmeister, and M. Völker. “Cross-Paradigm Pretraining of Convolutional Networks Improves Intracranial EEG Decoding”. In *Proc. IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1046–1053. IEEE, 2018.

[291] A. K. Jaiswal, H. Liu, and P. Tiwari. “Towards Subject Agnostic Affective Emotion Recognition”. *CoRR: A Computing Research Repository*, vol. abs/2310.15189, 2023.

[292] A. M. Azab, L. Mihaylova, and K. K. Ang. “Weighted Transfer Learning for Improving Motor Imagery-Based Brain – Computer Interface”. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, pp. 1352–1359, 2019.

[293] J. Giles, K. K. Ang, and L. S. Mihaylova. “A Subject-to-Subject Transfer Learning Framework Based on Jensen-Shannon Divergence for Improving Brain-Computer Interface”. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3087–3091. IEEE, 2019.

[294] J. Adair, A. Brownlee, and F. Daolio. “Evolving Training Sets for Improved Transfer Learning in Brain Computer Interfaces”. In *Proc. International Workshop on Machine Learning, Optimization, and Big Data*, pp. 186–197, 2017.

[295] C.-S. Wei, M. Nakanishi, and K.-J. Chiang. “Exploring Human Variability in Steady-State Visual Evoked Potentials”. In *Proc. IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 474–479, 2018.

[296] I. Hossain, A. Khosravi, and I. Hettiarachchi. “Multiclass Informative Instance Transfer Learning Framework for Motor Imagery-Based Brain-Computer Interface”. *Computational Intelligence and Neuroscience*, vol. 2018, 2018.

[297] K. Zhang, G. Xu, and L. Chen. “Instance Transfer Subject-Dependent Strategy for Motor Imagery Signal Classification Using Deep Convolutional Neural Networks”. *Computational and Mathematical Methods in Medicine*, vol. 2020, 2020.

[298] O. ”Özdenizci et al. “Learning Invariant Representations From EEG Via Adversarial Inference”. *IEEE Access*, vol. 8, pp. 27074–27085, 2020.

[299] H. Zhao, Q. Zheng, and K. Ma. “Deep Representation-Based Domain Adaptation for Nonstationary EEG Classification”. *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, pp. 535–545, 2020.

[300] X. Tang and X. Zhang. “Conditional Adversarial Domain Adaptation Neural Network for Motor Imagery EEG Decoding”. *Entropy*, vol. 22, pp. 96, 2020.

[301] W. Wei, S. Qiu, and X. Ma. “A Transfer Learning Framework for RSVP-Based Brain Computer Interface”. In *Proc. IEEE International Conference of Engineering in Medicine & Biology Society (EMBS)*, pp. 2963–2968. IEEE, 2020.

[302] Y. Wang, S. Qiu, and X. Ma. “A Prototype-Based SPD Matrix Network for Domain Adaptation EEG Emotion Recognition”. *Pattern Recognition*, vol. 110, pp. 107626, 2021.

[303] L. Gao, Yunyuan et al. "Double Stage Transfer Learning for Brain – Computer Interfaces". *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 1128–1136, 2023.

[304] Z. Cai et al. "Multi-Layer Transfer Learning Algorithm Based on Improved Common Spatial Pattern for Brain – Computer Interfaces". *Journal of Neuroscience Methods*, vol. 415, pp. 110332, 2025.

[305] A. Smrdel. "Use of Common Spatial Patterns for Early Detection of Parkinson 's Disease". *Scientific Reports*, vol. 12, pp. 18793, 2022.

[306] S. Hendy and Y. Dar. "TL-PCA: Transfer Learning of Principal Component Analysis". *arXiv Preprint arXiv:2410.10805*, 2024.

[307] Jing, Longlong and Tian, Yingli. "Self-Supervised Visual Feature Learning with Deep Neural Networks: A Survey". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, pp. 4037–4058, 2020.

[308] T. Chen et al. "A Simple Framework for Contrastive Learning of Visual Representations". In *Proc. PMLR International Conference on Machine Learning (ICML)*, pp. 1597–1607, 2020.

[309] C. Wang et al. "BrainBERT: Self-Supervised Representation Learning for Intracranial Recordings". In *Proc. International Conference on Learning Representations (ICLR)*, 2023.

[310] S. Koelstra et al. "DEAP: A Database for Emotion Analysis; Using Physiological Signals". *IEEE Transactions on Affective Computing*, vol. 3, pp. 18–31, 2011.

[311] T. Zhang et al. "Cross-Domain EEG-Based Emotion Recognition Using Domain Adversarial Neural Network". *IEEE Transactions on Affective Computing*, 2023.

[312] X. Wu et al. "Multi-Source Domain Adaptation for EEG Emotion Recognition Based on Inter-Domain Sample Hybridization". *Frontiers in Human Neuroscience*, vol. Volume 18-2024, 2024.

[313] L. Xu, Li, et al. “Improved Regularization of Convolutional Neural Networks with Point Mask”. In *Proc. International Conference on Advances in Artificial Intelligence and Security (ICAIS)*, pp. 16–25, 2022.

[314] Z. Zhong et al. “Random Erasing Data Augmentation”. In *Proc. AAAI Conference on Artificial Intelligence*, volume 34, pp. 13001–13008, 2020.

[315] E. S. Suviseshamuthu et al. “EEG-Based Spectral Analysis Showing Brainwave Changes Related to Modulating Progressive Fatigue During a Prolonged Intermittent Motor Task”. *Frontiers in Human Neuroscience*, vol. 16, 2022.

[316] E. D. Cubuk et al. “Autoaugment: Learning Augmentation Strategies From Data”. In *Proc. IEEE / CVF Conference on Computer Vision and Pattern Recognition*, pp. 113–123, 2019.

[317] W. MacInnes et al. “BrainDecode: A Deep Learning Toolbox for Brain-Machine Interfaces”. https://github.com/braindecode/braindecode". Accessed: 2025-04-23.

[318] E. Jang, S. Gu, and B. Poole. “Categorical Reparameterization with Gumbel-Softmax”. In *Proc. International Conference on Learning Representations (ICLR)*, 2017.

[319] C. J. Maddison, A. Mnih, and Y. W. Teh. ”The Concrete Distribution: A Continuous Relaxation of Discrete Random Variables”. In *Proc. International Conference on Learning Representations (ICLR)*, 2017. URL https://arxiv.org/abs/1611.00712.

[320] E. J. Hu et al. “LoRA: Low-Rank Adaptation of Large Language Models”. *arXiv Preprint arXiv:2106.09685*, 2021.

[321] J. Fan et al. “EEG Data Augmentation: Towards Class Imbalance Problem in Sleep Staging Tasks”. *Journal of Neural Engineering*, vol. 17, pp. 056017, 2020.

[322] G. Tosato, C. M. Dalbagno, and F. Fumagalli. “EEG Synthetic Data Generation Using Probabilistic Diffusion Models”. *arXiv preprint arXiv:2303.06068*, 2023.

[323] T. Duan et al. “Ultra Efficient Transfer Learning with Meta Update for Cross Subject EEG Classification”. *arXiv Preprint arXiv:2003.06113*, 2021.

[324] T. Duan et al. “Meta Learn on Constrained Transfer Learning for Low Resource Cross Subject EEG Classification”. *IEEE Access*, vol. 8, pp. 224791–224802, 2020.

[325] T. N. Kipf and M. Welling. “Semi-Supervised Classification with Graph Convolutional Networks”. In *Proc. International Conference on Learning Representations (ICLR)*, 2017.

[326] P. Velič”ković et al. “Graph Attention Networks”. *arXiv Preprint arXiv:1710.10903*, 2017.

[327] J. Deng et al. “ArcFace: Additive Angular Margin Loss for Deep Face Recognition”. In *Proc. IEEE / CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4690–4699, 2019.

[328] S. Micheloyannis et al. “Small-World Networks and Disturbed Functional Connectivity in Schizophrenia”. *Schizophrenia Research*, vol. 87, pp. 60–66, 2006.

[329] A. Petrosian et al. “Recurrent Neural Network Based Prediction of Epileptic Seizures in Intra-and Extracranial EEG”. *Neurocomputing*, vol. 16, pp. 1120–1126, 2000.

[330] S. R. Sheikh et al. “Machine Learning Algorithm for Predicting Seizure Control After Temporal Lobe Resection Using Peri-Ictal Electroencephalography”. *Scientific Reports*, vol. 14, pp. 21771, 2024.

[331] J. Cai et al. “Application of Electroencephalography-Based Machine Learning in Emotion Recognition: A Review”. *Frontiers in Systems Neuroscience*, vol. 15, 2021.

[332] M. McCloskey and N. J. Cohen. “Catastrophic Interference in Connectionist Networks: the Sequential Learning Problem”. volume 24 of *Psychology of Learning and Motivation*, pp. 109–165. 1989.

[333] W. Ren et al. “Analyzing and Reducing Catastrophic Forgetting in Parameter Efficient Tuning”. *arXiv Preprint arXiv:2402.18865*, 2024.

[334] S. V. Eeckt et al. “Using Adapters to Overcome Catastrophic Forgetting in End-to-End Automatic Speech Recognition”. *arXiv Preprint arXiv:2203.16082*, 2022.

[335] C. M. Tyng et al. “The Influences of Emotion on Learning and Memory”. *Frontiers in Psychology, 8:1454*, 2017.

[336] P. J. Lang and M. M. Bradley. “Emotion and the Motivational Brain”. *Biological Psychology*, vol. 84, pp. 437–450, 2010.

[337] R. Duan, J. Zhu, and B. Lu. “Differential Entropy Feature for EEG-Based Emotion Classification”. In *Proc. IEEE / EMBS Conference on Neural Engineering (NER)*, pp. 81–84, 2013.

[338] L. Yang and J. Liu. “EEG-Based Emotion Recognition Using Temporal Convolutional Network”. In *Proc. IEEE International Conference on Data Driven Control and Learning Systems Conference (DDCLS)*, pp. 437–442. IEEE, 2019.

[339] Y. Zhou et al. “Enhancing Cross-Dataset EEG Emotion Recognition: A Novel Approach with Emotional EEG Style Transfer Network”. *IEEE Transactions on Affective Computing*, pp. 1–15, 2025.

[340] W. Cui et al. “Neuro-GPT: Towards A Foundation Model for EEG”. In *Proc. IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5, 2024.

[341] I. Jentzsch. “Independent Component Analysis Separates Sequence-Sensitive ERP Components”. *International Journal of Bifurcation and Chaos*, vol. 14, pp. 667–678, 2004.

[342] S. Khurana et al. “Neurology-as-a-Service for the Developing World”. In *Proc. International Conference on Knowledge Discovery & Data Mining, Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD)*, pp. 2665–2675, 2020.

[343] R. Fisher et al. “Electrical Stimulation of the Anterior Nucleus of Thalamus for Treatment of Refractory Epilepsy”. *Epilepsia*, vol. 51, pp. 899–908, 2010.

[344] Z. Mohamed et al. “Characterizing Focused Attention and Working Memory Using EEG”. *Sensors*, vol. 18, pp. 3743, 2018.

[345] S. Michelmann et al. “Leveraging Deep Learning for Robust EEG Analysis in Mental Health”. *Frontiers in Neuroinformatics*, vol. 18, pp. 1494970, 2024.

[346] M. J. Crosse, G. M. Di Liberto, A. Bednar, and E. C. Lalor. “The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli”. *Frontiers in Human Neuroscience*, vol. 10, pp. 604, 2016.

[347] N. Das and A. Bertrand. “Linear vs. Deep Learning Models for EEG based Auditory Attention Decoding”. In *European Signal Processing Conference, European Association for Signal Processing (EURASIP)*, pp. 1158–1162. IEEE, 2016.

[348] N. Hollenstein, C. Zhang, and N. Langer. “ZuCo, A Simultaneous EEG and Eye-Tracking Resource for Natural Sentence Reading”. *Scientific Data*, vol. 5, pp. 1–13, 2018.