Ain shams university
Faculty of computer and information sciences
Scientific Computing department

Skelly: a Mind Booster Game

BY

Ahmed Awad Sayed Hassan	[Scientific computing department]
Abdelrahman Emam Fadl Emam	[Scientific computing department]
Ezz Khalil Abdelhalim	[Scientific computing department]
Caroline Atta Samy	[Scientific computing department]
Mohammed Abdelraouf Ahmed Abdelshafy	[Scientific computing department]
Mohamed Mohamed Khalifa	[Scientific computing department]

Supervisors

Dr. Manal Mohsen Tantawy
Associate Professor in Scientific Computing Department
Faculty of computer and information sciences,
Ain shams university

TA. Aya Nasser
Scientific Computing Department
Faculty of computer and information sciences,
Ain shams university

Acknowledgments

We would like to express our profound gratitude to God for His blessings, guidance, and unwavering support throughout our journey. His presence has bestowed upon us the strength and resilience to navigate both the triumphs and challenges we encountered along the way.

Next, we extend our sincere appreciation to DR. Manal Mohsen Tantawy for her invaluable guidance, support, and mentorship. Her wealth of experience and unwavering dedication to our success have been instrumental in shaping our project and fostering our career development.

We would also like to express our heartfelt thanks to TA Aya Nasser for her exceptional assistance in the technical implementation of our project. Her expertise, patience, and unwavering support have played a significant role in expanding our skills and enhancing our overall experience.

Furthermore, we would like to extend our gratitude to our families and friends who stood by us during our moments of doubt and challenge. Their unwavering belief in us, words of encouragement, and unconditional support have provided us with the strength and motivation to overcome obstacles and pursue excellence.

Finally, we would like to acknowledge the contributions of all those who have directly or indirectly supported us in this endeavor. Their assistance, feedback, and encouragement have been invaluable, and we are deeply grateful for their presence in our lives.

Together, these individuals and forces have played a significant role in our growth, success, and personal development. We are truly grateful for the opportunities and experiences that have shaped us into the individuals we are today.

Abstract

This project introduces "Skelly," a brain-controlled game where players navigate through various maze levels using their brain signals (EEG). "Skelly" is designed to provide entertainment for individuals with disabilities who cannot use traditional controllers, support rehabilitation for injured individuals by engaging them in therapeutic tasks, and enhance the gaming experience by replacing conventional controls with Brain-Computer Interface (BCI) technology. This allows users to control the game character directly with their thoughts, offering a unique and inclusive way to interact with digital environments.

For this purpose, we utilized the BCI Competition IV dataset 2a, which includes data from nine subjects performing motor imagery tasks (imagining movements of the left hand, right hand, feet, and tongue). The dataset provided structured data cues that were essential for training our model.

We trained a deep learning model called ATCNet on this dataset. This model was chosen for its effectiveness in decoding EEG signals into control commands. It achieved an average accuracy of 87.16%, ensuring reliable interpretation of the brain's signals to navigate the game's mazes.

Overall, "Skelly" demonstrates the potential of integrating BCI technology into gaming, offering an innovative way for users to interact with digital environments using only their mind.

Table of Contents

Table of figures

Table of abbreviations

AAR	Automated Artifacts Removal
ATCNet	Attention Temporal Convolution Network
BCI	Brain Computer Interface
CNS	Central Nervous System
CSP	Common Spatial Pattern
CNN	Convolutional Neural Network
CAR	Common Average Reference
DL	Deep Learning
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
EEG	Electroencephalogram
EMD	Empirical Mode Decomposition
FC	Fully Connected
GRU	Gated Recurrent Unit
HHT	Hilbert-Huang Transform
ICA	Independent Component Analysis
LSTM	Long Short-Term Memory
MRI	Magnetic Resonance Imaging
MI	Motor Imagery
NN	Neural Network
PNS	Peripheral Nervous System
Power Spectral Density	Power Spectral Density
RNN	Recurrent Neural Network
WPD	Wavelet Packet Decomposition

List of Tables

Chapter 1
Introduction

1.1 Problem definition

Individuals with physical disabilities or injuries often face significant challenges in accessing and enjoying traditional video games and interactive digital content. Conventional gaming interfaces, such as controllers and keyboards, are not always accessible or feasible for these users, limiting their ability to participate in engaging digital experiences. Furthermore, current rehabilitation methods for injured individuals can lack the motivational and interactive elements that foster continuous engagement and effective recovery.

Skelly aims to address these issues by leveraging Brain-Computer Interface (BCI) technology to develop a game that can be controlled using brain signals (EEG). This approach not only provides an inclusive entertainment platform for individuals who are unable to use standard gaming controllers but also serves as a novel tool for rehabilitation. By enabling users to navigate through complex mazes using their thoughts, Skelly offers a stimulating and accessible way to interact with digital environments, thereby enhancing the quality of life and therapeutic outcomes for disabled and injured individuals.

1.2 Motivation

There are 1.3 billion people with disabilities in the world, which is about 16% of the total population. Some of them suffer from brain diseases such as ALS, which usually affects people between **40** and **70** years old, but can also occur at younger ages. ALS can be classified into two types: sporadic, which accounts for **90**-**95**% of all cases, and familial, which is inherited and affects a small number of people. ALS causes five deaths per **100,000** people aged **20** or older.

So this project aims to help people who have disabilities or brain injuries that impair their motor and communication abilities.It uses a type of BCI called MI

EEG signal, which can detect the user's intention from their brain activity when they imagine performing a movement. MI EEG signals can enable the users to control devices, communicate, or restore mobility of the existing MI EEG systems.

1.3 Objective

Design a game utilizing Brain-Computer Interface (BCI) technology to aid in the rehabilitation and enjoyment of individuals affected by brain injuries or disabilities.

1.4 Methodology

1.4.1 Overview

The methodology for developing "Skelly," a brain-controlled game, involves several key stages: data acquisition, data preprocessing, model training, game integration, and testing. The aim is to replace traditional gaming controls with a Brain-Computer Interface (BCI) to allow users, especially those with disabilities, to control the game character using their brain signals.

1.4.2 Data Acquisition

The BCI Competition IV dataset 2a was selected as the optimal dataset for this project. This dataset contains EEG data recorded from 9 subjects, each performing four motor imagery (MI) tasks: right hand, left hand, both feet, and tongue. The dataset includes two sessions per subject recorded on different days, each comprising six runs of 48 trials (12 for each MI class), totaling 288 trials per session. This dataset is ideal due to its:

- **Cue-based (synchronous) nature:** Subjects start the MI task after a specific cue.
- **Subject-dependent structure:** Training and testing are performed on data specific to each subject.

1.4.3 Data Pre-processing

Normalization: Z-score normalization is applied to standardize the EEG data, ensuring that each feature has a mean of 0 and a standard deviation of 1. This helps in stabilizing the learning process and improving model performance.

Filtering: A band-pass filter is used to retain relevant EEG signal frequencies (e.g., 0.5–35 Hz) while removing noise and irrelevant frequencies.
Segmentation: Each trial in the dataset is segmented into epochs from 0 to 6 seconds based on the event markers (cue presentation and fixation cross appearance).

1.4.4 Model Training

The ATCNet (Attention Temporal Convolutional Network) is chosen for its ability to effectively handle time-series EEG data and its robust performance in BCI applications.

The model is trained on the preprocessed EEG data from the BCI Competition IV dataset 2a. The input consists of epochs of EEG data, and the output is the classification of the motor imagery task.

1.4.5 Game Design

Maze Levels: "Skelly" consists of multiple maze levels where the player navigates the character through obstacles and paths using their brain signals.
User Interface: The game interface is designed to provide clear visual feedback and an engaging user experience.

1.4.6 Game Integration

Signal Mapping

- **Mapping Strategy:** Each motor imagery class (right hand, left hand, both feet, and tongue) is mapped to specific control signals for game navigation. For instance, right hand imagery moves the character right, left hand imagery moves left, etc.
- **Real-Time Processing:** EEG signals are continuously processed in real-time to translate user intentions into game actions.

1.5 Time Plan

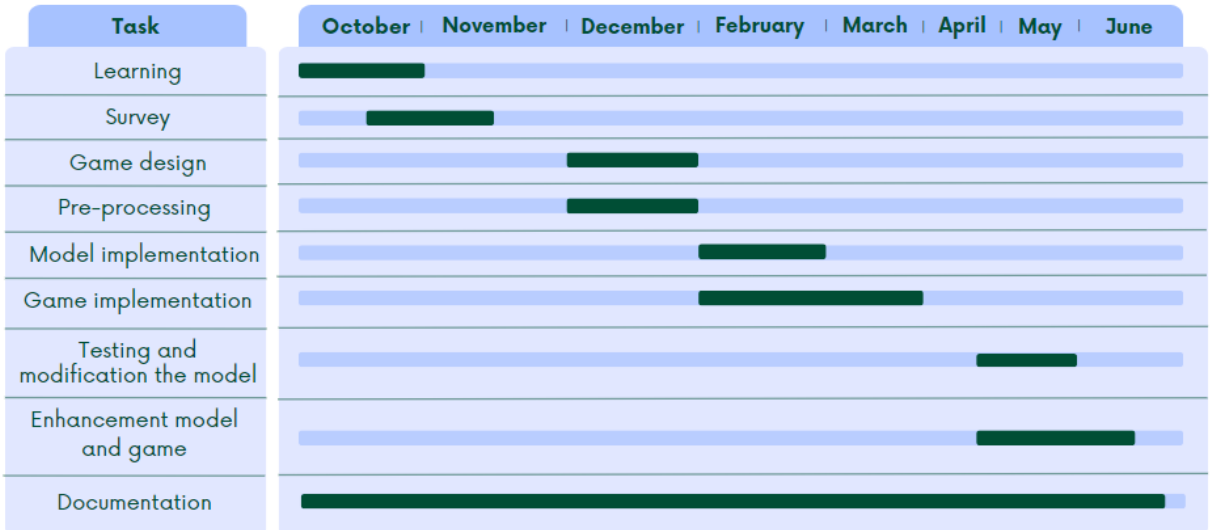

Figure 1.1 Time Plan

1.6 Thesis Outline

- Chapter 2 is background for the project which describes some concepts that should be known like: brain anatomy, brain-computer interface(BCI) and Electroencephalography (EEG).

- Chapter 3 deals with the general architecture of BCI applications and the related works in that field

- Chapter 4 describes our system architecture (the pre-processing and the classification models).

- chapter 5 deals with the dataset and its description and the results that we get by using the mentioned models in the previous section.

- chapter 6 deals with the User-Interface which the game (Skelly) ,it's Flow Board, menus and levels.

Chapter 2
Background

Since the mid-20th century, the study of the brain's complexities has become a focal point of scientific inquiry, driven by advancements in technology and theory. Initially focused on understanding neural architecture and basic cognitive processes, brain research has expanded significantly. Recent decades have seen the emergence of brain science as a leading interdisciplinary field, integrating neuroscience, psychology, biology, and computer science. Cutting-edge imaging techniques like functional MRI and PET scans have transformed our ability to study neural connectivity and brain plasticity, while computational modeling and artificial intelligence have provided new insights into neural computation and cognitive functions.

This interdisciplinary approach has yielded profound insights into neurological disorders, cognitive development, and the brain's ability to adapt through neuroplasticity and neurogenesis. Beyond academia, research in brain science influences clinical practices, guiding the development of treatments for neurological and psychiatric conditions. Moreover, it plays a pivotal role in advancing artificial intelligence, fostering innovations that replicate human cognitive processes. As brain science continues to evolve, it remains at the forefront of scientific exploration, offering promising avenues to unravel the complexities of human cognition, consciousness, and behavior. This chapter aims to provide a comprehensive overview of current research, highlighting key advancements, methodologies, and future directions in understanding one of the most intricate organs in the human body.

2.1 Brain anatomy

The human nervous system, comprising billions of neurons, processes sensory information via peripheral nerves and initiates responses through muscles and glands for voluntary actions. It consists of the central nervous system (CNS) — the brain and spinal cord — which coordinates bodily functions like movement, cognition, memory, and speech by interpreting sensory inputs and sending signals throughout the body. The peripheral nervous system (PNS) includes sensory, motor, cranial, and spinal nerves, with the autonomic nervous system regulating vital functions such as digestion, breathing, and heart rate to ensure optimal body function without conscious effort.

The human brain is anatomically divided into four lobes: the frontal, parietal, occipital, and temporal lobes. These lobes each play critical roles in various functions:

- **Frontal Lobe**: Responsible for problem-solving, emotional traits, reasoning, and personality.
- **Parietal lobe:** Involved in sensation, language processing, intelligence, and reading.
- **Temporal lobe:** Handles functions related to hearing, memory formation, understanding language, and regulating behavior.
- **Occipital lobe:** Primarily processes vision, including color perception and visual interpretation.

In addition to these lobes, several cortical areas are essential in brain-computer interface (BCI) research, including motor areas responsible for movement control, the somatosensory cortex for processing touch and proprioception, the posterior parietal cortex involved in spatial awareness, and the visual cortex responsible for visual processing. These areas collectively contribute to our understanding of how the brain processes information and controls bodily functions.

2.2 Nature form of communication

Natural communication and control processes involve the coordination of peripheral nerves and muscles, beginning with the user's intention. This intention initiates a complex sequence where specific brain regions are activated, sending signals via the peripheral nervous system to the relevant muscles. These muscles then execute the movements necessary for the intended communication or control task. This entire process is commonly referred to as motor output or **efferent output**.

Efferent output refers to the transmission of impulses from the central nervous system to the periphery, specifically to muscles or other effectors. In contrast, **afferent communication** describes the flow of information from sensory receptors to the central nervous system. This bidirectional communication—efferent and afferent—is fundamental to all forms of motor control and sensory feedback.

In motor control, the motor pathway is crucial for executing movements, while the sensory pathway plays a vital role in learning motor skills and performing intricate tasks, such as playing musical instruments or typing. Damage to the brain's communication pathways to muscles due to accidents can disrupt these signals, leading to impaired coordination and loss of motor control.

2.3 Electroencephalography (EEG)

EEG is a method used to record the spontaneous electrical activity of the brain. It predominantly captures electrical signals from the cerebral cortex. While EEG offers good temporal resolution (measured in milliseconds), its spatial resolution is generally poorer, typically in the square centimeter range.

It can be decomposed to 5 sub-bands:

- **Delta waves:** have a frequency range of 0.5-4 Hz and are detectable in babies and during slow-wave sleep in adults
- **Theta waves:** with a frequency of 4-8 Hz, are associated with drowsiness or “idling” in children and adults.
- **Alpha waves:** are electrical fluctuations in the range of 8-13 Hz and can be measured in EEG from the occipital region in awake persons when they are relaxed, or their eyes are closed
- **Beta waves:** with a frequency of 13-30 Hz, are detectable over the parietal and frontal lobes in a person who is alert and actively concentrating.

- **Gamma waves:** with a frequency of 30-100 Hz or more, have been reported in tasks involving short-term memory and language processing.

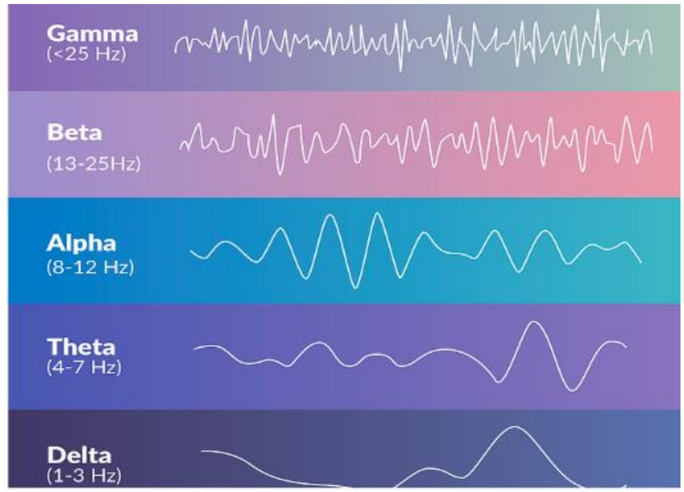

Figure 2.1 EEG Sub-Bands

2.4 Brain Computer Interface (BCI)*[1]*

It's an artificial system designed to bypass the body's normal efferent pathways, which include neuromuscular output channels.

Rather than relying on peripheral nerves and muscles, a BCI directly detects and measures brain activity associated with the user's intentions. This recorded brain activity is then translated into control signals for various BCI applications.

This translation process involves sophisticated signal processing and pattern recognition techniques, typically performed by a computer. Because the signals originate directly from the brain rather than from peripheral systems or muscles, the technology is referred to as a Brain-Computer Interface (BCI).

BCI has different perspectives:

1. **Mode of operation:**
    a. **Synchronous:** cue-based systems, which are restricted to a given time interval. In this scheme, the user receives a cue from the system and the system records the neural response to this cue. The features extracted from the signal are processed sequentially. This system permits the processing of another set of features only after the completion of the first instruction.
    b. **Asynchronous:** the user can produce any mental state whenever he/she wants without cue; therefore, it is called self-based BCI. In this scheme, features are extracted and processed continuously.
2. Measure brain activity:
    a. Invasive: refers to interfaces that require surgical implantation of devices or electrodes directly into the brain to enable communication between the brain and external devices or computers.
    b. **Non-invasive:** do not require surgical procedures and instead use external sensors to detect brain activity. Common non-invasive BCI methods include electroencephalography (EEG)
3. Considered mental strategy:
    a. **Stimulus-Evoked (Evoked Potential)**: Stimulus-evoked responses in EEG recordings include the P300 signal, which appears approximately 300 milliseconds after a rare or unexpected stimulus. This event-related potential is prominent over the parietal lobe and reflects brain activity linked to stimulus processing. Another type, the steady-state visually evoked potential (SSVEP), occurs when visual cortex neurons respond to a flickering stimulus like a checkerboard. EEG captures SSVEP as peaks in the power spectrum at the stimulus frequency and its harmonics, enabling Brain-Computer Interfaces (BCIs) to decode user choices based on these frequency patterns.

b. **Conditioned Responses**: In pioneering brain-computer interface research, Eberhard demonstrated how a single neuron in the primate motor cortex could be conditioned to control an analog meter's needle through its firing rate, leading to rewards when the needle reached a set threshold. This operant conditioning process highlighted the direct link between neural activity and external actions. Expanding on this principle, experiments with human subjects showed that after training sessions, individuals could modulate power in specific EEG frequency bands to manipulate a cursor on a computer screen, aiming to hit designated targets. These studies illustrate how conditioning techniques can leverage neural signals for precise control in brain-computer interface applications, bridging neuroscience with practical technological advancements.

c. **Motor imagery *[2]***: refers to the mental simulation or visualization of specific motor actions without physically executing them, it's widely used for brain-computer interfacing in humans is neural activity produced when a subject voluntarily imagines making a movement. Imagining a movement typically produces neural activity that is spatiotemporally like the activity generated during actual movement, but smaller in magnitude. Motor imagery is capable of discerning different classes of imagined movements, ranging from simple tasks like left-hand, right-hand, or foot movements to more intricate actions such as tongue movements, depending on the task complexity and the accuracy of the detection system employed. Classifiers are instrumental in distinguishing between various types of imagined movements, enabling each specific mental activity to be associated with a distinct control signal. For instance, this could involve mapping different imagined actions to precise commands, such as moving a cursor upwards or performing other desired tasks in brain-computer interface applications.

Motor imagery applications:

- Rehabilitation and Restoration
- BCI Speller
- Brain to Vehicle
- Image Reconstruction from Brain Data
- brain-to-text communication via handwriting
- Speech from Brain signals
- **Games and entertainment**

Figure 2.2 speech from brain signals

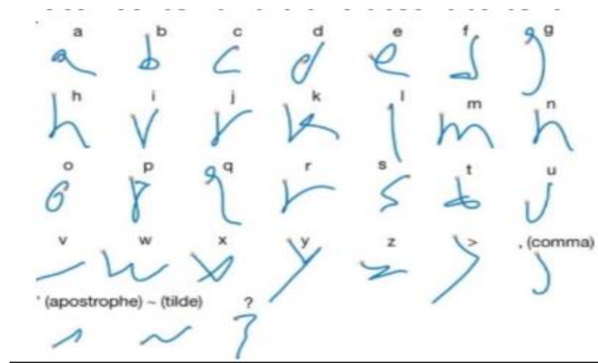

Figure 2.3 brain to text communication

Figure 2.4 games and entertainment

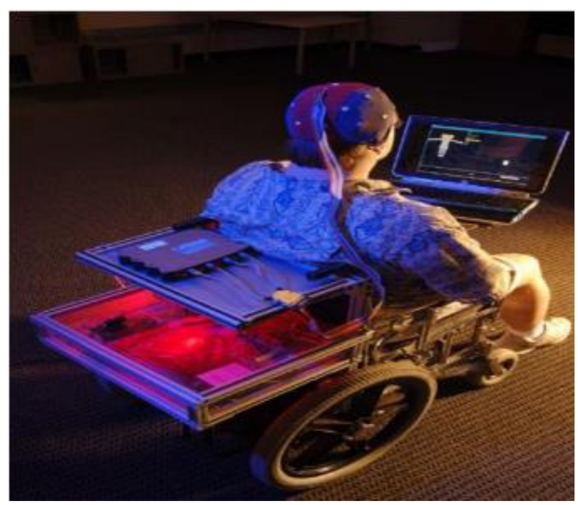

Figure 2.5 Rehabilitation and Restoration

Chapter 3

Literature Review

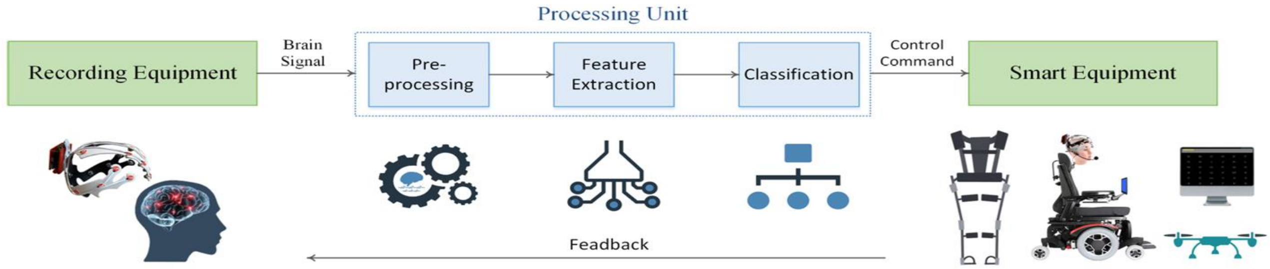

Figure 3.1 General architecture

Figure 3.1 describes the general architecture of the BCI applications which consists of: Signal recording, processing unit (pre-processing, feature extraction and classification) and the smart equipment or application that will be controlled using the classified signals

3.1 General Architecture

3.1.1 EEG signals recording

Mental activity in the central nervous system (CNS) produces continuous patterns that vary over time, known as neural oscillations or brain waves. During a mental activity, neurons in the brain communicate with each other, causing a change in the electrical current and blood flow in the brain, which can be measured using various techniques. Brain current can be measured using electrical and magnetic fields, while cerebral blood flow can be measured using optical and magnetic properties. EEG is a technique for recording electrical brain activities using a non-invasive electrophysiological method that measures voltage fluctuations induced by the ionic current within brain neurons. Because the ionic current produced inside the brain is recorded on the scalp, obstructions (such as the skull) significantly reduce the quality of the signal. The recorded EEG signal is only about 5% of the actual brain signal. Therefore, to improve signal quality, raw EEG signals are normally preprocessed before feature extraction and classification. EEG signals typically consist of a 2D matrix of real values (channel and time) that represent task-related brain potentials. These two dimensions represent the spatial and temporal information of the EEG signal. The spatial resolution refers to the spatial positions of the electrodes on the scalp (number of electrodes), while the temporal resolution represents the number of time points per second (i.e., sampling rate). The spatial resolution ranges from 1 to 256 electrodes; however, for research or clinical purposes, a range of 21 to 64 electrodes is typically used. The sampling rate of EEG signals typically ranges from 128 Hz to 1000 Hz.

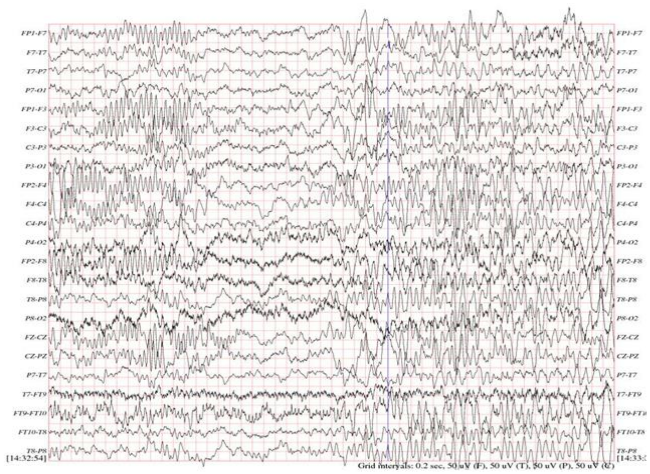

Figure 3.2 Sample of 23-channel EEG signal recorded for 10 ms and sampled at 256 Hz with 16-bit resolution

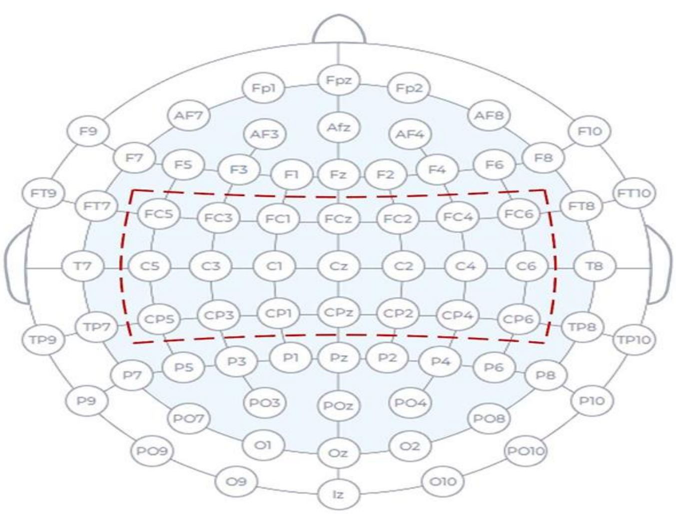

Figure 3.3 Locations of 74 electrodes on the scalp using the standard 10-20 system.

3.1.2 Pre-processing

In order to extract valuable MI components from EEG signals, pre-processing is generally performed in three main steps: channel selection, signal frequency filtering, and artifact removal. In channel selection, a subset of MI-EEG data is selected from a set of EEG electrodes that contain the most distinct MI features, helping to reduce system complexity, computational time, equipment cost, and possibly system performance. More than 79% of the studies reviewed used all EEG channels in the datasets, while 8 studies investigated the effect of channel selection on the accuracy of MI classification with different numbers of electrodes.

Signal frequency filtering was used as a pre-processing step for MI classification in the majority of the studies (91%) for two reasons: selecting the most valuable frequency bands for MI tasks and eliminating artifacts. For the MI-EEG signal, the ERD/ERS of sensorimotor rhythms mainly occurs in μ (8–12 Hz) and β (18–26 Hz) frequency bands. Therefore, almost all the studies reviewed in this survey (96%) included these two frequency bands in their analysis, as they contain the most distinctive features related to MI activities. With frequency filtering, a large portion of the noise can also be eliminated, such as low-frequency artifacts (e.g., EOG, caused by eye blinking) and high-frequency noise (e.g., EMG above 35 Hz). For this reason, 47% of the studies suggested using frequency bands in the range of 6–35 Hz. However, artifacts cannot be easily excluded using band-pass filters, as they may interfere with the effective ERD/ERS bands. Several other studies (35%) suggested using a wider frequency band than 6–35 Hz, in the range of 0–40 Hz, we identify three main strategies for removing artifacts in the reviewed papers: automatic removal (20%), manual removal (4%), without artifacts removal (40%), Most of the reviewed studies classified the MI-EEG signal without any artifact removal based on the fact that deep learning is able to extract useful features from raw and unfiltered data. Other studies employed artifact removal approaches before inputting the MI data into deep learning models. The most common method used in the reviewed papers was ICA, and common average reference (CAR), Some studies used more advanced tools to remove MI signal artifacts, such as the automatic artifact removal (AAR) toolbox.

3.1.3 Feature Extraction

In the feature-based input formulation, the process of MI classification is performed in two steps. First, conventional feature extraction approaches translate EEG signals into vectors. Then, the feature vectors are entered into a deep learning model that trains to classify the data associated with those features. The most popular features extracted from MI EEG data in the previous works were the CSPs. Luo et al. utilized FBCSP to extract spatial-frequency sequential time slices from MI-EEG signals and classified them using long short-term memory (LSTM) and gated recurrent unit (GRU) models. The proposed method achieved good results for both recurrent models. the authors also used the FBCSP approach to extract spatial features from MI data and fed them into a CNN model as a 2D matrix. Other types of features have also been used for MI-EEG classification with DL-based methods including frequency features(e.g., FFT, discrete cosine transform (DCT), and PSD), time-frequency features (e.g., wavelet packet decomposition (WPD), discrete wavelet transform (DWT), empirical mode decomposition (EMD), and Hilbert-Huang transform (HHT)), and temporal features (e.g., statistical measures).

3.1.4 Classification

Discriminative DL models: Discriminative DL models refer to DL architectures that can learn distinct features from input signals using nonlinear transformations and classify them using probabilistic prediction into pre-defined classes. Therefore, these techniques can be used for both feature extraction and classification. Discriminative models include CNN, RNNs (and their variations, GRU and LSTM), MLP, and ELM. A CNN is one of the most common models for deep learning that specializes in extracting local and spatial patterns. The CNN architecture consists of a group of neural networks arranged in a particular order with layers of different sizes, where each layer performs a particular task. The earlier layers learn low-level features while the deeper layers learn high-level features. CNNs typically consist of three structure blocks: convolutional layers (for feature extraction), pooling layers (for feature dimensionality reduction), and fully connected (FC) layers (for classification). A convolutional layer is an essential component of the CNN architecture that performs feature extraction. A pooling layer provides a typical downsampling operation that reduces network computation. The output feature maps of the pooling layer are typically flattened and are connected to one or more fully connected layers. CNN-based DL strategies were used in 78% of the studies reviewed (including standalone and hybrid CNNs).

Representative DL models Representative DL models refer to DL architectures that specialize in feature extraction in an unsupervised manner, which can be used for various tasks, such as clustering and classification. Representative DL models include deep AEs (D-AEs), deep RBMs (D-RBMs), and DBN. An autoencoder (AE) is a type of representative artificial neural network used to learn features in an unsupervised manner with efficient data coding. AE composes of three main components: an encoder, code, and decoder. The encoder compresses the input into a latent-space representation known as the code, which is then used by the decoder to reconstruct the input. There are many varieties of AE, and this review identifies three general types that differ significantly in how they function: D-AE, SAE, and VAE. D AE learns like normal AEs, where all layers in the network are trained at the same time, no matter how many layers in the network. In SAE, different stacked AE blocks are trained separately, where the representation of each block (code) is used as input for the next block. SAE is defined as an AE-based DBN (DBN-AE) and will be explored next with DBNs. VAEs, proposed in 2013, differ from other AEs in that they have a layer of data means and standard deviations at their core, allowing for easy interpolation and random sampling. VAE is one of the most powerful generating methods. VAE is explored with generative DL models. Typically, AEs, other than VAEs, are used for feature extraction; therefore, AE models are usually combined with other discriminative DL models to create hybrid model.

Generative DL models: Generative DL models are typically used to augment and improve training data. The most popular generative DL models are GAN and VAE. Several studies in this review used traditional data augmentation approaches, i.e., non-DL, to increase the size of training data, such as noise addition. sliding window, and amplitude perturbation. Two of the reviewed studies introduced DL-based data augmentation using GAN and VAE networks.

The results from these studies showed that using GAN models for MI data augmentation significantly increased classification performance. Zhang et al proposed a four-layer GAN model for MI data augmentation and compared its performance with VAE and other traditional augmentation methods such as geometric transformation and noise addition. The results revealed that both GAN and VAE outperformed the traditional methods, while the GAN had the best performance. The study found that the performance of a CNN model trained on MI data augmented with GANs improved by 17% and 21% for the BCI-C IV-2a.

3.2 Studies

This section will discuss different studies about EEG classification and motor imagery for different references [1:22]

3.2.1 Two-class Studies:

Feng Li et al.[3], selected the public BCI Competition IV Dataset 2b to validate the effectiveness of our proposed algorithm. This dataset was collected from nine subjects using electrodes C3, Cz, and C4 at a sampling frequency of 250 Hz, with each subject participating in five sessions. We analyzed data from the last two sessions (04E and 05E), each containing 160 trials. We proposed a CWT-SCNN algorithm to classify motor imagery (MI) EEG signals, The raw MI-EEG signals were first filtered within a 4–35 Hz range, then transformed into time-frequency images using continuous wavelet transform (CWT), focusing on the mu and beta rhythms for SCNN training. Our SCNN comprised six layers: an input layer, two convolutional layers (each with batch normalization and ReLu), a flatten layer, a fully connected layer, and an output layer. The C2 convolution layer had eight filters extracting frequency features, while the C3 layer's 16 filters extracted time-domain features. After flattening and fully connecting these features, the softmax layer predicted the output class probabilities. Left and right-hand MI-EEG signals, crucial for brain-computer interface (BCI) systems, present classification challenges due to low signal-to-noise ratios. Deep learning has recently shown promise in pattern recognition, yet few algorithms effectively apply it to MI-based BCI. Our algorithm combined

CWT with a simplified convolutional neural network (SCNN), mapping MI-EEG signals to time-frequency images for feature extraction and classification. Testing on the BCI Competition IV Dataset 2b demonstrated an average classification accuracy of 83.2% and a mean kappa value of 0.651, outperforming the competition champion by 11.9%. This CWT-SCNN algorithm not only enhanced classification performance but also reduced training time, suggesting its potential for real-time BCI systems, particularly for aiding disabled individuals

TARIQ SADIQ et al.[4], proposed a robust and straightforward automated multivariate empirical wavelet transform (MEWT) algorithm for decoding different motor imagery (MI) tasks. The study's contributions are four-fold: first, we utilize a multiscale principal component analysis method in the preprocessing module to enhance noise robustness. Second, we introduce a novel automated channel selection strategy, which is validated through comprehensive comparisons among three different strategies for decoding channel combination selection. Third, we adopt a sub-band alignment method using MEWT to obtain joint instantaneous amplitude and frequency components for the first time in MI applications. Fourth, a robust correlation-based feature selection strategy is applied to significantly reduce system complexity and computational load. Extensive experiments were conducted for both subject-specific and subject-independent cases using three benchmark datasets from BCI Competition III to evaluate the proposed method's performance with typical machine learning classifiers. For subject-specific cases, experimental results showed an average sensitivity, specificity, and classification accuracy of 98% using multilayer perceptron neural networks, logistic model tree, and least-square support vector machine (LS-SVM) classifiers, respectively, achieving an improvement of up to 23.50% in classification accuracy compared to existing methods. For subject-independent cases, average sensitivity, specificity, and classification accuracy were 93%, 92.1%, and 91.4%, respectively, using the LS-SVM classifier for all datasets, with an improvement of up to 18.14% over existing methods. Additionally, the proposed algorithm achieved a classification accuracy of 100% for subjects with small training sizes in subject-specific cases and for subject-independent cases using a single source subject. These results demonstrate the MEWT algorithm's great potential for practical MI EEG signal classification.

Ikhtiyor Majidov and Taegkeun Whangbo [5], focus on single-trial motor imagery (MI) classification, a critical aspect of brain-computer interface (BCI) applications, by leveraging Riemannian geometry-based feature extraction methods and deep-learning techniques. Riemannian geometry, primarily used with covariance matrices in information theory, preserves frequency and spatial information when combined with a filterbank approach. This study aims to integrate deep-learning classifiers with BCI applications using common spatial patterns (CSP) and Riemannian geometry feature extraction, alongside particle swarm optimization (PSO) for feature selection. The CSP method was utilized for multiclass classification with a single classifier, combining power spectrum density features with covariance matrices mapped onto a Riemannian manifold's tangent space. PSO facilitated training by penalizing poor features, and a moving windows method was used for data augmentation. The pre-processed data was classified using a convolutional neural network (CNN), significantly improving classification accuracy on the BCI Competition IV 2a dataset.

The proposed method included several notable algorithms. Data augmentation was performed using a sliced windows method, enhancing the number of training samples. Following noise reduction through independent component analysis, the CSP algorithm was applied, and spatially filtered data were concatenated. The data was then transformed into square matrices using covariance matrices. Deep networks were employed despite challenges with noise and data scarcity. A 1D CNN was used for data classification, as deeper networks led to overfitting. Two CNN architectures were tested: one with CNN and softmax layers, and another with CNN followed by fully connected layers. Training utilized a filter size of seven, rectified linear unit activation functions, and the Adam optimizer, with a variable learning rate and a 20-epoch limit to avoid overfitting. This comprehensive approach demonstrates the potential for improved MI EEG signal classification in practical BCI applications.

Brain-computer interaction (BCI) based on electroencephalography (EEG) can assist patients with limb dyskinesia in daily life and rehabilitation training. However, due to the low signal-to-noise ratio and significant individual differences, EEG feature extraction and classification often suffer from low accuracy and efficiency. To address these challenges, ***Xiongliang Xiao and Yuee Fang[6]***, proposed a motor imagery EEG signal recognition method based on a deep convolutional neural network (CNN). Initially, to tackle the issue of low-quality EEG signal characteristic data, short-time Fourier transform (STFT) and continuous Morlet wavelet transform (CMWT) are employed to preprocess the collected experimental datasets based on their time series characteristics. This preprocessing yields EEG signals with distinct time-frequency characteristics. The improved CNN network model is then used for efficient recognition, enabling high-quality EEG feature extraction and classification. This approach enhances the quality of EEG signal feature acquisition and ensures high accuracy and precision in EEG signal recognition. The proposed method was validated using the BCI Competition dataset and laboratory-measured data. Experimental results demonstrated an accuracy of 0.9324, a precision of 0.9653, and an AUC of 0.9464, indicating good practicality and applicability.

The improved CNN structure is mainly divided into five layers. The first layer is the input layer (I1), with an input sample data size of 1 × N, where N represents the number of features obtained after the motor imagery EEG is processed by the common spatial patterns (CSP) algorithm (N = 4 × m). The second layer (C2) and the third layer (C3) are both convolutional layers designed for feature extraction from the input sample data. The C2 layer contains i2 convolution kernels of size 1 × n2, while the C3 layer has i3 convolution kernels of size 1 × n3. Due to the short length of the input sample data, the downsampling layer is omitted in this CNN architecture. The fourth layer (F4) and the fifth layer (O5) form a single-layer perceptron in a fully connected manner. After processing the output of the C3 layer, the classification result is produced, demonstrating the method's efficiency in recognizing EEG signals.

3.2.2 Four-class Studies

Hassanpour et al. *[7]*, a novel gradient deep learning (GDL) methodology is proposed to leverage the end-to-end (e2e) property of deep learning models, requiring only minimal objective-free preprocessing steps. To address the complexity of multi-class motor imagery (MI) EEG signals, an innovative multilevel GDL-based classification scheme is introduced. The robustness of the proposed model against noisy MI-EEG signals is evaluated using two different GDL models: a deep belief network (DBN) and a stacked sparse autoencoder (SSAE), both in an e2e manner. Experimental results indicate that the proposed methodology improves accuracy compared to the widely used filter bank common spatial patterns (FBCSP) algorithm.

To measure the robustness of GDL models against noisy high-dimensional MI-EEG signals, this study explores two architectures with different learning procedures: (a) the DBN model, which consists of multiple restricted Boltzmann machine (RBM) blocks and a softmax regression (SR) layer, and (b) the SSAE model, which comprises several sparse autoencoder (SAE) blocks and an SR layer. The SR layer outputs a vector specifying the probability of each input sample belonging to each class.

In the preprocessing block, to exploit the e2e learning property of deep learning models, all stages—preprocessing, feature extraction, and classification—are integrated into one deep learning block, termed the multilevel binary classifier (MBC) block. This approach eliminates steps with specific objective functions, such as common spatial patterns (CSP), noise reduction, and feature extraction, which are typical in traditional methods. Instead, a few objective-free steps, including data augmentation, band-pass filtering, and converting signals to the frequency domain, are retained.

The MBC employs several basic classifiers (BC) distributed in a binary tree structure to simplify the classification of multi-class noisy high-dimensional MI-EEG problems. This divide-and-conquer strategy divides the multi-class problem into binary subproblems, grouping the most similar classes within the same subproblem using a newly introduced similarity metric and a decomposition process to determine the binary tree's architecture. Building on the one-versus-the-rest common spatial pattern (OVR-CSP) algorithm, an M × M similarity matrix (SM) is derived to approximate the similarity between multiple classes. By using whitening transformations and analyzing the transformed eigenvalues of each class, the study establishes a similarity measure that aids in the effective classification of MI-EEG signals.

Xu et al. ***[8],*** Recognition of motor imagery intention is a prominent focus in brain-computer interface (BCI) research, particularly for aiding patients with physical dyskinesia in conveying movement intentions. Despite recent breakthroughs in using deep learning for motor imagery task recognition, ignoring key features related to motor imagery can degrade algorithm performance. This paper proposes a novel deep multi-view feature learning method for classifying motor imagery electroencephalogram (EEG) signals to address this issue.

To capture more representative motor imagery features in EEG signals, a multi-view feature representation approach was introduced, based on the distinct characteristics of EEG signals and the differences among various features. Different feature extraction methods were used to separately extract the time domain, frequency domain, time-frequency domain, and spatial features of EEG signals, allowing these features to cooperate and complement each other. Subsequently, a deep restricted Boltzmann machine (RBM) network, enhanced by t-distributed stochastic neighbor embedding (t-SNE), was employed to learn the multi-view features of EEG signals. This approach effectively eliminated feature redundancy while considering the global characteristics in the multi-view feature sequence reduced the dimension of the multi-visual features, and enhanced feature recognizability. Finally, a support vector machine (SVM) was used to classify the deep multi-view features.

Applying this proposed method to the BCI Competition IV 2a dataset yielded excellent classification results. The findings demonstrate that the deep multi-view feature learning method significantly improved the classification accuracy of motor imagery tasks, highlighting its potential for enhancing BCI applications.

Mahamune et al. ***[9]***, a novel framework proposed to classify four motor imagery (MI) tasks—left hand, right hand, feet, and tongue—using two-dimensional (2D) images for a convolutional neural network (CNN). The framework begins by pre-processing the MI-based EEG data with a multi-class common spatial pattern (CSP) method and then decomposing each trial using a continuous wavelet transform (CWT) filter bank to form 2D images. These images are used to train the CNN model for classification.

The proposed framework was evaluated using the publicly available BCI Competition IV dataset 2a, calculating the classification accuracy for all subjects. The results demonstrated that this framework achieved better classification accuracy compared to some existing CNN-based and conventional machine-learning approaches. The average time required to train the CNN using this framework was 12.67 seconds, making it suitable for online MI-based BCI applications.

The spatially filtered EEG signal is obtained using the CSP algorithm, followed by applying the CWT filter bank to extract detailed information at every frequency and scale. The raw EEG data from Sessions 1 and 2 are band-pass filtered between 8 and 30 Hz using a fifth-order Butterworth finite impulse response (FIR) filter. Trials are then segmented, and the data from each trial are selected within a time slot starting from 2.5 to 5.5 seconds. The CSP method decomposes the raw EEG signals into spatial patterns with maximum differences in variance, reducing the dataset with 22 EEG channels to eight dimensions.

After pre-processing with multi-class CSP, all trials from each channel are normalized using z-score normalization to enhance features and identify outliers. Feature extraction is performed using CWT, transforming the one-dimensional signals into a two-dimensional matrix in the frequency domain. This transformation, known as a scalogram, consists of the absolute value of the wavelet coefficients at different scales. The scalogram images are then used to classify motor imagery signals. For the classification of four classes, CSP is extended to the multi-class paradigm using the one-versus-rest scheme.

Finally, the multi-class CSP filtered signal is decomposed using a CWT filter bank to develop 2D images, which are then used to train the CNN. The number of samples in each trial is 751. This comprehensive process ensures effective classification of MI tasks, demonstrating the practicality and applicability of the proposed framework in BCI applications

3.3 Summary and Observations

Table 3.1 summarizes the key findings from the related studies discussed.

Study	Pre-processing	Classification	accuracy
Hassanpour et al. [15]	Band-pass filter, Data augmentation, Converting signals to Freq. domain	Multi-level binary classification	87.99%
Xu et al. [29]	—	Deep multi-view feature learning	78.5%
Mahamune et al. [7]	Band-pass filter, CSP , Segmentation, Z-score	CNN Model	71.25%
Xiongliang Xiao and Yuee Fang [24]	STFT, CMWT	CNN model	93.24%
Ikhtiyor Majidov and Taegkeun Whangbo [17]	CSP, PSO	CNN model	82.39%
TARIQ SADIQ et al.[16]	PCA, channel selection, MEWT	LS-SVM	91.4%
Feng Li et al. [14]	Band-pass filter, CWT	SCNN	83.2%

Across all studies, deep learning models, including CNNs, SCNNs, RBMs, and SSAEs, consistently showed significant improvements in classification accuracy compared to traditional methods. The ability of deep learning to automatically extract and learn features from raw EEG data is a key factor in these improvements.

Effective pre-processing techniques, such as CSP, CWT, and MEWT, play a crucial role in enhancing the quality of EEG signals and the performance of classification algorithms. Methods that integrate time-domain, frequency-domain, and spatial features tend to produce better results.

Strategies for noise reduction and handling individual differences in EEG signals, such as multiscale principal component analysis and robust feature selection methods, are vital for achieving high classification accuracy. Robustness against noisy signals is a common theme in successful algorithms.

Several studies emphasize the efficiency and speed of the proposed methods, making them viable for real-time BCI applications. Quick training times and high accuracy rates are essential for practical deployment in assisting patients with motor impairments.

Combining multiple feature extraction methods to capture a wide range of signal characteristics enhances the classification performance. Multi-view feature learning and multi-class CSP methods exemplify the benefits of comprehensive feature extraction.

these studies collectively indicate that integrating advanced signal processing techniques with deep learning models significantly enhances the classification accuracy of MI-EEG signals. This progress is critical for developing effective and practical BCI systems that can greatly benefit individuals with physical impairments.

Chapter 4

System Architecture and Methods

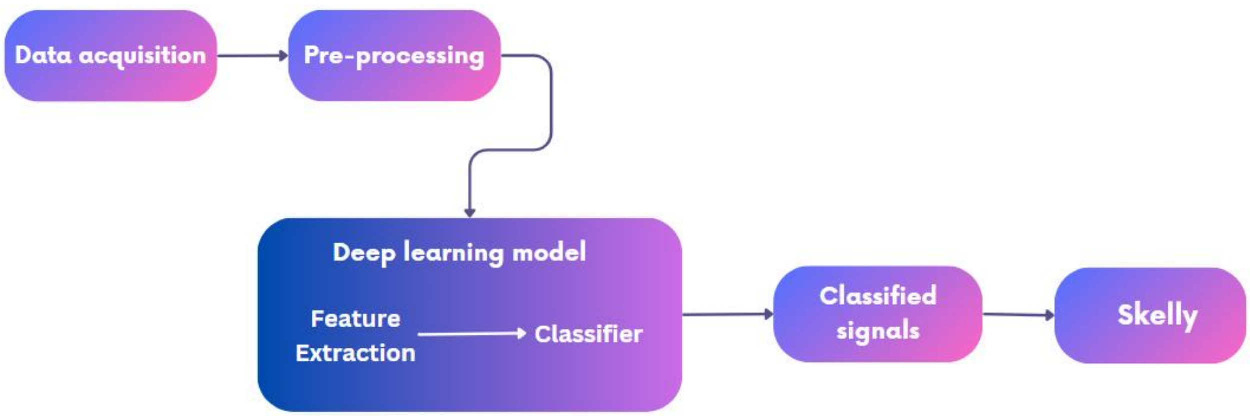

Figure 4.1 system architecture

Figure 4.1 describes our system architecture which is the same as the general architecture of the BCI applications that consists of : data acquisition, pre-processing, deep learning model (feature extraction then classification) and the game (skelly) which will be controlled by the classified EEG signals

4.1 Pre-processing

1. Filtering using the band-pass filter in the range of 8-30 Hz to capture alpha and beta bands
2. Z-score normalization standardizes data to have a mean of 0 and a standard deviation of 1, which enhances the performance of machine learning algorithms by ensuring features contribute equally
3. Segmentation from 1,5-6 sec & 3-6 sec

4.2 Implemented Models

4.2.1 Resnet50

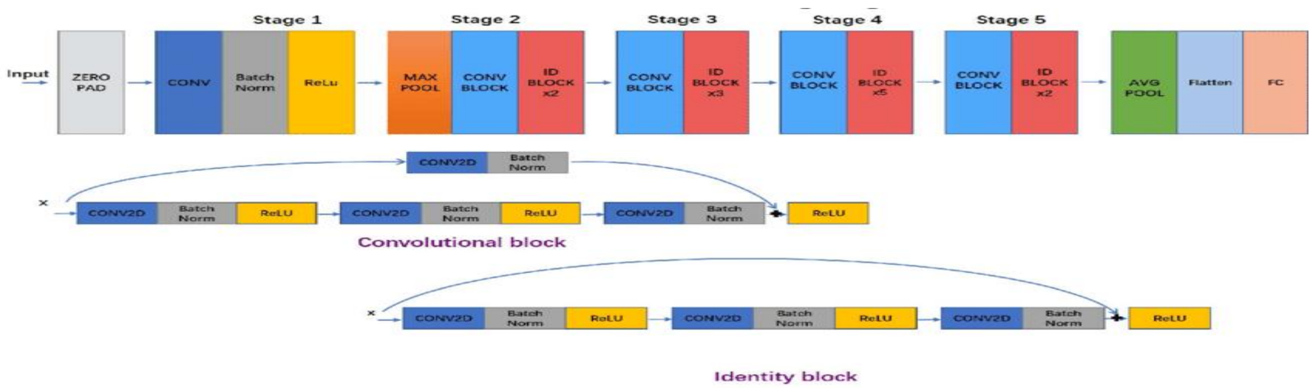

Figure 4.2 Resnet50 model

ResNet50*[10]*, short for Residual Network with 50 layers, is a deep convolutional neural network (CNN) architecture that has become a popular choice for various image recognition tasks due to its remarkable performance and innovative design. Developed by Kaiming He and his colleagues at Microsoft Research, ResNet50 was introduced as part of the deeper ResNet family, which won the 2015 ImageNet competition by achieving unprecedented accuracy.

The core innovation of ResNet50 is its use of residual learning, which addresses the degradation problem encountered when training very deep neural networks. In traditional deep networks, as the number of layers increases, the accuracy often saturates and then degrades rapidly. ResNet50 overcomes this by incorporating residual blocks, which allow the network to learn residual functions with reference to the layer inputs. Essentially, these blocks let the network skip one or more layers, forming a shortcut connection that helps preserve the identity mapping and mitigates the vanishing gradient problem.

The architecture of ResNet50 consists of 50 layers, including convolutional layers, batch normalization layers, activation layers (ReLU), and pooling layers. The network is divided into several stages, each comprising multiple residual blocks. Here is a breakdown of the main components and structure:

Initial Layers:

The network starts with a 7x7 convolutional layer with 64 filters and a stride of 2, followed by a 3x3 max pooling layer with a stride of 2. This reduces the spatial dimensions of the input image while capturing essential features.

Residual Blocks:

ResNet50 contains a series of residual blocks, each comprising three convolutional layers. The first layer of a residual block is a 1x1 convolution that reduces the dimensionality, the second layer is a 3x3 convolution, and the third layer is another 1x1 convolution that restores the original dimensionality. Batch normalization and ReLU activation are applied after each convolution.

The shortcut connection bypasses these three layers and directly adds the input of the block to the output of the third convolution, forming the residual connection.

Stages:

The network is organized into four stages, each with a different number of residual blocks. The first stage has 3 blocks, the second has 4 blocks, the third has 6 blocks, and the fourth has 3 blocks. Each stage processes the feature maps at different scales and depths, allowing the network to learn hierarchical representations of the input data.

Final Layers:

- After the last stage, the network applies a global average pooling layer, which reduces each feature map to a single value by averaging its spatial dimensions. This is followed by a fully connected (dense) layer with 1000 units (for classification tasks involving 1000 classes) and a softmax activation function to produce the probability distribution over the classes.
- The use of residual connections in ResNet50 not only allows the network to be significantly deeper than its predecessors but also

makes it easier to train. The residual blocks enable gradient flow through the network even as the number of layers increases, leading to better performance and faster convergence during training. ResNet50's architecture has been widely adopted and extended in various applications, from image classification to object detection and beyond, making it a cornerstone of modern deep learning models.

4.2.2 EEGNet

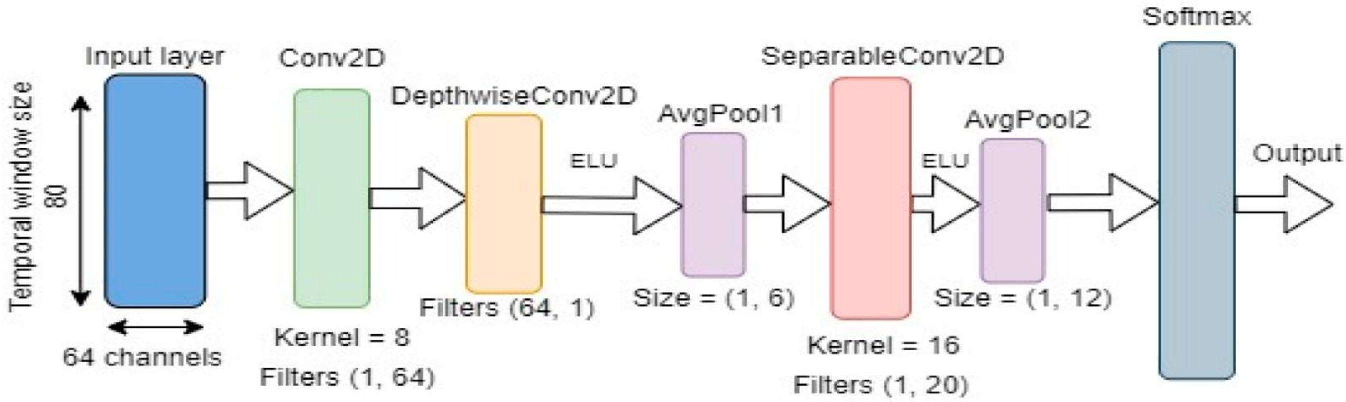

Figure 4.3 EEGNet model

Here we introduce EEGNet***[11]***, a compact CNN architecture for EEG-based BCIs that (1) can be applied across several different BCI paradigms, (2) can be trained with very limited data and (3) can produce neurophysiologically interpretable features.

The network starts with a temporal convolution (second column) to learn frequency filters, then uses a depthwise convolution (middle column), connected to each feature map individually, to learn frequency-specific spatial filters. The separable convolution (fourth column) is a combination of a depthwise convolution, which learns a temporal summary for each feature map individually, followed by a pointwise convolution, which learns how to optimally mix the feature maps together.

In Block 1, we perform two convolutional steps in sequence. First, we t F1 2D convolutional filters of size (1,64), with the filter length chosen to be half the sampling rate of the data.

(here, 128Hz), outputting F1 feature maps containing the EEG signal at different band-pass frequencies. Setting the length of the temporal kernel at half the sampling rate allows for capturing frequency information at 2Hz and above. We then use a Depthwise Convolution of size (C1) to learn a spatial filter. In CNN applications for computer vision, the main benefit of a depthwise convolution is reducing the number of trainable parameters to t, as these convolutions are not fully connected to all previous feature maps. Importantly, when used in EEG-specific applications, this operation provides a direct way to learn spatial filters for each temporal filter, thus enabling the efficient extraction of frequency-specific spatial filters. A depth parameter D controls the number of spatial filters to learn for each feature map. This two-step convolutional sequence is inspired in part by the Filter-Bank Common Spatial Pattern (FBCSP) algorithm and is similar in nature to another decomposition technique, Bilinear Discriminant Component Analysis We keep both convolutions linear as we found no significant cant gains in performance when using nonlinear activations. We apply Batch Normalization along the feature map dimension before applying the exponential linear unit (ELU) nonlinearity To help regularize or model, we use the Dropout technique We set the dropout probability to 05 for within-subject classification to help prevent over-fitting when training on small sample sizes, whereas we set the dropout probability to 025 in cross-subject classification, as the training set sizes are much larger We apply an average pooling layer of size(1,4) to reduce the sampling rate of the signal to 32Hz. We also regularize each spatial filter by using a maximum norm constraint of 1 on its weights;

In Block2, we use separable convolution, which is a Depthwise Convolution (here, of size (116), representing 500ms of EEG activity at 32Hz) followed by (1,1)Pointwise Convolutions The main benefits of separable convolutions are (1) reducing the number of parameters to fit and (2) explicitly decoupling the relationship within and across feature maps by first learning a kernel summarizing each feature map individually, then optimally merging the outputs afterward When used for EEG-specific applications this operation separates learning how to summarize individual feature maps in time(the depthwise convolution)with how to optimally combine the feature maps(the pointwise convolution). An Average Pooling layer of size(1,8) is used for dimension reduction.

In the classification block, the features are passed directly to a SoftMax classification with N units, N being the number of classes in the data. We omit the use of a dense layer for feature aggregation before the SoftMax classification layer to reduce the number of free parameters in the model.

4.2.3 ATCNet(Attention Temporal Convolutional Network)

we propose an attention-based temporal convolutional network, ATCNet*[12]*, to decode MI-EEG brain signals. This research utilizes SciML to address domain-specific MI-EEG data challenges, which results in a robust, interpretable, and explainable DL model specifically designed for decoding MI-EEG brain signals.The proposed DL model processes the MI-EEG data in three steps: first, encode the MI EEG signal into a sequence of high-level temporal representations using conventional layers, then, highlight the most valuable information in the temporal sequence using an attention layer, and finally, extract high-level temporal features from the highlighted information using a temporal convolutional layer. The proposed model utilizes a multi-head self-attention and convolutional-based sliding window to boost the performance of MI classification.

1. We propose a high-performance ATCNet model, which utilizes the powerful of TCN, SciML, attention mechanism, and convolutional-based sliding window.

2. Performing sliding-window using convolution helps augment MI data and efficiently enhance accuracy, by parallelizing the process and reducing computations.

3. Self-attention helps the DL model to attend to the most effective MI information in the EEG data, and the multiple heads help to focus on multiple positions, resulting in multiple attention representations

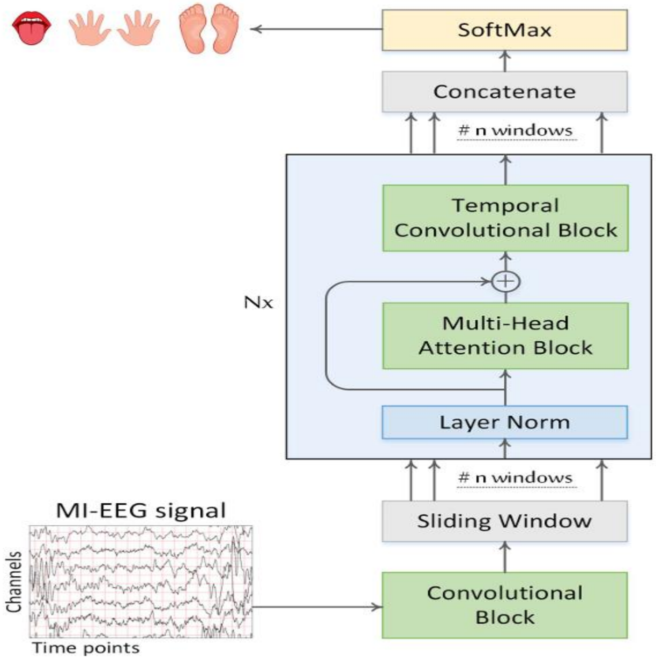

Figure 4.4 ATCNet model

The proposed ATCNet model consists of three main blocks: convolutional (CV) block, attention (AT) block, and temporal convolutional (TC) block,

CV block encodes low-level spatio-temporal information within the MI EEG signal through three convolutional layers: temporal, channel depth-wise, and spatial convolutions. The output of the CV block is a temporal sequence with a higher-level representation. The AT block then highlights the most important information in the temporal sequence using a multi head self-attention (MSA). Finally, the TC block extracts high-level temporal features within the temporal sequence using TCN and feeds them into a fully connected (FC) layer with a SoftMax classifier. The temporal sequence, output from CV block, can be split into multiple windows and each is fed to AT/TC blacks separately. The output of

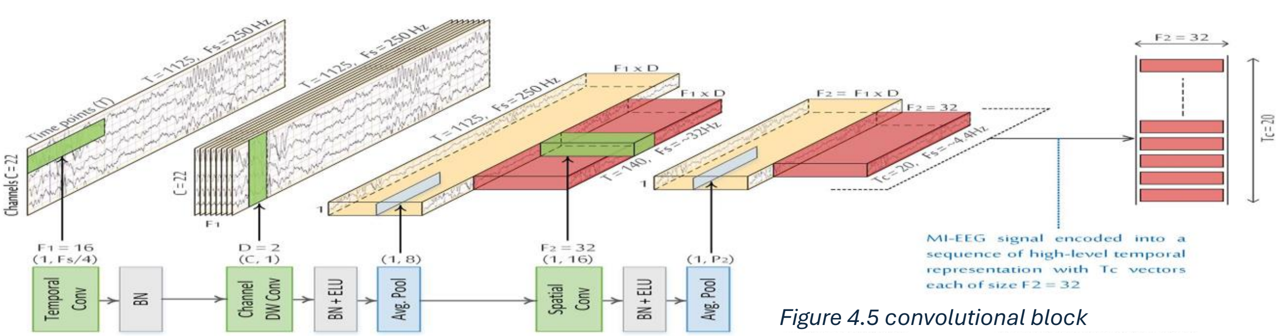

Figure 4.5 convolutional block

all windows is then concatenated and fed to a SoftMax classifier. This helps efficiently augment the data and enhances accuracy.

CV block consists of three convolutional (conv) layers, The first layer performs a temporal convolution using F filters of size (1,K), where K is the filter length in the time axis. K was set to be one-fourth of the sampling rate (64 for BCI-2a). This allows the filters to extract temporal information associated with frequencies above 4 Hz.

The output of this layer is F temporal feature maps. The second layer is a depth-wise convolution with F filters of size (C,1), where C is the number of EEG channels. Using depth-wise convolution, each filter extracts spatial features (i.e., related to EEG channels) from a single temporal feature map. Therefore, the output of this layer is $F\times D$ feature maps, map. Therefore, the output of this layer is $F\times D$ feature maps, where D is the number of filters linked to each temporal feature map in the previous layer. D is set empirically to 2. $F\times D$ determines the output dimension of the CV block. The depth-wise convolution is followed by an average pooling layer of size (1, 8) to abstract the temporal data by a factor of 8. This reduces the sampling rate of the signal to ~32Hz. The third convolutional layer consists of F filters of size (1,K). K was set to 16 to decode MI activities within 500 ms (for 32 Hz sampled data). Finally, a second average pooling layer with a size of (1, P) is used to reduce the sampling rate to ~32/P Hz. P is used to control the length of the temporal sequence produced by CV block. The second and third conv layers are followed by batch normalization to speed up network training and then by exponential linear unit (ELU) activation for nonlinearity.

Instead of entering the whole T samples of z to the later layers, a sliding window has been used to divide the temporal sequences into multiple windows. This helps to augment the data and enhance the decoding accuracy.

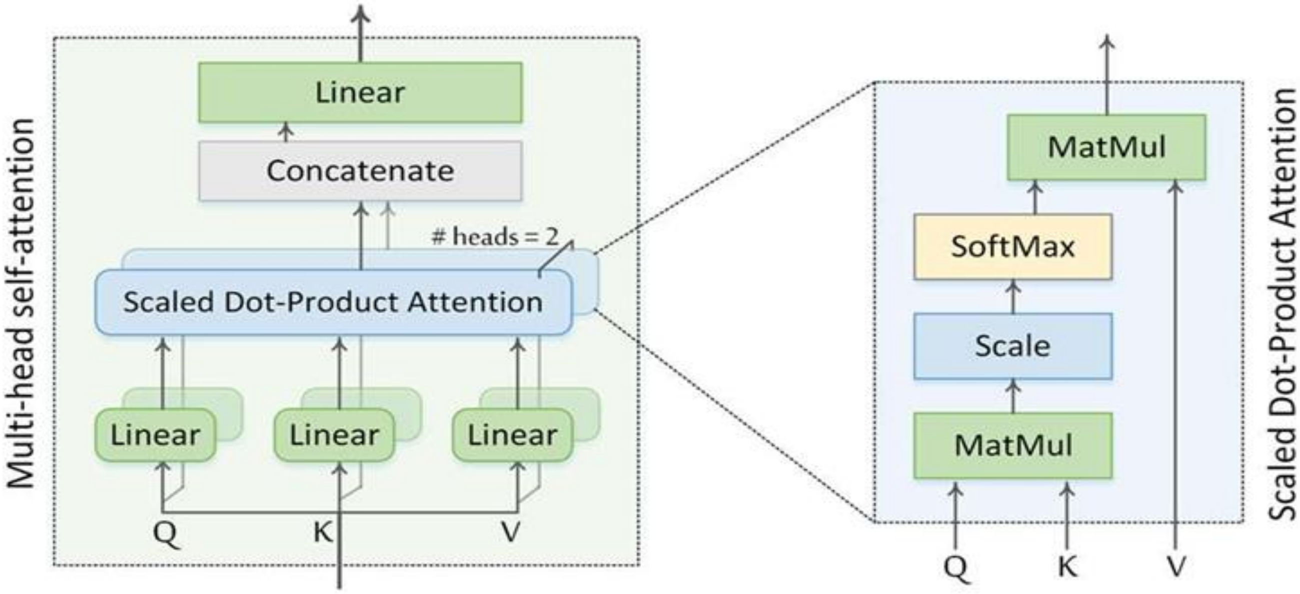

Figure 4.6 Attention block

the attention block consists of an MSA layer. MSA consists of several self-attention layers (i.e., scaled dot-product attention) called heads, as shown in the above Figure. Each self-attention layer consists of three main components: query Q, keys K, and values V. Interactions between query and keys produce attention scores that guide selection bias over values.

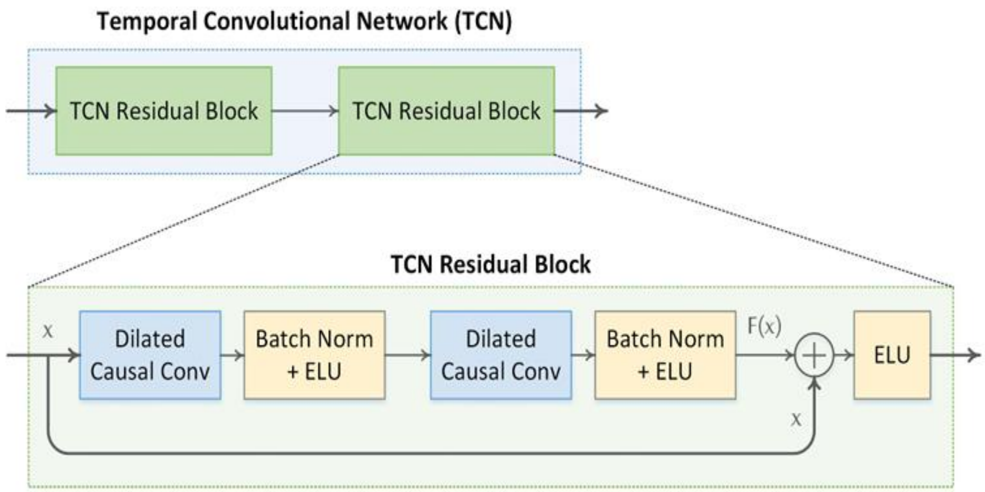

Figure 4.7 temporal convolutional network block

The TC block has the same architecture as the TCN described in [22]. TCN consists of a stack of residual blocks. The residual block is composed of two dilated causal convolutional layers, each one followed by batch normalization and ELU activation, as shown in the above Figure.

Chapter 5

Dataset and Results

5.1 Dataset

Upon comprehensive analysis of available datasets, we identified the ***BCI Competition IV dataset 2a****[13]* as the optimal choice for our research objectives. This dataset was particularly advantageous because it includes four distinct motor imagery classes—right hand, left hand, both feet, and tongue—which align perfectly with the directional controls in our application. Furthermore, its cue-based (synchronous) design ensures that the motor imagery tasks are initiated in response to specific visual cues (such as an arrow in our data), enhancing the precision of our experiments. Additionally, the subject-dependent nature of the dataset allows for tailored training and testing on each individual's data, thereby increasing the relevance and accuracy of our findings.

The BCI Competition IV dataset 2a comprises data from 9 subjects, each recorded over two separate sessions on different days. Each session includes six runs, with short breaks in between. In every run, there are 48 trials, consisting of 12 trials for each of the four classes (right hand, left hand, both feet and tongue). This results in a total of 288 trials per session.

During each trial, subjects sat comfortably in an armchair facing a computer screen. Initially (t = 0s), a fixation cross appeared on a black screen, accompanied by a brief acoustic warning tone. Two seconds later (t = 2s), a directional cue in the form of an arrow (pointing left, right, down, or up) appeared, corresponding to the motor imagery task for either the left hand, right hand, both feet, or tongue. This arrow remained visible for 1.25 seconds, prompting the subjects to engage in the indicated motor imagery task. Subjects were instructed to continue the task until the fixation cross disappeared at t = 6s, after which a short break ensued with the screen turning black.

5.2 Results

To evaluate the effectiveness of different models in interpreting EEG signals for controlling the "Skelly" game, we trained and tested three deep learning models using the BCI Competition IV dataset 2a. Each model's performance was assessed based on its ability to classify motor imagery tasks into corresponding control signals for the game character.

The training parameters:

1. Batch-size: 64
2. Patience: 100
3. Learning rate: 0.001
4. Number of Epochs: 500
5. Optimizer: Adam
6. Loss function: Categorical Cross-Entropy

Models evaluated:

1. **ResNet50:**

Table 5.1 presents the results obtained using ResNet-50 model. It reveals that no subject achieved an accuracy greater than 50%, except for subject 9.

Subject	Accuracy
Subject 1	30%
Subject 2	32%
Subject 3	35%
Subject 4	40%
Subject 5	41%
Subject 6	45%
Subject 7	50%
Subject 8	55%
Subject 9	60%
Avg. Accuracy	42%

2. EEGNet:

Table 5.2 below presents the results obtained using the EEGNet model. It reveals a discrepancy between the paper's reported accuracy (72.2%) and our achieved accuracy (55.5%).

Subject	Paper Accuracy	Accuracy
Subject 1	80%	72%
Subject 2	60%	52%
Subject 3	90%	72%
Subject 4	65%	48%
Subject 5	68%	52%
Subject 6	57%	33%
Subject 7	86%	81%
Subject 8	75%	48%
Subject 9	73%	34%
Avg. Accuracy	72.2%	55.5%

3. ATCNet:

Table 5.3 represents the highest-performing model we worked on, showing a comparison between the paper's reported accuracy (85.4%) and our achieved accuracy (87.16%). Notably, subject 7 demonstrated the highest accuracy among all subjects.

Subject	Paper Accuracy	Accuracy
Subject 1	88.5%	84.48%
Subject 2	70.5%	77.59%
Subject 3	97.5%	89.66%
Subject 4	81%	81.03%
Subject 5	83%	87.93%
Subject 6	73.6%	77.59%
Subject 7	93.1%	96.55%
Subject 8	90.3%	94.83%
Subject 9	91%	94.83%
Avg. Accuracy	85.4%	87.16%

Each model's performance was measured by its classification accuracy on the test set. The results indicate that ATCNet achieved the highest accuracy, with an impressive 85.4% in translating EEG signals into game control commands. This demonstrates ATCNet's superior ability in mapping brain activity to game controls, outperforming the other models tested. The findings confirm the feasibility of using EEG signals for controlling game mechanics, showing that ATCNet is particularly effective for this purpose.

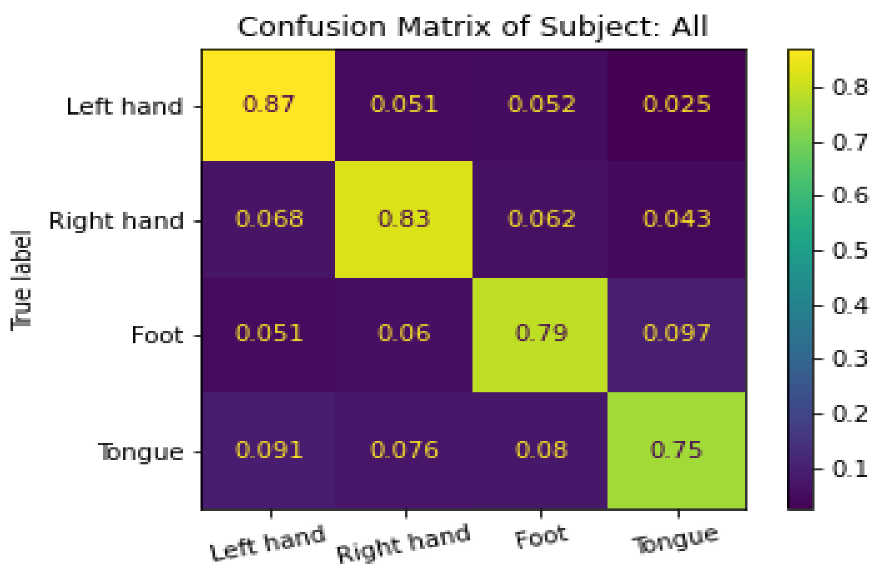

Figure 5.1 ATCNet confusion matrix

Figure 5.1: Confusion matrix showing the classification performance of the model. The diagonal entries represent the number of correct predictions for each class, indicating high accuracy. Off-diagonal entries are minimal, reflecting low misclassification rates. The matrix demonstrates that the model effectively distinguishes between the different classes, contributing to its overall robust performance.

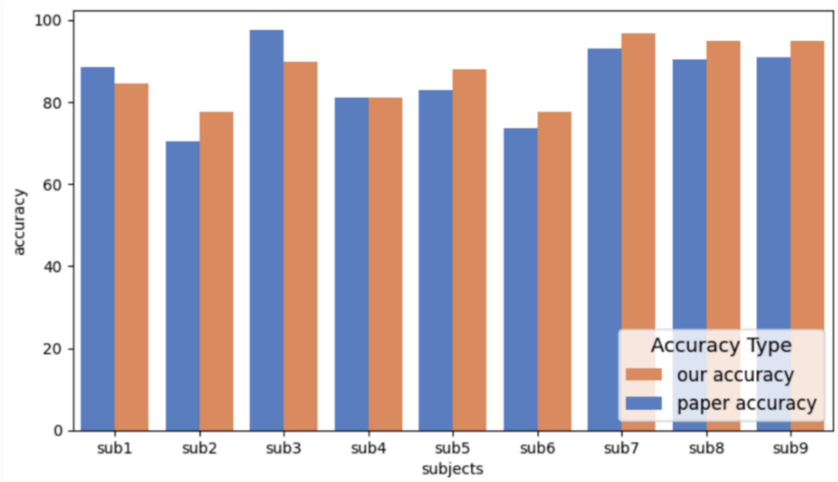

Figure 5.2 difference between implemented ATCNet and paper's accuracy

Figure 5.2: Comparison of model accuracy across subjects between our implementation and the paper's reported results. Our model's accuracy (orange bars) consistently outperforms or closely matches the paper's accuracy (blue bars) for most subjects. Notably, for subjects 3 and 7, our model significantly surpasses the reported accuracy, highlighting improvements in our model's performance for these subjects.

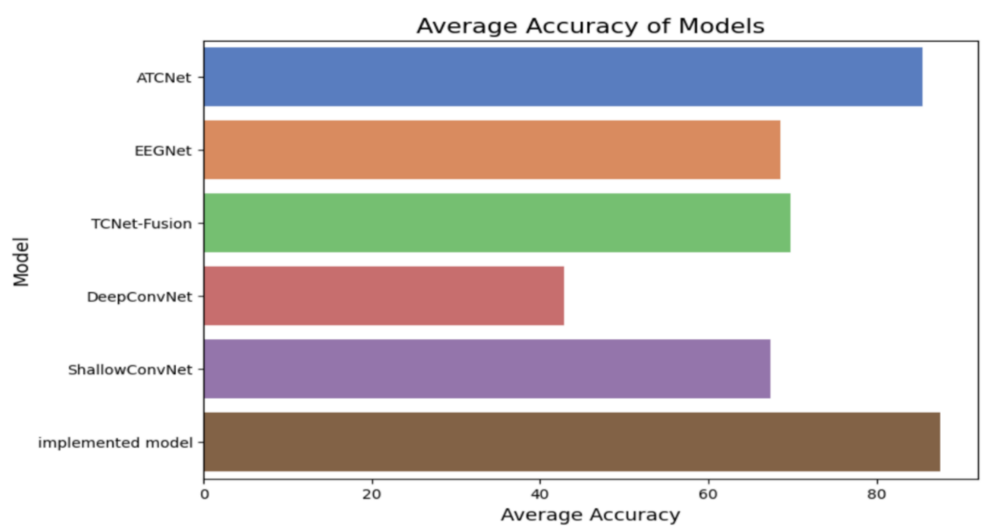

Figure 5.3 difference between implemented model and other DL models

Figure 5.3: Average accuracy of various models, highlighting that our implemented model achieves the highest accuracy compared to others.

Chapter 6
Skelly

6.1 What is skelly?

Skelly is a game that uses brain signals (EEG) using electrodes attached to the scalp to create a fun and engaging gaming experience. It's a type of brain-computer interface (BCI) application, which is a technology that allows users to communicate or control devices or applications with their thoughts. so it can be used for rehabilitation purposes, as it provides feedback, motivation, and stimulation to users who need to improve their abilities or recover from brain injuries, In a number of levels graded in difficulty.

6.2 Flow Board

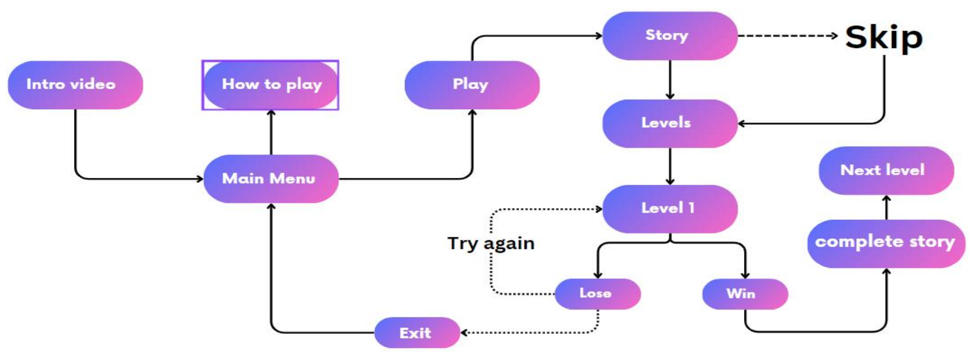

Figure 6.1 skelly flow board

→Upon launching "Skelly," players are greeted by a minimalist loading screen featuring the game's title in a contemporary font, alongside a stylized logo.

Figure 6.2 opening screen

→Following the loading screen, the main menu of "Skelly" emerges, presenting players with three distinct options: play, how to play, exit

Figure 6.3 main menu

→Upon selecting the Play button from the main menu of "Skelly," players are seamlessly transported into the story of the game.

Figure 6.4 game's story

→Choosing the How to Play button from the main menu of "Skelly" directs players to a detailed instructions page, providing comprehensive guidance on navigating the game's mazes using EEG technology.

Figure 6.5 how to play screen

→Exit will close the game

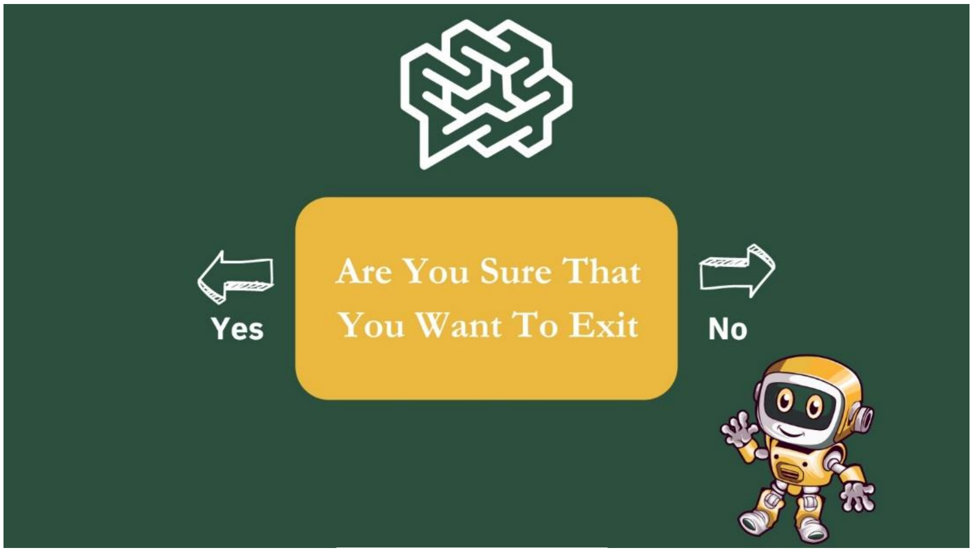

Figure 6.6 Exit screen

→In "Skelly," you can choose to skip the story screens and go straight to playing levels using a skip button. If you prefer, you can also watch the entire story and then automatically move to the levels menu.

Figure 6.7 skip button

In the levels menu of "Skelly," you can navigate through different levels and choose which one to play. Simply select a level to start playing and dive into the maze challenges ahead.

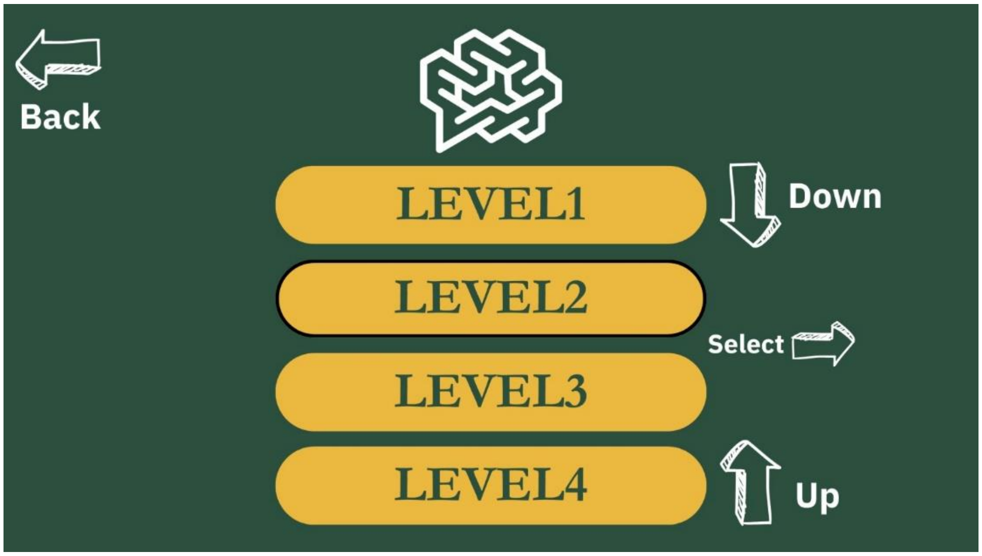

Figure 6.8 levels menu

And here is a snapshots from the game levels

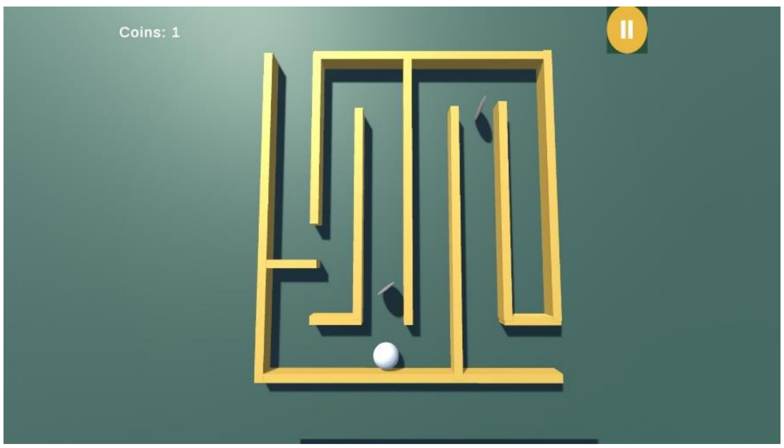

Figure 6.9 level 1 of skelly

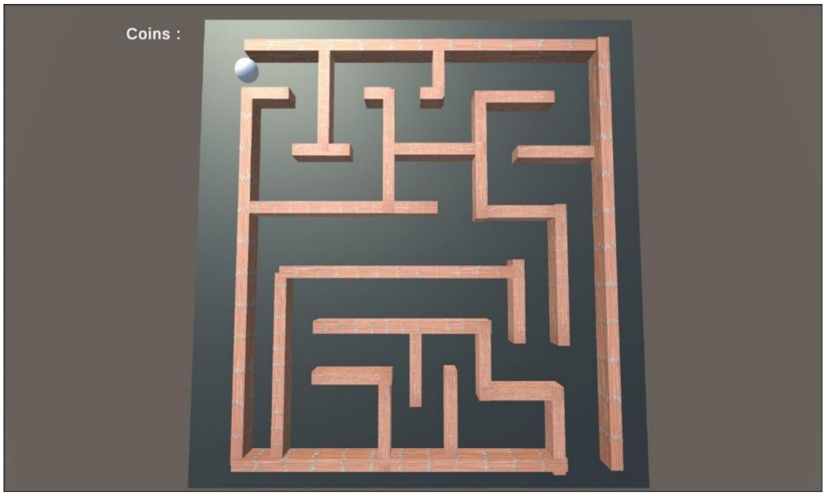

Figure 6.10 level 2 of skelly

Figure 6.11 winning screen

Chapter 7

Conclusions and next steps

7.1 Conclusion

- The "Skelly" project integrates Brain-Computer Interface (BCI) technology into gaming, allowing users to control characters using EEG signals. This innovation enables players to navigate through mazes using thoughts alone.

- "Skelly" provides an accessible platform for individuals with disabilities or those undergoing rehabilitation, offering a new way to interact with digital environments without traditional input devices.

- Utilizing the BCI Competition IV dataset 2a, the ATCNet model was trained to interpret EEG data. This deep learning model achieved 87.16% accuracy in translating users' thoughts into game commands, demonstrating reliability and precision.

- "Besides being fun, 'Skelly' helps develop cognitive and motor skills, showcasing how BCI technology can engage users and improve their abilities."

- Skelly's success highlights BCI's potential in assistive tech and rehab. Future improvements could offer more control options, better accuracy, and additional sensory feedback, making BCI useful in daily life.

7.2 Future Work

1. Turn the game to real time mode once the availability of the headset
2. Increase the number of levels
3. Add more challenges in the game
4. Add some improvements in movement algorithm

References

1. Tangermann, M., Müller, K. R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., ... & Blankertz, B. (2019). Review of the BCI competition IV. *Frontiers in neuroscience*,
2. Liao, W., Li, J., Zhang, X., & Li, C. (2023). Motor imagery brain–computer interface rehabilitation system enhances upper limb performance and improves brain activity in stroke patients: a clinical study. *Frontiers in Human Neuroscience*
3. Li, F., He, F., Wang, F., Zhang, D., Xia, Y., & Li, X. (2020). A novel simplified convolutional neural network classification algorithm of motor imagery EEG signals based on deep learning. *Applied Sciences*,
4. Sadiq, M. T., Yu, X., Yuan, Z., Zeming, F., Rehman, A. U., Ullah, I., ... & Xiao, G. (2019). Motor imagery EEG signals decoding by multivariate empirical wavelet transform-based framework for robust brain–computer interfaces.
5. Majidov, I., & Whangbo, T. (2019). Efficient classification of motor imagery electroencephalography signals using deep learning methods. *Sensors*,
6. Wang, X., Hersche, M., Magno, M., & Benini, L. (2024). Mi-bminet: An efficient convolutional neural network for motor imagery brain–machine interfaces with eeg channel selection. *IEEE Sensors Journal*.
7. Hassanpour, A., Moradikia, M., Adeli, H., Khayami, S. R., & Shamsinejadbabaki, P. (2019). A novel end-to-end deep learning scheme for classifying multi-class motor imagery electroencephalography signals
8. Xu, J., Zheng, H., Wang, J., Li, D., & Fang, X. (2020). Recognition of EEG signal motor imagery intention based on deep multi-view feature learning.

9. Mahamune, R., & Laskar, S. H. (2021). Classification of the four-class motor imagery signals using continuous wavelet transform filter bank-based two-dimensional images. *International Journal of Imaging Systems and Technology*
10. Hassan, E., Hossain, M. S., Saber, A., Elmougy, S., Ghoneim, A., & Muhammad, G. (2024). A quantum convolutional network and ResNet (50)-based classification architecture for the MNIST medical dataset. *Biomedical Signal Processing and Control*,
11. Li, M., Li, J., Song, Z., Deng, H., Xu, J., Xu, G., & Liao, W. (2024). EEGNet-based multi-source domain filter for BCI transfer learning. *Medical & Biological Engineering & Computing*,
12. Deng, H., Li, M., Li, J., Guo, M., & Xu, G. (2024). A robust multi-branch multi-attention-mechanism EEGNet for motor imagery BCI decoding. *Journal of Neuroscience Methods*, *405*, 110108
13. Fumanal-Idocin, J., Vidaurre, C., Fernández, J., Gómez, M., Andreu-Perez, J., Prasad, M., & Bustince, H. (2024). Supervised penalty-based aggregation applied to motor-imagery based brain-computer-interface. *Pattern Recognition*,
14. More, V., & George, K. (2024, February). Incorporating Motor Imagery-Controlled Gaming into Paralysis Rehabilitation. In *2024 IEEE First International Conference on Artificial Intelligence for Medicine, Health and Care (AIMHC)* (pp. 72-77). IEEE
15. Barron, E. N. (2024). *Game theory: an introduction*. John Wiley & Sons.
16. I. Ahmed, G. Jeon, and F. Piccialli, “From artificial intelligence to explainable artificial intelligence in industry 4.0: a survey on what, how, and where,” IEEE Trans. Ind. Informatics, vol. 18, no. 8, pp. 5031–5042, 2022.
17. H. Altaheri et al., “Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: a review,” Neural Comput. Appl., pp. 1–42, 2021.
18. D. Zhang, K. Chen, D. Jian, and L. Yao, “Motor imagery classification via temporal attention cues of graph embedded EEG signals,” IEEE J. Biomed. Heal. informatics, vol. 24, no. 9, pp. 2570–2579, 2020.

19. G. A. Altuwaijri, G. Muhammad, H. Altaheri, and M. Alsulaiman, "A Multi-Branch Convolutional Neural Network with Squeeze-and-Excitation Attention Blocks for EEG-Based Motor Imagery Signals Classification," Diagnostics, vol. 12, no. 4, p. 995, 2022.
20. S. U. Amin, H. Altaheri, G. Muhammad, M. Alsulaiman, and W. Abdul, "Attention based Inception model for robust EEG motor imagery classification," in 2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2021, pp. 1–6.
21. J. Xu, H. Zheng, J. Wang, D. Li, and X. Fang, "Recognition of EEG signal motor imagery intention based on deep multi-view feature learning," Sensors, vol. 20, no. 12, p. 3496, 2020
22. T. Liu and D. Yang, "A Densely Connected Multi-Branch 3D Convolutional Neural Network for Motor Imagery EEG Decoding," Brain Sci., vol. 11, no. 2, p. 197, 2021.