CNN-LSTM-Based EEG Eye Movement Classification for Assistive IoT Control in Locked-In Syndrome

University of Turku
Department of Computing
Master of Science (Tech) Thesis
Data Analytics
December 2025
Shoaib Azad Tonmoy

Supervisors:
Tuukka Panula (University of Turku)
Katri Karhinoja (University of Turku)

UNIVERSITY OF TURKU
Department of Computing

Shoaib Azad Tonmoy: CNN-LSTM-Based EEG Eye Movement Classification for Assistive IoT Control in Locked-In Syndrome

Master of Science (Tech) Thesis, 61 p.
Data Analytics
December 2025

Locked In Syndrome (LIS) is a rare neurological condition in which individuals remain fully conscious but experience near total paralysis, often retaining only limited eye movement. For such patients, even simple interaction with their surroundings can be profoundly challenging. This thesis investigates the feasibility of a low cost, non invasive brain–computer interface (BCI) capable of translating voluntary eye movements, detected through electroencephalography (EEG), into control commands for Internet of Things (IoT) devices.

A publicly available consumer grade EEG dataset was utilised, focusing on saccadic eye movement paradigms of varying complexity. The Muse S2 headset, with four frontal and temporal electrodes, provided the recording framework. The preprocessed EEG data provided by the dataset were directly segmented using a sliding-window approach for feature extraction and model training. Two feature extraction strategies were considered: handcrafted statistical and spectral features for classical machine learning models, and end to end feature learning via a hybrid Convolutional Neural Network and Long Short Term Memory (CNN+LSTM) architecture.

Comparative experiments demonstrated that the CNN+LSTM model achieved the highest performance, with accuracies of 75.9% on structured Level 1 Saccades data and 69.0% on the more complex Level 2 Saccades data using a four class command configuration. The trained model was integrated into a simulated IoT environment, enabling eye movement driven control of common household devices such as lights, fans, and televisions.

The results confirm that reliable eye movement classification is feasible with consumer grade EEG and deep learning methods, offering a practical pathway toward accessible assistive technology for individuals with severe motor impairments. The proposed system provides a foundation for future work involving real time deployment, expanded device control, and integration with user specific calibration to enhance accuracy and usability in real world settings.

Keywords: EEG, BCI, Eye-movement classification, Saccades, CNN–LSTM (hybrid deep learning), LIS, Assistive IoT control, Consumer-grade EEG

Contents

List of Figures

List of Tables

1 Introduction

Locked-In Syndrome (LIS) is a severe neurological condition that is incredibly rare. It is characterized by quadriplegia and anarthria, which leads to the loss of voluntary motor control over nearly all skeletal muscles. The main exception is eye movements, which are typically limited to vertical motion and blinking. Cognitive processing and consciousness are wholly unaffected by this substantial physical impairment. The primary cause of the condition is brainstem strokes, specifically infarction of the ventral pons. However, it may also result from traumatic brain injury, tumors, or progressive neurological disorders, such as ALS. Patients with LIS are fully cognizant but incapable of communicating through conventional means, which leads to a substantial reduction in quality of life and complete reliance on their caregivers. LIS is predominantly categorized into three subtypes: classical LIS, which is characterized by retained vertical eye movement and blinking; incomplete LIS, in which patients exhibit limited voluntary movement; and total LIS, in which all voluntary movement, including eye movement, is entirely absent. The necessity for non-invasive, efficient, and accessible assistive technology is underscored by the existence of substantial communication barriers with patients. Traditional communication devices for LIS patients, such as eye-tracking systems, head switches, or partner-assisted scanning boards, may require residual motor ability, rendering them unsuitable for individuals with total body paralysis. These tools are not appropriate for patients who have completely lost voluntary eye movement. In such cases, eye-tracking systems cannot function, and the EEG-based approach used in this thesis would also be ineffective because it relies on detecting ocular activity. Additionally, the implementation of these devices in domestic or institutional settings may be technically difficult, substantially invasive, or exceedingly expensive. Although LIS is traditionally divided into classical, incomplete, and total forms, this study focuses on individuals who retain some degree of voluntary eye movement. This corresponds mainly to classical LIS and the less severe forms of incomplete LIS, as the proposed method relies on detecting intentional ocular activity from EEG signals. Patients with total LIS, who lack any measurable eye movement, fall outside the scope of this work. The aim of this work is to examine whether eye-movement activity recorded with a consumer-grade EEG device can be reliably classified and used as the basis for simple control commands in an assistive context.

1.1 Background

LIS was first described in 1966 as a "false coma due to supranuclear motor deafferentation" [1]. Although individuals with LIS retain full awareness of their surroundings and remain cognitively intact, they are unable to move or speak because of bilateral disruption of motor pathways in the ventral brainstem, most frequently caused by stroke, which accounts for more than half of all cases [2]. Plum and Posner formalised the clinical definition in 1972, classifying LIS as one of the most severe neurological disorders in terms of functional limitation. Vertical gaze and eyelid movements are often preserved, but voluntary control over the limbs and speech musculature is almost entirely lost [1].

The preserved state of consciousness, combined with the inability to express needs or intentions through voluntary movement, creates an extreme barrier to communication. This situation highlights the need for assistive tools that can bypass the damaged motor pathways and translate remaining physiological signals—whether brain activity or ocular movements—into functional communication channels [2].

Patients typically retain vertical eye movement, which becomes the primary method of intentional control [3]. The characteristic lesion of LIS is located in the pons, interrupting corticospinal and corticobulbar connections while sparing the cerebral cortex. As a result, individuals experience near-total paralysis but remain fully aware and capable of thought. In severe cases, even eye movements may be minimal or absent, although higher cortical functions continue to operate normally. EEG recordings often show preserved patterns of cortical activity and regular sleep–wake cycles, further reinforcing that cognition remains unaffected. Respiratory compromise may occur depending on the extent of brainstem involvement, occasionally requiring ventilatory support. These clinical features explain why assistive technologies must rely on signals that do not depend on intact motor pathways [4].

The dissociation between intact cognitive processing and severe motor impairment is clearly visible in neuroimaging. Figure 1.1 presents a sagittal MRI image of a patient with LIS, where lesions in the ventral pons can be observed. This region connects the motor cortex to the spinal cord, and its damage disconnects voluntary motor commands from the rest of the body. Despite this loss of movement, the cerebral cortex remains structurally unharmed, allowing individuals to think normally even though they cannot act on those thoughts [3]. This preserved cortical function forms the basis for non-muscular communication methods, including EEG-based systems.

Because motor output pathways are interrupted rather than cortical areas, LIS leaves higher-order cognitive abilities intact. This distinction is central to understanding why several neurological disorders with similar motor limitations—such as cerebral palsy, muscular dystrophy, multiple sclerosis, Parkinson's disease, and brain tumours—face comparable communication barriers despite differing underlying causes [2]. The common challenge across these conditions is the absence of

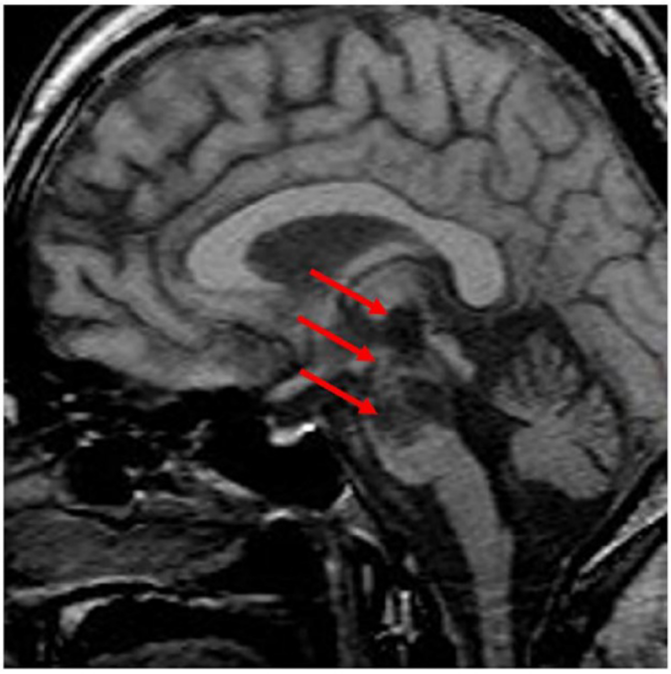

Figure 1.1: MRI of a patient with LIS showing lesions in the brainstem [3]

voluntary muscle control, which necessitates assistive systems capable of interpreting non-muscular signals.

EEG provides one such pathway. Eye blinks and lateral eye movements produce characteristic potentials that appear prominently in frontal EEG channels [5]. Although these components are considered artifacts in typical EEG analysis, they are valuable in situations where eye movement becomes the only controllable signal, as in LIS. For this reason, EEG-based brain–computer interface (BCI) systems offer a promising avenue for restoring communication. In the present study, these ocular EEG components serve as the primary features for distinguishing intentional eye-movement patterns [6].

A BCI—also referred to as a Brain–Machine Interface (BMI)—is a system that allows users to control external devices using signals derived from brain activity rather than muscular movement. BCIs have been applied in contexts such as advanced prosthetic control, where they enable actions not achievable through conventional

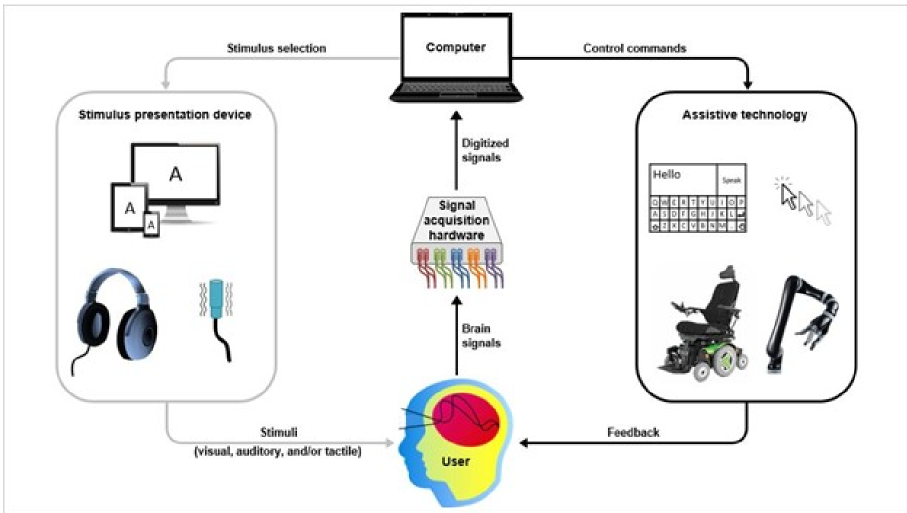

Figure 1.2: Diagram of human brain and computer interface
[11]

assistive technologies [7]. The core objective of a BCI is to identify patterns in neural activity that correspond to a user's intention [8]. Brain responses to stimuli generate voltage fluctuations that reflect synaptic activity, and these potentials can be captured non-invasively using surface electrodes. Various techniques exist for recording neural signals, including invasive methods such as electrocorticography (ECoG) and non-invasive alternatives such as EEG, magnetoencephalography (MEG), and functional near-infrared spectroscopy (fNIRS) [9], [10]. EEG is used in this thesis because it is portable, safe, and highly suitable for detecting eye-movement-related activity. An overview of the basic components of a brain–computer interface is illustrated in Figure 1.2.

The EEG signal results from the summed electrical activity of neuronal populations and forms the foundation for most non-invasive BCIs [8]. By relying on cortical activity rather than residual muscle function, EEG-based systems offer communication possibilities for individuals who cannot use conventional modalities such as speech or gestures [12]. Motor neuron damage prevents voluntary commands from reaching the muscles, making EEG a practical alternative for decoding user intent. Although BCIs face technical limitations related to signal quality and the inherent challenges of EEG, advancements in device design have improved usability and reduced setup complexity [8].

EEG activity is typically described in terms of frequency bands—delta, theta, alpha, beta, and gamma [7]. These rhythms vary with cognitive state, attention, and task engagement [8]. Standardized electrode placement systems, such as the international 10–20 configuration, allow consistent sampling of occipital, parietal, central, and frontal regions. Research aimed at improving classification performance has investigated reducing electrode count while focusing on regions relevant to visual and attentional processing.

EEG signals naturally contain various artifacts, including muscle activity, motion-related noise, environmental interference, and ocular events. Removing these unwanted components is essential for accurate analysis, with Independent Component Analysis (ICA) and filtering being common tools for isolating genuine neural activity. In LIS-oriented EEG studies, experimental designs often rely on structured visual tasks with repeated cues to elicit stable, classifiable patterns. Such protocols help differentiate intentional activity from background rhythms, thereby supporting more reliable decoding of user intent [13].

1.2 Problem Description

Patients living with LIS face extreme restrictions in daily interaction because their residual motor activity is often too limited to operate conventional assistive tools. Many existing communication systems—such as camera-based eye trackers, head switches, or specialized scanning interfaces—require precise movement control, stable posture, or costly technical setups, which are not always feasible in home environments. As a result, a large number of individuals with severe motor impairment remain dependent on caregivers for even basic tasks.

In contrast, EEG captures the electrical activity generated by extraocular muscles and the surrounding cortical regions, which can be present even when voluntary eye control is weak or inconsistent. This makes EEG-based methods more suitable for individuals with severe motor impairment. However, most EEG-driven communication solutions still face barriers to practical use. Many depend on laboratory-grade equipment, rigid stimulus protocols, or interfaces that support only very simple forms of communication, making them unsuitable for everyday control of common devices. Furthermore, existing systems often struggle with noisy recordings, inconsistent signal patterns across users, and difficulties in distinguishing subtle eye-related EEG activity from background brain rhythms.

These limitations highlight the need for a more accessible solution that can interpret eye-movement-related EEG patterns at a level suitable for functional device control. The central challenge addressed in this study is the accurate detection and classification of these EEG patterns using consumer-grade hardware. Meeting this challenge requires robust preprocessing, reliable classification methods, and a design that maintains usability outside controlled laboratory settings. The overarching problem, therefore, is to develop a system capable of converting complex EEG signals into practical, meaningful commands that can support everyday interaction for individuals with profound motor disabilities.

1.3 Aim and Objectives of the Study

The main aim of this study is to determine whether eye-movement patterns can be accurately classified using consumer-grade EEG recordings in order to support communication and interaction for individuals with severe motor impairment. Although EEG-based BCIs have been widely explored in controlled laboratory environments, their practical use with low-cost equipment remains limited. This research therefore focuses on examining the feasibility of using frontal EEG channels to detect distinct ocular patterns and exploring whether these signals can support functional, real-world interaction for individuals who cannot rely on conventional motor-based input methods. The study adopts an applied perspective, seeking not only to evaluate the classification performance of the EEG-based approach but also to illustrate how the classified eye-movement signals could be mapped to basic control actions within an assistive-technology framework.

To pursue this aim, the study first examines how directional eye movements manifest in EEG recordings from the four electrodes available on the Muse S2 headset. These signals are known to carry strong ocular components, and the study investigates whether these components can be consistently distinguished from background neural activity across different trial conditions. A key objective is to isolate the EEG segments that best represent horizontal and vertical eye movements and to determine the extent to which these segments can be used to form distinct, classifiable patterns. In this study, the analysis is carried out on data that have already undergone preprocessing, allowing the focus to be placed on evaluating the classification performance rather than on developing a new preprocessing pipeline.

Another objective is to design and evaluate classification models capable of learning these patterns with dependable accuracy. Both machine-learning and deep-learning approaches are considered in order to compare their strengths and limitations when working with low-density EEG data. Particular attention is given to hybrid architectures that combine convolutional and recurrent layers, as these models are well suited to capturing both the spatial structure and temporal evolution of EEG signals. The performance of these models is assessed across datasets of varying complexity to understand how classification reliability changes under different task demands. Beyond classification accuracy, the study also aims to demonstrate how the resulting outputs can be mapped to functional control actions. A conceptual demonstration is used to illustrate how classified eye-movement directions could support simple control actions in smart-home environments.

Together, these aims and objectives outline a complete pathway from EEG acquisition to practical IoT control. By integrating signal analysis, model development, and functional demonstration, the study evaluates whether a consumer-grade EEG device can support an accessible and low-cost approach to assistive interaction. The broader intention is to assess whether such a system could form the basis for future communication or environmental-control tools designed for individuals with severe motor impairments.

1.4 Significance of the Study

The significance of this study lies in its examination of whether consumer-grade EEG signals can be used to distinguish simple eye-movement patterns that may support communication for individuals with severe motor impairment. While many assistive technologies require consistent motor control or specialised hardware, EEG-based systems offer an alternative pathway that does not rely on visible movement. By analysing eye-movement-related components recorded from low-density frontal electrodes, this work provides insight into the extent to which such signals can be reliably classified using lightweight equipment.

Although the study does not implement a full assistive device, its findings contribute to the ongoing evaluation of consumer-grade EEG as a practical tool for basic interaction. The conceptual demonstration included in this thesis illustrates one possible way that classified signals could be mapped onto simple control actions, offering a foundation for future development rather than a complete solution. In this regard, the study's significance is primarily methodological: it assesses feasibility, highlights limitations, and identifies directions for further research in the context of accessible EEG-based communication systems.

1.5 Research Gap

Despite the growing interest in EEG-based interaction methods, there is limited evidence on whether consumer-grade headsets can reliably distinguish multiple eye-movement directions using only a small number of frontal electrodes. Most prior work has focused either on laboratory-grade devices or on simplified tasks, leaving uncertainty about the feasibility of performing multi-class ocular classification with lightweight, accessible hardware. This gap motivates the present study, which examines the classification performance achievable using low-density consumer EEG.

A major gap relevant to this study is the absence of real-time EEG acquisition and testing. Because the current work relies on publicly available pre-recorded datasets, it cannot fully assess how fluctuations in user behavior, environmental conditions, and moment-to-moment signal quality might affect system reliability. Real-time testing is essential for evaluating latency, responsiveness, and the user's ability to adapt to the interface—factors that play a decisive role in the practical deployment of assistive technologies.

Another limitation in current research is the relatively small number of studies that examine end-to-end pipelines capable of linking eye-movement classification directly to functional device control. Many investigations focus on classification accuracy alone, without demonstrating how these outputs integrate into systems that users can operate independently. This leaves a gap between technical performance in controlled datasets and real-world usability in home or clinical environments.

There is also a broader methodological gap concerning generalizability across users. EEG signals vary widely between individuals, and most existing studies either rely on subject-specific training or limit their evaluation to small participant groups. More work is needed to understand how well classification models trained on consumer-grade EEG data can adapt to new users without extensive calibration.

Taken together, these gaps underscore the need for research that not only evaluates classification performance but also explores the feasibility of connecting consumer-grade EEG signals to practical, reliable IoT control mechanisms. The present study seeks to contribute to this area by examining whether a low-cost EEG device can support a functional interaction framework, even in the absence of real-time data collection

1.6 Writing Process and Use of AI Tools

In the preparation of this thesis, artificial intelligence (AI) tools were employed solely to support the writing process in a limited and transparent manner. Specifically, language assistance tools such as OpenAI's ChatGPT were used to improve sentence structure, enhance grammatical correctness, and maintain clarity across sections. No technical content, experimental design, data analysis, or interpretation was generated or influenced by AI. All conceptual development, dataset handling, algorithm implementation, and critical evaluations were conducted independently by the author. The AI assistance served only as a supplementary aid for refining existing text and aligning the overall writing style with academic conventions. Care was taken to ensure that all content reflects the author's original understanding and contributions. Furthermore, no AI-generated text was used without verification, and all scientific claims, figures, and references were manually curated and validated against original sources.

2 Literature Review

2.1 Introduction

The chapter represents the impact of the most recent communication techniques for LIS patients. Potential BCI paradigms and previously tested BCIs developed systems. Several studies have been conducted on EEG-based BCIs, focusing on enhancing communication, control, and interaction for individuals with neurological impairments, particularly through the use of non-invasive EEG signal processing and machine learning techniques. It is a very effective communication technology that does not rely on neuromuscular or muscle pathways to accomplish communication, command, and, hence action. While thinking with intention, the subject generates brain signals that are converted to commands for an output device [8].

2.2 Related Work on ML-Based EEG Systems and IoT Systems

Machine learning (ML) is a critical component of the development of BCI systems that are designed to assist LIS patients by enabling communication and control without the use of muscle movement. The objective of ML in this study [14] on EEG-based MI-BCI systems is to identify brain patterns that are associated with a specific activity, without relying on traditional statistical methods. The application of machine learning to interpret EEG data, discern user intent, and improve system adaptability has been the subject of numerous studies. The EEG signals are feeble, susceptible to interference and noise, non-stationary for the same individual, and variable among various participants and sessions, according to this study report [15]. As a result, the development of a universal machine learning model for an EEG-based BCI system that is optimal across a variety of individuals, sessions, devices, and activities presents substantial challenges. The study in [15] demonstrates progress toward this objective through the implementation of a variety of machine learning methodologies, including active learning and transfer learning (TL). The research study [8] illustrates that machine learning algorithms can classify a variety of affective states by utilizing EEG-based BCI. Additionally, the authors [16] explore the identification of EEG paradigms, such as motor imagery, using machine learning and deep learning methodologies. LDA is a suitable classifier in BCI systems, notably for constrained training datasets. Three benchmark machine learning classifiers, namely SVM, k-NN, and ANN, are extensively employed in MI-BCI systems. Convolutional neural networks (CNNs) are the primary method used by deep learning to classify motor imagery signals in brain-computer interface devices. The studies reviewed here and related works primarily involve healthy participants rather than individuals with LIS.

This research [17] introduces a hybrid signal processing method that utilizes power spectral density and digital filters to extract EMG-like characteristics from EEG data. This method simplifies the administration of domestic appliances and other IoT devices. The prototype execution was successful, and the approach was centered on immobile individuals. A non-invasive communication system based on EEG was developed in this study [18] through the use of fuzzy logic pattern identification. Their solution included a 14-channel EEG apparatus, a tablet interface that demonstrated six fundamental demands, and a bidirectional communication system with caregivers via a mobile application. The device efficiently converted EEG data generated by eye movements and color concentration into auditory warnings and instantaneous notifications. The studies reported in [19] and [20] investigated the categorization of EEG responses to visual color stimuli using both conventional machine learning and CNN-based methods. Despite the fact that the research was not explicitly IoT-enabled, it emphasized real-time, user-centric interaction paradigms that are suitable for LIS patients, thereby laying the groundwork for an eventual integration with intelligent settings. These studies collectively offer effective methods for closing the communication gap for LIS patients through EEG-driven systems, thereby enabling the potential control of IoT devices and assistive communication. The IoT aims to simplify the process of data exchange among global systems by utilizing advanced communication technologies. IoT-based solutions have become a practical reality as a result of recent technological advancements. The contemporary digital landscape will be enriched by a variety of potential transformative benefits that the IoT will offer. The healthcare sector is a prospective area for leveraging the benefits of IoT in the future, as it will be more personalized and interconnected. This paper investigates the technological requirements for the development of healthcare applications, including the specific requirements for the implementation of comprehensive end-to-end solutions for each application. The survey investigates the fundamental application-specific requirements from the perspective of communication technology [21].

2.3 EEG-Based BCI Systems for LIS

EEG-based BCI systems designed for individuals with LIS typically rely on well-established paradigms such as P300, Steady-State Visual Evoked Potentials (SSVEP), and Slow Cortical Potentials (SCP). P300-based BCIs rely on detecting an event-related potential that appears around 300 ms after a user perceives a target stimulus, making them suitable for individuals who can maintain visual attention. SSVEP systems, in contrast, use visual flickers at specific frequencies, which evoke steady-state responses in the occipital cortex when the user focuses on a chosen stimulus. Hybrid P300–SSVEP systems combine both responses within a single interface, enabling higher accuracy and improved reliability, particularly for users with limited voluntary movement, such as individuals with LIS. These systems share several essential components: an EEG acquisition device, a stimulus or task presentation module that elicits measurable neural responses, a signal-processing pipeline to extract relevant features, and a classifier that translates these features into selectable commands. Their relevance to LIS has been highlighted in clinical and review studies [22]. P300-based BCIs are widely researched for LIS [23] because the evoked response remains detectable even in individuals with limited voluntary movement. SSVEP systems, which rely on frequency-specific visual stimuli, provide high information-transfer rates [24], whereas SCP-based approaches support slower but more deliberate communication [25]. Together, these paradigms form the foundation of current EEG-based communication systems for LIS, each with strengths and limitations depending on the individual's residual abilities.

EEG is a well-known and, in some cases, the most established brain imaging technique. It offers a comprehensive understanding of a variety of cognitive processes, such as thought, learning, perception, and affective arousal [26]. The Eye-Movement EEG system delineates a distinctive approach that employs a BCI to acquire EEG data from human participants during eye movement. The data is subsequently classified into six categories using a random forest classification algorithm [27].

There are some limitations in Real-Life Deployments. The System's functionality is restricted by certain constraints, such as the necessity for patient training and the requirement for the headpiece electrodes to remain consistently damp [18]. The proposed system is user-friendly, transportable, and enables patients to articulate themselves without the need for physical exertion or effort. Contemporary technical approaches are markedly sluggish, resulting in a time lag of 10-15 minutes for responses from a LIS patient to various stimuli, such as quips or remarks. Consequently, some limitations are also acknowledged in BCI techniques [7].

2.4 Available Databases

Most publicly available EEG datasets used in BCI research do not focus on eye-movement patterns, as many were created for tasks such as P300 spellers, motor imagery, or deception detection. These datasets are therefore unsuitable for the present study, which requires EEG signals that contain clear ocular activity. For this reason, the analysis in this thesis relies on a dataset specifically designed for eye-movement categorisation using frontal EEG channels, ensuring that the recorded signals are directly relevant to the intended classification task.

EEG data from ten participants who completed a Concealed Information Test (CIT) paradigm were utilised to analyse P300 event-related potentials [26]. The CIT protocol involved presenting known and unknown visual stimuli, and the recorded EEG signals were processed on a subject-wise, single-trial basis for classification and feature extraction. The data were acquired using a 16-channel EEG system. Although the original dataset was designed for P300 analysis, it has also been adapted in subsequent work to support the development of low-sensor communication interfaces based on real-time eye-movement categorisation using continuous wavelet transform features. The classification methods, which included support vector machines, linear discriminant analysis, and k-nearest neighbors, were evaluated in this study [28] using EEG datasets. In addition, studies such as [29] used publicly available EEG datasets or experimental EEG recordings that were collected for the purpose of analyzing visual stimuli, motor imagery, or specific cognitive tasks. The EEG signals from healthy participants were typically obtained at 128–256 Hz in these datasets, which ranged from low-channel to high-density electrode arrangements. No medical documentation of any psychiatric condition was found for the individual, and they had normal or corrected vision. A succinct summary of the entire experimental protocol is provided to participants. Individuals have submitted written consent for their involvement in this investigation prior to the initiation of the experimental method.

The present study uses the Consumer-Grade EEG-Based Eye Tracking dataset published by Afonso and Heinrichs (2025) [30]. This dataset contains simultaneous EEG and eye-tracking recordings collected from 113 participants across 116 sessions using the Muse S2 headband, which provides four frontal and temporal electrodes (AF7, AF8, TP9, TP10) with Fpz as the reference. The recordings were conducted under four controlled eye-movement paradigms—level-1 saccades, level-1 smooth pursuit, level-2 saccades, and level-2 smooth pursuit—designed to capture horizontal and vertical gaze shifts of varying complexity.

2.5 EEG Classification Algorithms

In BCI devices for patients with LIS, the proficient categorization of EEG data is essential. Support Vector Machines (SVM) and Random Forests (RF) are examples of conventional machine learning techniques that have been extensively employed. RF has consistently exhibited efficacy by utilizing factors such as energy distribution and fractal dimensions derived from Variational Mode Decomposition (VMD). The study reported in [19][20] applied Random Forest classifiers to distinguish between target and non-target visual stimuli in an ERP-like task. CNNs have recently become increasingly popular in the field of deep learning. Conventional methods are surpassed by convolutional neural networks, such as EEGNet, which can autonomously extract spatial-temporal information from EEG data.

SVM, KNN, Decision Trees are some of the traditional machine learning meth-

Table 2.1: Analysis of Related Work

Author Name	**Methodology**	**Purpose**	**Dataset**	**Result**
Shehieb et al. (2017)	Rule-based fuzzy logic classifier	Interprets eye movement signals captured via an Emotiv EPOC EEG headset	Custom data collected randomly	Clear command mapping through eye movements; enabled patients to communicate basic needs via mobile app
Tobias Moe (2021)	Subject-independent CNN (EEG-Net)	Classify EEG signals related to visual stimuli like colored icons	Public and custom datasets	Avg. accuracy of 60-65% in 4-class classification
Aman Kurapa et al. (2020)	EEG + hybrid digital filters (MATLAB); PSD analysis	Extract EMG from EEG for IoT control in immobile users	8 male subjects; EEG tasks (eye, hand, jaw movement)	EMG signals successfully extracted; enabled device control
Evangelos Antoniou et al. (2021)	EEG with band energy features; Random Forest classifier	Classify eye movements for EEG-based wheelchair control	10 subjects; 219 EEG samples with 6 eye directions	85.39% accuracy with Random Forest; best among tested models
Sara Åsly (2019)	EMD + ML classifiers; CNN with raw EEG	Classify RGB visual stimuli via EEG for BCI applications	7 participants; dataset with RGB color stimuli	each subject: 0.58(NB), all subject: 0.37(NB); gray vs color: 0.99

ods. SVM, which was initially devised for data categorization, has gained significant prominence in BCI research. A binary classifier that is frequently employed in supervised learning is the SVM. The objective is to minimize the expected risk while identifying the optimal hyperplane for partitioning the training data. This study evaluates the proposed system in comparison to the following classification techniques: Naïve Bayes, Bayesian Networks, KNN, MLP, SVM, Decision Tree, and Bagging [27]. CNN, LSTM, EEGNet are some of the deep learning methods. A growing number of researchers regard deep learning as a significant methodology for the detection of emotions in EEG. The CNN architecture described in this study [31] is based on the EEGNet framework, which is publicly available. EEGNet is a CNN architecture that is expressly intended for the categorization of EEG data and is concise [32].

Model evaluation metrics include Accuracy, F1, Specificity. Evaluation metrics are quantitative measures used to assess the performance and effectiveness of a statistical or machine learning model. These metrics provide insights into how well the model is performing and help in comparing different models or algorithms.

Subject-Specific classification involves training and testing a model using data from the same individual. These models generally achieved higher accuracy. Some subjects' models (e.g., subjects 7, 8, 18, 32, 34) performed particularly well, showing that personalized models can effectively learn individual EEG patterns. Cross-Subject classification involves training a model using data from multiple participants and testing it on a new, unseen subject. While this approach is more scalable for real-world applications, its accuracy was noticeably lower than that of subject-specific models [30]. This drop in performance is due to variability in EEG signals across individuals, such as brain anatomy, signal noise, and eye movement differences.

2.6 Signal Processing for Eye Movement Detection

The systematic identification of reliable neural patterns from chaotic biosignals is a critical component of eye movement recognition in EEG-based BCI systems, which is facilitated by signal processing. The initial phase entails the acquisition of raw EEG data using non-invasive scalp electrodes, which are typically arranged in accordance with the international 10-20 system. The focus is on the frontal and occipital regions, where eye movements induce substantial brain activity [27]. Preprocessing is essential for the evaluation of EEG data, particularly in BCI systems that detect eye movements. The primary objective is to mitigate noise and eradicate unwanted artifacts that are the result of environmental influences, muscle contractions, and ocular blinks. Standard techniques involve the use of bandpass filters, which are typically configured between 0.5 Hz and 30 Hz, in order to preserve frequency ranges that are pertinent to ophthalmic signals. Furthermore, the utilization of notch filters at 50 or 60 Hz is employed to reduce power line noise [17].

Bandpass and notch filters are the most frequently used filters in the processing of EEG data [28]. A band-pass filter transmits sounds within a specific range while attenuating frequencies that exceed that range. The unprocessed data and the artifact-free data are both subjected to a band-pass filter. A notch filter is a filter that is substantially attenuated at a specific frequency and has a very narrow bandwidth. The optimal frequency response of a notch filter is demonstrated in this research study [17].ICA method is used to decompose blended signals into their independent, non-Gaussian constituents. It aims to identify a linear transformation of data that maximizes the statistical independence of the components [33]. The utilization of ICA is a critical methodology in this investigation. This entails blind source separation algorithms that linearly divide EEG data into distinct source components.

Shallow features are attributes that are painstakingly constructed in a variety of analytical domains, such as the frequency domain, time domain, and time-frequency domain [34]. The classifier type and the features selected during the feature extraction phase should be determined in accordance with the specific BCI type. SSVEP BCIs are more appropriate for frequency domain features, while P300 is suitable for time domain or time-frequency domain features, such as wavelets [28]. SSVEP signals consist of continuous oscillatory responses that occur at the same frequency as the flickering visual stimulus, which makes their information content primarily frequency-specific. For this reason, SSVEP BCIs are typically analysed using frequency-domain features such as power spectra or harmonic components. In contrast, the P300 is an event-related potential. Since this response is transient and has a well-defined temporal shape rather than a stable frequency pattern, it is more appropriately captured through time-domain measurements or time-frequency techniques such as wavelet transforms, which preserve both timing and signal morphology.

Event-related potentials (ERPs) are voltage fluctuations in the EEG signal that occur in response to specific sensory, cognitive, or motor events, and they are characterised by their precise timing and stereotyped waveform patterns. Within the broader field of EEG-based BCI research, ERPs form one of the six commonly recognised paradigms explored in recent transfer-learning studies, alongside approaches such as motor imagery and SSVEP [15]. These paradigms represent distinct frameworks through which neural activity can be interpreted for communication and control purposes.

2.7 IoT Integration in BCI Systems

Smart home systems integrated with BCIs can perform basic tasks like turning lights on/off, controlling the thermostat, or sending alerts to caregivers. These systems usually use a lightweight EEG headset like Emotive and connect to tablets or computers that interpret the signal and relay it to IoT devices via wireless protocols or databases. Several studies, such as [18], have developed IoT-enabled communication platforms that rely on EEG signals, but these systems were tested primarily with healthy participants rather than individuals with LIS. Although not evaluated in clinical LIS populations, these works demonstrate how basic EEG-derived commands can be integrated with mobile applications or smart-home devices using lightweight acquisition hardware. They developed communication platforms where patients could express basic needs such as food, sleep by focusing on screen elements. The selected command, detected using fuzzy logic from EEG signals, is sent to an Android app as a pop-up and voice alert. Other systems use similar techniques to control smart bulbs, wheelchairs, or even appliances.

Key challenges in integrating BCIs with IoT include real-time responsiveness, reliable wireless data transfer using Firebase or Bluetooth, and energy constraints in wearable devices. Signal noise, variability, and the need for continuous power to maintain connectivity and processing also pose limitations.

BCIs require individual calibration to adjust to each user's unique brain signal patterns. This step is essential for accuracy but can be time-consuming and mentally taxing, especially for patients. Sustained use of BCIs can cause fatigue, reducing performance and reliability. In the EEG experiments from Tobias Moe's project, minimizing mental effort was a design priority to avoid exhausting the user while still enabling communication. Effective BCIs use feedback to confirm that a command was received. For instance, in [18], visual feedback (GUI) and auditory messages ensured the patient knew their input was acknowledged, improving confidence and usability. BCI systems designed for clinical use must prioritize informed consent, data privacy, and emotional well-being. Since users are often in vulnerable conditions, ethical concerns also include long-term psychological impact and the potential for overdependence on technology.

3 Methodology

3.1 Dataset Description

This research employs the Consumer-Grade EEG-Based Eye Tracking Dataset, including synchronized EEG and webcam-based eye-tracking records. The dataset was created to enhance research on EEG-based eye tracking (EEG-ET) and to aid in the development of practical, cost-effective BCI systems. The collection comprises 113 subjects documented throughout 116 sessions, yielding roughly 11 hours and 45 minutes of EEG and eye-tracking data [30].

Participants were primarily healthy adults, with a distribution of 91 males (81%) and 22 females (19%). Most were right-handed (88%), while 11% were left-handed, and 1% ambidextrous. Approximately 48% of participants used vision correction, and a small fraction reported neurological conditions (5%) or color blindness (2%). All participants voluntarily participated without financial compensation and provided informed consent in accordance with ethical guidelines [30]. The dataset consists of four experimental paradigms, progressively increasing in complexity:

- Level-1 Saccades (simple left-right-up-down eye movements)
- Level-1 Smooth Pursuit (smooth movement along a single axis)
- Level-2 Saccades (random saccades across a 5×5 grid)
- Level-2 Smooth Pursuit (smooth movement along randomized complex trajec-

tories)

This thesis examines data from both the Level-1 and Level-2 saccades paradigms. The Level-2 task requires participants to track a stimulus that moves across 25 randomized positions within a 440 mm $\times$ 220 mm display area at 1.5-second intervals, whereas the Level-1 task consists of more basic directional saccades with a smaller set of fixed target positions. This simulates complex menu navigation scenarios suitable for IoT device control using eye movements [30].

EEG signals were recorded using the Muse S2 headband, a consumer-grade EEG headset equipped with four dry electrodes (TP9, TP10, AF7, AF8) and a reference electrode at Fpz, sampled at 256 Hz. Simultaneously, a Logitech StreamCam webcam captured eye gaze positions at 60 frames per second using GazePointer software. Data streams from the EEG device, eye tracker, and stimulus presentation software were synchronized via the Lab Streaming Layer (LSL) protocol [30].

Participants were seated approximately 60 cm from a 24-inch monitor (2560$\times$1440 resolution). Sessions were conducted in naturalistic classroom environments to reflect realistic, non-laboratory conditions. Each session included a calibration phase, followed by an experimental phase where participants were instructed to follow on-screen targets as accurately as possible using eye movements. During Level-2 Saccades, the target moved unpredictably between 25 grid locations with visual cues (e.g., countdown circles, directional lines) to assist gaze accuracy [30].

The dataset is available in both CSV and XDF file formats. Each session file contains time-synchronized recordings of several key data streams. These include the EEG signals recorded from four electrode positions: TP9, AF7, AF8, and TP10, labeled respectively as EEG_TP9, EEG_AF7, EEG_AF8, and EEG_TP10. Alongside these EEG channels, the dataset also provides two-dimensional gaze coordinates (Gaze_x and Gaze_y) corresponding to the participant's eye movement during each trial. In addition, the position of the visual stimulus presented on the screen is recorded through the Stimulus_x and Stimulus_y fields. All of these data streams are indexed with precise timestamps, allowing for accurate temporal alignment during analysis.

The preprocessed versions of the dataset incorporate several key signal conditioning steps. First, a bandpass filter ranging from 0.5 to 40 Hz was applied to retain relevant EEG frequencies while suppressing slow drifts and high-frequency noise. In addition, notch filters targeting both 50 Hz and 60 Hz frequencies were used to eliminate power-line interference common in EEG recordings. To address gaps in the data, missing values were imputed using Kalman smoothing in combination with Seasonal AutoRegressive Integrated Moving Average (SARIMA) models, as described in [35]. This structured design facilitates machine learning research while addressing challenges of consumer-grade EEG data quality.

Unlike prior datasets designed solely for medical research, this dataset uniquely enables the exploration of real-time IoT control systems. The Level-2 Saccades paradigm particularly simulates eye-based interaction with menu-driven interfaces, representing a practical scenario for assistive technology targeting individuals with limited or no motor control.

3.2 Data Preprocessing

All experiments in this thesis were carried out on the preprocessed EEG files provided with the Consumer-Grade EEG-Based Eye Tracking dataset. The dataset authors released, alongside the raw recordings, a set of “cleaned” CSV files in which the main artifact removal and signal conditioning steps had already been applied [30]. After loading the preprocessed files, a study-specific preprocessing pipeline was implemented in Python to transform the continuous signals into fixed-length examples suitable for CNN–LSTM training. Only the four Muse S2 EEG channels (TP9, AF7, AF8, TP10) were retained, as these correspond to the frontal and temporal regions most sensitive to eye-movement-related artifacts. For each recording session, only trials belonging to the Saccades paradigms were selected. Separate preprocessing scripts were used for Level-1 Saccades and Level-2 Saccades, but the core steps were identical: channel selection, epoching with an overlapping sliding window, label generation, and normalization.

The first study-specific step was epoch segmentation. The continuous EEG stream for each participant was divided into short time windows using a sliding-window procedure with overlap, implemented exactly as in the python notebooks. For Level-1 Saccades, which contains simple gaze shifts in four cardinal directions, windows of 2 seconds (512 samples at 256 Hz) were used. For Level-2 Saccades, which contains more irregular and spatially distributed target jumps, windows of 3 seconds (768 samples) were chosen to capture more temporal context around each stimulus. Although each stimulus in the Level-2 Saccades paradigm was displayed for 1.5 seconds, a 3-second window was selected to ensure that the full temporal evolution of the eye movement was captured. Level-2 Saccades contain larger and more irregular gaze shifts than the Level-1 paradigm, and participants often require additional time to initiate, complete, and stabilize their gaze position. Extending the window, therefore, includes the preparatory activity, the saccadic event itself, and the post-movement settling period, all of which contribute discriminative information to the classifier. In both cases, the window advanced with a 75% overlap, so successive epochs shared most of their samples. This choice increased the number of training examples and allowed the model to capture transitions between gaze positions rather than only isolated snapshots.

Class labels were then generated from the stimulus coordinates included in the preprocessed dataset. These coordinates specify the on-screen position of the target at every time point and were converted into discrete direction labels for classification. In the Level-1 Saccades paradigm, each epoch was assigned one of four labels—Left, Right, Up, or Down—according to the predominant stimulus movement within the window. For the Level-2 Saccades paradigm, the original 5×5 grid of target locations was reduced to four broader directional zones. This was done by grouping grid positions according to their horizontal and vertical offset from the centre of the grid: positions to the left of the central column were assigned to the Left Control class, positions to the right to the Right Control class, positions above the central row to the Up Control class, and those below it to the Down Control class. Each epoch received the label corresponding to the region in which the stimulus appeared most frequently during the epoch window. This approach mirrors a simple menu-based control structure in which each direction triggers a distinct IoT command.

Once epochs and labels were created, the EEG segments were normalized to reduce inter-subject variability and stabilize network training. For every epoch, each channel was transformed using z-score normalization: the channel mean within that epoch was subtracted and the result divided by the corresponding standard deviation. This operation centred the data around zero and scaled it to unit variance on a per-epoch, per-channel basis, which helped the CNN–LSTM model converge more reliably across different participants and recording sessions.

The final set of preprocessed and normalized epochs was divided manually into training and test sets using a stratified split to preserve class balance across both subsets. Each example was stored as a three-dimensional array with shape (samples, channels, 1) for compatibility with the one-dimensional convolutional layers, and the class labels were one-hot encoded for multi-class softmax output. A stratified 80/20 split was applied so that all classes were proportionally represented in both training and test data. No additional filtering, artifact correction, or feature extraction was performed at this stage; the CNN–LSTM architecture was designed to learn spatial–temporal patterns directly from these minimally processed yet already denoised EEG epochs.

Stage	Description	Purpose
Channel Selection	Extraction of the four Muse S2 channels (TP9, AF7, AF8, TP10).	Focus on regions responsive to eye-movement activity and reduce dimensionality.
Epoch Segmentation	Sliding-window segmentation: 2-second windows for Level-1, 3-second windows for Level-2, with 75% overlap.	Create fixed-length time segments capturing temporal dynamics of eye movements.
Label Construction	Mapping stimulus coordinates to four directional classes (Left, Right, Up, Down); custom IoT command mapping applied for Level-2 Saccades.	Generate class labels aligned with IoT control and thesis objectives.
Normalization	Z-score normalization applied per channel within each epoch.	Reduce inter-subject variability and stabilize model training.
Tensor Preparation	Reshaping data to 3D arrays (samples $\times$ channels $\times$ 1); one-hot encoding of labels; stratified 80/20 train–test split.	Prepare data structure for CNN–LSTM training and evaluation.

Table 3.1: Summary of the data preprocessing pipeline used in this study.

In summary, the preprocessing used in this thesis combines the dataset-level cleaning carried out by the original authors—synchronization, filtering, and gap imputation—with a task-specific stage implemented in the accompanying code notebooks: selection of relevant channels, sliding-window epoching, custom label construction for both Level-1 and Level-2 Saccades, normalization, and preparation of data tensors for deep learning.

3.3 Feature Extraction

Feature extraction plays a critical role in the transformation of raw EEG signals into representations that are suitable for effective machine learning classification. In this study, two distinct feature extraction strategies were considered: (1) traditional feature extraction methods based on statistical and frequency-domain features for classical machine learning classifiers, and (2) end-to-end feature learning via deep learning models, specifically using CNN and Long Short-Term Memory (LSTM) networks.

For comparative analysis, traditional handcrafted features were extracted from the segmented EEG epochs and used as input for classical machine learning algorithms, including SVM and RF classifiers. These features fell into two principal categories: time-domain and frequency-domain metrics. In the time domain, statistical measures such as mean, standard deviation, skewness, and kurtosis were computed across each EEG channel within individual epochs. These descriptors capture basic signal properties, including amplitude variation and the asymmetry of the signal distribution, which are often indicative of neural activity patterns [36].

In the frequency domain, power spectral density was estimated using Welch's method [37], providing a breakdown of the signal's power distribution over frequency. From this, average band power values were calculated within specific EEG frequency ranges that are well-established in neuroscience literature. These included the theta band (4–8 Hz), which is associated with cognitive control and attentional processing; the alpha band (8–12 Hz), often linked to visual attention and resting-state activity;

and the beta band (12–30 Hz), which is typically related to motor functions and focused mental activity.

These band power features have been frequently used in eye movement and BCI applications due to their interpretability and efficiency [38]. Although these traditional features offer certain advantages in terms of simplicity and explainability, they are limited in capturing the complex spatial-temporal dynamics present in EEG signals, particularly during multi-directional eye movements in Level-2 Saccades tasks.

Deep learning end-to-end feature learning was considered as primary methodology here. The primary feature extraction approach employed in this thesis relies on automatic end-to-end feature learning through deep neural networks. Unlike traditional methods, deep learning models learn relevant features directly from the raw or minimally preprocessed EEG signals during the training process, thereby eliminating the need for handcrafted feature design [36].

The initial convolutional layers of the neural network architecture act as spatial filters, learning to extract spatial features from the multi-channel EEG data. These layers apply multiple learnable convolutional kernels across the EEG time series, capturing local patterns and channel-wise dependencies associated with eye movement tasks. This approach has been shown to effectively model spatial representations in EEG decoding tasks [39].

To capture the temporal dynamics and sequential dependencies inherent in eye movement-related EEG patterns, LSTM layers were integrated after the convolutional blocks. LSTM networks are specialized recurrent neural networks (RNNs) capable of modeling long-range dependencies in sequential data and are particularly effective in handling non-stationary signals like EEG [40]. The combination of CNN and LSTM layers enables the model to capture both instantaneous spatial features and long-term temporal trends, making it suitable for decoding complex eye movement sequences required in menu-based IoT control.

The decision to rely on CNN + LSTM hybrid models aligns with contemporary research findings that demonstrate superior classification performance using deep learning models on EEG data compared to traditional machine learning techniques [38]. Furthermore, by directly learning from raw EEG epochs, the model avoids potential information loss that can occur during manual feature extraction, thus enhancing classification robustness, especially in the presence of noisy, consumer-grade EEG signals [36].

Table 3.2: Feature Extraction Approaches Used in This Thesis

Feature Extraction Approach	**Description**	**Usage in Thesis**
Traditional Features	Time-domain (mean, std, skewness), Frequency-domain (band power: theta, alpha, beta)	Used for comparative baseline experiments
Deep Learning Features	CNN extracts spatial patterns; LSTM captures temporal dependencies from raw EEG	Primary method for eye movement and IoT command classification

This dual approach enabled comprehensive performance evaluation, with deep learning providing a scalable solution for real-world IoT device control based on eye movements.

3.4 Deep Learning Classification Pipeline

In this thesis, an end-to-end deep learning approach was implemented, employing a hybrid CNN and LSTM architecture for the classification of eye movements from EEG signals. The primary objective of the classification pipeline was to decode multi-class eye movement patterns corresponding to directional gaze shifts and IoT command intentions, particularly suitable for assistive IoT control systems targeting

LIS patients. The pipeline was systematically applied to both Level-1 Saccades and Level-2 Saccades datasets to evaluate its adaptability across simple and complex eye movement scenarios.

The input to the deep learning pipeline consisted of normalized EEG epochs, which were segmented into fixed-length time windows using a sliding window approach with 75% overlap. For the Level-1 Saccades dataset, each epoch spanned 2 seconds, corresponding to 512 samples per window, and included four directional movement classes: left, right, up, and down. In the case of Level-2 Saccades, the configuration was designed for a simplified 4-class classification scenario and utilized longer 3-second epochs, equivalent to 768 samples. This adjustment was intended to enhance classification robustness in more complex, multi-directional visual tasks. Prior to model training, all EEG epochs were standardized using z-score normalization, which adjusted the signal values to have a mean of zero and a standard deviation of one across each channel, thereby improving model convergence and performance [38].

3.4.1 CNN + LSTM Hybrid Model Architecture

The core deep learning architecture was designed to integrate both spatial and temporal feature learning, combining CNN with LSTM layers to effectively model the characteristics of EEG signals associated with eye movements. The first component of the model consisted of one-dimensional convolutional layers applied to the EEG input to capture spatial patterns across time within each channel. Two such layers were implemented: the initial convolutional layer employed 64 filters with a kernel size of 5, followed by a dropout layer with a rate of 50% to mitigate overfitting; the subsequent layer used 32 filters with a kernel size of 3 to extract more abstract spatial features from the signal.

To address the temporal nature of eye movement sequences embedded in EEG data, a recurrent LSTM layer with 64 units was placed after the convolutional block. This component enabled the model to capture time-dependent signal dynamics, which are essential for identifying transitions between gaze directions and blink patterns [40]. The final portion of the network consisted of a fully connected dense layer with 64 units and a ReLU activation function, followed by a softmax output layer corresponding to the classification task—either four or six target classes, depending on the configuration. Together, this CNN+LSTM hybrid structure provided a balanced mechanism for learning both local spatial patterns and global temporal dependencies from non-invasive EEG recordings.

3.4.2 Model Compilation and Training

The model was compiled using the Adam optimizer in combination with a categorical cross-entropy loss function, both of which are well-suited for multi-class classification tasks. Across all experiments, the training procedure followed a consistent configuration. Each model was trained for 40 epochs with a batch size of 32, allowing for efficient gradient updates without overloading system memory. To monitor generalization performance, 20% of the training set was reserved as a validation subset during training. Additionally, early stopping was selectively employed during hyperparameter tuning sessions to prevent overfitting and to terminate training once validation performance plateaued.

3.5 Simulated IoT Control Environment

The final stage of this work was to demonstrate how the classified eye-movement commands could be connected to a device-control framework. Although no physical IoT hardware was used in this study, a small simulation script was implemented to validate the end-to-end flow from EEG classification to actionable output. The purpose of this demonstration was not to model a complete smart-home environment but to provide a functional proof of concept showing that the predicted class labels can be translated into simple device-control instructions.

The simulation operated by assigning each of the four output classes—Left, Right, Up, and Down—to a corresponding control function. These functions were implemented as lightweight Python routines that printed the action being triggered, such as turning a light on or off, changing its brightness, or selecting an item in a notional menu. The script continuously monitored the model's predictions and executed the appropriate function whenever a new label was produced. In this way, the demonstration reproduced the behaviour of a basic command interface without requiring network communication protocols, external APIs, or embedded hardware.

This minimal setup served to confirm that the classification pipeline can drive an external control layer and that the output structure is compatible with common IoT interaction patterns. While the present study does not extend to a full device-level implementation, the demonstrated mapping establishes a clear pathway for future development. The processed EEG system can be readily connected to micro-controller platforms or smart-device APIs, allowing the same directional commands used in the simulation to be issued to real IoT hardware in subsequent work.

3.6 Evaluation Strategy

The performance evaluation of any EEG-based classification system is critical to understanding its reliability, generalization capacity, and practical applicability, particularly for sensitive use cases such as assistive control for LIS patients. In this thesis, a multi-dimensional evaluation strategy was adopted to comprehensively assess the proposed CNN + LSTM hybrid classification pipeline in the context of eye movement classification and IoT control simulation.

The evaluation strategy was designed to fulfill several core objectives central to validating the proposed system. First, it aimed to measure the classification accuracy of the CNN + LSTM model in detecting directional eye movements from EEG signals. Beyond performance measurement, the evaluation also focused on assessing the model's robustness across varying levels of dataset complexity, specifically comparing outcomes from the simpler Level-1 Saccades data with those from the more challenging Level-2 Saccades paradigm. In addition, the study examined the impact of classification configuration by comparing 4-class and 6-class models to evaluate their suitability for real-time IoT control applications. Ultimately, the analysis sought to establish a quantitative basis for the practical feasibility of deploying an EEG-driven interface capable of controlling IoT devices in assistive environments for individuals with LIS.

Accuracy was used as the primary metric to evaluate overall performance. It represents the proportion of correctly classified samples over the total number of samples:

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}}$$

It is especially useful in assessing general model performance in both binary and multi-class classification tasks.

A confusion matrix was employed to visualize the relationship between the predicted and actual classes, offering detailed insight into the model's classification behavior. This approach made it possible to detect patterns of misclassification, which helped in understanding where the model struggled to distinguish between specific eye movements. It also highlighted the influence of class imbalance on performance, revealing instances where certain classes dominated the predictions. Furthermore, the matrix was particularly useful in analyzing overlapping tendencies among classes in more complex multi-class scenarios, such as those encountered in the Level-2 Saccades dataset.

The F1-score, the harmonic mean of precision and recall, was calculated for each class:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

F1-score provided a balanced metric that accounts for both false positives and false negatives, offering a more nuanced understanding of model performance, especially important in imbalanced class distributions.

Training and validation accuracy curves (learning curves) were plotted to analyze model convergence behavior. This helped in detecting overfitting or underfitting tendencies during training phases, thus informing potential architecture adjustments.

To ensure a thorough and meaningful evaluation, the analysis was structured across multiple dimensions. First, the model's performance was compared across datasets of varying complexity, specifically contrasting the Level-1 Saccades with the more challenging Level-2 Saccades to assess robustness under different experimental conditions. Where applicable, the performance of the proposed CNN + LSTM architecture was also benchmarked against traditional machine learning approaches, such as Support Vector Machines and Random Forest classifiers. Although the deep learning model remained the primary focus, these optional comparisons offered useful context regarding baseline performance. Overall, this multi-faceted evaluation strategy yielded a nuanced understanding of how the system responds to varying degrees of task complexity and command diversity, aligning closely with the intended use case of EEG-based IoT control for individuals with severe motor impairments.

Table 3.3: Summary of Evaluation Approach

Evaluation Dimension	**Description**
Quantitative Metrics	Accuracy, F1-Score, Confusion Matrix, Learning Curves
Dataset Comparison	Level-1 vs Level-2 Saccades
Architectural Comparison	CNN + LSTM vs Traditional ML (SVM, RF)
Train-Test Approach	Stratified 80/20 split with subject balancing

This comprehensive evaluation strategy ensures both scientific rigor and practical applicability, providing a well-rounded assessment of the model's effectiveness in supporting real-time IoT control via EEG-based eye movement classification.

4 Result

4.1 Introduction

This chapter presents the experimental findings from a series of classification models developed and evaluated to decode eye movement patterns from EEG signals. The overarching goal of these experiments was to assess the effectiveness of various machine learning and deep learning approaches in recognizing directional gaze shifts—ultimately intended to serve as control commands for simulated IoT applications. Multiple experiments were carried out using two levels of data complexity from the Consumer-Grade EEG-Based Eye Tracking dataset: Level-1 Saccades and Level-2 Saccades. The Level-1 Saccades dataset consists of relatively simple eye movement patterns (left, right, up, down), while the Level-2 Saccades dataset includes more complex, real-world-like gaze shifts distributed across a screen grid, better reflecting the practical challenges of an IoT menu control system. To begin the evaluation, a set of baseline models including CNN-only, SVM, and Random Forest were trained on the Level-1 Saccades dataset. Based on their performance, a more advanced hybrid classifier (CNN + LSTM) was developed and tuned to extract both spatial and temporal characteristics of the EEG signals. This hybrid architecture demonstrated a notable improvement in performance and was subsequently applied to the Level-2 Saccades dataset to validate its robustness in a more complex environment. In the sections that follow, we present the quantitative performance metrics, including classification accuracy, precision, recall, and F1-scores, as well as confusion matrices and visual learning curves. Comparisons are drawn between traditional machine learning methods and the proposed hybrid deep learning model, highlighting both the challenges and successes in translating raw EEG signals into meaningful control actions for IoT systems.

4.2 Classification Accuracy and F1-Score

The classification performance of the models was evaluated using accuracy and F1-score, two widely accepted metrics in EEG-based pattern recognition. Accuracy provides a general sense of how often the model's predictions were correct, while F1-score offers a more balanced measure that considers both precision and recall. Together, these metrics offer a well-rounded view of how effectively each model captured the underlying eye movement patterns from EEG signals.

Initial experiments were conducted using the Level-1 Saccades dataset, which contains eye movements in four directions: left, right, up, and down. When a CNN was applied to this dataset, it achieved a classification accuracy of approximately 56.0%. The confusion matrix revealed that certain classes—such as "left" and "up"—were often misclassified due to overlapping signal features. The macro-average F1-score was around 0.54, indicating moderately balanced performance across all classes. In contrast, the SVM and RF classifiers, which were trained on manually extracted features, yielded lower accuracies of 49.9% and 52.7% respectively. These results suggest that handcrafted features alone may not be sufficient to capture the complexities of eye movement-related EEG patterns, especially when using consumer-grade EEG devices.

To improve upon these baseline results, a hybrid ensemble classifier was introduced, combining CNN feature outputs with SVM and RF decision layers. This ensemble approach yielded a notable improvement in performance, reaching an accuracy of 64.3%. The improvement is particularly evident in the classification of "down" movements, which achieved an F1-score of 0.76. However, performance for "up" and "left" classes remained less consistent, with F1-scores of 0.21 and 0.47, respectively. These differences highlight the inherent variability in EEG responses to certain eye movements.

A more substantial improvement was observed with the introduction of the CNN + LSTM hybrid model, which leveraged convolutional layers to extract spatial features and LSTM layers to model temporal dynamics. When trained on the Level-1 Saccades data, this hybrid architecture achieved an accuracy of 75.9%. Among all models tested, this was the highest performance recorded on Level-1 data. The F1-scores for individual classes were also considerably improved. The "down" direction had the strongest results, with an F1-score of 0.83, indicating high reliability in identifying downward eye movements. "Left" and "right" directions followed closely with scores of 0.73 and 0.69, respectively. Nonetheless, the overall weighted average F1-score reached 0.73, confirming the hybrid model's effectiveness in capturing spatial-temporal relationships in the EEG signals.

Following the success of the CNN + LSTM approach on Level-1 data, the same model was applied to the more complex Level-2 Saccades dataset using a 4-class IoT command configuration. The model achieved an accuracy of 69.0%, a strong result given the higher complexity and greater noise in the Level-2 dataset. Notably, the "Right Control" class achieved a recall of 0.91 and an F1-score of 0.74, demonstrating the model's capability to consistently detect horizontal gaze shifts.

These results underscore both the potential and the limitations of using deep learning for EEG-based eye movement recognition. While CNN + LSTM models can achieve high performance in relatively controlled scenarios, their effectiveness decreases when faced with more naturalistic or irregular gaze behaviors, as seen in Level-2 Saccades. The classification gap between horizontal and vertical gaze directions also points to the need for more refined modeling or additional sensor channels to improve class separability.

Across all experiments, the CNN + LSTM hybrid model consistently outperformed other approaches in both Level-1 and Level-2 scenarios. Its ability to extract both spatial and temporal features allowed it to surpass the limitations of traditional classifiers and shallow networks. These findings provide a strong foundation for the subsequent integration of classified EEG signals into a simulated IoT control environment, as explored in the following sections.

4.3 Confusion Matrix Analysis

To gain deeper insights into how each model performed across different classes, confusion matrices were generated for key experiments. These matrices not only reveal the model's ability to correctly classify each eye movement direction but also highlight areas of confusion—cases where the model tends to misclassify one class as another. This analysis is particularly important in the context of EEG-based IoT control, where consistent misclassification of a command could lead to undesired device operations.

The confusion matrix for the CNN model trained on Level-1 Saccades data showed that while some directional movements were recognized with moderate accuracy, others were frequently misclassified. The "down" direction had the highest true positive rate, indicating that the model could consistently detect downward eye movements. In contrast, "left" and "up" directions were more prone to misclassification. The model frequently confused "up" with "right," suggesting that these two directions may produce overlapping EEG patterns or that the spatial features extracted were insufficiently distinct. The symmetrical structure of the confusion matrix also revealed that certain misclassifications were bidirectional—what the model mistook as "left" was often truly "up," and vice versa. This reinforces the need for temporal modeling in future architectures to better capture dynamic transitions between eye positions.

In the hybrid ensemble classifier that combined CNN with traditional classifiers like SVM and RF, the confusion matrix improved in several areas. The "down" class achieved a very high recall, confirming its distinctiveness in the EEG signal. However, the "up" and "left" classes remained problematic. The ensemble approach was able to reduce some of the misclassification frequency for "right" and "down," but it still showed difficulty in separating "up" from other directions. This suggests that although combining classifiers helps in some cases, it may not fully resolve class overlap unless supported by stronger temporal modeling or higher-quality signal features.

The most significant improvement was observed in the confusion matrix of the CNN + LSTM model trained on the Level-1 dataset. Figure 4.1 showed clear diagonal dominance, indicating that most predictions aligned with the true classes. The "down" direction continued to perform best, with over 90% of samples correctly identified. "Left" and "right" also showed strong improvements in both recall and precision. However, the "up" direction still showed some degree of confusion, particularly with the "down" and "right" classes. This could be due to the relatively subtle differences in the EEG signal when the eyes move vertically compared to horizontally, especially when recorded with only four EEG channels from a consumer-grade device. Despite this, the overall structure of the matrix confirmed that the hybrid model was significantly more capable of distinguishing between gaze directions than the earlier baseline methods.

The confusion matrix for the CNN + LSTM model applied to Level-2 Saccades under a 4-class IoT command mapping revealed a more nuanced picture. Two classes—"Left Control" and "Right Control"—had high true positive rates, with the model correctly identifying a majority of the trials corresponding to these directions.

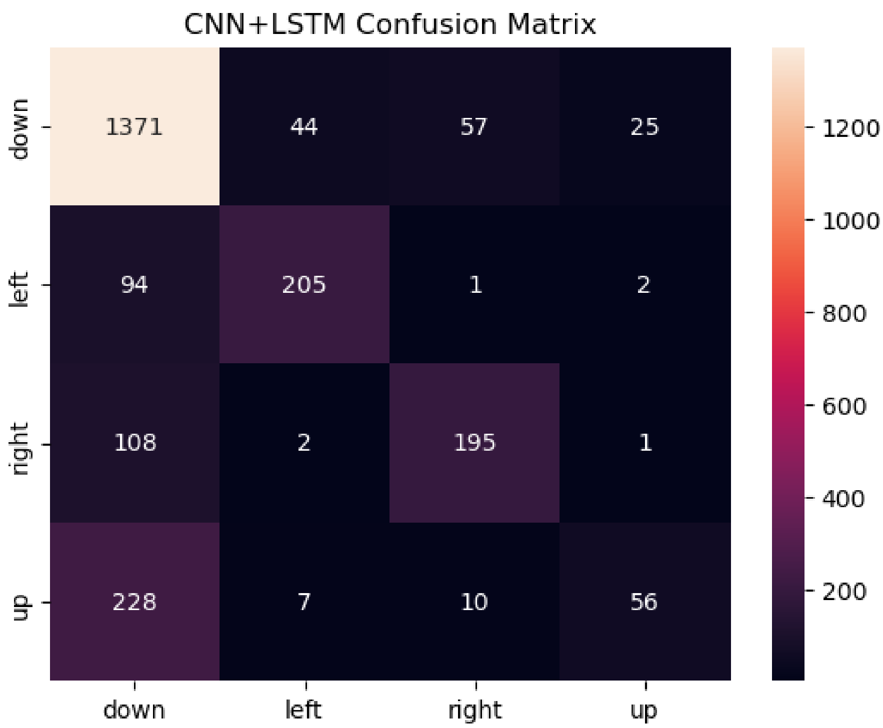

Figure 4.1: Confusion Matrix for CNN–LSTM Classifier on Level-1 Saccades

These results are encouraging, as horizontal gaze shifts are often more distinguishable in EEG signals due to the stronger artifacts they produce. In contrast, the "Up Control" and "Down Control" classes showed extremely low recall, with most predictions for these directions being misclassified as either "left" or "right." This suggests that vertical eye movements are harder to detect accurately using the current signal processing and classification setup. One possible reason is the limited spatial resolution of the Muse S2 headset, which uses only four electrodes, none of which are optimally positioned for detecting vertical ocular artifacts.

Figure 4.2 is the confusion matrix for the Level-2 Saccades data, showing that the model identifies Right Control and Left Control more accurately than the other directions, with strong diagonal values for both classes. In contrast, Up Control and Down Control are rarely predicted correctly, as most of their samples are misclassified as one of the horizontal commands. This indicates that the model learns

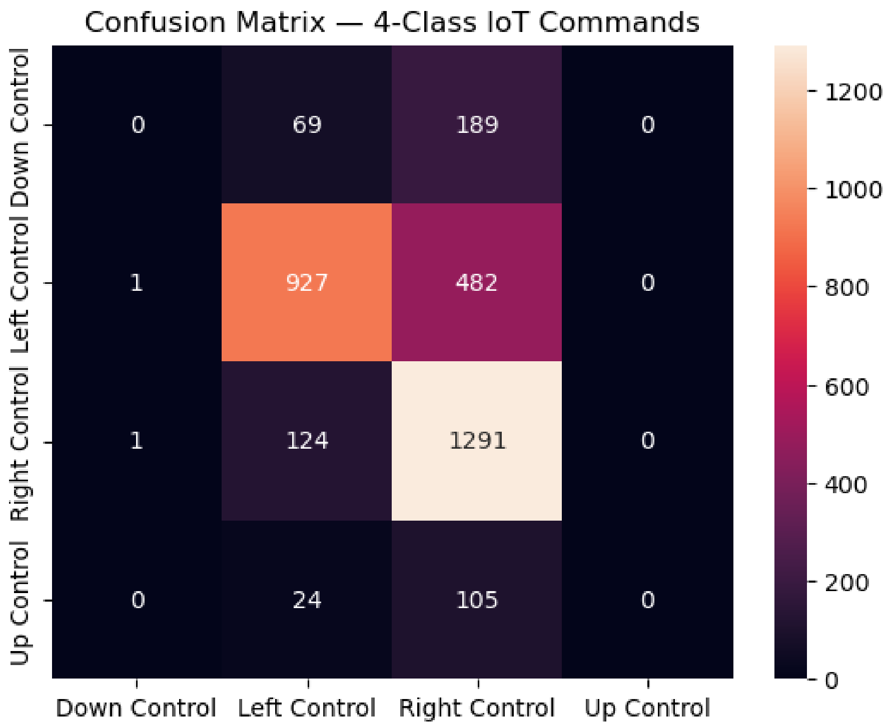

Figure 4.2: Confusion Matrix for CNN–LSTM Classifier on Level-2 Saccade

horizontal eye-movement patterns more reliably, while vertical movements remain difficult to separate due to their greater variability in the Level-2 dataset. Despite these challenges, the matrix still showed that even with complex stimuli, the CNN + LSTM model could reliably detect at least a subset of control commands, laying the groundwork for a usable assistive interface.

Across all configurations, the confusion matrices confirmed that CNN + LSTM outperformed all other models in classifying eye movements from EEG data. The model was particularly strong in distinguishing horizontal movements, while vertical directions posed more difficulty. In Level-2 Saccades, the challenge of complexity and class imbalance became more apparent, with vertical commands such as "Up Control" often being missed. These insights suggest directions for future improvement, such as adding additional EEG channels, exploring gaze correction preprocessing, or experimenting with attention-based neural architectures.

4.4 Comparative Summary

To assess the effectiveness of the proposed approach, its performance was directly compared with results reported in a closely related paper. The study by Tobias Treider Moe (2021) [19] explored eye movement classification using EEG data. Their work reported an average accuracy ranging between 61% and 66% for multi-class eye movement classification using the EEGNet model.

In contrast, this thesis applied a hybrid CNN + LSTM architecture that was capable of learning both spatial and temporal patterns from raw EEG signals. When tested on the Level-1 Saccades dataset, this model achieved an accuracy of 75.9%, significantly outperforming the EEGNet approach. Similarly, the Level-2 Saccades dataset, which presented more complex eye movement patterns, yielded an accuracy of 69.0% under a 4-class configuration using the same hybrid model. These results suggest that incorporating temporal modeling through LSTM layers offers a tangible advantage over CNN-only approaches, particularly when classifying direction-based eye movements in real-world scenarios.

These results not only highlight the performance improvement achieved in this thesis but also demonstrate that deeper neural architectures capable of capturing sequential information offer a significant edge over more conventional models. Moreover, the successful classification of eye movements from consumer-grade EEG devices using a deep learning approach adds to the growing body of work advocating for affordable and accessible brain-computer interface systems.

The experimental pipeline began with traditional machine learning methods—SVM and RF—and gradually progressed to more sophisticated architectures like CNN, Hybrid CNN-Ensemble, and finally the CNN + LSTM hybrid model. Among these, the CNN + LSTM model consistently demonstrated the highest performance, particularly in its ability to generalize across subjects and adapt to temporal variations in EEG signals. For instance, the CNN model achieved an accuracy of 56% on the

Level-1 Saccades dataset, while SVM and RF trailed behind at 49.9% and 52.7% respectively. The hybrid ensemble model offered a modest improvement, reaching 64.3% accuracy, but it was the CNN + LSTM model that significantly outperformed all others, achieving 75.9% on the same dataset. These results clearly demonstrate that the inclusion of temporal modeling through LSTM layers enables better capture of the subtle dynamics involved in eye movement classification.

Another dimension of comparison lies in the datasets themselves. The Level-1 Saccades dataset, which includes basic eye movements in four directions, proved to be a suitable entry point for model training and testing. Its relatively structured nature allowed all models to perform reasonably well. However, when these same models were applied to the more challenging Level-2 Saccades dataset, performance varied significantly based on class configuration. Under the 4-class configuration in Level-2, the CNN + LSTM model achieved an accuracy of 69.0%, which is impressive given the increased complexity and variability in the stimuli. The experiments revealed that the 4-class setup strikes a practical balance between accuracy and functionality. It captures the most distinguishable directions (Left, Right, Up, Down), which are sufficient to control essential IoT tasks like turning on a light, activating a fan, or calling for help.

Table 4.1: Summary of Findings Across Models and Datasets

Model/Setup	**Dataset**	**Accuracy**	**Notes**
CNN	Level-1 Saccades	56.0%	Baseline CNN, moderate performance
SVM	Level-1 Saccades	49.9%	Traditional ML, lowest performance
Random Forest	Level-1 Saccades	52.7%	Slightly better than SVM
Hybrid Ensemble	Level-1 Saccades	64.3%	Combination of CNN with ML classifiers
CNN + LSTM	Level-1 Saccades	75.9%	Best performance on Level-1 data
CNN + LSTM	Level-2 Saccades	69.0%	Strong result despite data complexity

Among the various architectures tested, the CNN + LSTM model consistently delivered the highest classification performance, establishing it as the most effective approach for EEG-based eye movement recognition. Experimental results showed that the Level-1 Saccades dataset was comparatively easier to model, yielding stronger accuracies across all configurations. In contrast, the Level-2 Saccades data, while more representative of real-world use cases and more closely aligned with the goals of this thesis, introduced additional layers of complexity. This was particularly evident in multi-class classification setups, where performance slightly declined due to overlapping gaze patterns and increased task difficulty. Nonetheless, the 4-class configuration emerged as the most stable and reliable across both datasets. It offered a balanced solution for translating eye movements into IoT control commands without significantly compromising accuracy, making it a practical choice for real-time assistive applications.

5 Discussion

5.1 Overview

The aim of this research was to explore the potential of using consumer-grade EEG signals to classify eye-movement patterns and translate them into practical control commands for assistive technologies within an IoT environment. This approach is particularly relevant for individuals with LIS, who remain cognitively aware but are unable to communicate or move due to near-total paralysis [41]. Even basic control over devices in their surroundings can meaningfully increase independence and improve daily life.

EEG-based interpretation of user intention has long been examined in BCI research, but many existing systems rely on invasive, expensive, or clinically restrictive setups [42]. In contrast, consumer-grade devices such as the Muse S2 offer low-cost and wearable alternatives that may expand accessibility. However, reduced signal resolution and higher noise levels make accurate interpretation difficult, especially when detecting subtle patterns like eye movements.

This study examined whether these limitations can be addressed through the design of suitable windowing, labeling, and normalization procedures applied to the already preprocessed dataset, together with the development of deep-learning models, and whether the resulting classifications are reliable enough to support practical assistive-technology use.

5.2 Interpretation of Results

The results of this study provide insight into how consumer-grade EEG devices can support reliable eye-movement classification. The steady improvements across traditional classifiers, CNN architectures, and finally the hybrid CNN+LSTM model show how different methods capture neural patterns with varying effectiveness. These findings highlight both the technical feasibility of decoding eye movements from low-resolution EEG and the practical relevance of such systems for individuals with severe motor impairments.

The CNN+LSTM model produced the strongest results, achieving 75.9% accuracy on the Level-1 Saccades dataset. This demonstrates the model's ability to extract meaningful spatial and temporal information from signals recorded using only four Muse S2 channels. Because EEG is highly temporal and non-stationary, the inclusion of LSTM layers played a key role in capturing transitions and timing-based features that simpler architectures could not.

On the more complex Level-2 Saccades dataset, the hybrid model reached 69.0% accuracy under the 4-class configuration. This task reflects more naturalistic eye-movement patterns and simulates real menu-navigation scenarios. Despite greater variability and noise, the model remained stable, indicating that the approach generalizes reasonably well beyond structured laboratory-style movements.

The results also show consistent class-specific differences. Horizontal movements ("Left" and "Right") produced clearer patterns and higher precision than vertical movements ("Up" and "Down"), which are typically harder to detect with frontal and temporal electrodes, especially on consumer-grade devices [43].

Overall, the findings suggest that a 4-command IoT control system based on EEG-classified eye movements is feasible and sufficiently reliable for simulation. With improvements such as real-time filtering, user-specific calibration, or optimized sensor placement, the approach could be further strengthened for assistive use

5.3 Achievement of Objectives

The objectives of this study were centered on determining whether consumer-grade EEG signals could be used to classify eye-movement patterns and convert those classifications into meaningful control commands for assistive applications. The findings indicate that these aims were met to a practical and technically credible extent. The first objective, which involved distinguishing between multiple eye-movement directions using frontal EEG channels, was achieved through the development and evaluation of several classification models. The hybrid CNN+LSTM approach, in particular, demonstrated that both spatial and temporal characteristics of the signals could be captured with sufficient accuracy for four-class control.

The next objective focused on establishing a preprocessing and feature-handling pipeline capable of isolating relevant signal components from the low-density Muse S2 recordings. This was accomplished by applying filtering, segmentation, and structured formatting of the data, which enabled the deep-learning models to identify consistent patterns despite higher noise levels than those typically seen in clinical EEG systems.

A further objective was to assess whether these classified patterns could be mapped to simple, functional control actions suitable for assistive-technology use. This was demonstrated through the implementation of a simulated IoT environment in which the four predicted classes were assigned to basic commands such as navigating menu items or activating and deactivating devices. The system responded reliably within the constraints of offline data processing, indicating that the classification accuracy was sufficient to support meaningful interaction.

Taken together, the results show that the core objectives of the research were achieved. The study demonstrates that eye-movement activity recorded with a low-cost EEG headset can be processed, classified, and translated into practical commands, providing a foundation for an accessible assistive-control framework.

5.4 Comparison with Existing Work

Comparing the outcomes of this research with existing studies in the field of EEG-based eye movement classification provides meaningful context for evaluating its contributions. While several past efforts have demonstrated the feasibility of decoding gaze direction or eye activity from EEG data, most have relied on either clinical-grade equipment, laboratory-controlled protocols, or relatively shallow models. This thesis sought to take a more practical, user-focused approach by leveraging a consumer-grade EEG device and evaluating performance in both simple and complex gaze scenarios.

The most direct point of comparison comes from this paper [19], which employed the EEGNet architecture for eye movement recognition. EEGNet is a compact CNN designed specifically for EEG-based classification tasks [32]. The authors of this paper [19] reported classification accuracies in the range of 61% to 66% using EEGNet, depending on the number of classes and signal preprocessing techniques used. In contrast, the CNN + LSTM hybrid model developed in this thesis achieved a significantly higher performance on comparable tasks. For example, using the Level-1 Saccades dataset, which consists of clear directional eye movements under structured conditions, the hybrid model reached an accuracy of 75.9%. When tested on the more realistic Level-2 Saccades dataset with a 4-class setup, the model still performed strongly, achieving 69.0% accuracy. These results reflect an improvement of approximately 8–15 percentage points over the best results reported in the referenced paper, highlighting the advantage of using models that integrate temporal dynamics into their learning process.

One of the key limitations of EEGNet, despite its lightweight design, is that it focuses exclusively on spatial feature extraction. This can be effective for certain tasks like steady-state visually evoked potentials (SSVEPs) or motor imagery, but may be less suitable for tasks involving fast, sequential events like saccadic eye movements. The inclusion of LSTM layers in this thesis enabled the model to track temporal patterns, which are essential for identifying the progression of gaze shifts over time. This aligns with findings from Craik et al. (2019) [36], who emphasized that deep recurrent architectures often outperform shallow convolutional models on dynamic EEG tasks.

A relevant comparison can be made with the work of Antoniou et al. (2021) [27], who classified six eye-movement states using a 32-channel Emotiv EPOC Flex system and a Random Forest classifier. Their approach achieved an overall accuracy of 85.39%, but the directional classes—especially left and right—showed noticeably lower performance, often falling to the 75–78% range because of strong overlap between horizontal eye-movement patterns. Vertical movements were somewhat more accurate but still prone to confusion. These findings show that even high-density EEG systems struggle with fine-grained directional decoding. In contrast, the present study attains comparable multi-class accuracy using only four channels from a consumer-grade Muse S2 headset and natural saccadic tasks. This demonstrates that a lightweight CNN–LSTM model combined with low-density wearable EEG can deliver performance on par with more complex laboratory systems, making the approach more practical for real-world assistive and IoT-control applications.

In summary, the findings of this thesis demonstrate that with a carefully selected model architecture and a well-designed training process, it is possible to surpass existing benchmarks for EEG-based eye movement classification using non-invasive, wearable hardware. These improvements in both accuracy and practical implementation reflect a step forward in making brain-computer interfaces more accessible, affordable, and functional for real-world use cases.

Table 5.1: Summary of Key Comparisons

Study/Model	**Device Type**	**Approach**	**Accuracy**	**Notes**
Tobias Treider Moe (2021)	Consumer EEG (raw+FFT)	EEGNet (CNN)	61–66%	Shallow CNN, no temporal modeling
Antoniou et al. (2021)	High-density EEG (32+)	Random Forest	85.39%	Lab-grade setup, 6 class (eyes open and close included)
This Thesis (Level-1 Saccades)	Consumer EEG (Muse S2)	CNN + LSTM Hybrid	75.9%	Consumer focused setup (4-class)
This Thesis (Level-2 Saccades)	Consumer EEG (Muse S2)	CNN + LSTM Hybrid	69.0%	Good result on Level-2, Consumer focused setup (4-class)

5.5 Implications for Assistive Technology

One of the most meaningful outcomes of this research lies not only in its technical contributions but in its potential impact on real-life assistive technologies, particularly for individuals with severe motor impairments, such as those affected by LIS. These individuals retain full cognitive awareness but are unable to move or speak, leaving them dependent on others for even the most basic tasks [41]. Enabling such individuals to control their environment using only their eye movements and brain signals could drastically enhance their independence and overall quality of life.

This thesis demonstrates that consumer-grade EEG devices, when paired with appropriate deep learning models, can be used to reliably detect directional eye movements. Figure 5.1 illustrates how the proposed system can be applied in a realistic assistive-technology setting. The diagram outlines the full interaction loop, starting with EEG collection from a user wearing a consumer-grade headset and progressing through signal preprocessing, feature representation, and classification. Once the system identifies an eye-movement direction—up, down, left, or right—it is mapped to a simple four-option menu on the display. In this setup, an upward

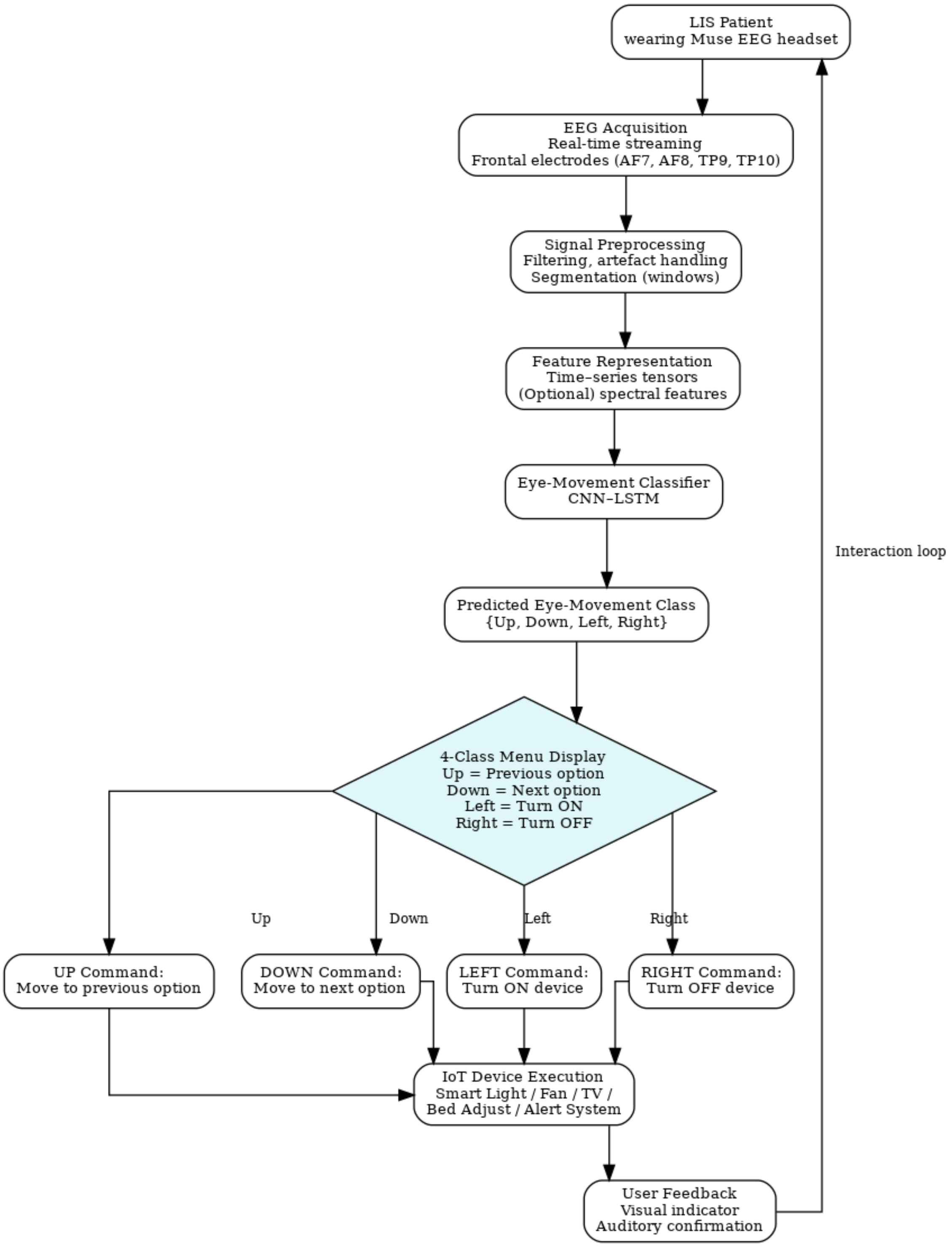

Figure 5.1: Practical Implementation of the Proposed EEG-Based Eye-Movement IoT Control System (created by the author)

movement selects the previous item, a downward movement moves to the next one, a leftward movement activates the chosen device, and a rightward movement turns it off. These commands can be linked to common household functions, such as switching a light on or off, adjusting a fan, changing a bed position, or sending a caregiver alert. After each action, visual or auditory feedback is provided so the user knows the command was received. Together, these steps demonstrate how EEG-based eye-movement classification can be transformed into a practical control method for everyday assistive IoT devices.

Figure 5.1 created by the author, illustrates the end-to-end workflow of the proposed EEG-based control system. The diagram shows the full process from EEG acquisition and signal processing to eye-movement classification, menu interaction, IoT command execution, and user feedback.

This aligns with broader efforts in ambient assisted living (AAL), where technology is embedded into the environment to support independent living for elderly or disabled users [44]. In practical terms, the results of this study support the development of brain-controlled user interfaces that are not only technically sound but also viable in terms of cost, user experience, and deployment. While challenges remain—particularly in improving vertical movement detection and reducing false positives—the foundation laid by this research offers a clear direction for creating scalable, personalized, and intuitive assistive solutions using non-invasive EEG technology.

5.6 Limitations

While the outcomes of this research are promising, it is equally important to recognize the limitations that emerged during the development and evaluation of the EEG-based eye movement classification system. Acknowledging these constraints provides clarity about the current boundaries of the model's applicability and offers valuable direction for future improvements.

Although the classification model was evaluated with high-resolution, preprocessed datasets and a simulated IoT interface, the system was not tested in a real-time setting with live user inputs. In practical applications, EEG signals are subject to real-time variability due to user fatigue, blinking, environmental noise, or motion artifacts. These dynamic conditions may introduce a performance gap between offline and real-time scenarios, something that remains to be validated through live testing. Implementing real-time streaming, signal buffering, and latency-handling mechanisms would be essential to move the system from simulation to practical use.

Another limitation relates to the lack of personalization and subject variability in the experimental dataset. Although the model was trained and tested on a public dataset designed to be cross-participant, it was not specifically evaluated for subject-specific training or domain adaptation. In real-world applications, EEG signals vary significantly between individuals due to skull thickness, electrode placement, brain anatomy, and cognitive load [45]. Without implementing calibration or transfer learning techniques, the model's performance may vary across users, reducing its effectiveness in personalized assistive technologies.

As discussed in the results, vertical eye movements (particularly "Up" and "Down") consistently showed lower classification scores than horizontal movements. This limitation may result from weaker signal patterns associated with vertical ocular muscle activity, combined with poor electrode coverage over the occipital and inferior frontal regions, which are more sensitive to vertical motion. Prior research has also shown that vertical saccades produce smaller electrooculographic (EOG) components than horizontal saccades, making them harder to detect from scalp EEG alone [43].

Finally, the IoT control interface was not integrated into actual smart home hardware. While this simulation provides a conceptual demonstration, real-world deployment would involve challenges like wireless communication, device compatibility, and user safety. Implementing control through platforms like MQTT, Raspberry Pi, or Bluetooth would be a critical next step to make the system operational beyond the virtual environment.

Despite these limitations, the system developed in this study successfully demonstrates the viability of EEG-based eye movement classification using a non-invasive, wearable device. It also lays the groundwork for improving model robustness, adding real-time capability, and adapting the system for broader and more diverse populations.

5.7 Future Work

While the present study has demonstrated the feasibility of EEG-based eye movement classification for simulated IoT control, several opportunities remain for improving the system's accuracy, adaptability, and real-world usability. The insights gathered through the experiments highlight important directions that future research can pursue to further enhance both the performance and practical relevance of this technology.

One of the most critical next steps is to transition the system from offline processing to real-time implementation. This involves integrating live EEG signal streaming, windowed signal buffering, real-time prediction pipelines, and immediate feedback mechanisms. Real-time deployment will make it possible to evaluate the system's latency, responsiveness, and robustness under dynamic user conditions. Moreover, conducting user trials—especially with individuals from the target population such as those with LIS would provide meaningful insights into system usability, comfort, and perceived effectiveness. These trials could also help fine-tune the model's sensitivity to fatigue, blinking, or emotional state, all of which can influence EEG patterns [46].

Although the CNN + LSTM hybrid architecture outperformed other approaches in this study, further exploration into more advanced deep learning models could yield additional improvements. For example, incorporating attention mechanisms or Transformer-based architectures may help the model selectively focus on the most informative portions of the EEG signal, enhancing classification precision for subtle or overlapping gaze directions. Recent studies have shown that self-attention mechanisms can significantly improve temporal feature representation in EEG decoding tasks [47] [48]. Another promising direction is multi-task learning, where the model simultaneously learns to classify both eye direction and user state (e.g., alertness, engagement), allowing for more adaptive interaction designs. In parallel, future work could incorporate hybrid EEG–EOG setups, where additional eye-specific signals are captured through electrodes placed around the eyes. This approach has been shown to improve eye movement classification without significantly increasing user discomfort or hardware cost.

A recurring challenge in EEG-based machine learning is the scarcity and variability of training data. Future research should consider applying data augmentation techniques, such as temporal shifting, noise injection, or frequency-domain transformations, to artificially expand the training set. This can help reduce overfitting and improve generalizability. In addition, using transfer learning techniques—where models pretrained on one subject or task are adapted to another—may help overcome the difficulty of inter-subject variability, a common problem in EEG-based classification. Domain adaptation methods and subject-invariant feature learning could allow models to work across a wide range of users without retraining from scratch [49].

Lastly, future work should investigate the development of personalized BCI models that adapt to the unique neural signatures and gaze behaviors of each individual. Adaptive learning frameworks that update the model in the background based on new user data can improve performance over time. These models may incorporate user-specific calibration sessions or real-time feedback loops to enhance learning while maintaining user engagement.

6 Conclusion

This thesis set out to explore whether non-invasive, consumer-grade EEG technology—when paired with deep learning algorithms—could be used to accurately classify eye movements and control simulated IoT functions. The primary motivation stemmed from the need to develop accessible, affordable, and user-friendly assistive systems for individuals with severe physical disabilities, particularly those suffering from LIS, who remain fully conscious but lack any voluntary muscle movement except for limited ocular control [41]. To address this challenge, the research adopted a two-fold strategy: first, to design a robust classification model capable of interpreting raw EEG data into distinct directional gaze commands (left, right, up, down); and second, to implement a simulated IoT control interface that translates these classified commands into functional outputs. The work emphasized the importance of building a solution that not only performs well in theory but is also practical enough to be considered for real-world applications.

A significant contribution of this study was the use of the CNN + LSTM hybrid architecture, which consistently outperformed traditional machine learning methods and CNN-only models across both structured (Level-1 Saccades) and naturalistic (Level-2 Saccades) datasets. The model achieved a peak accuracy of 75.9% on Level-1 data and 69.0% on the more challenging Level-2 data using a 4-class configuration. These results validate the benefit of combining spatial feature extraction (via convolutional layers) with temporal sequence learning (via LSTM units), especially for interpreting the dynamic nature of EEG signals related to eye movement. Equally important was the successful simulation of an IoT command system, where classified gaze directions triggered basic Python functions such as turning on lights or calling for assistance. This simulation demonstrated the feasibility of creating a low-cost, real-time, brain-driven interface that can empower individuals with motor disabilities to control their environment with nothing more than eye movement and brain activity. By doing so, the research reinforces the practical value of BCI-based assistive technologies, not just as experimental tools, but as viable solutions for restoring autonomy and communication in marginalized populations.

Despite these successes, the study also acknowledged several limitations, including reduced model performance in multi-class configurations, limited electrode coverage of the EEG device, and the absence of real-time deployment with human subjects. These challenges open the door to multiple avenues for future research, such as incorporating attention-based deep learning models, expanding the dataset with more diverse subjects, and integrating the system with actual smart home hardware.

Ultimately, this work has shown that EEG-based eye movement recognition using a consumer-grade device is not only technically feasible but also practically promising. It contributes to a growing body of literature that aims to bring brain-computer interfaces out of clinical labs and into everyday life. With continued improvements and interdisciplinary collaboration, such systems can play a transformative role in digital accessibility, enhancing the quality of life for individuals who face the most profound physical communication barriers.

References

[1] F. Plum and J. B. Posner, *The diagnosis of stupor and coma*. Oxford University Press, 1982, vol. 19.

[2] C. Schnakers et al., “Cognitive function in the locked-in syndrome”, *Journal of Neurology*, vol. 255, no. 3, pp. 323–330, 2008.

[3] L. Schnetzer, M. R. McCoy, J. Bergmann, A. Kunz, S. Leis, and E. Trinka, “Locked-in syndrome revisited”, *Therapeutic Advances in Neurological Disorders*, vol. 16, p. 17 562 864 231 160 873, 2023. DOI: `10.1177/17562864231160873`.

[4] F. Zilio et al., “Altered brain dynamics index levels of arousal in complete locked-in syndrome”, *Communications Biology*, vol. 6, p. 757, 2023. DOI: `10.1038/s42003-023-05109-1`.

[5] S. S. Gupta et al., “Detecting eye movements in eeg for controlling devices”, in *2012 IEEE International Conference on Computational Intelligence and Cybernetics (CyberneticsCom)*, IEEE, 2012.

[6] K. Karas et al., “Brain-computer interface for robot control with eye artifacts for assistive applications”, *Scientific Reports*, vol. 13, no. 1, p. 17 512, 2023.

[7] L. F. Nicolas-Alonso and J. Gomez-Gil, “Brain computer interfaces, a review”, *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012. DOI: `10.3390/s120201211`.

[8] E. H. Houssein, A. Hammad, and A. A. Ali, “Human emotion recognition from eeg-based brain–computer interface using machine learning: A comprehensive

review", *Neural Computing and Applications*, vol. 34, no. 15, pp. 12 527–12 557, 2022.

[9] E. Niedermeyer and F. L. da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 5th ed. Lippincott Williams & Wilkins, 2005.

[10] M. Teplan, "Fundamentals of eeg measurement", *Measurement Science Review*, vol. 2, no. 2, pp. 1–11, 2002.

[11] M. Fried-Oken, *Brain–computer interface: Current research efforts and considerations for implementation*, Presentation to the Chicago Public School Assistive Technology Group, Accessed: 2025-08-19, Oct. 2023. [Online]. Available: https://www.ohsu.edu/sites/default/files/2023-11/REKNEW-Brain-Computer-Interface-Current-Research-Efforts-and-Considerations-for-Implementation.pdf.

[12] M. A. Lebedev and M. A. L. Nicolelis, "Brain–machine interfaces: Past, present and future", *Trends in Neurosciences*, vol. 29, no. 9, pp. 536–546, 2006. DOI: 10.1016/j.tins.2006.07.004.

[13] J. R. Wolpaw and E. W. Wolpaw, Eds., *Brain–Computer Interfaces: Principles and Practice*. Oxford, UK: Oxford University Press, 2012.

[14] R. Dhiman, "Machine learning techniques for electroencephalogram based brain-computer interface: A systematic literature review", *Measurement: Sensors*, vol. 28, p. 100 823, 2023.

[15] D. Wu, Y. Xu, and B.-L. Lu, "Transfer learning for eeg-based brain–computer interfaces: A review of progress made since 2016", *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 1, pp. 4–19, 2020. [Online]. Available: https://lab.bciml.cn/wp-content/uploads/2020/08/

`Transfer-Learning-for-EEG-Based-Brain-Computer-Interfaces-A-Review-of-Progress-Made-Since-2016.pdf`.

[16] A. N. Belkacem, D.-H. Shin, H. Kambara, N. Yoshimura, and Y. Koike, "Classification of four eye directions from eeg signals for eye-movement-based communication systems", *Journal of Medical and Biological Engineering*, vol. 34, no. 6, pp. 581–588, 2014.

[17] A. Kurapa et al., "A hybrid approach for extracting emg signals by filtering eeg data for iot applications for immobile persons", *Wireless Personal Communications*, vol. 114, pp. 3081–3101, 2020.

[18] W. Shehieb, S. Alansari, and N. Jadallah, "Eeg-based communication system for patients with locked-in syndrome using fuzzy logic", in *2017 10th Biomedical Engineering International Conference (BMEiCON)*, IEEE, 2017.

[19] T. T. Moe, "Designing an eeg based communication system for patients with locked-in syndrome", Norwegian University of Science and Technology (NTNU), Technical Report, 2021.

[20] T. T. Moe, "Towards a communication system for patients with locked-in syndrome based on eeg and visual perception", Master's thesis, Norwegian University of Science and Technology (NTNU), 2022.

[21] M. M. Alam et al., "A survey on the roles of communication technologies in iot-based personalized healthcare applications", *IEEE Access*, vol. 6, pp. 36 611–36 631, 2018.

[22] A. Kübler and N. Birbaumer, "Brain–computer interfaces and communication in paralysis: Extinction of goal-directed thinking in completely locked-in state?", *Clinical Neurophysiology*, vol. 119, no. 11, pp. 2658–2666, 2008.

[23] L. A. Farwell and E. Donchin, “Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials”, *Electroencephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.

[24] Z. Xu, J. Li, R. Gu, and B. Xia, “Steady-state visually evoked potential (ssvep)-based brain-computer interface (bci): A low-delayed asynchronous wheelchair control system”, Lecture Notes in Computer Science, vol. 7663, T. Huang, Z. Zeng, C. Li, and A. C. Leung, Eds., pp. 305–314, 2012. DOI: 10.1007/978-3-642-34475-6_37.

[25] N. Birbaumer et al., “A spelling device for the paralysed”, *Nature*, vol. 398, pp. 297–298, 1999.

[26] A. Bablani, D. R. Edla, D. Tripathi, et al., “An efficient concealed information test: Eeg feature extraction and ensemble classification for lie identification”, *Machine Vision and Applications*, vol. 30, pp. 813–832, 2019. DOI: 10.1007/s00138-018-0950-y.

[27] I. Antoniou, P. Tzimourta, I. Mporas, and G. Grigoriadis, “Eeg-based eye movement recognition using random forests”, *Sensors*, vol. 21, no. 7, p. 2339, 2021. DOI: 10.3390/s21072339.

[28] S. Amiri, A. Rabbi, L. Azinfar, and R. Fazel-Rezai, “A review of p300, ssvep, and hybrid p300/ssvep brain-computer interface systems”, in *Brain-Computer Interface Systems - Recent Progress and Future Prospects*, R. Fazel-Rezai, Ed., IntechOpen, 2013. DOI: 10.5772/56135. [Online]. Available: https://www.researchgate.net/publication/284569024_A_Review_of_P300_SSVEP_and_Hybrid_P300SSVEP_Brain-_Computer_Interface_Systems.

[29] A. N. Belkacem, M. Abo-Zahhad, N. A. Al-Ahdal, A. Alqarni, and A. S. Malik, “Online classification algorithm for eye-movement-based communication

systems using two temporal eeg sensors", *Biomedical Signal Processing and Control*, vol. 16, pp. 40–47, 2015.

[30] T. V. Afonso and F. Heinrichs, "Consumer-grade eeg-based eye tracking: A dataset for evaluating eye movement classification with low-cost hardware", *Preprint*, 2025, Available at Zenodo: `https://zenodo.org/records/14860668`.

[31] S. Åsly, M. Gilde, L. A. Moctezuma, and M. Molinas, "Towards eeg-based signals classification of rgb color-based stimuli", M.S. thesis, 2019, pp. 61–70. DOI: `10.3217/978-3-85125-682-6-61`.

[32] V. J. Lawhern, A. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: A compact convolutional neural network for eeg-based brain–computer interfaces", *Journal of Neural Engineering*, vol. 15, no. 5, p. 056 013, 2018.

[33] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications", *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000. DOI: `10.1016/S0893-6080(00)00026-5`.

[34] P. Samal and M. F. Hashmi, "Role of machine learning and deep learning techniques in eeg-based bci emotion recognition system: A review", *Artificial Intelligence Review*, vol. 57, no. 3, p. 50, 2024.

[35] R. J. Hyndman and Y. Khandakar, "Automatic time series forecasting: The forecast package for r", *Journal of Statistical Software*, vol. 27, no. 3, pp. 1–22, 2008.

[36] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (eeg) classification tasks: A review", *Journal of Neural Engineering*, vol. 16, no. 3, p. 031 001, 2019.

[37] P. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms", *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[38] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: A systematic review", *Journal of Neural Engineering*, vol. 16, no. 5, p. 051 001, 2019.

[39] R. T. Schirrmeister et al., "Deep learning with convolutional neural networks for eeg decoding and visualization", *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[40] S. Gong, K. Xing, A. Cichocki, and J. Li, "Deep learning in eeg: Advance of the last ten-year critical period", *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 2, pp. 348–365, 2022. DOI: 10.1109/TCDS.2021.3079712.

[41] G. Bauer, F. Gerstenbrand, and E. Rumpl, "Varieties of the locked-in syndrome", *Journal of Neurology*, vol. 221, no. 2, pp. 77–91, 1979.

[42] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control", *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[43] T. O. Zander and C. Kothe, "Towards passive brain–computer interfaces: Applying brain–computer interface technology to human–machine systems in general", *Journal of Neural Engineering*, vol. 8, no. 2, p. 025 005, 2011. DOI: 10.1088/1741-2560/8/2/025005.

[44] R. Stodczyk and U. Fischer-Hirchert, "Ambient assisted living: An overview of current applications, end-users and acceptance", *Biomedical Journal of Scientific & Technical Research*, vol. 30, no. 4, pp. 23 706–23 713, 2020. DOI: 10.26717/BJSTR.2020.30.004949.

[45] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for eeg-based brain–computer interfaces: A 10-year update", *Journal of Neural Engineering*, vol. 15, no. 3, p. 031 005, 2018. DOI: 10.1088/1741-2552/aab2f2.

[46] J. N. Mak and J. R. Wolpaw, "Clinical applications of brain–computer interfaces: Current state and future prospects", *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 187–199, 2009. DOI: 10.1109/RBME.2009.2035356.

[47] A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention is all you need", in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[48] C. Li, S. H. Pun, J. W. Li, and F. Chen, "Eeg-based emotion recognition using graph attention network with dual-branch attention module", 2024. DOI: 10.1109/EMBC53108.2024.10782334.

[49] H. He and D. Wu, "Transfer learning for brain–computer interfaces: A euclidean space data alignment approach", *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, pp. 399–410, 2020.