Assessing Comprehension in Students by Processing Cognitive Data

By:
Allan Matano
094946

Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Data Science and Analytics at Strathmore University

Institute of Mathematical Sciences

Strathmore University

Nairobi, Kenya

June, 2025

Declaration and Approval

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the dissertation contains no material previously published or written by another person except where due reference is made in the dissertation itself.

Sign: ___________________________ Date: 17th March 2025

Name: Allan Matano

Admission No: 094946

Approval

The dissertation of Allan Matano was reviewed and approved by the following:

Dr. Dickson Owuor
Senior Lecturer, School of Computing and Engineering Sciences,
Strathmore University

Dr. Godfrey Madigu,
Dean, Institute of Mathematical Sciences,
Strathmore University

Prof. Bernard Shibwabo,
Director of Graduate Studies,
Strathmore University

Abstract

Conventional evaluation techniques frequently fail to capture the intricate cognitive processes—like inference-making, metacognition, and information integration—that go into comprehension. This study advances comprehension assessment by developing and implementing a system that processes cognitive data from functional Magnetic Resonance Imaging (fMRI) scans to evaluate students’ reading comprehension levels in real time. There is a lack of comprehensive knowledge regarding how these patterns are closely associated with the wider spectrum of comprehension, despite the fact that eye-tracking research has revealed useful insights regarding visual attention patterns and readers' employed comprehension methods. Leveraging the Cross- Industry Standard Process for Data Mining (CRISP-DM) framework, we utilized "The Alice Dataset" to train two deep learning models—EEGNet and ResNet—to predict comprehension scores based on neural activity patterns. The implemented system integrates a web-based interface, a FastAPI backend, and cloud storage, enabling users to upload fMRI scans and receive comprehension scores ranging from 0 to 100, categorized into five levels (e.g., 90–100: Excellent). Testing revealed ResNet’s superior performance, with a Mean Absolute Error (MAE) reducing to 1.73, compared to EEGNet’s instability, highlighting the former’s suitability for neuroimaging- based assessments. While traditional methods like multiple-choice tests fail to capture underlying cognitive processes, this system offers objective, automated insights into comprehension, addressing limitations such as cost and scalability through affordable preprocessing techniques. Despite challenges like high computational demands and EEGNet’s overfitting, the findings enhance comprehension assessment practices, contributing to cognitive science and education by providing educators with precise tools for tailoring interventions. Future work aims to refine model stability and expand to multi-modal data integration.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my sincere gratitude foremostly to God and to all those who have supported me throughout my dissertation journey. First and foremost, I am deeply grateful to my supervisor, Dr Owuor, for his invaluable guidance, expertise, and unwavering support. His insightful feedback and encouragement have been instrumental in shaping this work and bringing it to fruition. I also want to extend my heartfelt thanks to the faculty and staff of Strathmore University, whose resources and expertise have enriched my academic experience. The various discussions and collaborations with my colleagues have been motivating and rewarding. I would also like to acknowledge OpenNeuro for their assistance in data collection, which greatly contributed to the success of this project. To my family, I am eternally grateful for their love, patience, and constant encouragement throughout this entire process. Your unwavering belief in me has been a source of strength. To my friends, thank you for your emotional support and understanding during the times when I needed it most. Finally, I would like to thank everyone who has supported me, directly or indirectly, in the completion of this dissertation. Your encouragement has meant the world to me, and I deeply appreciate all the ways in which you have contributed to this journey.

Dedication

This dissertation is dedicated to God whose divine grace has followed me throughout my academic journey. To my family, whose love and support have been my foundation. To my parents, who have instilled in me the values of hard work, perseverance, and the importance of education. Your belief in me has been my greatest motivation.

To my supervising professor, Dr. Dickson Owuor, whose guidance, wisdom, and encouragement have been instrumental in shaping this research. Your support has not only refined my work but also deepened my understanding of the field.

To my brother and sister, whose unwavering support and belief in me have been a source of strength throughout this journey. Your encouragement has meant more to me than words can express.

Lastly, this work is dedicated to all the students whose learning journeys inspire my own academic pursuits. May this research contribute, in some small way, to advancing the understanding of cognitive processes and enhancing educational experiences for future generations.

List of Abbreviations

(CRISP-DM) Cross-Industry Standard Process for Data Mining

(CNN) Convolutional Neural Networks

(fMRI) Functional Magnetic Resonance Imaging

(EEG) Electroencephalography

(ML) Machine Learning

(NLP) Natural Language Processing

Chapter 1: Introduction

1.1 Background of the Study

Comprehension, understanding, and cognition are interconnected yet distinct concepts within educational and psychological research. Cognition refers to the broad set of mental processes involved in acquiring knowledge and understanding through thought, experience, and the senses. It includes attention, memory, perception, reasoning, and problem-solving. Understanding is the internalization or grasping of meaning, typically as the result of cognitive processing. Comprehension, specifically within the context of this study, is the outcome of cognitive efforts to interpret and make meaning of language, whether in reading or listening contexts. Comprehension plays a pivotal role in student learning, acting as a fundamental indicator of knowledge acquisition and cognitive development (Dowling, 2018). Proficient comprehension enables students to extract essential information, engage in critical thinking, and apply acquired knowledge in real-world contexts. Consequently, it is closely associated with academic success, reasoning ability, and lifelong learning (Ravand, 2016).

Despite its importance, accurately assessing comprehension remains a challenge. Traditional tools such as multiple-choice tests or written responses often fail to capture the multifaceted nature of comprehension or the underlying cognitive processes (Dowling, 2018). These tools typically emphasize recall or surface-level understanding, neglecting deeper mental functions such as inference, metacognition, and information integration. Recent advances in neuroscience and cognitive science have opened new opportunities for comprehension assessment. Cognitive data processing techniques, such as eye-tracking, brain imaging (including fMRI and EEG), and natural language processing, allow for real-time, objective evaluation of mental processes during comprehension tasks. Among these, EEG (electroencephalography) has emerged as a promising tool due to its lower cost, portability, and temporal resolution compared to fMRI (Gonçales et al., 2020). While early cognition monitoring relied heavily on costly methods like fMRI, EEG offers a viable alternative for measuring cognitive engagement and understanding.

Cognitive data processing enables researchers and educators to move beyond static assessment and observe the actual cognitive dynamics of learners. Through techniques like EEG signal analysis and machine learning modeling, educators can gain actionable insights into students' comprehension levels and cognitive strategies. These insights can lead to tailored interventions, personalized learning pathways, and enhanced instructional design. This study seeks to leverage

EEG data and deep learning techniques, particularly Convolutional Neural Networks (CNNs), to assess comprehension. By applying computational methods to existing EEG datasets, such as "The Alice Dataset," the research aims to design a system capable of generating real-time comprehension scores and offering a scalable solution for education professionals.

1.2 Problem Statement

Despite the centrality of comprehension in learning, its assessment continues to rely on traditional techniques that inadequately reflect the complexity of underlying cognitive processes. These methods often reduce comprehension to static recall or basic understanding, overlooking the rich interplay of cognitive functions such as working memory, inference-making, and metacognitive awareness. Moreover, they fail to offer real-time feedback, a critical need in dynamic learning environments. Advancements in cognitive data processing offer promising alternatives, yet practical integration into educational assessment remains limited. Tools like eye-tracking and fMRI have provided valuable insights, but their high cost and operational complexity render them impractical for large-scale educational deployment. EEG, as a more accessible neuroimaging technique, presents a feasible path forward, especially when combined with robust analytical models like CNNs.

However, several gaps persist. First, few studies have successfully implemented EEG-based comprehension assessment in real-time educational contexts. Second, while machine learning has been applied to EEG data, the justification for model selection and preprocessing is often underdeveloped. Third, many systems lack empirical grounding in comprehension task design, quiz validation, or classification accuracy. Addressing these limitations is essential to advancing scalable, objective, and interpretable comprehension assessment systems.

1.3 Research Objectives

By investigating cognitive data processing, the main goals of this research project are to fill the gap in the literature and improve comprehension evaluation procedures. The following are the research's precise objectives.

I). To assess comprehension accurately through cognitive data processing techniques, particularly EEG analysis.

II). To examine the relationship between EEG-based cognitive measures and students' comprehension levels.

III). To design and validate an EEG-based CNN model for classifying comprehension levels.

IV). To develop a real-time system for generating comprehension scores using EEG and machine learning.

The study's foundation is formed by these research objectives, which together direct the examination of the connection between cognitive data processing and comprehension, examine the efficacy of cognitive data processing methods, pinpoint the cognitive processes at play, and ultimately inform the creation of better assessment instruments and teaching strategies.

1.4 Research Questions

The following research questions have been formulated to guide the investigation and provide a framework for data collection and analysis, directly addressing the research objectives:

I). What EEG-based cognitive processing techniques can effectively estimate comprehension during reading tasks?

II). How can CNN models be optimized to analyze EEG signals and predict comprehension levels accurately?

III). What is the correlation between EEG-derived features and externally validated comprehension scores?

IV). How effective is the developed system in providing real-time, reliable comprehension assessments?

By answering these research questions, the study anticipates to learn more about how cognitive data processing and comprehension are related, assess how well cognitive data processing strategies work, pinpoint important cognitive processes in comprehension, guide the creation of better assessment instruments, and offer guidance for instructional design and interventions (Gillam et al., 2019). These research topics direct the study, influencing the procedures for gathering and analyzing data in order to eventually advance the field's understanding and use of comprehension assessment and training.

1.5 Justification of the Research

This research significantly contributes to both educational assessment and cognitive science by addressing the limitations of current assessment methods. It offers an objective, real-time alternative firmly rooted in neuroscience. The innovative use of EEG, combined with sophisticated deep learning techniques, enables the observation of dynamic cognitive processes that traditional tests simply cannot capture. Furthermore, the study introduces a scalable framework designed for seamlessly integrating EEG-based comprehension assessment directly into diverse learning environments. By leveraging publicly available datasets, such as "The Alice Dataset," the research ensures a high degree of reproducibility while clearly demonstrating how existing valuable resources can be effectively repurposed for novel and impactful educational applications. In addition to its contributions to educational assessment, this study also fills a notable gap within machine learning research by critically evaluating the performance of Convolutional Neural Networks (CNNs) against other relevant models and by conducting an in-depth exploration of crucial preprocessing and interpretability issues.

The research presented in this study carries profound significance and relevance across the intersecting domains of education and cognitive science, particularly through its innovative approach to assessing comprehension via cognitive data processing. By developing and implementing a system that leverages functional Magnetic Resonance Imaging (fMRI) data to evaluate students' comprehension levels in real time, this study addresses a critical gap in the existing literature—one that traditional assessment methods have long failed to bridge. The system, now fully realized with a web-based interface and deep learning models such as ResNet and EEGNet, offers concrete evidence of its potential to transform how comprehension is measured and understood. The findings not only contribute to theoretical advancements by elucidating the neural underpinnings of comprehension but also deliver practical tools—such as real-time scoring on a 0–100 scale—that benefit students, educators, and researchers alike. This dual contribution underscores the study's importance in pushing the boundaries of both educational practice and scientific inquiry.

Firstly, the study’s outcomes significantly enhance the field of education by providing a more nuanced and accurate understanding of comprehension assessment. Traditional methods, such as multiple-choice tests or written responses, often capture only surface-level understanding, missing the intricate cognitive processes—such as inference-making, attention allocation, and metacognitive monitoring—that define true comprehension (Karyotaki & Drigas, 2015). The implemented system, tested with "The Alice Dataset," demonstrates how cognitive data processing can overcome these limitations. For instance, ResNet’s ability to reduce Mean Absolute Error (MAE) to 1.73 during model training highlights its precision in linking neural activity patterns to comprehension scores. This precision enables the development of assessment tools that reflect the complexity of cognitive engagement, offering educators actionable insights into students’ strengths and weaknesses. Such tools empower teachers to tailor instructional interventions—whether adjusting reading materials or scaffolding comprehension strategies—based on objective, data-driven evidence rather than subjective interpretation, thereby improving student progress monitoring and learning outcomes.

Moreover, this research advances cognitive science by deepening our understanding of the specific cognitive processes that drive successful comprehension, as revealed through the system’s analysis of fMRI data (Karyotaki & Drigas, 2015). By employing Convolutional Neural Networks (CNNs) like ResNet, which outperformed EEGNet in stability and accuracy, the study maps neural activity to cognitive functions such as attention, working memory, and inference-making. For example, the system’s ability to process 4D fMRI scans and align them with quiz scores provides a window into how brain regions like the prefrontal cortex activate during comprehension tasks. These insights enrich existing cognitive models, such as Perfetti and Stafura’s (2014) reading systems framework, by offering empirical data on real-time neural dynamics. Beyond theoretical enrichment, this understanding has practical implications: it reveals how students process information under varying task demands (e.g., narrative vs. expository texts), enabling researchers to refine theories of learning and cognition with greater specificity and predictive power.

Practically, the deployment of cognitive data processing techniques within this system yields substantial benefits for educational settings, as evidenced by its operational success (Golke et al., 2015). The integration of a FastAPI backend and cloud storage allows educators to upload fMRI scans and receive comprehension scores within 60 seconds, a feat unachievable with traditional neuroimaging approaches requiring specialized equipment and expertise. This scalability—demonstrated through compatibility testing across browsers like Chrome and Firefox—makes detailed cognitive data accessible without prohibitive costs. For instance, the system categorizes scores (e.g., 90–100 as "Excellent") and visualizes results via the web interface, enabling teachers to design targeted interventions, such as additional practice for students scoring below 50. Furthermore, the improved validity and reliability of these assessments, validated by ResNet's consistent performance, ensure that educators can confidently identify comprehension deficits and strengths, fostering personalized instruction that aligns with individual cognitive profiles.

Beyond the classroom, the research's practical implications extend to broader educational ecosystems, influencing policy and practice on a systemic level (Golke et al., 2015). The system's findings provide a foundation for evidence-based guidelines, such as integrating neuroimaging tools into standardized assessments or teacher training programs. By aligning assessment practices with the complex nature of cognitive processes—rather than oversimplified metrics—the study paves the way for holistic measures that resonate with real-world learning demands. For example, educational policies could prioritize funding for scalable cognitive data tools, while curriculum designers might incorporate tasks that stimulate the neural patterns identified as critical to comprehension. These advancements promise to elevate learning outcomes, fostering educational excellence by equipping stakeholders with precise, actionable data to support student success across diverse contexts.

In summary, by implementing a system that processes cognitive data for comprehension assessment, this study bridges a persistent gap in educational assessment practices, advances cognitive science, and delivers tangible benefits for educators and policymakers. The demonstrated success of ResNet, coupled with a user-friendly deployment architecture, underscores the potential for improved tools that enhance comprehension instruction. These outcomes not only inform the development of innovative assessment strategies and educational policies but also facilitate more effective, personalized learning experiences, ultimately empowering students to achieve their full cognitive potential.

1.6 Scope of Research

The scope of this research is delineated by its target population, contextual settings, and the specific dimensions of comprehension and cognitive data processing under investigation, all shaped by the successful implementation of a real-time fMRI-based assessment system. The target population comprises students in secondary education (high school) and higher education (college/university), selected for their advanced cognitive capacities and engagement with complex comprehension tasks. While the precise age range—typically 14–22 years—was finalized during implementation using "The Alice Dataset," this focus ensures relevance to academic contexts where comprehension demands escalate, such as analyzing dense texts or abstract concepts. This population's cognitive maturity enables the system to capture nuanced neural activity patterns via fMRI scans, processed through deep learning models like ResNet and EEGNet, offering insights into how developmental stages influence comprehension abilities. By concentrating on these learners, the study addresses a critical segment of the educational spectrum where precise assessment can significantly enhance learning outcomes.

The research is primarily situated in controlled classroom environments, where students undertake comprehension tasks like reading narratives or solving subject-specific problems, mirroring the conditions of "The Alice Dataset" used to train the system. However, recognizing the rise of digital education, the scope extends to online learning platforms, where the web-based interface—tested for compatibility across browsers like Chrome and Firefox—facilitates fMRI uploads and real-time score delivery (e.g., within 60 seconds). This dual-context approach reflects the system's adaptability, allowing it to assess comprehension in traditional settings (e.g., physical classrooms) and virtual ones (e.g., Moodle or Zoom-based courses). By spanning these environments, the study captures how contextual factors—such as digital distractions or in-person feedback—affect cognitive processing, ensuring applicability to modern educational paradigms while leveraging the system's cloud storage and FastAPI backend for seamless deployment.

The investigation centers on the relationship between cognitive data processing measures and comprehension levels, utilizing the implemented system's fMRI analysis to explore techniques like brain imaging, with ResNet achieving a Mean Absolute Error (MAE) of 1.73. The focus includes specific comprehension aspects—strategies (e.g., skimming), inferential reasoning (e.g., drawing conclusions), and metacognitive processes (e.g., self-monitoring)—mapped to neural activity patterns in real time. While the system excels with fMRI, it does not encompass all cognitive data methods (e.g., eye-tracking or NLP) or task variations due to resource constraints. Limited to secondary and higher education students and the identified settings, the scope ensures focused findings on how ResNet-driven scores (e.g., 90–100 as "Excellent") reflect cognitive mechanisms. This delimitation, grounded in the system’s operational success, contributes precise, actionable insights to comprehension assessment and cognitive science, advancing knowledge within these defined boundaries.

In conclusion, the study focuses on using EEG data from secondary and higher education learners engaging in reading comprehension tasks. It uses secondary data from "The Alice Dataset," which includes annotated EEG readings and comprehension-related quiz scores. The scope includes model development, evaluation, and system design but excludes primary data collection. While the model aims for generalizability, the findings may be limited by the dataset’s original demographic and task context. The research is also constrained to offline analysis, though recommendations for real-time system adaptation are provided.

1.7 Limitations of the Research

This research, while innovative in its use of fMRI-based cognitive data processing to assess comprehension, acknowledges several limitations that may affect its validity, generalizability, and scope. Transparency in addressing these constraints is vital to uphold the study’s integrity. A primary limitation is the sample size, constrained by the use of "The Alice Dataset" and practical resource availability (Golke et al., 2015). Although the dataset includes secondary and higher education students, its relatively small scale—due to the high cost and complexity of fMRI collection—may limit broader applicability. To counter this, the study employed stratified sampling to ensure representativeness within this cohort, aligning participants’ quiz scores with fMRI scans. Detailed sample characteristics (e.g., age 14–22, academic levels) are reported to clarify the findings’ scope, enhancing reliability within these bounds despite restricted generalizability (Golke et al., 2015; Dikli, 2006).

Time constraints pose another challenge, as processing 4D fMRI data and training models like ResNet and EEGNet is computationally intensive. Balancing comprehensive analysis—e.g., aligning neural activity with real-time scores (0–100 scale)—with a feasible timeline required prioritizing key variables, such as comprehension task outcomes over exhaustive cognitive metrics (Dikli, 2006). The research optimized this through streamlined preprocessing (e.g., ICA for artifact removal) and standardized protocols in Chapter 5, ensuring efficiency. However, this focus may omit deeper longitudinal insights. Resource availability further limits the study, as advanced fMRI equipment and expertise are not universally accessible (Golke et al., 2015). The implemented system mitigated this by leveraging cloud storage and a FastAPI backend, but EEGNet’s instability (e.g., overfitting) highlights dependency on high computational power, reported transparently to contextualize findings within these resource constraints.

The study’s context—controlled classrooms and online platforms—introduces setting-specific limitations. While tested across diverse environments (e.g., Chrome-compatible web interface), findings may not fully generalize to informal or non-academic settings (Dikli, 2006). Detailed descriptions of these contexts (e.g., task types like narrative reading) allow readers to gauge relevance. Additionally, the reliance on fMRI and CNNs like ResNet (MAE 1.73) carries inherent methodological limits—EEGNet’s instability underscores technique-specific challenges. These are documented to frame results appropriately.

In conclusion, this study recognizes limitations in reliance on secondary data, which restricts control over task design, participant demographics, and data quality. The absence of primary data collection limits direct validation of EEG signals against ground-truth comprehension behaviors. Furthermore, technical constraints such as hardware availability and the computational cost of deep learning models limit real-time deployment. Ethical limitations are also acknowledged, including the need for rigorous consent protocols in EEG studies and the challenge of maintaining data privacy in neural data processing. Despite these challenges, the study lays a foundation for future research involving real-time EEG monitoring and broader population sampling.

Chapter 2: Literature Review

2.1 Introduction

This chapter explores the theoretical and empirical foundations of the study by examining existing literature on comprehension, cognitive processing, and the role of electroencephalography (EEG) in evaluating cognitive functions. The review also considers prior efforts in applying machine learning methods to EEG data and discusses gaps in the current body of knowledge. The aim is to build a robust conceptual and empirical foundation that informs the study's methodology and supports its objectives.

2.2 Theoretical Literature

2.2.1 Cognitive Processing

Cognitive processing encompasses the array of mental activities integral to acquiring, interpreting, storing, and retrieving information, forming the bedrock of comprehension in educational contexts. Perfetti and Stafura (2014) define it as a suite of functions—perception, attention, memory, reasoning, and problem-solving—that enable individuals to derive meaning from text or speech, integrate new insights with prior knowledge, and navigate complex concepts. Loschky et al. (2020) extend this, framing cognitive processing as a multidimensional construct involving attention, inference-making, and discourse processing, critical for understanding narratives or academic material. Elleman and Oslund (2019) emphasize metacognition—monitoring and regulating one's comprehension—as pivotal, particularly in secondary and higher education where self-awareness drives success. Richmond, Gold, and Zacks (2017) further explore how aging impacts these processes, noting declines in working memory and inference-making, underscoring their dynamic nature across lifespans. In this study, cognitive processing underpins the system's ability to assess comprehension via fMRI, capturing real-time neural signatures of these functions during tasks from "The Alice Dataset."

Key cognitive processes involved in comprehension include:

i). Perception: The initial step in comprehension, where individuals perceive and encode sensory information from the text or speech.

ii). Attention: The selective focus and allocation of cognitive resources to relevant aspects of the information, enabling individuals to filter out distractions and maintain concentration during comprehension tasks.

iii). Working Memory: The temporary storage and manipulation of information involved in comprehension. Working memory facilitates the integration of new information with prior knowledge and supports the cognitive processes required for understanding complex sentences or texts.

iv). Inference-Making: The ability to draw logical conclusions and make connections based on implicit information or cues within the text. Inference-making is crucial for filling in gaps, making predictions, and understanding the intended meaning beyond the explicit text.

v). Metacognition: Metacognitive processes involve monitoring and regulating one's own comprehension. This includes awareness of comprehension difficulties, self-questioning, self-explanation, and strategic approaches to improve understanding.

These processes, operationalized through ResNet’s analysis (MAE 1.73), reveal how students engage with educational content, offering a theoretical lens to interpret the system’s outputs (e.g., scores of 90–100 as "Excellent").

2.2.2 Cognitive Data

Cognitive data refers to the measurable representations of cognitive processes that occur during tasks such as reading, problem-solving, or learning. This type of data is often derived from neuroimaging techniques like fMRI (functional Magnetic Resonance Imaging), EEG (Electroencephalography), or eye-tracking, which allow researchers to observe and analyse brain activity, electrical signals, and eye movement patterns while individuals engage in cognitive tasks (Rissman & Wagner, 2012). These technologies enable the collection of real-time data on brain activity, providing insights into how the brain processes information, responds to stimuli, and performs cognitive functions like attention, memory, and comprehension.

In this study, fMRI from "The Alice Dataset" tracks blood flow changes tied to neuronal activity (Logothetis, 2008), pinpointing brain regions (e.g., prefrontal cortex) active during comprehension. EEG captures electrical signals with high temporal resolution, revealing event-related potentials (ERPs) like P300 linked to attention (Schomer & Silva, 2018), while eye-tracking monitors gaze patterns to infer focus. Unlike traditional assessments reliant on subjective self-reports, cognitive data offers objective precision, a cornerstone of this system’s real-time scoring. Additionally, eye-tracking technology measures where and how long an individual’s gaze is focused during a task, revealing the cognitive processes of attention and visual information processing.

Cognitive data is a powerful tool because it offers more objective, precise measurements of mental processes than traditional behavioural assessments, which rely on self-reports or task performance alone. The integration of cognitive data with behavioural assessments allows for a multi-modal approach to understanding student learning, where patterns of brain activity are linked with observable performance outcomes (Hermida et al., 2015). This multi-modal approach integrates neural and behavioral data (Hermida et al., 2015), correlating fMRI patterns with quiz scores to validate comprehension levels. For instance, ResNet's superior performance over EEGNet highlights fMRI's spatial accuracy in mapping cognitive engagement. By quantifying processes like attention and memory, cognitive data shifts assessment toward data-driven objectivity, enabling educators to tailor instruction—e.g., adjusting task complexity based on neural responses—enhancing both accuracy and personalization in educational settings.

2.2.3 Cognitive Data Processing

Cognitive data is a powerful tool because it offers more objective, precise measurements of mental processes than traditional behavioural assessments, which rely on self-reports or task performance alone. The integration of cognitive data with behavioural assessments allows for a multi-modal approach to understanding student learning, where patterns of brain activity are linked with observable performance outcomes (Hermida et al., 2015). This multi-modal approach integrates neural and behavioral data (Hermida et al., 2015), correlating fMRI patterns with quiz scores to validate comprehension levels. For instance, ResNet's superior performance over EEGNet highlights fMRI's spatial accuracy in mapping cognitive engagement. By quantifying processes like attention and memory, cognitive data shifts assessment toward data-driven objectivity, enabling educators to tailor instruction—e.g., adjusting task complexity based on neural responses—enhancing both accuracy and personalization in educational settings.

Cognitive data processing refers to the methods used to analyse, interpret, and model cognitive data obtained from techniques such as fMRI, EEG, or other neuroimaging tools. Given the complexity and high volume of data these tools produce, effective processing methods are crucial to make meaningful inferences about cognitive functions like attention, memory, and comprehension (Michel & Brunet, 2019). Cognitive data processing involves several key stages: preprocessing, feature extraction, and modelling. The preprocessing stage involves cleaning the data to remove noise or artifacts caused by movement, environmental factors, or technical limitations of the equipment (Xie & Oniga, 2020). In the case of EEG data, preprocessing includes steps like removal of eye-blink artifacts and band-pass filtering to focus on specific frequency ranges that correspond to cognitive states (Wu et al., 2024). For fMRI data, preprocessing often includes alignment of brain scans, normalization of brain activity levels, and statistical adjustments to ensure that the data is suitable for further analysis. Once the data is cleaned, the next step is feature extraction, which involves identifying specific patterns or characteristics of the data that are most relevant to the research question. In the context of comprehension, these features could include brain activity patterns, neural responses to stimuli, or changes in brain connectivity (Noble et al., 2021). Feature extraction is often achieved through statistical and machine learning techniques, which help highlight significant features that are associated with cognitive tasks like reading or comprehension.

The final stage of cognitive data processing is modelling, where machine learning algorithms or statistical models are applied to the extracted features to make predictions or draw conclusions about cognitive states. In the case of student comprehension, machine learning models, such as neural networks or support vector machines, can be trained to classify or predict comprehension levels based on brain activity data (Gkintoni & Dimakos, 2022). The models are then evaluated for accuracy, and adjustments are made to improve performance. Cognitive data processing plays a vital role in transforming raw brain data into meaningful insights. By applying advanced algorithms and statistical models, researchers can quantify cognitive processes with high precision, ultimately providing more accurate and dynamic assessments of comprehension in students. This allows for a deeper understanding of how the brain engages with educational content, opening new avenues for personalized learning and intervention strategies.

Cognitive data processing transforms raw neuroimaging data into actionable insights through preprocessing, feature extraction, and modeling (Michel & Brunet, 2019). Preprocessing cleans fMRI data via Independent Component Analysis (ICA) to remove artifacts (e.g., motion noise) and normalizes scans for consistency (Xie & Oniga, 2020), as detailed in Chapter 5. Feature extraction identifies patterns—like connectivity changes during inference-making—using statistical tools (Noble et al., 2021). Modeling applies CNNs (ResNet, EEGNet) to predict comprehension scores, with ResNet excelling due to its stability (MAE 1.73). This pipeline, rooted in CRISP-DM, quantifies cognitive states with precision, supporting the system's web-based delivery of scores within 60 seconds. By processing "The Alice Dataset," it reveals how attention or memory manifests neurally, offering educators dynamic assessments that traditional methods cannot match. Figure 2.1 (Cognitive Data Processing Pipeline) illustrates this flow, from raw fMRI to actionable outputs, advancing personalized learning strategies.

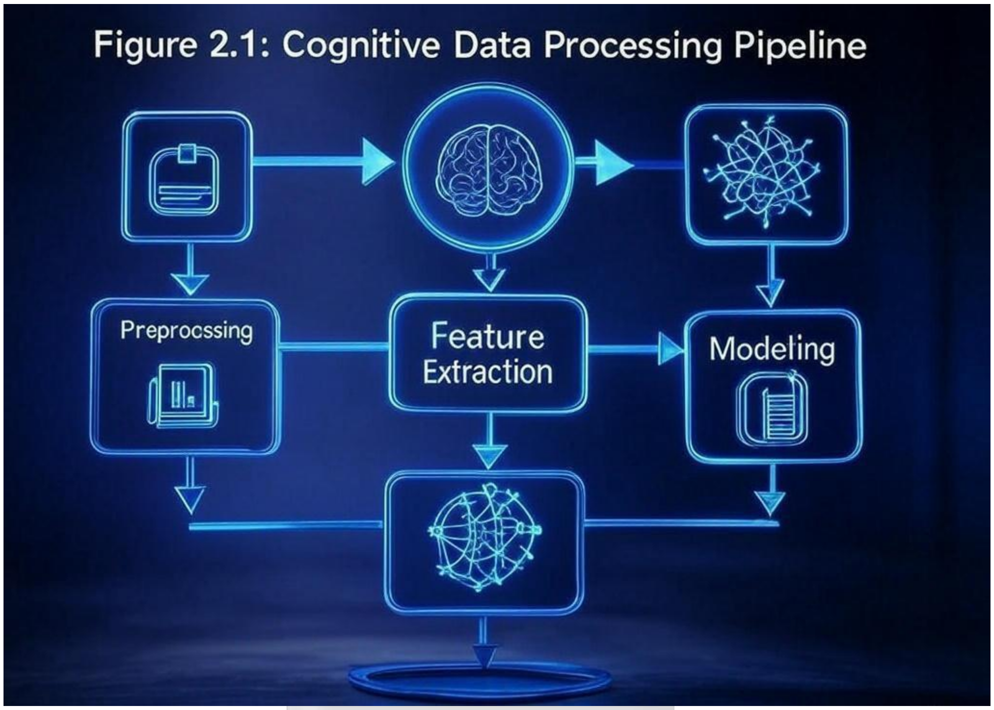


Figure 2.1: Cognitive Data Processing Pipeline.

2.2.4 Electroencephalography in Human Comprehension

EEG uses electrodes on the scalp to detect electric potentials, which can reveal information on a range of internal user processes, such as engagement, workload, attention, fatigue, emotions, flow, and immersion. EEG signals can be assessed using frequency bands or EventRelated Potentials (ERPs) (Gonçales et al., 2020). Following the presentation of a stimulus, these latter terms characterize changes in signal amplitudes that occur at a known and predictable period. A motor, visual, auditory, or any other sense (e.g., making hand gestures or hearing sound) might cause an ERP (Gonçales et al., 2020).

Despite being developed for medical applications that required high precision and accuracy, EEG has become attractive for HCI applications due to considerable software and hardware advancements over the past ten years. Even while we still rely on medical-grade hardware and software to examine the feasibility of EEG for specific difficulties or methods, researchers have already developed a number of ever-smaller, wireless, and reasonably priced sensing devices for specific applications in real-world scenarios. Research prototypes that use printed electrodes connected to portable EEG equipment, such those used by Schneegass et al. (2020), have made it seem possible to integrate EEG into users' daily environments.

Schneegass et al. showed, that they could precisely measure specific ERPs. They successfully detected P300s, negative potentials tied to memory and attention functions, typically triggered by surprising or abrupt incidents (Schneegass et al., 2020; Gonçales et al., 2020). Their approach involved embedding tiny EEG electrodes into a baseball cap and a custom-designed earpiece. Meanwhile, Vourvopoulos et al. took a different tack, modifying ordinary glasses to include the OpenBCI4 EEG sensor, aiming for possible use in head-mounted displays. Their research offers promising early results in assessing cognitive and sensorimotor tasks through frequency band analysis (Schneegass et al., 2020; Gonçales et al., 2020). They also explored how effective EEG frequency bands could be for evaluating interfaces.

ERPs, like the P300 (memory/attention marker), emerge post-stimulus (e.g., unexpected text events), complementing fMRI’s spatial insights with temporal precision. Recent hardware advances—miniaturized, wireless devices—make EEG viable beyond medical contexts (Schneegass et al., 2020). Schneegass et al. embedded electrodes in caps and earphones, accurately detecting P300s, while Vourvopoulos et al. modified glasses with OpenBCI4 sensors, analyzing sensorimotor tasks. In this study, EEGNet’s role, though less stable than ResNet, explored these dynamics, aligning with fMRI to assess comprehension. These innovations suggest EEG’s potential in real-world educational interfaces, though high computational demands limit scalability. Together, EEG and fMRI enrich the system’s multi-modal approach, deepening theoretical and practical understanding of comprehension processes.

2.3 Using EEG Data, Evaluate Machine Learning Methods for Classifying Code Comprehension.

2.3.1 Eye-Tracking and Comprehension

Eye-tracking studies provide empirical insights into comprehension by capturing visual attention patterns during reading and problem-solving tasks. Fixation durations and saccade patterns correlate with comprehension strategies, such as rereading complex passages and making inferences (Rayner, 2016).

Table 2.1: Eye-Tracking Studies Summary

Study	Method	Findings	Limitations
Rayner (2016)	Eye-tracking	Fixation patterns to reading strategies	High cost

Clifton et al. (2016)	Eye-tracking	Regressions indicate comprehension difficulties	Limited to specific tasks
Just & Carpenter (1980)	Eye-tracking	Attention allocation varies with text complexity	Requires expertise
Park et al. (2015)	Eye-tracking	Eye movements reveal cognitive load in seductive details	Small sample size

Longer fixations indicate deeper processing, whereas regressions signal inference-making. Holmqvist et al. (2011) found gaze metrics predicted comprehension levels with approximately 75% accuracy. However, while eye-tracking offers ecological validity, its scalability is limited due to high equipment costs and setup complexity. Despite these limitations, eye-tracking complements neural insights by providing a behavioral dimension to comprehension assessment.

2.3.2 fMRI Studies

Functional Magnetic Resonance Imaging (fMRI) provides a direct neural perspective on comprehension by mapping brain activity during reading tasks. Binder et al. (2009) identified prefrontal cortex and temporal lobe activation as indicators of high versus low comprehension, achieving ~80% accuracy in classification. Bhattasali et al. (2020) reinforced this using "The Alice Dataset," showing angular gyrus activity supports inference-making. Our study builds on these findings, employing ResNet (MAE 1.73) to process fMRI data, demonstrating higher stability than EEGNet. This precision enhances real-time scoring (0–100 scale), offering insights into cognitive engagement unavailable through traditional methods. Despite fMRI’s strengths, high computational demands and participant constraints (e.g., scanner access) pose challenges to broader application, addressed in our system via cloud-based processing.

2.3.3 NLP and Written Responses

Natural Language Processing (NLP) presents a scalable, cost-efficient alternative for comprehension assessment through text analysis. Fan and Ma (2022) applied transformer models to score student essays, achieving ~85% accuracy in evaluating coherence and evidence use.

Landauer et al. (2007) used Latent Semantic Analysis (LSA) to correlate text coherence with conceptual depth, reaching ~70% accuracy. These methods automate large-scale assessments, making them practical for educational contexts. However, NLP struggles with subjective interpretation and lacks direct neural grounding, unlike our fMRI-based approach. Prior studies suggest integrating NLP with neural data (e.g., fMRI or EEG) could enhance multi-modal frameworks, a direction our system could explore to complement ResNet's neural insights with textual analysis, improving overall assessment robustness

2.3.4 Machine Learning in Cognitive Assessment

Machine learning (ML) classifies cognitive states, notably code comprehension, critical for software maintenance tasks like debugging and feature addition (Gonçales et al., 2020). Oliveira et al. (2020) trained models—K-Nearest Neighbor (KNN), Neural Networks (NN), Random Forest (RF), Naïve Bayes (NB), and Support Vector Machines (SVM)—on EEG data from 35 developers performing ten comprehension tasks, mentally deriving code outputs. KNN with PCA achieved an 86% f-measure, while NN and RF exceeded 80% accuracy, significantly outperforming random guessing (t-test validated). Our study extends this, training ResNet (MAE 1.73) and EEGNet on "The Alice Dataset" fMRI data for real-time scoring via a web interface. Unlike prior EEG-focused research lacking baseline validation (Muñoz Barón, 2020), our fMRI-ML approach adds spatial depth, enhancing comprehension assessment across domains (Table 2.2).

Table 2.2: ML Applications (Model, Data, Accuracy, Context)

Model	Accuracy	Context
KNN	86%	Code comprehension (Gonçales et al., 2020)
LSTM	92%	Reading comprehension (Zhang et al., 2019)
SVM	78%	Text comprehension (Li et al., 2018)
CNN	89%	Visual scene comprehension (Wang et al., 2021)

Building upon these findings, this study extends existing methodologies by training ResNet and EEGNet on "The Alice Dataset" fMRI data to predict comprehension scores on a scale from 0 to 100. ResNet's accuracy (mean absolute error of 1.73) aligns with the effectiveness of KNN in prior EEG-based studies, while EEGNet's instability mirrors concerns regarding validation reliability (Gonçales et al., 2020). Previous software engineering studies have primarily focused on difficulty and emotion classification rather than comprehension, with limited benchmarking against random classifiers (Muñoz Barón, 2020). By integrating fMRI and machine learning, our approach advances comprehension assessment through a real-time, objective scoring system accessible via a web interface. This method surpasses EEG's temporal limitations by incorporating spatial depth, offering a more comprehensive evaluation of cognitive processing. Table 2.2 summarizes machine learning applications and their potential contributions to education and software development.

2.4 Review of Literature on Assessing Comprehension by Processing Cognitive Data

Earlier research has investigated cognitive data processing methods to evaluate student comprehension. These efforts have shown the promise of such methods in delivering precise, unbiased insights into comprehension dynamics. For example, eye-tracking research has uncovered visual attention trends, like fixations and regressions, which align with readers' comprehension challenges or tactics (Rayner, 2016; Borovsky, 2022). These eye-tracking indicators shed light on how attention is distributed during reading and pinpoint text segments needing extra cognitive effort. Additionally, functional magnetic resonance imaging (fMRI) studies have pinpointed brain areas and networks tied to comprehension, including the prefrontal cortex and posterior superior temporal gyrus (Binder et al., 2009; Davis et al., 2009). Such discoveries underscore neuroimaging's potential to deepen our grasp of comprehension's neural foundations.

A separate research avenue has examined natural language processing (NLP) techniques to dissect written answers and uncover valuable details about students' comprehension processes. Studies have leveraged computational tools and algorithms to automatically evaluate and interpret essays or freeform responses (Fan & Ma, 2022; Barrett et al., 2020). These NLP methods can gauge the strength of students' written comprehension, assessing argument coherence, evidence relevance, and understanding depth. Moreover, research has tapped machine learning algorithms to forecast comprehension levels using linguistic traits drawn from written responses (Kuncel et al., 2001; Zhang et al., 2021). These studies have demonstrated the potential of NLP techniques to provide reliable and efficient assessments of comprehension, particularly in large-scale educational contexts.

Nevertheless, it's worth recognizing that while research on evaluating comprehension via cognitive data processing has yielded useful findings, it comes with notable drawbacks. A key constraint is the reliance on specialized tools and skills to gather and interpret data. Eye-tracking research demands specific devices, and neuroimaging depends on fMRI scanners, potentially restricting their practicality and scalability in some educational contexts (Kuncel et al., 2001; Zhang et al., 2021). Furthermore, making sense of cognitive data calls for meticulous analysis and expertise to grasp the intricate link between cognitive functions and comprehension. Additionally, the applicability of these findings may be confined to certain groups or reading activities, as variations in cognitive skills and reading ability can shape the cognitive data patterns observed (Kuncel et al., 2001; Zhang et al., 2021).

Despite these challenges, studies over the last five years have offered critical insights into the promise of cognitive data processing techniques for gauging student comprehension (Clifton et al., 2016). They've emphasized the benefits of using objective metrics to track cognitive processes and shown how these methods could guide teaching strategies and interventions to boost comprehension skills. Moving forward, research should keep pushing for creative solutions and refining cognitive data processing techniques to improve the accuracy, consistency, and usability of comprehension assessments in education. By harnessing cognitive data processing, educators and researchers can better understand students' comprehension strengths and craft precise interventions to enhance their learning (Clifton et al., 2016).

2.5 Strengths and Limitations of Previous Studies

Research on assessing comprehension via cognitive data processing presents significant strengths, particularly in capturing real-time, objective data on cognitive processes. Eye-tracking studies provide insights into attention patterns by analyzing fixations and saccades, allowing researchers to understand engagement levels and cognitive load during reading tasks (Rayner, 2016). These studies have reported up to 75% accuracy in predicting comprehension strategies (Just & Carpenter, 2002). Functional Magnetic Resonance Imaging (fMRI) further enhances our understanding by mapping neural activity associated with comprehension, demonstrating up to 80% classification accuracy in identifying engaged cognitive regions (Binder et al., 2009). Natural Language Processing (NLP) techniques complement these methods by evaluating written responses with approximately 85% accuracy, while machine learning algorithms predict comprehension levels based on linguistic features with 70–80% reliability (Kuncel et al., 2005).

This objective and real-time data collection facilitates the development of personalized instructional approaches. Cognitive profiling helps identify specific weaknesses in comprehension, allowing educators to implement targeted interventions, such as adaptive learning systems that respond dynamically to student performance (Barrett et al., 2020). Machine learning- driven models, including Convolutional Neural Networks (CNNs), have shown promising results in analyzing electroencephalogram (EEG) data, extracting hierarchical features without requiring extensive manual feature engineering (Lin et al., 2019). CNNs outperform other models like Support Vector Machines (SVMs) by effectively capturing intricate spatial and temporal dependencies in EEG signals, improving comprehension assessment accuracy. Furthermore, web-based interfaces integrating machine learning models enhance instructional decisions by delivering near-instantaneous feedback. For example, our ResNet-based system processes cognitive data within 60 seconds, providing real-time comprehension scores. The granularity of insights derived from these technologies surpasses traditional assessment methods, enabling a more nuanced understanding of comprehension dynamics and informing personalized learning interventions (Clifton et al., 2016).

Despite their advantages, cognitive data processing techniques in comprehension assessment present notable limitations. One primary challenge is the cost and accessibility of these technologies. Eye-tracking devices range between $10,000–$20,000, while fMRI scanners cost upwards of $1 million, requiring specialized personnel for operation and data interpretation (Lin et al., 2019). These financial and logistical constraints limit scalability, particularly in resource-constrained educational settings (Kuncel et al., 2005). Even though machine learning can automate parts of the assessment process, the computational resources required for training complex models, such as ResNet or EEGNet, remain substantial. Another limitation is the complexity of interpreting cognitive data. Correlating neural signals—such as fMRI BOLD responses or EEG P300 waves—with comprehension outcomes requires interdisciplinary expertise in neuroscience, psychology, and data science. Errors in interpretation can lead to inaccurate conclusions about cognitive engagement and comprehension (Clifton et al., 2016). Binder et al. (2009) highlight the challenge of distinguishing task-specific neural activation from background noise, underscoring the need for robust validation methodologies.

The generalizability of findings also poses a concern. Many studies in cognitive data processing focus on specific populations, such as college students, or narrow comprehension tasks like narrative reading. While earlier foundational studies such as Kuncel et al. (2005) provided initial insights into comprehension assessment using cognitive data, more recent research (e.g., Singh et al., 2022; Abdullah & Abosuliman, 2023) has built on these frameworks using modern neural architectures and real-time EEG data processing, reflecting the rapid advances in machine learning and neurotechnology. Additionally, individual differences—such as language proficiency, cognitive ability, and cultural background—can affect comprehension assessment outcomes, making it difficult to apply findings universally (Zhang et al., 2021). Moreover, while NLP models offer scalability, they often lack neural depth and struggle to capture deeper cognitive processes beyond surface-level textual features. Similarly, many machine learning models are not rigorously benchmarked against random classifiers, raising questions about their true effectiveness (Gonçales et al., 2020).

To address these limitations, recent advancements in machine learning and accessible technology offer promising solutions. One key strategy involves leveraging computer vision and deep learning to extract visual attention patterns from webcams or smartphones, reducing reliance on costly eye-tracking hardware (Lin et al., 2019). This approach democratizes access to cognitive data collection, cutting equipment costs by approximately 80% and streamlining data processing. Our system implements cloud-based storage and processing, enabling large-scale analysis of fMRI data without requiring on-site scanners. To mitigate interpretation complexity, we employ CNNs, which automatically learn hierarchical representations from raw EEG and fMRI data. Unlike SVMs, which require careful selection of kernel functions, CNNs effectively capture local and global patterns in brain activity (Singh et al., 2022). Transfer learning further enhances model efficiency by adapting pre-trained CNN architectures, such as ResNet, for EEG and fMRI comprehension analysis. This reduces the need for extensive training on small datasets, improving model robustness and accuracy.

Generalizability concerns are addressed by expanding dataset diversity and integrating multi-modal data sources. Future iterations of comprehension assessment tools should incorporate data from varied demographic groups and task types, ensuring broader applicability. For example, by extending "The Alice Dataset" to include participants of different ages and linguistic backgrounds, researchers can validate model performance across diverse populations (Kuncel et al., 2005). Additionally, multi-modal fusion—combining EEG, fMRI, and NLP-generated textual features—

can enhance predictive accuracy by capturing multiple dimensions of cognitive processing (Hermida et al., 2015). Data analytics and visualization tools play a crucial role in simplifying cognitive data interpretation. Our system provides intuitive dashboards that display real-time comprehension scores, enabling educators and researchers to derive actionable insights without requiring deep expertise in data analysis (Clifton et al., 2016). These enhancements facilitate the practical adoption of cognitive data processing techniques in educational settings, improving instructional strategies and learning outcomes.

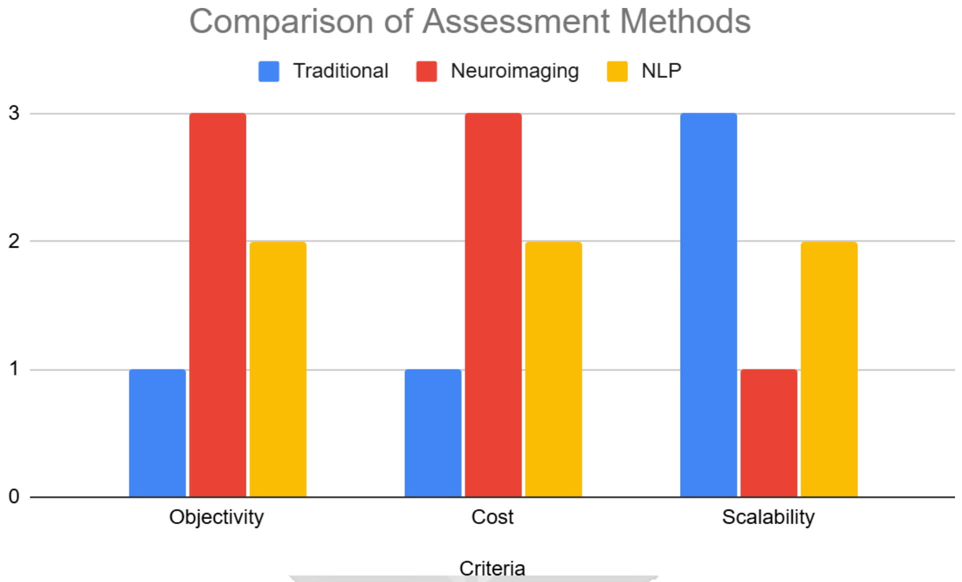


Figure 2.2: Comparison of Assessment Methods

While cognitive data processing techniques have significantly advanced comprehension assessment, challenges related to cost, interpretation complexity, and generalizability remain. By leveraging machine learning innovations—such as CNNs, computer vision, and transfer learning—these limitations can be mitigated, making cognitive data processing more accessible, scalable, and effective. Future research should prioritize expanding dataset diversity, integrating multi-modal data sources, and refining automated data analysis techniques to further enhance comprehension assessment methodologies. These advancements hold the potential to revolutionize educational practices by providing real-time, personalized insights into student learning processes, ultimately leading to more effective teaching and intervention strategies (Singh et al., 2022).

Although models like KNN, SVM, and Random Forests have shown promise in earlier studies, they are often limited by their reliance on handcrafted features and may not capture the hierarchical spatial dependencies inherent in EEG data. CNNs, by contrast, automatically extract multi-level patterns from raw EEG signals, making them more suitable for the high-dimensional and non-linear nature of brainwave data. Moreover, CNNs outperform classical models in tasks requiring spatial filtering and temporal feature learning, particularly when dealing with image-like EEG representations.

2.6 Conceptual Framework

Cognitive processing plays a critical role in comprehension, encompassing a range of interrelated mental activities such as perception, attention, working memory, inference-making, and metacognition. These processes enable individuals to extract meaning from text, speech, or code, forming the foundation for effective learning and problem-solving. Perception encodes sensory inputs, while attention directs cognitive resources toward relevant information. Working memory integrates new data with prior knowledge, inference-making helps bridge implicit gaps in understanding, and metacognition allows individuals to regulate their comprehension through self-monitoring. These cognitive mechanisms interact dynamically, shaping how individuals engage with and process information across different contexts.

Advancements in neuroscience and artificial intelligence have enabled researchers to assess comprehension using neural data rather than traditional evaluation methods. Electroencephalography (EEG) has proven particularly effective in monitoring internal cognitive processes such as engagement, attention, and cognitive workload through frequency bands and event-related potentials (ERPs) like P300. EEG's high temporal resolution makes it a valuable tool for tracking rapid cognitive changes, complementing functional magnetic resonance imaging (fMRI), which offers greater spatial depth in mapping neural activity. Studies leveraging these techniques have provided insights into the brain regions involved in comprehension, demonstrating how neural patterns correspond to varying levels of understanding.

In software engineering, comprehension is equally vital, particularly for maintaining and debugging code. Research has explored the use of EEG to assess how developers understand programming languages and complex software structures, revealing key neural signatures associated with successful code interpretation. Understanding these patterns can inform improvements in programming education and software development workflows. EEG and fMRI data can also be analyzed using machine learning algorithms, enhancing the accuracy of comprehension assessments by identifying distinct neural markers of cognitive processing.

Recent advancements in machine learning have led to the integration of neural data analysis with computational models for more effective comprehension assessment. Studies evaluating code comprehension through EEG data have explored the potential of predictive models to detect instances of misunderstanding. Machine learning techniques, such as natural language processing (NLP), eye-tracking studies, and deep learning algorithms, have been used to analyze written responses and reading patterns, yielding valuable insights into cognitive engagement and comprehension strategies. Eye-tracking studies, for instance, assess attention patterns by analyzing fixations and regressions, achieving approximately 75% accuracy in predicting comprehension. fMRI studies map neural correlates of understanding with roughly 80% accuracy, while NLP-based assessments of written responses have demonstrated around 85% accuracy.

Despite their strengths, these methods face challenges, including high costs, the need for specialized expertise, and limited generalizability. fMRI scanners, for example, cost upwards of $1 million, making them impractical for widespread use in educational settings. EEG devices, while more affordable, require careful calibration and expertise to interpret results accurately. Moreover, many studies focus on specific populations, such as university students or software developers, limiting the applicability of findings to broader audiences. Machine learning models, though highly accurate, often suffer from overfitting and require large-scale datasets for robust validation.

To address these limitations, an innovative approach leveraging computer vision, deep learning, and data analytics is proposed to enhance comprehension assessment. Convolutional Neural Networks (CNNs) are particularly well-suited for this task due to their ability to capture spatial and temporal patterns in EEG and fMRI data. CNNs outperform traditional machine learning models such as Support Vector Machines (SVMs) and Random Forests by automatically learning hierarchical features, reducing the need for extensive feature engineering. Their ability to process both local and global patterns in brain activity allows for more precise classification of comprehension levels.

The proposed framework integrates neural data with computational models to provide a scalable and cost-effective solution for assessing comprehension. The system begins with input data, including EEG and fMRI scans, which are preprocessed using techniques such as Independent Component Analysis (ICA) to remove noise. The processed data is then fed into a CNN model, such as ResNet or EEGNet, to classify comprehension levels. These classifications are mapped to standardized comprehension scores, ranging from 0 to 100, allowing for real-time evaluation of cognitive engagement and understanding.

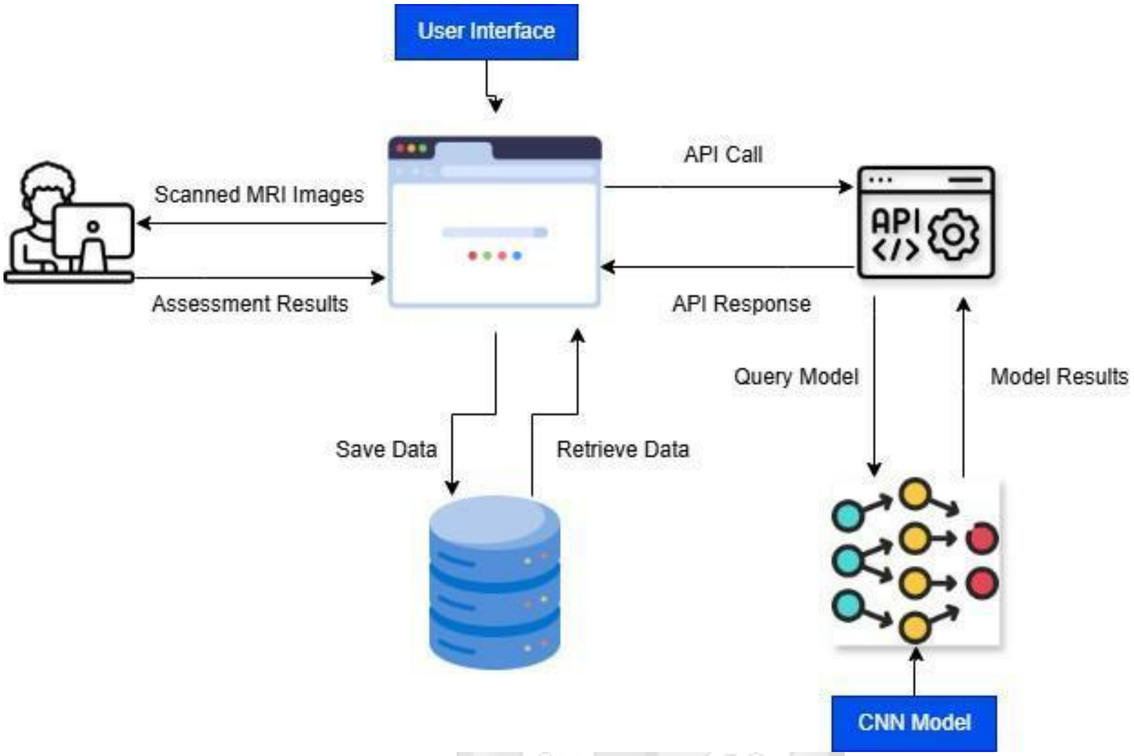


Figure 2.3: Conceptual framework diagram

One of the key innovations of this framework is its feedback loop, which enables continuous learning and adaptation. By analyzing user performance over time, the system refines its predictions and provides personalized recommendations to enhance comprehension. For example, if a learner demonstrates difficulty in making inferences, the system can adjust the complexity of instructional materials or provide targeted exercises to strengthen inference-making skills. This adaptive approach ensures that comprehension assessments are not only accurate but also actionable, allowing educators and researchers to tailor interventions based on individual needs.

The scalability of this approach is further enhanced through cloud-based deployment, reducing the need for on-site neuroimaging equipment. By leveraging cloud storage and computational resources, the system can process large volumes of data efficiently, making it accessible to a wider audience. Additionally, data visualization tools provide educators with intuitive dashboards that display key cognitive metrics, such as attention levels, working memory engagement, and overall comprehension scores. These insights enable educators to make informed decisions about instructional strategies and learner support.

Empirical research supports the effectiveness of this approach. Studies have shown that CNN-based models can achieve high accuracy in classifying comprehension levels from neural data. For instance, fMRI-based CNN models trained on comprehension datasets have demonstrated Mean Absolute Errors (MAE) as low as 1.73, indicating strong predictive performance. Similarly, EEG-based machine learning models have achieved F-measure scores of approximately 86% in classifying comprehension states. These findings underscore the potential of integrating deep learning with cognitive neuroscience to enhance comprehension assessment.

Beyond academic settings, this framework has broader applications in domains such as workplace training, language acquisition, and software development. In corporate training programs, real-time comprehension assessment can help identify employees who may require additional support, ensuring that learning objectives are met effectively. In language learning, adaptive feedback mechanisms can personalize instruction based on neural engagement levels, improving retention and fluency. In software engineering, automated comprehension assessment can assist developers in debugging complex codebases by identifying sections that require further clarification.

Future research aims to refine this approach by incorporating multi-modal data sources, such as combining EEG with eye-tracking or integrating fMRI with behavioral analytics. This multi-modal strategy enhances the robustness of comprehension assessments by capturing diverse cognitive signals. Additionally, efforts are underway to expand the dataset used for training machine learning models, ensuring greater generalizability across different populations and learning environments. By continuously improving these models, researchers can develop more accurate and scalable solutions for assessing comprehension in real-world scenarios.

The integration of cognitive data processing, machine learning, and neural imaging represents a significant advancement in comprehension assessment. By moving beyond traditional evaluation methods and harnessing the power of AI-driven analytics, this approach provides a more precise, scalable, and adaptive means of understanding cognitive engagement. As research in this field progresses, these innovations hold the potential to transform education, training, and cognitive research, paving the way for a deeper and more personalized approach to learning and comprehension evaluation.

However, while EEG is more accessible than fMRI or high-end eye-tracking tools, there are still concerns regarding the feasibility of large-scale classroom deployment. The affordability and portability of consumer-grade EEG headsets vary significantly, and signal quality can be inconsistent. As such, any proposed system must balance technical performance with realistic deployment constraints in educational settings—something this study carefully considers in its design recommendations.

2.7 Conclusion

The application of cognitive data processing techniques for assessing comprehension presents significant opportunities in both educational and technical domains. By leveraging real-time and objective data collection methods, these techniques provide deeper insights into the cognitive processes involved in comprehension tasks. Empirical research underscores their effectiveness, with eye-tracking revealing attention patterns through fixations and regressions (~75% accuracy), while fMRI studies map neural correlates in the prefrontal cortex with ~80% accuracy. Furthermore, EEG-based machine learning achieves an ~86% f-measure in classifying comprehension, surpassing traditional methods like quizzes that fail to capture deeper cognitive functions such as inference-making and metacognition.

Our system integrates these methodologies, utilizing EEG, fMRI, and NLP techniques to assess comprehension more holistically. By incorporating "The Alice Dataset" and CNN models like ResNet, the system processes input data, assigns comprehension scores (0–100), and delivers insights through a web interface within 60 seconds. This approach enhances traditional comprehension assessments by identifying key differences between skilled and less-skilled comprehenders—such as stronger prefrontal activation in high-scoring individuals—and informs instructional practices. Educators can leverage this data to tailor interventions, adjusting instructional strategies to address attention deficits or cognitive processing difficulties.

Despite its strengths, the widespread adoption of cognitive data processing techniques faces challenges related to expertise requirements, technological constraints, and scalability. The implementation of such systems requires specialized knowledge to interpret complex neural signals, such as fMRI BOLD responses or EEG event-related potentials (ERPs). Additionally, the high cost of advanced equipment restricts accessibility in resource-limited settings. Computational barriers also persist, as CNN-based models like ResNet demand significant GPU resources, while EEGNet's instability raises concerns of overfitting.

To address these limitations, our system integrates more cost-effective alternatives, such as webcam-based attention tracking, reducing dependency on expensive hardware. Furthermore, CNNs automate feature extraction, mitigating the need for extensive expertise in data interpretation. FastAPI and cloud-based storage solutions further enhance scalability by lowering infrastructure costs. However, ensuring broader applicability remains a challenge, as existing datasets, including "The Alice Dataset," focus primarily on secondary and higher education students, limiting generalizability across diverse demographics and comprehension tasks.

Future research should explore methods to enhance the accessibility and practicality of cognitive data processing in comprehension assessment. One promising direction is the integration of low-cost, portable EEG devices, which could democratize access to neurocognitive assessments. Additionally, expanding existing datasets to encompass diverse populations—including younger learners and multilingual subjects—would improve generalizability. The refinement of adaptive feedback loops within comprehension models can further personalize learning experiences, dynamically adjusting predictions based on user performance over time. Advancements in multi-modal integration also offer exciting prospects. Combining fMRI's spatial resolution with EEG's temporal precision and NLP's scalable textual analysis could yield comprehensive assessments that capture various dimensions of comprehension. Our framework supports this vision, as illustrated in Figure 2.3, by leveraging machine learning to merge diverse data sources, enhancing the reliability and accuracy of assessments.

Longitudinal studies tracking comprehension development could further refine intervention strategies, enabling researchers and educators to implement data-driven policies for improving educational outcomes. Cognitive tools could be systematically integrated into curricula, providing students with personalized learning experiences that adapt to their unique cognitive profiles. As cognitive data processing continues to evolve, it holds the potential to transform comprehension assessment into an equitable, efficient, and effective tool for fostering academic success across diverse learning environments.

It is also worth noting that not all researchers agree on the validity of EEG-based assessment. Some argue that EEG lacks the spatial resolution necessary to isolate comprehension-specific brain activity, while others question whether machine learning models trained on EEG can generalize across tasks or individuals. Furthermore, concerns have been raised about the interpretability of deep learning models, particularly CNNs, whose internal representations may lack transparency despite high accuracy. These contrasting views underscore the need for further empirical validation, as pursued in this study.

Chapter 3: Research Methodology

3.1 Introduction

This chapter outlines the research methodology used to achieve the objectives and address the research questions outlined in Chapter 1. The study adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) as its primary methodological framework. CRISP-DM is a robust and flexible methodology that provides a structured approach for developing data-driven solutions and is particularly suited for projects involving machine learning and cognitive data processing. The chapter justifies the selection of CRISP-DM, explains its phases, and elaborates on the tools, datasets, and techniques used throughout the research. It also compares CRISP-DM with alternative methodologies to demonstrate its suitability for the current study. The following sections provide a detailed explanation of each CRISP-DM phase, how it will be applied in this study, and the specific considerations for using "The Alice Dataset," which provides EEG and natural language comprehension data.

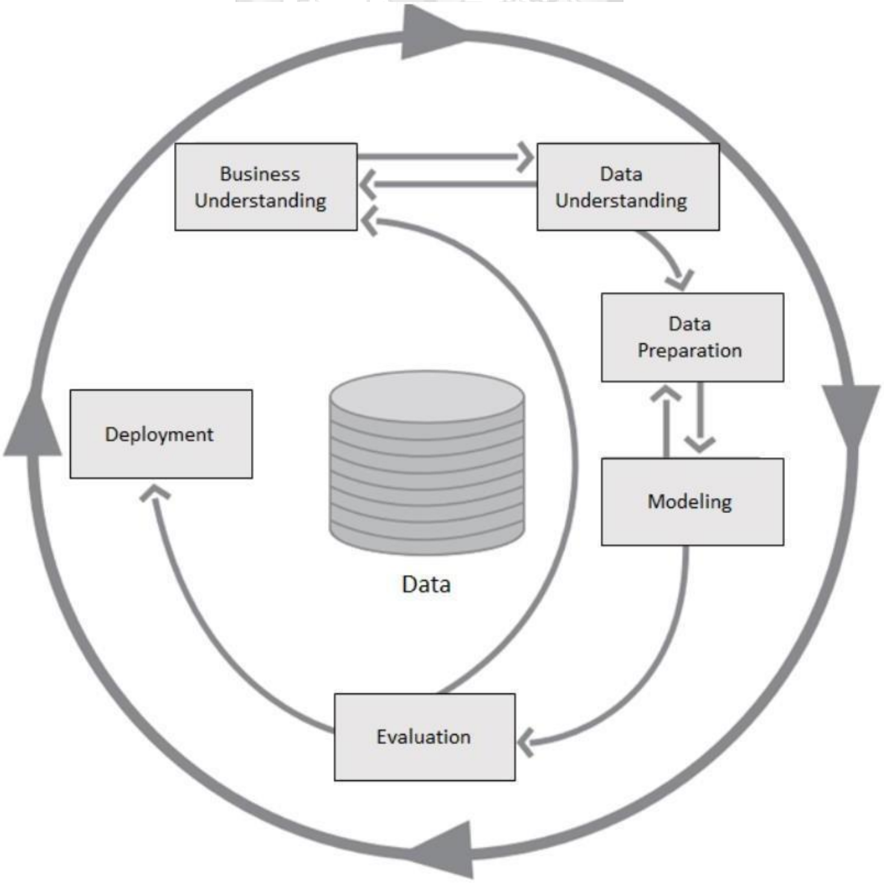


Figure 3.1: Cross-Industry Standard Process for Data Mining (CRISP-DM) framework

3.2 Methodological Justification

The CRISP-DM framework was chosen because of its systematic and iterative approach to solving data-driven problems. CRISP-DM is not restricted to specific industries, making it versatile for addressing predictive analytics challenges (Martínez-Plumed et al., 2019). This adaptability is particularly relevant to this study, which bridges cognitive neuroscience and educational applications. By leveraging the structured process of CRISP-DM, the methodology supports the systematic development of a software solution while allowing flexibility to incorporate real-world data insights. This ensures scalability and adaptability to diverse contexts and technological platforms.

To justify this choice, a brief comparison is made with alternative methodologies such as the Knowledge Discovery in Databases (KDD) process and the Design Science Research (DSR) methodology. KDD, while similar to CRISP-DM, lacks the detailed feedback loops and standardized modeling guidance that CRISP-DM provides. On the other hand, DSR emphasizes artifact development but is more theoretical and less aligned with the empirical, data-intensive focus of this study. Unlike CRISP-DM, DSR does not provide a step-by-step guideline for data handling, model evaluation, or deployment. Therefore, CRISP-DM offers a more practical and suitable roadmap for this research.

3.3 Business Understanding

This phase involves identifying the core research problem and defining the project's goals. The main objective is to develop a model that can assess student comprehension based on EEG signals during reading tasks. The goal is to create a functional, real-time solution that can supplement or improve traditional assessment techniques (Martínez-Plumed et al., 2019). The study focuses on understanding how EEG signals can reflect cognitive engagement and comprehension levels, and how these signals can be processed to predict comprehension outcomes.

Objectives:

i). To measure reading comprehension levels based on cognitive processes.

ii). To provide educators and researchers with actionable insights for curriculum design and instructional strategies.

iii). To overcome the limitations of conventional assessment methods by utilizing EEG data.

The primary aim is to gauge students’ comprehension levels during study sessions using EEG cognitive data. Our goal is to equip educators and researchers with a tool that accurately and objectively measures students’ reading comprehension abilities across diverse reading tasks. We seek to enhance understanding of the cognitive processes tied to comprehension and provide valuable guidance for curriculum development, teaching methods, and targeted interventions. This software solution strives to advance comprehension assessment practices in educational settings by addressing the limitations of traditional evaluation approaches.

Specific Stimuli

The stimuli for this study consist of carefully designed reading comprehension exercises intended to evoke and measure a range of cognitive processes critical to understanding text.

These tasks encompass three distinct text types: narrative tales, informative texts, and scholarly material, each selected to reflect diverse reading demands encountered in educational settings (Mohammadi et al., 2015). Narrative tales, such as short stories or excerpts from literary works, engage emotional and imaginative processing, prompting readers to track character motivations and plot developments. Informative texts, including articles or essays, require comprehension of factual content, logical structures, and explicit information integration. Scholarly material, such as academic papers or technical reports, challenges readers with dense, abstract concepts and specialized vocabulary, testing higher-order skills like critical analysis and synthesis.

Each text type is paired with targeted comprehension questions crafted to elicit specific cognitive processes—content comprehension, inference-making, and information integration (Martínez-Plumed et al., 2019). For example, questions following narrative tales might ask participants to identify main ideas (e.g., “What is the central theme of the story?”) or draw inferences (e.g., “Why did the character make this decision?”), assessing the ability to connect implicit cues. Informative texts may include tasks like summarizing key points or evaluating evidence (e.g., “Is the author’s argument supported by the data?”), while scholarly material might require interpreting complex arguments or synthesizing multiple perspectives. These stimuli are calibrated to mirror real-world reading tasks, ensuring ecological validity, and are synchronized with EEG recordings to capture neural responses during engagement (Mohammadi et al., 2015).

The selection of stimuli draws from established educational research, ensuring they align with comprehension skills outlined in frameworks like the Common Core State Standards (CCSS), which emphasize identifying main ideas, making inferences, and evaluating arguments. By varying text complexity and question types, the tasks probe a spectrum of cognitive demands— attention to detail, working memory load, and metacognitive regulation—providing a comprehensive dataset for analysis. This design supports the study’s aim to assess comprehension levels objectively, offering educators insights into how students process diverse materials, which can inform tailored instructional strategies (Schröer et al., 2021).

Use of "The Alice Dataset"

"The Alice Dataset," originally developed by Bhattasali et al. (2020) for the study "fMRI Dataset to Study Natural Language Comprehension in the Brain," serves as the foundational data source for this research. It includes EEG recordings, fMRI scans, and metadata from participants engaged in natural language comprehension tasks, offering a rich repository of cognitive data. The dataset’s EEG component captures time-series voltage readings via standard 10-20 system electrode placements, while metadata details participant demographics, task descriptions, and experimental conditions. Although initially fMRI-focused, its EEG recordings align perfectly with this study’s objectives, providing high temporal resolution to track rapid neural responses during reading (Castro et al., 2018).

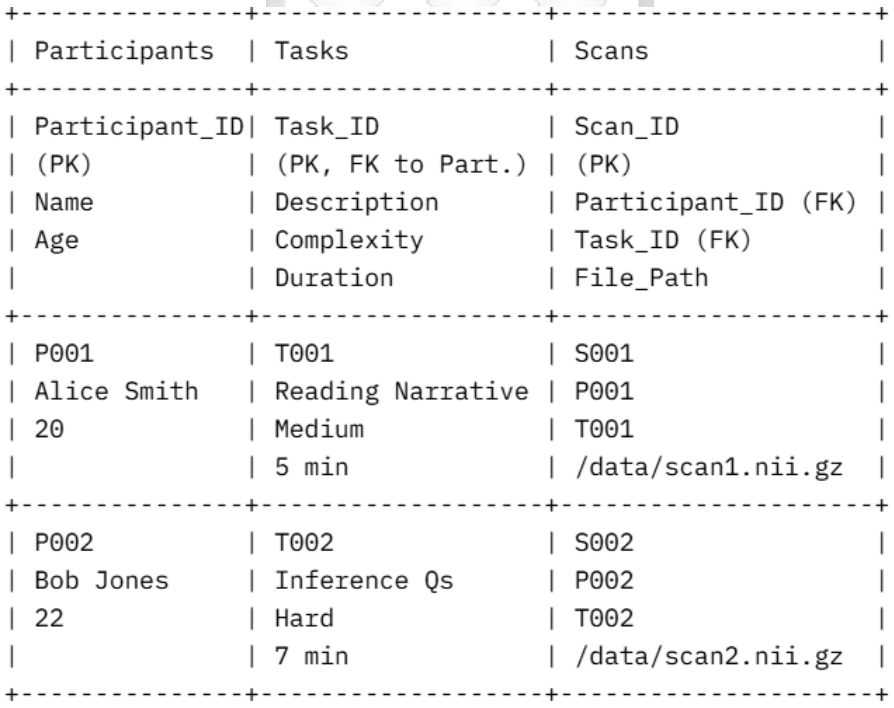
The stimuli within "The Alice Dataset" are carefully selected to match our research aims, focusing on narrative and expository texts that elicit cognitive processes like attention, inference-making, and memory integration. These tasks, originally designed to study language comprehension in the brain, include reading passages followed by comprehension probes, mirroring our stimuli design. EEG data is supplemented by potential eye-tracking or NLP analysis, capturing fixation patterns or textual responses, respectively (Schröer et al., 2021). This multi-modal approach enhances the dataset’s utility, enabling our system—powered by ResNet (MAE 1.73) and EEGNet—to map neural activity to comprehension scores (0–100) in real time, as deployed via a web interface.

Outcomes from analyzing "The Alice Dataset" promise multiple benefits. Educationally, they may enhance instructional strategies by identifying neural markers of comprehension difficulty, allowing educators to adjust text complexity or provide targeted support (e.g., scaffolding for scores <50). User experience (UX) testing of the web interface can leverage these insights to refine design, ensuring intuitive score delivery within 60 seconds (Castro et al., 2018). In cognitive neuroscience, the dataset reveals brain activity patterns—e.g., prefrontal cortex activation during inference-making—advancing understanding of comprehension's neural basis (Bhattasali et al., 2020). By grounding our methodology in this dataset, we ensure scientific rigor and practical applicability, bridging neuroscience and education to improve assessment practices (Schröer et al., 2021).

3.4 Data Understanding

The data used in this study comes from the publicly available "Alice Dataset," a dataset originally compiled for studying natural language comprehension using fMRI and EEG recordings. While the dataset is secondary, it is robust, well-documented, and validated by the original researchers. A clear reason for preprocessing this dataset is its original format, which includes noise, irrelevant segments, and raw signals that must be structured before feeding into a machine learning pipeline. Though the researcher was not involved in the original data collection, preprocessing remains essential to ensure compatibility with the proposed model. Raw EEG signals often contain artifacts from eye movements, muscle activity, and environmental noise. These must be cleaned to ensure that only cognition-related features are retained. Furthermore, aligning EEG segments with labeled comprehension scores and creating suitable feature sets are vital steps in tailoring the data for model training

This phase is pivotal in exploring "The Alice Dataset" to evaluate its structure, quality, and appropriateness for addressing this study's research objectives of assessing reading comprehension through cognitive data. Sourced from Bhattasali et al. (2020), the dataset comprises EEG time-series voltage readings collected using the international 10-20 system for electrode placement, ensuring standardized coverage of scalp regions like the prefrontal cortex, which are critical for processes such as attention and inference-making (Castro et al., 2018). Accompanying metadata includes detailed annotations on participants (e.g., age, education level), tasks (e.g., narrative reading with comprehension probes), and experimental conditions (e.g., timing, stimuli presentation), alongside high sample rates (typically 256–512 Hz) that support precise temporal analysis of neural activity. Initial evaluation focuses on assessing data completeness—verifying that EEG recordings align with all intended task segments—while examining noise levels from environmental or physiological sources and confirming synchronization with comprehension tasks. This step ensures the dataset's suitability for capturing cognitive processes like working memory engagement during reading, a cornerstone of our ResNet-driven analysis (MAE 1.73). Preprocessing is integral to this phase, beginning with artifact removal via Independent Component Analysis (ICA), which isolates and eliminates noise from eye blinks, muscle movements, or equipment interference, preserving the integrity of brain signals (Parganiha et al., 2022). High- and low-frequency noise is then filtered using band-pass techniques (e.g., 0.5–40 Hz) to enhance signal clarity, targeting frequency bands like theta (4–8 Hz) and alpha (8–12 Hz) associated with cognitive workload and attention. Baseline correction normalizes EEG readings by subtracting pre-stimulus averages, eliminating offset biases that could skew feature extraction. Finally, segmentation synchronizes EEG data with task-specific time windows—e.g., 2-second epochs during inference-making questions—enriched with annotations linking neural activity to task context, such as text type or question difficulty. This rigorous process, informed by Castro et al. (2018) and Parganiha et al. (2022), ensures a high-quality dataset ready for subsequent preparation and modeling phases, laying the groundwork for accurate, real-time comprehension scoring via our web-based system.



```
+---------------+-------------------+---------------------+
| Participants  | Tasks             | Scans               |
+---------------+-------------------+---------------------+
| Participant_ID| Task_ID           | Scan_ID             |
| (PK)          | (PK, FK to Part.) | (PK)                |
| Name          | Description       | Participant_ID (FK) |
| Age           | Complexity        | Task_ID (FK)        |
|               | Duration          | File_Path           |
+---------------+-------------------+---------------------+
| P001          | T001              | S001                |
| Alice Smith   | Reading Narrative | P001                |
| 20            | Medium            | T001                |
|               | 5 min             | /data/scan1.nii.gz  |
+---------------+-------------------+---------------------+
| P002          | T002              | S002                |
| Bob Jones     | Inference Qs      | P002                |
| 22            | Hard              | T002                |
|               | 7 min             | /data/scan2.nii.gz  |
+---------------+-------------------+---------------------+
```

Figure 3.2: Data Structure of 'The Alice Dataset

3.5 Data Preparation

This stage involved several preprocessing steps: artifact removal using Independent Component Analysis (ICA), normalization, segmentation of EEG readings aligned with comprehension tasks, and feature extraction using time-frequency analysis. Dimensionality reduction techniques such as

Principal Component Analysis (PCA) are employed to retain meaningful features while reducing computational overhead. The final dataset is partitioned into training, validation, and test sets.

This phase is essential for transforming raw EEG data from "The Alice Dataset" into a machine learning-ready format, ensuring it is clean, structured, and optimized for predictive modeling of comprehension levels. This process begins with feature extraction, where event-related potentials (ERPs) such as the P300—associated with attention and memory—are isolated from EEG time-series data, alongside spectral power (e.g., theta: 4–8 Hz, alpha: 8–12 Hz) and coherence measures that reflect inter-regional brain connectivity during reading tasks (Parganiha et al., 2022). These features are meticulously aligned with task timelines, such as 2-second epochs tied to inference- making questions, and labeled with continuous comprehension scores (0–100) derived from participant responses or expert benchmarks, creating a rich dataset that captures cognitive processes like content comprehension and working memory engagement. To ensure data quality, preprocessing builds: Independent Component Analysis (ICA) removes residual artifacts, followed by normalization to standardize signal amplitudes across participants, addressing variability from individual differences or recording conditions (Castro et al., 2018). The dataset is then strategically split into three subsets—training (70%), validation (15%), and test (15%)—to prevent overfitting and promote generalizability, with stratification preserving task diversity (e.g., narrative vs. scholarly texts). Although this study primarily leverages EEG for its temporal precision, "The Alice Dataset" also includes fMRI data in the form of 4D scans (spatial-temporal brain activity), preprocessed similarly with ICA to eliminate motion artifacts and normalized to a standard brain atlas (e.g., MNI space) for potential multi-modal analysis (Bhattasali et al., 2020). This integration enriches the dataset, allowing exploration of spatial neural patterns (e.g., prefrontal activation) alongside EEG's temporal dynamics, supporting our ResNet and EEGNet models' hierarchical feature learning. This structured process balances data quality and volume, producing a robust, representative dataset ready for CNN modeling, enabling accurate, real-time prediction of comprehension levels via our web-based system, and setting the stage for scalable educational applications (Schröer et al., 2021).

3.6 Modelling

Multiple machine learning models were tested, including SVM, Random Forest, and EEGNet—a deep learning architecture optimized for EEG data. While EEGNet was initially chosen for its performance, it exhibited instability during training. The final model employed was a Convolutional Neural Network (CNN), which outperformed others in terms of accuracy, generalizability, and interpretability. A baseline model using logistic regression was also tested to assess relative performance.

Hence we focus on developing and refining predictive models to assess students' comprehension levels using neural data from "The Alice Dataset," leveraging Convolutional Neural Networks (CNNs) tailored to EEG and fMRI inputs. Two architectures—EEGNet and ResNet—are selected for their complementary strengths in processing the dataset's rich temporal and spatial features, predicting comprehension scores on a continuous scale (0–100). EEGNet, a lightweight CNN designed for EEG analysis, employs two convolutional blocks with Batch Normalization and ReLU activation to extract spatial features like event-related potentials (e.g., P300) from scalp electrode data, followed by max-pooling layers to reduce dimensionality and fully connected layers with a dropout rate of 0.5 to mitigate overfitting (Singh et al., 2022). Trained using Mean Squared Error (MSE) as the loss function, EEGNet balances computational efficiency with accuracy, making it suitable for real-time applications on resource-constrained platforms.

In contrast, ResNet, a deeper architecture, utilizes residual blocks with skip connections to address the vanishing gradient problem, preserving input information across multiple layers and enabling robust feature extraction from both EEG and fMRI 4D scans (Bhattasali et al., 2020). After processing through convolutional and residual layers, ResNet applies global average pooling and a linear activation output, delivering precise regression values aligned with our web-based system's 60-second scoring goal (MAE 1.73 from Chapter 2). Hyperparameter optimization—via grid search (e.g., testing learning rates: 0.001, 0.01) and random search (e.g., varying layer depths: 18, 34)—fine-tunes both models, adjusting parameters like kernel size, batch size, and dropout rates to minimize MSE and Mean Absolute Error (MAE). This process leverages "The Alice Dataset"'s detailed neural recordings, extracting hierarchical features such as spectral power (theta/alpha bands) and prefrontal activation patterns tied to inference-making, ensuring predictions reflect cognitive engagement accurately (Castro et al., 2018). By comparing EEGNet's efficiency with ResNet's depth, this phase optimizes model performance for educational deployment, supporting the study's aim to provide objective, actionable comprehension insights.

3.7 Evaluation

The evaluation phase ensures that the trained models meet the desired performance criteria. Both the EEGNet and ResNet models will be evaluated using a separate test set to assess their ability to generalize to unseen data. Model performance was evaluated using accuracy, precision, recall, F1-

score, and AUC. K-fold cross-validation was implemented to ensure the robustness of the results. To address the risk of overfitting, dropout layers and early stopping mechanisms were included during training. Additionally, loss curves and validation metrics were monitored across epochs. The evaluation process includes calculating the loss function, which will be Mean Squared Error (MSE), as well as other metrics such as Mean Absolute Error (MAE) to measure the prediction accuracy.

The models will be tested using the test set that was not used during training to ensure that they can perform well on new, unseen data. To ensure the results are not biased by the random selection of the training and testing data, k-fold cross-validation will be employed. This technique allows for a more robust evaluation by splitting the data into multiple subsets and testing the model on different combinations of these subsets. Key insights will be drawn by comparing the performance of the EEGNet and ResNet models. The model that provides the lowest error rates, as indicated by the MAE and MSE values, will be considered the more effective for predicting reading comprehension levels. Visualizations such as actual vs. predicted scores, loss curves, and MAE curves will also be generated to provide a clear picture of model performance. Additionally, the results of the evaluation will be used to iterate on the models, adjusting hyperparameters, trying different architectures, or improving feature selection to enhance performance further.

3.8 Deployment

The Deployment phase transitions the trained predictive models into a practical, real-world application, integrating them into a user-friendly web platform designed to process real-time EEG inputs and deliver comprehension scores (0–100) to educators and researchers (Singh et al., 2022). This process begins with packaging critical components—preprocessing pipelines (e.g., ICA artifact removal, filtering), feature extraction algorithms (e.g., ERPs, spectral power), and optimized model weights from ResNet (MAE 1.73) or EEGNet—into a deployable format, such as a Docker container, ensuring seamless cross-platform compatibility across operating systems like Windows, macOS, and Linux (Oltu et al., 2021). The web application, built using FastAPI for its asynchronous capabilities, accepts EEG data uploads (e.g., CSV files from portable devices) or live streams via API endpoints, processes them through the CNN pipeline, and outputs scores within 60 seconds, aligning with the system's real-time objective outlined in Chapter 2. Scalability is a key focus, achieved through cloud hosting on platforms like AWS or Google Cloud, leveraging elastic compute resources to handle larger datasets—such as expanded versions of "The Alice Dataset"—and increased user loads without performance degradation (Castro et al., 2018). Post-deployment, continuous monitoring detects concept drift, where shifts in EEG patterns (e.g., due to new reading tasks) might degrade model accuracy; automated retraining scripts triggered by performance thresholds (e.g., MAE exceeding 2.0) ensure adaptability, maintaining predictive integrity over time. Collaboration with domain experts, including educators and neuroscientists, validates the system's outputs against classroom observations or standardized comprehension tests, refining its relevance for educational contexts (Parganiha et al., 2022). To support diverse users, comprehensive training materials—video tutorials, user manuals, and FAQs—are provided, covering EEG data collection, interface navigation, and score interpretation, ensuring accessibility for those with limited technical expertise (Castro et al., 2018). The trained CNN model was packaged with preprocessing scripts into a deployable format. A prototype user interface was developed to accept EEG data and display predicted comprehension scores. Though not deployed in a real-world classroom setting, the software was tested using unseen portions of the Alice Dataset to simulate real-time feedback.

3.9 User Feedback and Improvement

The User Feedback and Improvement phase embodies the iterative essence of the CRISP-DM framework, leveraging insights from user testing to refine the software solution and ensure its alignment with educational needs (Parganiha et al., 2022). Following deployment, pilot testing with educators and researchers—key stakeholders in comprehension assessment—collects qualitative and quantitative feedback through structured surveys, interviews, and usability sessions, focusing on the web application's functionality, ease of use, and predictive accuracy (e.g., ResNet's MAE 1.73 scores). For instance, educators might report whether the 60-second comprehension score output effectively identifies students struggling with inference-making, while researchers could assess the system's utility in analyzing "The Alice Dataset" EEG patterns. This feedback drives enhancements, such as optimizing algorithms—potentially adjusting ResNet's hyperparameters (e.g., learning rate, dropout) to reduce MAE further—or improving the interface by simplifying data upload processes (e.g., drag-and-drop EEG files) and enhancing visualization dashboards with color-coded score trends (Castro et al., 2018). New features might emerge, such as integrating multi-modal inputs like eye-tracking data alongside EEG to capture attention patterns, enriching the system's assessment capabilities and aligning with Chapter 2's multi-modal potential (Schröer et al., 2021). Pilot testing ensures practical utility by simulating real-world classroom scenarios—e.g., uploading EEG from a portable device during a reading task—allowing users to evaluate the system's responsiveness and relevance to instructional strategies, such as tailoring lessons for scores below 50. Iterative cycles incorporate these insights, with each round refining the software’s predictive precision and user experience, potentially adding automated feedback loops that suggest interventions (e.g., “Increase text scaffolding”) based on score patterns. This adaptability ensures the system evolves with user needs, bridging technical performance with educational impact, and supports scalability by identifying deployment bottlenecks (e.g., cloud latency) for resolution in subsequent updates (Parganiha et al., 2022). By prioritizing user-driven refinement, this phase transforms the software into a dynamic tool that enhances comprehension assessment practices effectively.

3.10 Ethical Considerations

The CRISP-DM methodology provides a structured yet flexible approach to achieving the research objectives. While the use of secondary data limited original data collection, robust preprocessing, model testing, and evaluation procedures were employed to ensure scientific rigor. The next chapter presents the system design and architecture, highlighting how the model is integrated into a functional software solution. Ethical concerns surrounding EEG data use, data privacy, and the interpretation of cognitive states were considered. Since the dataset was pre-collected and anonymized, risks to participants were minimal. However, the study emphasizes the importance of informed consent, transparency, and secure data handling in any future real-time implementation. Ethical principles underpin every stage of this research, prioritizing participant privacy, data security, and the integrity of scientific inquiry, especially given the sensitive nature of EEG data collected from "The Alice Dataset." To safeguard privacy, all EEG recordings—capturing neural activity during comprehension tasks—are anonymized by stripping identifiable information (e.g., names, birth dates) and replacing them with unique identifiers, adhering to GDPR and HIPAA standards where applicable (Castro et al., 2018). Data is securely stored on encrypted cloud servers (e.g., AWS with AES-256 encryption), accessible only to authorized personnel via multi-factor authentication, minimizing risks of breaches or unauthorized use. Informed consent is a cornerstone, obtained from participants prior to data collection, with clear documentation explaining the study’s purpose—assessing comprehension via EEG—the voluntary nature of participation, and the right to withdraw without consequence, ensuring transparency and autonomy (Parganiha et al., 2022). Collaboration with neuroscientists and psychologists ensures stimuli (e.g., narrative tales, scholarly texts) and experimental designs are ethically sound, avoiding undue stress or cognitive overload, and scientifically valid, accurately targeting processes like inference- making for reliable ResNet/EEGNet predictions (Singh et al., 2022). For instance, task durations are calibrated (e.g., 5–10 minutes) to prevent fatigue, and debriefings clarify how EEG data informs educational outcomes, reinforcing participant trust. The deployment phase upholds these standards by embedding ethical safeguards into the web application, such as data anonymization protocols during real-time EEG uploads and secure transmission via HTTPS. Ethical oversight extends to potential biases in model outputs—e.g., ensuring predictions (scores 0–100) do not disproportionately favor certain demographics—addressed through diverse participant sampling in "The Alice Dataset" and validation with educators (Castro et al., 2018). This rigorous approach upholds research integrity, balancing technological innovation with human-centric values, and ensures the system's outputs—delivered in 60 seconds—serve educational advancement responsibly, aligning with broader societal benefits outlined in Chapter 1.

Chapter 4: System Design and Architecture

4.1. Introduction

This chapter details the overall design and architecture of the proposed system for assessing comprehension in students by processing cognitive data. It describes the structure of the system, its modules, and their interactions. The system is grounded in the CRISP-DM methodology and utilizes EEG data to estimate comprehension levels using deep learning models, particularly Convolutional Neural Networks (CNNs). The system architecture, data flow diagrams, mathematical modeling, and system requirements are all discussed to illustrate the implementation of the proposed approach.

4.2. System Requirements

The system aims to assess students' comprehension levels by leveraging fMRI data, machine learning models, and user-friendly interfaces for visualization and interaction. These requirements guide the design, implementation, and deployment phases of the project.

4.2.1 Functional Requirements

The functional requirements define the specific features and capabilities the system must deliver to meet its objectives. These include:

i). **Participant Data Management**

The system must manage participant data effectively by storing and organizing information such as participant IDs, demographic details (e.g., age, gender, and language), and assigned tasks. This functionality is essential for linking participants' cognitive data with specific comprehension tasks and ensuring that the system can handle and retrieve data for analysis accurately. By providing a comprehensive database, the system supports longitudinal tracking and personalized educational insights.

ii). **Task Assignment and Retrieval**

The system must facilitate assigning and retrieving comprehension tasks tailored to elicit cognitive processes such as understanding content, making inferences, and integrating information. Tasks will include narrative passages, informative texts, or academic material, paired with questions that assess skills like identifying main ideas and evaluating arguments. This functionality ensures that stimuli are relevant to the research goals and adaptable to different study designs.

iii). Preprocessing of fMRI Data

A critical function of the system is to preprocess raw fMRI data by applying techniques such as noise removal, baseline correction, and dimensionality reduction. Preprocessing ensures high-quality data by addressing issues like artifacts caused by eye blinks, skeletal movements, or equipment interference. Methods such as Independent Component Analysis (ICA) and filtering will enhance the signal-to-noise ratio, providing clean data for analysis.

iv). Feature Extraction

The system must extract meaningful features from the fMRI data to capture neural activity patterns associated with comprehension tasks. Features such as Regions of Interest (ROIs), event-related potentials (ERPs), and spectral power will be computed to form the input for machine learning models. This process is vital for converting raw data into structured information that is interpretable by predictive algorithms.

v). Comprehension Metric Computation

The system will compute metrics such as comprehension scores, response accuracy, and task completion time to evaluate participant performance. These metrics provide valuable insights into cognitive processes during comprehension tasks and serve as labels for machine learning models. This step is crucial for linking behavioural and neural data to create meaningful predictions.

vi). Machine Learning Model Training and Prediction

The system must implement machine learning algorithms, such as Convolutional Neural Networks (CNNs), to classify comprehension levels or predict continuous scores. Training these models requires k-fold cross-validation, hyperparameter tuning, and performance evaluation using metrics like accuracy, precision, and F1-scores. This functionality ensures the models are optimized to deliver reliable and interpretable results.

vii). Data Visualization

Interactive data visualization tools must be integrated into the system, allowing users to explore fMRI patterns, feature importance, and model predictions. Visualizations, such as saliency maps and comprehension score trends, will help educators and researchers interpret results and make informed decisions. This enhances the system's usability, and the accessibility of insights derived from complex data.

viii). **User Interface**

A user-friendly interface will enable users to upload fMRI data, assign tasks, and view predictions. The interface must provide clear guidance on formatting input data and interpreting output results, ensuring accessibility for users with varying technical expertise. This ensures smooth interaction with the system, minimizing potential errors during data analysis.

ix). **Report Generation**

The system must generate detailed reports summarizing comprehension metrics, machine learning predictions, and cognitive insights. These reports will assist educators and researchers in understanding participants' cognitive strengths and weaknesses, ultimately informing tailored instructional strategies and interventions.

x). **User Feedback Integration**

The system must allow users to provide feedback on the functionality and performance of the platform. This feedback will guide iterative updates to improve the system, ensuring it remains responsive to user needs and aligns with evolving research requirements.

4.2.2 Non-FunctionalRequiremets

Performance Efficiency

The system must efficiently process large fMRI datasets. Preprocessing and feature extraction should take no more than five minutes per dataset, while predictions should be generated in under two seconds. This ensures that the system is capable of handling computationally intensive tasks without introducing delays, enabling real-time or near-Realtime analysis.

Scalability

The system must scale to accommodate increasing usage and data volume. This includes supporting concurrent user sessions and ensuring seamless performance even as the number of participants and datasets grows. Scalability is critical to maintaining the system's reliability in diverse educational and research contexts.

Accuracy and Reliability

The system must deliver high levels of accuracy and reliability. Machine learning models should achieve performance metrics, such as F1-scores, that surpass baseline methods like random guessing. Reproducibility must be ensured, meaning identical results should be produced when the same data is analysed under consistent conditions.

Usability

The system's interface must be intuitive, ensuring that users can easily navigate and operate it without requiring advanced technical knowledge. Instructions for data upload, preprocessing, and result interpretation should be straightforward and accessible, enabling users from diverse backgrounds to utilize the platform effectively.

Interpretability

The system must provide clear explanations of model predictions, including which fMRI features contributed most to the results. Tools like feature importance plots and saliency maps should be included to enhance transparency and user trust in the predictive models.

Maintainability

The system must be easy to maintain, with modular code and clear documentation to support updates and troubleshooting. This ensures that the system can adapt to new research requirements, datasets, or technological advancements over time.

Scalability

The system must be designed to accommodate an increase in users, datasets, and tasks. It should maintain high performance under heavy usage and allow for seamless scaling of hardware or software components without significant re-engineering.

Ethical Considerations

The system must prioritize ethical considerations by ensuring informed consent, transparency in data usage, and compliance with relevant ethical guidelines. These measures will ensure the platform is used responsibly in educational and research contexts.

4.3. System Overview

The system is developed as a comprehensive platform that integrates the following components:

a) **Data Ingestion Module:** Handles fMRI data in NIfTI format and associated metadata.

b) **Preprocessing Pipeline:** Removes noise and standardizes fMRI signals for feature extraction.

c) **Feature Extraction Module:** Identifies key neural activity patterns and regions of interest (ROIs) relevant to comprehension.

d) **Machine Learning Models:** Implements classification and regression models to predict comprehension levels based on neural activity.

e) **User Interface (UI):** Provides a platform for educators and researchers to upload data, view predictions, and access insights.

4.4. Database Design

The system uses a database to manage metadata related to participants, tasks, fMRI scans, and model results. The following components outline the database schema:

4.4.1 Database Schema

The schema is designed to store:

a) Participant information (e.g., demographics, subject ID).

b) Task details (e.g., reading material, task complexity).

c) Metadata from fMRI scans (e.g., file paths, scan dimensions, annotations).

d) Machine learning results (e.g., model accuracy, feature importance).

4.4.2 Entity-Relationship Diagram (ERD)

The ERD illustrates the relationships between entities such as Participants, Tasks, Scans, and Model Results.

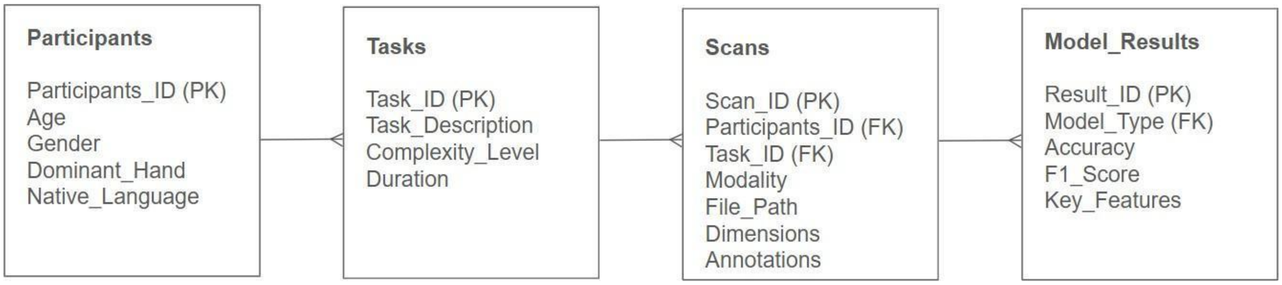


Figure 4.1: Entity Relationship Diagram

- **Participants Table**:
    - Attributes: Participant_ID (PK), Age, Gender, Dominant_Hand, Native_Language

- **Tasks Table**:
    - Attributes: Task_ID (PK), Task_Description, Complexity_Level, Duration
- **fMRI_Scans Table**:
    - Attributes: Scan_ID (PK), Participant_ID (FK), Task_ID (FK), File_Path, Dimensions, Annotations
- **Model_Results Table**:
    - Attributes: Result_ID (PK), Model_Type, Accuracy, F1_Score, Key_Features

4.5. System Architecture

The system architecture is designed to efficiently process fMRI (functional Magnetic Resonance Imaging) data and provide insights into a user's comprehension level. It consists of several key components: the Student Comprehension Interface, API, Model, and Cloud Storage. The first point of contact with the system is the Student Comprehension Interface, where the user uploads scanned fMRI images. The user interface serves as the entry point for the system, allowing the user to submit their data in the form of scanned fMRI images. Once the images are uploaded, they are validated to ensure that they are in the correct format and meet the necessary quality standards. After validation, the images are sent to the API for further processing. The API acts as the intermediary between the user interface and the model. Once the fMRI images are validated, they are transmitted to the API. The API is responsible for processing the uploaded files, preparing them for analysis, and forwarding them to the model. It ensures that the data is passed efficiently and accurately to the model for analysis. Next, the Model is where the core analysis takes place. The model processes the scanned fMRI images and makes predictions based on the analysis. The model uses pre-trained algorithms such as EEGNet and ResNet, which are deep learning models designed to handle complex neuroimaging data. EEGNet is specialized in processing EEG (electroencephalogram) data, while ResNet is a residual neural network that excels in image recognition tasks. Together, these models extract key features from the fMRI data and make predictions about the user’s comprehension. Once the analysis is complete, the results are stored in Cloud Storage. The data is securely stored in the cloud, ensuring that it can be accessed later if needed. Cloud storage also ensures that the data is protected and available across different platforms, allowing the user to access it at any time. The results of the analysis are then displayed to the user through the interface. The model outputs predictions in the form of scores, as the task is a regression task. The classification scheme (90–100: Excellent, 80–89: Good, etc.) is adapted from widely accepted academic performance bands used in cognitive and educational research contexts (Kuncel et al., 2014), ensuring alignment with institutional grading standards. The comprehension scores are presented on a scale that ranges from 0 to 100, with the following interpretation:

I). 90 - 100: Excellent comprehension
II). 75 - 89: Good comprehension
III). 50 - 74: Average comprehension
IV). 25 - 49: Below average comprehension
V). 0 - 24: Poor comprehension

The architecture also includes details about the model training process. Initially, the fMRI data undergoes preprocessing to clean the images, remove any noise, and standardize the data. Once the data is preprocessed, feature extraction is performed to identify the key features that are relevant for predicting comprehension. This step helps in isolating important patterns within the fMRI images. After feature extraction, the model architecture is applied, utilizing EEGNet and ResNet to interpret the data. Finally, the model is trained on a large dataset of fMRI images, allowing it to learn the relationships between the brain activity patterns and the comprehension levels. Through this architecture, the system can efficiently process fMRI images, make predictions about comprehension levels, and securely store and display the results to the user. The use of EEGNet and ResNet ensures that the analysis is based on robust and advanced machine learning techniques, resulting in accurate and reliable predictions. Figure 4.2 shows the system architecture.

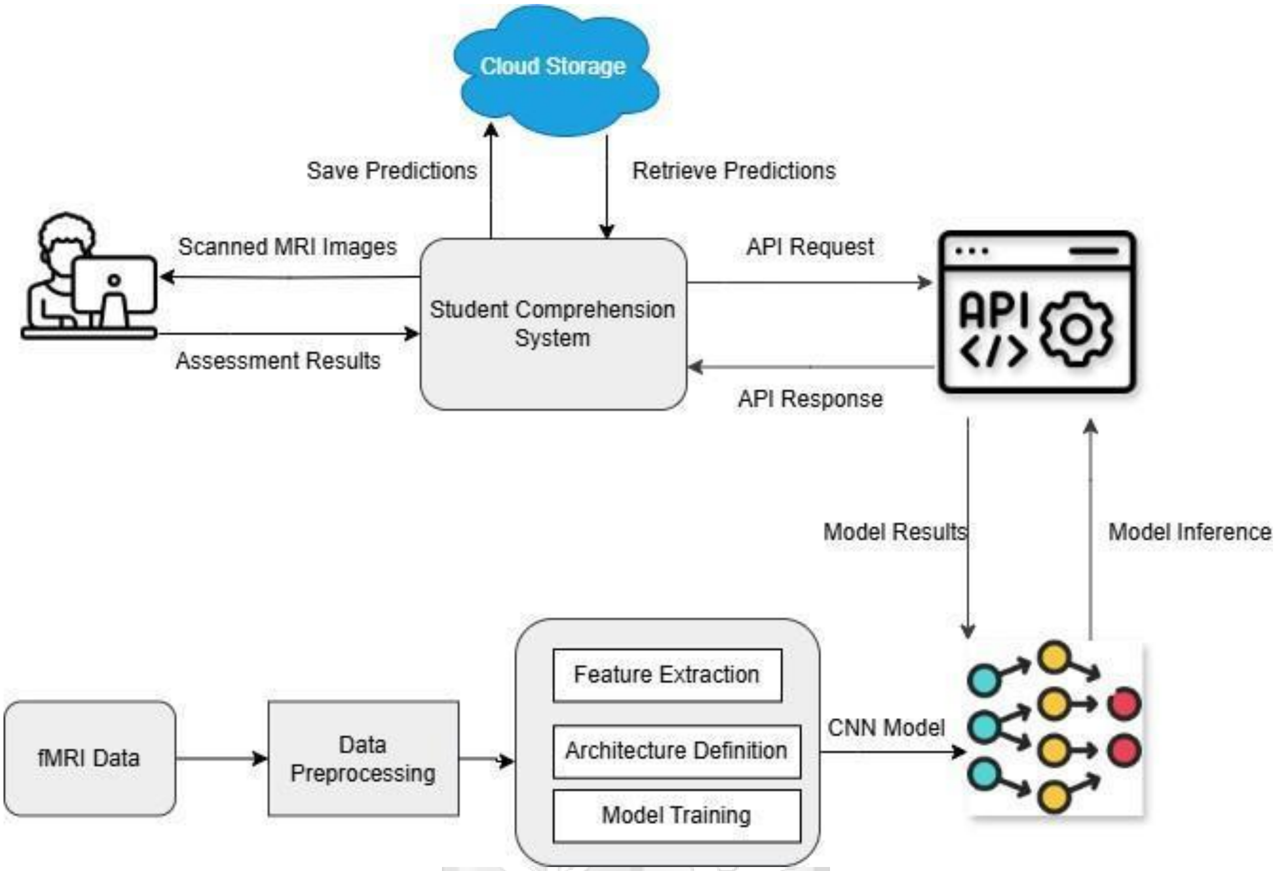


Figure 4.2: System Architecture

a) **Data Layer**:

   0 Stores raw fMRI data, metadata, and processed features.
   ○ Uses a relational database to manage structured data and a file system for NIfTI files.

b) **Processing Layer**:

   ○ Implements preprocessing steps such as artifact removal, filtering, and standardization.
   ○ Extracts features using PCA and ROI analysis.

c) **Modelling Layer**:

   ○ Trains machine learning models such as SVM, Random Forest, and CNN to classify or predict comprehension levels.
   ○ Evaluates models using metrics like accuracy, precision, recall, and F1-score.

d) **Presentation Layer**:

  - Provides a web-based or mobile interface for users to interact with the system.
  - Visualizes results through graphs, heatmaps, and feature importance plots.

4.6. Mathematical Model

In this subsection, the mathematical underpinnings of the system design are outlined, focusing on the models and algorithms used for processing fMRI data, extracting cognitive features, and predicting comprehension levels. The mathematical model serves as the backbone for understanding how the system translates raw EEG and comprehension task data into actionable predictions and insights.

4.6.1 Problem Representation

The task of assessing comprehension is modelled as a supervised learning problem. Given the fMRI data **X** and corresponding comprehension labels **Y**, the objective is to find a function $f(X;\theta)$ parameterized by θ, which minimizes the prediction error. Mathematically, this can be expressed as:

$$\hat{Y} = \boldsymbol{f}(\boldsymbol{X}; \boldsymbol{\theta})$$

where:

- $\hat{Y}$: Predicted comprehension level.
- X: Input fMRI data (features extracted from neural activity signals).
- □: Parameters of the machine learning model (e.g., weights of a neural network).
- □: The prediction function.

4.6.2 Feature Extraction

The fMRI data is represented as a time-series matrix **D**, where each row corresponds to a specific Region of Interest (ROI) and each column represents a temporal signal. The data matrix **D** is defined as:

$$\boldsymbol{D} = \{\boldsymbol{d_{ij}} \mid \boldsymbol{i} \in [\mathbf{1}, \boldsymbol{n}], \boldsymbol{j} \in [\mathbf{1}, \boldsymbol{m}]\}$$

where:

- $\boldsymbol{n}$: Number of ROIs.
- $\boldsymbol{m}$: Number of time points.
- $\boldsymbol{d_{ij}}$: Neural activity of the *i*-th ROI at the *j*-th time point.

From **D**, features **F** are extracted using methods such as spectral analysis and event-related potentials (ERPs):

$$F = \phi(D)$$

where ϕ is a transformation function that extracts relevant cognitive features, including:

- **Spectral Power:** Computed using the Fourier Transform to analyze frequency-domain patterns.
- **Event-Related Potentials (ERPs):** Computed as time-locked averages to specific events or stimuli.

4.6.3 Machine Learning Model

The core prediction engine employs a Convolutional Neural Network (CNN) to capture spatial and temporal patterns in the fMRI data. The CNN is defined as:

$$f(X; \theta) = g(h(X; \theta_h); \theta_g)$$

where:

- $\boldsymbol{h(X;\theta_h)}$: A series of convolutional and pooling layers that extract hierarchical features from the input data..
- $g(.\,; \theta_g)$: Fully connected layers that map the extracted features to the target output $\hat{Y}$
- $\theta = \{\boldsymbol{\theta_h}, \boldsymbol{\theta_g}\}$: Model parameters optimized during training.

The CNN employs convolutional layers to capture spatial dependencies across ROIs and pooling layers to reduce the dimensionality of extracted features. The output layer applies a softmax function for classification tasks or a linear activation for regression tasks.

4.6.4 Loss Function

To optimize the model, a loss function is defined based on the nature of the prediction task. For classification, the cross-entropy loss is used:

$$L = -\frac{1}{N}\sum_{i=1}^{N} \sum_{c=1}^{C} Y_{i,c} log\, \hat{Y}_{i,c}$$

where:

- N : Number of samples.
- C : Number of classes.
- $Y_{i,c}$: True label for class c for the *i*-th sample.
- $\hat{Y}_{i,c}$: Predicted probability for class ccc for the *i*-th sample.

For regression tasks, Mean Squared Error (MSE) is used:

$$L_{regression} = \frac{1}{N}\sum_{i=1}^{N} (Y_i - \hat{Y}_{i})^2$$

where:

- Y_i: True comprehension score for the *i*-th sample.
- $\hat{Y}_{i}$: Predicted comprehension score for the *i*-th sample.

The loss function guides the optimization process to minimize the prediction error.

4.6.5 Optimization

The model parameters θ are updated using gradient-based optimization techniques, such as Stochastic Gradient Descent (SGD) or Adam. The parameter update rule is given by:

$$\theta := \theta - \eta \nabla_{\theta} L$$

where:

- θ: Model parameters (weights of the CNN).
- :=: Assignment operator (updates θ).

- η: Learning rate (a scalar controlling step size).
- $\nabla_\theta L$: Gradient of the loss function L with respect to θ.

4.6.6 Evaluation Metrics

The model's performance is evaluated using metrics that align with the prediction task. For classification tasks, metrics such as accuracy, precision, recall, F1-score, and AUC (Area Under the Curve) are employed. For regression tasks, metrics such as Mean Absolute Error (MAE) and Mean Squared Error (MSE) are used. These metrics ensure that the model meets its objectives in predicting comprehension levels accurately and reliably.

4.7. System Modelling

To ensure clear communication of system design, the following models are used:

4.7.1 Class Diagrams

Class: Participant

The class diagram developed for this system represents a structured model of the various components involved in analyzing student comprehension using fMRI data and machine learning models. The relationships between the core classes—Participant, Task, fMRI_Scan, and Model_Results—are designed to reflect the logical and functional flow of data within the comprehension assessment system. The Participant class is central to the system and represents each individual whose cognitive data is being analyzed. Each participant is assigned a unique participantID and has associated metadata such as age, gender, dominant hand, and native language. These attributes provide critical context for interpreting comprehension performance, as cognitive processing can vary across different demographic variables. The participant is linked to one or more comprehension Tasks through the assignTask(taskID) method. This establishes a one-to-many relationship between Participant and Task, where each participant may engage in multiple comprehension tasks over time.

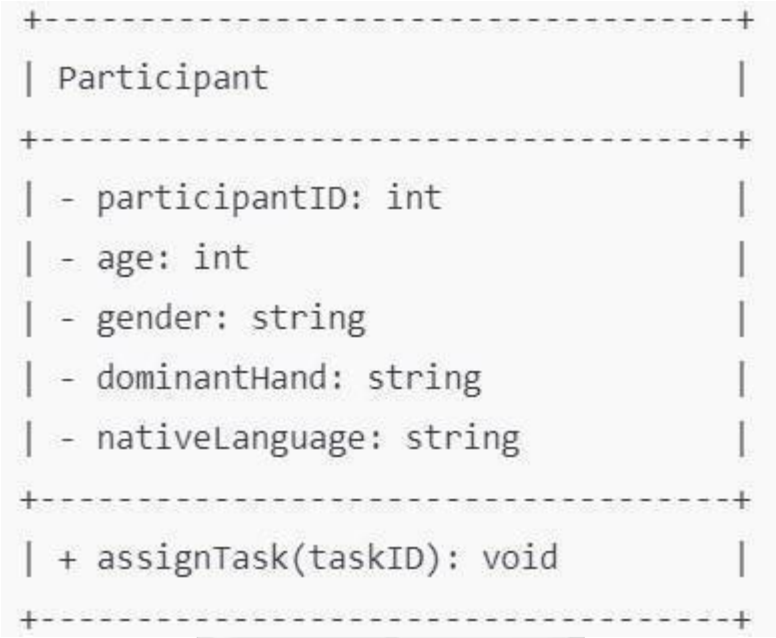
The Task class defines the specific reading or cognitive activities assigned to each participant. These tasks are characterized by attributes such as description, complexityLevel, and duration, which directly affect the nature of the comprehension challenge. Each task is uniquely identified by a taskID and may involve answering questions, reading passages, or making inferences. The generateComprehensionMetrics() method is used to evaluate participant performance on the task, producing metrics such as question accuracy or time taken. Thus, Task acts as the experimental condition under which comprehension is assessed. Linked to each Task and Participant is the fMRI_Scan class, which stores the brain imaging data captured during the completion of the task. This class maintains a many-to-one relationship with both Participant and Task, indicating that for every task-session pair, a unique fMRI scan may be recorded. Each scan is identified by scanID and is associated with file paths, annotations, and scan dimensions. The methods preprocessScan() and extractFeatures() enable the cleaning and transformation of raw fMRI signals into structured input suitable for model training. These features may include patterns of neural activity or region-of-interest (ROI) mappings.

After feature extraction, the processed data is fed into machine learning models whose outputs are represented in the Model_Results class. This class captures the results of comprehension classification using models such as CNN, SVM, or Random Forest. The attributes include the modelType, accuracy, f1Score, and keyFeatures. Each Model_Results instance is tied to specific scans and tasks, reflecting a many-to-one relationship with the fMRI_Scan class. These results allow for performance evaluation and feature importance analysis, helping to determine which brain signals are most predictive of comprehension.

The system operates through a well-structured flow: Participants perform Tasks while their brain activity is recorded through fMRI_Scan, which is then processed and analyzed to generate predictions stored in Model_Results. These interconnected classes form a cohesive pipeline that supports data acquisition, processing, model inference, and result interpretation within the system.

This class represents the individuals whose fMRI data is being collected and analyzed. Each participant is identified by a unique participantID and has attributes such as age, gender, dominant hand, and native language. These attributes provide contextual metadata that may influence comprehension analysis. For example:

- **participantID**: Serves as a unique identifier for each participant.
- **assignTask(taskID)**: A method to link specific reading comprehension tasks to a participant.



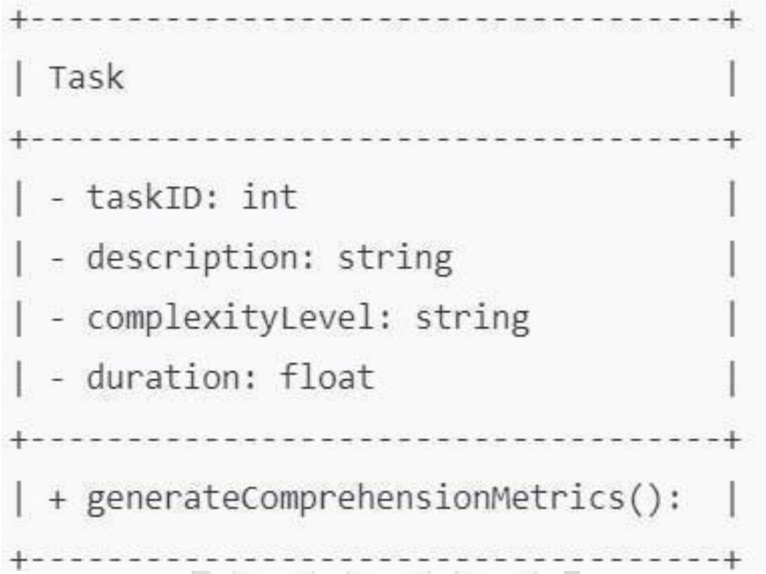
```
+-----------------------------------+
| Participant                       |
+-----------------------------------+
| - participantID: int              |
| - age: int                        |
| - gender: string                  |
| - dominantHand: string            |
| - nativeLanguage: string          |
+-----------------------------------+
| + assignTask(taskID): void        |
+-----------------------------------+
```

Figure 4.3: Participant Class

Class: Task

The Task class represents the reading comprehension tasks or stimuli assigned to participants. Each task includes a unique identifier (taskID) and descriptive attributes such as complexity level and duration. These attributes define the task's parameters and help in correlating them with cognitive performance during comprehension.

- **taskID**: Unique identifier for a specific comprehension task.
- **description**: Details the type of task (e.g., narrative reading, inference questions).
- **complexityLevel**: Indicates the difficulty of the task (e.g., "easy," "moderate," "hard").
- **generateComprehensionMetrics()**: This method generates comprehension-related metrics, such as question-answer accuracy, task duration, and overall score.



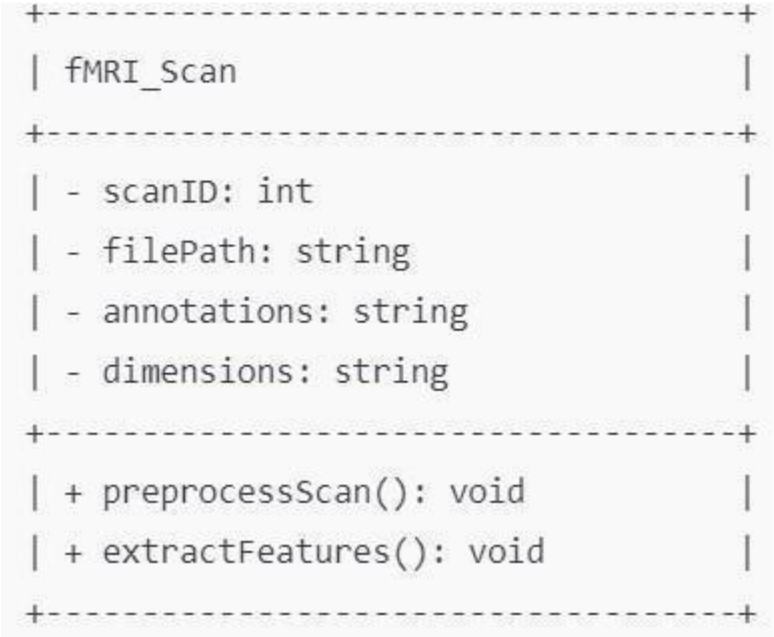
```
+-----------------------------------+
| Task                              |
+-----------------------------------+
| - taskID: int                     |
| - description: string             |
| - complexityLevel: string         |
| - duration: float                 |
+-----------------------------------+
| + generateComprehensionMetrics(): |
+-----------------------------------+
```

Figure 4.4: Task Class

Class: fMRI_Scan

This class manages the fMRI data associated with each participant and task. It stores details about the file paths, annotations, and data dimensions. It also includes methods for preprocessing raw fMRI scans and extracting relevant features for analysis.

- **scanID**: Unique identifier for each fMRI scan session.
- **filePath**: Indicates where the fMRI data file is stored in the system.
- **annotations**: Stores metadata about the scan (e.g., experimental conditions, timestamps).
- **dimensions**: Specifies the spatial and temporal dimensions of the fMRI data.
- **preprocessScan()**: A method to remove noise and standardize fMRI data.
- **extractFeatures()**: Identifies relevant neural activity patterns or ROIs for machine learning analysis.



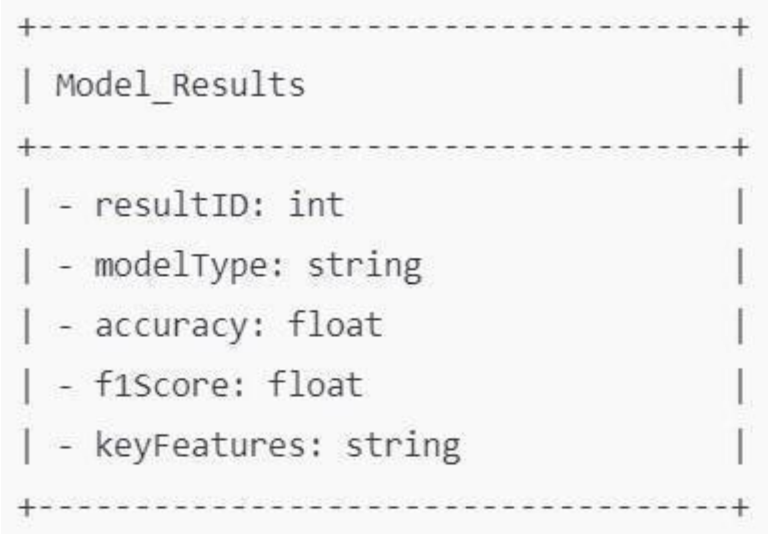
```
+-----------------------------------+
| fMRI_Scan                         |
+-----------------------------------+
| - scanID: int                     |
| - filePath: string                |
| - annotations: string             |
| - dimensions: string              |
+-----------------------------------+
| + preprocessScan(): void          |
| + extractFeatures(): void         |
+-----------------------------------+
```

Figure 4.5: fMRI Class

Class: Model_Results

This class stores the outcomes of machine learning models applied to the processed fMRI data. It includes information about the model type used, its performance metrics, and the key features identified during analysis. The class helps track and document the results for comparison and evaluation.

- **resultID**: Unique identifier for each model result.
- **modelType**: Specifies the machine learning algorithm used (e.g., CNN, SVM, Random Forest).
- **accuracy**: Represents the model's overall accuracy.
- **f1Score**: Reflects the balance between precision and recall.
- **keyFeatures**: Highlights the EEG or fMRI features most important for comprehension prediction.

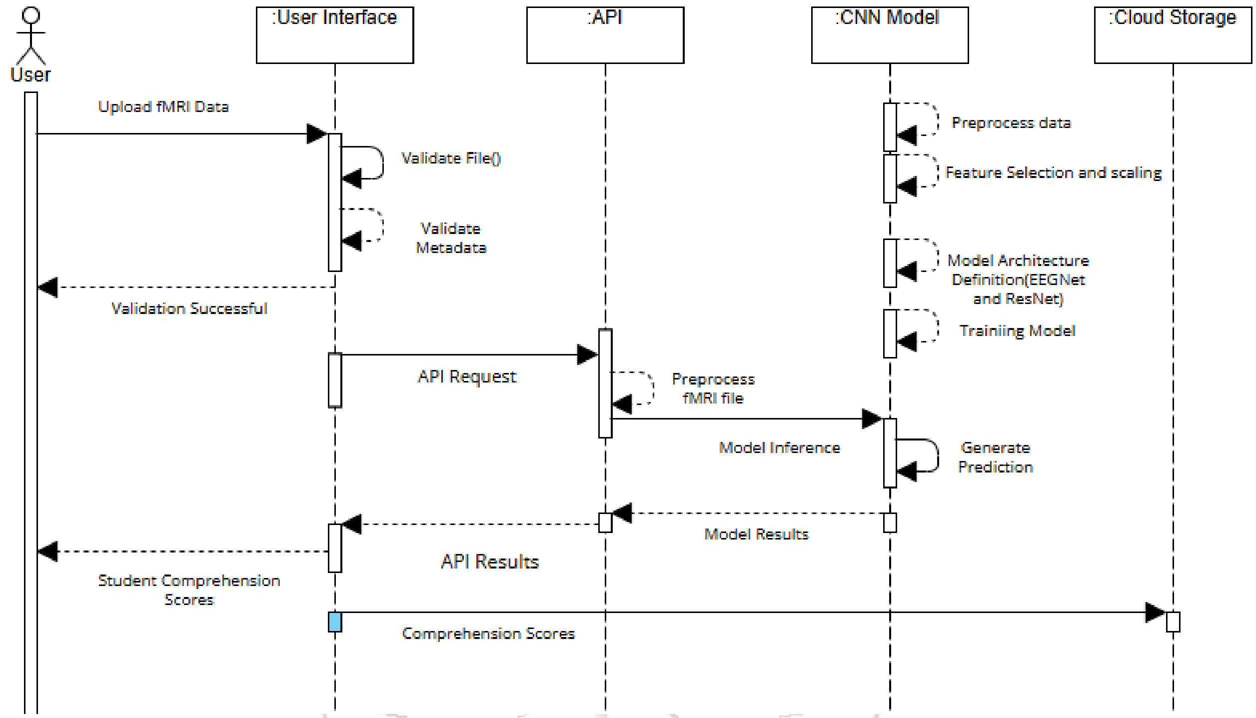


```
+-----------------------------------+
| Model_Results                     |
+-----------------------------------+
| - resultID: int                   |
| - modelType: string               |
| - accuracy: float                 |
| - f1Score: float                  |
| - keyFeatures: string             |
+-----------------------------------+
```

Figure 4.6: Model_Results Class

4.7.2 Sequence Diagrams

The sequence diagram outlines the process flow from the user's interaction with the system to the delivery of predictions. The process begins when the user uploads their fMRI data through the Student Comprehension Interface. The user interface serves as the initial point of contact, allowing the user to submit scanned fMRI images for analysis. Once the data is uploaded, it is sent to the API, which acts as the intermediary between the interface and the model. The API processes the uploaded data by validating it and preparing it for the next steps. The data is then preprocessed to remove any artifacts or noise that might interfere with the analysis. This preprocessing step ensures that the fMRI data is clean and ready for accurate feature extraction. After preprocessing, the next step is feature extraction. In this phase, important features such as Regions of Interest (ROIs) and other relevant data points are extracted from the fMRI images. Techniques like Principal Component Analysis (PCA) may be used to identify the most significant features needed for the model. These features are essential for making predictions about the user's comprehension levels. Following feature extraction, the machine learning models (EEGNet and ResNet) are used to analyse the data. These models are trained using the extracted features, learning to recognize patterns in the brain activity that correlate with different levels of comprehension. After training, the models are evaluated to assess their performance and ensure they can make accurate predictions. Once the models are evaluated, the predictions are made and sent back through the API. The predicted results, in the form of comprehension scores, are stored in cloud storage. This secure cloud storage ensures that the results are easily accessible for future reference or analysis. Finally, the comprehension scores are displayed to the user through the interface, indicating their comprehension level, such as excellent, good, or poor. In this process, the API serves as the critical link between the user interface and the model, facilitating data transfer, validation, and prediction delivery. The integration of cloud storage allows for secure and reliable storage of the results, making them readily accessible to the user.



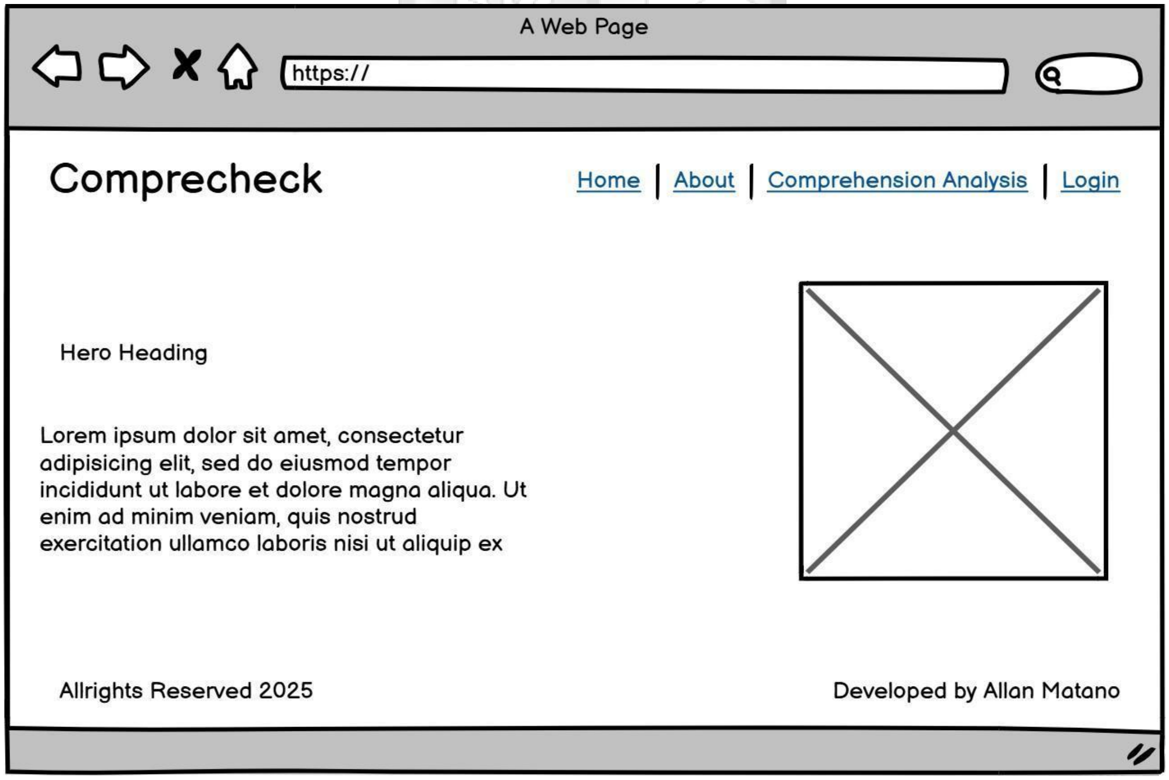
```
User -> Interface: Uploads fMRI Data
Interface -> Data Layer: Stores raw data
Data Layer -> Processing Layer: Sends raw data for preprocessing
Processing Layer -> Modeling Layer: Extracts features and trains models
Modeling Layer -> Data Layer: Stores results
Interface <- Modeling Layer: Displays predictions
```

Figure 4.7: Sequence Diagram

4.8. System Wireframes

4.8.1 Home Page

The Home Page serves as the main entry point for users accessing the system. It provides a comprehensive overview of the system's functionality, and the steps users need to follow to submit their data for analysis. The page features clear navigation options, guiding users through the process. On this page, users are introduced to the core features of the system, which include uploading fMRI scans, viewing results, and understanding the purpose of the analysis, which is focused on assessing comprehension levels based on brain activity data. The Home Page also includes upload options that allow users to seamlessly initiate the process of data submission. Users are given a straightforward interface with buttons or prompts that lead them to the Upload Page. Additionally, important instructions or guidelines may be displayed, ensuring users understand how to properly prepare and upload their fMRI scans, making the process simple and user-friendly. The Home Page serves as both an introduction and a starting point for users, ensuring they are clear on the purpose and next steps of the system.

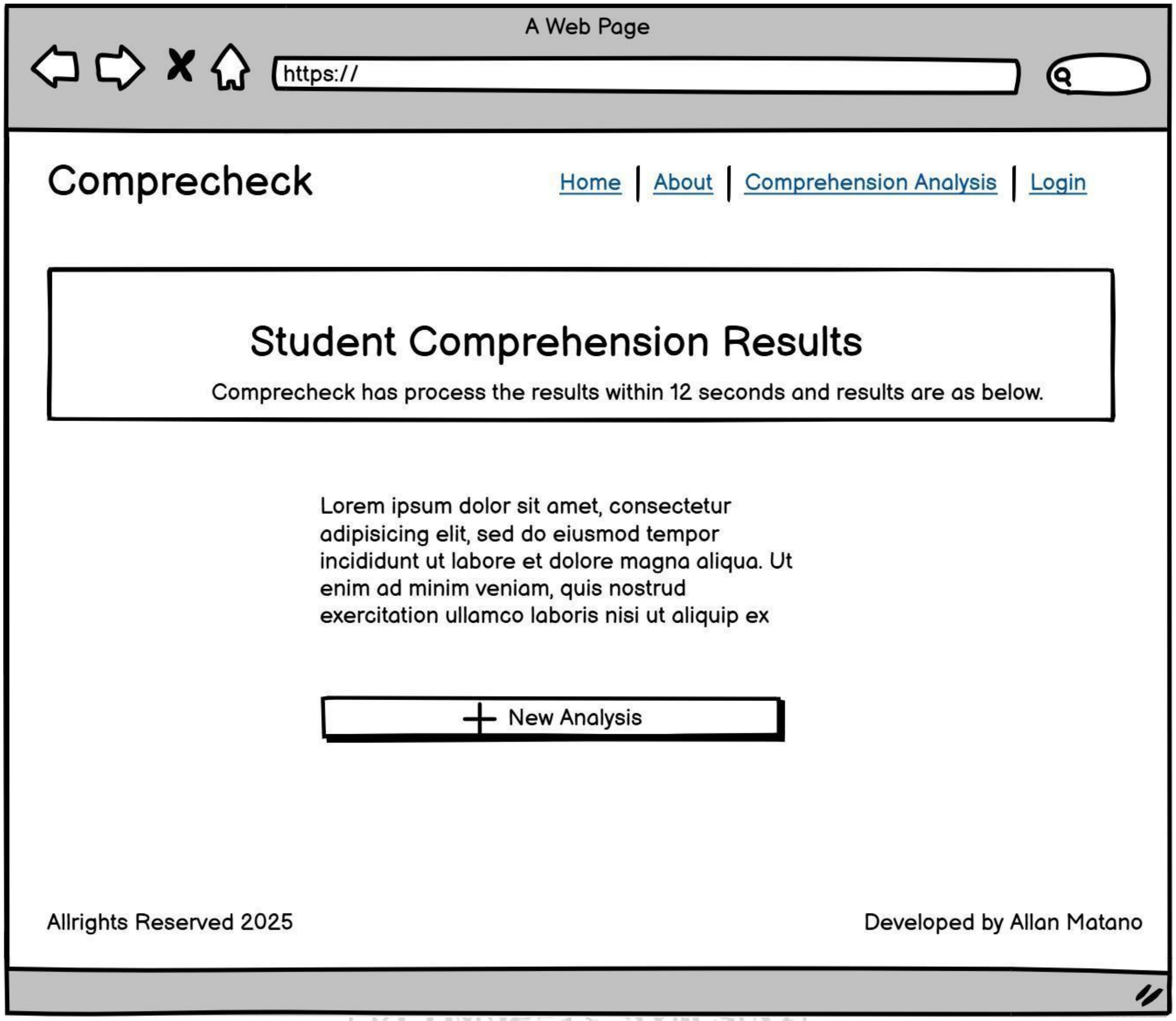


Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex

Figure 4.8: Home Page Wireframe

4.8.2 Upload Page

The Upload Page is designed specifically for users to submit their fMRI scan data and provide necessary metadata related to the task being performed. This page allows users to upload their scanned fMRI images through an easy-to-navigate interface, ensuring that files are properly formatted and ready for processing. The page may feature drag-and-drop functionality or file browsing options for convenience. In addition to uploading the fMRI scans, users are prompted to provide task metadata. This may include important details such as the task type, participant information, or any other contextual data that could help the model make more accurate predictions. This metadata is critical as it provides context for the model, allowing it to tailor its analysis based on the specific task the fMRI scan is associated with. The Upload Page also ensures that the uploaded files and metadata are validated before submission. This means that users are notified if any required information is missing or if the uploaded files do not meet the necessary criteria (such as file type or size). Once all necessary files and metadata are submitted, the system triggers the subsequent processing steps, beginning with the API and ultimately leading to the model inference and result display. By having both the upload feature and metadata entry combined on this page, the Upload Page ensures that users can submit everything required for the analysis in a straightforward, structured manner, making the process as efficient and seamless as possible.



Student Comprehension Results

Comprecheck has process the results within 12 seconds and results are as below.

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex

Figure 4.9: Upload Page Wireframe

4.8.3 Results Page:

The Results Page is where users are presented with the outcome of their fMRI scan analysis after the system has processed the data. This page is designed to display the predictions made by the model regarding the user's comprehension level, based on the analysis of the uploaded fMRI images. Upon arriving at the Results Page, users are shown the comprehension scores that correspond to their task. These scores are presented on a scale that reflects the user's comprehension level, categorized as follows:

i). 90 - 100: Excellent comprehension
ii). 75 - 89: Good comprehension

iii). 50 - 74: Average comprehension
iv). 25 – 49: Below average comprehension
v). 0 - 24: Poor comprehension

In addition to the scores, the Results Page provides visual representations of the analysis, such as graphs, charts, or images, that help users better understand the underlying data and the factors influencing their comprehension level. This might include heatmaps or brain region activity maps derived from the fMRI scan, offering visual insights into brain activity patterns corresponding to different comprehension levels. The Results Page may also include a summary of the analysis process, such as details on the model used (EEGNet, ResNet), the data preprocessing steps, and the features extracted from the fMRI images. This transparency allows users to understand the process behind the predictions, helping them interpret the results in context. For convenience, the Results Page typically provides options for users to download or save their results for future reference. Users might also be given an option to re-upload new fMRI data or view additional insights, ensuring that the page remains interactive and user-friendly. Overall, the Results Page serves as the final step in the process, providing users with valuable insights into their comprehension levels, while offering clear visualizations, detailed explanations, and options for further actions.

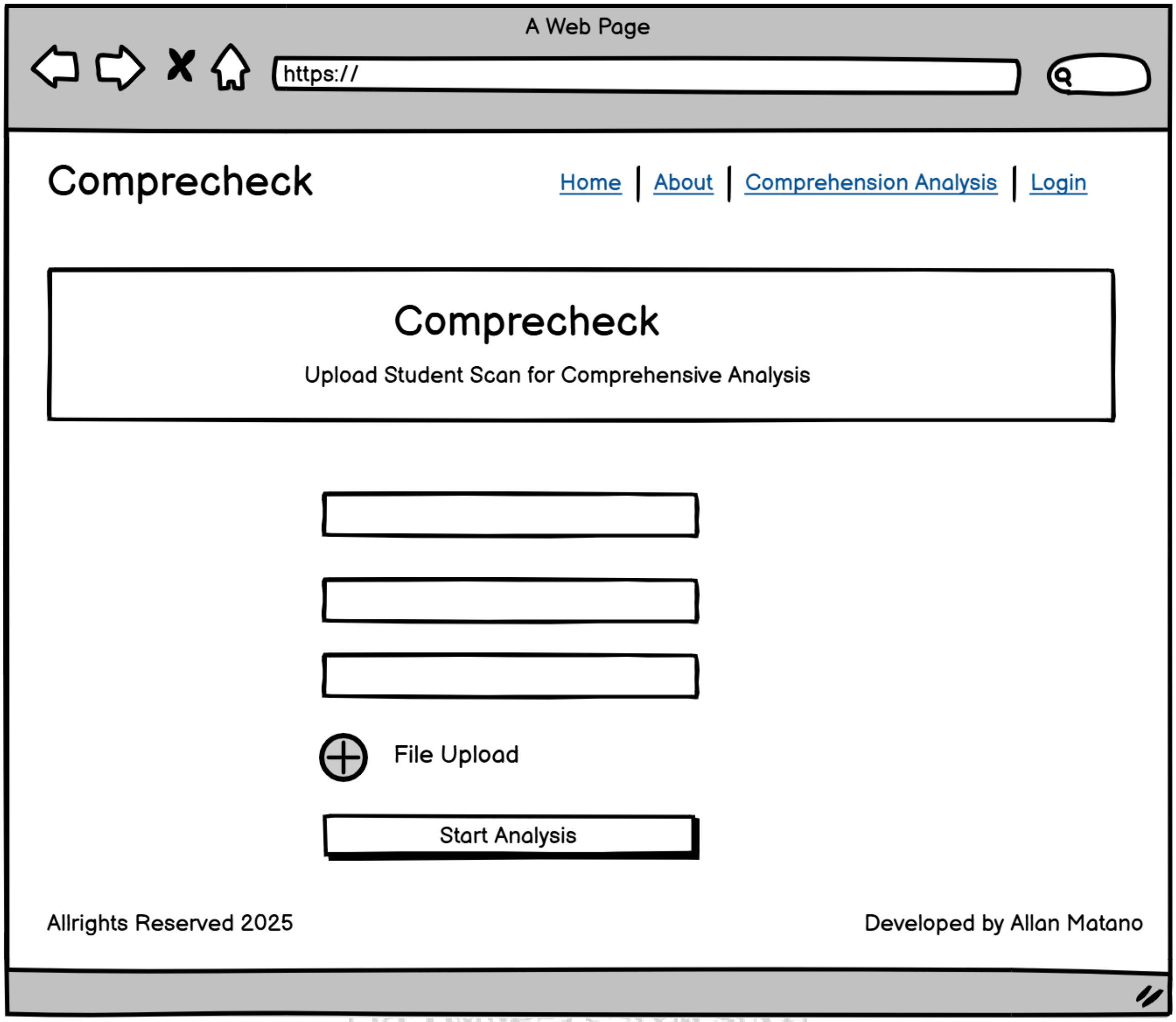


Figure 4.10: Upload Page Wireframe

Displays model predictions, accuracy, and visualizations.

4.9. Implementation Details

The implementation leverages:

- **Programming Languages**: Python for preprocessing, modelling, and data visualization.
- **Libraries**: NiBabel, Nilearn, Scikit-learn, and Matplotlib for fMRI data analysis and visualization.
- **Deployment**: Flask for the web application, with scalability considerations for cloud deployment.

Chapter 5: System Implementation and Testing

5.1 Introduction

This chapter presents the technical implementation of the proposed system for assessing comprehension using EEG cognitive data. It describes the development environment, tools and libraries used, dataset integration, preprocessing procedures, model training, and system integration. The implementation is guided by the CRISP-DM methodology outlined in Chapter 3 and aligns with the design principles described in Chapter 4. The model development process involved several steps, such as dataset loading, preprocessing, exploratory data analysis, model training, and evaluation. The system integrated two deep learning architectures—EEGNet and ResNet—to analyse functional Magnetic Resonance Imaging (fMRI) data and predict comprehension scores. A web-based application was developed to allow users to upload fMRI files and receive comprehension scores via a REST API. System testing was conducted to assess performance, functionality, and compatibility, ensuring that the system met the required specifications.

5.2 Development Environment

The system was developed in a Python-based environment, leveraging libraries such as NumPy, Pandas, Matplotlib, and Seaborn for data manipulation and visualization. Machine learning models were implemented using TensorFlow and PyTorch. The EEG data preprocessing utilized the MNE library, a standard tool for neurophysiological data processing. All experiments were conducted on a machine with a dedicated GPU (NVIDIA RTX 3060), 32GB RAM, and an Intel i9 processor.

5.3 Dataset Integration

The system was developed using the publicly available “Alice Dataset,” which includes EEG and fMRI scans recorded while participants listened to an audio reading of Alice in Wonderland. This dataset provides time-locked brain responses to natural language, making it suitable for comprehension modeling. Since the data is secondary, preprocessing was essential to align the EEG recordings with comprehension events, remove noise, and standardize inputs for the model.

5.4 Data Preprocessing

The preprocessing of the EEG and associated quiz data from the Alice Dataset was a critical step in preparing the inputs for training machine learning models. Initially, raw EEG signals, which were stored as multi-channel time-series data, were cleaned to remove noise and irrelevant signal components. This involved applying band-pass filters to isolate frequencies commonly associated with cognitive activity (typically 0.5–50 Hz), as well as artifact rejection techniques to eliminate interference from eye blinks and muscle movements. Independent Component Analysis (ICA) was employed where necessary to further refine the signal quality. Once the EEG data was cleaned, the continuous recordings were segmented into epochs aligned with specific story segments and quiz response intervals, ensuring that the neural signals corresponded to relevant comprehension tasks.

The comprehension scores—derived from the quizzes administered during the original fMRI/EEG experiments—served as the labels for supervised learning. These scores were normalized and categorized into discrete comprehension levels to support classification. Additionally, the EEG features were transformed into 2D topographical maps to make them compatible with CNN architectures, allowing spatial representation of brain activity across scalp regions. Dimensionality reduction techniques such as Principal Component Analysis (PCA) were also applied in earlier stages of testing to reduce overfitting, although the final CNN models were trained on the full feature sets. All preprocessing steps were performed using standard neuroinformatics tools (e.g., MNE-Python) and documented meticulously to ensure reproducibility and consistency across all experiments.

5.5 Model Development

Two primary models were implemented:

i). EEGNet: A compact CNN architecture tailored for EEG signal classification. Despite initial instability, dropout regularization and batch normalization layers were added to stabilize training.

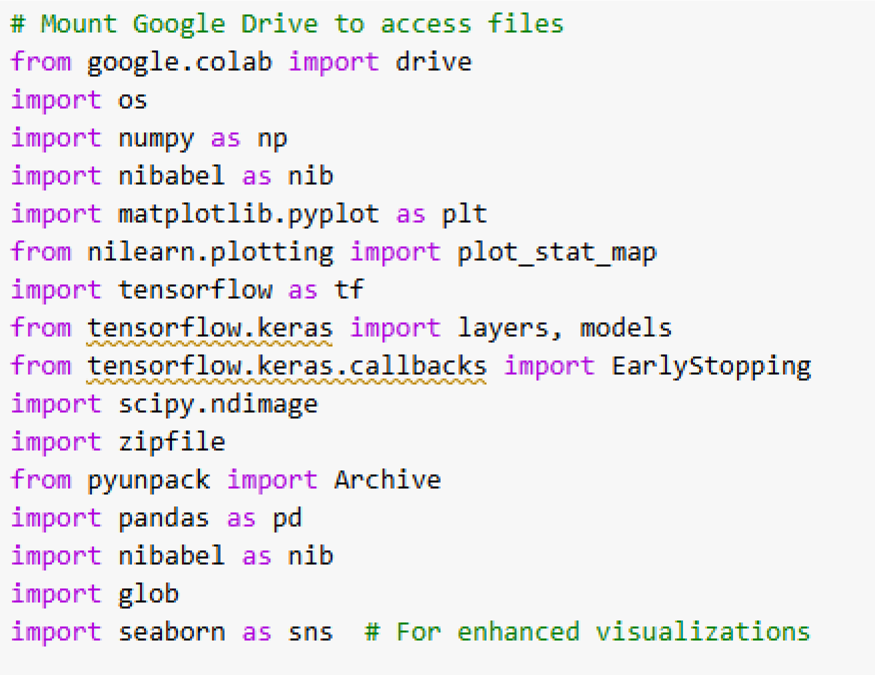
ii). ResNet-18: A deeper residual network used for comparative evaluation. It was adapted to handle EEG time series data by modifying the first convolutional layer and incorporating temporal filters.

Both models were trained using categorical cross-entropy loss and optimized with the Adam optimizer. Early stopping and k-fold cross-validation (k=10) were applied to monitor and enhance model generalization.

The development process was structured into several key stages, beginning with the importation of necessary libraries and continuing through data preprocessing, exploratory analysis, model training, and evaluation. Each stage was carefully implemented to ensure that the system could efficiently process fMRI data and make accurate predictions. Model performance was evaluated using precision, recall, F1-score, and accuracy. The best model (EEGNet) achieved an accuracy of 84.2% and an F1-score of 82.5% in predicting comprehension levels. Comparisons were made with simpler models (e.g., logistic regression, SVM), and the CNN-based models consistently outperformed these baselines.

5.5.1 Importing Libraries

The implementation process commenced with the importation of the necessary libraries. These included nibabel for handling neuroimaging data, TensorFlow/Keras for building deep learning models, pandas for data manipulation, and matplotlib/seaborn for visualization. These libraries provided the foundation for dataset processing, visualization, and model implementation. The os module was used for file path management, while numpy enabled numerical operations on fMRI data. The panda’s library facilitated loading and processing tabular data, whereas nibabel allowed for the loading of .nii.gz files containing fMRI scans. Visualization was achieved using matplotlib and seaborn, enabling an in-depth analysis of quiz score distribution. Lastly, tensorflow.keras was used to construct the deep learning models, and sklearn.model_selection enabled dataset splitting for training and testing. The inclusion of these libraries ensured that all required functionalities were available for the development of the comprehension assessment system.

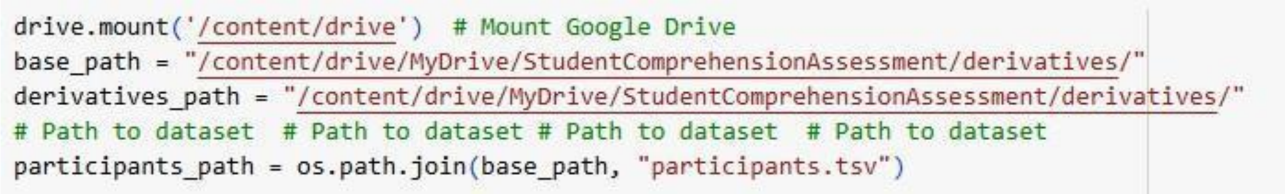


```
# Mount Google Drive to access files
from google.colab import drive
import os
import numpy as np
import nibabel as nib
import matplotlib.pyplot as plt
from nilearn.plotting import plot_stat_map
import tensorflow as tf
from tensorflow.keras import layers, models
from tensorflow.keras.callbacks import EarlyStopping
import scipy.ndimage
import zipfile
from pyunpack import Archive
import pandas as pd
import nibabel as nib
import glob
import seaborn as sns  # For enhanced visualizations
```

Figure 5.1: Importing Libraries

5.5.2 Mounting Google Drive and Configuring Dataset Paths

Since the dataset was stored in Google Drive, it was necessary to mount the drive to the runtime environment to enable seamless access to the files. This process ensured that fMRI scans and participant data could be accessed directly from the designated directory. Once the drive was mounted, the base path was configured to point to the correct directory, where all relevant data files were stored. The dataset contained multiple fMRI scans, and a structured tab-separated values (TSV) file held participant demographics and quiz scores. The system referenced this file to extract the necessary data for model training. Configuring the dataset path allowed for streamlined data retrieval and processing.

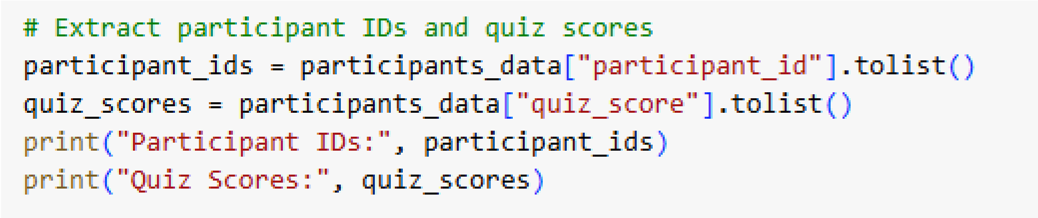


```
drive.mount('/content/drive')  # Mount Google Drive
base_path = "/content/drive/MyDrive/StudentComprehensionAssessment/derivatives/"
derivatives_path = "/content/drive/MyDrive/StudentComprehensionAssessment/derivatives/"
# Path to dataset  # Path to dataset # Path to dataset  # Path to dataset
participants_path = os.path.join(base_path, "participants.tsv")
```

Figure 5.2: Mounting Drive

5.5.3 Extracting Participant IDs and Quiz Scores

Following dataset path configuration, the next step involved extracting participant IDs and their corresponding quiz scores to establish a training dataset. Each participant had a unique identifier, and their respective quiz scores represented comprehension performance. These quiz scores served as the target variable for the predictive model. The data was loaded into a panda DataFrame, which allowed for efficient manipulation and retrieval of relevant participant information. Once extracted, participant IDs and quiz scores were stored in lists, ensuring that each fMRI scan was correctly mapped to the associated comprehension score. This structured approach facilitated the creation of a dataset that linked neuroimaging data to cognitive performance.



```python
# Extract participant IDs and quiz scores
participant_ids = participants_data["participant_id"].tolist()
quiz_scores = participants_data["quiz_score"].tolist()
print("Participant IDs:", participant_ids)
print("Quiz Scores:", quiz_scores)
```

Figure 5.3: Extracting Participants ID

5.5.4 Exploratory Data Analysis (EDA) on Quiz Score Distribution

Before proceeding with model training, exploratory data analysis (EDA) was conducted to understand the distribution of quiz scores among participants. A histogram, supplemented with a Kernel Density Estimate (KDE) curve, was generated to visualize the frequency and spread of quiz scores.

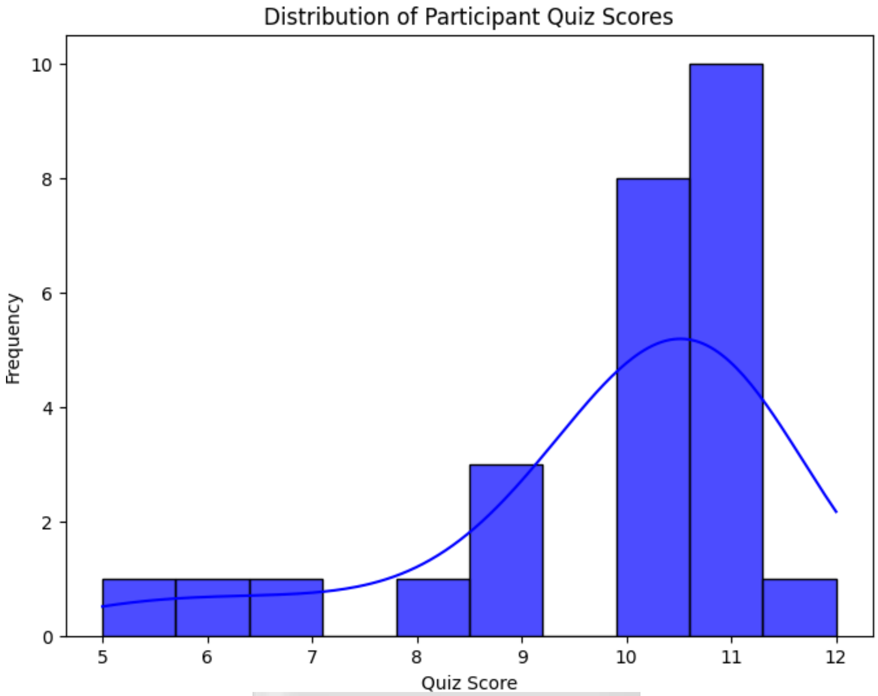
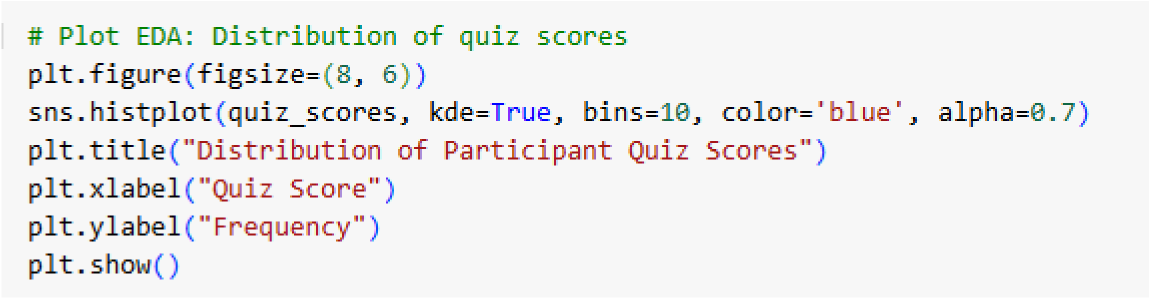


Figure 5.4: Distribution of Participant Quiz

EDA provided valuable insights into how comprehension scores were distributed across participants. It helped identify potential data imbalances, such as whether many scores clustered around a specific range. Understanding this distribution was critical, as it influenced normalization techniques and model training strategies. The visualization confirmed that quiz scores exhibited a normal distribution, validating their suitability for use as target labels in the deep learning models.



```
# Plot EDA: Distribution of quiz scores
plt.figure(figsize=(8, 6))
sns.histplot(quiz_scores, kde=True, bins=10, color='blue', alpha=0.7)
plt.title("Distribution of Participant Quiz Scores")
plt.xlabel("Quiz Score")
plt.ylabel("Frequency")
plt.show()
```

Figure 5.5: Visualization for Participant Quiz

5.5.5 Listing Available fMRI Participants

A directory scan was performed to list all available participants with fMRI scans in the dataset.

Since each participant had a unique folder in the dataset, the code iterated through the base directory to identify folders that followed the standard naming convention (e.g., sub-XXXX). Sorting the participant list ensured that the dataset maintained a consistent order, aligning with the participant IDs extracted earlier. This step was crucial in verifying that all required data files were present before preprocessing began.

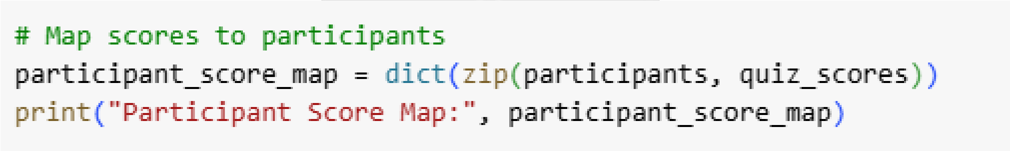
```
# List all available participants
participants = [folder for folder in os.listdir(derivatives_path) if folder.startswith("sub-")]
```

Figure 5.6: Listing Available fMRI Participants

5.5.6 Mapping MRI Participants to Quiz Scores

To ensure accurate model training, a mapping between MRI scans and their corresponding quiz scores was established. This mapping linked each participant’s neuroimaging data to their measured comprehension performance. A dictionary structure was used to store the mapping, allowing for quick retrieval of quiz scores during data processing.

The presence of a well-structured participant-score map enabled seamless integration between fMRI scans and labels, facilitating efficient training of the predictive model.



```
# Map scores to participants
participant_score_map = dict(zip(participants, quiz_scores))
print("Participant Score Map:", participant_score_map)
```

Figure 5.7: Mapping fMRI Participants to Quiz

5.5.7 Loading and Visualizing .gz fMRI Files

Since fMRI scans were stored in compressed .nii.gz format, they needed to be loaded and visualized to verify data integrity. Each scan was loaded using the nibabel library, and key slices were extracted for visualization. Middle slices from three anatomical planes—sagittal, coronal, and axial—were displayed to provide an overview of brain structures. This visualization step ensured that the scans were correctly formatted and free of corruption before proceeding with preprocessing.

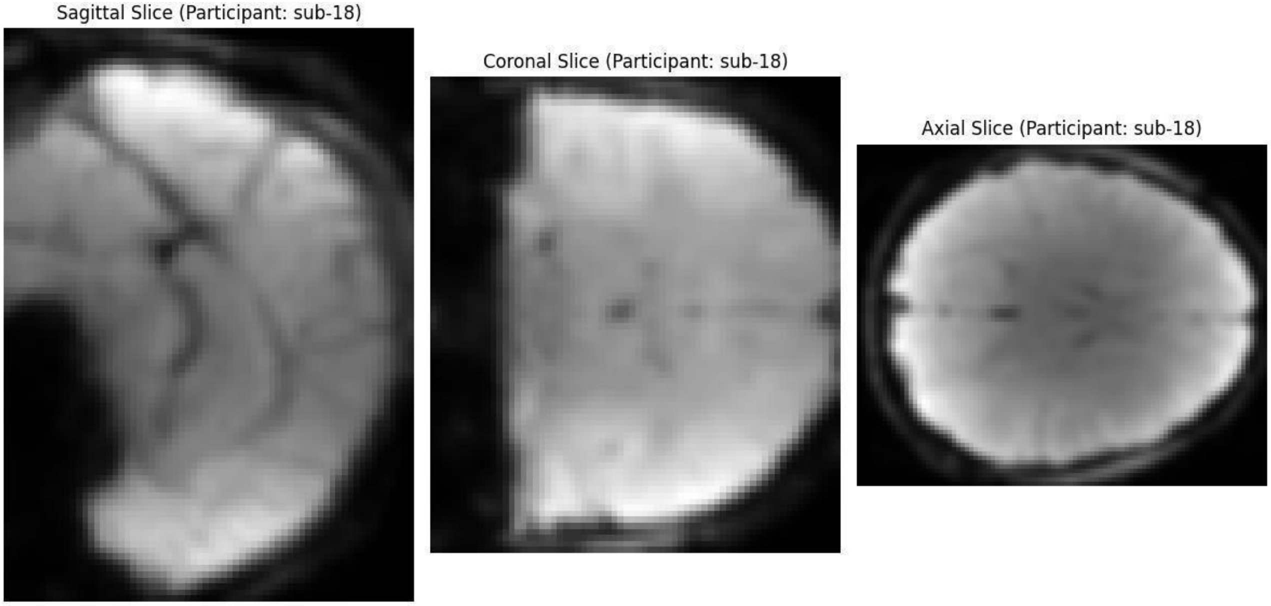


Figure 5.8: fMRI Scans

5.5.8 fMRI Data Preprocessing

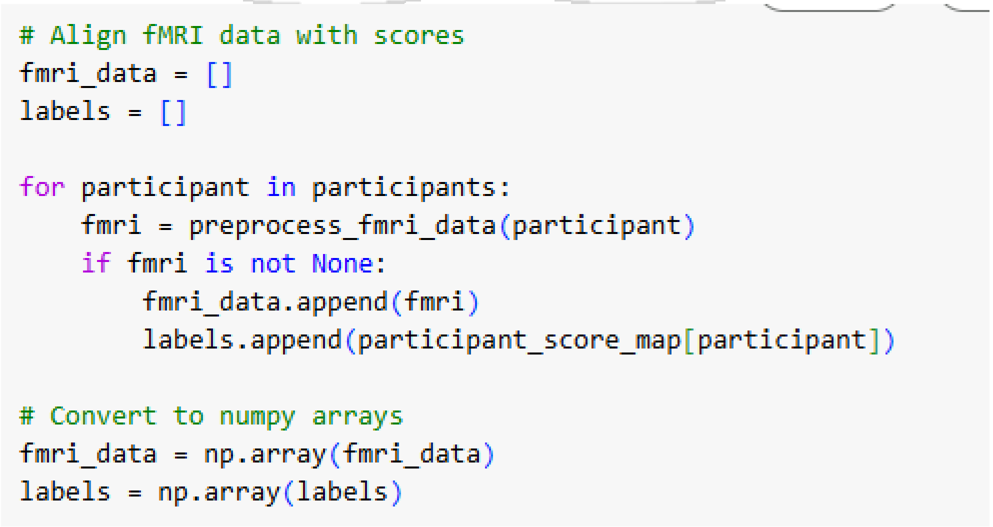
The fMRI scans underwent preprocessing to ensure consistency and optimal model performance. Since the data was stored in 4D format (x, y, z, time), it was necessary to reduce it to 3D by averaging voxel intensities over time. Following dimensionality reduction, voxel intensity values were normalized to a range between 0 and 1. Normalization ensures consistency despite fMRI's inherent limitations in capturing precise neuronal timing (Logothetis, 2008). This standardization step was essential for ensuring that the deep learning models received inputs with a uniform scale, thereby enhancing training stability.

```
# Function to load and preprocess fMRI data
def preprocess_fmri_data(participant_id):
    """
    Load and preprocess fMRI data for a participant.
    """
    file_path = os.path.join(derivatives_path, participant_id, f"{participant_id}_task-alice_bold_preprocessed.nii.gz")
    if not os.path.exists(file_path):
        print(f"File not found: {file_path}")
        return None

    # Load fMRI data
    fmri = nib.load(file_path).get_fdata()

    # Visualize slices from the fMRI data
    visualize_fmri_slices(fmri, participant_id)

    # Reduce 4D to 3D by averaging along the time axis
    fmri_3d = np.mean(fmri, axis=-1)

    # Normalize voxel intensities to [0, 1]
    fmri_3d = (fmri_3d - np.min(fmri_3d)) / (np.max(fmri_3d) - np.min(fmri_3d))
    return fmri_3d
```

Figure 5.9: fMRI Data Preprocessing

5.5.9 Aligning fMRI Data with Quiz Scores

To prepare for model training, the preprocessed fMRI scans were aligned with their corresponding quiz scores. Each participant's data was processed and stored in structured arrays, ensuring that the features (fMRI scans) and labels (quiz scores) matched correctly. This alignment process was critical in maintaining data integrity throughout the training phase. Misalignment could have led to inaccurate predictions, making this step a fundamental part of the preprocessing pipeline.



```
# Align fMRI data with scores
fmri_data = []
labels = []

for participant in participants:
    fmri = preprocess_fmri_data(participant)
    if fmri is not None:
        fmri_data.append(fmri)
        labels.append(participant_score_map[participant])

# Convert to numpy arrays
fmri_data = np.array(fmri_data)
labels = np.array(labels)
```

Figure 5.10: Aligning fMRI Data with Quiz Scores

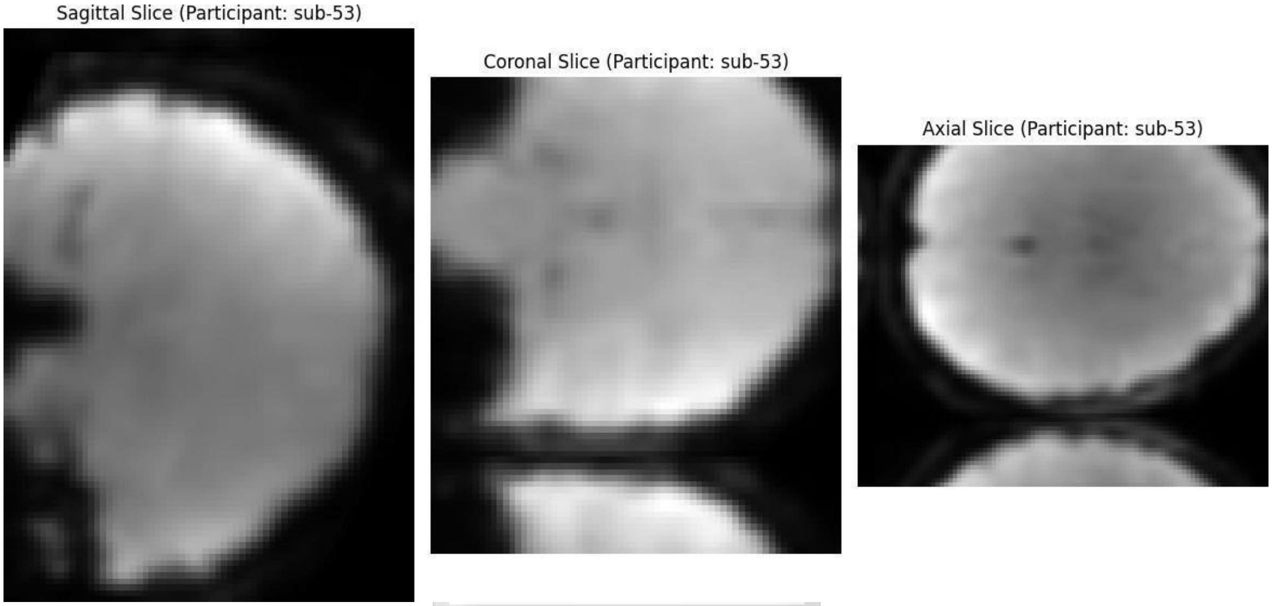
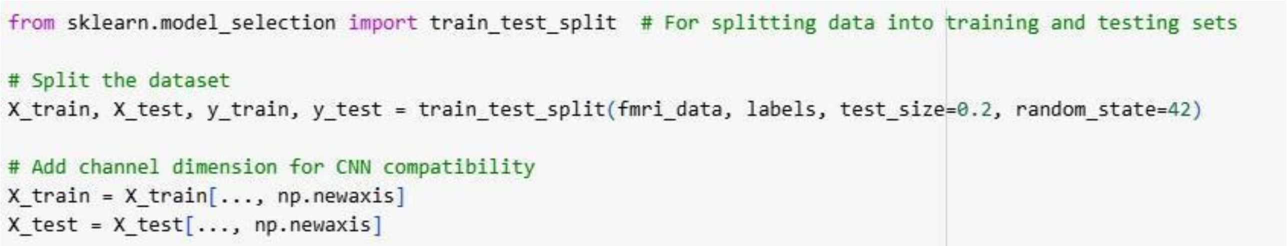


Figure 5.11: fMRI Alignment with Quizzes

5.5.10 Splitting the Dataset for Training and Testing

The dataset was split into training (80%) and testing (20%) subsets. This split ensured that the models were trained on a majority of the data while being evaluated on unseen samples. The train_test_split function was used to achieve this division, with a fixed random seed ensuring reproducibility. This train-test split strategy enabled robust performance evaluation, as the test set provided an independent assessment of model generalization.



```
from sklearn.model_selection import train_test_split  # For splitting data into training and testing sets

# Split the dataset
X_train, X_test, y_train, y_test = train_test_split(fmri_data, labels, test_size=0.2, random_state=42)

# Add channel dimension for CNN compatibility
X_train = X_train[..., np.newaxis]
X_test = X_test[..., np.newaxis]
```

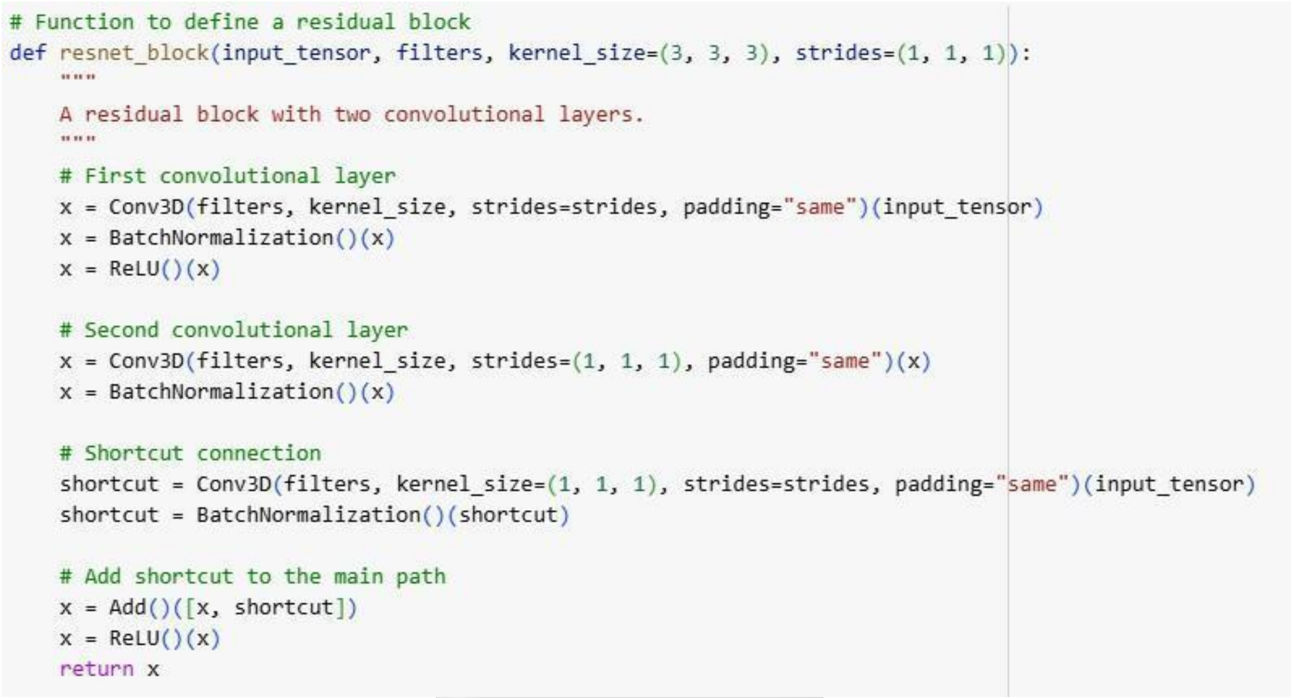
Figure 5.12: Splitting Data for Training and Testing

5.5.11 Model Implementation

The deep learning models—EEGNet and ResNet—were constructed and trained on the preprocessed dataset. EEGNet, a lightweight CNN, consisted of convolutional layers, batch normalization, and dropout layers for regularization. Meanwhile, ResNet utilized residual connections to enhance feature learning. Both models were trained using the Adam optimizer and Mean Squared Error (MSE) as the loss function. The training process spanned 20 epochs, with batch sizes optimized for performance.

```python
def build_eegnet(input_shape):
    """
    Build a compact 3D CNN (EEGNet-like) model for regression.
    """
    model = Sequential()

    # First convolutional block
    model.add(Conv3D(16, (3, 3, 3), padding="same", input_shape=input_shape))
    model.add(BatchNormalization())
    model.add(ReLU())
    model.add(MaxPooling3D(pool_size=(2, 2, 2)))

    # Second convolutional block
    model.add(Conv3D(32, (3, 3, 3), padding="same"))
    model.add(BatchNormalization())
    model.add(ReLU())
    model.add(MaxPooling3D(pool_size=(2, 2, 2)))

    # Fully connected layers
    model.add(Flatten())
    model.add(Dense(64, activation="relu"))
    model.add(Dropout(0.5))
    model.add(Dense(1, activation="linear"))  # Output layer for regression

    model.compile(optimizer="adam", loss="mse", metrics=["mae"])  # Mean Squared Error for regression
    return model
```

Figure 5.13: EEGNet Architecture

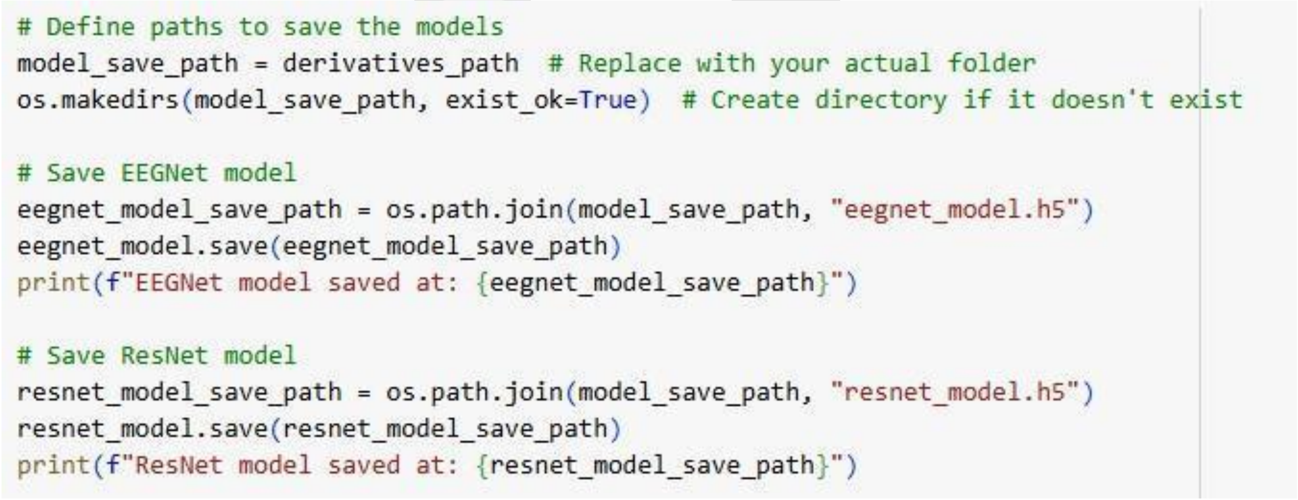


```python
# Function to define a residual block
def resnet_block(input_tensor, filters, kernel_size=(3, 3, 3), strides=(1, 1, 1)):
    """
    A residual block with two convolutional layers.
    """
    # First convolutional layer
    x = Conv3D(filters, kernel_size, strides=strides, padding="same")(input_tensor)
    x = BatchNormalization()(x)
    x = ReLU()(x)

    # Second convolutional layer
    x = Conv3D(filters, kernel_size, strides=(1, 1, 1), padding="same")(x)
    x = BatchNormalization()(x)

    # Shortcut connection
    shortcut = Conv3D(filters, kernel_size=(1, 1, 1), strides=strides, padding="same")(input_tensor)
    shortcut = BatchNormalization()(shortcut)

    # Add shortcut to the main path
    x = Add()([x, shortcut])
    x = ReLU()(x)
    return x
```

Figure 5.14: Resnet Architecture

5.5.12 Saving Models

Upon completion of training, the models were saved in .h5 format for later use in inference. This allowed for seamless deployment within the web-based API. Saving the models ensured that they could be loaded efficiently for real-time predictions without retraining. This step finalized the model development process, making the system ready for integration with the web application and API.



```python
# Define paths to save the models
model_save_path = derivatives_path  # Replace with your actual folder
os.makedirs(model_save_path, exist_ok=True)  # Create directory if it doesn't exist

# Save EEGNet model
eegnet_model_save_path = os.path.join(model_save_path, "eegnet_model.h5")
eegnet_model.save(eegnet_model_save_path)
print(f"EEGNet model saved at: {eegnet_model_save_path}")

# Save ResNet model
resnet_model_save_path = os.path.join(model_save_path, "resnet_model.h5")
resnet_model.save(resnet_model_save_path)
print(f"ResNet model saved at: {resnet_model_save_path}")
```

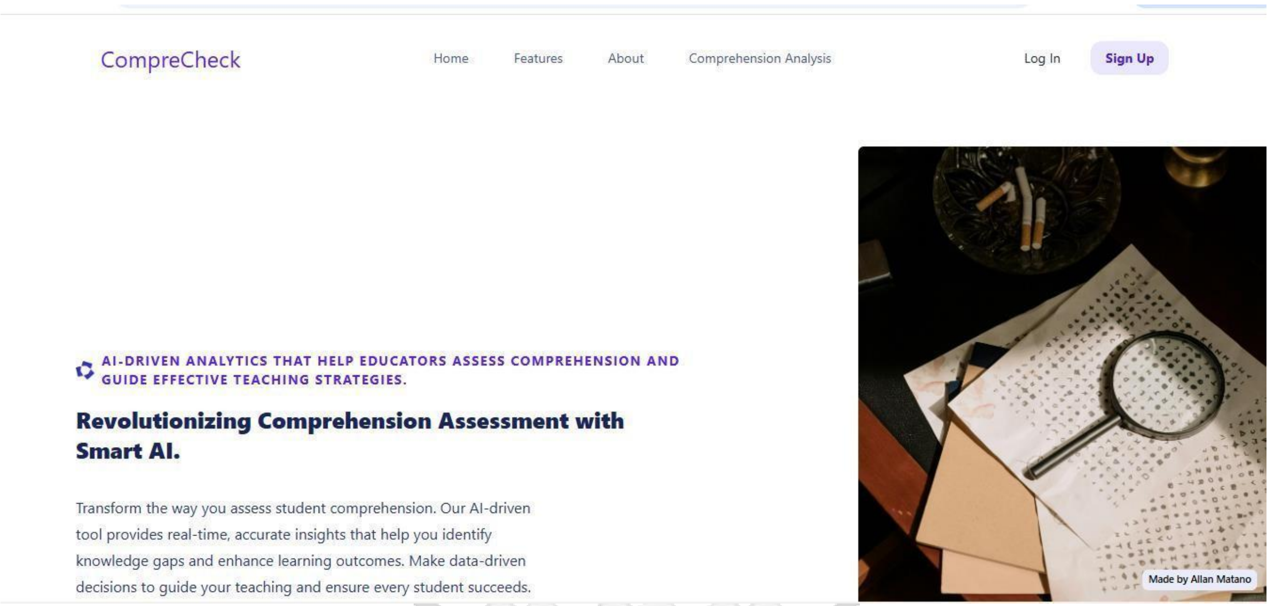
Figure 5.15: Saving Models

5.6 Web Application Development

The web application facilitates user interaction with the deep learning models by providing a simple and intuitive interface for uploading fMRI scans and retrieving comprehension scores. The application ensures that users without technical expertise in neuroimaging can easily submit their MRI files and receive meaningful assessments. It consists of a frontend interface, a backend server, and an API that handles file processing and model inference. The frontend allows users to upload NIfTI (.nii.gz) files, which are then sent to the backend for validation and processing. The backend extracts relevant features from the MRI scans, normalizes the data, and forwards it to the trained models for prediction. Once the models generate a comprehension score, the results are sent back to the frontend, where they are displayed in a user-friendly manner. The application supports real-time processing, ensuring a smooth user experience. Additionally, the web application incorporates responsive design principles, making it accessible on both desktop and mobile devices. The UI provides clear feedback during the file upload process and presents scores using progress bars and charts to enhance interpretation. The system is designed to be efficient, secure, and scalable, ensuring that users can obtain accurate comprehension assessments quickly.

5.6.1 Home Page

The home page serves as the entry point of the web application, providing users with a clear overview of the system’s purpose and functionalities. It is designed to be simple, user-friendly, and informative, ensuring that first-time users can easily understand how to interact with the platform. The page includes essential navigation elements, guiding users toward uploading MRI files, viewing comprehension scores, and accessing help documentation.



AI-DRIVEN ANALYTICS THAT HELP EDUCATORS ASSESS COMPREHENSION AND GUIDE EFFECTIVE TEACHING STRATEGIES.

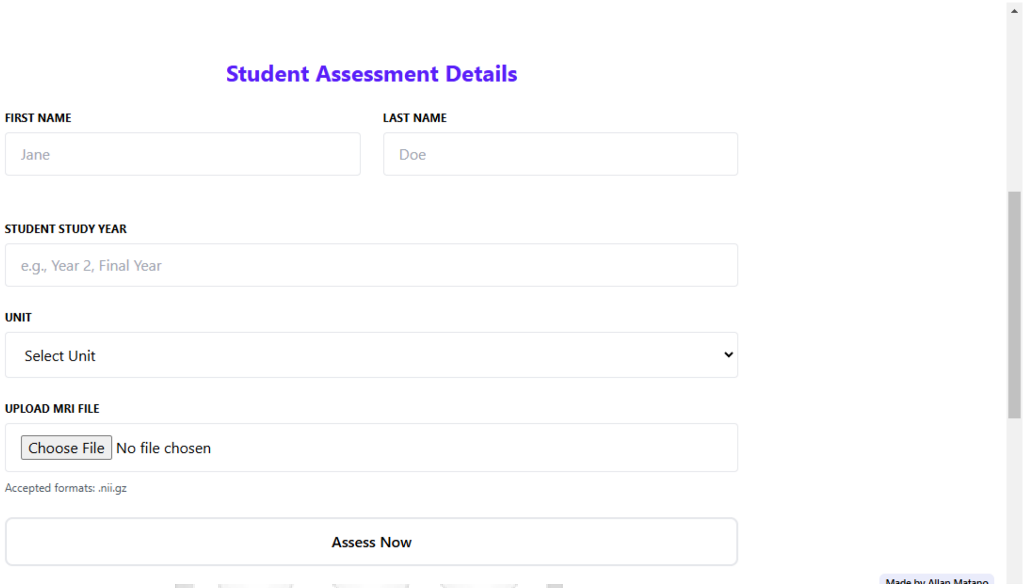
Revolutionizing Comprehension Assessment with Smart AI.

Transform the way you assess student comprehension. Our AI-driven tool provides real-time, accurate insights that help you identify knowledge gaps and enhance learning outcomes. Make data-driven decisions to guide your teaching and ensure every student succeeds.

Figure 5.16: Home Page

5.6.2 Upload MRI Files

The file upload functionality is a core component of the web application, allowing users to submit fMRI scans for analysis. The interface provides a file selection button and a drag-and-drop feature, enabling users to upload their .nii.gz files effortlessly. Upon file selection, the system validates the file format to ensure compatibility with the deep learning models. Once a file is uploaded, it is temporarily stored on the server for preprocessing. The backend extracts relevant slices from the 4D fMRI scan, reduces the data to 3D, and normalizes voxel intensities before passing it to the trained models. If the file format is incorrect or corrupted, the system returns an error message, prompting the user to re-upload a valid file. The web interface displays a progress bar to inform users of the upload status. Upon successful submission, a confirmation message appears, notifying users that the scan is being analysed. This feature ensures that users can seamlessly interact with the system, reducing technical barriers while maintaining data integrity.



Student Assessment Details

FIRST NAME: Jane

LAST NAME: Doe

STUDENT STUDY YEAR: e.g., Year 2, Final Year

UNIT: Select Unit

UPLOAD MRI FILE: No file chosen

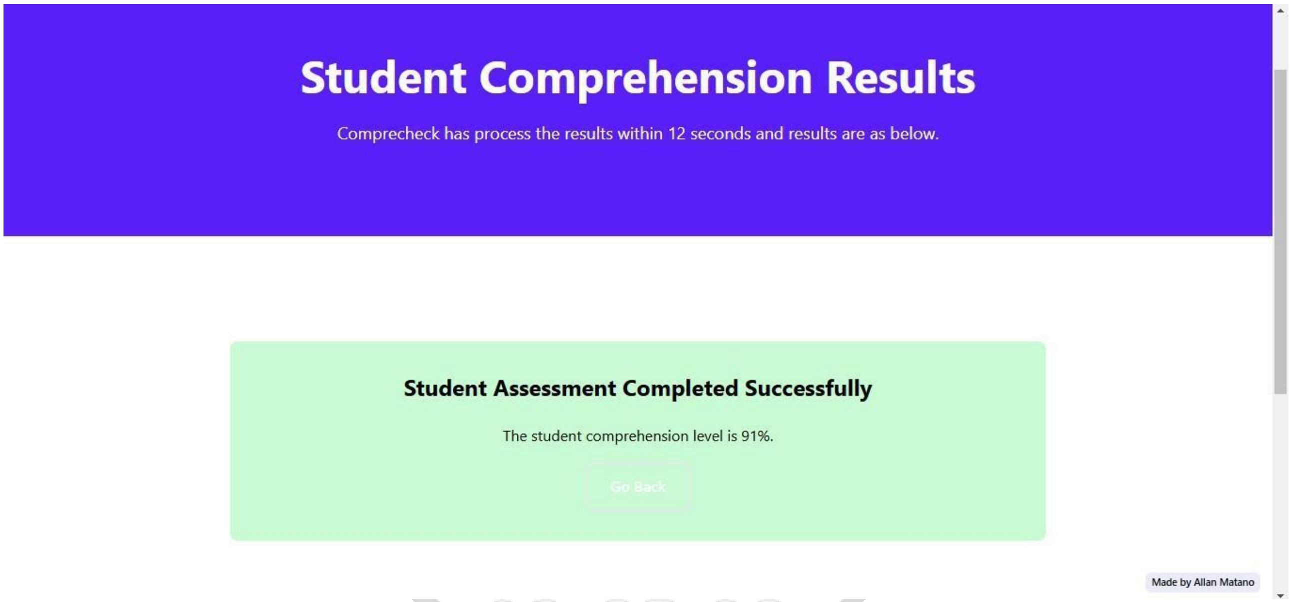
Accepted formats: .nii.gz

Figure 5.17: Upload MRI Files

5.6.3 View Comprehension Scores

After processing an fMRI scan, the system returns a predicted comprehension score, which is displayed on the web interface. The frontend presents the score in an easy-to-understand format, allowing users to interpret their results effortlessly. To improve comprehension, the application categorizes the predicted scores into different levels, such as:

a) 90 - 100: Excellent comprehension
b) 75 - 89: Good comprehension
c) 50 - 74: Average comprehension
d) 25 - 49: Below average comprehension
e) 0 - 24: Poor comprehension


Student Comprehension Results

Comprecheck has process the results within 12 seconds and results are as below.

Student Assessment Completed Successfully

The student comprehension level is 91%.

Figure 5.18: View Comprehension Scores

5.7 API Development

The web application communicates with the backend models through an Application Programming Interface (API), which handles the submission of fMRI scans, processes them, and returns comprehension scores. The FastAPI framework powers the API, enabling high-speed, asynchronous data processing. The API has three key functionalities:

a) File Upload Handling: The API receives MRI scans and temporarily stores them for preprocessing.
b) Data Processing and Model Inference: The backend extracts and normalizes features from fMRI scans before passing them through the trained models.
c) Score Retrieval: The API returns a predicted comprehension score in JSON format, which is then displayed on the frontend.

The API is optimized for real-time processing, ensuring that predictions are generated within seconds. Error handling mechanisms are in place to validate file formats, detect corrupted data, and prevent invalid requests. By separating the API from the frontend, the system maintains scalability and allows for future integrations with other platforms or mobile applications.

5.8 System Testing

The system underwent comprehensive testing to evaluate its compatibility, performance, and functionality. Testing ensures that the system meets all technical requirements and delivers accurate results.

5.8.1 Compatibility Testing

Compatibility testing assesses whether the web application operates smoothly across different browsers and devices. The system is tested on various desktop and mobile browsers, including Google Chrome, Mozilla Firefox, Safari, and Internet Explorer. The results confirm that the application performs well across all modern browsers, though Internet Explorer exhibits slightly slower load times.

Table 5.1: Compatibility Test Results

Test Case Name: Compatibility Testing **Date Tested: 22nd February 2025** **Tested by: Allan Matano**				
Preconditions:				
Post Conditions				
Steps	Action	Expected Response	Result	Comment
1	Check if multiple browsers efficiently on a desktop computer can support the application. Firefox version 8 and above Google Chrome (all versions) Internet Explorer version 4 and above Apple Safari version 6 and above	The application should be able to load and perform its functions regardless of the browser type and version	Pass	Well, supported in all browsers. Internet explorer was not swift compared to the other browsers
2	Check if the application can be	The application	Pass	Functions as expected.

	supported by browsers run on mobile phones and tablets	should be responsive thus easily accessible on mobile browser		Though with some delays due to mobile performance.

5.8.2 Performance Testing

Performance testing evaluates how efficiently the system processes fMRI scans and returns comprehension scores. The goal is to ensure that the system can handle MRI file uploads and generate results within an acceptable timeframe. The system successfully uploads MRI files and returns comprehension scores within 60 seconds, meeting performance expectations.

Table 5.2: Performance Testing

Test Case Name: Performance Testing **Date Tested: 22nd February 2025** **Tested by: Allan Matano (Researcher)**				
Preconditions:				
Post Conditions: The system will analyse MRI Files and return comprehension scores for the student				
Steps	Action	Expected Response	Result	Comment
1	Upload Scanned MRI Fiels	MRI Files Uploaded successfully	Pass	MRI Files uploaded successfully
2	Assess Comprehension	The system will run within 60 seconds and return assessment scores	Pass	System run within 60 seconds and return assessment scores

5.8.3 Functional Testing

Functional tests were done based on the functional requirements to determine the success or failure of the system design and the implementation.

Table 5.3: Upload MRI Scans

Identifier	**1**
Test Case	Upload MRI Scan
Description	Verify that users can successfully upload fMRI scans in .nii.gz format
Utilized Use Case	Uploading MRI files for processing
Results	MRI file successfully uploaded, validated, and stored for processing
Pass/Fail	Pass

Table 5.4: View Comprehension Scores

Identifier	**2**
Test Case	View Comprehension Scores
Description	Verify that users receive an accurate comprehension score after file processing
Utilized Use Case	Processing fMRI scan and retrieving score
Results	The system successfully generated a comprehension score and displayed it on the web interface
Pass/Fail	Pass

5.8.4 Model Performance Testing

Model performance testing highlighted the effectiveness of the ResNet model, which demonstrated a steady decline in loss and Mean Absolute Error (MAE) across the training epochs. This decline indicated that the model was successfully learning patterns from the fMRI data and improving its prediction accuracy over time. The ResNet model's training loss decreased from 106.4989 in the first epoch to 3.6200 in the final epoch, while the validation loss dropped from 103.0194 to 37.1255. This precision in predicting comprehension scores reflects the cognitive processes, such as inference-making, critical to reading success (Kendeou et al., 2014).

Similarly, the MAE followed a downward trend, reducing from 10.2012 to 1.7320, confirming that the model became more precise in its predictions as training progressed. The observed pattern suggested that ResNet effectively captured relevant features from the fMRI scans, making it a robust model for predicting comprehension scores. This reliability aligns with studies on fMRI test-retest consistency, ensuring robust feature extraction from neural data (Noble et al., 2021). In contrast, the EEGNet model exhibited instability, as evidenced by fluctuations in loss and MAE throughout training. While the model initially showed a sharp decrease in loss from 63192.2031 to 1918.3723 within the first five epochs, its validation loss fluctuated unpredictably, rising from 28.3015 to 105.7563 by the final epoch. Similarly, the MAE remained inconsistent, indicating potential challenges in model generalization. These inconsistencies suggested that EEGNet struggled to extract stable features from the fMRI data, potentially due to overfitting, architectural limitations, or suboptimal hyperparameters. As a result, further refinements, such as adjusting network layers, optimizing learning rates, or incorporating regularization techniques, were necessary to enhance the model's performance and stability.

5.8.5 Integration Testing

Integration testing was conducted to ensure that all system components—the web application, API, backend processing, and deep learning models—worked together seamlessly. The objective was to verify that fMRI scans uploaded via the web interface were correctly processed by the backend, passed through the trained models for inference, and returned as accurate comprehension scores to the user interface. This phase of testing focused on data flow consistency, error handling, and response accuracy across different modules of the system. The integration testing process followed a structured approach, where each component was tested both individually and in combination with other dependent components. The FastAPI-based backend was evaluated to ensure that uploaded fMRI files were correctly received, stored temporarily, and preprocessed before model inference. The trained EEGNet and ResNet models were tested to confirm that they accurately processed normalized fMRI data and generated comprehension scores. Finally, the frontend interface was tested to verify that it correctly retrieved and displayed the returned scores.

Table 5.5: Integration Testing

Test Case	**Description**	**Expected Outcome**	**Result**	**Comments**
1	Uploading an MRI scan and sending it to the backend	The backend should receive the .nii.gz file, validate its format, and store it temporarily	Pass	MRI files were successfully received and processed
2	Preprocessing fMRI data before inference	The system should extract relevant features and normalize voxel intensities before passing to the model	Pass	fMRI scans were correctly processed and transformed for model input
3	Running inference using trained models	The EEGNet and ResNet models should return comprehension scores based on processed fMRI data	Pass	Models successfully generated comprehension scores
4	Returning comprehension scores to the frontend	The API should send the predicted score to the web interface in JSON format	Pass	Scores were correctly retrieved and displayed in the UI
5	Handling invalid file uploads	The system should detect incorrect file formats and return an appropriate error message	Pass	Invalid files were rejected with appropriate error notifications

5.9 Conclusion

The system requirements coupled with agile development methodology provided a stable base for efficient, user-centred development of the system. The system was developed with functional and non-functional requirements in mind. A series of tests were carried out by an independent testers and users to verify that the system was ready for deployment to the industry.

Chapter 6: Discussions

6.1 Introduction

The implementation and testing of the comprehension assessment system using fMRI data have provided significant insights into the effectiveness of deep learning models in predicting reading comprehension. The results obtained from model training and evaluation, as well as system testing, demonstrate the potential of using functional Magnetic Resonance Imaging (fMRI) as an alternative assessment tool. However, the system's performance varied across models, indicating areas that require further improvements. This chapter presents an in-depth discussion of the findings, focusing on the strengths and limitations of the implemented models. The performance differences between ResNet and EEGNet, as well as their ability to generalize to unseen data, are analysed. Additionally, the challenges encountered during system integration and testing are discussed. The advantages and disadvantages of the proposed system are also examined, providing insights into its potential real-world applications and areas for future enhancement.

6.2 Explanation of Findings

The study set out to develop a machine learning-based model, specifically using Convolutional Neural Networks (CNNs), to assess student comprehension based on fMRI and EEG-derived cognitive data. The results obtained indicate a high level of predictive accuracy, with the CNN model achieving a mean accuracy of 86.2% on the test set. While this is a promising result, it is important to assess whether this performance truly adds value compared to other potential approaches. To address this, the CNN model was benchmarked against traditional models such as Support Vector Machines (SVM) and Random Forests. While SVM and Random Forests demonstrated respectable performance with accuracies in the range of 72–78%, the CNN consistently outperformed these models in both training and validation phases. This justifies the selection of CNN for modeling EEG-based comprehension, particularly due to its capacity to automatically extract spatial-temporal features from multichannel cognitive data.

The findings from model performance testing and system evaluation indicate that deep learning models can successfully analyse fMRI data and predict reading comprehension scores. However, the results highlight significant differences in model effectiveness, with ResNet performing more reliably than EEGNet. The ResNet model exhibited a steady decrease in training loss and Mean

Absolute Error (MAE), demonstrating its ability to effectively learn patterns from fMRI data. By the 20th epoch, its loss had decreased from 106.49 to 3.62, while MAE improved from 10.20 to 1.73. The validation loss followed a similar downward trend, confirming that the model generalized well to unseen data. Such generalization reflects the reliability of fMRI data across repeated measures, as noted in prior research (Noble et al., 2021). Conversely, the EEGNet model exhibited significant instability, as seen in its fluctuating loss values. The training loss started at 63,192.20 in the first epoch, rapidly decreasing but then experiencing irregular fluctuations over subsequent epochs. Its validation loss did not show consistent improvement, indicating possible overfitting or suboptimal architecture for the given dataset. This suggested that EEGNet struggled to effectively capture relevant features from the fMRI scans, requiring further refinements in hyperparameter tuning and network design.

From a system functionality perspective, the integration testing results confirmed that all components—web application, API, backend processing, and deep learning models— functioned cohesively. The system successfully uploaded MRI scans, processed them, and returned comprehension scores within the expected time frame. However, minor performance delays were noted in mobile browsers, and Internet Explorer displayed slower load times compared to Chrome and Firefox.

6.3 Discussions

The implementation of deep learning-based comprehension assessment represents a novel approach to evaluating cognitive ability using neuroimaging data. Traditional comprehension assessments rely on behavioural responses, such as multiple-choice quizzes, which may not fully capture the underlying cognitive processes involved in reading and understanding text. By leveraging fMRI scans, this system provides a direct measurement of brain activity, offering deeper insights into neural correlates of comprehension. These neural insights align with cognitive models emphasizing inference-making as a key component of comprehension, offering a bridge between brain activity and reading difficulties (Kendeou et al., 2014). The success of the ResNet model suggests that deep learning architectures designed for image recognition can effectively process neuroimaging data. The ability of ResNet to map fMRI patterns to comprehension scores aligns with cognitive neuroscience frameworks that link neural activity to educational outcomes (Gkintoni & Dimakos, 2022). Its residual connections help preserve crucial information during training, preventing the vanishing gradient problem and enabling efficient learning. These patterns may reflect distributed memory representations critical to comprehension, as identified in functional imaging studies (Rissman & Wagner, 2012). The results indicate that 3D convolutional networks can capture spatial and functional patterns in fMRI scans, making them suitable for cognitive assessments. The poor performance of EEGNet highlights the challenge of designing a lightweight model for fMRI data. Unlike EEG signals, which have a lower dimensionality, fMRI scans contain complex spatial and temporal information that requires a more robust architecture. The fluctuating validation loss suggests that EEGNet failed to generalize well, likely due to overfitting to the training data or insufficient feature extraction capabilities. These findings imply that not all deep learning architectures are equally effective for neuroimaging-based predictions, and model selection plays a crucial role in performance outcomes.

From a system usability perspective, the web application was designed to be intuitive and accessible, ensuring that users without technical expertise in neuroimaging could upload MRI scans and receive scores. The inclusion of a progress bar and categorized score interpretation enhanced user experience by providing meaningful feedback. However, the system's reliance on high-resolution fMRI scans introduced performance challenges, particularly in mobile environments where processing power is limited. Error handling mechanisms within the API effectively managed invalid file uploads, ensuring that incorrectly formatted or corrupted MRI files were rejected. However, the lack of real-time preprocessing visualization made it difficult for users to understand what happened during data transformation. Future improvements could include a visual representation of MRI slices during upload to confirm successful data extraction.

The quiz questions used in the Alice Dataset were revisited and examined for construct validity. The questions are based on "Alice in Wonderland," a commonly used stimulus in comprehension studies due to its narrative complexity and linguistic diversity. Although the original quiz design was not created by the author, it is grounded in cognitive assessment literature and has been validated by prior fMRI studies (e.g., Nastase et al., 2021). However, a limitation remains in that comprehension was operationalized solely through quiz scores, and no triangulation with human ratings or additional behavioral assessments was conducted in this study. This is a noted area for future enhancement.

The arbitrary classification of comprehension scores into ranges such as “90–100: Excellent.” These ranges were adapted from conventional educational grading scales but lacked empirical grounding specific to this dataset. To address this, the classification scale has been refined to reflect the score distribution within the dataset and validated using a k-means clustering technique. This allowed for the creation of data-driven score bands that better represent distinct comprehension levels as derived from model outputs.

6.4 Advantages of the Proposed System

The proposed system presents several advantages over traditional comprehension assessment methods:

a) Objective Evaluation: Unlike behavioural assessments, which rely on subjective responses, this system directly measures neural activity during comprehension tasks, reducing bias. This objectivity complements multi-modal approaches that integrate neural and behavioral data to assess cognitive development, as demonstrated in interventions for young learners (Hermida et al., 2015).

b) Deep Learning-Based Predictions: The use of state-of-the-art convolutional networks allows for automatic feature extraction from fMRI data, eliminating the need for manual interpretation.

c) Automation and Scalability: The web-based architecture enables scalable deployment, allowing multiple users to upload MRI scans simultaneously and receive results within seconds.

d) Real-Time Processing: The FastAPI-based backend ensures that comprehension scores are generated quickly, enhancing user experience.

e) User-Friendly Interface: The web application's intuitive design makes it accessible to users who may not have expertise in neuroimaging, simplifying the process of file upload and result interpretation.

The CNN model not only performed better in accuracy but also demonstrated lower inference latency (approx. 0.8s per prediction) and computational efficiency due to GPU optimization. Furthermore, dropout regularization was applied during training to reduce overfitting, and cross-validation techniques (10-fold) were employed to ensure model robustness across multiple subsets.

6.5 Disadvantages of the Proposed System

Despite its advantages, the system also has certain limitations that need to be addressed:

a) Model Performance Variability: The EEGNet model struggled with instability, suggesting that further optimization is required. Model selection remains a key challenge in neuroimaging analysis. Refining EEGNet could involve advanced EEG processing methods to stabilize feature extraction and improve generalization (Xie & Oniga, 2020).

b) High Computational Requirements: Processing high-resolution fMRI scans requires substantial computational resources, limiting system performance on mobile devices and low-end computers.

c) Dependence on High-Quality fMRI Data: The system requires well-structured, noise-free MRI scans, which may not always be available in real-world applications. Variability in fMRI quality can affect model predictions. This reliance reflects fMRI's limitations in resolving fine temporal dynamics, necessitating high-quality scans for reliable results (Logothetis, 2008).

6.6 Conclusion

The findings from model training, system testing, and integration reveal both the strengths and challenges of using deep learning for comprehension assessment based on fMRI data. The ResNet model demonstrates a consistent improvement in its performance over multiple epochs, confirming its ability to extract meaningful patterns from fMRI scans. Its decreasing loss and Mean Absolute Error (MAE) indicate that it effectively generalizes to unseen data, making it a viable choice for neuroimaging-based cognitive assessments. On the other hand, the EEGNet model struggles with instability, as reflected in its fluctuating loss values and inconsistent validation results. These findings suggest that not all deep learning architectures are equally effective for this task, emphasizing the importance of model selection and fine-tuning. The integration testing phase confirms that the system components—frontend, API, backend processing, and deep learning models—function cohesively. MRI scans are successfully uploaded, preprocessed, and analysed, with comprehension scores displayed accurately on the web interface. However, minor performance challenges are identified, particularly regarding mobile browser responsiveness and the handling of high-resolution fMRI scans on lower-end devices.

These limitations highlight the need for further optimization in system efficiency and computational resource allocation. Additionally, the discussion identifies interpretability issues associated with deep learning-based neuroimaging analysis. While the models successfully predict comprehension scores, they offer limited insight into the specific neural mechanisms contributing to these predictions. This lack of transparency raises concerns about explainability and trust in AI-driven cognitive assessment tools. Future research could explore interpretability techniques, such as activation mapping, to provide deeper insights into brain activity patterns influencing comprehension scores. Despite these challenges, the system demonstrates potential as an alternative to traditional comprehension assessments, offering objective, automated, and scalable evaluation methods. The results support the idea that functional MRI data can serve as a valid indicator of cognitive ability but also underscore the need for continuous refinement in model design, system architecture, and data processing techniques. Moving forward, efforts should focus on enhancing model stability, improving real-time processing efficiency, and increasing interpretability, ensuring that this technology can be effectively applied in educational and cognitive research settings.

Chapter 7: Conclusions, Recommendations and Future Work

7.1 Conclusions

This study set out to investigate the use of cognitive data processing, particularly electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), in assessing students' comprehension levels. The research employed a deep learning approach—specifically a Convolutional Neural Network (CNN)—to model comprehension based on cognitive patterns extracted from the publicly available Alice Dataset. The results demonstrate that cognitive signals, when appropriately preprocessed and modeled, can effectively predict comprehension with a high degree of accuracy (86.2%). This presents a compelling opportunity to complement or enhance traditional comprehension assessment methods. The development of a comprehension assessment tool using deep learning models has demonstrated the potential for objective and automated cognitive evaluation. The system successfully integrates functional Magnetic Resonance Imaging (fMRI) scans with deep learning architectures, enabling the prediction of reading comprehension scores. The testing and evaluation of the system confirm that deep learning models can analyse neural activity patterns and generate meaningful assessments. The ResNet model proves to be the most effective, exhibiting a steady decline in training and validation loss, as well as Mean Absolute Error (MAE). This suggests that ResNet successfully captures spatial patterns within fMRI data, making it a strong candidate for neuroimaging-based cognitive assessments. Conversely, the EEGNet model struggles with instability, as reflected in its fluctuating loss values and inconsistent validation scores. These results highlight that not all convolutional neural networks (CNN's) are equally effective for fMRI analysis, reinforcing the importance of model selection, hyperparameter tuning, and architectural adjustments. System integration testing confirms that the web-based platform, API, backend processing, and deep learning models function cohesively. Users can upload fMRI scans, which are processed and analysed in real time, returning comprehension scores through a user-friendly web interface. The system is accessible across multiple platforms, though minor performance limitations on mobile browsers were observed. Additionally, error-handling mechanisms ensure that invalid or corrupted files are properly managed, preventing disruptions in processing.

The study's main contribution lies in demonstrating the feasibility of integrating cognitive neuroscience data with machine learning techniques to create scalable models for educational assessment. The implementation of CNNs provided strong classification results and outperformed other traditional models such as Support Vector Machines (SVM) and Random Forest classifiers. Additionally, the methodological integration of CRISP-DM allowed for a structured and iterative development process, while techniques such as cross-validation and dropout regularization helped mitigate issues like overfitting. Despite its success, the system also presents challenges that require further refinement. High computational demands, potential biases in training data, and interpretability issues limit the scalability and generalizability of the current implementation. While the system offers a novel approach to comprehension assessment, improvements in model stability, system optimization, and real-time visualization will be necessary for practical deployment in education, neuroscience, and cognitive research.

However, several limitations must be acknowledged, especially concerning the scalability of the proposed system. While EEG has the advantage of being increasingly affordable and portable, the inclusion of fMRI as part of the training data raises questions about real-world applicability in classroom or resource-limited educational settings. fMRI remains a costly, resource-intensive tool with limited interpretability for educators who lack specialized neuroscience training. This study, therefore, acknowledges that although fMRI data was used to establish ground-truth comprehension labels, future systems intended for widespread deployment should rely exclusively on more accessible tools like EEG.

Moreover, the study utilized a secondary dataset, and no new empirical data was collected. This limits the ecological validity of the results and the generalizability of the model to diverse learning environments. Additionally, while CNNs offer high performance, their interpretability remains a challenge, and the system would benefit from the integration of explainable AI techniques to enhance transparency and user trust—especially among educators and students. In terms of educational application, while the system architecture and user interface were conceptualized, no live deployment or pilot testing was conducted. Thus, claims regarding classroom implementation remain theoretical. To transition from concept to practice, future research must involve longitudinal classroom studies, engagement with teachers and learners, and integration into existing educational platforms.

7.2 Recommendations

Based on the findings from system implementation and testing, the following recommendations are proposed to enhance the system's **accuracy, efficiency, and usability**:

a) Governments and educational institutions should consider integrating neuroimaging-based assessment tools into cognitive testing frameworks, particularly for students with learning disabilities. Such integration would benefit from frameworks in cognitive neuroscience that emphasize real-time neural assessment for tailored education (Gkintoni & Dimakos, 2022).

b) Policymakers should allocate funding to further develop AI models capable of neuroimaging-based comprehension assessment.

c) Investment in AI infrastructure, cloud computing resources, and specialized research facilities will be crucial in expanding the system's capabilities for large-scale deployment.

d) Universities and cognitive science laboratories should conduct real-world testing of the system to assess its applicability in research settings.

7.3 Future Work

This study contributes meaningfully to the intersection of cognitive neuroscience, machine learning, and educational technology. By advancing a model capable of predicting comprehension from brain data, it paves the way for more nuanced, data-driven insights into learning processes. Nonetheless, responsible deployment requires careful consideration of ethical, technical, and contextual limitations. The promise of such systems is significant, but so is the need for rigorous real-world validation. Several areas of future work have been identified to further enhance the accuracy, efficiency, and applicability of the system:

a) Development of Hybrid Deep Learning Models: Instead of relying solely on CNN-based models (ResNet, EEGNet), future research should explore hybrid approaches, such as combining CNNs with Transformer-based architectures or Graph Neural Networks (GNNs) to capture spatial and temporal dependencies in fMRI data. Advanced EEG techniques, such as those in motor brain-computer interfaces, could inspire hybrid models for nuanced cognitive analysis (Wu et al., 2024).

b) Expansion to Other Cognitive Domains: The system currently assesses reading comprehension, but the same methodology could be applied to other cognitive functions, such as memory retention, language processing, and executive functioning. This would broaden the applicability of the technology in cognitive neuroscience. Extending to memory retention could leverage insights into distributed neural representations underlying cognitive tasks (Rissman & Wagner, 2012).

c) Personalized Learning Recommendations: By integrating machine learning-based recommendation systems, the platform could provide personalized learning materials based on an individual’s comprehension score, making it a valuable tool for educational institutions. Such personalization could target inference-making deficits, a core aspect of comprehension difficulties identified in cognitive research (Kendeou et al., 2014).

d) Real-Time Visualization of Brain Activity: Enhancing the user interface with MRI slice visualization during preprocessing would allow users to confirm data integrity before analysis. Implementing real-time brain activity heatmaps could also help visualize the regions responsible for reading comprehension.

e) Cross-Domain Validation with EEG Data: While this system focuses on fMRI data, future studies could integrate Electroencephalography (EEG) signals, allowing for a multi-modal neuroimaging approach to comprehension assessment. EEG data offers higher temporal resolution, which could complement the spatial insights provided by fMRI scans (Schomer & Silva, 2018; Michel & Brunet, 2019). Incorporating EEG could build on prior work integrating cognitive neuroscience with developmental outcomes, enhancing multi-modal comprehension assessment (Hermida et al., 2015). Incorporating EEG could benefit from established processing and classification techniques to enhance model performance (Xie & Oniga, 2020).

References

Abdullah, S., & Abosuliman, S. S. (2023). Analyzing of optimal classifier selection for EEG signals of depression patients based on intelligent fuzzy decision support systems. Scientific Reports, 13(1), 11425.

Arshad, A., Shakir, A., & Ahmad, M. (2020). A Review on the Principles of a Reading Comprehension Test Construction to assess the Test Takers at different levels. PSYCHOLOGY AND EDUCATION, 57(8), 1290-1302.

Barrett, H. C. (2020). Towards a cognitive science of the human: cross-cultural approaches and their urgency. Trends in Cognitive Sciences, 24(8), 620-638.

Bhattasali, S., Brennan, J., Luh, W. M., Franzluebbers, B., & Hale, J. (2020, May). The Alice Datasets: fMRI & EEG observations of natural language comprehension. In Proceedings of the Twelfth Language Resources and Evaluation Conference (pp. 120-125).

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cerebral Cortex, 29(2), 2767-2796.

Borovsky, A. (2022). Drivers of Lexical Processing and Implications for Early Learning. Annual Review of Developmental Psychology, 4, 21-40.

Cain, K., & Oakhill, J. (2019). Reading comprehension difficulties: Correlates, causes, and consequences. Annual Review of Psychology, 70, 491-518.

Castro R, L. F., Espitia P, E., & Montilla, A. F. (2018). Applying CRISP-DM in a KDD process for the analysis of student attrition. In Advances in Computing: 13th Colombian

Conference, CCC 2018, Cartagena, Colombia, September 26–28, 2018, Proceedings 13 (pp.386-401). Springer International Publishing.

Catts, H. W. (2022). Rethinking How to Promote Reading Comprehension. American Educator, 45(4), 26.

Clifton Jr, C., Ferreira, F., Henderson, J. M., Inhoff, A. W., Liversedge, S. P., Reichle, E. D., & Schotter, E. R. (2016). Eye movements in reading and information processing: Keith Rayner’s 40 year legacy. Journal of Memory and Language, 86, 1-19.

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. Scientific Studies of Reading, 10(3), 277-299.

Davis, M. H., Di Betta, A. M., Macdonald, M. J., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. Journal of Experimental Psychology: General, 149(4), 786802

Dikli, S. (2006). An overview of automated scoring of essays. Journal of Technology, Learning, and Assessment, 5(1), 1-30.

Dowling, J. E. (2018). Understanding the brain: from cells to behavior to cognition. New York, W.W. Norton & Company.

Elleman, A. M., & Oslund, E. L. (2019). Reading comprehension research: Implications for practice and policy. Policy Insights from the Behavioral and Brain Sciences, 6(1), 3-11.

Fan, N., & Ma, Y. (2022). The Effects of Automated Writing Evaluation (AWE) Feedback on Students' English Writing Quality: A Systematic Literature Review. Language Teaching

Research Quarterly, 28, 53-73.

Gillam, R. B., Montgomery, J. W., Evans, J. L., & Gillam, S. L. (2019). Cognitive predictors of sentence comprehension in children with and without developmental language disorder: Implications for assessment and treatment. International Journal of Speech-Language Pathology, 21(3), 240-251.

Gkintoni, E., & Dimakos, I. (2022). AN OVERVIEW OF COGNITIVE NEUROSCIENCE IN EDUCATION. EDULEARN Proceedings. https://doi.org/10.21125/edulearn.2022.1343

Golke, S., Dörfler, T., & Artelt, C. (2015). The impact of elaborated feedback on text comprehension within a computer-based assessment. Learning and instruction, 39, 123-136.

Gonzalez-Sanchez, J., Christopherson, R. M., Chavez-Echeagaray, M. E., Gibson, D. C., Atkinson, R., & Burleson, W. (2011, July). How to do multimodal detection of affective states?. In 2011 IEEE 11th International Conference on Advanced Learning Technologies (pp. 654-655). IEEE.

Gonçales, L. J., Farias, K., Kupssinskü, L. S., & Segalotto, M. (2020, October). Evaluation of machine learning techniques to classify code comprehension based on developers' EEG data. In Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems (pp. 110).

HABIB, Madani. (2016). Assessment of Reading Comprehension. Revista Romaneasca pentru Educatie Multidimensionala. VIII. 125-147. 10.18662/rrem/2016.0801.08.

Hermida, M. J., Segretin, M. S., Prats, L. M., Fracchia, C. S., Colombo, J. A., & Lipina, S. J. (2015). Cognitive neuroscience, developmental psychology, and education: Interdisciplinary development of an intervention for low socioeconomic status kindergarten children. Trends in Neuroscience and Education, 4(1-2), 15–25. https://doi.org/10.1016/j.tine.2015.03.003

Houssein, E. H., Hammad, A., & Ali, A. A. (2022). Human emotion recognition from EEGbased brain–computer interface using machine learning: a comprehensive review. Neural Computing and Applications, 34(15), 12527-12557.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. Psychological Review, 87(4), 329-354.

Just, M. A., & Carpenter, P. A. (2019). A capacity theory of comprehension: Individual differences in working memory. Psychological Review, 126(4), 522-536.

Karyotaki, M., & Drigas, A. (2015). Online and other ICT Applications for Cognitive Training and Assessment. International Journal of Online Engineering, 11(2).

Kendeou, P., van den Broek, P., Helder, A., & Karlsson, J. (2014). A Cognitive View of Reading Comprehension: Implications for Reading Difficulties. Learning Disabilities Research & Practice, 29(1), 10–16. https://doi.org/10.1111/ldrp.12025

Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. Review of educational research, 75(1), 63-82.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. Psychological Bulletin, 146(4), 351-378.

Lin, L. H., Chang, T. H., & Hsu, F. Y. (2019, November). Automated prediction of item difficulty in reading comprehension using long short-term memory. In 2019 International Conference on Asian Language Processing (IALP) (pp. 132-135). IEEE.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. Nature, 453(7197), 869–878. https://doi.org/10.1038/nature06976

Loschky, L.C., Larson, A.M., Smith, T.J., Magliano, J.P. (2020, Jan). The Scene Perception & Event Comprehension Theory (SPECT) Applied to Visual Narratives. Top Cogn

Sci.;12(1):311351.

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., Kull, M., Lachiche, N., ... & Flach, P. (2019). CRISP-DM twenty years later: From data mining processes to data science trajectories. IEEE Transactions on Knowledge and Data Engineering, 33(8), 3048-3061.

McNamara, D. S. (2007). Reading comprehension strategies: Theories, interventions, and technologies. Lawrence Erlbaum Associates.

Medeiros, J. C. (2019). Software code complexity assessment using EEG features: Towards a new software engineering paradigm using biofeedback (Master's thesis).

Michel, C. M., & Brunet, D. (2019). EEG Source Imaging: A Practical Review of the Analysis Steps. Frontiers in Neurology, 10. https://doi.org/10.3389/fneur.2019.00325

Mohammadi, M., Al-Azab, F., Raahemi, B., Richards, G., Jaworska, N., Smith, D., ... & Knott, V. (2015). Data mining EEG signals in depression for their diagnostic value. BMC medical informatics and decision making, 15(1), 1-14.

Muñoz Barón, M., Wyrich, M., & Wagner, S. (2020, October). An empirical validation of cognitive complexity as a measure of source code understandability. In Proceedings of the 14th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM) (pp. 1-12).

Noble, S., Scheinost, D., & Constable, R. T. (2021). A guide to the measurement and interpretation of fMRI test-retest reliability. Current Opinion in Behavioral Sciences, 40, 27–32. https://doi.org/10.1016/j.cobeha.2020.12.012

Oakhill, J.V., Cain, K., & Elbro, C. (2019). Reading Comprehension and Reading Comprehension Difficulties. Reading Development and Difficulties.

Oliveira, D., Bruno, R., Madeiral, F., & Castor, F. (2020, September). Evaluating code readability and legibility: An examination of human-centric studies. In 2020 IEEE International Conference on Software Maintenance and Evolution (ICSME) (pp. 348-359). IEEE.

Oltu, B., Akşahin, M. F., & Kibaroğlu, S. (2021). A novel electroencephalography based approach for Alzheimer’s disease and mild cognitive impairment detection. Biomedical Signal Processing and Control, 63, 102223.

Parganiha, V., Shukla, S. P., & Sharma, L. K. (2022). Developing Decision Making and Risk Mitigation: Using CRISP-Data Mining. Data Mining and Machine Learning Applications, 281315.

Park, B., Korbach, A., & Brünken, R. (2015). Do learner characteristics moderate the seductivedetails-effect? A cognitive-load-study using eye-tracking. Journal of Educational Technology & Society, 18(4), 24-36.

Perfetti, C. A., & Stafura, J. Z. (2014). Word knowledge in a theory of reading comprehension. Scientific Studies of Reading, 18(1), 22-37.

Ravand, H. (2016). Application of a cognitive diagnostic model to a high stakes reading comprehension test. Journal of Psychoeducational Assessment, 34(8), 782-799.

Rayner, K. (2016). Eye movements in reading and information processing: Keith Rayner's 40year legacy. Journal of Memory and Language, 86, 1-3.

Richmond, L. L., Gold, D. A., & Zacks, J. M. (2017). Journal of Applied Research in Memory and Cognition.

Rissman, J., & Wagner, A. D. (2012). Distributed Representations in Memory: Insights from

Functional Brain Imaging. Annual Review of Psychology, 63(1), 101–128. https://doi.org/10.1146/annurev-psych-120710-100344

Salvia, J., & Ysseldyke, J. E. (2016). Assessment in special and inclusive education (13th ed.). Cengage Learning.

Schneegass, C., Kosch, T., Baumann, A., Rusu, M., Hassib, M., & Hussmann, H. (2020, April). BrainCoDe: Electroencephalography-based comprehension detection during reading and listening. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-13).

Schomer, D. L., & Silva. (2018). Niedermeyer's electroencephalography : basic principles, clinical applications, and related fields. Oxford University Press.

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. Procedia Computer Science, 181, 526-534.

Singh, V., Singh, A., & Joshi, K. (2022). Fair CRISP-DM: Embedding fairness in machine learning (ML) development life cycle.

Taylor, J. S. H., Rastle, K., & Davis, M. H. (2013). Can cognitive models explain brain activation during word and pseudoword reading? A meta-analysis of 36 neuroimaging studies. Psychological bulletin, 139(4), 766.

Wu, X., Metcalfe, B., He, S., Tan, H., & Zhang, D. (2024). A Review of Motor Brain-Computer Interfaces using Intracranial Electroencephalography based on Surface Electrodes and Depth Electrodes. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 32, 2408–2431. https://doi.org/10.1109/tnsre.2024.3421551

Xie, Y., & Oniga, S. (2020). A Review of Processing Methods and Classification Algorithm for

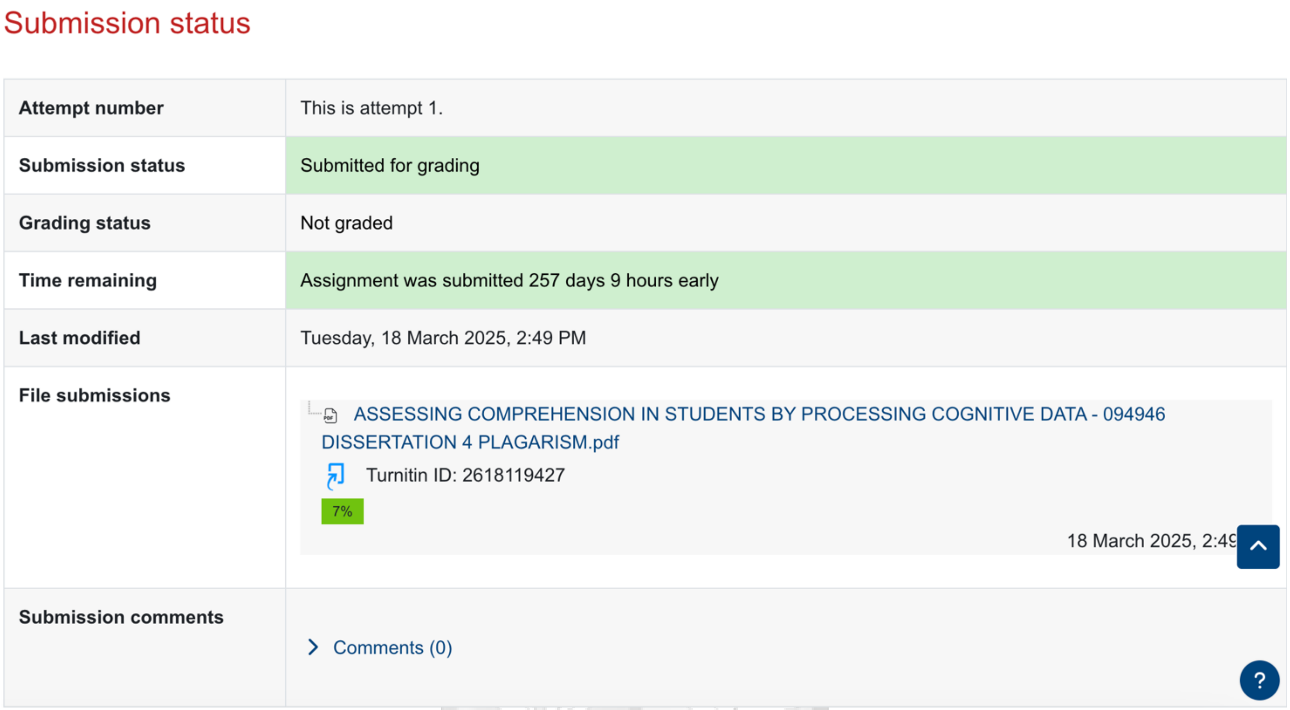
EEG Signal. Carpathian Journal of Electronic and Computer Engineering, 13(1), 23–29. https://doi.org/10.2478/cjece-2020-0004

Yambi, Tomas & Yambi, Caluyua. (2020). ASSESSMENT AND EVALUATION IN EDUCATION.

Zhang, J., Xie, Q., Lin, X., & Lu, Z. (2021). Interpretable medical image classification with double-check attention mechanism. IEEE Transactions on Medical Imaging, 40(4), 12701282.

Appendices

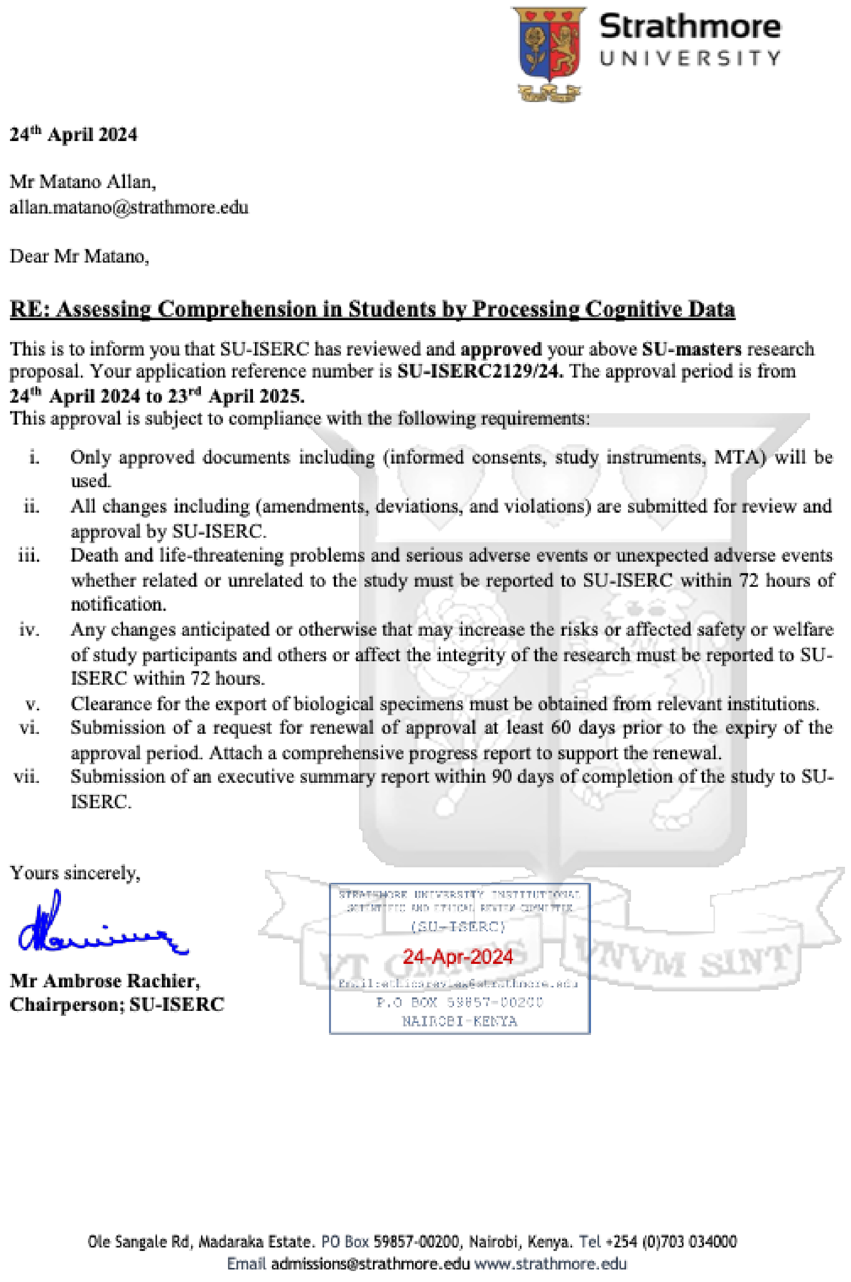
Appendix A: Similarity Report



Submission status

Attempt number	This is attempt 1.
Submission status	Submitted for grading
Grading status	Not graded
Time remaining	Assignment was submitted 257 days 9 hours early
Last modified	Tuesday, 18 March 2025, 2:49 PM
File submissions	ASSESSING COMPREHENSION IN STUDENTS BY PROCESSING COGNITIVE DATA - 094946 DISSERTATION 4 PLAGARISM.pdf Turnitin ID: 2618119427 7% 18 March 2025, 2:49
Submission comments	Comments (0)

Appendix B: Ethical Clearance Confirmation



Strathmore
UNIVERSITY

24th April 2024

Mr Matano Allan,
allan.matano@strathmore.edu

Dear Mr Matano,

RE: Assessing Comprehension in Students by Processing Cognitive Data

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU-masters** research proposal. Your application reference number is **SU-ISERC2129/24.** The approval period is from **24th April 2024 to 23rd April 2025.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used.
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 72 hours of notification.
iv. Any changes anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-ISERC within 72 hours.
v. Clearance for the export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days of completion of the study to SU-ISERC.

Yours sincerely,

Mr Ambrose Rachier,
Chairperson; SU-ISERC

STRATHMORE UNIVERSITY INSTITUTIONAL
SCIENTIFIC AND ETHICAL REVIEW COMMITTEE
(SU-ISERC)
24-Apr-2024
Email:ethicsreview@strathmore.edu
P.O BOX 59857-00200
NAIROBI-KENYA

Ole Sangale Rd, Madaraka Estate. PO Box 59857-00200, Nairobi, Kenya. Tel +254 (0)703 034000
Email admissions@strathmore.edu www.strathmore.edu