fDecoding Morse Code via Eye Blinks using CNN and LSTM Machine learning Algorithms

A Thesis Project Presented to the

Faculty of the College of Computing Studies

Eulogio "Amang" Rodriguez Institute of

Science and Technology Nagtahan, Sampaloc Manila.

In Partial Fulfillment

of the Requirements for the Degree of

Bachelor of Science in Computer Science

By:

Cruz, Josh Harold H.

Narciso, Victor Price

Palagam, Erika Van Imeren A.

Rosales, Aries P.

2025

CHAPTER 1

THE PROBLEM AND ITS BACKGROUND

Introduction

Communication is a fundamental aspect of human life, essential for expressing feelings, needs, and building relationships. However, individuals facing speech and movement challenges encounter significant hurdles in interacting with their environment. According to the World Report on Disability from the World Health Organization, the global population aged 60 and over is expected to double from 11% to 22% by 2050, with an estimated 15% of people worldwide living with disabilities. Severe disabilities, such as quadriplegia, limit individuals to basic functions like blinking or moving their cheeks, confining them to a bed for most of the year. They often find themselves without the ability to move independently or communicate effectively.

For many individuals with severe disabilities, eye blinks provide one of the few remaining voluntary movements, making blink-based interfaces a promising means of transmitting information. However, traditional blink-based systems often fall short due to their reliance on simple blink duration thresholds or limited switch-based encoding schemes. These systems can struggle to perform reliably in real-world conditions, subject to variations in lighting, camera angles, and user fatigue. By focusing on enhancing the robustness of these interfaces, the study can significantly improve communication opportunities for those facing severe disabilities.

To address these limitations, this thesis proposes a modular deep-learning pipeline designed to translate intentional eye blinks into Morse code and then into textual output. By leveraging the strengths of convolutional neural networks (CNNs) for spatial eye-state classification and long short-term memory networks (LSTMs) for temporal blink-event modeling, the research approach significantly enhances reliability. A CNN processes video frames captured via a webcam or pre-recorded footage to accurately classify the eye as “open” or “closed,” even under challenging visual conditions. These frame-level classifications are then fed into an LSTM, which recognizes the temporal pattern of a complete blink (open → closed → open), measures its duration, and categorizes it as a Morse “dot” or “dash.” Finally, a rule-based module compiles these symbols into letters and words, decoding them against the International Morse alphabet for display as on-screen text or through synthesized speech.

Background of the Study

Manual eye-blink interpretation systems, often involving time-based thresholding or mechanical switches, have long posed limitations in accurately translating user intention into meaningful communication, particularly for individuals with severe motor impairments. These traditional approaches typically rely on rudimentary signal detection, such as prolonged closures to indicate a “dash” or brief ones for a “dot” without contextual awareness of temporal patterns or spatial noise.

With the advent of deep learning frameworks, particularly convolutional neural networks (CNNs) and recurrent architectures like long short-term memory (LSTM), there is now the opportunity to construct intelligent, context-aware systems that not only detect blinks with spatial precision but also understand their temporal dynamics for semantic interpretation. However, as noted by Xiong et. al. (2025) in A Review of Deep Learning Blink Detection, many implementations either employ CNNs in isolation, focusing solely on frame-based classification without accounting for sequence continuity, or utilize LSTMs without robust spatial feature anchoring, leading to temporal misclassification.

Studies indicate that hybrid CNN-LSTM architectures have yielded promising results in multimodal decoding environments, such as lip-reading and sign language interpretation. For instance, Deocampo et. al. (2023) proposed a hybrid CNN-LSTM-based visual decoding technique with independent video preprocessing for lip-reading in Tagalog, demonstrating the effectiveness of combining spatial and temporal features. These results suggest that a modular pipeline combining spatial and temporal learning can achieve superior accuracy in gesture-based communication systems. However, little attention has been given to using this hybrid architecture for Morse code decoding via eye blinks, particularly in real-time, webcam-based setups optimized for individuals with quadriplegia or similar conditions.

This thesis addresses these gaps by developing and evaluating a deep learning pipeline that translates intentional eye blinks into Morse code using a CNN for eye-state classification and an LSTM for blink-event recognition. By integrating MediaPipe's face-mesh detection for precise feature localization and leveraging a rule-based decoder to interpret dot/dash sequences, the system aims to produce reliable, latency-conscious textual output. In doing so, it reduces reliance on hardware-intensive setups, avoids heuristic threshold tuning, and brings low-cost, scalable communication tools closer to individuals in need. The proposed architecture also contributes to the broader field of assistive AI by offering a generalizable framework that can be adapted to diverse non-verbal communication contexts.

Theoretical Framework

Eye-blink detection is a computer vision task that identifies voluntary eye closures from video input. According to Xiong et al., 2025, this capability has become an important aspect of human-computer interaction (HCI) and assistive technologies . In various applications from Ezzat et al., 2023 study, they found that in cases similar to fatigue monitoring and communication aids, detecting blinks offers a non-invasive and intuitive means of communication for users with severe motor impairments .

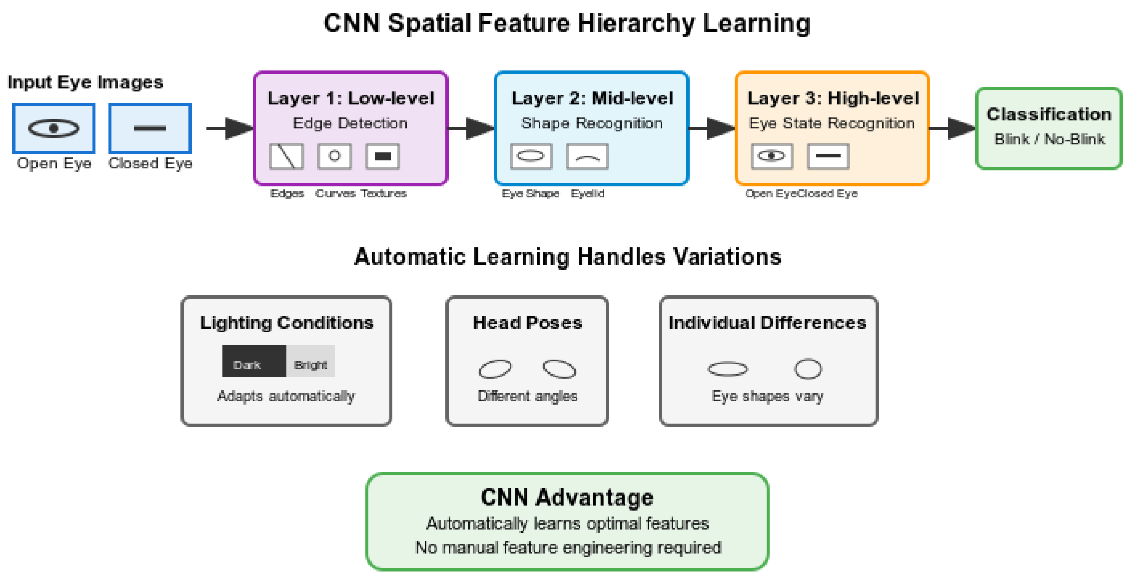

Figure 1.1: CNN Learning Hierarchy for distinguishing Open and Closed Eyelids

Convolutional Neural Networks (CNNs) are particularly well-suited for this visual recognition task, as they automatically learn spatial feature hierarchies from image grids. A CNN can be trained to extract distinguishing features of the eye region, such as differentiating between open and closed eyelids in each frame. This results in highly accurate blink detection, even under varying lighting conditions or head pose. Deep CNNs have been shown to outperform traditional heuristic methods due to their ability to adaptively learn features, achieving superior accuracy in blink recognition. Therefore, a CNN front-end provides robust frame-by-frame eye-state classification (blink vs. no-blink), which is essential for effective downstream decoding.

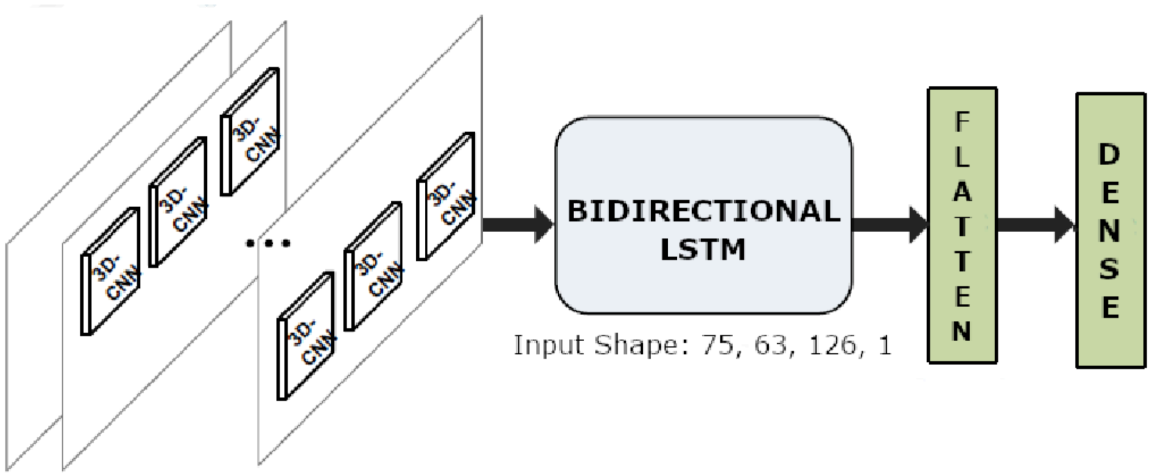

Figure 1.2: CNN and LSTM Connection Diagram. Adapted from De Ocampo et al.

Long Short-Term Memory (LSTM) networks, a specialized form of recurrent neural network, are particularly well-suited for tasks that require the handling of sequences. LSTMs address the vanishing gradient problem found in standard recurrent neural networks (RNNs) and effectively retain information over extended sequences, making them relatively insensitive to the length of gaps in data and capable of learning long-term dependencies. In practice, the LSTM processes outputs from a Convolutional Neural Network (CNN)—specifically, a time series of blink events and their durations—to identify temporal patterns that correspond to Morse code “dots” and “dashes.” In this system, the CNN functions as a spatial feature extractor for each video frame, while the LSTM layer is responsible for sequence prediction necessary for decoding Morse symbols.

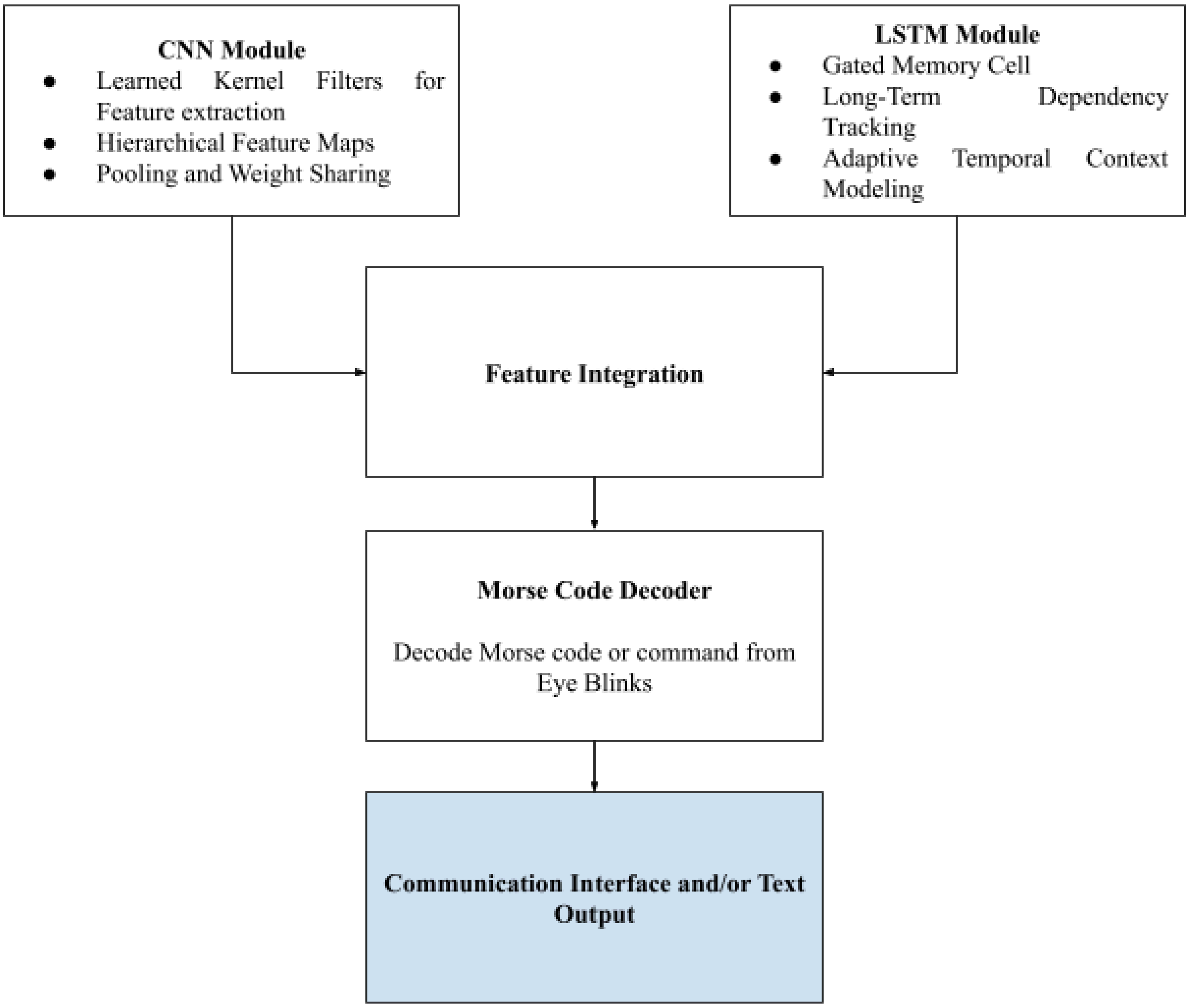

Figure 1.3: The System's Theoretical Diagram

This hybrid CNN-LSTM architecture has shown versatility across various domains; for instance, hybrid models that integrate CNN layers for feature extraction with LSTM layers for sequence interpretation have been applied in tasks such as activity recognition and image captioning. By modeling both the spatial characteristics of blinks and their temporal sequence, the CNN with an LSTM framework aligns seamlessly with the objectives of Morse code decoding.

Prior research demonstrates that CNN with LSTM systems can achieve remarkable accuracy; for example, a CNN with an LSTM model attained an accuracy of 94.43% in pre-miRNA classification, as reported by Koochaki et al. In assistive interfaces, reliable gaze and eye-tracking, as well as blink mapping, are crucial. The capacity to accurately estimate gaze direction, detect blinks, and convert eye movements into communication commands is vital for the functionality of virtual keyboards and real-time assistive devices, as demonstrated in a study by Ezzat et al. (2023). In this context, the CNN guarantees robust blink detection even under noisy or variable conditions, while the LSTM ensures precise temporal decoding.

Conceptual Framework

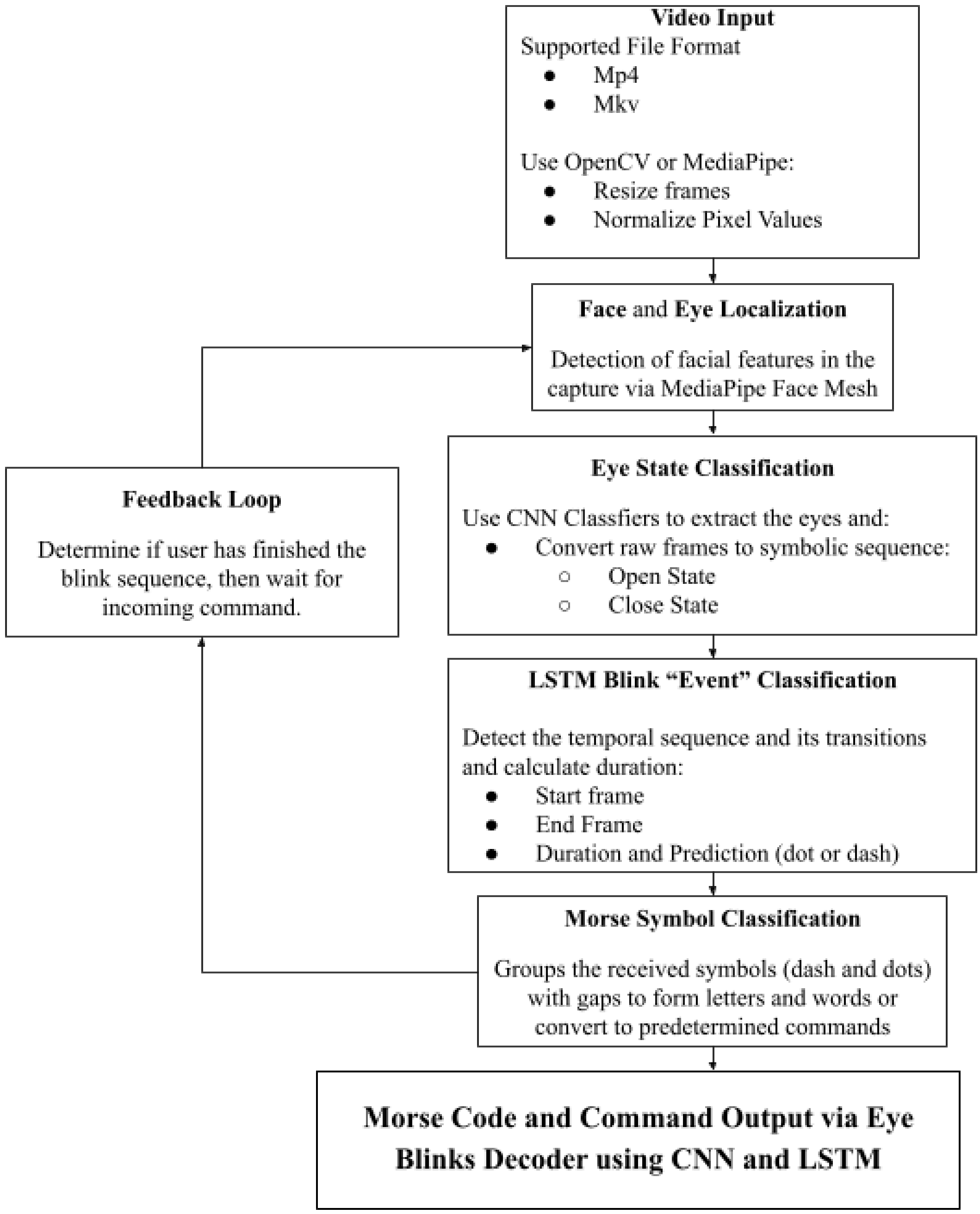

Figure 1.4: Block Diagram of the Proposed Study

The study uses a face-mesh algorithm (e.g., MediaPipe Face Mesh) to locate and track the user's eye regions in each incoming video frame, providing spatially aligned crops for downstream analysis. These crops feed into a convolutional neural network (CNN), which has been trained to distinguish "open" versus "closed" eye states on a per-frame basis; its learned convolutional filters encode robust eyelid features that generalize across lighting variations and head-pose changes.

Next, the CNN's binary state outputs are buffered into a temporal sequence and passed to a long short-term memory (LSTM) network. The LSTM leverages its gated recurrent architecture to recognize the characteristic "open → closed → open" transition of a blink, measure its duration, and classify each blink as a Morse "dot" or "dash." In doing so, it inherently learns the temporal thresholds and contextual dependencies that differentiate intentional blinks from involuntary closures or noise. Finally, a rule-based decoder groups the resulting dot/dash symbols according to standard inter-symbol and inter-word timing conventions, mapping them through a Morse-code lookup to alphanumeric characters or a different lookup table with a pre-determined set of sentences and commands. This modular CNN → LSTM → decoder pipeline thus transforms raw video into reliable textual output, balancing spatial feature learning with temporal sequence interpretation.

Statement of the Problem

This study aims to answer the following questions:

1. How can the CNN-LSTM model accurately detect the eye region and classify its state (open vs. closed) across varied lighting, head poses, and individuals via precision, recall, and F1-score?
2. What timing thresholds for dots, dashes, and inter-blink gaps does the LSTM learn, and how consistent are these thresholds over time or under changing conditions?
3. What is the system's overall symbol error rate and word-level decoding accuracy when converting a stream of blinks to text or commands, and how does latency impact real-time usability?
4. What would be an effective set of Morse code to text-to-speech commands that could assist the user in their day-to-day life?

Significance of the Study

By demonstrating a low-cost, webcam-based interface that converts simple eye blinks into readable text, this work raises awareness of how everyday technologies can be repurposed for inclusive communication. It underscores the broader societal value of computer vision and deep learning, showing that innovations originally developed for entertainment or surveillance can also empower users with diverse abilities. Specifically for those who are:

1. **Individuals with Communication Impairments:** For nonverbal or speech-impaired users, this system offers an accessible channel for expression. Rather than relying on expensive eye-tracking hardware or laborious switch-based inputs, a familiar webcam and natural blinking become a conduit to full alphabetic communication. This enhances autonomy, reduces frustration, and can facilitate richer social interaction.
2. **Caregivers:** Caregivers benefit from faster, more reliable understanding of the user's needs and intentions. This adds another way of interpretation to receive clear, real-time textual (or spoken) feedback, which can streamline daily routines, medical decision-making, and emotional support. The modular design also allows caregivers to calibrate and adjust system sensitivity to individual blink patterns.
3. **The broader Academic Community:** This modular CNN and LSTM pipeline adds to the benchmarks for a clear methodological blueprint for subsequent work in assistive communication. Researchers can extend or replace individual modules by experimenting with alternative CNN architectures, attention-based sequence models, or end-to-end CTC approaches and directly compare performance metrics (e.g., blink-event detection rates, decoding accuracy) against the published baseline. This fosters rapid iteration and continual improvement in blink-based interfaces.

Scope and Delimitation of the Study

The development, testing, and evaluation were carried out in Metro Manila, Philippines, in the academic year 2024 - 2025, utilizing publicly available datasets sourced from online and locally sourced datasets for model training and validation. The proposed system features a real-time or pre-recorded video, camera-based communication by translating intentional eye blinks into alphanumeric text through a machine-learning pipeline. It employs a lightweight face-mesh detector to localize and crop the user's eyes in each video frame and a convolutional neural network (CNN) to classify the eye state (open vs. closed) with high accuracy under varied lighting and moderate head movement. For better reliability, the system buffers these binary classifications into a temporal sequence and applies a long short-term memory (LSTM) network to detect complete blink events (open → closed → open), automatically measure their durations, and distinguish between "dots" and "dashes" without manual parameter tuning. The system then groups the resulting Morse symbols into letters and words according to standardized inter-symbol and inter-word timing conventions and then performs a direct lookup to produce readable text. Finally, it can optionally synthesize the decoded text into speech for auditory feedback. Collectively, these modular components operate end-to-end on commodity RGB cameras, offering user-calibrated, adaptive blink-to-text translation suitable for individuals with severe motor impairments.

The system will not be able to process and understand commands and sequences outside the known International Morse Code Standard. The system also does not cover commands that involve commands that involve the direction of the eye (e.g., whether the user is looking left or right).

Objective of the Study

1. Measure how well the CNN–LSTM detects eyes and classifies open or closed states under different lighting, poses, and users using precision, recall, and F1-score.
2. Find the dot, dash, and blink-gap times the LSTM learns and check if they stay steady over time and conditions.

3. Assess symbol error rate, word accuracy, and latency when decoding blinks, and see how latency affects real-time use.
4. Create and test simple Morse-to-speech commands for everyday tasks, focusing on speed, accuracy, and user comfort.

Definition of Terms

Assistive Communication Technology. A category of tools designed to support individuals with communication impairments, such as systems that enable blink-to-text translation.

Blink Event. A voluntary eye movement involving a full open–closed–open transition used as the primary input gesture for communication in the system.

CNN (Convolutional Neural Network). A type of deep learning model designed for spatial pattern recognition is used to classify the eye state (open or closed) in each frame.

Dot and Dash Classification. The process of labeling detected blinks as short-duration (dot) or long-duration (dash) events based on temporal characteristics.

Eye Aspect Ratio (EAR). A computed ratio based on eye landmarks is used to detect blink events by measuring eyelid openness in real time.

Frame-Level Classification. The process of evaluating each video frame individually to determine the user's eye state without considering previous or future frames.

Latency. The time delay between the user's blink and the system's output, crucial for assessing real-time communication performance.

LSTM (Long Short-Term Memory). A recurrent neural network (RNN) architecture capable of learning time-dependent patterns for identifying blink sequences and classifying them as Morse code dots or dashes.

MediaPipe Face Mesh. A real-time face tracking model used to detect and align facial landmarks, particularly the eye region, for consistent input to the CNN.

Morse Code. A character encoding system that represents letters and numbers using short (dots) and long (dashes) signals, adapted here for blink-based **communication.**

OpenCV (Open Source Computer Vision Library). A computer vision library used for video capture, image processing, and integration of the webcam feed into the blink detection pipeline.

Quasi-Experimental Design. A research approach that assesses intervention outcomes without random participant assignment, suitable for real-world system evaluation.

Rule-Based Decoder. A logic-based system that translates dot and dash sequences into alphanumeric characters using standard Morse code conventions.

Symbol Error Rate (SER). A metric that measures the accuracy of decoded Morse code symbols by comparing predicted sequences to the ground truth.

Temporal Sequence Modeling. The use of sequential data analysis techniques to understand blink timing patterns over time allows for accurate Morse code interpretation.

CHAPTER 2

REVIEW OF RELATED LITERATURE AND STUDIES

Review of Related Literature

This section presents a review of the foreign literature, sourced from academic journals, research articles, and scholarly publications, that is relevant to this study. It focuses primarily on the development and application of AI-enhanced assistive communication technologies, including Morse code-based translation systems, eye-tracking methods, blink detection mechanisms, and innovative frameworks for technological progress.

a. Foreign Literature

Understanding of Convolutional Neural Network (CNN): A Review

Purwono et al. (2022) presented a comprehensive review of Convolutional Neural Networks (CNNs), detailing their core architecture—including convolutional, pooling, fully connected, and activation layers—and examining popular models such as LeNet, AlexNet, and VGG. They discussed CNN applications across image classification, segmentation, object detection, video processing, NLP, and speech recognition, highlighting how architectural choices (e.g., kernel size, pooling strategies, activation functions) influence performance and computational requirements. The review also addressed challenges and future directions for CNN adoption in various domains, emphasizing its role in advancing deep learning capabilities.

A review of deep learning in blink detection

Xiong et al. (2025) conducted a comprehensive review of deep learning methods for eye‑blink detection, evaluating classical architectures such as CNNs and RNNs across varied application contexts like human–computer interaction, fatigue monitoring, and emotion analysis. They categorized existing work by algorithm types, data sources, and evaluation metrics, noting strong performance overall but highlighting ongoing challenges related to class imbalance, environmental variability, real-time processing, and device limitations. Their synthesis concludes with practical recommendations for improving data diversity, model efficiency, and real-world reliability in blink detection systems.

Face Recognition Based on Deep Learning: A Comprehensive Review

Dakhil et al. (2024) conducted a comprehensive review of deep learning–based face recognition, analyzing 150 peer-reviewed papers from major databases. They evaluated key architectures—including CNNs, autoencoders, and ResNet variants—and discussed their performance under challenges such as lighting variation, occlusion, and facial expressions. The study also emphasized ethical concerns like privacy and bias while outlining future research directions toward robust, efficient, and fair face recognition systems.

Innovation Alchemy: Unravelling the Secrets of Technological Progress

Chaturvedi et al. (2024) described a structured framework in *Innovation Alchemy* to decode the mechanics behind technological innovation. They introduced the concept of "codes"—systemic elements such as socio-economic enablers, behavioral loops, and adaptive learning systems—that collectively influence innovation as both a process and an outcome. Using interdisciplinary case studies, the authors provide a replicable model for understanding and applying innovation strategies across various contexts, offering practical guidance for researchers and technology practitioners.

Morse Code Translator Using Camera Feed and Blinks

Akbuluteren et al. (2021) introduced a computer vision-based Morse code translator that uses webcam footage to detect eye blinks—categorized by duration as short or long—and convert them into Morse alphabet characters in real time. The system leverages MediaPipe's Face Mesh to track 468 facial landmarks, calculating the Eye Aspect Ratio (EAR) to accurately distinguish blink types, and then maps sequences of blinks to corresponding Morse code symbols for message translation. This practical prototype demonstrates the potential of accessible, low-cost tools for assistive communication through blink-controlled input.

b. Local Literature

A Hybrid CNN-LSTM-Based Visual Decoding Technique and Independent Video Preprocessing for Lip-Reading in Tagalog

Deocampo et al. (2023) introduced a hybrid deep learning model combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks to enhance lip-reading accuracy for the Tagalog language. Using a dataset of 450 videos with 50 distinct phrases spoken by nine native speakers, the model extracted spatial and temporal features to decode lip movements into text. The system achieved high frame-level and phrase-level accuracy, demonstrating its potential for applications in accessibility, human-computer interaction, and surveillance.

Gesture Recognition of Filipino Sign Language Using Convolutional and Long Short-Term Memory Deep Neural Networks

Cayme et al. (2024) developed a real-time Filipino Sign Language (FSL) gesture recognition system using a hybrid CNN–LSTM model to interpret both expressive and transactional FSL gestures from live video input. The model was trained on preprocessed frames involving cropping, contrast enhancement, grayscale conversion, resizing, and normalization, achieving around 98% average accuracy, precision, and recall—an 11.3% improvement over earlier systems. They also created a lightweight version of the model using quantization, maintaining up to 99% accuracy on low-resource devices like the Raspberry Pi.

A Lip-Reading Model for Tagalog Using a Multimodal Deep Learning Approach

Deocampo et al. (2024) created a Tagalog-specific lip-reading system using a multimodal deep learning approach that combines visual cues like facial movements with linguistic context through a CNN–LSTM framework. Trained on Tagalog speech videos for over 80 epochs, the model reached 89.5% validation accuracy and processed inputs at least 25% faster. This research highlights the potential of tailored AI tools to improve accessibility for the deaf and hard of hearing, with adaptability to other languages and speaking environments.

Smart Face Shield: A Sensor-Based Wearable Face Shield Utilizing Computer Vision Algorithms

Delos Santos et al. (2022) developed a wearable face shield enhanced with computer vision algorithms to promote social distancing and raise awareness of health protocols. The device integrates a transparent polycarbonate shield, sensors, a camera, and an on-screen display, utilizing OpenCV for object detection and TensorFlow for distance measurement. This innovation aims to augment traditional face shields with real-time monitoring capabilities, contributing to public health safety efforts.

Identity Verification through Face Recognition Implemented on Raspberry Pi Framework

Virata et al. (2022) implemented a cost-effective, offline face recognition framework using a Raspberry Pi 3 B+ that connects wirelessly to a mobile phone for image processing. The system uses Local Binary Patterns (LBP) for face detection and verification across a dataset of 3,000 images, achieving 87.5% accuracy, 88% precision, 88% recall, and an 86% F1 score. Despite hardware constraints, the project demonstrated feasibility for practical applications such as voter ID systems and checkpoints within government services.

Review of Related Studies

A. Foreign Studies

Creating an AI-Enhanced Morse Code Translation System Based on Images for People with Severe Disabilities

Wu et al. (2023) developed an AI-enhanced Morse code translation system (AIMcT) designed to assist individuals with severe physical impairments in their communication efforts. By utilizing facial motion recognition technology, the system captures facial movements through a webcam and translates them into Morse code using sophisticated algorithms. The system achieved an impressive accuracy rate of 99.83% with expert data and 98.6% with data collected from individuals with disabilities. These results demonstrate its effectiveness in facilitating independent communication and significantly enhancing the quality of life for users.

A CNN and LSTM Network for Eye-Blink Classification from MRI Scanner Monitoring Videos

Bennett et al. (2021) developed a two-stage convolutional recurrent neural network to classify eye states from low-quality MRI monitoring videos, addressing the critical need for accurate eye-blink detection during resting-state fMRI. Their model combined an Inception-v3-based convolutional neural network (CNN) with long short-term memory (LSTM) networks to capture temporal dependencies, achieving a precision of 0.739 and a recall of 0.835 on an independent dataset. This non-invasive approach enables eye state tracking as a covariate in neuroimaging studies, improving the reliability of functional connectivity analyses and contributing to more precise interpretations of brain activity.

Hands-Free Computing: Virtual Mouse and Keyboard with Eye Gaze

Shanthini et al. (2024) created a hands-free computing system that enables individuals with physical disabilities to operate a virtual mouse and keyboard by detecting eye blinks and tracking gaze in real time. The system employs the Eye Aspect Ratio (EAR), Haar cascade classifiers, and Support Vector Machine (SVM) to accurately recognize and classify eye movements. This method eliminates the requirement for external sensors and facilitates efficient human-computer interaction using just a regular webcam.

Morse Code-Based Assistive Communication System for Individuals with Partial Paralysis

Pravin Kumar S K et al. (2025) developed a Morse code-based assistive communication system that enables individuals with partial paralysis to communicate through eye blinks. The system employs a wearable device equipped with an infrared sensor to detect these movements, translating them into Morse code signals that are displayed as text on a mobile application. By emphasizing user-friendliness and accessibility, this design allows individuals with severe physical impairments to engage in meaningful interactions, highlighting the potential of adaptive technologies to enhance the quality of life for those facing communication challenges.

An efficient machine learning approach for volunteer eye-blink detection in real-time using webcam

Wei et al. (2022) present an original study proposing an enhanced face liveness detection method that utilizes an optimized LeNet-5 convolutional neural network (CNN). The authors introduce modifications to the standard LeNet-5 architecture by increasing the size of the convolution kernels and incorporating global average pooling, aiming to improve the model's ability to distinguish between live and spoofed facial images. Experimental results demonstrate that the proposed model achieves a recognition accuracy of 99.95%, surpassing the 96.67% accuracy of the Support Vector Machine (SVM) model and the 98.23% accuracy of the standard LeNet-5 model, underscoring the effectiveness of the enhanced CNN approach in improving the security of face recognition systems by accurately identifying liveness in facial images.

Eye Blink-based Morse Code Communication Tool for ALS Patients

Salvi et al. (2024) system uses Dlib's 68-point facial landmark detector and OpenCV to compute an Eye Aspect Ratio (EAR) for each frame, then applies fixed thresholds on EAR-derived blink durations to distinguish short blinks (dots) from long blinks (dashes). It converts the resulting dot/dash sequence via a lookup table into letters, which are displayed on an LCD or sent over WhatsApp via PyWhatKit.

Creating an AI-Enhanced Morse Code Translation System Based on Images for People with Severe Disabilities

Wu et. al. (2023) employ AI-driven facial-feature extraction (via Dlib) to detect mouth or cheek movements, then use a fuzzy-time-recognition (FTR) algorithm to dynamically adjust thresholds for short vs. long motion intervals, generating dot/dash signals which map into Morse characters and control external devices wirelessly.

B. Local Studies

Facial recognition technology: A case study on its adaptation for local traffic management use in the Philippines

Turingan et al. (2024) conducted a case study on implementing facial recognition technology for traffic management in the Philippines' Land

Transportation Office. They evaluated system integration challenges and identified policy, infrastructure, ethical, and privacy considerations critical for adoption. Their findings highlight the potential benefits of facial recognition for improving identity verification and road safety, while underscoring the need for robust data protection and user acceptance measures.

CCTV-Based Surveillance System with Face Recognition Feature

Lumaban et al. (2020) developed a CCTV-based surveillance system enhanced with facial recognition capabilities, employing algorithms such as LBPH, Eigenface, and Fisherface for accurate identity verification. The system integrates real-time video monitoring with automatic face detection and recognition, providing alerts when an unauthorized person is detected in a secured area. With LBPH achieving a recognition accuracy of 95.9%, the study demonstrates the practical viability of combining machine learning algorithms with surveillance hardware to enhance security systems in institutional environments.

A Convolutional Neural Network Study on Depression and Eye Blink Analysis

Dadiz et al. (2023) developed a CNN-based method for detecting eye blinks using the Closed Eyes in the Wild dataset, achieving 99.24% training accuracy over 50 epochs. When validated on BDI-II-labeled datasets, the model reached 61.09% accuracy and showed a moderate correlation ($R^2 \approx 0.34$) between blink frequency and depression severity. The results suggest that blink patterns may serve as early indicators of depression, despite the relatively weak relationship.

Overcoming Communication Barriers and Building Facilitative Patterns for Learners with Disabilities in Self-Contained Classes

Marinduque et al. (2025) conducted a descriptive study on overcoming communication barriers and establishing facilitative patterns among learners with disabilities in self-contained classes. Based on survey responses from 100 special education teachers, the study found that speech and language difficulties, sensory and behavioral issues, and emotional and cognitive barriers were “Sometimes Observed.” Effective strategies identified included the use of augmentative and alternative communication (AAC) tools, structured teaching, and peer-assisted learning. While no significant overall relationship was found between communication barriers and facilitative patterns (p = 0.348), significant differences were noted based on teacher sex, experience, education, and training.

Liveness Detection Based on an Improved Convolutional Neural Network for Face Recognition Security

Wei et al. (2022) proposed an improved CNN-based face liveness detection method for biometric security in their paper. They optimized LeNet-5 by increasing the convolutional kernel size and applying global average pooling, achieving 99.95% accuracy, outperforming both the original LeNet-5 (98.23%) and SVM-based approaches (96.67%) in distinguishing live from spoofed face inputs using test datasets.

Facial recognition technology: A case study on its adaptation for local traffic management use in the Philippines

Edgar R. Eslit's study (2024) evaluates the effectiveness and ethical concerns of AI plagiarism detection tools in higher education. It finds these tools often misinterpret context, leading to inaccuracies. Emphasizing the need for human judgment, the research argues that technology should support, not replace, academic integrity. The study advocates for a hybrid approach that blends technological solutions with human oversight.

Synthesis

The review of related literature and studies uncovers notable shortcomings in current assistive communication technologies for individuals with severe motor impairments. Purwono et al., 2022 and Xiong et al., 2025 noted that although Convolutional Neural Networks (CNNs) have consistently proven effective in extracting spatial features from visual data, and deep learning–based approaches continue to deliver "strong overall performance" in blink detection, there is still no cohesive framework that integrates both spatial and temporal modeling for real-time Morse code translation using eye blinks. Existing solutions often address isolated components such as "facial motion recognition" (Wu et al., 2023) or simple blink duration detection (Akbuluteren et al., 2021) yet fall short of capturing the sequential dynamics of blinks required for reliable Morse decoding under real-world conditions like shifting illumination, varied head orientations, or involuntary twitching.

Hybrid CNN-LSTM approaches have shown great promise in closely related areas. Deocampo et al. (2023, 2024) successfully combined CNNs and LSTMs for Tagalog lip-reading, reporting “89.5% validation accuracy,” while Cayme et al. (2024) applied a similar model to Filipino Sign Language recognition and achieved “98% accuracy.” These works highlight the benefits of combining CNN-driven spatial modeling with LSTM-driven temporal learning for complex language-related visual tasks. However, when applied to eye-blink detection, such as in the study by Bennett and Joshi (2021), results were modest, with “0.739 precision and 0.835 recall,” showing that intentional and unintentional blinks remain difficult to distinguish without stronger sequence modeling.

Research in face recognition and other vision-based assistive systems provides additional insights relevant to this gap. Dakhil and Abdulazeez (2024) reviewed CNN-based methods in face recognition, underscoring their accuracy but also pointing out challenges with generalization across diverse environments. Local initiatives further demonstrate feasibility: Virata and Festijo (2022) implemented recognition systems on lightweight Raspberry Pi devices, while Santos et al. (2023) introduced a computer vision–driven wearable for protection and monitoring. Similarly, Wei et al. (2022) stressed the necessity of liveness detection for security, a principle that can equally enhance the robustness of blink-based communication systems.

Meanwhile, studies focusing directly on Morse code–based communication highlight the continuing limitations of existing methods. Wu et al. (2023) developed an AI-enhanced Morse code translation tool for individuals with severe disabilities, while K (2025) introduced a blink-driven system for those with partial paralysis. These contributions validate Morse code as a viable medium for communication but still rely on fixed thresholds and basic detection frameworks, reducing adaptability to different user conditions. Earlier efforts, such as the prototype from Akbuluteren et al. (2021), further illustrate the potential of blink-based Morse systems but also expose their lack of temporal robustness for continuous real-time interaction.

The proposed study aims to bridge these shortcomings through a modular CNN-LSTM pipeline specifically tailored to blink-driven Morse code communication. In this design, CNNs provide accurate open/closed eye classification despite variations in lighting or head pose, while LSTMs capture the full blink sequence (open → closed → open) and learn personalized timing patterns for dot–dash interpretation. This eliminates the dependence on rigid thresholds, improves resilience to fatigue and involuntary blinks, and allows for modular testing of both spatial and temporal components. By drawing from advances in multimodal deep learning applied in lip-reading, gesture recognition, and vision-based assistive devices, this framework not only improves decoding accuracy but also contributes a scalable and adaptive communication tool designed for individuals with severe motor impairments.

CHAPTER 3

METHODS AND PROCEDURES OF THE STUDY

Methods of Research Used

This study adopts a quantitative, quasi-experimental framework. Quantitative methods form the backbone of this investigation by providing precise, objective measurements of each pipeline component's performance. By operationalizing key outcomes such as CNN precision, recall, and F_1-score for eye-state classification, LSTM blink-detection rate and false-positive rate, symbol error rate, and end-to-end latency, the study can employ statistical analyses (e.g., confidence intervals, hypothesis tests) to determine whether observed improvements are reliable rather than due to chance. These numeric indicators facilitate direct comparisons across lighting conditions, head poses, and participant groups, and their replicability ensures that subsequent researchers can validate or build upon the research findings under similar or novel settings.

A quasi-experimental design enables causal inference without demanding random assignment, where participants are naturally grouped (healthy vs. motor-impaired) and not randomly assigned. Together, these approaches strike a balance between methodological rigor and ecological relevance. The structured, data-driven nature of quantitative research yields transparent, statistical evidence of system performance, while the quasi-experimental design ensures that those findings reflect the lived experiences of the intended user population. By uniting these designs, the study not only quantifies the CNN and LSTM pipeline's capabilities but also validates its practical utility as a robust, adaptable communication aid. The research will also employ a Guided Survey to assess the behavior, performance, and accuracy of the system's output.

The evaluation of the system and the CNN and LSTM models to determine which is beneficial in development and answering the research question is as follows:

1. Frame-Level Classification Performance
2. Blink-Event Detection Accuracy
3. Temporal Parameter Stability
4. End-to-End Decoding Metrics

5. User‑Centric Measures

Locale of the Study

This study was conducted on different groups of individuals from Manila, Philippines. It combines a large, heterogeneous urban population with readily accessible research and clinical infrastructures, thereby maximizing both external validity and participant diversity. Furthermore, the concentration of universities, rehabilitation centers, and assistive‑technology clinics within Metro Manila facilitates efficient recruitment of both neurologically healthy volunteers and individuals with motor impairments, ensuring that the researcher's sample reflects the full spectrum of end‑user capabilities.

Respondents of the Study

Because intentional eye blinking is a universally preserved motor function, the study will primarily recruit a broad sample of adult Metro Manila residents (aged 18–65) with normal or corrected‑to‑normal vision. This healthy group will serve as a baseline to evaluate the accuracy of CNN-based eye-state classification and LSTM-based blink-event detection across varying lighting conditions, head positions, and individual blinking habits. Their data will inform model calibration, threshold refinement, and robustness testing under conditions representative of the general population.

In addition, to assess real‑world applicability among end‑users, the study will include any available individuals living with severe motor impairments (e.g., early‑stage ALS or locked‑in syndrome) who retain voluntary control of eye blinks. Performance metrics gathered from this secondary cohort will demonstrate the system's resilience and utility for its intended assistive communication purpose. All data collection and evaluation will occur in controlled laboratory and affiliated clinical settings throughout Metro Manila using standardized webcam setups and indoor lighting.

Research Instrument

The purpose of the study's research instruments is to systematically capture, standardize, and annotate participants' eye-blink video data through controlled webcam recordings, automated eye-region extraction, and manual frame-level labeling to provide reliable ground-truth inputs for training and evaluating the CNN and LSTM modules.

Hardware, Software, and Environment Specifications

Hardware:

The testing and part of the development of the system were done on a local Personal Computer with the following specifications:

1. **CPU:** Ryzen 5 5600 (6 cores, 12 threads, running at 4.6 GHz)
2. **GPU:** Nvidia RTX 4060 8GB GDDR6
3. **RAM:** 32 GB DDR4
4. **Storage:** WD SN770 2TB SSD

Software:

1. Video Capture Hardware: Standard USB webcams record participants' eye movements.
2. OpenCV - For video capture (webcam or file), frame preprocessing (resizing, normalization), and rapid prototyping of image pipelines.
3. MediaPipe Face Mesh - Used to identify regions of a face in each frame for consistent input to the CNN.
4. Prebuilt Datasets and Annotation Tools - Publicly available eye-blink image/video collections and custom annotation utilities and blink event timestamps in the recordings, establishing ground-truth.
5. Microsoft Excel - Used to store datasets for training and enhancing pre-trained models, such as BERT and SciBERT.
6. Google Colab - Used for model training, testing, and experimentation in a cloud-based, GPU-supported environment.
7. Android Studio Emulator - Emulating and Testing the mobile application deployment of the hybrid model.

8. Flutter - Used to develop the cross-platform web and mobile application, ensuring a consistent interface for the model.

The research also employs a survey as the research instrument. This survey aims to capture user experience, ease of use, and the performance of the proposed system to capture participants' subjective experiences, such as perceived system responsiveness, ease of use, and mental or physical fatigue that cannot be measured by quantitative model metrics alone. Administered immediately after each testing block, the survey uses standardized Likert‑scale items and open‑ended questions.

Data Gathering Procedure

To obtain the necessary data for analysis in this study, the research will utilize a mix of publicly accessible datasets and custom-collected video recordings. Public blink datasets from platforms such as GitHub will form the foundation for initial training and benchmarking of the convolutional neural network (CNN). These datasets provide pre-annotated labels for eye states and blink sequences across varying conditions of lighting, head pose, and subject diversity, allowing the model to effectively generalize to different user scenarios.

In addition, the research will gather participant-recorded videos through a controlled setup using standard webcams. Participants will intentionally perform blink sequences to encode specific Morse characters, with their eye movements captured. The research will preprocess these recordings with MediaPipe Face Mesh to ensure consistent extraction of the eye region, followed by manual verification using a custom annotation tool to label frame-level eye states and blink events. The resulting annotated data will be structured into well-organized datasets for training and evaluation.

All collected data will undergo cleaning, merging, and partitioning using tools like NumPy and Pandas to create distinct training, validation, and test sets. This process guarantees that both the CNN and LSTM components are trained on high-quality, diverse, and clearly labeled inputs, enabling rigorous quantitative analysis of their performance using standard evaluation metrics.

Data Analysis Tools

The following tools and platforms were utilized for data analysis and model implementation:

1. Python - Used for data preprocessing, model development, annotation processing, and evaluation.
2. TensorFlow – Framework for loading, fine-tuning, and deploying transformer models.
3. Visual Studio Code - Used for developing the front-end interface, code management, as well as running and testing the proposed system locally.
4. NumPy - Computes evaluation metrics (e.g., blink durations, frame-level timing) using logical operations and vectorized arithmetic.
5. Pandas - Enables filtering and statistical analysis (e.g., grouping blinks by user, summarizing durations, calculating average latency).

Statistical Treatment of Data

CNN Eye-State Detection and Performance Metrics

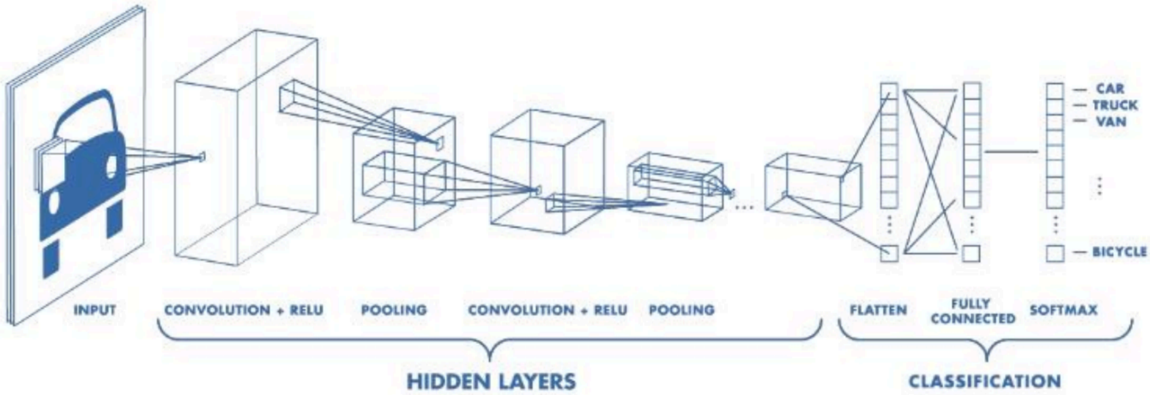

Figure 3.1: CNN Architecture Diagram, adapted from Understanding of Convolutional Neural Network (CNN): A Review by Purwono et al.

Figure 3.1 depicts the Convolutional Neural Network (CNN) architecture that functions as the foundational spatial feature extractor. Each cropped eye image (the input) is subjected to a series of convolutions and ReLU (Rectified Linear Unit) layers, followed by pooling layers, to learn robust representations of eyelid contour, texture, and contextual information.

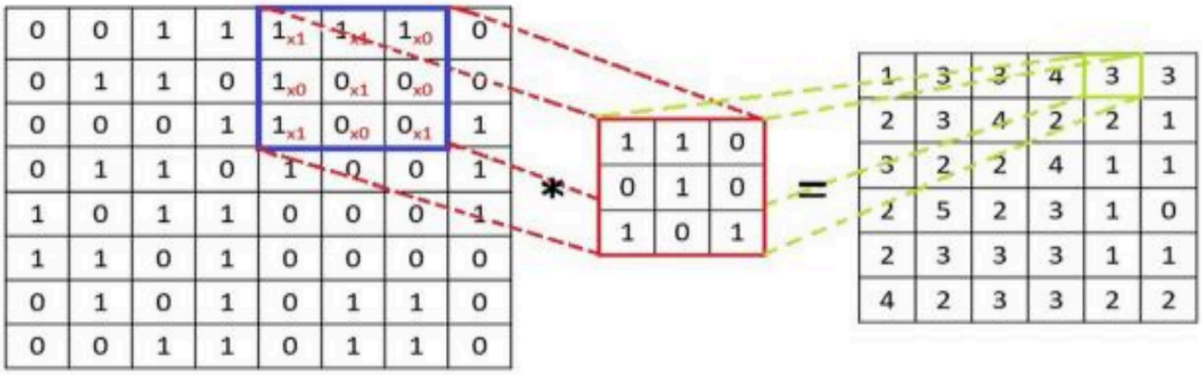

Figure 3.2: Convolutional Layer Innerworkings, adapted from Understanding of Convolutional Neural Network (CNN): A Review by Purwono et al.

A convolutional layer operates within a Convolutional Neural Network (CNN) to enhance image processing. A small filter or kernel (illustrated in red) systematically moves over the input image, executing element-wise multiplication before summing the results to create a single value in the activation map (output). This process is vital for capturing local features such as edges, textures, and shapes within the image. The kernel's movement is governed by the stride, while padding can be applied around the input to maintain spatial dimensions. Following the convolutional step, CNNs implement a pooling layer that minimizes the spatial size of the feature maps while preserving critical information. Pooling operates by moving a small window (for instance, 2×2) across the activation map, selecting either the maximum (max pooling) or average (average pooling) value from each window. This approach not only optimizes computation and enhances spatial invariance but also aids in mitigating overfitting. Although pooling is not depicted in Fig. 3, it is a common practice in most CNN architectures after the convolution process.

This process culminates in a final softmax output that classifies each frame as either "open" or "closed" with a high degree of confidence. By transforming raw pixel data into reliable per-frame eye-state predictions—even amidst variations in lighting and head pose—the CNN front-end ensures that subsequent modules receive clean and invariant signals, thereby minimizing the impact of noisy image data.

To quantify the frame-level reliability of our CNN in distinguishing open and closed eye states, the study computes standard classification metrics (precision, recall, and F_1-score) across controlled lighting and pose conditions. These measures will establish the model's ability to generalize spatially learned eyelid features under varied environmental and subject-specific factors.

For the test set of *N* frames, let:

TP = true positives ("closed" correctly predicted),

FP = false positives ("open" predicted as "closed"),

FN = false negatives ("closed" predicted as "open").

Then, to calculate Precision, Recall, and F1

Precision = TP / (TP + FP)

Recall = TP / (TP + FN)

F1 = 2 x (Precision x Recall) / (Precision + Recall)

LSTM Blink-Event Detection & Classification

Building upon the CNN's frame-level labels, the LSTM's temporal modeling performance is evaluated through its blink detection rate and false-positive rate, as well as a blink-type F_1-score derived from confusion matrices comparing dots and dashes. This suite of statistics illuminates the network's effectiveness in identifying complete open→closed→open transitions and accurately categorizing their durations as Morse signals.

For Computing dot and dash precision, recall, and overall blink type F1 score:

$$F_{1,\text{blink}} = \frac{\text{Precision}_d \times \text{Recall}_d + \text{Precision}_D \times \text{Recall}_D}{(\text{Precision}_d + \text{Precision}_D) + (\text{Recall}_d + \text{Recall}_D)}$$

For the dot and dash Confusion Matrix classification:

Table 3.1: *Confusion Matrix:*

	Predicted Dot	Predicted Dash
True Dot	TP_d	FN_d
True Dash	FP_d	TP_D

End-to-End Decoding & Latency

The research will measure the system's holistic communication accuracy by calculating symbol error rate and word-level decoding accuracy via Levenshtein distance, and we will profile real-time performance through average and 95th-percentile latency metrics. This end-to-end evaluation will confirm whether the pipeline meets practical usability requirements for blink-based Morse transcription.

The end-to-end decoding performance is quantified by the *Symbol Error Rate* (SER), defined as the sum of substitutions, deletions, and insertions (D_{sub} + D_{del} + D_{ins}) and normalized by the length of the ground-truth symbol sequence | S_{ref} |.

$$SER = \frac{D_{sub} + D_{del} + D_{ins}}{|S_{ref}|}$$

4-Point Likert Scale

Scale used for the Guided Survey given to the respondents of the Study. The number that corresponds to the answer will be used internally for Statistical Evaluation of the Result:

1. Completely Accurate - 4
2. Mostly Accurate - 3
3. Somewhat Accurate - 2
4. Not Accurate - 1

Statistical Evaluation

Metrics for measuring the aggregate of user feedback:

1. *Mean:* For finding the average user rating from the survey
2. *Frequency Distribution and Mode:* To identify the distribution of the responses and the central tendency of the answers

CHAPTER 5

SUMMARY OF FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS

This chapter presents the summary of the thesis work undertaken, the conclusions drawn, and the recommendations made as an outgrowth of this study. This study examines the performance and accuracy of decoding Morse code via eye blinks using Convolutional Neural Networks and Long Short-Term Memory machine learning algorithms.

Summary of Findings

This research presents a comprehensive approach to assistive communication technology through a deep learning pipeline that translates intentional eye blinks into Morse code and subsequently into textual output. The research addresses critical communication barriers faced by individuals with severe motor impairments, particularly those with quadriplegia who retain only basic functions like blinking. According to the World Health Organization's World Report on Disability cited in the study, the global population aged 60 and over is expected to double from 11% to 22% by 2050, with an estimated 15% of people worldwide living with disabilities, making this research increasingly relevant to a growing demographic need.

The proposed system employs a modular three-stage architecture that integrates spatial and temporal learning approaches. The first component utilizes MediaPipe Face Mesh for precise eye region localization and tracking across 468 facial landmarks, coupled with a convolutional neural network trained to classify eye states as "open" or "closed" on a per-frame basis. This CNN demonstrates robustness across varying lighting conditions, camera angles, and head poses while processing video frames captured via standard webcam or pre-recorded footage. The second stage employs Long Short-Term Memory networks to recognize complete blink patterns through open-to-closed-to-open transitions, automatically measuring blink duration and categorizing events as Morse "dots" or "dashes" without requiring manual threshold tuning. The final component implements a rule-based decoder that compiles dot and dash symbols into letters and words using International Morse Code standards, providing on-screen text display and optional text-to-speech synthesis.

The research framework adopts a quantitative, quasi-experimental approach focusing on frame-level classification performance using precision, recall, and F1-score metrics, blink-event detection accuracy and false-positive rates, temporal parameter stability across varying conditions, end-to-end decoding metrics including Symbol Error Rate and word-level accuracy, and real-time latency measurements for usability assessment. The study targets individuals with severe motor impairments while maintaining broad applicability, with primary users including adults with conditions like quadriplegia, early-stage ALS, or locked-in syndrome, and testing conducted among Metro Manila residents aged 18-65 with normal or corrected vision.

The comprehensive literature review revealed several critical gaps in existing assistive communication technologies. Most current implementations focus on either CNN-based spatial classification or LSTM-based temporal modeling in isolation, while traditional blink-based systems rely on simple duration thresholds or mechanical switches without considering real-world conditions affecting users with severe disabilities. The review found that related hybrid architectures showed promising results, with Deocampo et al. (2023) achieving 89.5% validation accuracy in Tagalog lip-reading and Cayme et al. (2024) reaching 98% average accuracy in Filipino Sign Language recognition. However, Bennett et al. (2021) applied CNN-LSTM to eye-blink classification and achieved only 0.739 precision and 0.835 recall under challenging conditions, indicating room for improvement in blink-based temporal modeling approaches.

Conclusion

This research addresses a significant gap in assistive communication technology by developing an integrated CNN-LSTM pipeline specifically optimized for Morse code translation through eye blink detection. The hybrid approach offers key advantages over traditional threshold-based systems by demonstrating improved resilience to environmental variations and providing sophisticated temporal understanding that reduces false positives from involuntary eye movements.

For individuals with severe motor impairments, this webcam-based system represents an accessible alternative to expensive eye-tracking hardware, potentially enhancing autonomy and social interaction for users with conditions like quadriplegia or ALS. The modular design enables future researchers to experiment with alternative architectures or expand beyond Morse code to more efficient communication protocols.

While the study demonstrates methodological rigor appropriate for assistive technology research, certain limitations should be acknowledged, including the restriction to International Morse Code which may limit communication speed, and the need for empirical validation with the intended user population to establish real-world clinical efficacy. The research represents a meaningful step toward democratizing assistive communication technology, leveraging consumer hardware and advanced machine learning to create potentially transformative tools for individuals with severe disabilities.

REFERENCES:

Purwono, P., Ma'arif, A., Rahmaniar, W., Fathurrahman, H. I. K., Kusuma Frisky, A. Z., & Mazhar ul Haq, Q. (2022). Understanding of Convolutional Neural Network (CNN): A Review. *International Journal of Robotics and Control Systems, 2*(4), 739–748. https://doi.org/10.31763/ijrcs.v2i4.888

Xiong, J., Dai, W., Wang, Q., Dong, X., Ye, B., & Yang, J. (2025). A review of deep learning in blink detection. *PeerJ Computer Science*, 11, e2594 https://pubmed.ncbi.nlm.nih.gov/39631108/

Dakhil, N., & Abdulazeez, A. M. (2024). *Face recognition based on deep learning: A comprehensive review*. *Indonesian Journal of Computer Science, 13*(3), 3779–3794. Retrieved from http://ijcs.net/ijcs/index.php/ijcs/article/view/4037

Chaturvedi, V., Singh, G. D., Kursija, M., Singh, G., & Saha, A. (2024). Innovation Alchemy: Unravelling the Secrets of Technological Progress. In *Innovation Alchemy: Unravelling the Secrets of Technological Progress*. Parab Publications. https://parabpublications.com/books/pdf/innovation-alchemy-unravelling-the-secrets-of-technological-progress.pdf#page=61

Akbuluteren, E., Güngören, A., & Yılmaz, B. (2021, June 15). *Morse code translator using camera feed and blinks*. Eren Akbulut's Blog. Retrieved from https://blog.akbuluteren.com/blog/morse-code-translator-using-camera-feed-and-blinks

Deocampo, N. J. E., West Visayas State University, Laguna State Polytechnic University, & Isabela State University. (2023). A hybrid CNN-LSTM-Based visual decoding technique and independent video preprocessing for Lip-Reading in Tagalog. In *International Journal of Engineering Research in*

Computer Science and Engineering. https://www.researchgate.net/publication/380585071

Cayme, K. J., Retutal, V. A., Salubre, M. E., Astillo, P. V., Cañete, L. G., & Choudhary, G. (2024). Gesture recognition of Filipino sign language using convolutional and Long Short-Term memory deep neural networks. *Knowledge*, *4*(3), 358–381. https://doi.org/10.3390/knowledge4030020

Deocampo, N. J., Villarica, M., & Vinluan, A. (2024). A Lip-Reading model for Tagalog using multimodal deep learning approach. *International Journal of Computing Sciences Research*, *8*, 2796–2808. https://doi.org/10.25147/ijcsr.2017.001.1.186

Santos, M. L. D., Tinio, R., Diaz, D., & Tolosa, K. E. (2023). Smart Face Shield: A Sensor-Based wearable face shield utilizing computer vision algorithms. *International Journal of Computing Sciences Research*, *7*, 1578–1593. https://doi.org/10.25147/ijcsr.2017.001.1.118

Virata, A. J., & Festijo, E. (2022). *Identity verification through face recognition implemented on Raspberry Pi framework*. *International Journal of Communication Networks and Information Security*, 11(3). https://doi.org/10.17762/ijcnis.v11i3.4254

Wu, C., Chen, Y., Chen, S., & Zheng, S. (2023). Creating an AI-Enhanced Morse Code Translation System Based on Images for People with Severe Disabilities. *Bioengineering*, *10*(11), 1281. https://doi.org/10.3390/bioengineering10111281

Bennett, R., & Joshi, S. H. (2021). A CNN and LSTM Network for Eye-Blink Classification from MRI Scanner Monitoring Videos. In IEEE Engineering in Medicine & Biology Society, *43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (Vols. 1–1, p. 3463). https://paperhost.org/proceedings/embs/EMBC21/files/1178.pdf

Shanthini, S., Sathiya Priya, A., & Banuroopa, K. (2024). Hands-Free Computing: Virtual Mouse and Keyboard with Eye Gaze. In N.G.P. Arts and Science College, *International Journal of Advanced Trends in Engineering and Management* (pp. 831–838). https://www.researchgate.net/publication/387031009

K, P. K. S. (2025). Morse Code-Based Assistive Communication System for Individuals with Partial Paralysis. *International Journal for Research in Applied Science and Engineering Technology*, *13*(4), 597–601. https://doi.org/10.22214/ijraset.2025.68330

Medeiros, P. A. D. L., Medeiros, Silva, G. V. S. D., Silva, Fernandes, F. R. D. S., Fernandes, Sánchez-Gendriz, I., Sánchez-Gendriz, Lins, H. W. C., Lins, Barros, D. M. D. S., Barros, Nagem, D. A. P., Nagem, Valentim, R. A. D. M., Valentim, & Laboratory of Technological Innovation in Health (LAIS), Federal University of Rio Grande do Norte (UFRN), Natal, RN, Brazil. (2021). Efficient machine learning approach for volunteer eye-blink detection in real-time using webcam. In *Expert Systems With Applications* (Vol. 188, p. 116073). https://doi.org/10.1016/j.eswa.2021.116073

Turingan, A. M., Nortado, R. A., Jr, & Jocson, J. C. (2024, May 15). *Facial recognition technology; A case study on its adaptation for local traffic* *management use in the Philippines.* https://journal.ijprse.com/index.php/ijprse/article/view/1055

Lumaban, M. B. P. (2020). CCTV-Based Surveillance System with Face Recognition Feature. International Journal of Advanced Trends in Computer Science and Engineering, 9(1.3), 349–355. *https://doi.org/10.30534/ijatcse/2020/5491.32020*

Dadiz, B. G., & IRegla, A. (2023). A Convolutional Neural Network Study on Depression and eye blink analysis. International Journal of Innovative Technology and Exploring Engineering, 12(5), 7–11. *https://doi.org/10.35940/ijitee.e9488.0412523*

Marinduque, K. K. T., & Marquez, M. F. (2025). Overcoming Communication Barriers and Building Facilitative Patterns for Learners with Disabilities in Self-Contained Classes. *Journal of Education, Learning, and Management*, *2*(1), 45–49. https://doi.org/10.69739/jelm.v2i1.245

Wei, Y., Machica, I. K. D., Dumdumaya, C. E., Arroyo, J. C. T., & Delima, A. P. (2022). Liveness detection based on improved convolutional neural network for face recognition security. International Journal of Emerging Technology and Advanced Engineering, 12(8), 45–53. *https://doi.org/10.46338/ijetae0822_06*