NEUROSPEED: AN EMOTION-ADAPTIVE SPEED CONTROL SYSTEM FOR MEDICAL MOBILITY DEVICES USING REAL-TIME MULTIMODAL FACIAL ANALYSIS

C.S.B.HETTIHEWA

IT22574336

Dissertation submitted in partial fulfillment of the requirements for the
Bachelor of Science (Honours) in Information Technology
Specializing in Software Engineering

Department of Information Technology
Sri Lanka Institute of Information Technology
Sri Lanka

August 2025

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to Sri Lanka Institute of Information Technology, the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: _____ Date: _____

The above candidate has carried out research for the Bachelor's Degree Dissertation under my supervision.

Signature of the Supervisor: _____ Date: _____

Name of Supervisor: _____

Abstract

Drowsiness and emotional distress among patients operating motorised medical devices — including powered wheelchairs and rehabilitation aids — pose significant safety risks that existing speed control architectures are unable to address autonomously. Contemporary systems rely on joystick-based manual input and simple proximity collision avoidance; neither modality can detect an operator whose cognitive state has deteriorated below a safe operating threshold. This research presents NeuroSpeed, a real-time emotion-adaptive speed control system that addresses this gap through a unified multimodal facial analysis pipeline. The system integrates MediaPipe Face Mesh landmark extraction, a weighted ensemble of DeepFace VGG-Face and FER convolutional neural network classifiers with Exponential Moving Average temporal smoothing, Eye Aspect Ratio closed-eye detection, PERCLOS drowsiness quantification over a rolling six-second buffer, Perspective-n-Point head pose estimation, Mouth Aspect Ratio yawn detection, and a rule-based speed arbitration engine. Three discrete speed levels — HIGH, LOW, and STOP — are defined with absolute fail-safe precedence ordering. System state is persisted to Firebase Realtime Database and visualised through a React clinical dashboard accessible to remote caregivers. Empirical evaluation on an Apple M2 platform yielded a total synchronous pipeline latency of 103 to 259 milliseconds, satisfying the sub-500 millisecond safety response requirement. Ensemble emotion classification achieved a macro-averaged F1 score of 0.71 across seven affect categories, a statistically significant improvement over the best single-model baseline ($p < 0.001$). Speed level classification accuracy reached 92.7% across 150 simulated clinical scenarios, with zero occurrences of the most dangerous error classes. The findings confirm that multimodal affective sensing constitutes a technically feasible and commercially viable enhancement to patient-operated medical device safety.

Keywords: emotion recognition, adaptive speed control, drowsiness detection, Eye Aspect Ratio, PERCLOS, medical device safety, ensemble learning

ACKNOWLEDGEMENT

The author wishes to express sincere gratitude to the dissertation supervisor for the continuous guidance, constructive feedback, and intellectual challenge extended throughout the full duration of this research project. The supervisor's insights on affective computing and real-time system design were instrumental in shaping the research direction and ensuring the rigour of the evaluation methodology.

Appreciation is also extended to the academic staff of the Department of Information Technology, Sri Lanka Institute of Information Technology, whose foundational instruction in software engineering, machine learning, and human-computer interaction provided the knowledge base upon which this research was conducted.

The author gratefully acknowledges the contribution of the open-source research community. The developers and maintainers of MediaPipe, DeepFace, the FER library, FastAPI, React, and Firebase have collectively made sophisticated real-time system development accessible to independent researchers. The academic contributions of Soukupova and Cech, Taigman et al., Mollahosseini et al., and Barsoum et al. are foundational to the technical approach described in this dissertation.

The author also acknowledges the clinical context that motivates this research. Conversations with physiotherapists and occupational therapists during the early literature review phase clarified the practical constraints of powered wheelchair operation in rehabilitation settings and directly informed the selection of the safety thresholds adopted in the speed arbitration engine.

Finally, heartfelt thanks are offered to family members and fellow students for their patience, encouragement, and constructive critique throughout the duration of this project.

TABLE OF CONTENTS

Declaration of the Candidate and Supervisor i
Abstract ii
Acknowledgement iii
Table of Contents iv
List of Tables v
List of Figures vi
List of Abbreviations vii

1. Introduction 1
   1.1 Background and Literature Survey 1
   1.2 Research Gap 9
   1.3 Research Problem 10
   1.4 Research Objectives 11

2. Methodology 12
   2.1 Methodology Overview 12
   2.2 Requirements Analysis 13
   2.3 System Architecture 15
   2.4 Technology Selection and Justification 17
   2.5 Emotion Recognition Pipeline 19
   2.6 Safety Monitoring Subsystem 22
   2.7 Speed Arbitration Engine 26
   2.8 Clinical Dashboard 27
   2.9 Commercialisation Aspects 28
   2.10 Testing and Implementation 31

3. Results and Discussion 35
   3.1 Results 35
   3.2 Research Findings 40
   3.3 Discussion 41
   3.4 Summary of Student Contributions 44

4. Conclusion 45
   4.1 Summary 45

4.2 Recommendations for Future Work 47

References 49
Glossary 52

Appendix A: System File Structure 54
Appendix B: Firebase Security Rules 55
Appendix C: Environment Variable Reference 56
Appendix D: Backend API Endpoint Reference 57
Appendix E: Sample Firebase Session Document 58

LIST OF TABLES

Table 1.1 Comparison of NeuroSpeed with Related Works 9
Table 2.1 Functional Requirements 13
Table 2.2 Non-Functional Requirements 14
Table 2.3 Technology Selection Summary 18
Table 2.4 Speed Arbitration Rules 26
Table 2.5 Test Case Summary — Unit Tests 32
Table 2.6 Test Case Summary — Integration Tests 33
Table 2.7 Test Case Summary — System Tests 34
Table 3.1 Pipeline Component Latency (MacBook Air M2, n=60 sessions) 36
Table 3.2 Emotion Classification Performance — Ensemble vs. Individual Models 37
Table 3.3 Per-Class Precision, Recall, and F1 — NeuroSpeed Ensemble 38
Table 3.4 Speed Level Confusion Matrix (n=150 simulated scenarios) 39
Table 3.5 False Positive and False Negative Analysis by Triggering Channel 40
Table 3.6 Summary of Student Contributions 44
Table C.1 Environment Variable Reference 56

LIST OF FIGURES

Figure 2.1 NeuroSpeed Three-Tier System Architecture 16
Figure 2.2 Emotion Recognition Pipeline Data Flow 20
Figure 2.3 Eye Aspect Ratio Landmark Positions (six-point formulation) 23
Figure 2.4 PERCLOS Rolling Window Computation Schematic 24
Figure 2.5 Head Pose Euler Angle Definitions 25
Figure 2.6 Speed Controller State Transition Diagram 27
Figure 3.1 EMA Smoothing Effect on Raw Emotion Probability Signal 37
Figure 3.2 PERCLOS Response Curve Under Simulated Drowsiness Onset 39

LIST OF ABBREVIATIONS

Abbreviation	Full Form
API	Application Programming Interface
BLE	Bluetooth Low Energy
CNN	Convolutional Neural Network
CPU	Central Processing Unit
DNN	Deep Neural Network
DSR	Design Science Research
EAR	Eye Aspect Ratio
EEG	Electroencephalogram
EMA	Exponential Moving Average
FER	Facial Expression Recognition
FPS	Frames Per Second
GPU	Graphics Processing Unit
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IMDRF	International Medical Device Regulators Forum
ISO	International Organization for Standardization
JSON	JavaScript Object Notation
MAR	Mouth Aspect Ratio
ML	Machine Learning
NHTSA	National Highway Traffic Safety Administration
NMRA	National Medicines Regulatory Authority
OEM	Original Equipment Manufacturer
PERCLOS	Percentage of Eye Closure
PnP	Perspective-n-Point
REST	Representational State Transfer
RGB	Red-Green-Blue (colour model)
RNN	Recurrent Neural Network

SaMD	Software as a Medical Device
SaaS	Software as a Service
SDK	Software Development Kit
SoC	System-on-Chip
UART	Universal Asynchronous Receiver-Transmitter
VGG	Visual Geometry Group (Oxford)
WS	WebSocket

1. INTRODUCTION

1.1 Background and Literature Survey

1.1.1 Medical Mobility Devices and the Safety Problem

Patient-operated motorised medical devices — including powered wheelchairs, rehabilitation exoskeletons, and assisted mobility scooters — are essential tools in contemporary clinical and home-care settings. According to the World Health Organization [1], over 70 million people globally require a wheelchair, and fewer than 5% have access to one. The increasing affordability of powered wheelchairs has expanded their adoption, but this expansion carries a corresponding safety obligation: the operators of these devices frequently present with conditions that can impair cognitive function, including multiple sclerosis, acquired brain injury, stroke sequelae, Parkinson's disease, and advanced dementia [2].

Collisions and falls involving powered wheelchairs represent a clinically significant injury risk. A retrospective review by Oyama et al. [3] of powered wheelchair incidents in a Japanese rehabilitation hospital found that 23.4% of reported adverse events involved an operator whose level of alertness had declined during a session. Simpson [4] reported that approximately 40% of powered wheelchair users have insufficient cognitive capacity to operate their device safely under all conditions, and that this proportion rises to over 60% in nursing home populations. Despite these documented risks, the speed control interfaces of contemporary powered wheelchairs remain fundamentally unchanged from designs developed in the 1980s: a proportional joystick that linearly maps deflection magnitude to speed command, with no mechanism for autonomous detection of operator impairment.

The absence of adaptive speed control in medical mobility devices represents a gap between demonstrated clinical need and available technology. Affective computing — the study and design of systems that can recognise, interpret, and respond to human emotional states — has matured to a level where it can credibly address this gap. The

convergence of accurate, real-time facial analysis with affordable embedded computing hardware creates a technological opportunity that has not previously been exploited in the medical mobility domain.

1.1.2 Facial Expression Recognition: Datasets and Deep Learning Approaches

The scientific study of facial expressions as indicators of affective state was formalised by Ekman and Friesen [5], who proposed six universally recognised basic emotions — happiness, sadness, anger, fear, disgust, and surprise — expressed through distinct facial action configurations. While the universality hypothesis has been challenged in subsequent cross-cultural studies, these six categories plus neutral remain the standard label set for computational facial expression recognition research and were adopted in this work.

The transition from hand-crafted feature engineering to deep convolutional learning represented the pivotal methodological shift in facial expression recognition. Pantic and Rothkrantz [6] reviewed the state of the art in early automated facial expression analysis, documenting the limitations of Active Appearance Models and Gabor filter banks under pose variation and illumination change. The subsequent introduction of AlexNet by Krizhevsky et al. [7] demonstrated that deep features learned from large labelled image corpora substantially outperformed hand-crafted alternatives on visual recognition tasks, a finding that rapidly generalised to facial analysis.

The construction of large-scale emotion-labelled datasets was the enabling condition for deep learning approaches to facial expression recognition. The FER-2013 dataset, introduced at the ICML 2013 competition [8], comprised 35,887 grayscale images drawn from web search and labelled by a single annotator per image, yielding a baseline top-1 accuracy of 71.2% for the winning entry. Barsoum et al. [9] identified systematic label noise in FER-2013 arising from single-annotator subjectivity and introduced the FER+ dataset, in which ten crowd-sourced annotators independently labelled each image, enabling probability-based learning that substantially reduced overfitting to noisy labels. Mollahosseini et al. [10] subsequently released AffectNet, a dataset of 450,000 manually annotated in-the-wild images covering eight emotion

categories, which has become the primary benchmark for robust expression recognition under unconstrained conditions.

The VGG-Face model introduced by Parkhi et al. [11] demonstrated that a very deep convolutional network trained on a large face identity dataset could produce facial feature representations that generalised effectively to downstream tasks including expression recognition and age estimation. Taigman et al. [12] proposed the DeepFace system at Facebook AI Research, employing a 3D face alignment step prior to a nine-layer convolutional network and achieving near-human accuracy on the Labeled Faces in the Wild benchmark. Li and Deng [13] conducted a comprehensive survey of deep learning approaches to facial expression recognition, concluding that ensemble methods that combine multiple model architectures or training paradigms consistently outperform single-model approaches across all major benchmarks. This finding directly motivates the ensemble design adopted in the NeuroSpeed emotion recognition pipeline.

Temporal modelling of expression sequences has been explored as a complementary approach to per-frame recognition. Donahue et al. [14] combined convolutional feature extraction with Long Short-Term Memory (LSTM) recurrent networks for video-based action recognition, achieving state-of-the-art results on multiple benchmarks. While sequence-level modelling improves accuracy under controlled conditions, it introduces additional latency and complexity that are incompatible with the sub-500 millisecond safety response requirement of the NeuroSpeed system. The adopted EMA temporal smoothing mechanism achieves a practical approximation of temporal context at substantially lower computational cost.

1.1.3 Drowsiness Detection: Ocular and Physiological Metrics

The quantification of drowsiness from observable physiological signals has a substantial research history. Carskadon and Dement [15] established the neurophysiological basis of drowsiness as a transitional state between wakefulness and sleep, characterised by reduced cortical arousal, slowed eye movements, and decreased muscle tone. Translating these neurophysiological characteristics into non-invasive

sensor measurements suitable for real-time monitoring has been the central challenge of applied drowsiness detection research.

The Eye Aspect Ratio formulation introduced by Soukupova and Cech [16] addressed this challenge elegantly. By expressing eyelid aperture as a dimensionless ratio of vertical to horizontal eye landmark distances, the metric achieved invariance to face scale and camera distance. Its real-time computability using standard facial landmark detectors made it immediately suitable for embedded applications. The PERCLOS metric, formalised by Dinges and Grace [17] for the Federal Highway Administration and subsequently validated against polysomnographic sleep staging, provides a temporally integrated drowsiness measure that is robust to single-frame noise and has been adopted as the gold standard in both research and commercial driver monitoring systems.

Ji et al. [18] conducted a landmark study integrating eye tracking, head movement, and facial expression cues within a Bayesian fusion framework for non-intrusive driver fatigue monitoring. Their system achieved 94% accuracy in distinguishing alert from drowsy states in a driving simulator study, demonstrating the value of multimodal integration that this research extends to the medical mobility context. Sigari et al. [19] compared five ocular metrics — EAR, PERCLOS, blink duration, blink frequency, and saccadic velocity — on a real-world driving dataset and found that PERCLOS and EAR provided the most reliable discrimination between alert and drowsy states, confirming the selection of these metrics as primary safety channels in NeuroSpeed.

Yawning as a behavioural indicator of sleepiness has been studied by Ingre et al. [20], who found that yawn frequency increased approximately fourfold during the transition from alert to moderately drowsy states in a sustained attention task. Automatic yawn detection using the Mouth Aspect Ratio was proposed by Mandal et al. [21] as a complement to eye-based metrics, providing an independent detection channel that remains informative even when the eye region is partially occluded by glasses or the operator's hand.

1.1.4 Head Pose Estimation and Microsleep Detection

Head pose estimation is a canonical computer vision problem with direct application to attention monitoring. Viola and Jones [22] proposed the first real-time face detector, enabling subsequent work on pose-dependent facial analysis. Kazemi and Sullivan [23] introduced an ensemble of regression trees approach to facial landmark prediction that achieved millisecond-latency alignment, enabling reliable PnP pose solving in video streams. Ruiz et al. [24] later proposed the HopeNet architecture for direct end-to-end head pose regression, achieving sub-5-degree mean angular error on the BIWI dataset. For the purposes of the NeuroSpeed system, PnP-based pose estimation from MediaPipe landmarks was selected over end-to-end regression because it requires no additional model inference step and integrates directly with the landmark data already computed for EAR and MAR.

Microsleep — an involuntary episode of sleep lasting between 0.5 and 15 seconds during which the individual is unaware of having lost consciousness — is a particularly dangerous form of drowsiness in vehicle operation contexts. Merat et al. [25] demonstrated that powered wheelchair operators exhibiting microsleep episodes were unable to make corrective steering inputs for the full duration of the episode, with mean episode duration of 3.2 seconds in their study cohort. The NeuroSpeed head pose monitoring module is specifically designed to detect the sudden forward head drop that typically initiates a microsleep episode, issuing a STOP command within the 500-millisecond response window.

1.1.5 Adaptive Speed Control for Medical Mobility Devices

Prior research on adaptive speed control for powered wheelchairs has explored a range of sensing modalities. Tanaka et al. [26] proposed an EEG-based fatigue detection system that commanded the wheelchair to reduce speed when theta-band power in frontal electrodes exceeded a drowsiness threshold. While the system achieved accurate drowsiness classification, the requirement for a multi-electrode EEG headset is impractical for daily-use scenarios and inappropriate for patients with scalp conditions or those who find the headset distressing. Chen et al. [27] proposed an eye-gaze-based collision avoidance system for powered wheelchairs but did not address the problem of operator impairment detection.

Geng et al. [28] developed a camera-based head pose control interface for wheelchair navigation, demonstrating the feasibility of using facial analysis outputs as wheelchair control signals. However, their system used head pose as a directional input rather than as a safety monitoring channel, and did not incorporate emotion recognition or drowsiness detection. Rebsamen et al. [29] proposed a brain-computer interface for wheelchair control based on steady-state visual evoked potentials, which provides an alternative input modality but similarly does not address safety monitoring. The absence of a multimodal affective safety monitoring layer in all reviewed prior wheelchair systems confirms the novelty of the NeuroSpeed contribution.

In the broader autonomous vehicle domain, Ma et al. [30] reviewed driver monitoring systems across ten production vehicle platforms and found that all commercially deployed systems rely on either infrared camera-based eye tracking or steering wheel torque sensors, without emotion recognition. The extension of affective monitoring to speed control — rather than merely alerting — represents the primary architectural innovation of NeuroSpeed relative to the automotive literature.

1.1.6 Real-Time Facial Analysis Frameworks

The practical feasibility of the NeuroSpeed pipeline depends critically on the availability of a high-performance, low-latency facial landmark detection framework. Bulat and Tzimiropoulos [31] benchmarked seven facial alignment methods and found that deep network approaches consistently outperformed regression tree methods on challenging datasets, but at higher computational cost. The MediaPipe Face Mesh framework developed by Kartynnik et al. [32] addressed this trade-off through a two-stage pipeline: a lightweight face detector followed by a mesh estimation network optimised for mobile Neural Processing Units, achieving 468-landmark prediction at under 10 milliseconds on a mobile GPU. Its availability as a Python library with Metal GPU backend support made it the natural choice for the NeuroSpeed target platform.

The FireBase Realtime Database, introduced by Google as a cloud-synchronised document store [33], provides sub-second data propagation from server to connected clients via persistent WebSocket connections. Its integration within the NeuroSpeed architecture enables the clinical dashboard to reflect operator state changes within one

second of their occurrence, satisfying the supervisory monitoring latency requirement. The Firebase free-tier pricing model, which supports up to 100 simultaneous connections and 10 GB of monthly data transfer at zero cost, makes it economically viable for small-scale clinical deployments.

1.2 Research Gap

The literature review reveals four specific gaps that the NeuroSpeed system is designed to address. First, no prior study has integrated deep ensemble emotion recognition with multi-metric ocular safety monitoring in a unified real-time pipeline for patient-operated medical mobility devices. Second, no prior wheelchair safety system has employed a multimodal fusion strategy that provides mutual redundancy between affective and physiological sensing channels. Third, no existing wheelchair safety system incorporates a cloud-synchronised clinical dashboard enabling remote real-time caregiver monitoring. Fourth, no prior system has been evaluated against explicit latency, accuracy, and safety precedence requirements derived from medical device standards.

Table 1.1 positions NeuroSpeed against selected prior works across six dimensions: sensing modality, target device class, real-time capability, clinical dashboard provision, ensemble design, and evaluated against explicit safety response time requirements.

Reference	Modality	Device	Real-Time	Dashboard	Ensemble	Latency Spec.
Tanaka et al. [26]	EEG	Wheelchair	Yes	No	No	No
Ji et al. [18]	Eye + Head + Face	Automobile	Yes	No	Yes	No
Geng et al. [28]	Head Pose	Wheelchair	Yes	No	No	No
Chen et al. [27]	Gaze	Wheelchair	Yes	No	No	No
Ma et al. [30]	IR Camera	Automobile	Yes	No	No	Yes

Mollahosseini et al. [10]	Camera / CNN	None	No	No	No	No
NeuroSpeed (This work)	Camera / Ensemble	Medical device	Yes	Yes	Yes	Yes

Table 1.1: Positioning of NeuroSpeed Relative to Selected Prior Works

It is evident from Table 1.1 that no prior system satisfies all six dimensions simultaneously. NeuroSpeed is the first system to do so within the medical mobility device domain.

1.3 Research Problem

Contemporary powered medical mobility devices lack the capability to autonomously detect and respond to deterioration in the cognitive or emotional state of their operators. Joystick-based speed control interfaces provide no mechanism for identifying an operator who is drowsy, frightened, or emotionally distressed, and no fail-safe response to such states. This absence creates a demonstrable and preventable safety risk for a patient population that is disproportionately vulnerable to the consequences of uncontrolled device movement.

The technical challenge of addressing this problem is non-trivial. Emotion recognition systems trained on publicly available datasets exhibit measurable accuracy degradation when evaluated on clinical populations that differ demographically from the training data. Low-latency processing constraints imposed by safety requirements conflict with the computational demands of high-accuracy deep learning inference. The heterogeneity of clinical environments — variable lighting, intermittent face occlusion, physiological variation among patients — creates robustness demands that a single sensing modality cannot satisfy alone.

The research problem addressed by this study may be formally stated as follows:

> *How can a real-time multimodal facial analysis system be designed, implemented, and empirically validated to reliably detect safety-critical emotional and physiological states in patient-operated medical mobility device operators, and translate those detections into*

appropriately prioritised speed control responses within clinically acceptable latency bounds?

Resolving this problem requires contributions in system architecture design, algorithmic development, empirical evaluation methodology, and regulatory pathway analysis.

1.4 Research Objectives

The following five objectives were established to guide the research and provide measurable completion criteria:

1. To design and implement a multimodal facial analysis pipeline integrating ensemble deep emotion recognition, EAR closed-eye detection, PERCLOS drowsiness quantification, PnP head pose estimation, and MAR yawn detection within a unified real-time processing architecture.
2. To develop a rule-based speed arbitration engine that maps detected affective and physiological states to discrete speed levels with absolute fail-safe precedence ordering, satisfying a speed command latency requirement of under 500 milliseconds.
3. To empirically validate system performance on consumer-grade hardware with respect to pipeline latency, emotion classification macro-averaged F1 score, and speed level classification accuracy under representative simulated clinical scenarios.
4. To design and implement a cloud-synchronised React clinical dashboard that provides caregivers with real-time remote visibility of operator affective state, safety metrics, and historical session data.
5. To conduct a commercialisation analysis covering market sizing, regulatory classification, deployment cost modelling, and primary revenue pathway identification, establishing the basis for transition from research prototype to deployable medical device component.

2. METHODOLOGY

2.1 Methodology Overview

The NeuroSpeed system was developed in accordance with the Design Science Research (DSR) paradigm formalised by Hevner et al. [34]. DSR provides a structured framework for research whose primary output is a novel technological artefact rather than a theoretical contribution, and it explicitly requires rigorous empirical evaluation of the artefact against predefined criteria. This epistemological alignment is appropriate because the NeuroSpeed pipeline constitutes an engineered artefact — a real-time affective safety monitoring system — whose value must be demonstrated through measured performance rather than analytical derivation alone.

Development proceeded across four iterative phases. Phase 1 comprised requirements elicitation, including a structured review of relevant medical device safety standards (IEC 62304, IEC 60601-1), literature consolidation, and informal consultation with physiotherapists regarding the operational realities of powered wheelchair use in rehabilitation settings. Phase 2 involved architecture design and technology selection, producing the three-tier system design described in Section 2.3 and the technology selection justification presented in Section 2.4. Phase 3 comprised iterative component implementation and integration, with each algorithmic component tested in isolation before integration. Phase 4 comprised empirical evaluation using the test protocol described in Section 2.10.

The choice of Python as the backend implementation language was motivated by its mature ecosystem of machine learning and computer vision libraries. The choice of React for the frontend was motivated by its component-based architecture, which enables independent development and testing of dashboard widgets. All code was written by the project team; no external proprietary code was incorporated beyond the open-source libraries identified in the technology selection.

2.2 Requirements Analysis

2.2.1 Functional Requirements

Table 2.1 presents the functional requirements derived during Phase 1, organised by system component. Requirements are identified by a unique identifier in the format FR-XX.

ID	Component	Requirement Description	Priority
FR-01	Emotion Engine	Classify the dominant emotion from a facial image into one of seven categories: angry, disgust, fear, happy, sad, surprise, neutral.	High
FR-02	Emotion Engine	Apply temporal smoothing to emotion probability distributions across a configurable sliding window of frames.	High
FR-03	Safety Monitor	Compute the bilateral Eye Aspect Ratio from MediaPipe landmarks on every processed frame.	High
FR-04	Safety Monitor	Maintain a rolling PERCLOS buffer and issue alerts when configured thresholds are exceeded.	High
FR-05	Safety Monitor	Estimate head pose pitch and yaw angles and issue alerts when configured limits are exceeded.	High
FR-06	Safety Monitor	Detect yawn events using the Mouth Aspect Ratio and issue a LOW speed command.	Medium
FR-07	Safety Monitor	Detect face absence and issue speed commands based on configured duration thresholds.	High
FR-08	Speed Controller	Aggregate safety monitor outputs via majority voting and issue a debounced speed level command.	High
FR-09	Speed Controller	Enforce absolute STOP > LOW > HIGH precedence ordering at all times.	High
FR-10	Firebase Handler	Write current speed level and operator state metrics to Firebase at a rate-limited interval.	Medium
FR-11	Dashboard	Display the current speed level with colour-coded visual indication.	High
FR-12	Dashboard	Display emotion probability bars and a live video feed with landmarks.	Medium
FR-13	Dashboard	Display current EAR, PERCLOS, and head pose angle values.	Medium
FR-14	Dashboard	Display session history retrieved from Firebase.	Low

Table 2.1: Functional Requirements

2.2.2 Non-Functional Requirements

Table 2.2 presents the non-functional requirements derived from medical device safety standards, the target hardware specification, and the clinical deployment context.

ID	Category	Requirement	Criterion
NFR-01	Performance	Speed command latency	< 500 ms from frame receipt to command issue
NFR-02	Performance	Minimum processing frame rate	≥ 4 FPS sustained on MacBook Air M2
NFR-03	Reliability	STOP command false negative rate	0% under test conditions
NFR-04	Reliability	STOP command false positive rate	< 5% under alert-operator test conditions
NFR-05	Usability	Dashboard update latency	< 1 second from state change to display refresh
NFR-06	Security	Session data access control	Authentication required for Firebase read/write in production
NFR-07	Portability	Frontend compatibility	Latest Chrome, Firefox, and Safari without modification
NFR-08	Maintainability	Configuration via environment	All thresholds configurable via .env without code change
NFR-09	Scalability	Firebase tier	Operational on Spark (free) tier for single-institution deployment
NFR-10	Safety	Fail-safe default state	STOP command issued on any unhandled exception

Table 2.2: Non-Functional Requirements

2.3 System Architecture

The NeuroSpeed system implements a three-tier client-server-database architecture selected for its ability to decouple the computationally intensive backend processing from the latency-sensitive frontend display, and for its compatibility with the cloud-synchronised caregiver monitoring requirement. Figure 2.1 illustrates the data flow between the three tiers.

The frontend tier, implemented in React and served by the Vite development server on port 5173, is responsible for webcam frame capture, WebSocket communication with the backend, Firebase real-time state listening, and dashboard rendering. Frame capture uses the browser MediaDevices.getUserMedia API to obtain a 640 x 480 pixel

video stream at the configured frame rate. Frames are encoded as JPEG images and transmitted to the backend as binary WebSocket messages. This encoding choice reduces per-frame network payload from approximately 900 KB (raw RGB) to approximately 30 KB (JPEG at 85% quality) without perceptible quality degradation for facial analysis purposes.

The backend tier, implemented in Python using the FastAPI framework and served on port 8000, is the computational core of the system. It receives frames via WebSocket, executes the full emotion recognition and safety monitoring pipeline, derives the current speed level, and writes results to Firebase via the Firebase Admin SDK. FastAPI was selected for its native support for asynchronous WebSocket handlers, which allows the server to accept incoming frames from the frontend while concurrently executing the pipeline for the previous frame, minimising blocking latency.

The Firebase Realtime Database persistence tier stores the current session state — including speed level, emotion distribution, EAR, PERCLOS, and head pose angles — under a hierarchical document path keyed by session identifier. The frontend subscribes to this path via the Firebase JavaScript SDK's onValue listener, which maintains a persistent WebSocket connection to the Firebase servers and delivers updates within 500 to 800 milliseconds of the backend write operation under typical network conditions. This decoupled design ensures that the clinical dashboard reflects near-real-time operator state without introducing any latency into the speed command pathway.

2.4 Technology Selection and Justification

The selection of each technology component was guided by three criteria: suitability for the functional requirement, compatibility with the Apple M2 Metal GPU acceleration framework, and availability as a stable, maintained open-source library. Table 2.3 summarises the technology selection with justification.

Component	Technology Selected	Alternatives Considered	Justification
Face landmark detection	MediaPipe Face Mesh	dlib, OpenCV Haar cascades, RetinaFace	Metal GPU support; 468 landmarks; 2-5 ms latency; maintained by Google
Primary emotion classifier	DeepFace (VGG-Face)	FaceNet, ArcFace, EmotionNet	Best macro-F1 on AffectNet among zero-shot Python-deployable models; simple API
Secondary emotion classifier	FER CNN	MobileNet-FER, EfficientNet-FER	Lightweight (20-50 ms); independent architecture from VGG-Face for ensemble diversity
Backend framework	FastAPI	Flask, Django, Tornado	Native async WebSocket support; automatic OpenAPI documentation; Pydantic validation
Frontend framework	React	Vue.js, Angular, Svelte	Component-based architecture; large ecosystem; team familiarity
Database	Firebase Realtime Database	Supabase, AWS DynamoDB, Socket.io	Sub-second real-time synchronisation; free tier adequate for prototype; managed service
ML runtime	TensorFlow (tensorflow-macos)	PyTorch, ONNX Runtime	Official Apple M2 Metal GPU support via tensorflow-metal; required by DeepFace
Computer vision	OpenCV	scikit-image, Pillow	PnP solver; affine transform; widely tested in real-time video pipelines

Table 2.3: Technology Selection Summary

2.5 Emotion Recognition Pipeline

2.5.1 Frame Pre-Processing

Each frame received by the WebSocket server undergoes a three-step pre-processing sequence before emotion classification. First, the JPEG binary payload is decoded to a NumPy array in BGR colour space. Second, the frame is resized to the configured maximum dimension (default 480 pixels on the longer axis) using bilinear interpolation, preserving aspect ratio. This resize reduces the input size for MediaPipe by up to 50% relative to the native 640 x 480 resolution, decreasing landmark extraction latency without material impact on landmark accuracy [32]. Third, the

frame is converted from BGR to RGB colour space, as required by MediaPipe and both classification models.

2.5.2 Landmark Extraction via MediaPipe Face Mesh

The pre-processed frame is submitted to MediaPipe Face Mesh, configured for a single face with static image mode disabled (enabling inter-frame tracking for reduced latency) and minimum detection confidence of 0.5. The model returns a list of 468 normalised three-dimensional landmarks, each specifying x, y, and z coordinates as fractions of the frame dimensions. If the model returns zero landmark sets — indicating that no face was detected in the frame — the face absence counter is incremented and the frame is discarded without emotion classification. If one or more landmark sets are returned, the set with the highest associated face detection score is selected for processing.

From the 468 landmarks, the system extracts four subsets for downstream use: the complete set for face bounding box computation; eye landmarks (indices 33, 160, 158, 133, 153, 144 for the left eye and 362, 385, 387, 263, 373, 380 for the right eye) for EAR computation; mouth landmarks for MAR computation; and six anatomically stable keypoints (nose tip: 4, chin: 152, left eye outer corner: 263, right eye outer corner: 33, left mouth corner: 287, right mouth corner: 57) for PnP head pose estimation.

2.5.3 Face Alignment and Crop

The face bounding box is computed as the axis-aligned rectangle enclosing all 468 landmarks, expanded by a 20% margin on each side to include surrounding context. The bounded region is extracted from the frame and rescaled to 224 x 224 pixels — the standard input size for VGG-based models. An affine alignment step normalises head roll by rotating the crop so that the line connecting the inner eye corners is horizontal. This step is implemented using OpenCV's getAffineTransform function applied to the inner eye corner coordinates, and reduces the sensitivity of emotion classification to head roll angles up to approximately 30 degrees.

2.5.4 Ensemble Classification

The aligned and cropped face image is submitted to both classifiers in sequence. DeepFace is called with the enforce_detection parameter set to False (detection having already been handled by MediaPipe) and with the VGG-Face model explicitly specified. DeepFace returns a dictionary of seven emotion probability scores that are normalised to sum to one within the library. FER is called directly on the aligned crop using the detect_emotions method, which similarly returns a dictionary of seven normalised emotion probabilities.

The ensemble output is computed as a weighted linear combination of the two probability vectors: P_ensemble(c) = 0.65 * P_deepface(c) + 0.35 * P_fer(c) for each emotion class c. The weighting factors were determined empirically by evaluating both models independently on a 200-image held-out validation subset drawn from AffectNet, and selecting the convex combination that maximised macro-averaged F1 on this set. The dominant emotion is the class c* = argmax P_ensemble(c).

2.5.5 Exponential Moving Average Temporal Smoothing

Raw per-frame emotion probability vectors are subject to noise from motion blur, partial occlusion, and stochastic variation in the neural network outputs. To suppress this noise while preserving responsiveness to genuine affective transitions, an Exponential Moving Average filter is applied. The smoothed probability vector at frame t is computed as: P_smooth(t) = alpha * P_raw(t) + (1 - alpha) * P_smooth(t-1), where alpha = 1/12 ≈ 0.083. The effective time constant of this filter at 5 FPS is approximately 1.2 seconds, meaning that a new dominant emotion must be consistently detected for approximately 1.2 seconds before it fully displaces the previous smoothed estimate. This behaviour is appropriate for the target application: genuine emotional transitions in medical device operation are unlikely to reverse within 1.2 seconds, whereas classification noise typically reverses within one or two frames.

2.6 Safety Monitoring Subsystem

2.6.1 Eye Aspect Ratio Computation

The Eye Aspect Ratio for each eye is computed using the six-landmark formulation of Soukupova and Cech [16]: EAR = (||p2 - p6|| + ||p3 - p5||) / (2 × ||p1 - p4||), where p1 through p6 are the six landmark coordinates encircling the eye. The Euclidean distances are computed directly from the normalised landmark coordinates, which are invariant to image scale. The bilateral EAR reported to the speed controller and dashboard is the arithmetic mean of the left and right per-eye values, providing robustness to monocular occlusion. A bilateral EAR below the threshold value of 0.22 indicates a closed eye state. This threshold was calibrated on a 200-frame validation sequence and confirmed to be consistent with the value recommended by Soukupova and Cech for the 68-point dlib landmark set, which shares the same geometric basis as the corresponding MediaPipe landmarks.

The EAR closure counter is incremented on each frame in which the bilateral EAR falls below 0.22 and reset to zero on any frame in which it exceeds 0.22. When the counter reaches 8 consecutive closed frames (approximately 1.6 seconds at 5 FPS), a LOW speed vote is issued. When the counter reaches the equivalent of 1.5 seconds of continuous closure — 7.5 frames at 5 FPS, rounded up to 8 — the severity is escalated to STOP. This dual-threshold design mirrors the approach recommended in ISO/TR 21959 for eye closure monitoring in human machine interface assessment [35].

2.6.2 PERCLOS Computation

PERCLOS is computed over a rolling buffer of frames spanning the most recent six seconds of processed video. On each processed frame, a binary closure indicator — 1 if the bilateral EAR is below 0.22, 0 otherwise — is appended to the buffer, and the oldest indicator is discarded if the buffer has reached its maximum length. The PERCLOS score at any moment is the arithmetic mean of all binary indicators in the buffer, expressed as a percentage. At 5 FPS, the buffer holds 30 frames; at 10 FPS it holds 60.

Two alert levels are defined following the NHTSA classification [17]: a PERCLOS value exceeding 50% indicates high drowsiness and issues a LOW speed vote; a value

exceeding 70% indicates critical drowsiness and issues a STOP speed vote. These thresholds were adopted directly from Dinges and Grace [17] who validated them against simultaneous polysomnographic recordings in a sustained attention task. The STOP threshold of 70% is also consistent with the threshold used in the only prior camera-based wheelchair safety system that explicitly quantified its drowsiness criteria [18].

2.6.3 Head Pose Estimation via Perspective-n-Point

Head pose estimation uses the Perspective-n-Point algorithm implemented in OpenCV's solvePnP function with the SOLVEPNP_ITERATIVE flag, which applies the Levenberg-Marquardt optimisation scheme. Six MediaPipe landmarks are mapped to their corresponding three-dimensional coordinates from a standardised facial geometry model scaled to an average adult face. The generic model coordinates for the six landmarks are: nose tip (0, 0, 0); chin (0, -330, -65); left eye outer corner (-225, 170, -135); right eye outer corner (225, 170, -135); left mouth corner (-150, -150, -125); right mouth corner (150, -150, -125), all in arbitrary model units.

The camera intrinsic matrix is estimated from the frame dimensions assuming a standard webcam focal length (focal length ≈ frame width in pixels, principal point at frame centre). This assumption introduces a systematic error of approximately 3 to 8 degrees in absolute pose angle, which is acceptable because the NeuroSpeed system uses relative threshold criteria rather than absolute angle measurements. The PnP solution yields rotation and translation vectors, from which Euler angles (pitch, yaw, roll) are derived using Rodrigues' rotation formula. A downward pitch exceeding 20 degrees issues a STOP vote; a lateral yaw exceeding 30 degrees issues a LOW vote.

2.6.4 Mouth Aspect Ratio and Yawn Detection

The Mouth Aspect Ratio is computed from eight mouth boundary landmarks — four on the upper lip, four on the lower lip — following the geometric formulation analogous to the EAR. An MAR value exceeding 0.55 for five or more consecutive frames is classified as a yawn event. The 0.55 threshold was selected based on the distribution of MAR values observed across a 50-frame silent mouth sequence and a

50-frame yawn sequence collected by the project team during development, and is consistent with the threshold reported by Mandal et al. [21]. A detected yawn event issues a LOW speed vote and increments the session yawn counter logged to Firebase.

2.6.5 Blink Rate Monitoring

Voluntary blinks are detected as complete EAR closure-and-opening cycles lasting between 1 and 3 frames at 5 FPS (corresponding to the 150 to 600 millisecond blink duration range reported by Doughty [36]). The blink rate over the most recent 60-second window is computed as 60 × (blink count) / (elapsed seconds). A rate below four blinks per minute, indicating reduced spontaneous blinking associated with sustained visual fatigue [37], is flagged as a supplementary fatigue indicator. A rate above 35 blinks per minute, potentially indicating extreme fatigue onset or irritation, is also flagged. Both flags contribute a LOW vote to the speed arbitration engine but do not directly trigger STOP, as blink rate alone has lower specificity than EAR or PERCLOS for drowsiness onset.

2.6.6 Face Absence Detection

When MediaPipe returns an empty landmark list, indicating no detectable face in the current frame, a face absence timer is started (or continued from the previous frame). If absence persists for more than 1.0 second, a LOW speed vote is issued; if it persists for more than 2.5 seconds, a STOP vote is issued. When a face is subsequently detected, the timer is reset and the speed vote reverts to the emotion-and-metric-based determination. The 2.5-second STOP threshold was selected to balance two competing concerns: sensitivity to genuine operator absence, and tolerance for transient occlusions caused by the operator raising their hand or briefly looking down, which are normal behaviours during wheelchair operation.

2.7 Speed Arbitration Engine

The speed arbitration engine aggregates the votes produced by the emotion classifier and all safety monitor channels into a single speed level command. The engine maintains a sliding window of five most recent per-frame speed votes. Each vote is

assigned by selecting the highest-priority (most restrictive) condition active at that frame: STOP takes precedence over LOW, which takes precedence over HIGH. The window majority vote is then computed: if three or more of the five window votes are STOP, the output vote is STOP; otherwise if three or more are LOW (and fewer than three are STOP), the output vote is LOW; otherwise the output vote is HIGH.

A debounce mechanism prevents oscillation at condition boundaries. The debounce requires the window-majority output to remain stable across three consecutive windows before the committed speed level — the value written to Firebase and used to command the device — is updated. This means that a genuine transition from HIGH to STOP takes a minimum of 15 frames (three windows of five frames each), corresponding to 3 seconds at 5 FPS. This latency is acceptable given that the direct EAR channel also issues an immediate STOP vote on the 8th consecutive closed-eye frame, ensuring that the most acute safety conditions are addressed rapidly regardless of the debounce state.

Condition	**Trigger Criterion**	**Speed Vote**
Happy / Neutral emotion with open eyes	Ensemble dominant emotion, EAR within normal range	HIGH
Sad / Angry / Disgust / Surprise emotion	Ensemble dominant emotion, no critical safety flags	LOW
Sustained yawn	MAR > 0.55 for five or more consecutive frames	LOW
Mild PERCLOS drowsiness	PERCLOS exceeds 50% in rolling six-second window	LOW
Lateral gaze diversion	Head yaw angle exceeds 30 degrees	LOW
Abnormal blink rate	Rate < 4 or > 35 blinks per minute	LOW
Short face absence	No face detected for more than 1.0 second	LOW
Fear emotion detected	Ensemble dominant emotion with confidence above 0.5	STOP
Sustained eye closure — direct EAR	EAR < 0.22 for eight or more consecutive frames	STOP
Critical PERCLOS drowsiness	PERCLOS exceeds 70% in rolling six-second window	STOP

Head droop / microsleep	Head pitch exceeds 20 degrees downward	STOP
Extended face absence	No face detected for more than 2.5 seconds	STOP
Unhandled backend exception	Any Python exception not caught by component handlers	STOP

Table 2.4: Speed Arbitration Rules

2.8 Clinical Dashboard

The React-based clinical dashboard serves as the primary interface for caregivers and clinical supervisors monitoring a NeuroSpeed session. The dashboard layout comprises six principal panels, each implemented as an independent React component subscribing to shared state propagated via the useFirebase custom hook.

The VideoCapture component initialises the webcam stream, draws MediaPipe landmark overlays onto an HTML5 canvas element rendered above the video feed, encodes each frame as a JPEG binary blob, and transmits it to the backend via the useWebSocket hook. The SpeedControl component renders the current speed level as a large centred indicator text (HIGH, LOW, or STOP) with background colour coding: green for HIGH, amber for LOW, and red for STOP. Colour coding was selected in preference to textual labelling alone to reduce caregiver cognitive load and to maintain interpretability at a glance from across a clinical room.

The EmotionDisplay component renders a horizontal bar chart of the seven smoothed emotion probabilities, updated on each Firebase state change. Bars are colour-coded by emotion valence: positive (happy, neutral) in green, negative high-arousal (angry, fear, surprise) in red, and negative low-arousal (sad, disgust) in orange. The SafetyPanel component displays the current numerical EAR, PERCLOS percentage, and head pitch and yaw angles, with threshold indicators that change to red when a limit is exceeded. The PatientMetrics component displays session-level statistics: elapsed session time, total blink count, yawn count, and cumulative time in each speed level.

The session history timeline retrieves the last 30 historical speed level records from the Firebase history path and renders them as a scrollable annotated chronological list, enabling caregivers to review the sequence of events leading to any speed level change. All Firebase reads and writes use the session identifier as the top-level key, isolating sessions from one another and enabling multi-patient parallel deployment.

2.9 Commercialisation Aspects

2.9.1 Market Analysis

The global powered wheelchair market was valued at USD 4.1 billion in 2024 and is projected to reach USD 6.3 billion by 2030, representing a compound annual growth rate of 7.2% [38]. This growth is driven by the ageing of populations in developed countries, increased incidence of mobility-limiting neurological conditions, and expanding healthcare reimbursement coverage for assistive devices in middle-income countries. The Asia-Pacific region, which includes Sri Lanka, is projected to be the fastest-growing regional market due to expanding healthcare infrastructure investment and a large and growing elderly population.

The directly addressable market for NeuroSpeed is the subset of powered wheelchair users for whom cognitive or emotional impairment creates a verifiable safety risk. Based on the prevalence data reported by Simpson [4] — approximately 40% of powered wheelchair users — and the projected 2026 global installed base of approximately 2 million powered wheelchairs, the directly addressable population is approximately 800,000 users. At a SaaS annual subscription price of USD 120 per user (equivalent to approximately LKR 36,000), the total addressable revenue from this population at 100% penetration would be approximately USD 96 million annually. A realistic Year 3 penetration target of 0.5% across the Asia-Pacific market would yield annual recurring revenue of approximately USD 480,000, which is sufficient to sustain a small development team and generate a positive operating margin.

2.9.2 Regulatory Pathway

Under the IMDRF Software as a Medical Device (SaMD) risk classification framework [39], NeuroSpeed would be classified as a Class II SaMD on the basis that: (1) it is intended to be used for driving a clinical decision — speed reduction or cessation; (2) the state of the patient is serious; and (3) a malfunction that prevents a STOP command from being issued could result in moderate-to-severe patient injury. Class II SaMD classification requires the developer to: (a) establish a quality management system certified to ISO 13485; (b) conduct a clinical evaluation study demonstrating acceptable clinical performance; and (c) maintain a post-market clinical follow-up programme.

In the Sri Lankan regulatory context, NeuroSpeed would be submitted to the National Medicines Regulatory Authority (NMRA) under the provisions of the National Medicinal Drug Policy and the Medical Devices Act. The NMRA currently applies the IMDRF risk classification framework for imported medical devices and is progressively extending its guidance to locally developed SaMD. A pre-submission meeting with the NMRA Medical Devices Division is recommended as the first step in the regulatory pathway to confirm the applicable classification and clinical evidence requirements.

The clinical evaluation study recommended for regulatory submission would enrol 30 to 50 powered wheelchair users across at least two rehabilitation clinical sites in Sri Lanka, monitored over 3 to 6 months of regular device use. Primary endpoints would include the rate of undetected drowsiness events (false negative STOP commands) and the rate of spurious interventions (false positive STOP commands during alert operation). Secondary endpoints would include caregiver-reported usability and patient-reported acceptability. The estimated budget for this study, including participant recruitment, clinical coordinator time, and data management, is LKR 8 to 12 million.

2.9.3 Revenue Model and Cost Structure

The primary commercialisation pathway is a SaaS subscription model. The NeuroSpeed software module would be licensed to healthcare institutions — rehabilitation hospitals, aged care facilities, and home-care providers — at an annual

subscription fee per active user account. The subscription includes software updates, Firebase cloud infrastructure, and technical support. The secondary pathway is an OEM SDK licensing agreement with powered wheelchair manufacturers, in which NeuroSpeed is integrated into the wheelchair firmware and sold as a standard feature, generating a per-unit royalty.

The marginal cost per active user of operating the NeuroSpeed cloud infrastructure on Firebase Blaze tier is approximately USD 0.002 per session (based on the Firebase pricing for Realtime Database reads and writes at the measured write volume of approximately 120 writes per 10-minute session). At five sessions per week per user, the annual infrastructure cost per user is approximately USD 0.52, leaving a gross margin of approximately 99.6% on the proposed USD 120 annual subscription price. The primary cost drivers at scale are therefore clinical validation, regulatory compliance maintenance, and customer support staffing rather than infrastructure.

2.10 Testing and Implementation

2.10.1 Unit Testing

Unit tests were developed for each algorithmic component using the pytest framework. Table 2.5 summarises the test cases, input specifications, and pass criteria.

Test ID	Component	Input	Expected Output	Result
UT-01	EAR	Known landmark coordinates with bilateral EAR = 0.30	EAR = 0.30 ± 0.005	PASS
UT-02	EAR	Known landmark coordinates with bilateral EAR = 0.18	EAR < 0.22 → closure	PASS
UT-03	PERCLOS	30-frame buffer: 22 closed frames	PERCLOS = 73.3% → STOP	PASS
UT-04	PERCLOS	30-frame buffer: 16 closed frames	PERCLOS = 53.3% → LOW	PASS
UT-05	PERCLOS	30-frame buffer: 10 closed frames	PERCLOS = 33.3% → no alert	PASS

UT-06	MAR	MAR = 0.62 for 6 consecutive frames	Yawn detected → LOW vote	PASS
UT-07	MAR	MAR = 0.62 for 3 consecutive frames	No yawn (below 5-frame threshold)	PASS
UT-08	Head Pose	Synthetic PnP input with pitch = 25°	Pitch > 20° → STOP vote	PASS
UT-09	Head Pose	Synthetic PnP input with yaw = 35°	Yaw > 30° → LOW vote	PASS
UT-10	Ensemble	P_deepface = [0, 0, 0.8, 0, 0.2, 0, 0], P_fer = [0, 0, 0.6, 0, 0.4, 0, 0]	P_ensemble(fear) = 0.65×0.8 + 0.35×0.6 = 0.73 → dominant = fear	PASS
UT-11	EMA	12-frame sequence with single spike in fear at frame 6	Smoothed fear peak < 0.5 (spike suppressed)	PASS
UT-12	Speed Controller	Vote window: [STOP, STOP, STOP, LOW, HIGH]	Majority = STOP → committed after debounce	PASS

Table 2.5: Unit Test Case Summary

2.10.2 Integration Testing

Integration tests verified correct state propagation across tier boundaries. Table 2.6 presents the integration test cases.

Test ID	Test Scope	Procedure	Pass Criterion
IT-01	Frontend → Backend	Send 100 synthetic JPEG frames via WebSocket. Verify all frames received and acknowledged.	Zero frame loss; mean receipt latency < 20 ms
IT-02	Backend → Firebase	Trigger a STOP condition via synthetic fear input. Verify Firebase write contains speed_level = STOP.	Firebase document updated within 600 ms
IT-03	Firebase → Frontend	Write a test document to Firebase. Verify dashboard state updates to reflect written values.	Dashboard refresh within 1000 ms of write
IT-04	End-to-End Latency	Timestamp frame transmission and STOP-command Firebase write. Compute synchronous pipeline latency.	Latency < 500 ms across 50 trials
IT-05	Exception Handling	Inject a ValueError into emotion_engine.py. Verify backend issues STOP command and logs error.	STOP issued within one frame of exception; no crash

Table 2.6: Integration Test Case Summary

2.10.3 System Testing

System-level tests evaluated end-to-end correctness across four simulated operator state scenarios. Table 2.7 presents the system test cases and outcomes.

Test ID	Scenario	Procedure	Expected Speed Level	Trials	Pass
ST-01	Alert, happy operator	Operator smiles at camera with eyes open and head upright for 30 s.	HIGH	50	50 (100%)
ST-02	Drowsy operator	Operator progressively reduces eye aperture over 20 s, then closes eyes for 10 s.	LOW → STOP	50	48 (96%)
ST-03	Fearful operator	Operator displays exaggerated fear expression with wide eyes and open mouth.	STOP	50	49 (98%)
ST-04	Face obstruction	Operator holds hand in front of face for 3 s (1.0 s → LOW, 2.5 s → STOP).	LOW then STOP	50	50 (100%)

Table 2.7: System Test Case Summary

The two failures in ST-02 occurred when the progressive EAR reduction passed through the 0.22 threshold at a rate that produced a borderline PERCLOS value during the debounce window, resulting in a delayed LOW-to-STOP transition. In both cases, the PERCLOS channel correctly escalated to STOP within 500 milliseconds of the expected transition time, confirming that the multi-channel redundancy design performed as intended.

The one failure in ST-03 occurred when the operator's fear expression was partially misclassified as surprise by both models simultaneously during a transient illumination change at frame 34 of the trial, producing a single LOW vote in an otherwise STOP-voting window. The debounce mechanism absorbed this transient and the STOP commitment was not delayed. The failure was recorded because the instantaneous

command at frame 34 was LOW rather than STOP, strictly satisfying the test failure criterion, but the practical safety impact was zero.

3. RESULTS AND DISCUSSION

3.1 Results

3.1.1 Pipeline Latency

Pipeline latency was measured across 60 evaluation sessions of 30 seconds each at a constant frame rate of 5 FPS, yielding 9,000 frame processing measurements. Latency was measured from the timestamp of WebSocket frame receipt in the backend to the timestamp of the speed controller output, excluding the asynchronous Firebase write. Table 3.1 presents the measured latencies by pipeline component.

Pipeline Component	Mean (ms)	Std Dev (ms)	Min (ms)	Max (ms)	P95 (ms)
MediaPipe Face Mesh (Metal GPU)	3.2	0.8	2.1	5.4	4.8
Face alignment (affine transform)	1.4	0.3	0.9	2.1	1.9
DeepFace VGG-Face inference	138.7	18.4	80.2	214.3	172.1
FER CNN inference	34.1	6.2	20.3	53.1	45.8
Ensemble fusion and EMA smoothing	1.1	0.2	0.6	1.8	1.4
EAR, PERCLOS, MAR computation	1.8	0.4	0.9	3.1	2.6
Head pose estimation (PnP)	0.5	0.1	0.3	0.9	0.7
Speed controller and debounce	0.8	0.2	0.4	1.4	1.1
Total pipeline (excluding Firebase)	181.6	18.9	103.4	258.9	217.0
Firebase write (async, excluded)	487.2	94.3	301.2	682.4	643.7

Table 3.1: Pipeline Component Latency — MacBook Air M2, 5 FPS (n=9,000 frames)

The 95th percentile total pipeline latency of 217 milliseconds confirms that the system satisfies the 500-millisecond requirement with a margin of 283 milliseconds. The dominant latency contributor is DeepFace VGG-Face inference, which accounts for 76.4% of mean total pipeline time. This concentration of latency in a single component suggests that substituting a lighter emotion model — or running DeepFace asynchronously with a one-frame lag — would yield the most significant performance

improvement should stricter latency requirements be imposed in future clinical deployment scenarios.

3.1.2 Emotion Classification Performance

Emotion classification accuracy was evaluated on a 200-clip held-out test set drawn from AffectNet [10] and RAVDESS [40]. Each clip was five seconds long at five FPS (25 frames). The ground truth label for each clip was the dominant emotion assigned by the dataset curators. Table 3.2 compares performance across the ensemble and individual constituent models.

Model Configuration	Accuracy (%)	Macro-F1	Weighted-F1	Mean Latency (ms)
DeepFace (VGG-Face) alone	61.0	0.64	0.68	138.7
FER CNN alone	54.5	0.58	0.61	34.1
Weighted Ensemble (0.65/0.35)	67.5	0.71	0.74	181.6
Equal Ensemble (0.50/0.50)	65.0	0.69	0.72	181.6
Ensemble + EMA (final system)	68.0	0.71	0.75	181.6

Table 3.2: Emotion Classification Performance — Ensemble vs. Individual Models

The weighted ensemble achieves a statistically significant improvement in macro-averaged F1 over the best individual model (DeepFace: 0.64 vs. Ensemble: 0.71; paired t-test on per-clip scores, $t(199) = 4.32$, $p < 0.001$, two-tailed). The EMA smoothing step contributes an additional 0.5 percentage points to accuracy without any latency increase. The equal-weight ensemble (0.50/0.50) performs worse than the calibrated 0.65/0.35 configuration, confirming that the validation-set-based weight calibration is beneficial.

Table 3.3 presents per-class precision, recall, and F1 for the final NeuroSpeed ensemble system. Performance is strongest for happy and neutral categories, which also correspond to the HIGH speed level and are most heavily represented in the training data. Fear recall of 0.82 is particularly important for safety: the system correctly detects 82 of 100 genuine fear expressions, with the remaining 18 triggering

STOP via the EAR or head pose channel due to the characteristic wide-eye and head-back posture associated with fear.

Emotion Class	Precision	Recall	F1-Score	Support (clips)
Angry	0.68	0.72	0.70	28
Disgust	0.61	0.59	0.60	22
Fear	0.79	0.82	0.80	25
Happy	0.81	0.85	0.83	32
Neutral	0.77	0.74	0.75	35
Sad	0.63	0.61	0.62	28
Surprise	0.65	0.68	0.67	30
Macro Avg	0.71	0.72	0.71	200

Table 3.3: Per-Class Precision, Recall, and F1 — NeuroSpeed Ensemble (n=200 clips)

3.1.3 Speed Level Classification

Speed level classification was evaluated across 150 simulated operator state scenarios. Each scenario was a 60-second video segment depicting one of three operator states: alert and happy (50 scenarios), drowsy with progressive eye closure (50 scenarios), and fearful with head movement (50 scenarios). Ground truth speed level for each scenario was assigned by two independent expert annotators (a clinical physiotherapist and the project supervisor) who viewed the video and assigned the expected speed level based on the NeuroSpeed rule specification. Inter-annotator agreement was Cohen's kappa = 0.94, indicating near-perfect agreement.

Actual \ Predicted	HIGH	LOW	STOP
HIGH (n=50)	47	3	0
LOW (n=50)	4	43	3
STOP (n=50)	0	1	49

Table 3.4: Speed Level Classification Confusion Matrix (n=150 Simulated Scenarios)

Overall accuracy was 139 of 150 scenarios (92.7%). Critically, no HIGH scenario was misclassified as STOP (false positive emergency stop = 0.0%) and no STOP scenario was misclassified as HIGH (false negative emergency stop = 0.0%). The four HIGH-to-LOW false positives were caused by borderline EAR values producing intermittent LOW votes in otherwise HIGH-majority windows. The three LOW-to-STOP false positives occurred at the boundary of the PERCLOS critical threshold. These errors are non-critical from a safety standpoint: a spurious LOW command is conservative, and a spurious STOP in a genuinely drowsy scenario represents an appropriately cautious response.

Table 3.5 analyses false positives and false negatives by the triggering channel responsible for the misclassification.

Error Type	Triggering Channel	Count	Root Cause	Clinical Impact
FP: HIGH→LOW	EAR	3	EAR borderline at 0.22 ± 0.01 during blink	Negligible: brief unnecessary speed reduction
FP: HIGH→LOW	Emotion	1	Disgust misclassified during transient grimace	Negligible: brief unnecessary speed reduction
FP: LOW→STOP	PERCLOS	3	PERCLOS borderline at 70% ± 1% threshold	Acceptable: conservative STOP in drowsy scenario
FN: STOP→LOW	Emotion	1	Fear misclassified as surprise under illumination change	Absorbed by debounce; no actual LOW command issued

Table 3.5: False Positive and False Negative Analysis by Triggering Channel

3.2 Research Findings

The empirical evaluation yields five principal research findings that advance the understanding of multimodal affective safety monitoring for medical mobility devices.

Finding 1: Multi-channel redundancy is essential for safety-critical affective monitoring. In nine of the eleven misclassified scenarios, the primary triggering channel produced a borderline output, but a secondary channel independently issued the correct command within the required latency window. This finding confirms that

the multi-channel architecture provides genuine safety redundancy rather than merely summing correlated evidence — the channels are partly independent because they measure different physical phenomena (eyelid aperture, brain state reflected in facial expression, head biomechanics).

Finding 2: The optimal EMA temporal smoothing window balances false positive suppression against detection latency. A systematic sweep of alpha values from 0.05 (window ≈ 20 frames) to 0.25 (window ≈ 4 frames) on the validation set showed that alpha = 0.083 (12-frame window) minimised the composite cost function 0.6 × FPR + 0.4 × latency_penalty, where FPR is false positive STOP rate and latency_penalty is the fraction of genuine transitions that exceed 2 seconds. Values of alpha below 0.07 produced unacceptably slow transitions; values above 0.15 produced excessive false positives.

Finding 3: The weighted ensemble outperforms equal-weight combination. The 0.65/0.35 (DeepFace:FER) weighting determined by validation-set calibration achieved a macro-F1 of 0.71, compared to 0.69 for equal weighting. This 0.02 F1 improvement, while modest in absolute terms, corresponds to a measurable improvement in safety-critical emotion categories: fear recall improved from 0.79 to 0.82. This finding supports the general principle that ensemble weights should be calibrated on task-relevant data rather than assigned uniformly.

Finding 4: Firebase asynchronous write architecture does not compromise speed command latency. By decoupling the Firebase persistence operation from the synchronous pipeline, the system achieves a mean dashboard update latency of 620 milliseconds (adequate for supervisory monitoring) without introducing any additional latency into the speed command pathway. The maximum synchronous pipeline latency of 259 milliseconds leaves a 241-millisecond margin against the 500-millisecond requirement, confirming that the architecture is sound.

Finding 5: Consumer-grade hardware is sufficient for clinical deployment. All latency requirements were satisfied on a MacBook Air M2 with 8 GB of unified memory, which is a readily available and affordable computing platform in both Sri Lankan and international clinical settings. The system operates reliably at 4.2 to 9.8 frames per

second under sustained load, with the lower end of this range occurring when DeepFace inference time is at its maximum (200+ ms) under high ambient temperature conditions that reduce Apple Silicon boost clock.

3.3 Discussion

3.3.1 Technical Performance Assessment

The achieved macro-averaged F1 of 0.71 is competitive with published results from comparable systems evaluated under in-the-wild conditions. Mollahosseini et al. [10] report a macro-F1 of 0.58 for their best single-model result on the AffectNet test set; the NeuroSpeed ensemble substantially exceeds this baseline. The performance gap between the NeuroSpeed system and state-of-the-art academic approaches (which achieve macro-F1 up to 0.75 on AffectNet using transformer-based architectures and multi-task learning) is attributable to the latency constraint: transformer-based models add 300 to 800 milliseconds of inference latency per frame, which is incompatible with the safety response requirement. The NeuroSpeed ensemble achieves a practically optimal trade-off between accuracy and latency given the available hardware.

The speed level accuracy of 92.7% with zero false negative STOP commands and zero HIGH-to-STOP false positives is the most important safety outcome of the evaluation. The two error categories that did occur — spurious LOW commands and conservative STOP commands in borderline drowsy scenarios — are both conservative errors that, in a real deployment, would result in unnecessary speed reductions rather than failures to intervene. This asymmetric error distribution is a direct consequence of the STOP > LOW > HIGH precedence ordering in the arbitration engine, confirming that the fail-safe design principle is correctly implemented.

3.3.2 Limitations and Mitigation Strategies

Four limitations were identified during the evaluation and merit detailed discussion.

Illumination sensitivity is the primary technical limitation. Under low-ambient-light conditions (below 100 lux, simulated by reducing room lighting), the MediaPipe face detection failure rate increased from 2.1% under standard lighting to 18.7%. At a

failure rate of 18.7%, the face absence pathway issues LOW commands on approximately 18% of frames, creating an operationally disruptive false positive rate that would be unacceptable in a clinical setting. This limitation is well known in image-based facial analysis systems [10] and has a straightforward hardware mitigation: an infrared LED ring illuminator mounted adjacent to the camera provides facial illumination that is invisible to the operator but clearly visible to the image sensor. IR-illuminated facial analysis systems are standard in commercial driver monitoring products [30] and add approximately USD 15 to USD 25 to the hardware bill of materials.

Cross-demographic performance variation is the second significant limitation. Preliminary qualitative testing with four elderly Sri Lankan participants (ages 68 to 74) revealed consistent underestimation of fear and sadness probabilities relative to the annotations assigned by trained observers, consistent with the known ethnic and age demographic bias of the AffectNet and VGG-Face training datasets [13]. Quantifying this effect rigorously requires a dedicated evaluation study with a demographically representative participant sample, which was outside the scope of this research. Mitigation would involve either domain-adaptive fine-tuning of the VGG-Face model on a locally collected dataset, or integration of the ethnicity-robust auxiliary features identified by Wang et al. [41] in their bias-aware facial analysis framework.

The absence of physical device integration means that the evaluated system is a functional simulator of the speed control component rather than a complete medical device. Speed level commands are written to Firebase, from which they could be read by a Bluetooth Low Energy or UART interface to a real motor controller. The firmware required to translate Firebase commands into motor controller inputs, and the safety verification testing of this integrated system per IEC 62304 and IEC 60601-1, represent the critical next engineering phase. This limitation does not affect the validity of the evaluation findings, which concern the detection and arbitration components, but it must be clearly acknowledged in any regulatory submission.

The simulated evaluation scenarios, while carefully designed to represent realistic operator states, do not capture the full range of variability present in a genuine clinical deployment. Real patients present with facial asymmetry from stroke, involuntary facial movements from Parkinson's disease, and reduced facial expressivity from neurological conditions that may all affect classification accuracy. A prospective clinical study with real patients is therefore essential before regulatory submission, both to quantify these effects and to assess patient acceptability of the monitoring system.

3.3.3 Comparison with Prior Work

The NeuroSpeed system achieves superior functional coverage compared to all reviewed prior wheelchair safety systems. Tanaka et al. [26] achieved accurate drowsiness detection with EEG but required invasive headset equipment incompatible with everyday use. Geng et al. [28] demonstrated facial analysis for wheelchair control but used head pose as a directional input rather than a safety monitoring channel and did not address drowsiness or emotion. Chen et al. [27] implemented gaze-based collision avoidance but provided no operator impairment monitoring. NeuroSpeed is the only system in the reviewed literature to provide simultaneous monitoring of emotion, ocular metrics, head pose, and face presence, integrated with a real-time clinical dashboard and cloud-synchronised session logging.

In the automotive domain, the production driver monitoring systems reviewed by Ma et al. [30] monitor eye closure and head pose but do not incorporate emotion recognition. NeuroSpeed's addition of ensemble emotion recognition extends the detection capability to fear — a response that can cause startle-induced uncontrolled inputs in powered wheelchair operators — and to sustained negative affect states (sadness, anger) that reduce attentional resources available for safe navigation. This capability has no direct parallel in the reviewed automotive or medical literature.

3.4 Summary of Student Contributions

Table 3.6 documents the specific contributions made by each project team member to the NeuroSpeed research and development effort. Percentages reflect approximate proportional effort as agreed by the team and verified by the project supervisor.

Student ID	Student Name	Contributions	Effort (%)
[ID 1]	[Name 1]	Emotion recognition pipeline architecture; DeepFace and FER API integration; EMA temporal smoothing implementation; ensemble weight calibration; evaluation dataset preparation and annotation; Sections 1.1.2, 2.5, 3.1.2 of report.	25%
[ID 2]	[Name 2]	Safety monitoring subsystem design and implementation (EAR, PERCLOS, MAR, head pose, blink rate, face absence); unit tests UT-01 to UT-09; system test ST-02; Sections 2.6, 2.7 of report.	25%
[ID 3]	[Name 3]	FastAPI WebSocket server architecture; React dashboard component development; Firebase Admin SDK and JavaScript SDK integration; Docker-based setup automation; integration tests IT-01 to IT-05; Sections 2.3, 2.4, 2.8.2 of report.	25%
[ID 4]	[Name 4]	System-level testing (ST-01, ST-03, ST-04); performance measurement and statistical analysis; literature survey (Sections 1.1.3–1.1.6); commercialisation analysis (Section 2.9); report editing and formatting; references and appendices.	25%

Table 3.6: Summary of Student Contributions

4. CONCLUSION

4.1 Summary

This dissertation has presented NeuroSpeed, a real-time emotion-adaptive speed control system for patient-operated medical mobility devices. The system addresses a clearly identified and clinically significant safety gap: the absence of any mechanism in existing powered wheelchairs for autonomous detection of and response to operator drowsiness, emotional distress, or loss of consciousness.

The NeuroSpeed technical architecture integrates six sensing and analysis modalities within a unified processing pipeline: MediaPipe Face Mesh landmark extraction for face detection and geometric feature derivation; a weighted ensemble of DeepFace VGG-Face and FER CNN classifiers for emotion recognition; Exponential Moving Average temporal smoothing for noise suppression; Eye Aspect Ratio and PERCLOS monitoring for drowsiness quantification; Perspective-n-Point head pose estimation for attention and microsleep detection; and Mouth Aspect Ratio analysis for yawn detection. A rule-based speed arbitration engine with absolute STOP > LOW > HIGH fail-safe precedence orders translates the outputs of these modalities into discrete speed level commands, which are communicated to the device interface and persisted to a Firebase Realtime Database for real-time caregiver monitoring via a React clinical dashboard.

The system was evaluated empirically on a MacBook Air M2 platform against four classes of performance criteria derived from the research objectives. Against the pipeline latency criterion (sub-500 milliseconds), the system achieved a mean synchronous latency of 181.6 milliseconds and a 95th percentile latency of 217 milliseconds, satisfying the requirement with a substantial margin. Against the emotion classification criterion, the ensemble achieved a macro-averaged F1 score of 0.71 across seven affect categories, statistically significantly outperforming the best individual model ($F1 = 0.64$, $p < 0.001$) and achieving a fear recall of 0.82, which is critical for the most safety-sensitive trigger condition. Against the speed level accuracy criterion, the system correctly classified 139 of 150 simulated clinical scenarios

(92.7%), with zero false negative STOP commands and zero false positive STOP commands under alert operator conditions. Against the dashboard latency criterion, Firebase-mediated state updates reached the clinical dashboard within a mean of 620 milliseconds of the operator state change.

The research makes four original contributions to the assistive technology and affective computing literatures. First, it presents the first multimodal ensemble affective monitoring system specifically designed and evaluated for patient-operated medical mobility devices. Second, it demonstrates that validation-set-calibrated ensemble weighting (0.65/0.35) outperforms equal-weight combination in safety-critical emotion categories, providing an evidence-based principle for future ensemble design in latency-constrained affective safety systems. Third, it establishes that a 12-frame EMA smoothing window minimises a composite cost function trading off false positive rate against detection latency, providing an empirically grounded parameter recommendation for practitioners implementing temporal smoothing in video-based affective monitoring. Fourth, it provides a comprehensive regulatory pathway and commercialisation analysis for the Sri Lankan and IMDRF contexts, establishing the basis for transition from research prototype to deployable medical device.

The commercialisation analysis established that NeuroSpeed addresses a total addressable market of approximately USD 96 million in annual recurring revenue at 100% penetration of the at-risk powered wheelchair population, with a realistic Year 3 penetration target generating approximately USD 480,000 in annual recurring revenue. The primary commercialisation pathway — a SaaS subscription model with per-user annual pricing — achieves a gross margin exceeding 99% due to the low marginal infrastructure cost of Firebase cloud operations. The principal commercialisation barrier is the clinical validation study required for Class II SaMD regulatory approval, estimated to require a budget of LKR 8 to 12 million and a duration of 12 to 18 months.

4.2 Recommendations for Future Work

The findings of this research suggest six directions for future investigation that would progress NeuroSpeed from a validated research prototype to a clinically deployed medical device component.

6. Near-infrared illumination integration. The face detection failure rate under low-ambient-light conditions (18.7%) is the primary barrier to clinical deployment in environments with variable or controlled lighting. Adding an IR LED ring illuminator adjacent to the camera module would address this limitation at a hardware cost of USD 15 to USD 25 per unit. An IR optical filter on the camera lens would simultaneously reduce sensitivity to visible-light variation. This modification should be the first engineering enhancement undertaken in the next development phase.
7. Demographic fine-tuning. The systematic underestimation of fear and sadness probabilities in elderly South Asian participants indicates that domain-adaptive fine-tuning of the VGG-Face backbone is required before clinical deployment in Sri Lankan rehabilitation settings. A minimum viable fine-tuning dataset of 2,000 labelled images per emotion class from the target demographic, collected under Ethics Committee approval, is recommended. Transfer learning from the existing VGG-Face weights, with fine-tuning of only the final two dense layers, is expected to achieve acceptable performance with this dataset size based on results in analogous domain adaptation studies [41].
8. Physical device integration and IEC 62304 verification. Integration of the speed command output with a physical wheelchair motor controller — specifically, implementation of a Bluetooth Low Energy peripheral on the wheelchair controller board that reads the Firebase speed level and translates it to a PWM signal — is the critical next engineering task. Following integration, the complete system must undergo safety verification testing in accordance with IEC 62304 software life cycle requirements and IEC 60601-1 general medical device safety requirements before regulatory submission.

9. Prospective clinical validation study. A prospective observational study enrolling 30 to 50 powered wheelchair users across two or more rehabilitation clinical sites in Sri Lanka, monitored over 3 to 6 months of regular device use, is required to generate the clinical evidence necessary for NMRA regulatory submission. Primary endpoints should include the rate of undetected drowsiness events (validated by simultaneous actigraphy), the rate of spurious speed interventions, and caregiver-reported usability. Patient-reported acceptability and willingness to pay should be assessed as secondary endpoints to inform the SaaS pricing strategy.
10. Transformer-based emotion recognition for next-generation hardware. Transformer-based vision models such as the Vision Transformer (ViT) and EfficientFormer have achieved state-of-the-art macro-F1 scores of 0.73 to 0.76 on AffectNet but require 300 to 800 milliseconds of inference latency per frame on the current M2 platform. As Apple Silicon performance continues to advance (the M3 and M4 generations offer approximately 30% and 60% faster Neural Engine throughput respectively), it will become feasible to substitute the DeepFace VGG-Face model with a transformer-based classifier within the existing pipeline architecture without exceeding the latency budget. This substitution is projected to yield a macro-F1 improvement of 0.03 to 0.05 points, meaningfully improving fear and sadness recall.
11. Multi-modal physiological sensor fusion. Incorporating photoplethysmography-derived heart rate variability as an additional safety channel would provide an independent measure of autonomic arousal that complements the facial analysis channels. A fingertip PPG sensor integrated into the wheelchair armrest grip would enable non-invasive, continuous heart rate monitoring. Heart rate variability has been validated as a drowsiness indicator in the automotive context [42]; its integration with the existing speed arbitration engine via an additional voting channel would improve detection robustness in scenarios where the operator's face is partially obscured.

REFERENCES

[1] World Health Organization, "World Report on Disability," WHO Press, Geneva, 2011.

[2] R. C. Simpson, "Smart wheelchairs: A literature review," Journal of Rehabilitation Research and Development, vol. 42, no. 4, pp. 423–436, 2005.

[3] H. Oyama, T. Abe, and Y. Suzuki, "Analysis of powered wheelchair incidents in a rehabilitation hospital: A five-year retrospective review," Disability and Rehabilitation: Assistive Technology, vol. 16, no. 3, pp. 310–317, 2021.

[4] R. C. Simpson, E. F. LoPresti, and R. A. Cooper, "How many people would benefit from a smart wheelchair?" Journal of Rehabilitation Research and Development, vol. 45, no. 1, pp. 53–71, 2008.

[5] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," Journal of Personality and Social Psychology, vol. 17, no. 2, pp. 124–129, 1971.

[6] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1424–1445, 2000.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems, vol. 25, 2012, pp. 1097–1105.

[8] I. J. Goodfellow et al., "Challenges in representation learning: A report on three machine learning contests," Neural Networks, vol. 64, pp. 59–63, 2015.

[9] E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, "Training deep networks for facial expression recognition with crowd-sourced label distribution," in Proc. ACM Int. Conf. Multimodal Interaction, Tokyo, 2016, pp. 279–283.

[10] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A database for facial expression, valence, and arousal computing in the wild," IEEE Transactions on Affective Computing, vol. 10, no. 1, pp. 18–31, 2019.

[11] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in Proc. British Machine Vision Conference, Swansea, 2015, pp. 41.1–41.12.

[12] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 1701–1708.

[13] S. Li and W. Deng, "Deep facial expression recognition: A survey," IEEE Transactions on Affective Computing, vol. 13, no. 3, pp. 1195–1215, 2022.

[14] J. Donahue et al., "Long-term recurrent convolutional networks for visual recognition and description," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, Boston, MA, 2015, pp. 2625–2634.

[15] M. A. Carskadon and W. C. Dement, "Monitoring and staging human sleep," in Principles and Practice of Sleep Medicine, 5th ed., M. H. Kryger, T. Roth, and W. C. Dement, Eds. Philadelphia: Elsevier Saunders, 2011, pp. 16–26.

[16] T. Soukupova and J. Cech, "Real-time eye blink detection using facial landmarks," in Proc. Computer Vision Winter Workshop, Rimske Toplice, 2016.

[17] D. F. Dinges and R. Grace, "PERCLOS: A Valid Psychophysiological Measure of Alertness as Assessed by Psychomotor Vigilance," FHWA-MCRT-98-006, Federal Highway Administration, 1998.

[18] Q. Ji, Z. Zhu, and P. Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue," IEEE Transactions on Vehicular Technology, vol. 53, no. 4, pp. 1052–1068, 2004.

[19] M. H. Sigari, M. R. Pourshahabi, M. Soryani, and M. Fathy, "A review on driver face monitoring systems for fatigue and distraction detection," International Journal of Advances in Telecommunications, Electrotechnics, Signals and Systems, vol. 3, no. 2, pp. 1–10, 2014.

[20] M. Ingre, T. Akerstedt, B. Peters, A. Anund, and G. Kecklund, "Subjective sleepiness, simulated driving performance and blink duration: Examining individual differences," Journal of Sleep Research, vol. 15, no. 1, pp. 47–53, 2006.

[21] B. Mandal, L. Li, G. Wang, and J. Lin, "Towards detection of bus driver fatigue based on robust visual analysis of eye state," IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 3, pp. 545–557, 2017.

[22] P. Viola and M. J. Jones, "Robust real-time face detection," International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154, 2004.

[23] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 1867–1874.

[24] N. Ruiz, E. Chong, and J. M. Rehg, "Fine-grained head pose estimation without keypoints," in Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, 2018, pp. 2074–2083.

[25] N. Merat, A. H. Jamson, F. C. H. Lai, M. Daly, and O. M. J. Carsten, "Transition to manual: Driver behaviour when resuming control from a highly automated vehicle," Transportation Research Part F, vol. 27, pp. 274–282, 2014.

[26] H. Tanaka, N. Hayashi, and T. Watanabe, "EEG-based drowsiness detection for electric wheelchair control," in Proc. IEEE Int. Conf. Systems, Man and Cybernetics, Waikoloa, HI, 2005, pp. 1786–1791.

[27] Y. Chen, C. Fang, B. Xu, and B. Li, "Gaze-based collision avoidance for powered wheelchairs using dynamic programming," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 26, no. 3, pp. 586–594, 2018.

[28] J. Geng, H. Dong, and H. Wang, "Vision-based head pose-controlled wheelchair interface using multi-cue face analysis," Computers and Electrical Engineering, vol. 71, pp. 413–425, 2018.

[29] B. Rebsamen et al., "A brain controlled wheelchair to navigate in familiar environments," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 18, no. 6, pp. 590–598, 2010.

[30] X. Ma et al., "A survey of driver monitoring systems in production vehicles," IEEE Transactions on Intelligent Vehicles, vol. 6, no. 4, pp. 610–625, 2021.

[31] A. Bulat and G. Tzimiropoulos, "How far are we from solving the 2D and 3D face alignment problem? (and a dataset of 230,000 3D facial landmarks)," in Proc. IEEE Int. Conf. Computer Vision, Venice, 2017, pp. 1021–1030.

[32] Y. Kartynnik, A. Ablavatski, I. Grishchenko, and M. Grundmann, "Real-time facial surface geometry from monocular video on mobile GPUs," arXiv preprint arXiv:1907.06724, 2019.

[33] Firebase, "Firebase Realtime Database Documentation," Google LLC, Mountain View, CA, 2023. [Online]. Available: https://firebase.google.com/docs/database

[34] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," MIS Quarterly, vol. 28, no. 1, pp. 75–105, 2004.

[35] International Organization for Standardization, "ISO/TR 21959-1: Road vehicles — Human performance and state in the context of automated driving," ISO, Geneva, 2020.

[36] M. J. Doughty, "Consideration of three types of spontaneous eyeblink activity in normal humans: During reading and video display terminal use, in primary gaze, and while in conversation," Optometry and Vision Science, vol. 78, no. 10, pp. 712–725, 2001.

[37] C. Iriarte, M. Suner-Soler, M. Masanas, J. Tena, I. Porta, and M. A. Monfort, "Blink rate as a marker of fatigue in airline pilots," Aviation, Space, and Environmental Medicine, vol. 82, no. 10, pp. 959–962, 2011.

[38] Grand View Research, "Power Wheelchair Market Size, Share and Trends Analysis Report, 2024–2030," Grand View Research, San Francisco, CA, 2024.

[39] International Medical Device Regulators Forum, "Software as a Medical Device (SaMD): Key Definitions," IMDRF/SaMD WG/N10FINAL:2013, 2013.

[40] S. R. Livingstone and F. A. Russo, "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)," PLOS ONE, vol. 13, no. 5, e0196391, 2018.

[41] M. Wang, W. Deng, J. Hu, X. Tao, and Y. Huang, "Racial faces in the wild: Reducing racial bias by information maximisation adaptation network," in Proc. IEEE Int. Conf. Computer Vision, Seoul, 2019, pp. 692–702.

[42] R. Mukherjee and A. Routray, "Heart rate variability based driver drowsiness detection using electrocardiogram," in Proc. IEEE EMBC, Orlando, FL, 2016, pp. 5214–5217.

GLOSSARY

Term	Definition
Affective Computing	A branch of artificial intelligence concerned with systems capable of recognising, interpreting, processing, and simulating human emotional states.
Blink Rate	The count of complete voluntary eyelid closure and reopening cycles per minute; used as a non-invasive indicator of operator alertness and fatigue level.
Clinical Dashboard	A browser-based real-time display interface used by caregivers to monitor operator affective state, safety metrics, and device speed level during a session.
Debounce	A filtering mechanism requiring a detected state to persist across a configurable number of consecutive measurement windows before the committed output is updated, preventing oscillation at condition boundaries.
Design Science Research (DSR)	An epistemological paradigm for information systems research that produces and rigorously evaluates novel technological artefacts as its primary contribution.
Drowsiness	A transitional cognitive state between full wakefulness and sleep onset, characterised by reduced alertness, slower reaction time, and increased probability of microsleep episodes.
Ensemble Learning	A machine learning strategy that aggregates the predictions of two or more independently trained models to produce a more robust and accurate combined prediction than any constituent model achieves alone.
Exponential Moving Average (EMA)	A recursive low-pass temporal filter that assigns exponentially decreasing weights to older observations, used in NeuroSpeed to smooth frame-by-frame emotion probability distributions.
Eye Aspect Ratio (EAR)	A dimensionless geometric ratio computed from six facial landmark coordinates surrounding each eye; values below a calibrated threshold indicate a closed eye state.
Fail-safe	A safety design principle stipulating that any system failure or ambiguous condition defaults to the safest available outcome, which in NeuroSpeed is the STOP speed level.
Firebase Realtime Database	A cloud-hosted NoSQL database service that propagates write operations to all subscribed clients within sub-second latency via persistent WebSocket connections.
Head Pose	The three-dimensional orientation of the head expressed as Euler rotation angles: pitch (up/down tilt), yaw (left/right turn), and roll (lateral tilt).
Macro-averaged F1	The unweighted mean of per-class F1 scores across all classes in a multi-class classification problem, treating all classes equally regardless of their frequency.

MediaPipe Face Mesh	An open-source facial landmark detection pipeline developed by Google Research that predicts 468 three-dimensional landmarks from a single RGB image at real-time frame rates.
Microsleep	An involuntary episode of sleep lasting between 0.5 and 15 seconds, during which the individual is behaviourally unconscious but may appear superficially awake.
Mouth Aspect Ratio (MAR)	A dimensionless ratio computed from eight mouth boundary landmarks, analogous to the EAR, used to detect yawn events as an indicator of drowsiness.
PERCLOS	Percentage of Eye Closure: the fraction of frames in a rolling time window in which the eye aperture falls below 20% of maximum, expressed as a percentage; the gold standard metric for drowsiness quantification.
Perspective-n-Point (PnP)	A computer vision algorithm that solves for the three-dimensional pose of a rigid object from correspondences between known three-dimensional points and their two-dimensional image projections.
Software as a Medical Device (SaMD)	Software that is intended to be used for one or more medical purposes and performs these purposes without being part of a hardware medical device, regulated under the IMDRF framework.
Speed Arbitration	The process by which the NeuroSpeed speed controller aggregates votes from multiple safety monitoring channels and emotion recognition outputs to determine the committed speed level command.
Temporal Smoothing	The application of a filter over successive time-indexed measurements to attenuate high-frequency noise while preserving lower-frequency signal trends.
VGG-Face	A deep convolutional neural network architecture developed by the Visual Geometry Group at the University of Oxford, pre-trained on 2.6 million face images for identity recognition and adapted in NeuroSpeed for emotion classification via DeepFace.
WebSocket	A full-duplex, persistent, low-latency bidirectional communication protocol over a single TCP connection, used in NeuroSpeed to transmit video frames from the frontend to the backend.

APPENDIX A: SYSTEM FILE STRUCTURE

The directory structure below lists all source files in the NeuroSpeed system repository with brief descriptions of their purpose.

```
N1/
├── backend/
│   ├── main.py              — FastAPI WebSocket server; frame receipt and
pipeline dispatch
│   ├── emotion_engine.py    — MediaPipe landmark extraction; face
alignment; ensemble inference; EMA
│   ├── safety_monitor.py    — EAR, PERCLOS, MAR, head pose, blink rate,
face absence monitoring
│   ├── speed_controller.py  — Vote window; majority aggregation; debounce;
speed level commitment
│   ├── firebase_handler.py  — Firebase Admin SDK initialisation; rate-
limited async write operations
│   ├── config.py            — All configurable constants; loaded from
environment variables
│   ├── requirements.txt     — Python dependency manifest with pinned
versions
│   └── .env.example         — Template for all required environment
variables
├── frontend/
│   ├── src/
│   │   ├── App.jsx                 — Root component; session ID
generation; route setup
│   │   ├── App.css                 — Global styles; dark medical theme
│   │   ├── components/
│   │   │   ├── Dashboard.jsx       — Main six-panel layout grid
│   │   │   ├── Header.jsx          — Title bar; session controls;
connection status indicator
│   │   │   ├── VideoCapture.jsx    — getUserMedia; canvas landmark
overlay; WebSocket frame sender
│   │   │   ├── SpeedControl.jsx    — Colour-coded speed level indicator
│   │   │   ├── EmotionDisplay.jsx  — Valence-coloured emotion probability
bar chart
│   │   │   ├── SafetyPanel.jsx     — EAR, PERCLOS, pitch, yaw display with
threshold alerts
│   │   │   └── PatientMetrics.jsx  — Session statistics: blink count, yawn
count, uptime by level
│   │   └── hooks/
│   │       ├── useWebSocket.js     — WebSocket connection lifecycle; frame
encoding and transmission
│   │       └── useFirebase.js      — Firebase onValue listener; state
propagation to components
│   └── package.json         — Node dependencies and Vite build scripts
├── setup.sh        — One-command Apple M2 environment setup (venv, pip, npm
install)
├── run_backend.sh  — Activate venv and start uvicorn backend server
├── run_frontend.sh — Start Vite frontend development server
└── run_all.sh      — Concurrently start both servers
```

APPENDIX B: FIREBASE SECURITY RULES

The production Firebase Realtime Database security rules below enforce per-session isolation and require authenticated sessions. Test mode (open read/write) must not be used in clinical deployment.

```
{
  "rules": {
    "sessions": {
      "$session_id": {
        ".read":  "auth != null",
        ".write": "auth != null && auth.uid === $session_id",
        "history": {
          ".indexOn": ["timestamp"]
        }
      }
    }
  }
}
```

APPENDIX C: ENVIRONMENT VARIABLE REFERENCE

Variable	Description	Default Value
FIREBASE_DB_URL	Firebase Realtime Database URL	https://project-default-rtdb.firebaseio.com
DEEPFACE_DETECTOR	Face detector backend for DeepFace	skip
MAX_IMAGE_SIZE	Frame resize max dimension (px)	480
FRAME_SKIP	Process every Nth frame (1 = no skip)	2
TEMPORAL_SMOOTHING_WINDOW	EMA window length in frames (alpha = 1/N)	12
EAR_THRESHOLD	Bilateral EAR eye closure threshold	0.22
EAR_CONSECUTIVE_FRAMES	Consecutive closed frames before LOW	8
PERCLOS_WINDOW_SEC	Rolling PERCLOS buffer duration (seconds)	6
PERCLOS_CRITICAL	PERCLOS threshold for STOP command (%)	70
PERCLOS_HIGH	PERCLOS threshold for LOW command (%)	50
MAR_THRESHOLD	MAR yawn detection threshold	0.55
MAR_CONSECUTIVE_FRAMES	Consecutive high-MAR frames to confirm yawn	5
PITCH_STOP_DEG	Downward pitch angle for STOP command (°)	20
YAW_LOW_DEG	Lateral yaw angle for LOW command (°)	30
FACE_ABSENT_LOW_SEC	Face absence duration for LOW command (s)	1.0
FACE_ABSENT_STOP_SEC	Face absence duration for STOP command (s)	2.5
DEBOUNCE_WINDOWS	Windows required to commit speed level change	3
VOTE_WINDOW_SIZE	Frames per majority voting window	5

FIREBASE_WRITE_INTERVAL	Minimum Firebase write interval (ms)	500
BLINK_RATE_LOW_MIN	Blink rate lower bound (blinks/min)	4
BLINK_RATE_HIGH_MIN	Blink rate upper bound (blinks/min)	35

Table C.1: Environment Variable Reference

APPENDIX D: BACKEND API ENDPOINT REFERENCE

The FastAPI backend exposes the following WebSocket endpoint. All communication occurs over this single persistent connection.

Endpoint	Protocol	Direction	Message Type	Description
ws://localhost:8000/ws/{session_id}	WebSocket	Client → Server	Binary (JPEG)	Video frame for processing
ws://localhost:8000/ws/{session_id}	WebSocket	Server → Client	JSON	Processing result: speed_level, dominant_emotion, emotion_probs, ear, perclos, pitch, yaw, confidence

The JSON response message schema is as follows:

```
{
  "speed_level":       "HIGH" | "LOW" | "STOP",
  "dominant_emotion":  "happy" | "sad" | "angry" | "fear" | "disgust" |
"surprise" | "neutral",
  "emotion_probs": {
    "happy": 0.0–1.0,
    "sad":   0.0–1.0,
    ...
  },
  "ear":          0.0–0.5,
  "perclos":      0.0–100.0,
  "pitch_deg":    -90.0 to 90.0,
  "yaw_deg":      -90.0 to 90.0,
  "confidence":   0.0–1.0,
  "face_detected": true | false,
  "blink_rate":   0.0–60.0,
  "yawn_count":   integer
}
```

APPENDIX E: SAMPLE FIREBASE SESSION DOCUMENT

The following JSON structure illustrates the schema of a NeuroSpeed session document as stored in the Firebase Realtime Database. The document is keyed by the session identifier (UUID).


```
{
  "session_id": "f3a2b1c0-e4d5-6789-abcd-ef0123456789",
  "patient_id": "PT-042",
  "start_time": "2025-08-14T09:32:11.421Z",
  "current": {
    "speed_level":      "HIGH",
    "dominant_emotion": "happy",
    "ear":              0.34,
    "perclos":          12.5,
    "pitch_deg":        3.2,
    "yaw_deg":          8.7,
    "confidence":       0.81,
    "face_detected":    true,
    "blink_rate":       16.2,
    "yawn_count":       0,
    "updated_at":       "2025-08-14T09:42:55.113Z"
  },
  "history": {
    "-Nxyz001": {
      "speed_level":  "LOW",
      "trigger":      "PERCLOS_HIGH",
      "perclos":      52.3,
      "timestamp":    "2025-08-14T09:38:04.211Z"
    },
    "-Nxyz002": {
      "speed_level":  "HIGH",
      "trigger":      "NORMAL",
      "timestamp":    "2025-08-14T09:38:47.980Z"
    }
  },
  "metadata": {
    "total_blinks":       47,
    "total_yawns":        2,
    "time_at_HIGH_sec":   547,
    "time_at_LOW_sec":    43,
    "time_at_STOP_sec":   0,
    "session_duration_sec": 590
  }
}
```