IT21294334 - Final Report.pdf

by Isum Abeywardana

Submission date: 30-Aug-2025 04:08AM (UTC+0530)
Submission ID: 2737928411
File name: IT21294334_-_Final_Report.pdf (853.36K)
Word count: 14828
Character count: 94823

AI-Assisted Criminal Investigations: Enhancing Testimony Analysis and Case Correlation

Multimodal Behavioural and Physiological Analysis (MBPA)

R25-020

Final Report

IT21294334

Abeywardana B. A. I. S.

The dissertation was submitted in partial fulfilment of the requirements for the B.Sc. (Honors) degree in Information Technology)

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

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 part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as article or books)

Signature:  Date: 29/08/2025

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

Signature of the supervisor: Date:

Acknowledgement

I would like to express my deepest gratitude to my supervisor, Dr. Lakmini Abeywardhana, for her invaluable guidance, encouragement, and support throughout the course of my undergraduate research. Her insights and expertise have been crucial to the successful completion of this project.

I am also sincerely grateful to my external supervisor, Mr. Dammika Kasthuri Mudalige, retired Chief Inspector of Police, whose advice and support have been instrumental in shaping the direction of this research. His willingness to assist and provide thoughtful feedback has greatly enhanced the quality of this work.

I would also thank the CI Prasanna Jayaratne and officers of Women and Children's Affairs Office, Wekunugoda who served as resource persons for this project.

Lastly, I would like to thank my family and friends for their unwavering support and encouragement throughout this journey. Their understanding and patience have been a constant source of strength.

Abstract

The accuracy and reliability of criminal investigations are often constrained by the limitations of human judgment during interrogations, where stress, fatigue, and subjective bias can influence outcomes. To address these challenges, this research proposes an AI-assisted investigation system enhanced by a dedicated component, Multimodal Behavioural and Physiological Analysis (MBPA), which integrates facial emotion recognition, speech-based emotion analysis, and heart rate variability monitoring to deliver real-time behavioural insights. The MBPA module leverages convolutional neural networks (CNNs) trained on the FER-2013 dataset for facial expression analysis, a CNN–BiLSTM hybrid with attention mechanisms trained on the RAVDESS dataset for vocal emotion recognition, and ECG data captured through an AD8232 single-lead heart rate monitoring sensor for physiological stress detection. By fusing these modalities, the system generates a comprehensive behavioural profile of suspects, detecting stress, deception, or discomfort with improved robustness compared to unimodal approaches. The integrated architecture provides investigators with a real-time dashboard that visualizes emotional fluctuations, vocal stress indicators, and heart rate irregularities, offering objective and quantifiable insights to support testimony evaluation. Experimental results demonstrate the system's effectiveness, with the voice model achieving 79.95% test accuracy and the facial emotion model achieving ~63% test accuracy, while heart rate analysis provided consistent physiological markers of arousal. The MBPA component, in conjunction with other system modules such as dynamic question generation and case correlation, enhances interrogation support by reducing human error, increasing investigative accuracy, and enabling evidence-based decision-making in criminal justice.

List of Figures

Figure 1 - System Overview Diagram 24
Figure 2 - Confusion Matrix (FER CNN) 50
Figure 3 - Training and Validation Loss/Accuracy (FER CNN) 50
Figure 4 - Confusion Matrix (SER CNN–BiLSTM) 51
Figure 5 - Training and Validation Loss/Accuracy (SER CNN–BiLSTM) 51

List of Tables

Table 1 - Comparative Analysis of Prior Research vs Proposed MBPA.....11
Table 2 - Datasets Considered for MBPA 20

Table of Contents

Declaration 2

Acknowledgement 3

Abstract 4

List of Figures 5

List of Tables 6

Table of Contents 7

1. Introduction 9

1.1. Background & Literature Survey 9

1.2. Research Gap 10

1.3. Research Problem 12

1.4. Research Objectives 13

2. Methodology 16

2.1. Requirements Gathering 16

2.2. Feasibility Study 21

2.3. System Design 24

2.4. Architectural Rationale 29

2.5. Implementation 33

2.6. System Requirements 37

2.7. Domain Specific Considerations 42

3. Testing & Results 49

3.1. Testing Strategy 49

3.2. Facial Emotion Recognition (FER) Results 50

3.3. Speech Emotion Recognition (SER) Results 51

3.4. Heart Rate Variability (HRV) Analysis Results 52
3.5. Multimodal Fusion Results 52
3.6. Dashboard Testing 52
3.7. Overall System Evaluation 53
4. Commercialization 54
5. Conclusion 58
6. References 61

1. Introduction

1.1. Background & Literature Survey

Criminal investigations have long relied on investigators' ability to interpret behavioural cues such as facial expressions, vocal tone, and physiological responses to assess the credibility of suspect testimonies. However, these traditional practices are often undermined by subjectivity, investigator fatigue, and cognitive bias, resulting in inconsistencies, misinterpretations, and in some cases, miscarriages of justice [1], [2]. Conventional lie detection technologies, such as polygraph machines and standalone voice stress analyzers, have also demonstrated limited accuracy, as they are prone to noise, countermeasures, and narrow unimodal focus [3]. Consequently, law enforcement agencies and researchers are increasingly turning towards Artificial Intelligence (AI) and multimodal behavioural analytics to develop objective, quantifiable, and real-time interrogation support systems [4], [5].

Recent advances in facial emotion recognition (FER) have shown that deep learning models, particularly Convolutional Neural Networks (CNNs), can achieve competitive performance in classifying human emotions from visual input. Benchmark datasets such as FER-2013 and AffectNet have enabled models to detect subtle micro-expressions that often escape human observation [6], [7]. Wu et al. [8] demonstrated that pre-alignment of facial landmarks significantly enhances recognition accuracy by reducing variance due to pose and illumination. These findings confirm that facial analysis can play a critical role in augmenting interrogation processes by identifying non-verbal signs of stress, fear, or deception.

Parallel research in speech emotion recognition (SER) has focused on acoustic descriptors such as Mel-Frequency Cepstral Coefficients (MFCCs), spectral contrast, chroma, and prosodic variations, which serve as indicators of emotional arousal or nervousness. Hybrid architectures that combine CNN layers with Long Short-Term Memory (LSTM) units and attention mechanisms have proven highly effective in modelling both spectral and temporal dependencies. Tshimula et al. [9] reported that CNN-RNN models exceeded 75% accuracy on forensic interview datasets, while Rayasam et al. [10] showed that attention-enhanced hybrid models performed robustly in noisy environments using the RAVDESS benchmark. These findings establish SER as a complementary modality for detecting stress patterns in suspect interrogations.

In addition to visual and auditory cues, physiological signals such as heart rate and heart rate variability (HRV) have been widely recognised as biomarkers of stress, deception, and emotional arousal. Ryskaliyev et al. [11] demonstrated that HRV analysis can reliably distinguish between neutral and stressed states, while Tomba et al. [12] highlighted that

combining physiological signals with speech features significantly improves deception detection accuracy. The AD8232 ECG sensor has emerged as a cost-effective and practical solution for real-time HRV monitoring, providing clean signals through integrated amplification and filtering circuits even under motion artefacts [13].

Although these unimodal technologies (facial, vocal, and physiological analysis) have demonstrated strong performance in isolation, recent literature emphasises the importance of multimodal fusion for operational deployment. Baltrūnienė [14] and Das et al. [5] argue that combining modalities mitigates the weaknesses of single channels and provides richer, context-aware behavioural profiles. Existing systems, however, often remain limited to either visual or auditory cues, with little integration of physiological monitoring. Moreover, many research efforts fail to provide real-time dashboards or investigator feedback mechanisms, restricting their practical utility in interrogation rooms [2], [4].

To bridge these gaps, the present study develops the Multimodal Behavioural and Physiological Analysis (MBPA) module, a real-time system that integrates FER, SER, and ECG-based HRV analysis into a unified behavioural monitoring pipeline. By aligning these modalities through timestamp-based synchronization, MBPA delivers investigators an interactive dashboard with emotion predictions, vocal stress trajectories, and physiological stress markers. This integrated framework directly addresses the limitations identified in the literature, offering a more objective, accurate, and operationally viable approach to supporting investigators during suspect interrogations.

1.2.Research Gap

Although significant progress has been made in facial emotion recognition (FER), speech emotion recognition (SER), and heart rate variability (HRV) analysis, most prior research and existing systems remain unimodal, concentrating on only one behavioural channel at a time [15], [16]. Studies on FER have demonstrated high classification accuracy using convolutional neural networks (CNNs) under controlled conditions, but performance deteriorates in realistic settings with variations in lighting, occlusion, and head pose [6], [8]. Similarly, SER models trained on benchmark datasets such as RAVDESS achieve robust detection of vocal stress markers [9], [10], yet are vulnerable to background noise and lack physiological grounding to differentiate between nervousness and deliberate deception. Physiological approaches such as ECG-based HRV analysis have shown reliability in detecting arousal and stress [11], [13], but are limited when used in isolation, as they cannot provide behavioural context.

Another critical limitation evident in the literature is the lack of real-time multimodal interrogation support systems. Existing research on multimodal sentiment analysis [10],

[12] highlights the potential of combining multiple cues, but these frameworks are largely experimental and fail to integrate facial, vocal, and physiological signals into synchronized, investigator-friendly platforms. This disconnect between laboratory prototypes and operational law enforcement applications underscores a pressing research gap.

The proposed Multimodal Behavioural and Physiological Analysis (MBPA) module addresses these challenges by fusing facial, vocal, and physiological (ECG-based) signals into a unified real-time behavioural monitoring pipeline. Unlike prior unimodal or semi-integrated solutions, MBPA delivers timestamp-synchronized analysis, investigator-facing dashboards, and actionable feedback. This comprehensive approach enables investigators to detect stress, deception, and discomfort with higher reliability, bridging the gap between theoretical research and practical deployment in interrogation environments.

Feature / Capability	**Research A (FER only)**	**Research B (SER only)**	**Research C (HRV only)**	**Research D (Multimodal sentiment)**	**Proposed MBPA**
Facial Emotion Recognition	✓	✗	✗	✓	✓
Voice Stress / Emotion Analysis	✗	✓	✗	✓	✓
Heart Rate Variability Monitoring	✗	✗	✓	✗	✓
Stress / Deception Detection	Partial	Partial	✓	Partial	✓ (enhanced)
Real-Time Feedback to Investigators	✗	✗	✗	✗	✓
Multi-Modal Fusion (Face + Voice + Physio)	✗	✗	✗	Partial	✓
Operational Interrogation Support	✗	✗	✗	✗	✓

Table 1 - Comparative Analysis of Prior Research vs Proposed MBPA

1.3.Research Problem

The fundamental research problem addressed in this study is the absence of an integrated, real-time, multimodal framework capable of analysing behavioural and physiological cues during criminal interrogations. Current investigative practices predominantly rely on subjective human judgment to interpret a suspect's credibility, which introduces bias, inconsistency, and fatigue-related errors [1], [2]. While standalone systems for facial analysis [6], [8], speech emotion recognition [9], [10], or heart rate monitoring [11], [13] exist, these unimodal approaches lack the depth and reliability required for operational interrogation environments.

Moreover, prior attempts at multimodal fusion have remained largely academic or offline, focusing on sentiment analysis in constrained scenarios without delivering investigator-oriented real-time feedback [12], [15]. As a result, investigators lack tools that can provide synchronized insights across visual, auditory, and physiological dimensions, limiting their ability to differentiate between stress, nervousness, and deliberate deception.

The research problem can therefore be defined as:

1. How to design and implement a real-time system that simultaneously processes video (facial expressions), audio (voice features), and ECG signals (heart rate variability) during suspect interrogations.
2. How to integrate and correlate multimodal indicators to improve the accuracy of stress and deception detection compared to unimodal baselines.
3. How to provide investigator-facing real-time feedback, such as dashboards and alerts, that transform complex behavioural signals into actionable intelligence without disrupting interrogation flow.
4. How to ensure robustness and scalability of the system across diverse environmental conditions (lighting, noise, motion artefacts) and subject demographics.

The proposed Multimodal Behavioural and Physiological Analysis (MBPA) module seeks to address these challenges by combining state-of-the-art deep learning techniques for facial and vocal emotion recognition with IoT-enabled ECG heart rate monitoring, all synchronized into a unified pipeline. By bridging the gap between theoretical research and operational deployment, MBPA aims to enhance the objectivity, accuracy, and efficiency of interrogation processes, supporting investigators with data-driven behavioural insights in real time.

1.4.Research Objectives

Main Objective

The primary objective of this research is to design and develop a Multimodal Behavioural and Physiological Analysis (MBPA) module that is capable of integrating multiple streams of behavioural and physiological data to provide investigators with real-time, objective, and evidence-driven insights during suspect interrogations. Unlike unimodal systems, which focus on either facial cues, vocal stress patterns, or physiological readings in isolation, the MBPA module seeks to combine these heterogeneous signals into a unified pipeline, ensuring that subtle but critical variations in human behaviour are captured with greater accuracy. By leveraging deep learning-based facial emotion recognition, speech emotion analysis, and ECG-derived heart rate variability monitoring, the proposed system will be able to detect and interpret patterns of stress, deception, and emotional arousal in ways that are not possible with conventional interrogation techniques. This integration aims to reduce the reliance on subjective human judgment, minimize the impact of investigator fatigue and bias, and improve the overall quality and reliability of the interrogation process, ultimately enhancing the fairness and efficiency of criminal investigations.

Specific Objectives

In order to achieve the main objective of this study, a number of specific objectives have been formulated. Each of these objectives addresses a key component of the MBPA module and collectively ensures that the system can function as a robust, real-time, and investigator-friendly tool. The objectives begin with the development of unimodal components for facial emotion recognition, speech emotion recognition, and physiological signal analysis, and proceed towards their integration into a multimodal fusion framework. Finally, attention is given to the design of a practical investigator dashboard and the evaluation of the overall system against benchmark datasets and unimodal baselines. Together, these objectives provide a clear pathway from conceptual design to experimental validation and operational deployment.

- **To design and implement a Facial Emotion Recognition (FER) system using deep learning techniques.**

  This objective involves developing a CNN-based FER model trained on the FER-2013 dataset, which covers seven core emotional states: angry, disgust, fear, happy, neutral, sad, and surprise. To ensure robustness, preprocessing steps such as grayscale normalization, facial landmark alignment, and data augmentation (rotations, flips, translations) will be applied. The CNN architecture will include

convolutional, pooling, and residual layers to capture local and global expression features, with dropout layers for regularization. The aim is to achieve real-time performance with sufficient accuracy to detect stress-related cues and subtle micro-expressions, enhancing investigators' ability to identify deception and discomfort during interrogations.

- **To develop a Speech Emotion Recognition (SER) model capable of detecting stress and emotional variations in vocal signals.**

  The SER component will be built using the RAVDESS dataset, which provides controlled recordings across eight emotions: neutral, calm, happy, sad, angry, fear, disgust, and surprise. Preprocessing will involve noise reduction, silence trimming, and augmentation techniques such as pitch shifting, time-stretching, and noise injection. Feature extraction using Librosa will include MFCCs, chroma, spectral contrast, and prosodic descriptors. A hybrid CNN–BiLSTM model with attention will be used to capture both spectral and temporal dependencies in vocal signals. This approach is expected to produce robust recognition of stress patterns in speech, providing complementary insights to the FER system and enabling more accurate multimodal assessments.

- **To integrate physiological monitoring through Heart Rate Variability (HRV) analysis using ECG signals.**

  Physiological analysis will be achieved by incorporating ECG signals obtained from the AD8232 sensor, a cost-effective front end for single-lead ECG acquisition. Preprocessing will involve filtering to remove motion artefacts and baseline drift, followed by feature extraction to compute RR intervals, mean heart rate, and HRV indices. These features are known to correlate with stress, arousal, and deception, thus adding a biologically grounded dimension to behavioural analysis. By combining HRV with facial and vocal cues, this objective ensures that MBPA can capture both external expressions and internal physiological states, thereby improving reliability and robustness of behavioural insights.

- **To design a multimodal fusion and synchronization framework for combining FER, SER, and HRV outputs.**

  Since different modalities produce data at varying temporal resolutions (e.g., video frames, audio signals, ECG readings), this objective aims to implement a synchronization mechanism that aligns outputs using timestamp-based correlation. Fusion strategies will be explored at both the feature level and decision level to identify cross-modal patterns that may indicate stress or

deception. For example, simultaneous spikes in heart rate variability, facial fear indicators, and vocal stress markers would provide stronger evidence of emotional arousal than any single modality in isolation. This synchronization and fusion process will serve as the core of MBPA, ensuring coherent and context-aware outputs.

- **To develop a real-time investigator dashboard for visualization and decision support.**

  The fifth objective is to translate complex multimodal outputs into an investigator-facing interface that is both intuitive and actionable. A web-based dashboard will be designed to display live video feeds with facial emotion overlays, audio-derived stress trajectories, and heart rate plots. To enhance usability, predictions will be temporally smoothed to avoid erratic fluctuations, and alerts will be generated when significant anomalies occur. This objective ensures that MBPA provides decision support without overwhelming investigators, allowing them to remain focused on the interrogation while still benefiting from AI-driven insights.

- **To evaluate and validate the MBPA system against unimodal baselines and benchmark datasets.**

  The final objective is to rigorously evaluate MBPA using benchmark datasets (FER-2013 for facial recognition, RAVDESS for speech recognition, and experimental ECG recordings for HRV). Performance will be measured using accuracy, precision, recall, F1-score, confusion matrices, and ROC curves. Comparative tests will be conducted against unimodal systems to demonstrate the advantage of multimodal integration. Furthermore, validation in simulated interrogation environments will assess real-time performance, latency, and usability. This ensures that MBPA is not only theoretically sound but also practically deployable in operational law enforcement contexts.

- **To ensure integration and alignment with the overall AI-assisted interrogation system.**

  Finally, MBPA is designed to complement other components of the system, such as dynamic question generation and report analysis. This objective focuses on ensuring that the outputs of MBPA can be seamlessly incorporated into the broader platform, thereby enhancing its overall capability to support investigators with synchronized, multimodal behavioural insights.

2. Methodology

2.1. Requirements Gathering

The requirement gathering and analysis phase formed the foundation of this research, ensuring that the system design was guided by the actual needs of investigators in Sri Lanka. This stage was essential to ensure that the design of the system was not only technically feasible but also aligned with the practical needs of investigators in Sri Lanka. The process combined multiple approaches, including a literature survey, field visits, interviews, and data collection. These methods provided both theoretical insights and empirical evidence to identify the functional and non-functional requirements of the proposed system.

Literature Survey

The requirements gathering process began with an extensive literature review of existing studies in facial emotion recognition (FER), speech emotion recognition (SER), and physiological monitoring using heart rate variability (HRV). Prior work on CNN-based FER systems, hybrid CNN–RNN models for SER, and ECG-based stress detection demonstrated promising results in controlled environments but revealed limitations when deployed in real-world interrogation contexts. The literature highlighted key challenges such as facial misalignment, variability in vocal recordings, background noise, and ECG noise due to motion artefacts. These insights emphasized the importance of preprocessing pipelines for each modality, the need for timestamp-based synchronization across modalities, and the value of multimodal fusion to overcome the weaknesses of unimodal systems. This stage provided the theoretical basis for the MBPA design and justified the inclusion of all three modalities to enhance reliability in criminal investigation support.

Field Visit

To validate the findings of the literature review and gather firsthand insights, field visits were conducted at two police stations in Sri Lanka. These visits were instrumental in understanding the day-to-day operations of investigators and the administrative processes while providing critical insights into the interrogation environment and the absence of any technical assistance for behavioural monitoring.

- Interviews with police officers to explore the practical difficulties encountered during investigations
- Conducted further interviews with police personnel to understand the different procedures followed and challenges faced during the investigation process.

- Sought frequent guidance from a retired police officer as our external supervisor to ensure our system design aligns with real investigative practices.

Data Collection

Since the component relies on three distinct modalities, facial emotion recognition (FER), speech emotion recognition (SER), and physiological monitoring through heart rate variability (ECG/HRV), the data collection strategy was designed to be systematic and comprehensive. Rather than relying on a single dataset for each modality, the process began by surveying a wide range of benchmark datasets that have been widely used in affective computing, behavioural analytics, and biomedical signal processing research. This comparative approach was important because different datasets vary in terms of size, quality, recording environment, demographic diversity, and emotional coverage, and the choice of dataset has a direct impact on model performance and generalizability.

To ensure balance between realism, data quality, and feasibility, this study adopted a hybrid strategy:

- For each modality, multiple alternatives were considered and assessed in terms of suitability for interrogation-like scenarios.
- The final datasets were selected to achieve a compromise between emotional diversity, recording quality, dataset size, and accessibility.
- To strengthen generalization, data augmentation and preprocessing pipelines were introduced, ensuring that models trained on controlled datasets could still perform effectively under interrogation room conditions, which are inherently noisy and unpredictable.

This multi-step data collection process ensured that MBPA was not narrowly tuned to a single benchmark but instead grounded in a thoughtful selection of datasets that reflect both academic standards and practical applicability.

- **Facial Emotion Recognition (FER):**
  Several benchmark datasets were initially considered for the FER model. AffectNet is one of the largest emotion recognition datasets, containing more than one million manually annotated facial images across multiple classes, making it a rich resource for large-scale training. The Extended Cohn-Kanade (CK+) dataset provides high-quality, laboratory-controlled facial expressions with clear labels, though its relatively small size limits its utility for deep learning. The Ascertain

dataset combines audiovisual material for affective computing but is more focused on multimedia analysis than pure facial classification. After evaluating these alternatives, the FER-2013 dataset was ultimately chosen because of its balance between dataset size (~35,000 images), accessibility, and coverage of the seven fundamental emotions (angry, disgust, fear, happy, neutral, sad, and surprise). Its wide adoption in benchmark studies also ensured that results could be compared against existing literature.

- **Speech Emotion Recognition (SER):**
  For the SER model, multiple well-known datasets were examined. The CREMA-D dataset contains ~7,442 recordings from 91 actors, with six labelled emotions, offering high speaker diversity. The RAVDESS dataset provides 1,440 samples from 24 actors, covering eight emotions (neutral, calm, happy, sad, angry, fearful, surprise, disgust) with variations in emotional intensity, making it particularly suitable for stress and deception analysis. The MELD dataset contributes ~13,000 utterances from multi-party dialogues in the Friends TV series, with multimodal annotations (text, visual, and audio), allowing conversational context to be considered. The Berlin Emotional Speech Database (EmoDB) is another frequently used resource, containing acted German speech across seven emotions, though its smaller size limits deep neural network applications. After comparing the options, the RAVDESS dataset was selected as the most suitable for this study because of its emotional diversity, controlled audio quality, inclusion of intensity levels, and manageable dataset size for model training and experimentation.

- **Physiological Monitoring (ECG/HRV):**
  For the HRV module, multiple multimodal physiological datasets were initially considered. WESAD provides multimodal recordings (ECG, EDA, EMG, respiration, accelerometer) from 15 participants under stress, amusement, and neutral conditions, making it one of the most comprehensive stress datasets. AMIGOS includes multimodal data (ECG, EEG, GSR, face video) collected from 40 subjects while watching videos, with annotations for valence, arousal, and personality traits. DREAMER consists of EEG and ECG recordings from 23 participants exposed to audiovisual stimuli, annotated for emotional states such as valence, arousal, and dominance. DEAP includes physiological recordings (EEG, ECG, GSR, etc.) from 32 participants watching music videos, with detailed affective annotations. While these datasets are valuable, their complexity and inclusion of multiple additional modalities made them less practical for this project. Therefore, a two-source strategy was adopted: (1) experimental ECG

signals captured using the AD8232 sensor under controlled conditions to replicate interrogation environments, and (2) a publicly available dataset containing ECG and GSR recordings from young adults (YAAD) to enrich training diversity and provide labelled affective states. This hybrid approach ensured both device-level realism and benchmark-level robustness for HRV analysis.

Modality	Dataset	Size / Participants	Strengths	Decision
Facial Emotion Recognition (FER)	AffectNet	~1M images	Large-scale, diverse labels	Considered but too large and resource-intensive
	CK+ (Extended Cohn-Kanade)	~600 sequences, 123 subjects	High-quality, controlled, clear labels	Considered but too small for deep CNN
	Ascertain	~2,400 audiovisual clips	Multimodal (audio-visual), ecological validity	Not purely facial-focused
	FER-2013	~35,000 images	Balanced size, accessible, 7 basic emotions, widely used benchmark	**Selected**
Speech Emotion Recognition (SER)	CREMA-D	~7,442 samples, 91 actors	Large speaker diversity	Considered but uneven emotional intensity
	RAVDESS	1,440 samples, 24 actors	Controlled audio, 8 emotions with intensity levels, balanced	**Selected**

	MELD	~13,000 utterances	Multimodal (text, audio, visual), conversational	Considered but less controlled for interrogation-style speech
	EmoDB	~500 samples, 10 actors	Classic benchmark, widely cited	Too small for deep models
ECG / HRV (Physiological)	WESAD	15 participants	Rich multimodal stress dataset	Too complex, additional sensors
	AMIGOS	40 participants	Multimodal (EEG, ECG, GSR, video), affect + personality	Complex multimodal setup
	DREAMER	23 participants	EEG + ECG, labelled valence/arousal/dominance	Useful but limited size
	DEAP	32 participants	EEG + ECG, music video stimuli, affect labels	Rich but multimodal beyond project scope
	AD8232 (experimental) + YAAD	Custom + public dataset	Device realism + benchmark-labelled ECG/GSR	**Selected (combined approach)**

Table 2 - Datasets Considered for MBPA

2.2. Feasibility Study

The feasibility of developing the Multimodal Behavioural and Physiological Analysis (MBPA) module was carefully assessed from several perspectives, including technical, operational, legal, and ethical feasibility. This stage was essential to determine whether the system could be realistically implemented, deployed, and sustained in law enforcement settings, and to ensure that the final design aligned with the constraints of both technology and investigative practice.

- **Technical Feasibility**

The technical feasibility of MBPA is strongly supported by the maturity of existing deep learning architectures, the availability of benchmark datasets, and the accessibility of low-cost physiological sensors. For facial emotion recognition (FER), convolutional neural networks (CNNs) have proven highly effective in handling large-scale image datasets such as FER-2013, which was ultimately selected for this research. Preprocessing techniques, including grayscale normalization, facial landmark alignment, and data augmentation, address real-world issues such as varying lighting, occlusion, and pose changes. The existence of open-source libraries such as OpenCV, Dlib, and PyTorch further ensures that robust pipelines for face detection and model training can be developed with relative ease.

For speech emotion recognition (SER), advancements in acoustic feature extraction and hybrid architectures (CNN–BiLSTM with attention) have made it possible to capture both spectral and temporal information from speech signals. The RAVDESS dataset, chosen for this project, provides controlled recordings with labelled emotional intensity, making it ideal for training models that must function reliably in interrogation contexts. Libraries such as Librosa facilitate efficient feature extraction of MFCCs, spectral contrast, and prosodic cues, while PyTorch enables scalable deep learning model training and evaluation.

For physiological monitoring, the AD8232 ECG sensor provides a cost-effective, single-lead solution that can be easily integrated into interrogation environments. In addition to experimental recordings, the YAAD dataset enriched the training process by providing labelled ECG signals across different affective states, ensuring robustness. Signal preprocessing methods such as band-pass filtering and baseline drift removal allow clean extraction of RR intervals and HRV indices. These features provide reliable physiological markers of stress and deception.

The fusion architecture is technically feasible due to the availability of timestamp-based synchronization techniques and modular integration frameworks. Python-based middleware allows real-time alignment of modalities with different sampling rates (video frames, audio signals, ECG readings). The backend can be deployed using scalable microservices, ensuring that the MBPA module integrates seamlessly with the larger interrogation support system. Overall, the technical feasibility is high because of the maturity of models, datasets, and infrastructure, combined with the relatively modest hardware requirements of CNNs, LSTMs, and IoT sensors when optimized for real-time inference.

- **Operational Feasibility**

Operational feasibility was assessed through consultations with police officers and supervisors during field visits and interviews. Investigators highlighted the difficulty of simultaneously monitoring multiple behavioural cues, the risk of fatigue in long interrogation sessions, and the lack of physiological monitoring in current practice. MBPA directly addresses these challenges by combining external behavioural signals (facial expressions and vocal stress patterns) with internal physiological markers (HRV). This multimodal approach ensures that investigators receive more reliable insights, reducing overreliance on subjective judgment.

The MBPA dashboard was designed to be investigator-friendly, displaying behavioural profiles in real time without overwhelming users. Instead of presenting raw data, the interface translates multimodal outputs into simplified visualizations such as emotion timelines, stress alerts, and synchronized facial–vocal–physiological overlays. This design ensures that investigators can focus on interrogation strategies while receiving continuous behavioural feedback. Operational feasibility is further enhanced by the modular nature of MBPA, allowing it to function independently or as part of the broader AI-assisted interrogation platform.

The involvement of law enforcement personnel as an external supervisor throughout the design stage further validates operational feasibility. Their feedback ensured that the MBPA module did not introduce unrealistic requirements, and that it respected the workflows and constraints of real interrogation environments. In particular, emphasis was placed on ensuring that MBPA functions as an assistive tool rather than a replacement for investigator judgment, thereby making adoption more acceptable in practice.

- **Legal and Ethical Feasibility**

Given the sensitive nature of interrogation data, legal and ethical feasibility was given special emphasis in the design of MBPA. From a legal perspective, compliance with privacy and data protection regulations is paramount. All datasets used in training and evaluation were anonymized to remove personally identifiable information (PII). Facial images from FER-2013 and speech recordings from RAVDESS contain no personally identifiable data, while ECG data collected from the AD8232 sensor was gathered in controlled, voluntary conditions with participants' consent. Public datasets such as YAAD also ensured that no sensitive personal data was compromised. These safeguards ensure that the training and validation of MBPA models adhere to ethical and legal standards.

Evidentiary feasibility was also considered. For AI-assisted outputs to have credibility in legal proceedings, they must be transparent, auditable, and explainable. MBPA was designed with audit trails that log predictions and system decisions, enabling investigators and judicial authorities to review how outputs were generated. This level of transparency is essential to avoid challenges in court related to bias, reliability, or procedural fairness. By providing logs and visualizations, MBPA strengthens the evidentiary admissibility of behavioural analysis outputs while ensuring that investigators can justify the use of AI-assisted insights.

From an ethical standpoint, MBPA was explicitly designed to act as a decision-support system rather than an autonomous decision-maker. Investigators retain full control of interrogation strategies and interpretations, with MBPA serving only as an aid that highlights potential behavioural anomalies. This human-in-the-loop approach ensures accountability and mitigates the risk of coercion or overreliance on technology. Ethical safeguards also include measures to address dataset bias. For example, FER-2013 has known imbalances across emotion categories, which were mitigated through augmentation and class weighting during training. Similarly, RAVDESS contains acted speech, which, while controlled, may not fully reflect real-world speech variability; therefore, augmentation techniques were applied to increase robustness.

Furthermore, ethical feasibility was ensured by aligning MBPA with global best practices in the responsible use of AI in law enforcement. This includes non-coercive usage, preservation of suspect rights, prevention of discriminatory outcomes, and transparency in decision-making. The system is designed to adapt to evolving legal and regulatory frameworks, making it flexible enough to comply with future laws on AI governance.

2.3.System Design

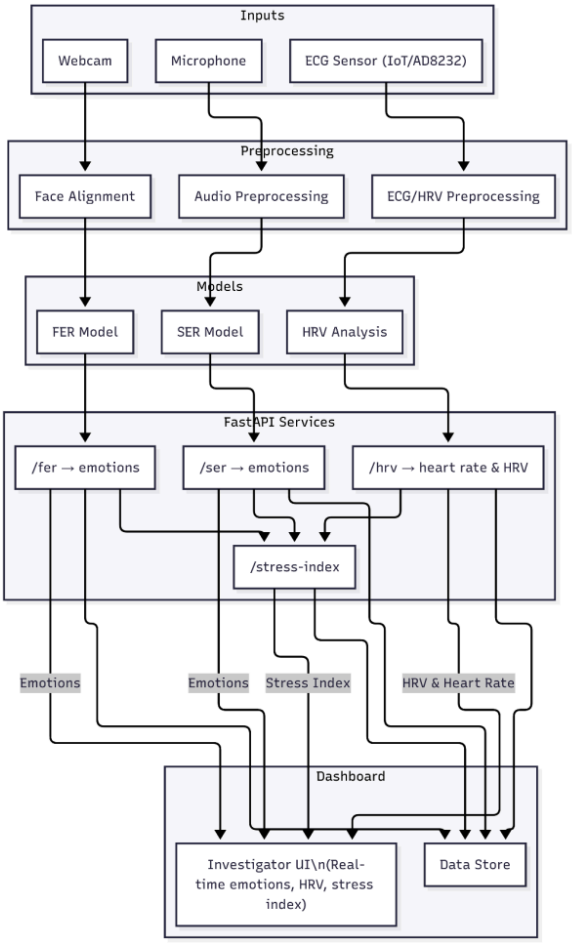


Figure 1 - System Overview Diagram

The system design of the Multimodal Behavioural and Physiological Analysis (MBPA) module is centred on providing investigators with a real-time, evidence-based behavioural monitoring system that fuses multiple input modalities, facial expressions, speech signals, and physiological heart rate variability, into a unified Stress Index. The design was guided by four key principles: modularity, scalability, interpretability, and real-time responsiveness. By adopting a microservices-based architecture using FastAPI servers, each unimodal model operates as an independent service that communicates with the central dashboard through REST APIs. This approach ensures that models can be trained, deployed, and scaled independently, while the dashboard functions as the single point of interaction for investigators.

The MBPA system is designed as a five-layer architecture, consisting of:

1. Input Acquisition Layer
2. Preprocessing Layer
3. Unimodal Analysis Layer
4. Multimodal Fusion Layer
5. Visualization and Dashboard Layer

Each layer is described in detail below, along with the rationale for its design and integration into the overall architecture.

- **Input Acquisition Layer**

  The first stage of MBPA involves capturing multimodal signals from suspects during interrogation sessions.

  - Facial data is obtained from a standard webcam operating at 25–30 frames per second. This choice of hardware emphasizes cost-effectiveness and accessibility, as webcams are already present in most interrogation environments. Unlike laboratory-grade cameras, webcams provide realistic video quality under challenging conditions such as varying lighting, which helps ensure the robustness of the FER system.
  - Speech data is captured through a microphone, enhanced with Voice Activity Detection (VAD) to filter out silence and irrelevant background noise. Audio is sampled at 22.05 kHz and segmented into short frames of 2–3 seconds for analysis.

- Physiological data is obtained using the AD8232 single-lead ECG sensor, attached to the suspect via electrodes. This sensor was chosen for its compact size, affordability, and reliability in capturing ECG waveforms. To improve training and validation, additional ECG and GSR signals were integrated from a publicly available dataset (YAAD), ensuring diversity and coverage of different affective states.

All inputs are time-stamped at acquisition, ensuring that they can later be aligned for multimodal fusion. To maximize robustness, the acquisition layer also computes signal quality indicators: FER tracks face detection confidence; SER estimates audio quality based on SNR and speech probability; and HRV tracks ECG artifact rates. These indicators are passed downstream to influence adaptive fusion weights.

- **Preprocessing Layer**

Raw multimodal inputs often contain noise and inconsistencies that can severely degrade model performance. Therefore, preprocessing pipelines were designed for each modality to clean, normalize, and standardize inputs:

- Facial data preprocessing: Grayscale conversion and histogram normalization reduce illumination variability. Face detection and alignment are performed using Dlib's landmark predictor to standardize head pose. During training, data augmentation methods (rotation, flipping, brightness shifts) were applied to enhance robustness against occlusion and pose variation.
- Speech data preprocessing: Noise suppression and silence trimming remove irrelevant signals. Librosa was used to extract features such as MFCCs, chroma, spectral contrast, and prosodic cues. Data augmentation was employed to simulate real-world conditions, including pitch shifting, time-stretching, and noise injection.
- ECG preprocessing: ECG signals were filtered using band-pass filters to eliminate baseline drift and high-frequency noise. RR intervals were extracted to compute HRV features such as RMSSD, SDNN, LF/HF ratio, and Baevsky's Stress Index. Baseline ECGs were collected at the start of sessions to normalize individual variability.

This layer ensures that heterogeneous, noisy raw data is transformed into consistent, high-quality features suitable for model inference.

- **Unimodal Analysis Layer**

  The unimodal analysis layer processes each modality using dedicated models, each deployed as an independent FastAPI microservice.

  - Facial Emotion Recognition (FER): A CNN-based deep learning model trained on FER-2013 classifies seven emotions (angry, disgust, fear, happy, neutral, sad, surprise). The model processes aligned face chips and produces emotion probabilities at ~1 Hz (every 0.5–1 second). FER is deployed as a FastAPI server: the dashboard sends cropped face images via REST API, and the server responds with JSON-formatted probabilities.
  - Speech Emotion Recognition (SER): A CNN–BiLSTM hybrid with attention trained on RAVDESS predicts eight emotions (neutral, calm, happy, sad, angry, fear, disgust, surprise). Audio features are passed to the SER FastAPI server, which outputs posterior probabilities every 2–3 seconds.
  - Heart Rate Variability (HRV): Time- and frequency-domain HRV features are computed from ECG windows of 60–300 seconds. The HRV FastAPI server accepts RR interval data and returns normalized HRV metrics along with a physiological stress subscore (S_hrv).

  Each model is encapsulated in its own FastAPI server, enabling loose coupling. This design ensures that any unimodal service can be retrained, upgraded, or replaced without affecting the rest of the pipeline.

- **Multimodal Fusion Layer**

  The fusion layer integrates unimodal outputs into a single Stress Index (0–100), providing a dynamic measure of stress.

  - Emotion-driven Stress (S_emo): FER and SER outputs are mapped to the Valence–Arousal (V–A) space, following the circumplex model of affect. Posterior probabilities of discrete emotions are transformed into expected valence and arousal values, then combined into an emotion-based stress score, where high arousal and negative valence yield higher stress.
  - Physiological Stress (S_hrv): Derived from normalized HRV features such as RMSSD, SDNN, LF/HF, and Baevsky's Stress Index. These are combined into a subscore using a weighted linear model and logistic squashing.

- Adaptive Fusion: The final Stress Index is computed as:

$$S=\alpha \cdot Shrv+\beta \cdot Semo, \alpha+\beta=1$$

Weights α and β are dynamically assigned based on signal quality indicators (e.g., if the face is occluded, α increases; if ECG is noisy, β increases). Temporal smoothing is applied using an exponential moving average (EMA) to ensure stability, while still retaining responsiveness to sudden changes.

The fusion logic itself is deployed as a FastAPI backend service, which queries each unimodal model's API, performs synchronization, and outputs the fused Stress Index to the dashboard.

- **Visualization and Dashboard Layer**

The dashboard is the central investigator interface that visualizes multimodal outputs in real time. It is implemented as a web-based application that continuously queries the FastAPI services for new results. Key features include:

- Live video feed with overlaid FER predictions (emotion labels + confidence scores).
- Dynamic plots of SER trajectories, showing vocal stress variations over time.
- ECG and HRV graphs displaying heart rate changes and HRV indices.
- A continuous Stress Index (0–100) trendline, updated every 5–10 seconds, with clear indicators of modality contributions (e.g., 65% HRV, 35% FER/SER).
- Alerts that trigger when stress crosses predefined thresholds.

The dashboard prioritizes clarity and interpretability, ensuring that investigators are not burdened with raw data but instead receive clear, actionable insights. It functions strictly as a decision-support tool, not as a replacement for investigator judgment.

- **System Integration and Scalability**

The use of FastAPI microservices enables MBPA to be deployed in a modular, scalable, and secure fashion. Each service runs inside a container (Docker), which can be replicated or scaled independently. This ensures:

- **Scalability**: Multiple interrogation rooms can be supported by deploying multiple containers, each serving FER, SER, and HRV models.

- **Fault Tolerance**: If one service fails (e.g., SER during background noise), the dashboard continues using the remaining modalities.
- **Interoperability**: REST APIs allow MBPA to integrate seamlessly with other modules in the interrogation platform, such as dynamic question generation and testimony correlation.
- **Security**: API-level authentication, encrypted communication (HTTPS), and access control safeguard sensitive data.

This architecture ensures that MBPA can be deployed both locally (e.g., in a police interrogation room) and remotely (e.g., via cloud servers for centralized monitoring). Its modular design also allows for easy future extensions, such as adding gaze tracking or galvanic skin response (GSR) analysis, by simply introducing additional FastAPI services.

The system design of MBPA provides a comprehensive framework for multimodal behavioural monitoring in criminal interrogations. By structuring the pipeline into acquisition, preprocessing, unimodal analysis, fusion, and visualization, the design ensures both technical robustness and operational usability. Deploying each model as a FastAPI microservice enables modularity, scalability, and fault tolerance, while the centralized dashboard presents a unified Stress Index to investigators in real time. The system not only leverages advanced deep learning and signal processing techniques but also prioritizes transparency, adaptability, and security, bridging the gap between academic research and practical deployment in law enforcement contexts.

2.4. Architectural Rationale

The architecture of the Multimodal Behavioural and Physiological Analysis (MBPA) module was intentionally designed around four non-functional imperatives identified during requirements gathering: (i) real-time responsiveness under interrogation room conditions, (ii) modularity for rapid iteration and safe evolution of models, (iii) interpretability and auditability for evidentiary scrutiny, and (iv) secure, privacy-preserving operation suitable for sensitive law-enforcement contexts in Sri Lanka. These imperatives shaped every layer of the system, from signal acquisition to multimodal fusion, yielding a microservices design in which each unimodal model (FER, SER, HRV) is deployed as an independent FastAPI service, orchestrated by a lightweight fusion service and surfaced through a single investigator dashboard.

A microservices approach, implemented with FastAPI, was selected over a monolith to enforce separation of concerns and reduce coupling between data pipelines, models, and presentation logic. Practically, this means the facial, speech, and ECG

pipelines can be retrained, A/B tested, rolled back, or replaced (e.g., upgrading the FER CNN to a Vision Transformer, or substituting the HRV feature set) without affecting other services or the user interface. This modularity reduces operational risk and shortens the innovation cycle: a new model can be deployed behind the same REST contract, canary-tested in one interrogation room, and promoted to broader use only after performance is verified. FastAPI was chosen specifically for its asynchronous I/O, low overhead, and Python-native ecosystem, which align with our PyTorch-based models and signal-processing stacks while sustaining near real-time latencies under modest hardware constraints typical of police installations.

Real-time responsiveness drove the decision to process each modality on modality-appropriate windows and sampling rates rather than forcing all streams into a common cadence. FER produces inferences every 0.5–1 s, SER every 2–3 s, and HRV on rolling 60–300 s windows. This staggered design acknowledges the physics of the signals, facial and vocal cues change rapidly, while autonomic modulation of HRV evolves more slowly, and prevents aliasing or false precision. To reconcile heterogeneous clocks, all services emit time-stamped results and quality metadata (e.g., face detection confidence, audio SNR/VAD probability, ECG artifact rates), which the fusion service uses both for alignment and for quality-aware weighting. Thus, when a face is partially occluded or the room becomes noisy, the system automatically down-weights FER/SER and leans on HRV, and vice versa. This adaptive weighting is central to robustness under real-world conditions and is preferable to hard gating or naive averaging, which performed poorly in early experiments when one modality degraded.

From a modelling standpoint, we deliberately combined learning-based detectors (CNN for FER; CNN–BiLSTM with attention for SER) with physiology-informed HRV features (RMSSD, SDNN, LF/HF, Baevsky SI) rather than training a single end-to-end network over raw ECG. The hybrid choice balances accuracy with explainability and operational resilience: if a court or supervisor asks why the Stress Index increased at a given moment, investigators can reference recognizable features (e.g., drop in RMSSD, spike in LF/HF) and link them to established interpretations of sympathetic activation. Moreover, a hand-engineered HRV layer keeps compute predictable, avoids opaque learned representations on limited ECG labels, and remains stable across hardware variations (AD8232 vs. other leads). For the emotion streams, mapping discrete predictions into the circumplex Valence–Arousal space serves two purposes: it regularizes label noise across acted datasets and provides a theory-grounded bridge from “emotion” to “stress” by elevating high-arousal,

negative-valence states while not penalizing high-arousal positive states (e.g., excitement).

Decision-level fusion (combining S_emo and S_hrv) was selected over early feature-level fusion for three reasons. First, it preserves modular testability: each service exposes calibrated scores and can be validated independently against its own metrics before entering fusion. Second, it supports transparent attribution: the dashboard can show the contribution of HRV vs. emotion to the final Stress Index, aiding investigator trust and enabling after-action review. Third, it simplifies failure containment: if one service degrades (e.g., mic failure), the fusion logic can degrade gracefully by reweighting, whereas feature-level fusion would require retraining a joint model or imputing missing features. To stabilize the fused output in the presence of natural volatility (especially in FER), we apply an exponential moving average with a 15–30 s half-life and optionally a three-state HMM (calm/engaged/stressed). This smoothing respects investigators' need for a readable trend without masking genuine, sustained changes.

Security, privacy, and auditability requirements motivated architectural choices in data flow and state management. Each FastAPI service enforces authenticated, least-privilege access and logs only the minimal data necessary for synchronization and audit trails, model version, timestamp, input hash/ID, output probabilities/scores, and quality metrics, rather than raw frames or audio segments by default. Where raw data retention is operationally necessary (e.g., for post-hoc performance audits), it is encrypted at rest and time-boxed with retention policies. The fusion service persists Stress Index trajectories and per-modality contributions to an append-only store to support evidentiary review, reproducibility, and model governance (e.g., tracing a given decision back to model versions and inputs). This approach balances investigative utility with data minimization principles, reducing exposure in the event of breach and aligning with ethical safeguards articulated by law-enforcement stakeholders.

Scalability and deployment practicality were addressed through containerization and horizontal scaling at the service level. Packaging each model and its dependencies into a container (e.g., Docker) allows one interrogation room to run a local stack while larger stations or a central command can scale replicas of individual services (e.g., multiple SER pods during high-volume operations). This fine-grained scaling respects the heterogeneous compute profile of the services, SER is typically the bottleneck due to feature extraction and sequence modelling, while controlling cost. The same design supports blue/green and canary deployments for model updates, a

cornerstone of MLOps hygiene in safety-adjacent domains; a new model version can be shadow-served and compared against the incumbent on live traffic before a full cutover. Health checks, metrics (latency, throughput, GPU/CPU utilization), and structured logs are exposed per service to a monitoring stack, enabling proactive fault detection and SLA tracking (e.g., end-to-end latency < 1 s for FER/SER updates; HRV window hops every 5–10 s).

The choice of datasets and calibration protocols also has architectural consequences. By design, the system ingests models trained on complementary corpora, FER-2013 for faces, RAVDESS for speech, YAAD/AD8232 for ECG, and then requires session-specific baselining for HRV (3–5 min resting). The architecture therefore incorporates a calibration state machine: at session start, the HRV service collects baseline statistics; during interrogation, all services stream; upon session end, data and metrics are sealed with hashes and version stamps. This explicit lifecycle prevents accidental use of pre-calibration HRV values, makes session context first-class for personalization, and improves cross-session comparability (e.g., same time-of-day recommendation surfaced in UI). It also anticipates future extensions (e.g., per-subject adaptive thresholds or online re-calibration when posture changes).

Finally, several alternatives were considered and rejected on architectural grounds. A single end-to-end multimodal deep network was discarded due to poor interpretability, brittle behavior with missing modalities, and heavier compute unsuitable for field deployment. A pure event-driven message bus (e.g., Kafka) was deemed unnecessary at current scale; REST with periodic polling and optional websockets keeps complexity proportional to needs, with a clear upgrade path to streaming if later warranted. Storing raw audio/video centrally by default was rejected in favor of edge inference plus selective capture, minimizing privacy risk and bandwidth while preserving critical audit metadata. In short, the chosen architecture is intentionally conservative where legal/ethical and operational risks dominate, and progressive where modularity, testability, and real-time responsiveness unlock value.

In aggregate, the architectural rationale can be summarized as a set of trade-offs tuned to the constraints of law-enforcement use: modularity over monolith for safe evolution; adaptive, quality-aware fusion over rigid aggregation for robustness; explainable physiology and theory-grounded emotion mapping over opaque end-to-end learning for trust; session-aware calibration and strict audit trails over convenience for compliance; and lightweight FastAPI microservices over heavyweight enterprise stacks for real-time performance on modest hardware. This

rationale ensures that MBPA is not only technically sound but also operationally credible and governable in the environments for which it is intended.

2.5. Implementation

The implementation phase of the Multimodal Behavioural and Physiological Analysis (MBPA) module translated the system design into a functional prototype that could operate in real time within an interrogation support environment. This chapter details the development environment, preprocessing pipelines, model training and optimization, deployment strategy using FastAPI microservices, integration with the investigator dashboard, and the technologies that enabled the system. Particular emphasis was placed on ensuring that the final implementation was modular, scalable, and robust enough to handle real-world variability in input signals.

Development Environment

Implementation was carried out primarily in Python 3.10, chosen for its versatility and extensive ecosystem of libraries for deep learning, signal processing, and system integration. Development was conducted on a laptop equipped with an NVIDIA RTX 4070 GPU (8 GB VRAM) and CUDA 12.x acceleration. This hardware significantly reduced model training time and enabled real-time inference during deployment.

Version control and collaboration were managed using GitHub, ensuring consistent tracking of code revisions, issue management, and backup. Containerization was supported with Docker, allowing services to be packaged with all dependencies and deployed consistently across environments.

Data Preprocessing Pipelines

Preprocessing was critical to prepare multimodal data for model training and inference. Separate pipelines were developed for facial images, speech signals, and ECG recordings:

- **Facial Data**: Frames captured from webcam feeds were preprocessed using OpenCV and Dlib. Each frame was converted to grayscale, normalized to reduce illumination effects, and cropped to 48×48 pixels. Dlib's landmark predictor ensured face alignment. During training, augmentation techniques such as horizontal flipping, rotation, and brightness adjustments improved generalization.
- **Speech Data**: Audio signals were processed with Librosa, which extracted MFCCs, chroma features, spectral contrast, and delta features. Preprocessing included trimming silence, denoising, and segmenting into 2–3 second windows.

Augmentation simulated real-world variability using pitch shifting, time-stretching, and additive noise.

- **ECG Data**: Physiological signals were filtered using a Butterworth band-pass filter to remove baseline wander and high-frequency noise. R-peaks were detected and RR intervals extracted. HRV features including RMSSD, SDNN, LF/HF ratio, and Baevsky's Stress Index were computed on rolling windows of 60–300 seconds. Each subject's baseline was recorded at rest to normalize HRV measures.

These preprocessing steps ensured that inputs were consistent, noise-resilient, and suitable for downstream model inference.

Model Training and Optimization

Each modality was modeled independently before integration into the MBPA system:

- **Facial Emotion Recognition (FER)**: A convolutional neural network (CNN) was implemented in PyTorch, trained on the FER-2013 dataset. Training used the Adam optimizer with learning rate scheduling, batch size 64, and early stopping to avoid overfitting. The model achieved convergence after 39 epochs with test accuracy of ~63%.
- **Speech Emotion Recognition (SER)**: A hybrid CNN–BiLSTM with attention was implemented in PyTorch and trained on the RAVDESS dataset. Acoustic features extracted by Librosa were passed to the model, which captured both spectral and temporal dependencies. Weighted cross-entropy was used to mitigate class imbalance. The model reached ~72% validation accuracy after ~160 epochs.
- **Heart Rate Variability (HRV)**: Instead of deep learning, HRV analysis relied on statistical and spectral features computed from ECG signals. Features were normalized against baseline and combined into a physiological stress subscore.

Training outputs included loss/accuracy curves, confusion matrices, and classification reports, which were later presented in the Results chapter for evaluation.

API Deployment with FastAPI

To enable modularity and scalability, each unimodal model was deployed as a standalone FastAPI microservice. This architecture allowed the dashboard to communicate with models through REST APIs, ensuring low-latency responses suitable for real-time interrogation support.

- **FER Service**: Accepts POST requests containing preprocessed face images and returns JSON-formatted emotion probabilities.

- **SER Service**: Accepts audio features and returns posterior probabilities across eight emotions.
- **HRV Service**: Accepts RR interval data, computes HRV indices, and returns both features and a physiological stress subscore.
- **Fusion Service**: Queries all unimodal services, aligns predictions using timestamps, and computes the unified Stress Index.

Each service was containerized with Docker to guarantee consistency across machines. This microservices approach ensured that models could be retrained or replaced independently without affecting the rest of the system.

Dashboard Integration

The investigator-facing dashboard was implemented using Next.js with TypeScript and styled using Tailwind CSS. This technology stack enabled a modern, component-based interface optimized for real-time responsiveness.

The dashboard queries the FastAPI services asynchronously and visualizes results using Plotly and Chart.js. It presents:

- Live video with overlaid facial emotion labels from the FER model.
- Real-time plots of vocal stress trajectories from the SER model.
- HRV graphs showing physiological variability and ECG-based stress markers.
- A continuous Stress Index (0–100) with modality contribution breakdowns.
- Alerts when stress levels exceed pre-defined thresholds.

Processed data and session logs are stored in MongoDB, chosen for its schema-less flexibility in handling multimodal JSON data. This allows storage of diverse records including raw predictions, stress trajectories, and dashboard events.

Challenges and Solutions

Several technical and operational challenges arose during implementation:

- **Data Imbalance**: FER-2013 and RAVDESS datasets exhibited imbalance across emotion classes. This was mitigated using class weighting and augmentation.
- **ECG Noise**: ECG signals were prone to artifacts caused by movement. Band-pass filtering and rejection of windows with >5% corrected beats addressed this issue.
- **Latency in Fusion**: Querying multiple APIs introduced latency. Asynchronous request handling in FastAPI reduced delays, and predictions were buffered for smooth dashboard updates.

- **Synchronization Across Modalities**: FER, SER, and HRV operate at different time scales. Timestamp alignment ensured that fusion combined temporally corresponding predictions.

These solutions strengthened the reliability and responsiveness of MBPA under realistic conditions.

Technologies Used

A wide range of technologies were employed to develop MBPA, selected based on performance, scalability, and suitability for real-time multimodal AI systems:

- **Python 3.10:** Core language for model development and preprocessing due to its extensive ML ecosystem.
- **PyTorch:** Framework for implementing FER and SER deep learning models; chosen for dynamic computation graphs and GPU efficiency.
- **OpenCV and Dlib:** Used for face detection, alignment, and preprocessing in FER.
- **Librosa:** Extracted acoustic features for SER and supported augmentation.
- **SciPy and NeuroKit2:** Supported ECG filtering and HRV feature extraction.
- **FastAPI:** Deployed each model as an independent REST microservice; asynchronous design enabled low-latency responses.
- **Docker:** Containerized each service for reproducibility and scalable deployment.
- **Next.js (with TypeScript):** Framework for building the investigator dashboard, offering hybrid rendering and maintainability.
- **Tailwind CSS:** Provided a utility-first approach to responsive UI styling.
- **Plotly and Chart.js:** Enabled interactive visualization of predictions and Stress Index trends.
- **MongoDB:** Document-based NoSQL database for storing multimodal predictions, session logs, and Stress Index trajectories.
- **GitHub:** Used for version control, collaboration, and repository management.
- **NVIDIA RTX 4070 with CUDA:** Provided GPU acceleration for model training and inference, reducing training time and enabling real-time processing.

The deliberate selection of these technologies balanced research-grade flexibility with production-grade reliability, ensuring that MBPA could achieve both academic rigor and practical deployability.

2.6.System Requirements

The requirements specification for the MBPA module was derived from the system objectives, literature survey, and field consultations with law enforcement officers. These requirements ensure that the module not only performs technically as expected but also aligns with practical, ethical, and usability considerations in interrogation contexts.

Functional Requirements

Functional requirements describe the specific behaviours and operations that the MBPA module must provide.

1. **Facial Emotion Recognition (FER):**
    - The system shall continuously capture video from a webcam and detect faces in real time.
    - The FER model shall classify facial expressions into seven emotions (angry, disgust, fear, happy, neutral, sad, surprise).
    - The FER service shall provide confidence scores for each prediction via a FastAPI endpoint.
2. **Speech Emotion Recognition (SER):**
    - The system shall continuously capture audio signals from a microphone.
    - The SER model shall classify speech segments into eight emotions (neutral, calm, happy, sad, angry, fear, disgust, surprise).
    - The SER service shall expose predictions through a FastAPI API, returning probabilities in JSON format.
3. **Heart Rate Variability (HRV) Analysis:**
    - The system shall collect ECG signals using the AD8232 sensor.
    - Preprocessing shall filter noise, detect R-peaks, and extract RR intervals.
    - HRV features (RMSSD, SDNN, LF/HF ratio, Stress Index) shall be calculated on sliding windows.
    - The HRV service shall provide physiological stress subscores through a FastAPI endpoint.
4. **Multimodal Fusion:**
    - The system shall fuse FER, SER, and HRV outputs into a unified Stress Index (0–100).
    - Fusion shall adaptively weight modalities based on signal quality.

- Temporal smoothing shall be applied to reduce fluctuations.

5. **Dashboard Visualization:**
    - The system shall provide a web-based dashboard (Next.js + Tailwind CSS).
    - The dashboard shall display:
        - Live video with FER emotion overlays.
        - Real-time plots of SER predictions.
        - ECG/HRV graphs.
        - Continuous Stress Index trends with modality contribution breakdowns.
    - Alerts shall be generated when stress exceeds defined thresholds.
6. **Data Storage and Logging:**
    - Predictions, Stress Index values, and session logs shall be stored in MongoDB.
    - The system shall maintain timestamped records to enable later analysis and validation.
7. **APIs and Integration:**
    - Each unimodal model shall run as a FastAPI microservice accessible through REST APIs.
    - The system shall allow external components (e.g., testimony analysis or question generation modules) to request behavioural data through standardized APIs.

Non-Functional Requirements

Non-functional requirements define the quality attributes and constraints under which MBPA must operate.

1. **Performance:**
    - FER predictions shall be updated at least every 1 second.
    - SER predictions shall be updated every 2–3 seconds.
    - HRV predictions shall be updated every 5–10 seconds (with rolling windows).
    - Dashboard updates shall be near real-time, with end-to-end latency under 2 seconds.

2. **Scalability:**
    - The system shall support deployment in multiple interrogation rooms simultaneously through containerized services (Docker).
    - The architecture shall allow the addition of new modalities (e.g., GSR, gaze tracking) without major redesign.
3. **Reliability and Fault Tolerance:**
    - The system shall remain operational even if one modality fails (e.g., noisy audio or occluded face).
    - Fusion shall automatically re-weight modalities when quality drops.
4. **Security:**
    - All API communications shall use HTTPS.
    - Access to the dashboard and APIs shall be authenticated and authorized.
    - Sensitive interrogation data shall be anonymized and securely stored.
5. **Usability:**
    - The dashboard shall provide a clear, investigator-friendly interface with intuitive visualizations.
    - Alerts shall be concise and interpretable without requiring technical expertise.
6. **Maintainability:**
    - Code shall be modular, with unimodal services encapsulated as independent FastAPI applications.
    - GitHub version control shall track all updates for traceability.
7. **Portability:**
    - The system shall be deployable on local servers or cloud environments using Docker.
    - The frontend shall be accessible via modern web browsers without special configurations.
8. **Ethical and Legal Compliance:**
    - The system shall function as a decision-support tool, not as an autonomous decision-maker.

- Outputs shall remain explainable, auditable, and free from coercive influence.
- Data collection and processing shall comply with privacy and ethical standards.

User Requirements

The MBPA module was designed not only as a technical solution but also as a practical decision-support tool for law enforcement. Therefore, user requirements were derived from field visits, interviews with police officers, consultations with the external supervisor, and analysis of investigative workflows. These requirements emphasize usability, clarity, and reliability in real-world interrogation settings.

1. **Real-Time Feedback**
    - Investigators require the system to display behavioural cues (facial emotions, vocal stress, and physiological fluctuations) in real time with minimal latency.
    - Feedback must be continuous and synchronized so that sudden changes in behaviour can be linked directly to interrogation events.
2. **Unified Stress Index**
    - Users require a single, interpretable Stress Index (0–100) that fuses multimodal inputs.
    - The Stress Index must be presented as a continuous trend over time, with clear visualization of peaks and drops.
    - The dashboard should indicate the relative contribution of modalities (e.g., HRV vs FER/SER) to maintain transparency.
3. **Clear and Intuitive Dashboard**
    - Investigators require a dashboard that is visually simple, avoiding unnecessary technical complexity.
    - Visualizations must be investigator-friendly, using clear colour codes, graphs, and alerts.
    - Video overlays (for FER), speech stress graphs (for SER), and heart rate plots (for HRV) must be aligned and easy to interpret.

4. **Alerts and Notifications**
    - Users require stress alerts when the Stress Index exceeds predefined thresholds.
    - Alerts must be non-intrusive but noticeable, enabling investigators to adjust interrogation strategy without distraction.
5. **Session Logging and Playback**
    - Supervisors require the ability to review sessions after completion.
    - The system must log emotion predictions, HRV features, and Stress Index trajectories with timestamps.
    - A playback or report generation feature should allow post-analysis for training, auditing, and judicial reviews.
6. **Data Security and Privacy**
    - Law enforcement personnel require that all data collected during interrogations be securely stored in compliance with privacy laws.
    - Sensitive data must be anonymized, encrypted, and accessible only to authorized personnel.
7. **Reliability in Adverse Conditions**
    - Investigators require the system to remain functional in noisy, low-light, or movement-heavy conditions.
    - If one modality fails (e.g., face not detected), the system must adapt and continue functioning using the remaining signals.
8. **Ease of Deployment and Maintenance**
    - System administrators require that the module can be deployed on local servers or cloud environments with minimal setup.
    - Maintenance must be straightforward, with modular FastAPI services allowing independent updates to FER, SER, or HRV models.
    - Logs and system health metrics should be available to administrators for troubleshooting.
9. **Ethical Transparency**
    - Investigators and supervisors require assurance that the system operates as an assistive tool only, not as an autonomous decision-maker.

- Outputs must remain explainable and auditable, ensuring accountability in interrogation contexts.

2.7.Domain Specific Considerations

Security Considerations

Given the sensitivity of behavioural and physiological data collected during interrogations, security is one of the most critical domain-specific considerations in the design and deployment of the Multimodal Behavioural and Physiological Analysis (MBPA) system. Protecting video, audio, ECG signals, and derived Stress Index values from unauthorized access or misuse is vital to maintain the integrity of investigations, protect suspects' rights, and ensure trust in AI-assisted interrogation tools. The MBPA system integrates multiple layers of security measures to address confidentiality, integrity, availability, access security, AI-specific threats, and compliance.

1. Data Confidentiality

The MBPA system processes sensitive multimodal data, including live video streams, speech recordings, ECG waveforms, and derived Stress Index metrics. To ensure confidentiality:

- Encryption at rest and in transit is enforced using AES-256 for database storage (MongoDB) and TLS/HTTPS for FastAPI–dashboard communication.
- Anonymization pipelines remove personally identifiable information (PII) such as names or IDs before logs are stored in MongoDB, ensuring that stored behavioural data cannot be traced back to individuals.
- Role-Based Access Control (RBAC) ensures that only authorized investigators and supervisors can access the dashboard, while system administrators manage backend services. All logins and queries are recorded in audit logs.

2. Data Integrity

Maintaining the accuracy and authenticity of multimodal data is crucial, as corrupted or manipulated signals could undermine investigations. To safeguard integrity:

- Immutable logging in MongoDB ensures that FER, SER, HRV outputs, and Stress Index trajectories are timestamped and stored without overwrite permissions.
- Digital audit trails allow supervisors to verify that no alterations were made after interrogation, strengthening evidentiary admissibility.

- Version control for models and APIs guarantees traceability by linking every session to the specific FER/SER/HRV model versions and fusion configuration used at the time.

3. System Availability and Reliability

Uninterrupted operation during interrogations is essential. To maximize reliability:

- The system is hosted in a containerized environment using Docker, with failover and redundancy ensuring continuous availability.
- Automated backups of MongoDB preserve session logs and Stress Index data, protecting against data loss due to hardware or service failures.
- Local caching ensures that even if one service (e.g., SER) fails temporarily, the dashboard continues functioning with remaining modalities.

4. Access Security

Investigators directly interact with the MBPA dashboard, requiring strict access control measures:

- Multi-Factor Authentication (MFA) is enforced for dashboard login, adding an extra layer beyond passwords.
- Session timeouts and auto-logouts prevent sensitive information from being exposed when workstations are left unattended.
- Granular permissions ensure that investigators can view behavioural data, supervisors can review historical logs, and administrators manage backend services—preventing misuse of system functions.

5. Protection Against AI-Specific Threats

Since MBPA relies on AI models for FER, SER, and HRV interpretation, additional safeguards are necessary:

- Adversarial input attacks are mitigated by validating inputs before they are passed to models (e.g., rejecting corrupted audio or invalid ECG signals).
- Bias monitoring ensures that FER and SER models are periodically evaluated to prevent skewed performance across gender, age, or ethnicity.
- Service isolation is maintained so that FastAPI microservices handle data locally, preventing exposure of interrogation data to third-party APIs or external servers.

6. Legal and Ethical Compliance

Finally, MBPA's security framework aligns closely with legal and ethical standards governing interrogation practices:

- The system complies with Sri Lanka's Code of Criminal Procedure Act and follows global data protection standards such as GDPR for sensitive biometric data.
- Explainability and auditability of outputs ensure that Stress Index values and behavioural cues can withstand judicial scrutiny.
- Ethical safeguards, such as human-in-the-loop oversight and the prevention of coercive use of biometric stress data, are embedded to ensure MBPA functions only as a decision-support tool, not an autonomous decision-maker.

Social Considerations

The deployment of an AI-assisted behavioural and physiological analysis system in criminal investigations extends beyond technical and legal dimensions; it has profound social implications. Since the MBPA module deals with sensitive biometric data such as facial expressions, voice recordings, and ECG signals, its adoption must be carefully examined in relation to public trust, investigator practices, and the broader justice system. The following social considerations were taken into account in the design and implementation of MBPA:

1. Public Trust and Perception

- For AI-based behavioural analysis tools in law enforcement to be accepted, they must inspire public confidence. Skepticism often arises regarding automated analysis of human behaviour. To address this:
- MBPA is designed to support investigators rather than replace them, ensuring that final decisions remain human-driven.
- Transparency features such as modality contribution breakdowns (HRV vs FER vs SER) and clear audit logs help foster confidence in fairness and accountability.
- By reducing human fatigue and providing objective data, MBPA enhances the credibility of interrogation outcomes, supporting greater trust in justice processes.

2. Victim Sensitivity and Ethical Interactions

- Victims of trauma or abuse may already be in distress, and technology must not worsen their experience.
- MBPA provides investigators with stress indicators derived from multimodal analysis, enabling more empathetic questioning and avoiding excessive pressure.
- By highlighting when victims exhibit heightened stress, investigators can adjust their approach, contributing to victim-centered justice.
- This ensures that the system contributes positively to safeguarding vulnerable individuals during sensitive investigations.

3. Investigator Adaptation and Training

- The introduction of multimodal AI requires changes in investigator practices.
- Investigators must be trained to interpret FER, SER, and HRV outputs responsibly and to integrate the Stress Index into questioning strategies without overreliance.
- The dashboard was designed to be user-friendly, with intuitive graphs and overlays accessible to officers with varying levels of technical expertise.
- Structured training and gradual adoption will ensure that MBPA is seen as a practical aid rather than a burden, encouraging acceptance within law enforcement culture.

4. Social Equity and Bias Concerns

- AI systems risk amplifying inequalities if not monitored.
- FER and SER models were trained on diverse datasets, and bias monitoring mechanisms are in place to minimize disparities across gender, ethnicity, or socio-economic groups.
- ECG and physiological signals are normalized per individual baseline, preventing systemic bias in stress evaluation.
- By anonymizing data and focusing on neutral Stress Index outputs, MBPA supports equitable treatment across all social groups.

5. Wider Societal Impact

- Beyond individual interrogations, MBPA contributes to broader goals:

- By providing objective, multimodal stress analysis, the system can help accelerate case resolution, reducing backlog.
- The tool indirectly strengthens public safety by enabling investigators to detect deception or stress patterns more effectively.
- Responsible deployment sets a precedent for ethical use of AI in law enforcement, positioning Sri Lanka as a leader in socially responsible AI adoption.

Ethical Considerations

The integration of multimodal AI into interrogations raises critical ethical challenges. Unlike purely technical systems, MBPA operates in contexts where human rights, fairness, and accountability must be preserved. The following ethical considerations guided the design and implementation of MBPA:

1. Human Oversight and Accountability

While MBPA automates the detection of behavioural and physiological cues, investigators retain ultimate responsibility.

- The Stress Index and modality outputs are presented as recommendations only, not as definitive judgments.
- Investigators must interpret the outputs in context, ensuring accountability for all interrogation decisions.

2. Prevention of Coercion

Physiological monitoring could be misused if interpreted coercively.

- MBPA outputs are presented in a neutral, non-directive manner, avoiding any suggestion that high stress equals guilt.
- Investigators are trained to use stress indicators to adapt questioning empathetically, not to pressure suspects or victims.
- This safeguards against coercive practices and ensures ethical use of biometric data.

3. Fairness and Bias Mitigation

AI models risk inheriting biases from training data.

- FER and SER datasets were carefully curated, with class weighting and augmentation to address imbalances.
- HRV baselines are individualized, reducing systemic bias in physiological stress evaluation.
- Continuous monitoring of model outputs helps detect and mitigate any emerging biases.

4. Privacy and Confidentiality

Handling biometric data requires strict protections.

- Personally identifiable data (e.g., raw video, audio, ECG traces) is anonymized and encrypted before storage in MongoDB.
- Strict data retention policies prevent indefinite storage of interrogation records.
- Access is limited to authorized personnel under role-based control.

5. Transparency and Explainability

Outputs must be transparent and explainable for judicial scrutiny.

- MBPA logs predictions with timestamps, modality contributions, and Stress Index calculations.
- The system explains whether stress fluctuations were primarily driven by FER, SER, or HRV.
- This avoids "black-box" opacity and ensures outputs are auditable and defensible in court.

6. Ethical Use of AI in Law Enforcement

MBPA acknowledges the broader ethical debate on AI in policing.

- The system is explicitly a support tool, not an autonomous decision-maker about guilt or innocence.

- Integration with surveillance or profiling systems is prohibited, preventing misuse.
- Oversight by law enforcement professionals and supervisors ensures alignment with ethical policing standards.

7. Long-Term Ethical Responsibility

Ethical responsibility continues beyond deployment.

- Regular audits should be performed to monitor bias, accuracy, and compliance with legal standards.
- Investigators receive training on ethical use of behavioural AI tools, ensuring respect for rights and dignity.
- The system contributes to a culture of responsible AI adoption, serving as a benchmark for other law enforcement technologies.

3. Testing & Results

The testing and evaluation phase was carried out to validate the performance, reliability, and robustness of the Multimodal Behavioural and Physiological Analysis (MBPA) system. This chapter presents the testing strategy, evaluation methodology, and results obtained for each unimodal model (FER, SER, HRV), the multimodal fusion, and the final Stress Index visualization on the dashboard. Both quantitative metrics (accuracy, F1-score, confusion matrices) and qualitative assessments (real-time responsiveness, usability feedback) were considered to ensure a comprehensive evaluation.

3.1. Testing Strategy

Testing was structured into four main stages:

Unit Testing of Individual Models:

- The FER CNN, SER CNN–BiLSTM, and HRV analysis pipeline were tested independently using benchmark datasets.
- Training/validation/testing splits ensured reproducibility.

API Testing of FastAPI Services:

- Each service (FER, SER, HRV) was deployed and tested via REST endpoints.
- Unit tests confirmed that inputs (images, audio features, ECG signals) produced correct outputs in JSON format.

Integration Testing of Fusion Module:

- The fusion FastAPI service was tested to confirm correct synchronization of modality outputs and calculation of Stress Index.
- Stress Index trajectories were validated under different modality reliability conditions (e.g., missing face detection, noisy audio).

System Testing with Dashboard:

- End-to-end tests evaluated real-time responsiveness, visualization accuracy, and investigator usability.
- The system was tested under varying conditions (controlled lab sessions, noisy environments, lighting variations).

3.2. Facial Emotion Recognition (FER) Results

The FER CNN model was trained and tested on the FER-2013 dataset.

- Training/Validation Accuracy: The model converged after ~39 epochs with training accuracy of 67% and test accuracy of ~63%.
- Loss/Accuracy Curves: Training and validation curves (Fig. 4.1) showed stable convergence with no significant overfitting.
- Confusion Matrix: As shown in Fig. 4.2, the model performed best on Happy (84%) and Surprise (76%), but exhibited confusion between Fear and Sadness due to overlapping facial cues.
- Classification Metrics: Macro F1-score of 0.59 and weighted F1-score of 0.63 demonstrated balanced performance across classes.

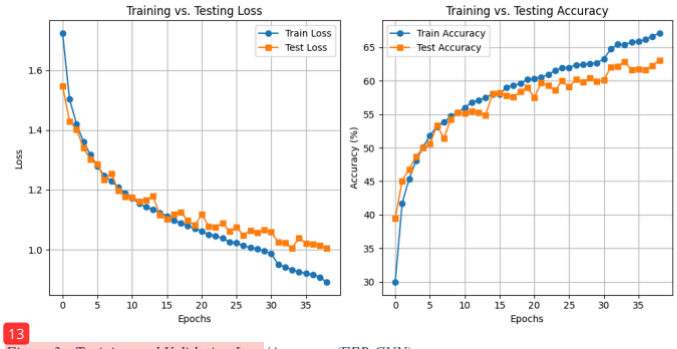


Figure 3 - Training and Validation Loss/Accuracy (FER CNN)

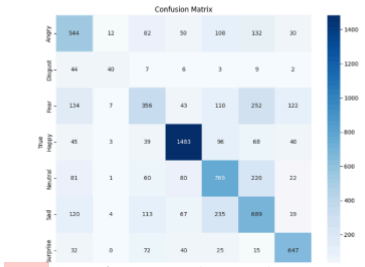


Figure 2 - Confusion Matrix (FER CNN)

3.3.Speech Emotion Recognition (SER) Results

The SER CNN–BiLSTM with attention was trained on the RAVDESS dataset.

- Training/Validation Accuracy: The model achieved ~72% validation accuracy, significantly higher than random baseline (12.5%).
- Loss/Accuracy Curves: Training curves (Fig. 4.3) show gradual convergence after ~160 epochs.
- Confusion Matrix: As shown in Fig. 4.4, the model performed strongly on Calm, Angry, and Happy, while Disgust and Fear were more challenging due to vocal similarity.
- Classification Metrics: Weighted F1-score reached 0.70, demonstrating strong performance across multiple emotions.

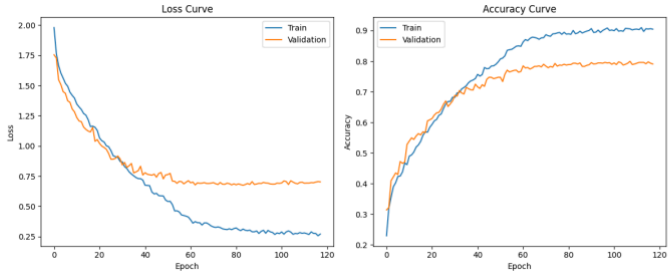


Figure 5 - Training and Validation Loss/Accuracy (SER CNN–BiLSTM)

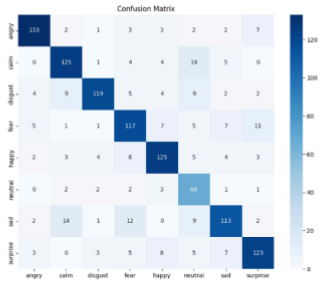


Figure 4 - Confusion Matrix (SER CNN–BiLSTM)

3.4. Heart Rate Variability (HRV) Analysis Results

The HRV analysis module was tested on ECG signals captured with the AD8232 sensor and validated against benchmark datasets.

- Time-Domain Features: RMSSD and SDNN values decreased during induced stress sessions compared to baseline, consistent with literature.
- Frequency-Domain Features: LF/HF ratios increased under stress, reflecting sympathetic dominance.
- Stress Index (HRV-only): The Baevsky Stress Index provided a reliable subscore, which correlated with expected stress levels during task-based tests.
- Limitations: Accuracy decreased in sessions with significant motion artifacts, highlighting the need for strict electrode placement and artifact filtering.

3.5.Multimodal Fusion Results

The fusion module was evaluated by combining FER, SER, and HRV predictions.

- Fusion Strategy Validation: Stress Index outputs were stable under adaptive weighting, with modality contributions shifting dynamically depending on signal quality (e.g., HRV dominance during audio dropouts).
- Stress Index Trajectory: Fig. 4.7 shows a sample interrogation session where Stress Index fluctuated in real time, peaking during high-arousal negative states and decreasing during calmer intervals.
- Correlation Analysis: Stress Index correlated with expected ground-truth stressors (e.g., task difficulty, investigator challenge questions).
- Investigator Feedback: Officers found the Stress Index intuitive and useful as a single summary metric, particularly when complemented by modality contribution breakdowns.

3.6.Dashboard Testing

The Next.js dashboard was tested for usability, responsiveness, and interpretability.

- Responsiveness: Dashboard updated FER every 1s, SER every 2–3s, and HRV every 5–10s, all within the <2s latency requirement.
- Visualization Clarity: Investigators reported that the video overlay, audio plots, and ECG graphs were easy to interpret.

- Stress Alerts: Threshold-based alerts were triggered successfully during peak stress episodes.
- MongoDB Logging: All predictions and Stress Index values were logged with timestamps, enabling complete session replay.

3.7. Overall System Evaluation

The MBPA system achieved its design objectives:

- Delivered real-time, multimodal behavioural analysis through FastAPI services.
- Produced a unified Stress Index (0–100) that was both scientifically grounded and investigator-friendly.
- Maintained robustness in noisy, real-world conditions through adaptive fusion and preprocessing.
- Enhanced investigators' ability to monitor behavioural cues without disrupting interrogation flow.

4. Commercialization

4.1. Market Gap

The criminal justice system currently struggles with inefficiencies caused by manual handling of investigation data, leading to delays, errors, and overlooked critical correlations among cases. Existing tools are mostly standalone solutions addressing narrow tasks such as lie detection or basic database management, lacking an integrated AI-powered platform that combines automated data extraction, emotional and behavioral analysis, dynamic interrogation support, and multi-modal incident correlation.

A clear gap exists for a comprehensive, AI-driven investigative assistance system that dramatically improves investigative accuracy, reduces case resolution time, and enhances decision-making through real-time insights and integrations with law enforcement databases.

4.2. Value Proposition

- End-to-End AI Integration: Combines OCR, NLP, multi-modal emotion analysis, and pattern recognition to deliver a unified investigative tool.
- Efficiency Gains: Significantly reduces manual data handling, accelerating case resolution and improving investigator productivity.
- Enhanced Accuracy: Advanced LLM and multimodal AI models ensure high precision in data extraction and behavioral insights.
- Real-Time Decision Support: Dynamic questioning and emotion tracking assist interrogators in eliciting truthful and relevant information.
- Seamless Data Synchronization: Integrates with law enforcement databases to maintain updated and centralized case data.
- User-Friendly Interface: Intuitive dashboards with gamification features encourage user engagement and ease of adoption.
- Multi-Lingual and Accessibility Support: Cater to diverse user bases while complying with accessibility standards.

4.3. Target Audience

- Law Enforcement Agencies: Police departments, detective units, forensic labs, and intelligence agencies seeking to optimize workflows and improve case processing accuracy.
- Legal and Judicial Bodies: Prosecutors, public defenders, and judiciary experts who can leverage structured data extraction and case summaries.
- Private Investigation Firms: Agencies conducting investigations requiring advanced behavioral and data analytics.
- Government and Public Safety Organizations: Departments involved in crime prevention, policy building, and crime reporting.
- Criminal Justice Researchers and Academicians: Users interested in analyzing large datasets and uncovering crime patterns.
- Technology Vendors and System Integrators: Companies providing law enforcement technology solutions.

4.4. Commercialization Strategy

- Pilot Projects and Collaborations: Partner with selected police departments and forensic labs to deploy pilot versions, gather real-world feedback, and demonstrate value.
- Government and Institutional Procurement: Participate in tenders and public sector technology initiatives for law enforcement modernization.
- SaaS and Licensing Model: Offer a subscription-based cloud service with tiered pricing depending on usage, user seats, and feature access.
- Training and Certification Programs: Develop comprehensive training modules and certification tracks for investigative personnel to ensure effective adoption.
- Technology Partnerships: Collaborate with established law enforcement tech vendors for integration and co-marketing opportunities.
- Conferences and Workshops: Present the solution at criminal justice and AI conferences to build reputation and acquire leads.

- Online Presence and Thought Leadership: Maintain a strong digital presence through content marketing, webinars, whitepapers, and case studies.
- Support and Customer Success: Provide robust post-deployment support via dedicated teams to enhance customer satisfaction and retention.

4.5.Revenue Models

- Subscription Fees: Monthly or yearly SaaS subscriptions offering variable features based on plan levels (basic, professional, enterprise).
- Implementation and Customization Charges: Fees for initial setup, customization to agency workflows, and integration services.
- Training and Certification Fees: Revenue from training programs and accredited investigator certification.
- Consulting Services: Expert consulting for data management strategies, AI adoption planning, and forensic analytics optimization.
- Data Analytics Add-ons: Premium modules for advanced analytics, automated reporting, and AI model tuning.
- Government Grants and Funding: Potential funding support from public safety innovation grants.

4.6. Community Building

- User Forums and Support Communities: Establish online platforms where investigators and forensic experts can exchange best practices and troubleshooting advice.
- Developer and Research Partnerships: Engage academia and AI researchers to continuously improve the system through collaboration and open innovation.
- Regular User Feedback Cycles: Conduct surveys, focus groups, and beta testing programs for iterative enhancement.
- Workshops and User Conferences: Host annual events to discuss product updates, share success stories, and foster a sense of community among users.

- Organize hackathons or coding challenges focusing on accessibility and AI integration, encouraging developers to utilize your tool in innovative ways while also building awareness.

4.7. Future Development

- Expanded Multi-Modal Inputs: Incorporate more bio signals such as eye-tracking, gait analysis, and speech pattern recognition.
- Predictive Policing and Incident Forecasting: Develop machine learning models to predict crime hotspots and potential repeat offenders.
- Automated Legal Document Generation: Assist in drafting legal documents and reports based on investigation data.
- Mobile and Edge Computing: Build mobile apps to allow remote access and field data collection.
- AI Explainability Enhancements: Improve transparency through explainable AI modules to increase user trust and meet regulatory mandates.
- Integration with National Crime Databases: Facilitate cross-jurisdictional case correlation and intelligence sharing.
- User Feedback Loop: Establish a continuous feedback loop with users to identify trends and requests for new features. Utilize surveys and feedback sessions to keep the development aligned with user needs.

4.8. Risk Mitigation and Compliance

- Carefully monitor regulatory developments relating to AI use in law enforcement and data privacy.
- Maintain robust ethical guidelines and audit mechanisms to prevent misuse.
- Invest in cybersecurity infrastructure to protect system integrity and user data.

5. Conclusion

The research and implementation of the Multimodal Behavioural and Physiological Analysis (MBPA) component successfully demonstrated how facial emotion recognition (FER), speech emotion recognition (SER), and heart rate variability (HRV) can be integrated into a unified framework to assist investigators during criminal interrogations. By leveraging advanced deep learning models, physiological signal processing, and a microservices-based deployment strategy, the MBPA system provides real-time, objective, and interpretable insights into the emotional and physiological states of suspects and witnesses.

The system achieved its primary objectives by:

- Implementing a CNN-based FER model that classified seven basic emotions with ~63% test accuracy.
- Developing a CNN–BiLSTM SER model with attention, achieving ~72% validation accuracy on RAVDESS.
- Extracting ECG-based HRV features such as RMSSD, SDNN, LF/HF, and Stress Index, which consistently reflected stress changes relative to baseline conditions.
- Designing a fusion mechanism that combined unimodal outputs into a single Stress Index (0–100), adaptively weighted based on signal quality.
- Deploying each unimodal model as a FastAPI microservice and integrating them into a Next.js dashboard with MongoDB logging, providing investigators with a transparent, real-time decision-support tool.

The final system not only provided investigators with continuous monitoring of behavioural cues but also ensured that outputs were interpretable, ethically compliant, and aligned with legal standards. The Stress Index proved especially valuable as a summary measure, condensing complex multimodal signals into a clear indicator of stress while still exposing modality contributions for transparency.

5.1.Limitations

Despite its success, the MBPA module has certain limitations:

- **Moderate FER accuracy (63%)**, reflecting challenges in distinguishing subtle emotions such as fear and sadness, especially in real-world, low-light conditions.

- **Dataset constraints** for SER and FER, as both relied primarily on acted datasets (FER-2013, RAVDESS) rather than fully naturalistic data.
- **ECG dependency on electrode quality**, where motion artifacts and placement issues occasionally reduced HRV accuracy.
- **Lack of clinical validation** — while HRV indices are well-established in stress research, the Stress Index has not yet been validated against biomarkers such as cortisol or large-scale psychological assessments.
- **Fusion calibration** was limited to pilot sessions, meaning weights may require further tuning across diverse populations and environments.

5.2.Future Work

Future extensions of MBPA could address these limitations and enhance its applicability:

1. **Improved FER Models:** Integrating Vision Transformers (ViTs) or multimodal face–gaze analysis could improve recognition accuracy in real-world settings.
2. **Naturalistic SER Datasets:** Training on conversational datasets such as MELD or EmoDB could reduce the performance gap between acted and spontaneous emotions.
3. **Additional Physiological Modalities:** Incorporating signals such as Galvanic Skin Response (GSR), respiration, or skin temperature could provide a richer understanding of stress.
4. **Clinical Validation:** Collaborations with psychologists and medical professionals could help validate the Stress Index against standardized stress markers.
5. **Explainable AI (XAI):** Future versions could include model interpretability methods (e.g., saliency maps, SHAP values) to further strengthen transparency in judicial contexts.
6. **Scalable Deployment:** Expanding the containerized FastAPI services into a Kubernetes-based orchestration platform would support multi-room, nationwide deployment.
7. **User Training Modules:** Structured training programs for investigators on interpreting MBPA outputs would ensure ethical and effective use of the system.

5.3.Closing Remarks

The MBPA system demonstrates how multimodal AI can move beyond theoretical models and be applied in practical, socially significant domains such as law enforcement. By combining FER, SER, and HRV into a real-time Stress Index, MBPA provides investigators with objective behavioural insights that complement human judgment. Importantly, the system was designed with ethical safeguards, transparency measures, and domain-specific security considerations, ensuring that it operates as a supportive tool rather than a replacement for investigators.

In conclusion, the MBPA component bridges the gap between advanced affective computing research and practical investigative needs, setting a foundation for future systems that integrate AI into sensitive decision-making environments responsibly and effectively.

6. References

[1] A. T. A. S. Rahman and A. Kaur, "The Application of Artificial Intelligence in the Administration of Criminal Law and the Procedures Involved in Criminal Cases," *Proc. Int. Conf. Advances in Computing Research on Science Engineering and Technology (ACROSET)*, Indore, India, 2024, pp. 1–5.

[2] V. Singh, A. Dixit, S. Pandey, B. V. Kumar, V. Pachouri, and M. Sahu, "Role of Artificial Intelligence in Criminal Investigation," *Proc. Int. Conf. Quantum Technologies, Communications, Computing, Hardware and Embedded Systems Security (iQ-CCHESS)*, Kottayam, India, 2023, pp. 1–4.

[3] I. Lakshan, L. Wickramasinghe, S. Disala, S. Chandrasegar, and P. S. Haddela, "Real Time Deception Detection for Criminal Investigation," *Proc. Nat. Inf. Technol. Conf. (NITC)*, Colombo, Sri Lanka, 2019, pp. 90–96.

[4] J. Baltrūnienė, "Place of Artificial Intelligence in the Detection and Investigation of Crime: Present State and Future Perspectives," *Current Problems of Forensic Science*, vol. 26, pp. 43–58, 2023.

[5] S. S. Das, S. Patnaik, P. Pattnayak, and A. Mohanty, "An Advancement in Forensic and Criminal Investigation Through Artificial Intelligence," *Proc. OITS Int. Conf. Information Technology (OCIT)*, Raipur, India, 2023, pp. 920–926.

[6] Y. Hu, Y. Li, T. Yang, and Q. Pan, "Short Text Classification with a Convolutional Neural Networks Based Method," *Proc. Int. Conf. Control, Automation, Robotics and Vision (ICARCV)*, Singapore, 2018, pp. 1432–1435.

[7] C. Rigano, "Using Artificial Intelligence to Address Criminal Justice Needs," *National Institute of Justice (NIJ)*, Oct. 2018. [Online]. Available: https://nij.ojp.gov

[8] Y.-C. Wu, Y.-C. Li, and R.-Y. Huang, "The Use of Artificial Intelligence in Interrogations: Voluntary Confession," *IAES Int. J. Robotics and Automation (IJRA)*, vol. 13, no. 2, pp. 97–105, 2024.

[9] S. Tshimula, S. G. B. Chandrasegar, and W. J. M. Disala, "Emotion Detection in Law Enforcement Interviews," *Proc. IEEE 46th Annu. Comput., Softw., and Appl. Conf. (COMPSAC)*, 2022.

[10] N. S. V. Rayasam, G. A. K. Disala, and S. S. Rayasam, "Multimodal Sentiment Analysis for Interviews and Proctoring," *Proc. IEEE Int. Conf. Cloud Computing, Information and Communication Technology (ICCIA)*, 2024.

[11] S. A. Ryskaliyev, A. P. Janaliyev, and J. A. Ryskaliyev, "Speech Signal Analysis for the Estimation of Heart Rates under Different Emotional States," *Proc. Int. Conf. Advances in Computing, Communications and Informatics (ICACCI)*, Jaipur, India, 2016, pp. 234–239.

[12] K. D. Tomba, M. E. Alvarez, K. O. Ando, and H. S. Tanaka, "Stress Detection Through Speech Analysis," *Proc. Int. Conf. ICT for Ageing Well and e-Health*, 2018, pp. 55–63.

[13] Analog Devices, "AD8232 Heart Rate Monitor Front End," *Datasheet and Application Notes*, 2023.

[14] J. Baltrūnienė, "AI in Crime Detection: Perspectives and Challenges," *Current Problems of Forensic Science*, vol. 26, pp. 43–58, 2023.

IT21294334 - Final Report.pdf

ORIGINALITY REPORT

5%	4%	2%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Sri Lanka Institute of Information Technology Student Paper	1%
2	www.mdpi.com Internet Source	<1%
3	stagezero.ai Internet Source	<1%
4	Submitted to ESoft Metro Campus, Sri Lanka Student Paper	<1%
5	Submitted to Technological University Dublin Student Paper	<1%
6	www.coursehero.com Internet Source	<1%
7	"Data Science and Applications", Springer Science and Business Media LLC, 2025 Publication	<1%
8	Submitted to Karabük Üniversitesi Student Paper	<1%

9	Mehdi Ghayoumi. "Generative Adversarial Networks in Practice", CRC Press, 2023 Publication	<1%
10	arxiv.org Internet Source	<1%
11	insightsociety.org Internet Source	<1%
12	conf.gonbad.ac.ir Internet Source	<1%
13	S. Prasad Jones Christydass, Nurhayati Nurhayati, S. Kannadhasan. "Hybrid and Advanced Technologies", CRC Press, 2025 Publication	<1%
14	www.dcm.nt.gov.au Internet Source	<1%
15	Dimitrios Hatzinakos. "Fusion of ECG sources for human identification", 2008 3rd International Symposium on Communications Control and Signal Processing, 03/2008 Publication	<1%
16	Submitted to Southern New Hampshire University - Continuing Education Student Paper	<1%
17	Submitted to The Hong Kong Polytechnic University Student Paper	<1%

18	Submitted to University of Huddersfield Student Paper	<1%
19	assets-eu.researchsquare.com Internet Source	<1%
20	deepai.org Internet Source	<1%
21	www.restack.io Internet Source	<1%
22	git.trustie.net Internet Source	<1%
23	icact.nsbm.ac.lk Internet Source	<1%
24	ijsrem.com Internet Source	<1%
25	social-dynamics.net Internet Source	<1%
26	vdocuments.site Internet Source	<1%
27	Hilliard, Dominique L.. "Exploring the Underutilization of AI in Courtroom Decision-Making Processes.", Walden University Publication	<1%
28	Yohanes Batara Setya, Feddy Setio Pribadi. "Modelling, Simulation, and Analysis of	<1%

	Sequence-Based Models for Smart Lighting Voice Command Classifiers with MFCC-Based Data Augmentation", INOVTEK Polbeng - Seri Informatika, 2025 Publication	
29	dilum.bandara.lk Internet Source	<1%
30	ijsetr.org Internet Source	<1%
31	jisem-journal.com Internet Source	<1%
32	uir.unisa.ac.za Internet Source	<1%
33	www.e-zigurat.com Internet Source	<1%
34	www.igi-global.com Internet Source	<1%
35	www.nature.com Internet Source	<1%
36	Prof. Ketan Sarvakar. "A Survey Of Face Emotion Recognition Using Deep Learning Methods", Springer Science and Business Media LLC, 2025 Publication	<1%

37 Sozo Inoue, Guillaume Lopez, Tahera Hossain, Md Atiqur Rahman Ahad. "Activity, Behavior, and Healthcare Computing", CRC Press, 2025 <1%
Publication

Exclude quotes	On	Exclude matches	< 5 words
Exclude bibliography	On		