Advanced personnel security system using behaviour and activity analysis

by

Nuzhat Farhin Rahman Charu
23341101
Rikth Humayun
22101895
Zabir Azmayan
21201297
Alve Ibrahim
21301065
Samira Hossain
21201567

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
December 2025

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at BRAC University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Rikth Humayun

22101895

Nuzhat Farhin Rahman Charu

23341101

Alve Ibrahim

21301065

Zabir Azmayan

21201297

Samira Hossain

21201567

Approval

The thesis/project titled "Advanced personnel security system for a nuclear power plant using behaviour and activity analysis" submitted by

1. Nuzhat Farhin Rahman Charu (23341101)
2. Rikth Humayun (22101895)
3. Zabir Azmayan (21201297)
4. Alve Ibrahim (21301065)
5. Samira Hossain (21201567)

Of Fall, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering on December 18, 2025.

Examining Committee:

Supervisor:
(Member)

Dr. Md. Ashraful Alam

Associate Professor
Department of Computer Science and Engineering
Brac University

Thesis Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi, PhD

Associate Professor and Chairperson
Department of Computer Science and Engineering
Brac University

Abstract

Safety is of utmost importance in workplaces, where human error or malicious intent can jeopardize employee safety. This thesis explores the development of an advanced personal security system, leveraging machine learning to enhance security protocols. The proposed system integrates body language detection, keystroke analysis, social media monitoring, and weapon detection within the facility to establish a robust security framework. Body language detection, powered by a Temporal Segment Network with a ResNet backbone and context fusion, precisely recognizes personnel behavior and emotional states by predicting 26 categorical emotions and 3 continuous dimensions, an GloVe-centered embedding loss enhances semantic consistency to help prevent risky actions. Social media monitoring, facilitated by web scraping with Auto-archiver, provides insights into potential security risks by analyzing the online behavior and communications of individuals with access to sensitive areas. This multi-faceted approach aims to proactively identify and mitigate threats, safeguarding the integrity of workplaces. The system aligns with IAEA standards, ensuring applicability in nuclear power plants and other high risk workplaces. By integrating machine learning with traditional security measures, this research sets new standards for security protocols in critical infrastructure, addressing evolving threats in an increasingly digital environment. The proposed personnel surveillance AI focuses on four core components: analyzing employee body language for warning signs in their emotions, scanning CCTV feed to detect weapons in the hands of employees, surveilling employee keystroke patterns to detect abnormalities in their typing and conducting sentiment analysis of social media activities. This innovative approach enhances the efficiency of security surveillance, offering a proactive solution to ensure the safety and security of workplaces.

Keywords: Deep Learning, Computer Vision, Supervised Learning, Natural Language Processing (NLP), Auto-archiver, Real Time, YOLO, Multi-modal, Personnel surveillance.

Table of Contents

Chapter 1

Introduction

1.1 Background

We think of Nuclear power plants (NPPs) and bank vaults and other critical workplaces as a basis to develop our advanced personnel security system. Especially, NPPs sustain the energy demands of an ever-increasing world. More so, the sensitive nature of these plants and the severe ramifications of serious breaches call for a unique level of safety and security. Historically, the key physical infrastructures for the security systems of NPPs or bank vaults have included surveillance cameras, biometric access control systems, and perimeter fencing. While these measures do address security from external threats, they are often inadequate against most insidious threats.
Comprising individuals with authorized access to critical areas, insider threats present unique challenges. These include deliberate acts, such as sabotage, as well as unintentional acts, such as negligence or human error. The growing complexities of these challenges require that the seucrity industry transition toward more advanced technologies with an emphasis on proactive and dynamic threat detection. Currently, advances in artificial intelligence (AI), machine learning (ML), and sensors offer new possibilities for bolstering security. Behavioral and activity analysis have proved useful for monitoring or observing human actions, detecting anomalies, and therefore forecasting threats. When integrated into personnel security, these approaches maintain a workplace's shift from static defense to adaptable and real-time threat management.

1.2 Rational of the Study or Motivation

Despite there being multiple object detection and video analysis based models to ananlyze weapons and threat to security, there does not exist multimodal alanysis system in the field of personnel surveillance. Our goal is to see if a multimodal analysis surveillance system can be developed to make security systems more robust and potentially flag insidious threats to make high security workplaces more secure. We are implementing machine learning based models so that the system can adapt to real life scenarios and reduce human error in surveillance.

1.3 Problem Statement

While classical methods of security can subdue almost all external threats to sensitive workplaces, it is weak to insider threats. Human surveillance is also prone to errors, especially in large workplaces. To reduce the possibility of human error and bias while improving surveillance quality, it is important to implement more advanced surveillance measures. However, despite the advancement in ML based solution, there is a lack of multimodal analysis for surveillance. Because of this, AI based surveillance technologies are missing out on various dimensions of data that might be necessary to make a system more robust and accurate at detecting criminal activities.

1.4 Objective

The primary objective of the study is to design and develop an advanced personnel security system (APSS) for sensitive work places employing behavior and activity analysis.

1. To detect and deter insider threats through continous monitoring and anomaly detection.
2. To make a smart threat level detection based on threat detection from CCTV feed, body language analysis, continuous user verification via keystroke dynamic analysis and sentiment analysis from social media.
3. Utilize advanced technologies of AI, ML and computer vision to conduct exhaustive threat assessment.
4. Provide security measures that are dynamic and adaptive instead of static while reducing surveillance fatigue for surveillance employees.
5. Enhance safety of personnel and security of critical infrastructure.
6. Compliance with industry regulations and standards can be assured through the system. The system has been designed after taking the safety guidelines from IAEA and expert opinion into account to ensure usability.

1.5 Methodology in Brief

For real time object tracking, weapon detection and body language study, we will be using YOLO. For emotion detection from body language, we will use a two-stream Temporal Segment Network (TSN). For social media monitoring we will use Bellingcat Auto-archiver for web-scrapping employee social media posts and we will implement deberta-v3-base-absa for aspect based sentiment tasks.

1.6 Scopes and Challenges

While the primary application of our model is security surveillance within high risk workplaces, its multimodal approach and the robustness we aim to achieve enable its adaptability to a wide range of security surveillance scenarios. The model can be deployed in various sensitive professional workplaces, such as critical infrastructure, government offices and research facilities. The implementation of multiple analytical models means we have to acquire sufficient data for each modality. The type of data includes images, videos and time-series data. As such, obtaining high-quality datasets for training is a significant challenge. Moreover, another challenge is making sure the results are interpretable by the end users. Lastly, mitigating harmful biases within the model is an important consideration, as harmful biases can lead to false positives and false negatives. Addressing these challenges is essential to ensure that the system is reliable.

Chapter 2

Literature Review

2.1 Preliminaries

Some workplaces are need to be heavily secured due to the role they play in the society. One of these workplaces are nuclear power plants (NPPs), which play a significant role in energy generation and are vulnerable to insider threats or safety breaches. Bank vaults, high risk research facilities, government workplaces and other high risk workplaces have similar threats. As such, maintaining the security of these facilities is a critical issue. Conventional security techniques are stressed and challenged against advanced attacks, which is why it is important to complement the surveillance and detection techniques with Artificial Intelligence (AI) and Deep Learning (DL).

Theories and Definitions:

Risk-Informed Physical Security (RIPS):

> By integrating security effectiveness in risk assessment models, the never-ending supplementary security protocols, added authorities and processes, are in the rear-view mirror and the journey towards dynamic, intelligence-leveraged, security strategies are well underway.

Predictive Maintenance and Human Error Detection:

> Mistakes by human operators can compromise the safety of workplace. For example, generative adversarial networks have been proposed and implemented in manual surveillance data vs. sensor logs anomaly detection, which enhances the security of such systems through early error detection.

Cyber Security Threat Models in High risk workplace:

> AI-based Intrusion Detection Systems (IDS) have shown to be much better than their traditional signature-based counterparts in discovering newly emerged cyber threats. Various Deep learning models (DNN, Autoencoders, SVMs, etc.) have been used for zero-day attacks detection.

Insider Threat Behavior Analysis:

Anomalous activities alerting red flags of actions by employees that may threaten servers and data can be detected by using user behavior machine learning analytics. Clustering algorithms (K-means, Bayesian classifiers, SVMs) to review user activity logs and predicting a potential security risk.

Concepts and Models:
AI for Live Video Surveillance

Traditional CCTV surveillance is not only labour-intensive but also fraught with error. Real-time tracking suggestions using models based on AI as:

- Behavioral analysis using Long Short-Term Memory (LSTM) networks
- Real-time object detection with YOLO (You Only Look Once). Using it to detect weapons and suspicious individuals.

Security technologies that are deployed on-site or in-house, such as artificial intelligence (AI) and machine learning (ML) tools, improve real-time threat recognition.

Social Media tracking for future threats deterrent

The planning of attacks by criminal networks on social media is a common occurrence. Social media posts are used for sentiment analysis and Named Entity Recognition (NER). AI-driven models, such as:

- Natural language Processing (NLP)
- Sentiment classifiers based on transformers assist in identifying suspicious actions and possible threats.

Application of Artificial Intelligence and Deep Learning Methods for Anomaly Detection in Workplaces

Such detection is necessary for cyber threats, equipment malfunctions and threats from insiders in workplaces. AI models and techniques such as:

- You Only Look Once (YOLO)
- Long Short-Term Memory (LSTM)
- Transformer based encoder
- 3D Convolutional Neural Network (3D CNN)
- Cross modal attention mechanism
- Aspect Based Sentiment Analysis (ABSA)
- Decoding-enhanced Bidirectional Encoder Representations from Transformers with Disentangled Attention (deBERTa)

2.2 Review of Existing Research

Ayodeji et al. (2022) [18] addresses the application of deep learning (DL) techniques to enhance the safety assessment of nuclear power reactors. The authors aim to improve the reliability, explainability, and efficiency of DL models in predicting complex nuclear reactor safety phenomena, such as thermal-hydraulic behavior, turbulence modeling, and severe accident management. The study highlights the limitations of traditional empirical and semi-mechanistic models, which are often expensive and limited by experimental conditions. By leveraging DL, the authors propose data-driven approaches to overcome these challenges, particularly in areas like turbulence modeling and reactor safety analysis. The methodology involves the development of a deep feedforward neural network (DFNN) as a surrogate model to predict turbulent eddy viscosity in Reynolds-averaged Navier-Stokes (RANS) simulations. The DFNN is compared with the conventional Spalart-Allmaras closure model to evaluate its performance in terms of convergence speed and accuracy. The authors also employ explainability tools, such as Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME), to interpret the predictions of the DFNN model. These tools help in understanding the contribution of individual features to the model's predictions, thereby improving trustworthiness in DL-based safety analysis. The study is conducted using Python frameworks like TensorFlow and OpenFOAM, with the DFNN model trained on simulation data generated from RANS simulations of airflow over a backward-facing step. The results demonstrate that the DFNN model significantly accelerates the convergence of RANS simulations, reducing the number of iterations required compared to the Spalart-Allmaras model. The DFNN achieves comparable accuracy in predicting turbulent eddy viscosity and velocity profiles, as validated by qualitative and quantitative assessments. The SHAP and LIME tools provide insights into the importance of input features, such as vertical and horizontal velocity components, in influencing the model's predictions. These explainability tools enhance the interpretability of the DL model, which is crucial for its application in safety-critical nuclear environments. The dataset used for training the DFNN model consists of simulation data generated from RANS simulations of airflow over a backward-facing step. The data includes variables such as horizontal and vertical velocities, coordinates, and step height, with Reynolds numbers ranging from 34,000 to 41,500. The dataset is divided into training and validation sets, with 90% used for training and 10% for validation.

Mandalapu et al. (2023) [30] addresses the application of machine learning (ML) and deep learning (DL) techniques for crime prediction, aiming to identify patterns and trends in criminal activities to enhance law enforcement strategies. The study systematically reviews over 150 articles to explore various ML and DL algorithms used in crime prediction, focusing on their effectiveness in predicting crime patterns, detecting criminal activities, and providing insights into factors influencing crime. The authors aim to bridge gaps in existing research by highlighting datasets, methodologies, and future directions for improving crime prediction accuracy. The methodology involves a systematic review of research articles from databases such as IEEE, Science Direct, and ACM. The authors use a combination of automated

and manual search techniques to identify relevant studies, applying filters based on keywords, document titles, and index terms. They analyze the selected papers to extract information on the algorithms used, datasets, and performance metrics. The study categorizes the approaches into ML-based regression and classification methods, as well as DL-based regression and classification techniques, to provide a comprehensive overview of the field. The results indicate that traditional ML algorithms, such as decision trees, random forests, and support vector machines, are effective in predicting crime patterns and identifying correlations between crime incidents and environmental factors. DL algorithms, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), show promise in handling complex datasets, such as video footage and social media data, for real-time crime prediction. The study highlights the importance of interpretable models and the need for high-quality, reliable data to improve prediction accuracy. Additionally, the authors identify challenges related to ethical concerns, privacy, and the scalability of these technologies in real-world applications. The datasets used in the reviewed studies include publicly available crime data from cities such as Chicago, New York, London, and Los Angeles, as well as global datasets containing CCTV footage, social media data, and weather information. These datasets provide valuable insights into crime patterns and are used to train and evaluate ML and DL models. The study also lists the datasets in a table, making them accessible for future research.

Shenoy et al. (2022) [22] addresses the challenge of real-time crime prediction and detection using an intelligent framework that leverages artificial intelligence (AI) and computer vision techniques. The authors aim to automate the monitoring of CCTV footage to identify criminal offenders, detect weapons, and analyze suspicious human behavior, thereby reducing the reliance on manual observation and improving public safety. The proposed system is designed to alert authorities when signs of criminal activity are detected, making it a proactive tool for crime prevention. The methodology involves the integration of several AI-based techniques, including object detection, face detection, face recognition, and suspicious behavior analysis. For object detection, the authors employ the SSD Mobilenet model, which is trained on the DaSCI Weapon Dataset to identify hazardous objects such as knives and guns. Face detection is performed using the Haar Cascade Classifier, while face recognition is implemented using a pre-trained model from the dlib library in Python, which generates face embeddings for comparison with a criminal database. Suspicious behavior analysis is conducted using a ResNet-GRU model, which combines a 50-layer ResNet for feature extraction and Gated Recurrent Units (GRUs) to analyze sequences of frames for suspicious actions. The system is integrated using multiprocessing to run these components in parallel, improving computational efficiency. The results demonstrate the effectiveness of the proposed framework. The SSD Mobilenet model achieves an accuracy of 82% in detecting weapons, while the face recognition system accurately identifies known individuals in real-time. The ResNet-GRU model outperforms LSTM and RNN models, achieving a 95.97% accuracy in predicting suspicious behavior. The use of multiprocessing significantly enhances the system's performance, allowing it to process up to 30 frames per second for object detection and 24 frames per second for crime prediction. The datasets

used in the study include the DaSCI Weapon Dataset for object detection, which contains images of 91 hazardous objects, and the CAVIAR Dataset for suspicious behavior analysis, which includes video clips of six different human actions. Additionally, a custom dataset of faces was created for face recognition, consisting of 30 images per individual captured via a webcam.

Wankadu et al. (2022) [24] presents a comprehensive survey on sentiment analysis, focusing on its methods, applications, and challenges. Sentiment analysis, also known as opinion mining, is the process of identifying and extracting subjective information from text to determine the sentiment polarity (positive, negative, or neutral) expressed by individuals. This is particularly useful for businesses, governments, and organizations to monitor public opinion, improve decision-making, and enhance customer satisfaction. The authors aim to provide a thorough overview of sentiment analysis techniques, evaluate their advantages and disadvantages, and identify future research directions. The methodology employed in the survey involves a detailed examination of various sentiment analysis approaches, including lexicon-based methods, machine learning techniques, and hybrid models. Lexicon-based approaches rely on predefined sentiment lexicons to assign polarity scores to words, while machine learning methods, such as Naive Bayes (NB), Support Vector Machines (SVM), and deep learning models like Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN), are used for supervised and unsupervised sentiment classification. Hybrid approaches combine lexicon-based and machine learning techniques to improve accuracy. The authors also discuss advanced methods like Aspect-Based Sentiment Analysis (ABSA) and Multimodal Sentiment Analysis (MSA), which incorporate additional data modalities such as audio and visual information. The results of the survey highlight the effectiveness of different sentiment analysis techniques across various domains. Machine learning methods, particularly deep learning models, have shown superior performance in handling complex sentiment analysis tasks, especially when combined with lexicon-based approaches. However, challenges such as detecting sarcasm, handling informal language, and managing domain-specific sentiment lexicons remain significant hurdles. The survey also identifies the need for more robust models to address these challenges and improve the accuracy of sentiment analysis in real-world applications. The datasets used in the studies reviewed in the survey are diverse, including social media data (e.g., Twitter, Facebook), product reviews (e.g., Amazon, Yelp), and domain-specific datasets (e.g., healthcare, stock market). These datasets are crucial for training and evaluating sentiment analysis models, as they provide a wide range of textual data with varying sentiment expressions. The authors emphasize the importance of high-quality, domain-specific datasets to achieve accurate sentiment classification.

Sivakumar et al. (2021) [15]addresses the challenge of real-time crime detection using deep learning algorithms to enhance public safety. The primary goal is to detect criminal activities and weapons in real-time video feeds from CCTV cameras, thereby enabling immediate alerts to law enforcement authorities. This system aims to reduce crime rates by identifying potential threats before they escalate, leveraging

advanced object detection techniques to monitor crowded and public spaces effectively. The methodology employed in this study is based on the YOLO (You Only Look Once) algorithm, a deep learning object detection framework that uses a single convolutional neural network to predict bounding boxes and class probabilities simultaneously. The YOLO algorithm is chosen for its efficiency and speed, making it suitable for real-time applications. The system is trained using a custom dataset comprising high-resolution images of criminals and weapons, annotated using the ybat annotation tool. The training process is conducted on Google Colaboratory, which provides access to a Tesla K80 GPU for accelerated computation. The Darknet framework is used to implement YOLOv4, and the model is trained for 20,000 iterations to maximize accuracy and minimize loss. The results demonstrate that the proposed system achieves a mean average precision (mAP) of 78.3%, significantly outperforming the Fast R-CNN model, which achieves an mAP of 62.4%. The YOLO-based system processes images at 45 frames per second, with an average detection time of 44 milliseconds per image. The precision and recall rates for YOLO are 98.56% and 91.21%, respectively, indicating high accuracy in detecting criminals and weapons even in complex and crowded scenes. The dataset used for training and testing consists of 100 high-resolution images per criminal, along with images of weapons such as knives, guns, and pistols. The images are annotated with bounding boxes and labels using the ybat tool, ensuring precision and consistency. The dataset is split into an 80:20 ratio for training and testing, respectively, to validate the model's performance on unseen data.

Khalid et al (2023) [28] trained YOLO based object detection model to detect weapons and suspicious person and achieved 95% accuracy. This proves that YOLO is reliable for weapon and threat detection and can offer high accuracy compared to other object detection algorithms. In their paper, they reviewed methods such as Fast and Faster R-CNN, SVM, DCN,Mobile SDD, Mask RCNN etc. Because of the high accuracy, low latency and real time detection capabilities, YOLO is an excellent choice for developing a weapon and threat detection model.

Filntisis et Al (2020) [11] trained a TSN-based architecture with three modalities on the BoLD (Body Language Dataset), achieving mAP of 0.1796, mRA of 0.6416 and mR^2 of 0.1141. They utilized spatial, context and optical flow modalities. Since it is quite difficult to predict human emotions from body language alone, GloVE (Global Vectors) vectors were used to map the 26 discrete emotions and show their relationship with one another to help the model give more accurate detections. This approach improved the model's performance so it shows that TSN-based architecture is good for the BoLD dataset.

Agarwal and Hall (2024)[37] explored the barriers to adopting artificial intelligence and machine learning in nuclear power. Compared to other energy producing industries, the nuclear power industry relies on a large workforce. Therefore, there is a huge incentive to adopt large scale advanced technologies, as this would lower the cost of production. AI/ML is one such technology that would lower the production

costs. The barriers that are discussed in the paper are: end-user acceptance, stakeholder readiness, historical, business and technical barriers. Historical barriers are due to past events such as Fukushima Daiichi, Chernobyl, Three Mile Island, slow pace of development and strong regulations. Because of this, the rate of adoption of AI/ML in nuclear power plants is slow. Technical barriers are due to potential cybersecurity risks, lack of governance framework to oversee AI/ML development and operation and finding the required technical expertise for adoption of the technology. Data access is also an issue, as much of the data is stored in closed loop systems within the plant, rendering it inaccessible. Data quantity is also a consideration, as a result in recent years NPPs have invested in digitizing infrastructure to collect previously unrecorded or manually recorded data (Agarwal et al., 2016; Ma et al.m 2022)[5][21]. The business barriers are mostly related to cost, such as upfront cost and expertise required for AI/ML adoption. It is difficult to gauge the cost-savings in the short term, as the benefits are realized in a long term timeframe. Barriers stemming from stakeholder readiness are due to the culture of the nuclear power industry, which prizes risk-aversion and safety. Lastly, end-user adoption considers user experience (UX), ethical barriers and individual differences between adoptions. The interface in the UX should be consistent with the task the user is to perform (Braseth et al., 2009)[1]. Moreover, the US Nuclear Regulatory Commission has outlined some characteristics that should be present in the information display (O'Hara et al., 2020)[13].

Saketh et al. (2023)[32] introduced a system called Intelligence Surveillance Support System (ISSS). This system supports real time monitoring of video data through monitoring the frame for motion, monitoring a portion of the frame for motion, identifying the person, detecting noises, detecting visitors, detecting changes in environment and general video recording. The purpose of such a system is to provide an efficient and fast solution for a surveillance system. This is opposed to the traditional manual surveillance systems, where it is time-consuming and error-prone. In contrast, the ISSS uses machine perception and real-time analysis to perform surveillance tasks. The system uses Tkinter, OpenCV, numpy and cameras in tandem to analyze the data. For night vision, the cameras use LED, flashlights or have night vision capabilities built into them. Their system has a GUI and buttons for each of its functions. The ISSS uses Structural Similarity Index Metric, which is used to measure how similar images are to each other. For facial recognition, the Local Binary Pattern Histogram (LBPH) technique and Haar cascades are used. Thresholding and contour detection is applied to consecutive frames in order to detect motion. Visitor detection is implemented through using OpenCV and date-time libraries, where the time stamped images are taken and saved into different folders depending on the direction of movement of the person. Sample results show that for frame motion monitoring, NO-MOTION is shown when no motion is identified, and MOTION is shown when motion is identified in the frame. For the subframe motion monitoring, the same result is seen, NO-MOTION and MOTION shown when no motion and motion are identified respectively. Moreover, for the person identification module, when a person entry in frame is detected, the image is saved along with timestamp in a specified folder. Lastly, the environment change detection module is able to detect the object being removed from the environment. The

authors of this paper used the Lamar University Database, AT&T database of faces and 5_Celebrity faces datasets for their model.

Chang et al. (2014)[2] addresses the interoperability challenges inherent in video and environment surveillance systems deployed across unattended power stations within smart grids. Due to the proliferation of manufacturers, devices employing different communication protocols, proprietary SDKs, and incompatible image compression standards (e.g., MPEG4, H.264) prevent seamless integration. These issues create isolated subsystems known as information islands, obstructing unified monitoring and increasing maintenance complexity. A simulation-based test system to validate protocol consistency and ensure interoperability among devices, thereby resolving communication bottlenecks between platforms such as video surveillance platforms (VESP), network video recorders (NVR), and IP cameras (IPC) is proposed. The test system substitutes physical devices with simulators, retaining only the unit under test (UUT) as real hardware. The tools used include a SQL Server for managing device configurations, test parameters, fault databases and protocol-specific simulators, which include Interface A (HTTP, SIP, RTSP), Interface B (HTTP, SIP, RTCP), 104 telecontrol protocol and camera simulators. The software architecture modularizes functions into protocol analysis, message parsing, error logging, and automated report generation. For instance, the camera simulator emulates 16 IP camera channels, configures H.264 video streams, and tests RTP/RTSP packet transmission to NVRs. The results show the system's efficacy in identifying protocol inconsistencies and interoperability flaws. A case study validating Interface B between NVR and VESAP utilized a DELL Alienware notebook running the simulator and physical NVR hardware. The simulator successfully tested functionalities such as resource acquisition, PTZ control, and event subscription, with errors logged and detailed in real-time. Moreover, the results confirmed resolution of intercommunication issues, enabling seamless deployment in operational environments. The framework also supports pre-bid compatibility assessments and periodic equipment audits. While the authors did not use external datasets, they synthesized test parameters and protocol specifications from grid surveillance standards (e.g., Q/GDW 517.1-2010). Simulated data structures include APDU (Application Protocol Data Unit) formats for 104 telecontrol protocols, encompassing APCI headers and I/S/U format control domains. Video streams leverage H.264 compression, and camera templates enable parameter customization for emulating diverse IPC models.

Katser et al. (2022)[20]proposed a machine learning (ML) framework to automate anomaly detection and diagnosis in nuclear power plant (NPP) power transformers. The system addresses the limitations of traditional chromatographic analysis of dissolved gas (CADG) methods, which rely on predefined gas concentrations thresholds and manual oversight, leading to delayed fault identification. In contrast, the system uses supervised learning to classify transformer states into four operational modes: normal operation, partial discharge, low-energy discharge, and low-temperature overheating. This enables real-time monitoring and timely intervention to prevent equipment failure. The ML algorithms used are logistic regression, decision trees, random forest, gradient boosting and multilayer perceptron (MLP), with

feature engineering using the tsfresh library to extract statistical properties (mean, max, std) from time-series gas concentration data. A StackingClassifier ensemble which combines LightGBM, Random Forest, and MLP models, with XGBoost as a meta-learner was implemented to enhance classification accuracy. The system processes 420-point time-series data, reduces dimensionality by removing correlated features (Pearson coefficient 0.9), and employs precision, recall, and F1-score metrics to address class imbalance. Key functionalities include distinguishing between normal and fault states, such as detecting partial discharges (local dielectric breakdowns) through elevated gas concentrations and identifying low-temperature overheating via coolant system inefficiencies. Sample results demonstrate the ensemble's superior performance, achieving an F1-score of 0.974 on the test set. For normal mode detection, the system attained 99% F1, while partial discharge classification achieved perfect recall (1.0). Misclassifications primarily occurred in low-energy discharge (78% F1) and low-temperature overheating (94% F1) cases, often confused with normal operation due to overlapping gas patterns. The authors introduced a publicly available Kaggle dataset, comprising 3,000 labeled instances of gas concentration time-series data across four operational modes. Each instance includes 12-hour interval measurements over 420 time points, preprocessed into 16 statistical features.

(Gursel et al., 2022)[27] discusses the essential topic of detecting human error (HE) in nuclear power plants (NPPs), which is a major concern due to the potential for serious consequences such as safety hazards and operational disruptions. The authors present an unsupervised anomaly detection system that use Generative Adversarial Networks (GANs) to discover inaccuracies in manually gathered surveillance data by finding mismatches between automatically recorded sensor data and manually collected data. The study's goal is to improve the safety and dependability of nuclear power plant operations by detecting potential human errors early on, allowing for appropriate corrective steps. Gursel et al. (2022) utilizes an unsupervised anomaly detection model that is based on a GAN and is trained to detect discrepancies between sensor data and manually collected surveillance data. The GAN model is evaluated against two cutting-edge unsupervised anomaly detection algorithms: One-Class Support Vector Machine (OCSVM) and Isolation Forest (iForest). Moreover, The GAN architecture consists of an autoencoder network that serves as a generator, an encoder network, and a discriminator network. In order to detect anomalies, the model utilizes adversarial, contextual and encoder loss afterwards being trained on non anomalous data. Whether adding a prior sensor data improves the anomaly identification performance or not, this research uses temporal data window to check it. For validating the model's performance are under numerous assumed rates of anomaly the test aims even though the true rate of anamoly for NPP dataset remain unknown. Additionally, Gursel et al. (2022) stated that in terms of G-Mean value, GAN model outperforms both iForest and OCSVM in anomaly detection. GAN model is validated through comparisons with baselines under a variety of assumptions for anonyrmity rates (1%, 2.5%, 5%, and 10%) for NPP unlabeled dataset. In the testbed dataset with known labels, the GAN technique demonstrates consistent superiority, regardless of whether the data definitions are location-specific or holistic. According to the Gursel et al. (2022), the GAN-based framework is suc-

cessful in the detection of human errors and provides a basis for establishing HITL systems into NPP operations. This study uses a dataset consisting of two parts: a real-world NPP dataset and a testbed dataset. The NPP dataset comprises automatically gathered sensor data and manually obtained surveillance data from a primary feed pump, encompassing vibration and turbine speed measurements from 2016 to 2020. The testbed dataset was acquired from a nuclear engineering testbed and comprises sensor data (temperature, coolant flow, and vibrations) as well as manually collected surveillance data, with intentional human errors introduced during data collection. The testbed dataset contains known labels, which can be used to validate anomaly detection models.

Chaudhary et al. (2024)[35] addresses the the significant issue of detecting and examining anomalies in nuclear power plants which occurs because of the cyberattacks. The key intention of the authors is to build a reliable anomalous behavior discovering system that can detect anomalous behavior in NPPs even prior to present protective actions, including Reactor Protection System (RPS). Again, Chaudhary et al. (2024) seeks to improve the security and safety of nuclear power plants by utilising deep learning models and explainable AI (XAI) to identify and analyse abnormalities in real-time, thus reducing the dangers posed by cyber threats to critical infrastructure. Moreover, The Asherah Nuclear Power Plant (NPP) simulator, a virtual testbed implemented in MATLAB/Simulink, is employed in the study to simulate realistic NPP operations and introduces three different kinds of cyberattacks aiming at various systems inside the plant. Anomaly detection is executed via a Bidirectional Long Short-Term Memory (Bi-LSTM) model, a deep learning framework adept at handling time-series data. The Bi-LSTM model is trained solely on regular operational data, allowing it to identify variations that suggest probable anomalies. The authors also use explainable artificial intelligence (XAI) methods—more especially, SHapley Additive exPlanations (SHap)—to understand the model's forecasts and pinpoint the aspects supporting anomaly detection. The model's decision-making process is strengthened with this method's clarity and comprehensibility through the technique. The Bi-LSTM model detected all three attack scenarios, designed for pressuriser spray valve, feed water pump, and condenser cooling pump accurately, even before starting the RPS. All three attack scenarios gained a recall value of 1.0, a 0.0 value for a false positive, and an accuracy value of 1.0, showcasing its high reliability in terms of anomalous behavior detection. XAI, in its use, facilitated significant analysis of factors responsible for generating anomalous behavior, such as a variation in flow and pressure, important for knowing the cause of detected anomalous behavior. Two models, namely, TranAD, a transformer model for anomalous behavior detection, and Bi-LSTM model, have been considered in work. Chaudhary et al. (2024) stated that as both have high performance, Bi-LSTM model is opted for its interpretability and simplicity. With separate datasets for both normal operations and each attack scenario, it has 17 features—including reactor power, coolant temperature, and pump speeds. The training dataset comprises 3000 instances, whereas each attack scenario dataset contains 1200 examples, each lasting around 2 minutes.

(Yavuz & Lüle, 2022)[26] deals with the potential for Artificial Intelligence to help make nuclear power plants safer and more efficient in their operations. More focus has been given to reactors of the VVER-1200 type. Here, the authors addresses real-time decision-making difficulties associated with reactor transients, which are key for safety or accident prevention. The project intends to improve the overall safety and reliability of NPP operations by using AI to reduce human malaise and computational burden in the identification and response to transients. Artificial Neural Networks were a major AI approach whereby skilled training was developed to identify complex patterns and relationships in large datasets.The study employed the use of Particle Swarm Optimisation to help optimise core fuel configurations for better reactor performance. (Yavuz & Lüle, 2022) propose a semi-autonomous AI control system to help reactor operators detect transients and suggest the appropriate corrective action. It has been engineered to deal with incoming data streams and to provide operators with actionable outputs. For training data generation, Yavuz and Lüle (2022) use RELAP5, which is a simulation tool for simulating NPP dynamics during accidental conditions. Steady-state operations and transients are simulated using RELAP5, therefore generating a dataset with temperature, pressure, and void fraction measurements at key reactor system locations. Then, neural networks are trained using this dataset for transient detection and response. In fuel failure detection, artificial neural networks surpassed conventional isotopic ratio techniques, providing swifter and more precise diagnoses. In transient identification, artificial neural networks attained satisfactory error margins, facilitating the creation of decision-support systems to aid operators.

Yang et al. (2024)[42] created a state-of-the-art model for aspect-based sentiment analysis. This was done using local sentiment aggregation (LSA). Local sentiment aggregation works using sentiment coherency, which is the phenomenon of people expressing similar sentiments towards related subjects. Sentiment coherency is extracted and aggregated to aid in sentiment classification. This results in state-of-the-art performance on popular sentiment analysis datasets such as Laptop14, Restaurant14, Restaurant15 and Restaurant16. Moreover, Yang et al. (2023)[34] authored a user-friendly framework to perform aspect-based sentiment analysis for any task using Python.

Yadav et al. (2022)[25] explores how to make security at nuclear power plants better by using a risk-based approach. The authors want to make these plants safer, more efficient, and less expensive to run by using risk-based methods to assess their security. Usually, security checks at these plants follow very strict rules. This can result in designs that are too cautious and not very efficient. Yadav et al. (2024) proposed methodology aims to solve these constraints by linking physical security effectiveness to risk surrogates such as core damage frequency, thereby facilitating a more sophisticated and accurate assessment of security postures. To attain this, a variety of techniques and tools are utilized. Overall, the model utilizes adversary sequence diagrams (ASDs) to represent attack routes, with a detection probability and a traversal time for each barrier in a facility. There is a dynamic modelling and simulation (M&S) technique, with Diverse and Flexible Mitigation Capability

Strategies (FLEX) gear being utilized with force-on-force (FOF) simulations. The FOF simulations, executed with commercial software, yield attack chronology data and target status, which are subsequently integrated into the Event Modelling Risk Assessment using Linked Diagrams (EMRALD) tool. EMRALD actively simulates the uncertainties in FLEX planning and execution, producing statistical distributions of deadlines for FLEX equipment deployment. Furthermore, thermal-hydraulic analysis utilising RELAP5 is conducted to assess the time to core damage, factoring in uncertainties related to safety system activation and operator efficacy. The study shows that combining FLEX techniques with traditional security measures greatly increases the safety margin of NPPs. The cumulative core damage probability (CCDP) decreased across multiple attack scenarios with the implementation of FLEX methods. Furthermore, the constant improvement of security personnel revealed possible cuts in guard stations without endangering public safety. The approach also enabled a change from binary security results to probabilistic assessments, therefore allowing complex evaluations of security efficacy. The dataset in this study contains made-up attack scenarios and specific FLEX methods for different plants. Attack timeline data is produced by the FOF simulations, and the dynamic uncertainties in FLEX preparation and execution are modelled by EMRALD. RELAP5's thermal-hydraulic analysis combines operator-action timeframes and uncertainty sources—including random failures in safety system components. Statistical distributions of time to core damage and evaluation of the efficacy of several security postures are generated from these datasets.

(Kumari et al., 2023)[29] stated that nuclear power plant security is an important topic due to the growing complexity and frequency of cyber-attacks. Examples of such threats are malware breaching the systems of these facilities, phishing attacks and zero-day exploits that can compromise safety, disrupt operations, or lead to leakage of sensitive data (Kumari et al., 2023). A combination of deep learning approach to handle non-linear data such as Intrusion detection and non-intrusive and non-signature-based methods that can detect and differentiate them from human irregularities. But the lack of this access, necessitates the deployment of advanced security mechanisms that can identify, analyze and respond to advanced cyber threats in real-time. To tackle this problem, the authors provide an efficient AI-based security framework that leverages machine learning (ML) and deep learning (DL) techniques to support cybersecurity countermeasures concerning nuclear power plants. Something can be done to capture a large amount of real-time security data via different network sensors and logs and that data can also be analyzed by a system that will aid in identifying the malicious activity. Unlike traditional Intrusion Detection System (IDS) functions, this AI-based model keeps learning from new attacks, and modifying itself accordingly, improving the accuracy of its detection over the years (Kumari et al., 2023) The study methodology uses supervised and unsupervised learning, autoencoders, DNN, and SVM algorithms. These steps include preprocessing of raw network data, feature extraction, and training of machine learning models to be able to classify cyber-attacks. Autoencoder based Anomaly Detection is very successful in identifying zero-day attacks as it identifies aberrations from the baseline open network activity. It also offers improvement in collective learning detection performance, like decision trees and random forests.

These experimental results clearly proved that the proposed hybrid of AI and IDS outperforms that with the conventional signature-based approach. The results of the deep learning models show great success in the identification and classification of cyber threats, reaching a detection accuracy of 98.9% and an F1-score of 97.7% Furthermore, the system below is also successful in reducing false positive, resulting in enhance reliability (Kumari et al., 2023). The NSL-KDD is commonly used as a benchmark data set for researching intrusion detection, which we use in this study for evaluation. This data is made up of (labeled) network traffic from normal and malicious activity, allowing the authors to train/test their effective model. In general, the results highlight the potential of deploying AI-driven tools to safeguard nuclear facilities against the next generation of cyber threats.

(Ejigu et al., 2023)[36] stated that in order to operate nuclear power plants safely and efficiently, their control systems must be monitored and tuned on a regular basis. The sheer volumes of data generated by these installations are overwhelming for traditional monitoring systems, which vastly increases response times to detect anomalies or potential issues. Advanced technology of nuclear reactors has also increased the need for predictive maintenance solutions (Ejigu et al., 2024). Traditional approaches cannot analyze data in real-time to help prevent such important mistakes. To address these limitations, the authors proposed implementing artificial intelligence by utilizing SCADA systems big data analysis in nuclear power plants control systems. One proposed solution involves using AI to sift through mountains of operational data to improve plant performance. With the help of predictive analytics, AI is capable of identifying system anomalies early on, enabling preventive maintenance while minimizing the chances of unexpected breakdowns (Ejigu et al., 2024). The research employs multiple AI techniques including the deep neural network (DNN) (also called deep learning networks), Bayesian networks (BN) and reinforcement learning (RL). Virtual representations that reflect in real time the operation of nuclear reactors, so-called digital twins, are also used to validate AI-improving optimization methods. And the methodology involves gathering historical sensor data from reactors, pre-processing the data through feature engineering, and training AI models to predict how the reactors will behave under different conditions. Results confirmed that facility AI-based predictive maintenance models show much higher fault detection accuracy. Theye used deep learning models with accuracy of 92% to predict reactor anomalies, while reinforcement learning models optimize reactor efficiency, achieving 15% better performance compared to traditional methods. AI-based automation also frees the burden of manual labor from the plant operators and increases the safety and efficiency of plants further (Ejigu et al., 2024). Datasets used in this study come from real-world nuclear power plant sensors, including pressurized water reactors (PWRs). These datasets are composed of time-series data that document reactor performance, temperature swings, and deviations from normal operations. This study illustrates the successful application of AI analytics in efforts to enhance the efficiency and reliability of nuclear plants.

(Wang et al., 2023)[33] elaborated that user behavior analysis is crucial in many domains, for example, smart libraries and personalized techniques. Traditional be-

havior analysis methods, however, are purely based on mathematical statistics and rule-based heuristics, and do not effectively capture complex behavioral patterns (Wang et al., 2023). These outdated approaches do not scale, they simply are not transferrable or a lot of data is needed from the footprint that most digital engagements generate these days. The innovative, adaptive nature of social media and new behavior patterns that propagate from user to user further complicate the detection of activity as fraudulent behavior or simply unusual behavior. To overcome these shortcomings, the authors suggest a big data analytical model of user behavior using machine learning-based frameworks for AI-based user behavior analysis. They have used clustering and settlement algorithms to group users based on digital engagements. The AI-based model can also learn from the changing patterns of user behavior and do some fine-tuning to its predictions (Wang et al., 2023). Herein, they use k-means clustering, support vector machines (SVM) and Bayesian classifiers through user interaction logs as the input. The raw data needs to be cleaned, features need to be extracted and then separate machine learning models built to classify user behavior. Classifiers for fraud detection are then trained using supervised learning techniques, and clustering techniques segment users based on similarities of behavior. Experimental Results on AI-based User Behavior Analysis These findings can be summarized at a very high level as shown in the below figure. In summary, the byte-level user behavior analysis which is based on the above AI algorithms mentioned has outperformed tremendously the traditional manual machine-based user behavior analysis commercial systems available as products in the marketplace. Segmentation using the clustering model reaches 87% accuracy, and the fraud detection system reduces false positive rates by 23% from traditional rule-based detection models. The system further personalizes recommendations, increasing user engagement and satisfaction (Wang et al., 2023). The study is based on real-world datasets obtained from digital library access records, online interactions, and transaction logs. These datasets contain extensive behavioral information, which allows the AI to make very accurate predictions and to easily detect abnormal behavior.

(Sufi & Khalil, 2024)[41] pointed out that visibly destructive events such as wildfires, earthquakes, floods or hurricanes are natural disasters that affect communities globally is a major issue. A good example of this is a social media platform is Twitter, which became one of the most important tools during disasters to receive real-time information. Conversely, conventional disaster monitoring methods include only 1% of the total tweets which are geotagged, thus conventional methods are not able to provide holistic information about some disasters (Sufi & Khalil, 2024). Most of these existing models are not able to parse tweets concerning disasters having other languages and thus are relatively less contributive in the global disaster-response being produced in all sorts of languages. To overcome these problems, the authors propose an AI-based system that combines NER, sentiment analysis, and a deep learning-based anomaly detection algorithm to mine disaster-related information from social media postings (Sufi and Khalil, 2024). The author broke the limit by predicting the geographic coordinates of tweets in no need of geotagged data in October 2023 and serving a comprehensive and holistic disaster intelligence system that totally models the evolution of disaster spread, at which time Dash became a

revolution against traditional systems. This is where AI-powered location intelligence can come in, one that is designed to analyze thousands of tweets per second, pinpointing the geographic areas impacted by disasters. This research comprises a three-step methodology: data preprocessing, feature extraction, and disaster classification. It implements Natural Language Processing (NLP) methods such as tokenization and part-of-speech tagging and allows cleaning of unstructured social media data and structuring it so that it can be analyzed. This includes anomaly detection models based on convolutional neural network (CNN) which are used to classify disaster-related tweets, and the Getis-Ord Gi* algorithm to guide accumulation of disasters hotspots. In addition to this, sentiment analysis validates the mood and feelings related to the disaster tweets, which assists in filtering out the disaster type classification. We validated the system on 67,528 tweets pulled from September 28-Oct 6, 2021, across 39 languages. The results of the experiment indicate that the accuracy and F1-score for disaster detection (finds both disaster events and mentions are relevant) are 97% and 0.90, respectively. As stated in Sufi and Khalil (2024), "the AI model outperformed the traditionally used geotag-based disaster monitoring systems by a wide margin, making it potentially a powerful source of real-time, multicolored-language, global-scale disaster intelligence". This study collects real-time social media datasets primarily from Twitter via API-based extraction methods. These tweets contains both structured and unstructured disaster information, on the basis of which the AI model capable to do the location intelligence and disaster severity. According to the authors, this demonstrates the potential of artificial intelligence disaster monitoring systems to improve emergency response and disaster preparedness.

(Nasry et al., 2023)[31] presented that the deployment of video surveillance on public and private properties accelerates, the need for automated person tracking to supply security and prevent vandalism is increasing. Traditional video surveillance systems depend on human observation—and human observation is a time-intensive resource that is not effective for the simultaneous monitoring of multiple cameras (Nasry et al., 2023). Challenging mechanical tracking approaches are prone to occlusion, multi-view-point camera arrangement, and very vibrant weather conditions, which degrade tracking performance. In order to increase the efficiency of surveillance vehicles, the authors develop an artificial intelligent based people tracking system that use deep learning techniques including (but not limited to) convolutional neural networks (CNNs) (for object detecting) and recurrent neural networks (RNNs) (for object tracking) (Nasry et al., 2023). This solution combines multi-camera tracking, enabling seamless re-identification of people across multiple surveillance feeds. This study methodology includes an entire process from the detection of objects to the extraction of features from them, including tracking and re-identification of individuals. The authors are using a combination of Faster RCNN for real-time identification of individuals and Yolov (You Only Look Once) object detection frameworks for GeoRT-Tracking. To make sure that two individuals are the same object seen in different perspectives from different cameras, re-identification employs Kalman filters and Siamese networks. The study also uses many predictive models that track a pedestrian's motion based on previously observed movement, enabling even more accurate tracking. Results from the experiments confirm that the AI-based track-

ing methods achieve over 90% accurate tracking, better than the tracking result of heuristic tracking methods. It employs the GOTURN tracking algorithm provides a 90–94% efficiency (minimizes false positives) and enhances real-time tracking capabilities (Nasry et al., 2023). In addition, the model performs effectively on difficulties such as occlusion and the various illumination circumstances often involve surveillance techniques. The system was validated against benchmark datasets such as the MOT17 (Multiple Object Tracking 2017) dataset, as well as the CHIL pedestrian tracking dataset. These datasets contain video surveillance sequences that have been captured in natural scenes, where the pedestrian flow has been marked. Hence, they provide a rich source for our model training and evaluation. The outcome show that systems utilized for object tracking based on deep learning greatly improve surveillance capabilities, making it a scalable solution for real-time security monitoring.

(Lee et al., 2017)[6] introduced that Conventional NPP safety systems are based on common-sense, shared-control techniques that use human operators with automation support. This complicates the goal of true autonomy because decisions made by a human can incur latency and variability in the process. Current automation systems cannot self-assess, diagnose an issue, or optimize performance on the fly. Finally, nuclear safety systems operate in nonlinear environments, and traditional control mechanisms are not suitable for dynamic environments in which unexpected incidents like an emergency state require unfolding actions. They propose a long short-term memory (LSTM) automaton to autonomously monitor and control nuclear power plant (NPP) safety systems (Lee et al., 2017). This method seeks to enable NPPs to move forward from shared control to complete autonomy by using the pattern recognition and predictive decision-making capabilities of AI with time-series data. The main methodology is Function-Based Hierarchical Modeling: which Decomposes safety systems into structured layers for better modeling and AI integration. LSTM Neural Networks: Used to train on time-series data related to processing safety and forecast ideal control decisions. Min-Max Normalization: Keep consistency of input data by scaling values to [0,1]. CNS Training Dataset-Scenarios include normal operations, loss-of-coolant accidents (LOCA), safety injection failures, and steam generator tube ruptures. Evaluation Metrics: RMSE (Root Mean Squared Error) was used to evaluate the model performance and compare it to human and automated control systems (Lee et al., 2017). Experiment results showed that the LSTM-based control system outperformed humans in directional and lane control in the following ways: Improved Core Safety: The algorithm kept coolant levels higher, preventing overheating. Quicker Decision: LSTM was able to respond quicker than human operators in emergencies. Enhanced Reactor Cooling: The AI system cooled the reactor core at a greater rate than standard control mechanisms. Increased Accuracy: Demonstrated lower RMSE, showcasing better control over safety parameters (Lee et al., 2017). To train the ML models, we used 168 operational variables and 94 control signals that were derived from CNS simulations of actual nuclear power plant operations. With more than 206 accident scenarios included in the dataset, the models were kept to robust training and evaluation measures. This trained LSTM generalizes beyond the training data and can effectively control plant safety functions, in a previously untrained LOCA (in-loop2 hot-leg) case (Lee et al., 2017).

(Sufi et al., 2022)[23] put forth that the anti-vaccination (Anti-Vax) social movement was the leading threat to public health endeavors, providing misleading information and influencing social disagreement, decreasing the vaccination rates. However, observing (and making sense of) these social movements is critical for strategic decision-making as current methods are not efficient and resource-demanding (Sufi et al., 2022). The research proposes a potential remedy using AI via sentiment analysis and Named Entity Recognition (NER) to Identify Anti-Vax and Pro-Vax communities set across time, space, and meaning, by tracking their social media activity. Information that policymakers and researchers could respond to. It is a sophisticated AI analysis of publicly available data from Twitter, 55 languages, from June 15 - Dec 31, 2021 (Sufi et al., 2022). Data Retrieval and Preprocessing: To do this, they extract tweets from each week via the Twitter API and translate them via Microsoft Translator into English. The Microsoft Azure Cognitive Services API was used to perform sentiment analysis and NER. SQL queries were customized to find Anti-Vax and Pro-Vax posts, given fields (e.g., certain words like "Hoax," "Ignorant," and "Confused"). NLP techniques for detecting negativity and subjectivity sentiment analysis. NER: Extracting entities such as locations, organizations, and major topics from those unstructured tweets. They used to do it through Microsoft Power BI and create dashboards available for several platforms (iOS, Android, Windows) (Sufi et al., 2022). Sentiment Analysis Results: Anti-Vax-related tweets are 72% more negative than control (COVID-19-related) tweets. Pro-Vax tweets were 65% more negative as well (Sufi et al., 2022).

(Katkar et al., 2022)[19] presented the idea that the threat of crime and intimidation has the potential to undermine social stability in modern cities, which has become an increasingly pressing challenge. CCTV cameras and other surveillance systems are widely used in public places to ensure safety. However, monitoring these surveillance cameras is a labor-intensive, resource-heavy, and often ineffective process, this is especially true in dense urban areas. And furthermore, scanning real-time CCTV footage to catch criminals is cumbersome and takes up too much time. Due to the non-automated manner of finding anomalies or unusual behavior, these systems have constraints in definition, which may likely NOT be able to identify criminals and crimes and ensure public safety (Katkar et al., 2022). They proposed an AI-based, deep learning-based system for the detection and tracking of crimes using CCTV networks. It includes the following key elements of the solution. Step 2: Conversion of video to frame: The surveillance videos are divided into frames, and these frames are resized and pre-processed further for analysis. Feature Extraction: Deep learning models like ResNet-50 and CNNs generate high-level feature maps. Behavioral Analysis: A temporal and spatial normalcy model is established to distinguish between normal and abnormal activities. Abnormal Behavior Detection: Frames consisting of aberrant behavior are marked, and real-time alerts are delivered to the authorities for timely action. Criminal Identification and Tracking, After detecting pedestrians behaving abnormally, the system uses face recognition to identify them and follows their movement across the horizon of the metal camera. This method combines CNN with LSTM networks, called a hybrid

approach so that the spatial data can be captured in CNN and the temporal data can be captured in LSTM networks which makes the solution robust for human activity detection and recognition (Katkar et al., 2022). Deep Learning Models: The system employs ResNet-50, CNNs, and LSTMs to analyze visual data and identify anomalies. Feature Extraction Techniques: High-level feature maps are extracted for better classification. Videos are split into minute pieces that allow precise detection of abnormal motion. Motion Detection Algorithms: They are used to monitor movement patterns and detect any suspicious activities. LSTM Layer: The Softmax Classifier for categorizing identified behaviors probabilistically Evaluation Metrics The proposed system is evaluated on the following metrics: Metrics for detecting abnormal activities: Precision, Recall and F1-Score ROC-AUC Curves: For verifying classification models. Real-Time Detection Rate: The rate and effectiveness of the detection of an anomaly in a live environment. True Positive and Negative Rates: To provide accurate results with the least possible mistakes possible (Katkar et al., 2022). The experimental results proved that the system can achieve the following: Real-time Detection of Abnormal Behaviors and Crimes Automate the process of monitoring to lessen manual intervention. There are been retained for minute tricky situations like crowded environments and quick action. Improve public well-being through immediate action to identify dangers The figures demonstrate the notable frames and identify atypical action with the system's practical functionalities in real-world situational awareness. In case of providing any alarm, the system sends an email to control rooms for immediate corrective action (Katkar et al., 2022). This system was trained and validated on video datasets that mimic real-world response situations of surveillance. These datasets include: A diverse set of human behaviors for effective training. For testing the model in crowded environments and how it performs in them (Katkar et al., 2022).

(Teh et al, 2013)[3] conducted a comprehensive survey of research on keystroke dynamics biometrics over nearly three decades and found that while keystroke authentication is not robust enough to replace other biometric verification method, because of the potential to implement it stealthily, its cheap cost of implementation, potential for user verification and continuous assessment makes it a great addition into a security system to supplement surveillance. This paper looks into the use of simple statistical analysis of early work and explores the implementation of machine learning algorithms for keystroke dynamics analysis.

(Yang & Qin, 2021) [17] found that keystroke dynamics can be used to analyse user emotions and mood. They have also found that keystroke analysis can identify a user's identity. While keystroke dynamics is a soft biometric measure, varying for even the same user based on their mood, typing proficiency or even injuries, it is still a good supplimentary measure in a security system, especially when implemented for continuous analysis for user identity verification. Hence, based on current amount of research done on this field, we can implement keystroke dynamics as a supplimentary system for user verification in a surveillance system.

2.3 Summary of Key Findings

The key research findings that serve as the basis for the proposed security enhancements for workplace security are discussed below based on the reviewed research papers:

Surveillance and video tracking powered by AI

- Real-time tracking can be done using hybrid CNN-LSTM models for human movement and object detection, also for abnormal behavior.
- In fact, security-sensitive environments with YOLOv5 have shown up to 95.43% accuracy in predicting suspicious behavior, outperforming human monitors.

Threat Intelligence from Social Media Monitoring

- Social media activity is analyzed by AI models in search of real-time intelligence regarding possible security threats.
- Sentiment analysis augmented with NER can forecast hostile intent, misinformation campaigns, and insider threats.

Employee Behaviour Anomaly Detection

- With trained sentiment analysis model, we can analyze employee body language and movement to screen for emotional distress or malicious intent.
- Combined with other proposed detection measures, employee sentiment analysis can fortify early detection and allow superiors to check up on employees who raise low threat flags which might help in preventing insider threats.

These results recommend that the inclusion of a mixture of AI-primarily based monitoring and social media analysis and biometrics will play a vital position within the safety of sensitive workplaces. Combining multimodal analysis might lead to possible prevention of crimes. Further work is necessary to integrate such AI models into a holistic centralized security model and better coordination of security.

Chapter 3

Requirements, Impacts and Constraints

3.1 Final Specifications and Requirements

Technical Requirements:
A computer with an NVIDIA RTX 4060Ti GPU and 32GB RAM was required to train the weapon detection model. A computer with an NVIDIA RTX 5070Ti GPU and 32GB RAM was used to train the keystroke analysis model and the emotion detection model.

Software Requirements:
Python 3.10, Google Colaboratory, CUDA, Pytorch, Tensorflow, Huggingface Transformers, Numpy, SciKit-learn, SciKit-Fuzzy, CV2, Pickle, MatPlotLib, Seaborn, TQDM, Auto Archiver, YOLO11 [38].

Data Requirements:
A publicly available dataset created by a team from Bahria University, Pakistan was used to train the weapon and suspicious personnel detection model. The dataset has 1922 images of suspicious individuals, 3279 images of regular individuals and 2862 images of weapons. BoLD (Body Language Dataset) which has 9,876 Clips, 13,239 Instances and 50-150 Frames Per Clip was needed for the facial expression model.

3.2 Societal Impact

This work can help reduce the stress and workload on security personnel at a given workplace by augmenting their ability to monitor the workplace. Moreover, the workplace safety would increase due to increase in the ability to detect threats and suspicious activity. This system would also work as a deterrence against crime in the workplace, increasing safety further.

3.3 Environmental Impact

Due to the multi-modal design of this proposed system, a substantial amount of computing is required to train the system. The increased computing power results in an increase in energy used to train the system, which will lead to a rise in the overall carbon footprint associated with the system. After deployment, the inference portion of the system is much lighter weight than the training process and is therefore relatively efficient at run time.

3.4 Ethical Issues

The intensive social media monitoring as part of the surveillance model raises significant privacy concerns for individuals. To tackle this problem and to uphold legal and ethical standards, we will take explicit consent from the personnel. By obtaining prior and informed consent, the system will respect personnel privacy while maintaining the integrity of the security system. This approach ensures that employees are fully aware of the monitoring practices and their scope, which will help in mitigating privacy-related issues. Moreover, the system would not store unneccesary data or share data of the employees to maintain their privacy.

3.5 Project Management Plan

We have 3 phases in our project:

- **Collecting data and analyzing viability (Month 1-3):** We collected data from various reputable sources for the training of our models. We extensively studied relevant literature to assess the viability of our system and improved the design accordingly.
- **Preprocessing data and developing initial design (Month 4-6):** We processed the data for each model and designed the initial system.
- **Testing and improving the design (Month 6-9):** We tested the design and made improvements on it based on expert opinions. We consulted police, lawyers and forensic psychologists to understand how the different aspects of the system should be weighed for optimal results.

  Budget: A computer with minimum hardware requirements to train and test- $1200
  Google Colab credit - $20

3.6 Risk Management

The main risk of the system would be data breaches. To mitigate such a risk, airgapping the system should be done to prevent access to the data storage through unsecured networks. If the data is stored in the cloud, the privilege to access data should only be granted to security personnel who are in charge of maintaining the system. The privilege should be granted to as little people as possible to reduce the chance of misuse of the data by the personnel.

3.7 Economic Analysis

Cost analysis: The project will require $5000 to set up hardware, software and cloud storage.
Benefit Estimation: The system, when set up with proper CCTV coverage can save an organization up to $8000 by cutting down labor for manual monitoring.

Chapter 4

Proposed Methodology

4.1 Design Process or Methodology Overview

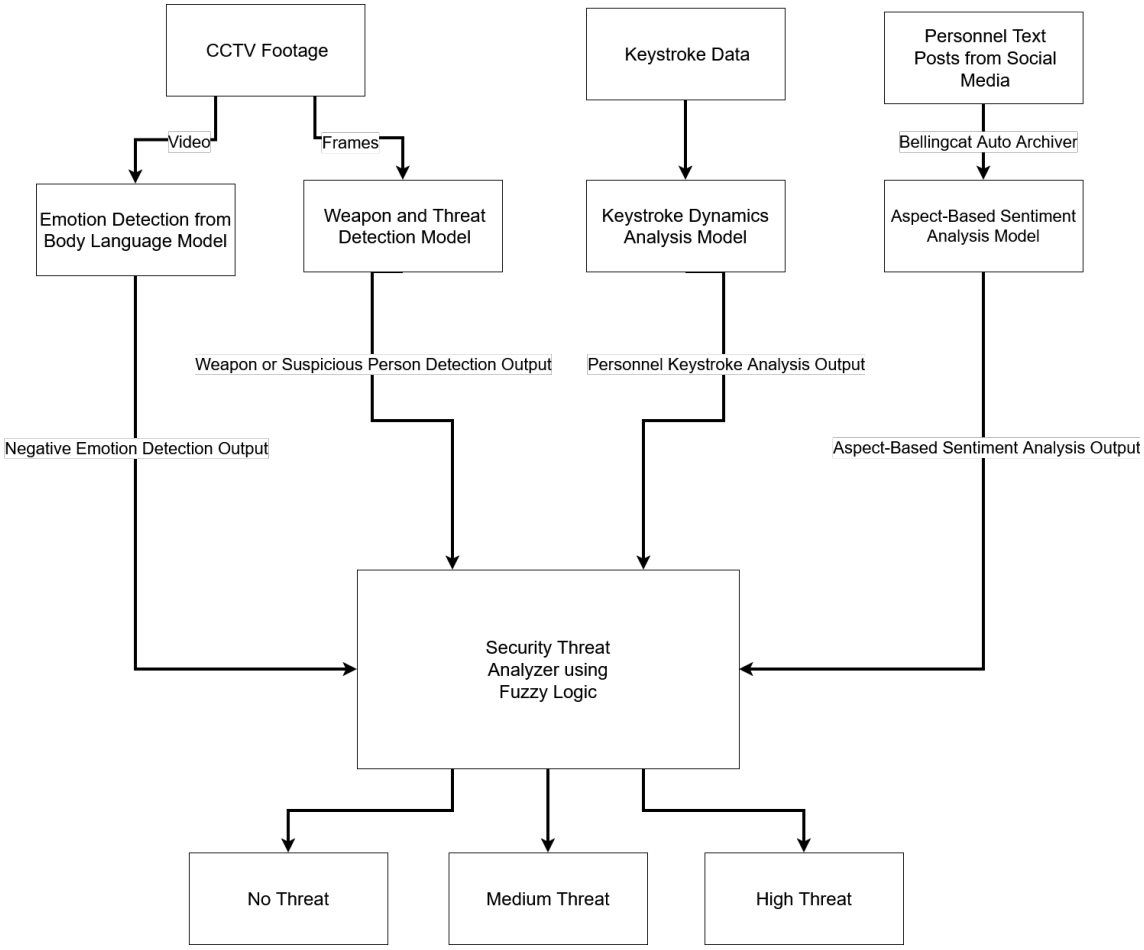

Figure 4.1: Overall System Flowchart

The security system is multifaceted and relies on multiple analysis for security scoring. The system will take inputs from four different analysis:

- Weapon detection
- Emotion detection from body language

- Employee key-stroke analysis
- Sentiment analysis of employee social media posts

These inputs are then combined in a scoring system to calculate a safety score for each employee.

One part of the security system is the weapon detection system. For the final score, the detection of weapons is a substantial indicator for the score to report suspicion. For this system, we trained a model that can distinguish among regular individuals, weapons and individuals with suspicious attributes (ski masks, carrying weapons, aggressive pose etc).

Another part of the security system is the emotion detection model, which detects human emotions through body language. The proposed system uses both body-focused visual aspects and context-based visual cues. The TSN (Temporal Segment Network) is used to analyze sequential video data in this approach. It consists of a ResNet backbone and two branches. The first is a RGB branch (Residual Network)-based body branch that captures individuals' posture and movement characteristics. The second is a context branch that analyzes visual cues that surround the individual being analyzed. After the two representations have been analyzed, they are combined using a temporal-consensus operation to create a unified behavioral embedding. This behavioral embedding is then enriched using optional semantic GloVe emotion vectors. Therefore, the model allows the detection of human emotions such as engagement, anger and peace.

Next, for the keystroke analysis part of the system, the model uses a deep learning method with LSTM (Long Short-Term Memory) autoencoders for behavioral biometric authentication. The approach focuses on detecting anomalies. During training, the system learns typical typing patterns, and any deviations from these patterns may suggest unauthorized access attempts.

Lastly, for sentiment analysis of social media posts, the aspect-based sentiment analysis model deberta-v3-base-absa will be deployed. The social media text will be scraped by using Auto Archiver, an open source tool written in Python by Bellingcat whose purpose is to archive content on the web. The sentiment will be calculated based on the workplace aspect in the text.

4.2 Preliminary Design

Threat Detection Model
The threat and weapon detection model was trained using YOLO11 [38] object detection algorithm. YOLO (You Only Look Once) [4] an object detection algorithm that does a single pass over the input image to predict class probabilities and bounding boxes. The YOLO algorithm consists of 3 parts, which are the backbone, the neck and the head. The backbone extracts features from the input image, the neck combines features and the head uses the features to output the bounding boxes and class labels for the image. For input, this model takes in video footage, e.g. CCTV

footage. The model outputs bounding boxes, labels and confidence scores for the given frame, with the labels being "Person", "Criminal" (meaning Suspicious Person) and "Weapon".

Body Language Analysis Model
The emotion detection model is designed using the BoLD (Body Language Dataset). Three architectures were planned for this model, those being: pure pose-based models, vision-only CNN architectures, and transformer-based temporal models. However, each of these designs had their own shortcomings. Iterative development ultimately led to the adoption of a design using TSN (Temporal Segment Network). This model includes: A ResNet backbone for extracting body-focused features, a parallel-context branch to facilitate environmental understanding, and temporal consensus for aggregating segment-level representations. A semantic alignment component utilizing GloVe embeddings is also included. For input, this model takes in video footage. The outputs are separate probabilities for each emotion and gives a list of the top 5 emotions detected within the video.

Keystroke Analysis
The keystroke dynamics authentication model is a biometric system that analyzes typing patterns to authenticate computer user identity. The individual model per user relies on an autoencoder model, and each user has one that extends a sequence of keystroke events (10 timesteps and 5 features) through a sequence of two LSTM layers that reduces the data to 64 dimensions using ReLU activation. This compressed representation is sent to a 32-unit bottleneck layer that encodes the necessary typing properties into a latent space, and a decoder, composed of two LSTM layers, is used to decode a sequence of original keystrokes to identify anomalies as potential indicators of unauthorized access. It also takes user ID embeddings as inputs to link the data to each user. The output is the percentage chance of the keystroke data being anomalous.

Sentiment Analysis from Social Media
For sentiment analysis, deberta-v3-base-absa is deployed. This is a state-of-the-art model specializing in aspect based sentiment analysis. The input text gathered via web scraping employee social media profiles using Auto-archiver from Bellingcat. Based on this input, aspect-based sentiment of the employee toward their workplace is given as output.

Fusion with Fuzzy Logic
To combine the outputs of our trained models we used fuzzy logic engine. We chose to use fuzzy engine because in real life, safety guidlines are not based on discrete logic or calculations. Rather it is based on the experience and expertise of the professionals in the field. Fuzzy logic mimics human reasoning and is more intutive to the end user than binary logic. Fuzzy logic also determines the degree of a threat more effectively compared to binary logic and reduces the chances of unnecessary alarm.

4.3 Dataset Description

Threat Detection Model
For the weapon and threat detection model, we collected data from a publicly available dataset made by a team from Bahria university on Roboflow [28]. The dataset included 4152 images including annotations from three classes: criminal, person and weapon. 3323 images were put in the train set, 416 in the validation set and 413 in the test set.

Body Language Analysis Model
The dataset used for emotion detection is BoLD (Body Language Dataset) [9] made by researchers from Pensylvania State University working on ARBEE (Automated Recognition of Bodily Expression of Emotion). It is composed of more than 9800 annotated video clips together with more than 13,000 human instances, which are labeled with 26 discrete emotions and 3 continuous emotional states (valence, arousal, dominance). The videos in the dataset were sourced from publicly available movie scenes and manually annotated using crowdsourcing.
To analyze the models performance on crime scenes, we collected 104 crime cctv clips from various sources and created a test set for inference.

Keystroke Analysis
The Keystroke Analysis model used the Keystroke Dynamics Dataset [40]. The dataset contains 55,925 keystroke records across 108 individual users. The records include the time taken for 5 distinct key press related events to occur, which includes dwell time (the length of time the key is pressed down) and flight time (the time difference between a key release and the next key press).

4.3.1 Data Preprocessing

Threat Detection Model
To make our threat detection model more robust and tolerant to low image quality or poor lighting conditions, we augmented the training set, doubling the size from 3323 images to 6646 images.

Augmentation Type	Value
Auto-Orient	Applied
Resize	Stretch to 640x640
Grayscale	Applied
Flip	Horizontal and Vertical
90° Rotate	Clockwise and Counter-Clockwise
Crop	0% to 20% Zoom
Rotation	-15° to +15°
Blur	10 pixels
Noise	5.01% of pixels

Table 4.1: Augmentations applied for Threat Detection Dataset

Body Language Analysis Model
Images were resized to 224x224 pixels. Greyscaling was also applied to the images.

Keystroke Analysis
The three csv of the dataset were combined and a source label was added to identify which csv file the keystroke data originated from. Next, rows containing missing data were removed. All the data features were then standardized.

4.4 Implementation of Selected Design

Threat Detection Model

After cleaning and preprocessing the data, we trained YOLO11 object detection model on our dataset until the model converged after 300 epochs.

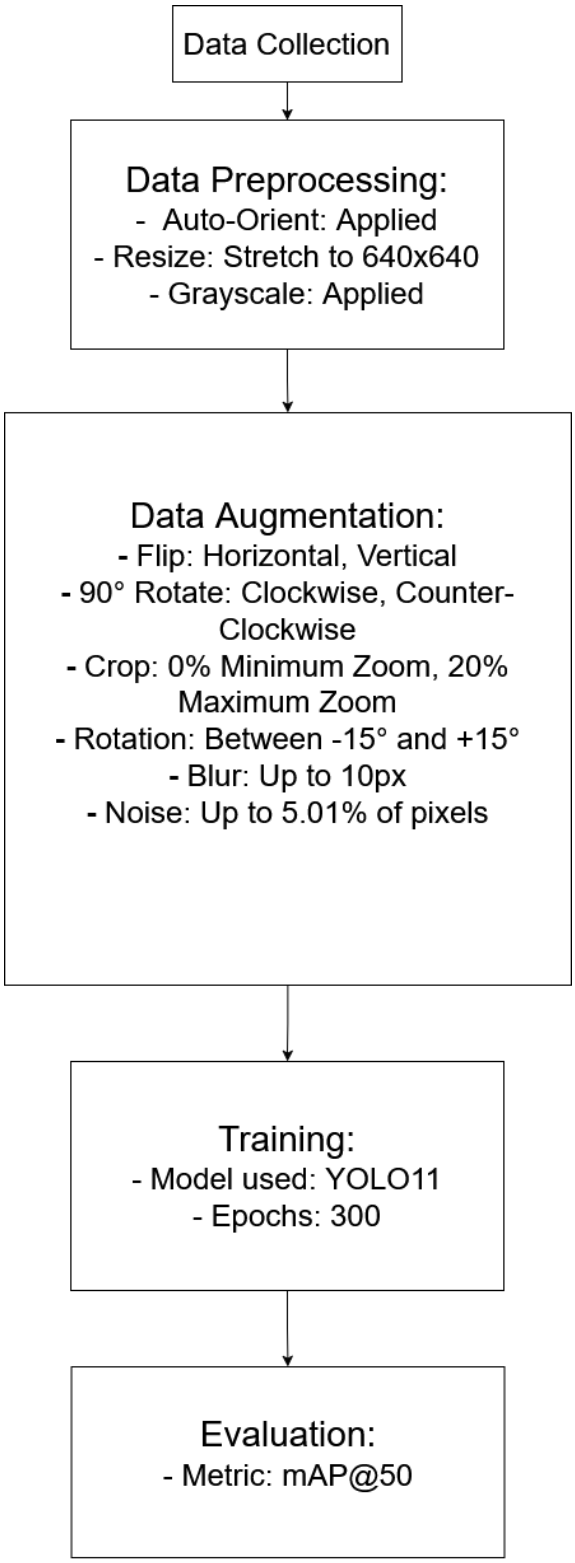

Figure 4.2: Work flow for weapon and threat detection

Body Language Analysis Model

For emotion detection from body language, the implementation of the proposed Hybrid TSN-based emotion-detection model included development of both Temporal Segment Networks (TSN) and dual visual streams, which are the Body Branch ResNet and Context Branch ResNet.These models include temporal-consensus func-

tions, data-loading pipelines, augmentation modules, and multi-head output models to produce categorical emotion labels and regression scores for Valence, Arousal, and Dominance. The model was trained using the preprocessed BoLD dataset using weighted binary cross-entropy to perform multi-label emotion classification and mean squared error (MSE) to create continuous prediction values. Our Training epochs were 60, warm-up epochs were 10. We used a batch size of 20 and SGD optimizer. We used cosine scheduling to set the learning rate.

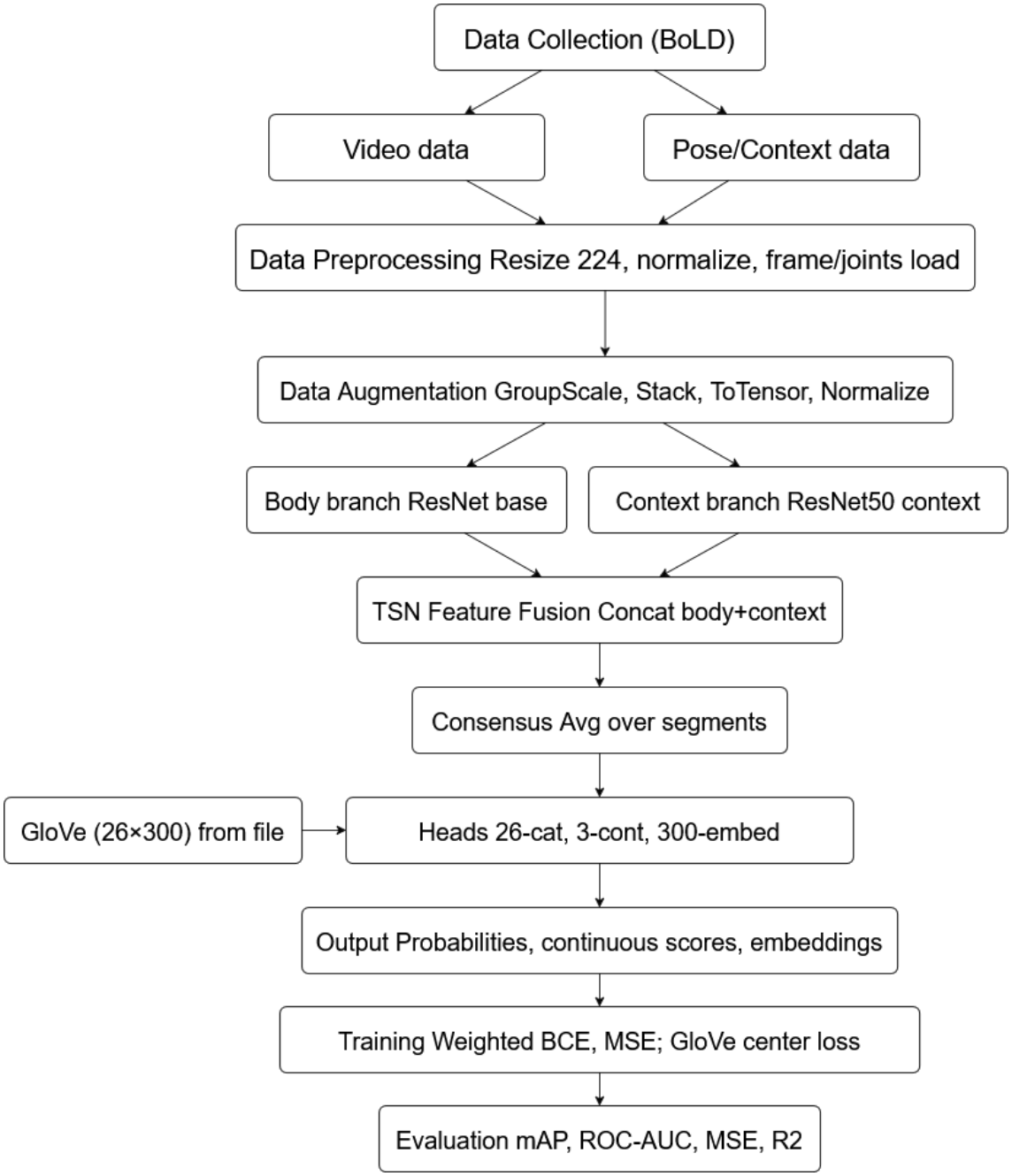

Figure 4.3: Work flow for body language analysis

Keystroke Analysis

The keystroke dynamics system uses deep learning based on autoencoders with LSTM. It is based on learning the natural typing pattern of the user and then

trying to detect anomalies in the typing pattern of that user. Data loading stage initiates with loading of three different datasets, which are: email, fullname, and phone inputs. The datasets are consolidated into one unified dataset source marking. The system measures dwell time, interval time, latency, flight time and up to up time which are the five important featured keystroke metrics. The data of keystroke is grouped in sequences to train the model with a sliding window. Training of individual models is based on 80/20 train, test split and early stopping is done to avoid overfitting. The underlying algorithm in anomaly detection is the mean squared error between original and reconstructed sequences, with thresholds the mean and three standard deviations of training errors. Any sequences beyond this benchmark are considered to be anomalous. Reconstruction error is used as performance measure, binary threshold comparisons and visualization. The LSTM Autoencoder was trained using 60 epochs, a batch size of 64, the Adam optimizer for learning rate optimization, and Mean Squared Error loss. The model includes two encoder LSTM layers (128 and 64 units, ReLU activation, with dropout 0.2), a 32-dimensional bottleneck, and symmetric decoder layers. A user embedding of size 10 and a sequence length of 10 were used. Thresholds were computed from the 95th percentile of training reconstruction errors with a 10% buffer.

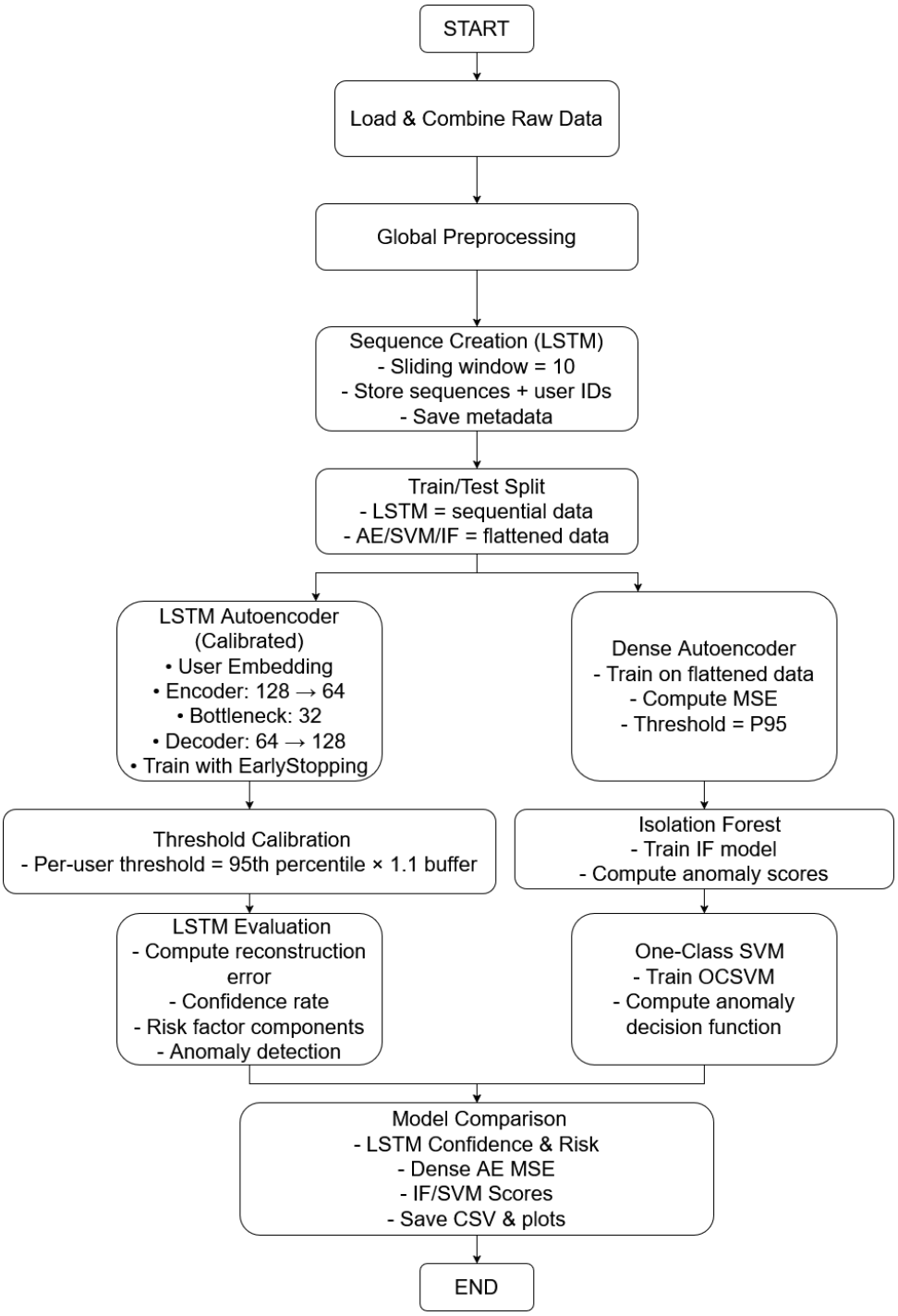

Figure 4.4: Work flow for keystroke analysis

Sentiment Analysis from Social Media

For the sentiment analysis part, text is scraped from the user's social media posts using Bellingcat Auto Archiver. Then, the text is stored locally and the DeBERTa-absa model (Yang et al, 2023) [34] (Yang et al, 2021) [16] is used to measure the aspect-based sentiment of the user's texts. The aspect measured is towards their workplace.

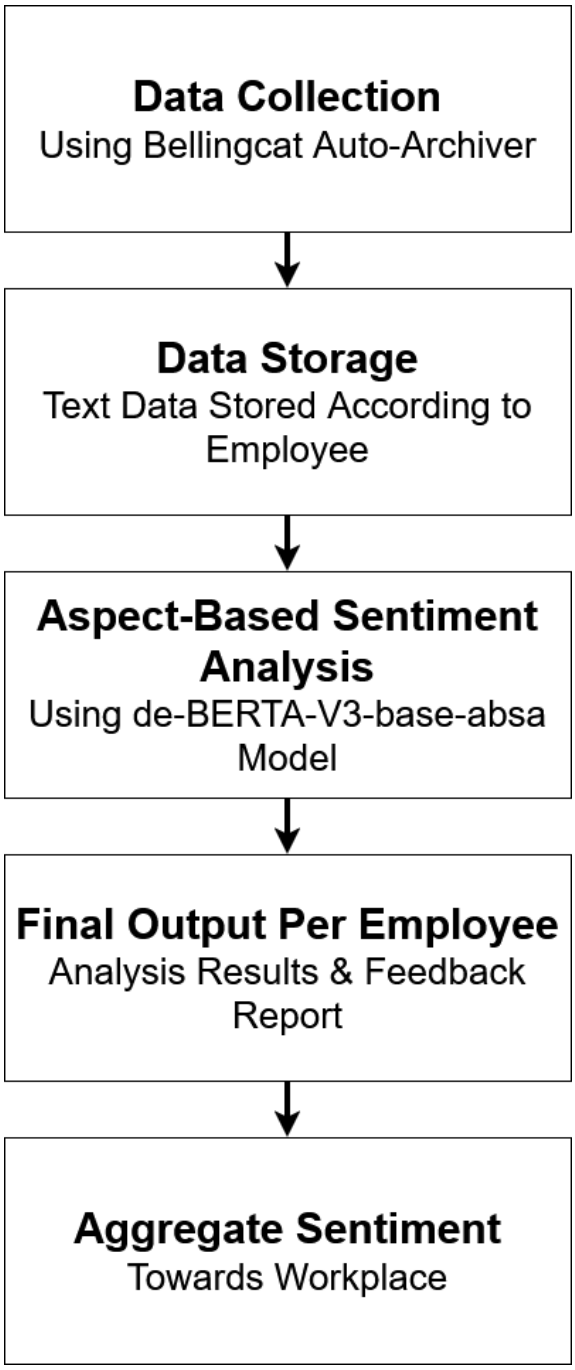

Figure 4.5: Work flow for sentiment analysis

Fusion with Fuzzy Logic

To fuse our outputs together, we created a fuzzy reasoning system. We first take the outputs from our four systems and put it into the fuzzifier. We used Mamdani inference system using triangular membership function for the fuzzification of the inputs.

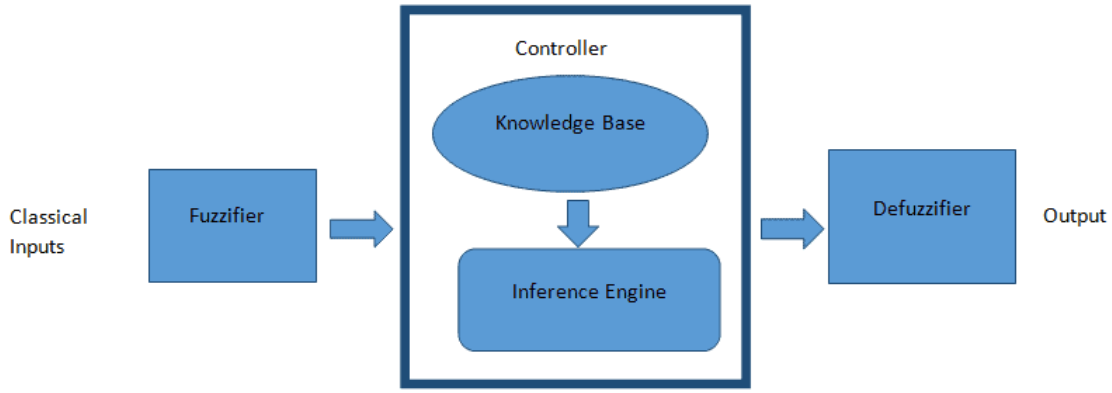

Figure 4.6: Mamdani Fuzzy Control System

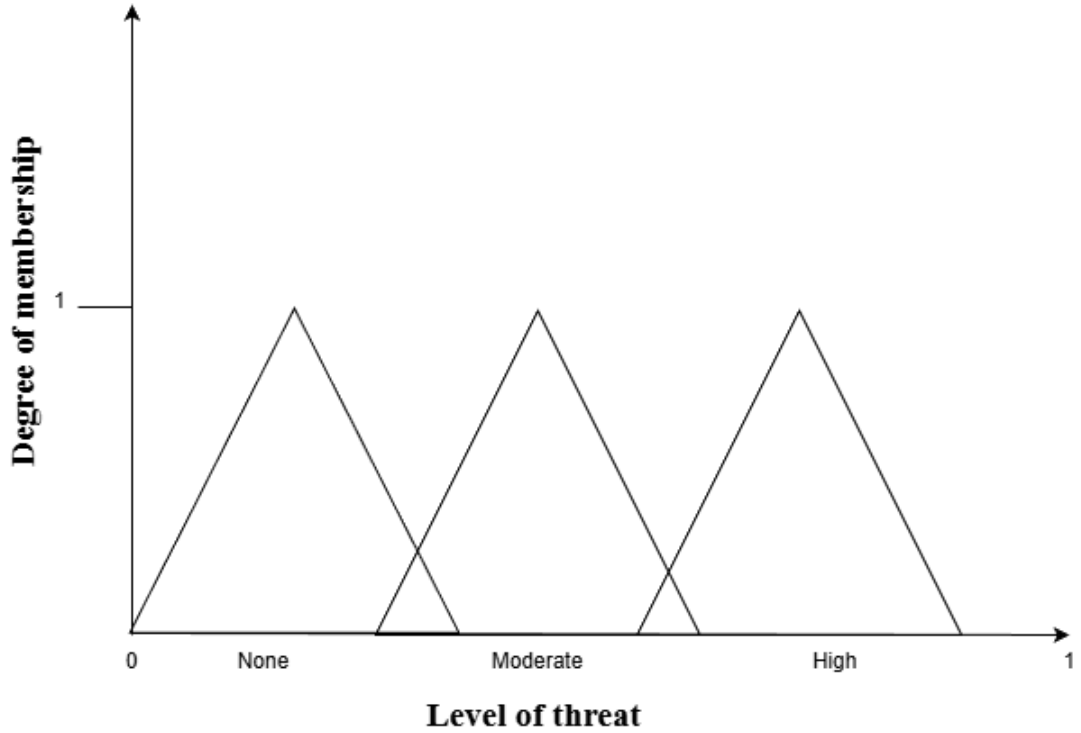

Figure 4.7: Triangular Membership Function

We set some rules for the knowledge base of our system. The inference engine followes the rules to evaluate the fuzzified inputs to determine the threat level membership of the input combination.

Rules for the knowledge base:

1. If Weapon/Suspicious person detection confidence is moderate (0.4-0.5), then the threat level is moderate.

2. If Weapon/Suspicious person detection confidence is moderate (0.4-0.5) and Body Language negativitiy detection is high, then the threat level is high.

3. If Sentiment from social media is negative and Body language negativity is high, then the threat level is high.

4. If Keystroke anomaly is high, then the threat level is moderate.

5. If Body Language negativity is high, then the threat level is moderate.

6. If Weapon/Suspicious person detection confidence is high (>0.5), then the threat level is high.
The defuzzifier then uses centroid function to create a crisp output that is mapped to the threat level.

Threat Level Interpretation

Threat Level	**Interpretation**
Low	Regular level of monitoring
Moderate	The systems firing alarm should be checked
High	Tangible threat has been detected, immediate actions necessary

Chapter 5

Result Analysis

5.1 Performance Evaluation

The comprehensive evaluation of the proposed advanced personnel security system demonstrates robust multi-modal performance suitable for workplace deployment.

In the computer vision stream for physical threat detection, the system implements YOLO-based real-time weapon detection to identify firearms, knives, and other contraband. The model was trained to detect three classes: suspicious individuals, weapons and regular individual. The model performs very well for detecting suspicious individuals (denoted as criminal in the figures) with the highest precision-recall across the classes. The model is excellent at discerning between suspicious and normal individuals and detecting guns. Across all classes the model achieved an accuracy level of 96%. Overall the model shows high confidence and detection capabilities that can be improved further with more training.

For keystroke dynamics authentication, the architecture used was a per-user LSTM autoencoder that learns an individual's latent typing signature with support for marking the user the data belongs to. The five timing features: dwell time, interval time, latency time, flight time, and up-to-up time-proved discriminative for capturing behavioral biometrics and enabling reliable anomaly-based authentication. Various methods were trialled for this modality. Among them, LSTM had the best performance and least false anomalies detected. The training dataset contained only legitimate user sequences without pre-labeled anomalies, so the anomaly detection rate was calculated based on false rejections of legitimate users in the test set.

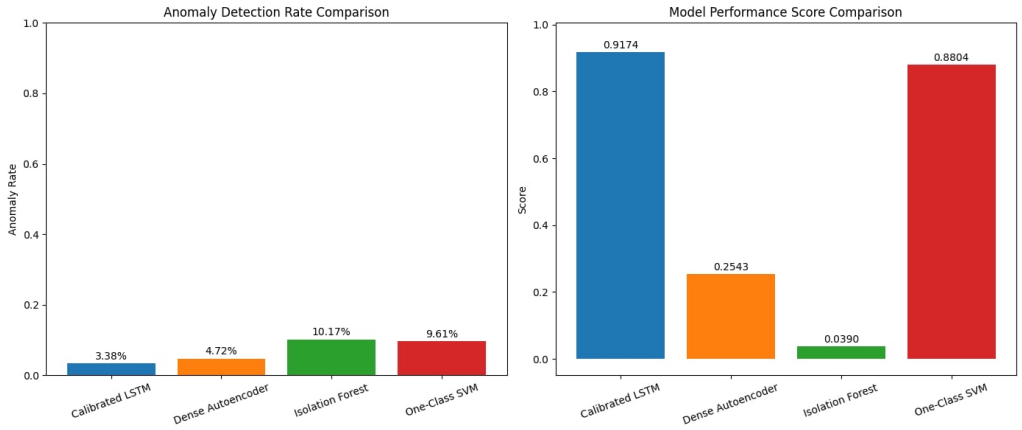

Figure 5.1: Comparison between various methods for keystroke analysis

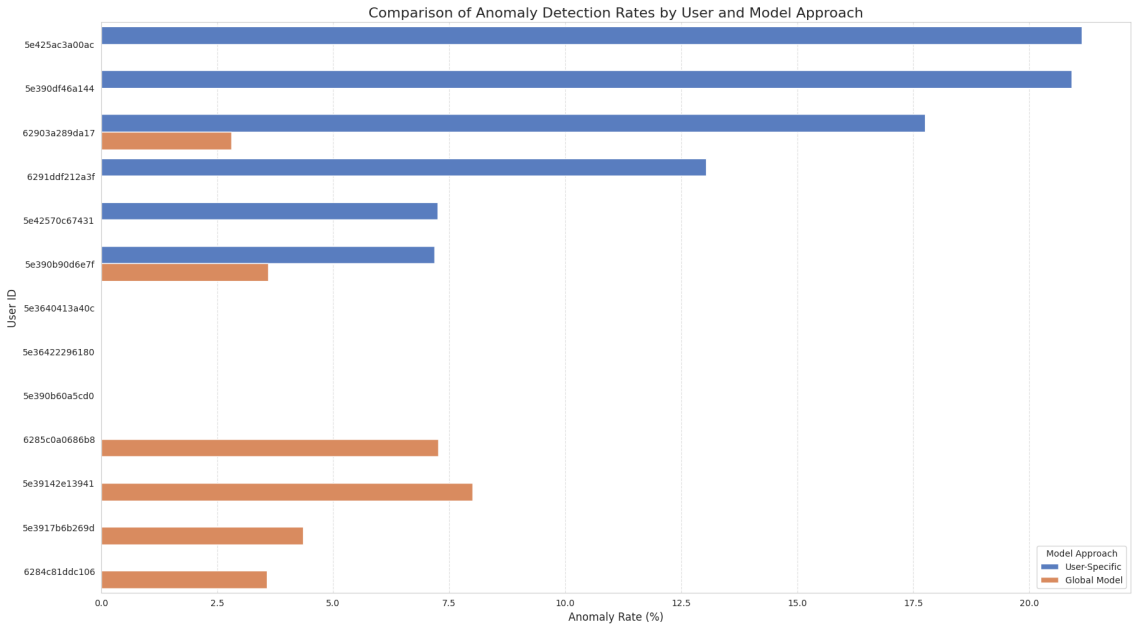

Figure 5.2: Key stroke analysis results comparison

The emotion detection model is built on a Temporal Segment Network with a ResNet body stream, an optional ResNet50 context stream, and an optional Flow body stream. The model is evaluated using mAP and ROC-AUC for multi-label categorical emotions and R^2 for continuous dimensions. On the validation split, the RGB+context model attains mAP = 0.125 and ROC-AUC = 0.576, while the Flow model performs slightly lower on these metrics. Late fusion of RGB+context with Flow further increases mAP and ROC-AUC and trends toward improved R^2, confirming that combining appearance/context with motion yields more reliable clip-level predictions. Moreover, there is room for improvement through threshold tuning.

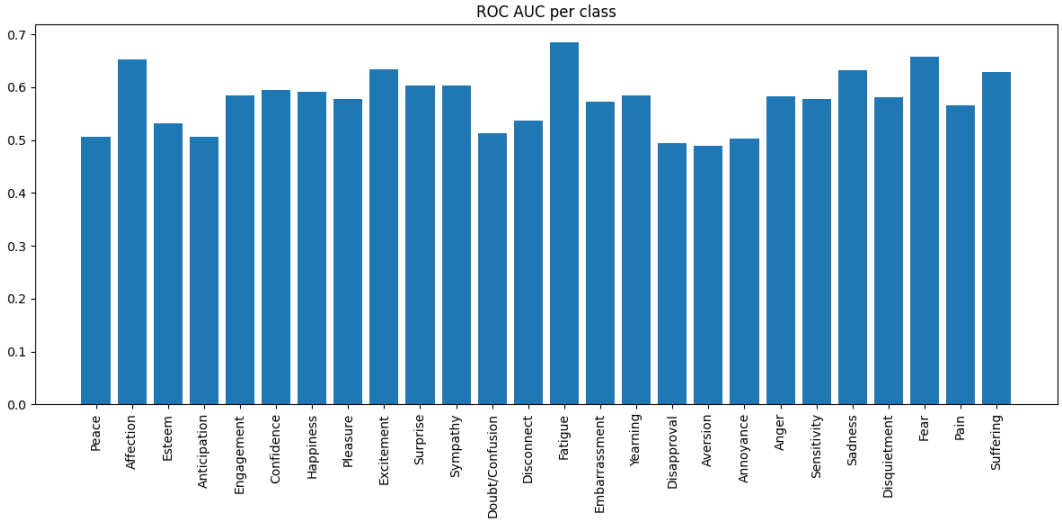

Figure 5.3: ROC metrics of Emotion Detection

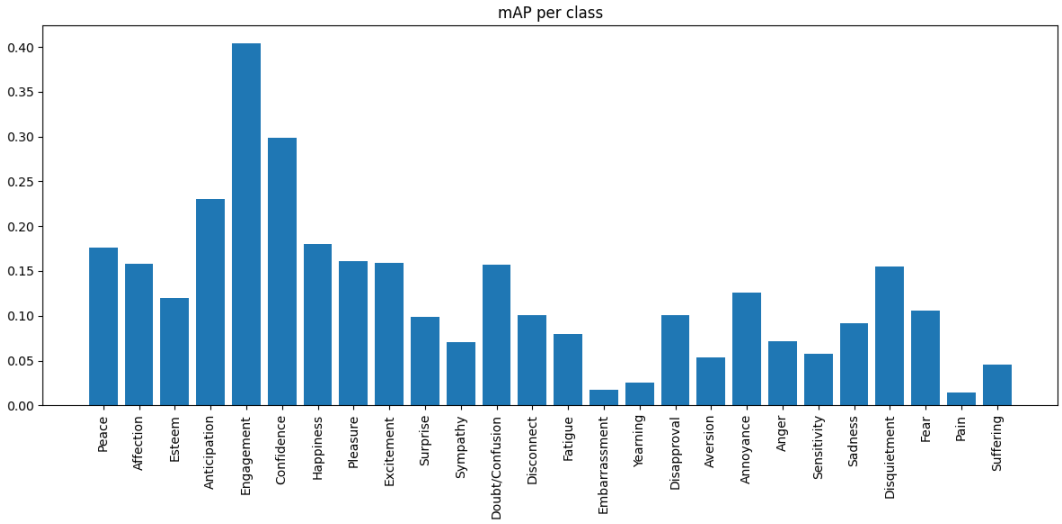

Figure 5.4: AUC metrics of Emotion Detection

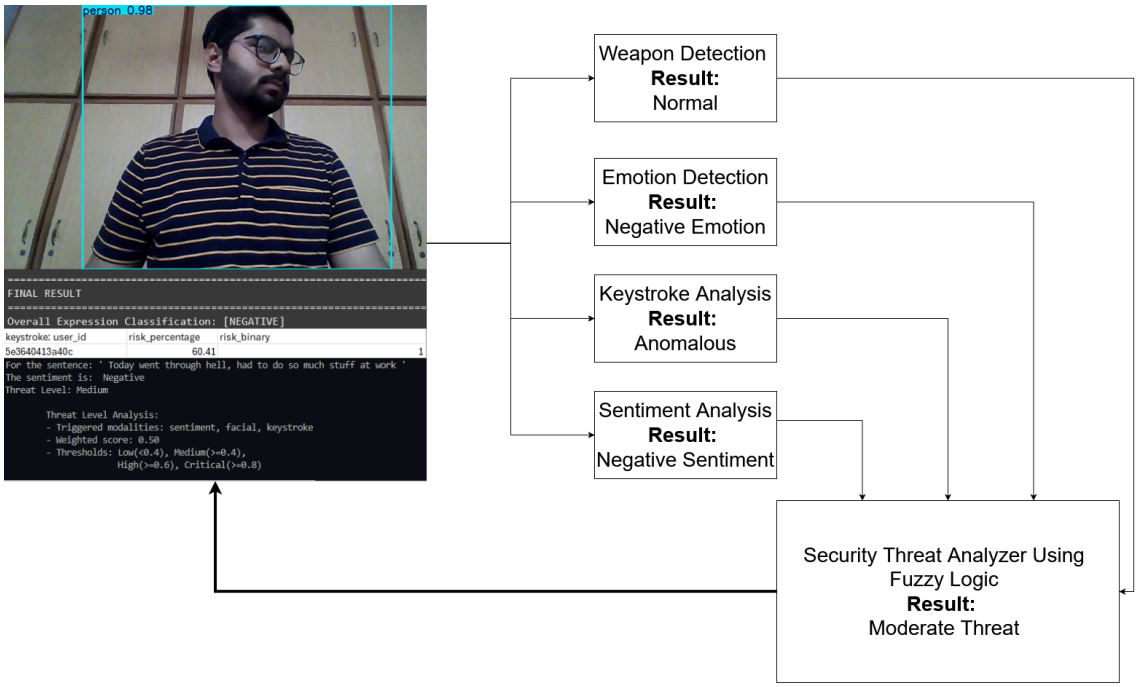

Figure 5.5: Sample full system moderate risk output example

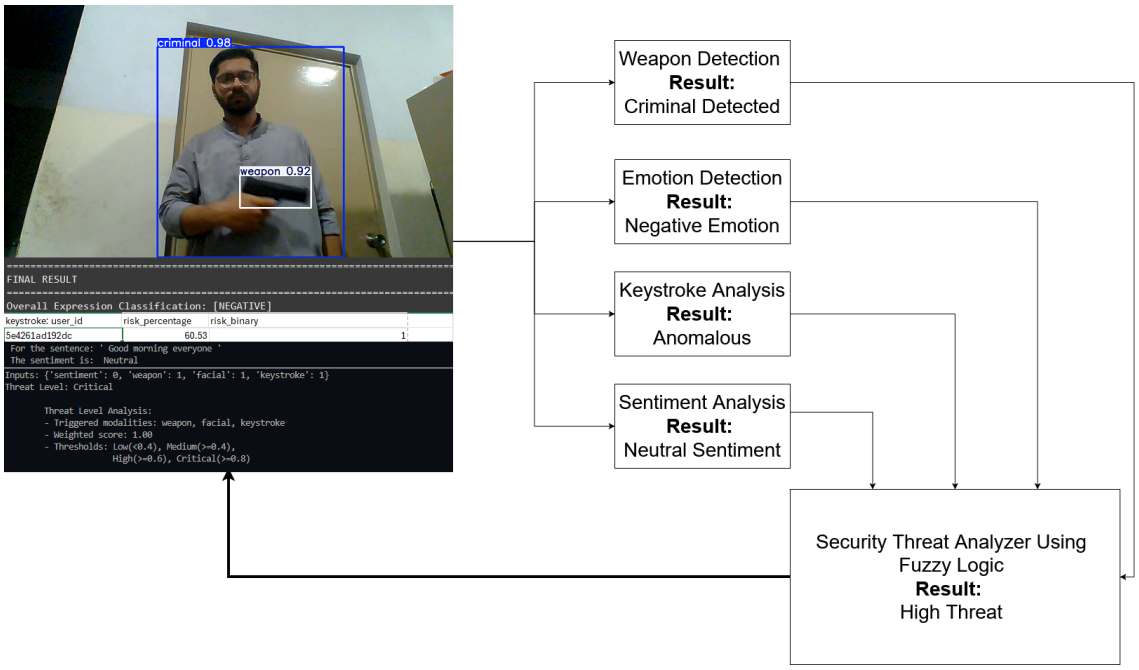

Figure 5.6: Sample full system high risk output example

5.2 Analysis of Design Solutions

For the final score, we elected to use a fuzzy system over other methods as there was little scope for improvement using other methods. This is because we did not have any basis to fuse the different datasets of the various modalities. We also did not have any dataset which had multimodal data that could be used to mimic the use case. As such, we chose to use a fuzzy system and set rules according to the importance of each modality.
Our system is multimodal in design so that we can use various different types of data for our behaviour and activity analysis. For the different modalities, we used weapon detection, emotion detection from body language, keystroke analysis and sentiment analysis from social media text posts.
Weapon detection was chosen because if someone is carrying weapons, it is a very solid indicator of tangible threats. Emotion detection was chosen as strong emotions could be an indicator for suspicious activity. Keystroke analysis was chosen to verify the identity of designated users, especially for sensitive PCs. Lastly, sentiment analysis from social media text posts were done to monitor the sentiment of the employee towards the workplace.
A strength of this system is the high accuracy of the weapon detection model. As such, highly suspicious and dangerous individuals will be frequently and accurately detected. Another strength is the multimodal nature of the system, which allows suspicious activity to be detected through more than one modality and reduce the chances of false alarms while creating opportunities for preventative surveillance.
A limitation of the system is the low accuracy of the emotion detection model. Due to emotion detection from body language being a difficult task, the accuracy of the model is lower compared to the other models that are being used in the system. But we have taken measures to improve detection by analyzing frequency of the top five emotions during criminal activities and tuning our model accordingly.

5.3 Final Design Adjustments

Three pivotal adjustments were made:

1. Introduced a fuzzy logic fusion engine to convert modality outputs into a single continuous risk score, improving interpretability and usability. The rules were based on expert opinion, where the experts interviewed were forensic psychologists, policemen and a lawyer. The majority of them decided that body language was more important than keystroke analysis and social media sentiment analysis. Sentiment analysis was thought of as the least important as according to them, employees in Bangladesh do not talk about their professional life on social media. The experts interviewed include Md Meraz Hossain, who is the Director at Bangladesh Institute of Forensic Psychology and Sciences and Golam Ahmed (Bulbul), who is an Advocate at the Appellate Division and High Court for the Supreme Court of Bangladesh. The rest have not been mentioned out of their own request.

2. Migrated social data acquisition from Selenium to Auto-archiver by Bellingcat for the workplace sentiment analysis pipeline. This was done as Auto-archiver is more user friendly and reliable than Selenium at scraping and storing data from the

web.

3. From analysing 104 crime clips we also found that in threatening situations the body language model consistently detected at least 2 negative emotion labels 64.35% of the time. Based on that we designed our system to take the predictions of the body language detection model if it triggers at least 2 negative emotions. This was done to balance out the poor accuracy of the model.

5.4 Statistical Analysis

Threat Detection Model
From our training of the threat detection model, we found the highest F1 score of 0.93 at confidence threshhold of 0.534. This proves that our threat detection model has a robust performance and a good balance between precision and recall. We achieved 96% accuracy with this model.

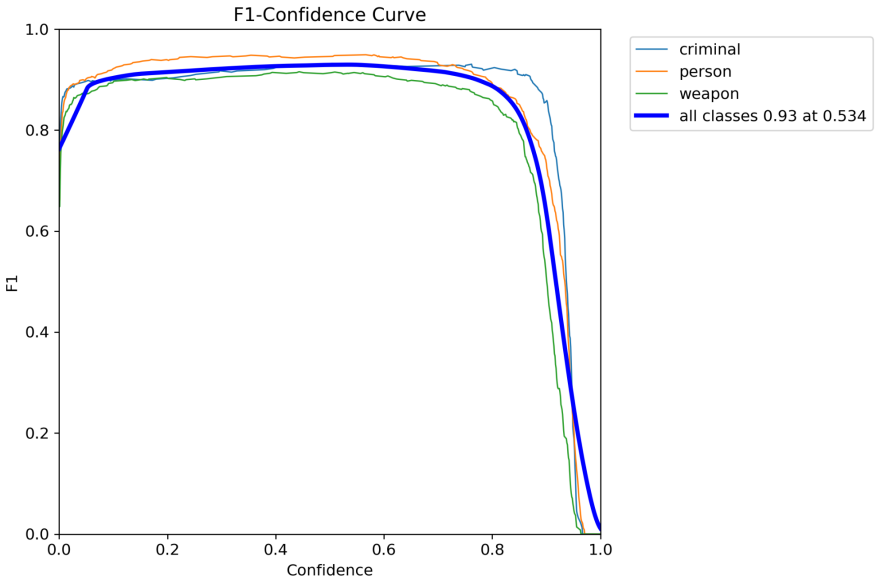

Figure 5.7: Weapon detection model F1-Confidence curve

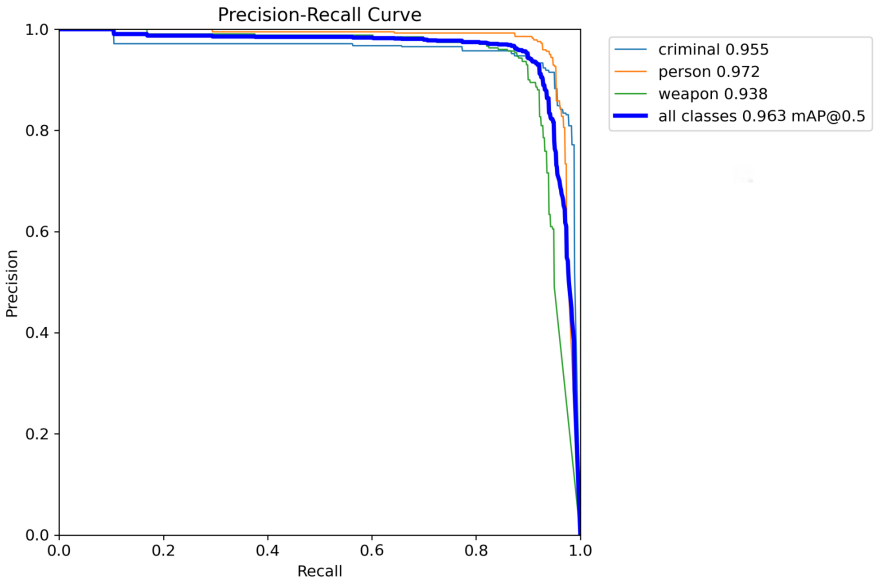

Figure 5.8: Weapon detection model Precision-Recall curve

Body Language Analysis Model

Classes such as Engagement, Confidence, and Anticipation show the strongest F1, indicating comparatively robust detection. Classes with very low F1 (e.g., Pain, Embarrassment, Yearning, Aversion) suffer from both low sensitivity and low reliability, reflecting sparse training examples or subtle visual-social cues.

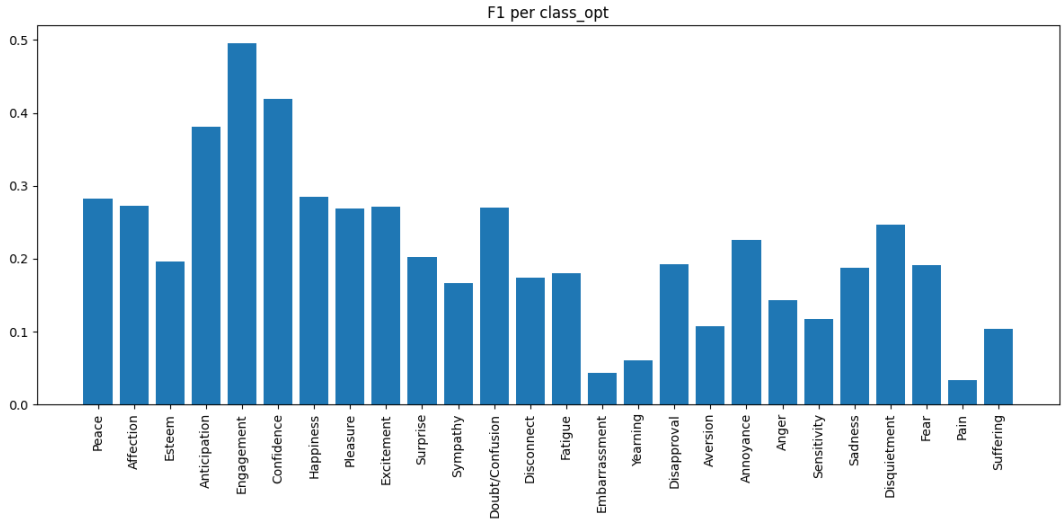

Figure 5.9: F1 Score of Emotion Detection

For the body language analysis model, the observed pattern for Recall per class is as follows: mid-range recall for several positive-valence states (e.g., Engagement, Confidence, Excitement, Anticipation). Lower recall for rarer or more ambiguous states (e.g., Embarrassment, Yearning, Pain, Aversion), implying many true cases are missed.

Moreover, the observed pattern for Precision per class is as follows: high precision for Peace, Engagement, Disapproval, Annoyance, Doubt/Confusion, Surprise, Pain i.e., when the model fires for these classes, it is usually correct. Lower precision for Pleasure, Sympathy, Fatigue, Anger, suggesting greater confusion with neighboring states.

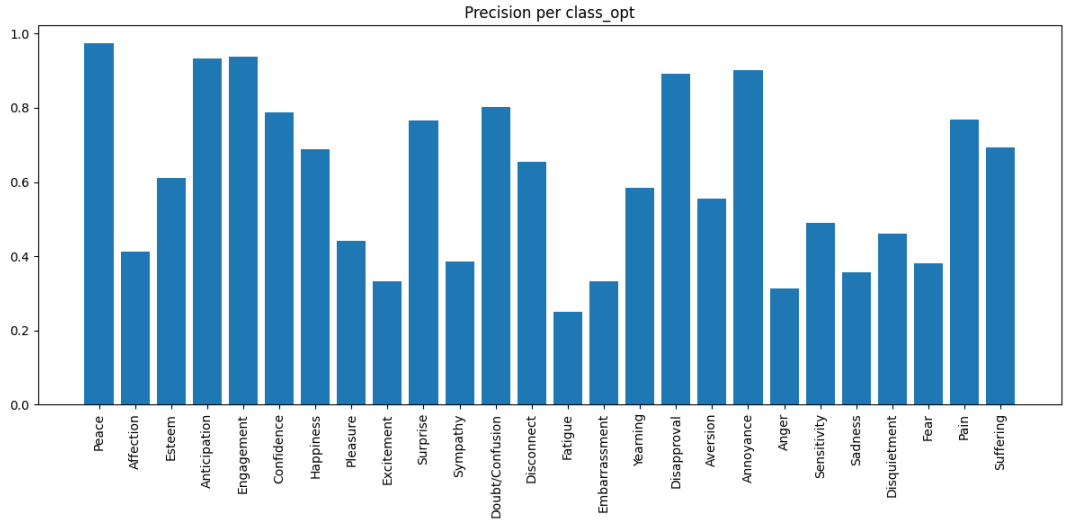

Figure 5.10: Precision metrics of Emotion Detection

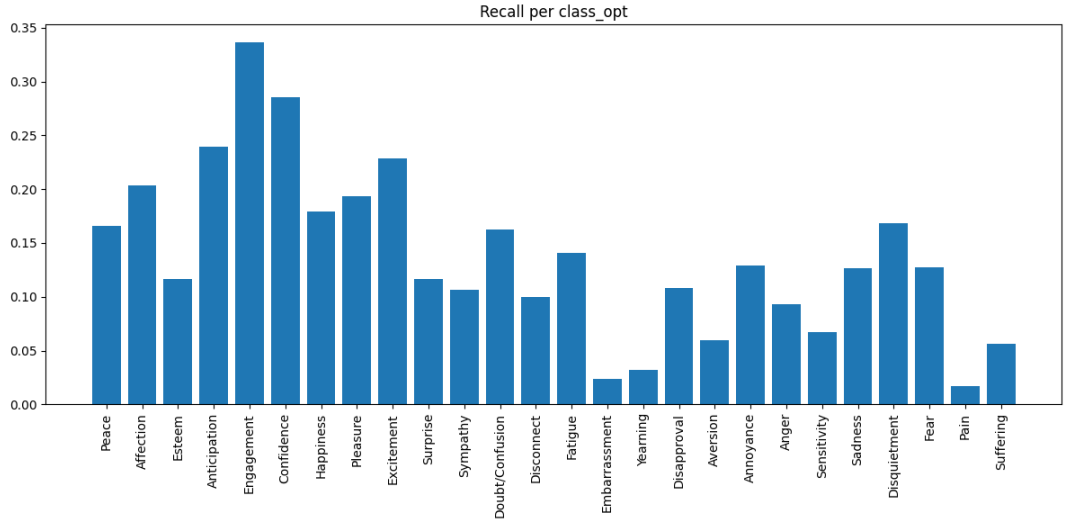

Figure 5.11: Recall metrics of Emotion Detection

To determine which type of emotions were being picked up by the body language detection system in negative scenarios, we inferred 104 clips of violent crimes such as robbery, gun violence, vandalism. Sensitivity and disquietment was most consistently triggered. From our analysis we also found that in threatening situations the body language model consistently detected at least 2 negative emotion labels 64.35% of the time.

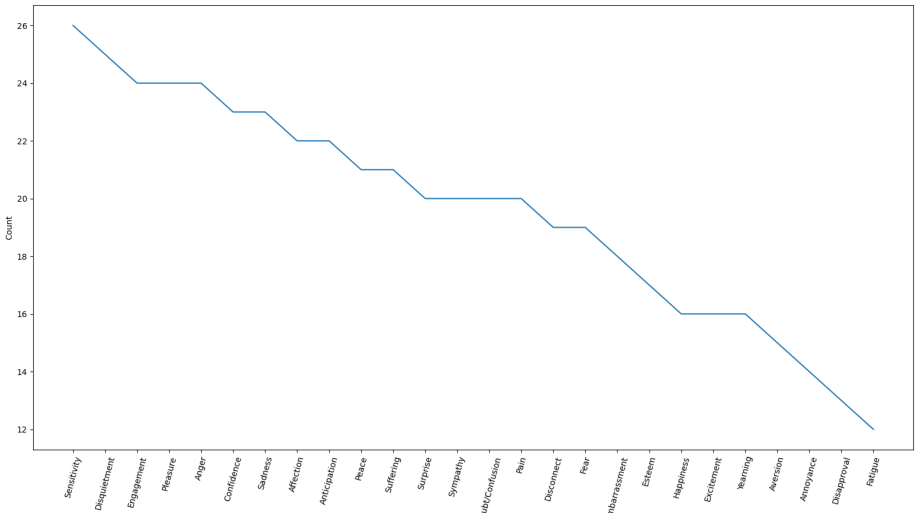

Figure 5.12: Frequency of top 5 emotions detected in crime footage

Keystroke Analysis

For the keystroke analysis dataset, the Dwell time shows a wide spread, while Interval time, Latency time, Flight time, and Uptoup time all display strong right-skewed patterns, indicating that most keystrokes are fast with occasional long delays.

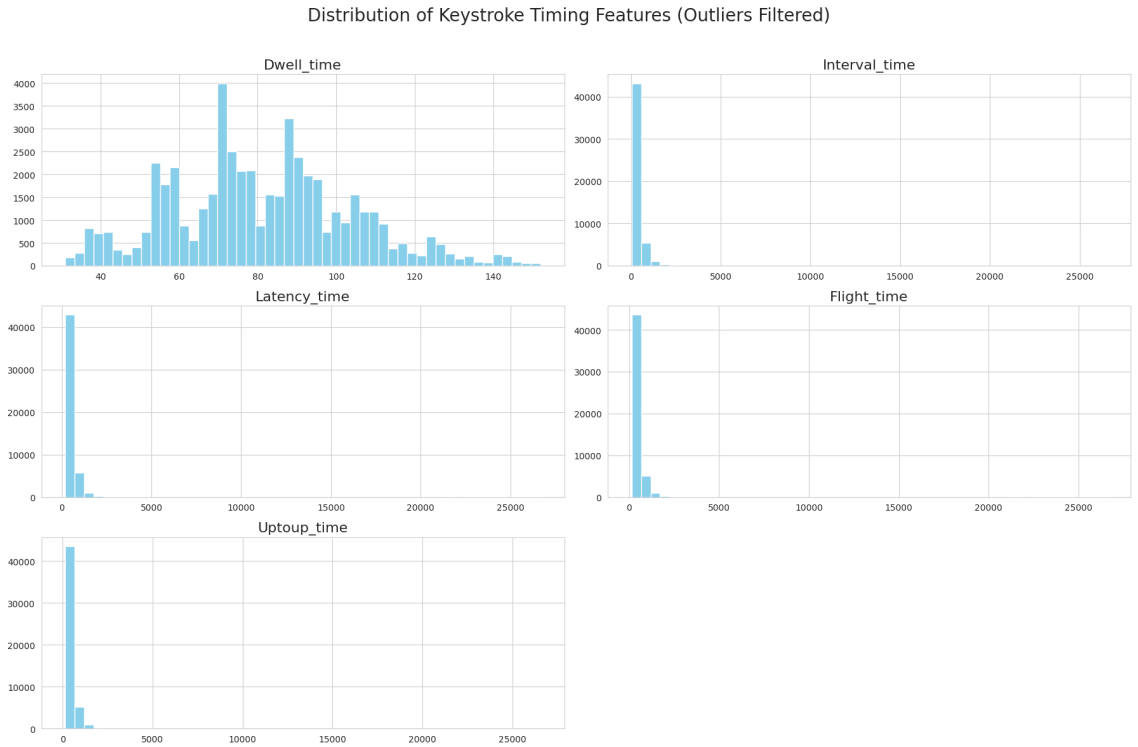

Figure 5.13: Distribution of keystroke timing features

5.4.1 Performance Metrics Across Modalities

Modality	Key Metric	Value	Interpretation
Weapon Detection	mAP	96	High Detection Accuracy, Suitable for Real-Time Deployment
Emotion Detection	mAP/ROC-AUC	0.125/0.576	Moderate Performance, Room for Improvement
Keystroke Analysis	Average Score/Anomaly Rate	0.917/0.033	LSTM Model, outperforms other ML techniques in learning the keystroke pattern

5.5 Comparisons and Relationships

From our literature review, we have determined the feasibility to implement YOLO algorithm to detect weapons and suspicious individuals. Khalid et al [28] has shown in their paper that YOLOv5 achieves an impressive accuracy of 95%. We trained their dataset on YOLO11 and achieved 96% accuracy on the test set. This rate of accuracy can help a trained personnel to detect weapons more effectively.

Paper	Technique	Accuracy(%) ↑
Verma et al[8]	SVM and DCN	89.9
Olmos et al[7]	Faster R-CNN	91.43
Mehta et al[12]	YOLOv3	90.3
Hashmi et al[14]	YOLOv4	82
Elmir et al[10]	Mobile-SDD	90
Khalid et al[28]	YOLOv5	95.43
Proposed	**YOLO11**	**96**

Table 5.1: Comparison of proposed approach with exisiting Weapon Detection approaches

Four different machine learning techniques were attempted for keystroke analysis: One-Class SVM, Dense Autoencoder, Isolation Forest, and Calibrated LSTM. There are two metrics for each technique, which are:

- Average Score: A performance metric where higher values indicate better performance
- Anomaly Rate: which measures the percentage of anomalous patterns detected

The Calibrated LSTM performed best with the highest average score (0.917) and lowest anomaly rate (0.033), indicating it was most effective at measuring employee

keystroke patterns while minimizing false anomaly detections. As such, the Calibrated LSTM approach was the most effective technique for keystroke authentication in this analysis.

Technique	Average Score↑	Anomaly Rate↓
One-Class SVM	0.880	0.096
Dense Autoencoder	0.254	0.047
Isolation Forest	0.038	0.101
Calibrated LSTM	**0.917**	**0.033**

Table 5.2: Comparison between various techniques for Keystroke Analysis

Institution	AP	ROCAUC	R^2	ERS
National Technical University of Athens	0.1796	0.6416	0.1141	0.2624
Beijing Normal University	0.5527	0.5000	−0.0048	0.2608
National Technical University of Athens	0.1792	0.1015	0.1141	0.2564
National Technical University of Athens	0.1746	0.0965	0.1141	0.2512
Brac University (TSN RGB and Context)	**0.1250**	**0.5760**	**-0.0086**	**0.1710**
Beijing Normal University	0.1185	0.5473	−0.0079	0.1625
Beijing Normal University	0.1076	0.5061	−0.0085	0.1492

Table 5.3: Comparison of proposed approach with exisiting Emotion Detection approaches based on the BOLD Dataset

The paper "A Risk Assessment Framework Based on Fuzzy Logic for Automotive Systems" (University of Pisa, ISTI-CNR, 2024) [39] presents a MATLAB-based fuzzy inference system to assess cybersecurity risks using feasibility and impact inputs aligned to ISO/SAE 21434 scenarios. This work corroborates the choice of

fuzzy methods for uncertainty handling and graded risk decisions. The present design extends that paradigm by ingesting multi-modal, real-time personnel-centric signals facial recognition, keystroke dynamics, sentiment, and weapon detection implemented in Python via skfuzzy and oriented to the nuclear security domain.

5.6 Discussions

The results collectively indicate that a multi-modal, real-time, and uncertainty-aware architecture can meaningfully improve personnel security in wide range of situations. Identity assurance is strengthened through concurrent face and keystroke verification, sentiment tracking provides context for early insider risk identification, and YOLO-based weapon detection contributes immediate physical risk coverage. The fuzzy fusion system uses these modalities to provide a single risk score which can be acted upon. Compared to existing AI surveillance literature, the proposed system is unique as it follows a multimodal approach to detect suspicion. For deployment, sufficient compute needs to be provisioned and proper management of data needs to be considered. The comparative analysis further validates the core design choices: transformer-based sentiment for nuanced workplace context, YOLO for low-latency visual interdiction, and fuzzy logic for decision fusion under uncertainty. Together, these components form a coherent, scalable, and ethically grounded approach to personnel security in critical infrastructure.

Chapter 6

Conclusion

6.1 Summary of Findings

Implementing multimodal analysis to improve security system is a realistic but difficult concept with current level of advancement in AI. Our work has found that:

1. Weapon and threat detection model based on YOLO object detection has 96% accuracy.

2. Keystroke analysis can continuously detect if a PC has been breached by an imposter or not or if the person using it is acting suspiciously.

3. Emotion detection from body language of a person, via CCTV footage can be used for preventative monitoring when combined with other analysis.

4. When outputs of the four different systems are combined in a fuzzy logic engine, it provides a smart threat rating that is easily interpretable for the human operator.

5.Multi-modal analysis increases robustness of the surveillance system and makes it multifaceted.

6.Well tuned rules decreases the possibilities of false alarms.

6.2 Contributions to the Field

The project can reduce the human labor in survellience by utilizing AI detection and by extension lower human error. It can even be used for home security so homeowners can have an extra line of defence.
Since the model combines multiple facets of individuals, the system can be used to study correlation between the different aspects like keystrokes, emotions, online sentiments to criminal tendencies.

6.3 Recommendations for Future Work

We plan on further training the model for the weapon and threat detection model for better predictions. Analyzing criminal body language to add possible predictive capabilities would create an even more robust system. We plan on adding a system that analyzes employee background to find their tendency to commit crime for professional environments.

Bibliography

[1] A. Braseth, C. NIHLWING, H. Svengren, Ø. VELAND, L. Hurlen, and J. Kvalem, "Lessons learned from halden project research on human system interfaces," *Nuclear Engineering and Technology*, vol. 41, Apr. 2009. DOI: 10.5516/NET.2009.41.3.215

[2] Z. Chang, D. Cai, W. Zhen, and Q. Huang, "Design and implementation of a measurement system for power station video and environment surveillance system," in *International Journal of Smart Grid and Clean Energy*, 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:55985673

[3] P. S. Teh, A. B. J. Teoh, and S. Yue, "A survey of keystroke dynamics biometrics," *The Scientific World Journal*, vol. 2013, no. 1, p. 408 280, 2013. DOI: https://doi.org/10.1155/2013/408280 eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1155/2013/408280. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1155/2013/408280

[4] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, *You only look once: Unified, real-time object detection*, 2016. arXiv: 1506.02640 [cs.CV]. [Online]. Available: https://arxiv.org/abs/1506.02640

[5] V. Agarwal, J. W. Buttles, L. H. Beaty, J. Naser, and B. P. Hallbert, "Wireless online position monitoring of manual valve types for plant configuration management in nuclear power plants," *IEEE Sensors Journal*, vol. 17, no. 2, pp. 311–322, 2017. DOI: 10.1109/JSEN.2016.2615131

[6] D. Lee and J. Kim, "Autonomous algorithm for safety systems of the nuclear power plant," in *Advances in Human Factors in Energy: Oil, Gas, Nuclear and Electric Power Industries*, P. Fechtelkotter and M. Legatt, Eds., Cham: Springer International Publishing, 2017, pp. 72–82, ISBN: 978-3-319-60204-2.

[7] R. Olmos, S. Tabik, and F. Herrera, *Automatic handgun detection alarm in videos using deep learning*, 2017. arXiv: 1702.05147 [cs.CV]. [Online]. Available: https://arxiv.org/abs/1702.05147

[8] G. Verma and A. Dhillon, "A handheld gun detection using faster r-cnn deep learning," Nov. 2017, pp. 84–88. DOI: 10.1145/3154979.3154988

[9] Y. Luo, J. Ye, R. B. A. Jr., J. Li, M. G. Newman, and J. Z. Wang, "ARBEE: towards automated recognition of bodily expression of emotion in the wild," *CoRR*, vol. abs/1808.09568, 2018. arXiv: 1808.09568. [Online]. Available: http://arxiv.org/abs/1808.09568

[10] Y. Elmir, S. Laouar, and L. Hamdaoui, "Deep learning for automatic detection of handguns in video sequences," Apr. 2019.

[11] P. Filntisis, N. Efthymiou, G. Potamianos, and P. Maragos, "Emotion understanding in videos through body, context, and visual-semantic embedding loss," in Jan. 2020, pp. 747–755, ISBN: 978-3-030-66414-5. DOI: 10.1007/978-3-030-66415-2_52

[12] P. Mehta, A. Kumar, and S. Bhattacharjee, "Fire and gun violence based anomaly detection system using deep neural networks," in *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 2020, pp. 199–204. DOI: 10.1109/ICESC48915.2020.9155625

[13] J. M. O'Hara and S. Fleger, "Human-system interface design review guidelines," Brookhaven National Lab. (BNL), Upton, NY (United States), Tech. Rep., Jul. 2020. DOI: 10.2172/1644018 [Online]. Available: https://www.osti.gov/biblio/1644018

[14] T. S. S. Hashmi, N. U. Haq, M. M. Fraz, and M. Shahzad, "Application of deep learning for weapons detection in surveillance videos," in *2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2)*, 2021, pp. 1–6. DOI: 10.1109/ICoDT252288.2021.9441523

[15] P. Sivakumar, J. V, R. R, and K. S, "Real time crime detection using deep learning algorithm," in *2021 International Conference on System, Computation, Automation and Networking (ICSCAN)*, 2021, pp. 1–5. DOI: 10.1109/ICSCAN53069.2021.9526393

[16] H. Yang, B. Zeng, M. Xu, and T. Wang, "Back to reality: Leveraging pattern-driven modeling to enable affordable sentiment dependency learning," *CoRR*, vol. abs/2110.08604, 2021. arXiv: 2110.08604. [Online]. Available: https://arxiv.org/abs/2110.08604

[17] L. Yang and S.-F. Qin, "A review of emotion recognition methods from keystroke, mouse, and touchscreen dynamics," *IEEE Access*, vol. 9, pp. 162 197–162 213, 2021. DOI: 10.1109/ACCESS.2021.3132233

[18] A. Ayodeji, M. A. Amidu, S. A. Olatubosun, Y. Addad, and H. Ahmed, "Deep learning for safety assessment of nuclear power reactors: Reliability, explainability, and research opportunities," *Progress in Nuclear Energy*, vol. 151, p. 104 339, 2022, ISSN: 0149-1970. DOI: https://doi.org/10.1016/j.pnucene.2022.104339 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0149197022002141

[19] N. Y. Katkar and V. K. Garg, "Detection and tracking the criminal activity using network of cctv cameras," in *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)*, 2022, pp. 664–668. DOI: 10.1109/ICOSEC54921.2022.9952104

[20] I. Katser, D. Raspopov, V. Kozitsin, and M. Mezhov, *Machine learning methods for anomaly detection in nuclear power plant power transformers*, 2022. arXiv: 2211.11013 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2211.11013

[21] Z. Ma, H. Bao, S. Zhang, M. Xian, and A. L. Mack, "Exploring advanced computational tools and techniques with artificial intelligence and machine learning in operating nuclear plants," Idaho National Lab. (INL), Idaho Falls, ID (United States), Tech. Rep., Feb. 2022. DOI: 10.2172/1847070 [Online]. Available: https://www.osti.gov/biblio/1847070

[22] R. Shenoy, D. Yadav, H. Lakhotiya, and J. Sisodia, "An intelligent framework for crime prediction using behavioural tracking and motion analysis," in *2022 International Conference on Emerging Smart Computing and Informatics (ESCI)*, 2022, pp. 1–6. DOI: 10.1109/ESCI53509.2022.9758281

[23] F. K. Sufi, I. Razzak, and I. Khalil, "Tracking anti-vax social movement using ai-based social media monitoring," *IEEE Transactions on Technology and Society*, vol. 3, no. 4, pp. 290–299, 2022. DOI: 10.1109/TTS.2022.3192757

[24] M. Wankhade, A. C. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review*, vol. 55, pp. 5731–5780, Feb. 2022. DOI: 10.1007/s10462-022-10144-1

[25] V. Yadav, R. Christian, S. Prescott, and S. Germain, "Risk-informed physical security assessment for nuclear power plants," English, in *16th International Conference on Probabilistic Safety Assessment and Management, PSAM 2022*, Funding Information: This research was supported through the U.S. Department of Energy's Light Water Reactor Sustainability program. Publisher Copyright: © 2022 Probabilistic Safety Assessment and Management, PSAM 2022. All rights reserved.; 16th International Conference on Probabilistic Safety Assessment and Management, PSAM 2022 ; Conference date: 26-06-2022 Through 01-07-2022, 2022.

[26] C. Yavuz and S. Şentürk Lüle, "The application of artificial intelligence to nuclear power plant safety," in *Artificial Intelligence for Knowledge Management, Energy, and Sustainability*, E. Mercier-Laurent and G. Kayakutlu, Eds., Cham: Springer International Publishing, 2022, pp. 117–127, ISBN: 978-3-030-96592-1.

[27] E. Gursel et al., "Using artificial intelligence to detect human errors in nuclear power plants: A case in operation and maintenance," *Nuclear Engineering and Technology*, vol. 55, no. 2, pp. 603–622, 2023, ISSN: 1738-5733. DOI: https://doi.org/10.1016/j.net.2022.10.032 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1738573322005137

[28] S. Khalid, A. Waqar, H. Tahir, O. Edo, and I. Tenebe, "Weapon detection system for surveillance and security," Mar. 2023, pp. 1–7. DOI: 10.1109/ITIKD56332.2023.10099733

[29] A. Kumari, R. Dubey, and I. Sharma, "Shnp: Shielding nuclear plants from cyber attacks using artificial intelligence techniques," Nov. 2023. DOI: 10.1109/AICERA/ICIS59538.2023.10420386

[30] V. Mandalapu, L. Elluri, P. Vyas, and N. Roy, "Crime prediction using machine learning and deep learning: A systematic review and future directions," *IEEE Access*, vol. 11, pp. 60 153–60 170, 2023. DOI: 10.1109/ACCESS.2023.3286344

[31] A. Nasry, A. Ezzahout, and F. Omary, "People tracking in video surveillance systems based on artificial intelligence," *Journal of Automation, Mobile Robotics and Intelligent Systems*, pp. 59–68, Dec. 2023. DOI: 10.14313/JAMRIS/1-2023/8

[32] M. Saketh, N. Nandal, R. Tanwar, and B. P. Reddy, "Intelligent surveillance support system," en, *Discov. Internet Things*, vol. 3, no. 1, Sep. 2023.

[33] T. Wang, Y. Ma, Z. Wang, X. Wu, and L. Li, "User behavior analysis based on big data and artificial intelligence," in Nov. 2023, ISBN: 9781643684444. DOI: 10.3233/FAIA230881

[34] H. Yang, C. Zhang, and K. Li, "Pyabsa: A modularized framework for reproducible aspect-based sentiment analysis," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM 2023, Birmingham, United Kingdom, October 21-25, 2023*, I. Frommholz et al., Eds., ACM, 2023, pp. 5117–5122. DOI: 10.1145/3583780.3614752 [Online]. Available: https://doi.org/10.1145/3583780.3614752

[35] A. Chaudhary, J. Han, S. Kim, A. Kim, and S. Choi, "Anomaly detection and analysis in nuclear power plants," *Electronics*, vol. 13, no. 22, 2024, ISSN: 2079-9292. DOI: 10.3390/electronics13224428 [Online]. Available: https://www.mdpi.com/2079-9292/13/22/4428

[36] D. A. Ejigu, Y. Tuo, and X. Liu, "Application of artificial intelligence technologies and big data computing for nuclear power plants control: A review," *Frontiers in Nuclear Engineering*, vol. 3, 2024, ISSN: 2813-3412. DOI: 10.3389/fnuen.2024.1355630 [Online]. Available: https://www.frontiersin.org/journals/nuclear-engineering/articles/10.3389/fnuen.2024.1355630

[37] A. Hall and V. Agarwal, "Barriers to adopting artificial intelligence and machine learning technologies in nuclear power," *Progress in Nuclear Energy*, vol. 175, p. 105 295, 2024, ISSN: 0149-1970. DOI: https://doi.org/10.1016/j.pnucene.2024.105295 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0149197024002452

[38] G. Jocher and J. Qiu, *Ultralytics yolo11*, version 11.0.0, 2024. [Online]. Available: https://github.com/ultralytics/ultralytics

[39] F. Merola, C. Bernardeschi, and G. Lami, "A risk assessment framework based on fuzzy logic for automotive systems," *Safety*, vol. 10, no. 2, 2024, ISSN: 2313-576X. DOI: 10.3390/safety10020041 [Online]. Available: https://www.mdpi.com/2313-576X/10/2/41

[40] N. Nanglae and P. Bhattarakosol, "Keystroke Dynamics Data," Oct. 2024. DOI: 10.6084/m9.figshare.27135000.v1 [Online]. Available: https://figshare.com/articles/dataset/Keystroke_Dynamics_Data/27135000

[41] F. K. Sufi and I. Khalil, "Automated disaster monitoring from social media posts using ai-based location intelligence and sentiment analysis," *IEEE Transactions on Computational Social Systems*, vol. 11, no. 4, pp. 4614–4624, 2024. DOI: 10.1109/TCSS.2022.3157142

[42] H. Yang and K. Li, "Modeling aspect sentiment coherency via local sentiment aggregation," in *Findings of the Association for Computational Linguistics: EACL 2024, St. Julian's, Malta, March 17-22, 2024*, Y. Graham and M. Purver, Eds., Association for Computational Linguistics, 2024, pp. 182–195. [Online]. Available: https://aclanthology.org/2024.findings-eacl.13