Smart Surveillance System for Public Mental Health Well Being

Keerthana R.
Dept. of CSE, School of Computer Science Programme
Dr. Ambedkar Institute of Technology
Bangalore, India
keerthana240904@gmail.com

Lavanya Santhosh
Dept. of CSE, School of Computer Science Programme
Dr. Ambedkar Institute of Technology
Bangalore, India
lavanyasanthosh.cs@drait.edu.in

Kavya Shree R
Dept. of CSE, School of Computer Science Programme
Dr. Ambedkar Institute of Technology
Bangalore, India
kavyaraj0413@gmail.com

Veena Potdar
Dept. of CSE, School of Computer Science Programme
Dr. Ambedkar Institute of Technology
Bangalore, India
veenapotdar.cs@drait.edu.in

Asha K.N.
Dept. of CSE, School of Computer Science Programme
Dr. Ambedkar Institute of Technology
Bangalore, India
ashakn.cs@drait.edu.in

Likitha H.M.
Dept. of CSE, School of Computer Science Programme
Dr. Ambedkar Institute of Technology
Bangalore, India
likithahm6@gmail.com

Nagaratna Biradar
Dept. of CSE, School of Computer Science Programme
Dr. Ambedkar Institute of Technology
Bangalore, India
nagaratna.b.tech@gmail.com

***Abstract*—Existing Smart Surveillance Systems usually are only designed for one task such as violence detection, or extract only one feature for analysis, thereby missing a lot of important context and data that could not only drastically improve the results of the application, but also provide in-depth analysis of the scene that humans alone cannot achieve. This is exactly what the Smart Surveillance System for Public Mental Wellbeing tries to achieve. This application makes use of lightweight, accurate and privacy preserving models to analyze various features detected in a surveillance footage, and provide a dashboard that updates in real time and not only reports anomalies, but also keeps track of both positive and negative states of the area, and provides an explained, in-depth analysis of the data collected.**

Keywords—Anomaly Detection, Artificial Intelligence, Crowd Analysis, Machine Learning, Smart Surveillance Systems.

I. Introduction

Machine Learning (ML) is a branch of Artificial Intelligence (AI) that is being used in various fields since it has a vast potential due to its capability to be able to do tasks that humans can't do, or can't accurately and efficiently complete. It reaches problems in various domains and test cases that humans alone could never imagine solving themselves.

There have been various innovations in the fields of Surveillance system management and public health and mental health monitoring in the past few years along with the growth of AI. The more we learn about the endless possibilities of AI in these fields, the more opportunities we see for improvement and innovations in this fields. Traditional surveillance systems typically consist of a huge network of surveillance cameras, and constantly produce live streams of real-time surveillance footages that realistically cannot be managed and analysed efficiently by human beings. Humans are bound to miss critical details from surveillance camera feeds, and surveillance footage is often only reviewed after an incident has occurred and damage has been caused.

This issue is prevalent not only due to the lack of adequate staffing for surveillance systems, but also because present smart surveillance systems focus mostly on one factor or incident, such as emotion or violence detection. A better approach would be to focus on multiple features to improve the system's understanding of the footage being analysed, and keeping track of positive outcomes in an area, which will in turn aid the systems to catch relevant anomalies before an adverse incident occurs, thereby helping prevent such incidents.

This is achieved by creating a smart surveillance system whose focus is not just violence detection or group emotion detection, but focuses on the group's overall mental health well-being. This way, events such as mass mood shifts, suspicious gatherings, etc., can be immediately flagged, so that the user of the smart surveillance system can interpret the urgency of the situation and take immediate action to mitigate it.

A. *Existing Systems*

Smart Surveillance Systems is a new, yet evolving field of application of AI and ML, and as a result, various different types of Surveillance Systems exist, and each of these perform a different operation on the surveillance footage. Some examples of existing Smart Surveillance Systems include Crowd Behavior Analysis and Forecasting, Violence Detection, Weapons Detection and Group Emotion Detection, and there also exist various system for public mental health monitoring, however, most of these systems focused highly on social media and textual data. Various public health detection systems also exist, and most of these systems focus on group emotion or body posture or one of the other extractable features from the surveillance footage.

B. *Proposed System*

The above-mentioned existing systems all shared one main characteristic – they all only considered one (or two, at maximum) features to focus on. They either focused on facial emotions, or crowd behavior characteristics, or body posture, thereby missing crucial data that can provide a lot of information about the area being surveilled.

To overcome this issue, we propose a Smart Surveillance System for Public Mental Well-being, that makes use of four well-trained core models including a crowd emotion detector, environmental factors detector, crowd analyser model and a body posture and language analysis model.

This system is designed to make smart decisions based on the input feed by:

- Detecting which features are extractable, and running analysis on the surveillance footage accordingly.
- To work in various contexts such as classrooms, concerts, and busy streets
- Provides an interactive dashboard displaying the states of the scene recorded by the surveillance camera with the help of an 8x8 chessboard-like grid, further assisting immediate action in cases of any detected anomalies.

II. Literature Survey

AI and computer-vision (CV) techniques have increasingly been applied to mental-health analysis and smart surveillance, although most existing systems treat these domains separately. Prior works vary widely in features used, modalities analyzed, and the degree of real-world applicability. This survey synthesizes contributions from 2021–2025 focusing on emotion recognition, crowd analysis, behavioral monitoring, and multimodal public mental-health assessment, with special emphasis on deployment in in-the-wild CCTV environments.

A. Early Work: Textual and Social Media Mental-Health Detection

Early models relied primarily on social-media text and profile metadata for depression and suicide-risk detection [1 -7]. Approaches such as sentiment-aware depression identification [4], psychology-informed suicide-risk modeling [5], and ensemble deep learning for suicidal ideation [10, 13] achieved promising results (AUC 76–90%). However, challenges included noisy labels, class imbalance, cross-platform bias, and limited clinical validation.

B. Clinical and Multimodal Depression Detection

Clinical multimodal studies using facial videos, audio recordings, and 3D facial dynamics demonstrated strong performance in controlled settings (78–90%) [8, 11, 12]. Although effective, these systems depend on high-quality, close-range recordings (e.g., Kinect, controlled interviews), limiting their application to far-field CCTV and public surveillance environments.

C. Behavioural and Crowd Analysis

Crowd-analysis and anomaly-detection systems demonstrated reliable surveillance capabilities. Research includes serverless city-surveillance platforms [2], trajectory-based personality mapping [3], density estimation using MCNN and multi-stage bilinear CNNs [25 ,36], and large-scale datasets such as HAJJv2 for abnormal behaviour detection [37]. Despite high benchmark accuracy (>85%), limitations include occlusion, illumination issues, and poor generalization to crowded, real-world public spaces.

D. Gait and Behaviometric Surveillance

Gait-based behaviometric identification using GEIs and hybrid CNN/GCN models enables privacy-preserving remote monitoring and shows strong recognition accuracy across varied environments [51]. These methods provide behavioural cues relevant to non-intrusive mental-wellbeing monitoring but require robustness against viewpoint changes and scene dynamics.

E. Real-time Surveillance, Privacy and Ethical Constraints

Recent works emphasize real-time anomaly detection and privacy-preserving AI models. Hybrid CNN–autoencoder–ESN models perform well on benchmark datasets [26]. Federated learning and differential privacy have been applied for surveillance under strict anonymity constraints [27]. Women's safety systems, cyber-physical surveillance models, and real-time alerting pipelines address practical safety needs but depend heavily on dataset quality and environmental consistency [28, 30]. Ethical frameworks highlight risks related to consent, bias, and societal impact [17], while acceptance studies reveal generational and cultural barriers in deploying mental-health AI assistants [14].

F. Bias, Interpretability and Domain Adaptation

Bias analyses show demographic and gender-related performance disparities in mental-health detection [18, 19]. Cross-modal distillation for emotion perception [24] demonstrates scalable training but suffers from noisy supervisory signals and limited domain transfer. These findings underline the need for transparent, interpretable models for public mental-wellbeing analysis.

G. Specialized Applications (2022-2024)

Applied systems span student mental-health platforms, urban perception analysis, wearable EEG emotion detectors, panic-attack prediction using wearables, and multilingual mental-health monitoring [38-46]. Though many report high accuracy (85–97%), they are often constrained by small datasets, cultural biases, lack of longitudinal data, or difficulty adapting to highly variable public environments.

H. Recent Machine Learning Foundations Relevant to Our Work

Gait-based behaviour recognition for surveillance [51], comparative ML studies for healthcare risk prediction [52, 54], and stroke/outcome modeling [53] demonstrate the effectiveness of lightweight supervised learning models (Naïve Bayes, SVM, Random Forest, KNN, Decision Trees) for high-risk domains. These insights guide our selection of interpretable, resource-efficient models for real-time mental-wellbeing surveillance.

Overall, prior research shows strong progress across isolated domains—social-media mental-health detection, clinical multimodal diagnosis, crowd analytics, gait behaviometrics, and privacy-aware learning. However, no existing system unifies these strengths into a real-time, privacy-preserving, CCTV-compatible framework for public mental-wellbeing analysis. Many existing models depend on controlled datasets, require high computational resources, or address a single task rather than holistic behavioral understanding. Motivated by these gaps, the proposed system aims to integrate lightweight supervised learning, multi-feature behavioral cues, and scalable privacy-preserving design to enable continuous, interpretable assessment of mental wellbeing in real-world public environments.

III. Proposed Methodology

A. System overview

Through the conducted literature survey, we discovered that most existing smart surveillance systems rely on single modal data and focus on only one type of feature, such as facial expressions or crowd analysis. Systems rarely used these models together, which causes the surveillance system to miss out on important details that could've been extracted if the system was trained to analyze all features from various modalities of data.

The main challenge, however, is that each of the existing models such as facial emotion detection, crowd analysis and environmental factors detection models are usually heavyweight themselves, and would require a lot of resources if they had to run in a system together at all times.

Hence, we propose a smart surveillance system framework that integrates crowd emotion analysis, environmental context detection and anomaly recognition with crowd behavior analysis to estimate public mental health trends in real time and adapts based on the context of use, making it a suitable replacement for most traditional surveillance systems.

B. System architecture

Fig. 1 is a system architecture diagram for the proposed smart surveillance system. It depicts the working of all the components in this system.

There are three main layers in this system – the Orchestrator Layer, Microservice Layer and the Ensemble Layer. The Orchestrator Layer is the key to our proposed system's functionality. Its main responsibility is to orchestrate the flow of data throughout the system. It's three main functions are:

- Deciding which services need to be used based on the input data. (Ex: For a mall surveillance system where faces are not clearly visible, it wouldn't be necessary to run the crowd emotion detection service.)
- Formatting input(s) and the outputs from the microservice layer to ensure the system is downstream and upstream compatible to a high degree.
- Combining all outputs in a way that makes graph formation and explanation easy.

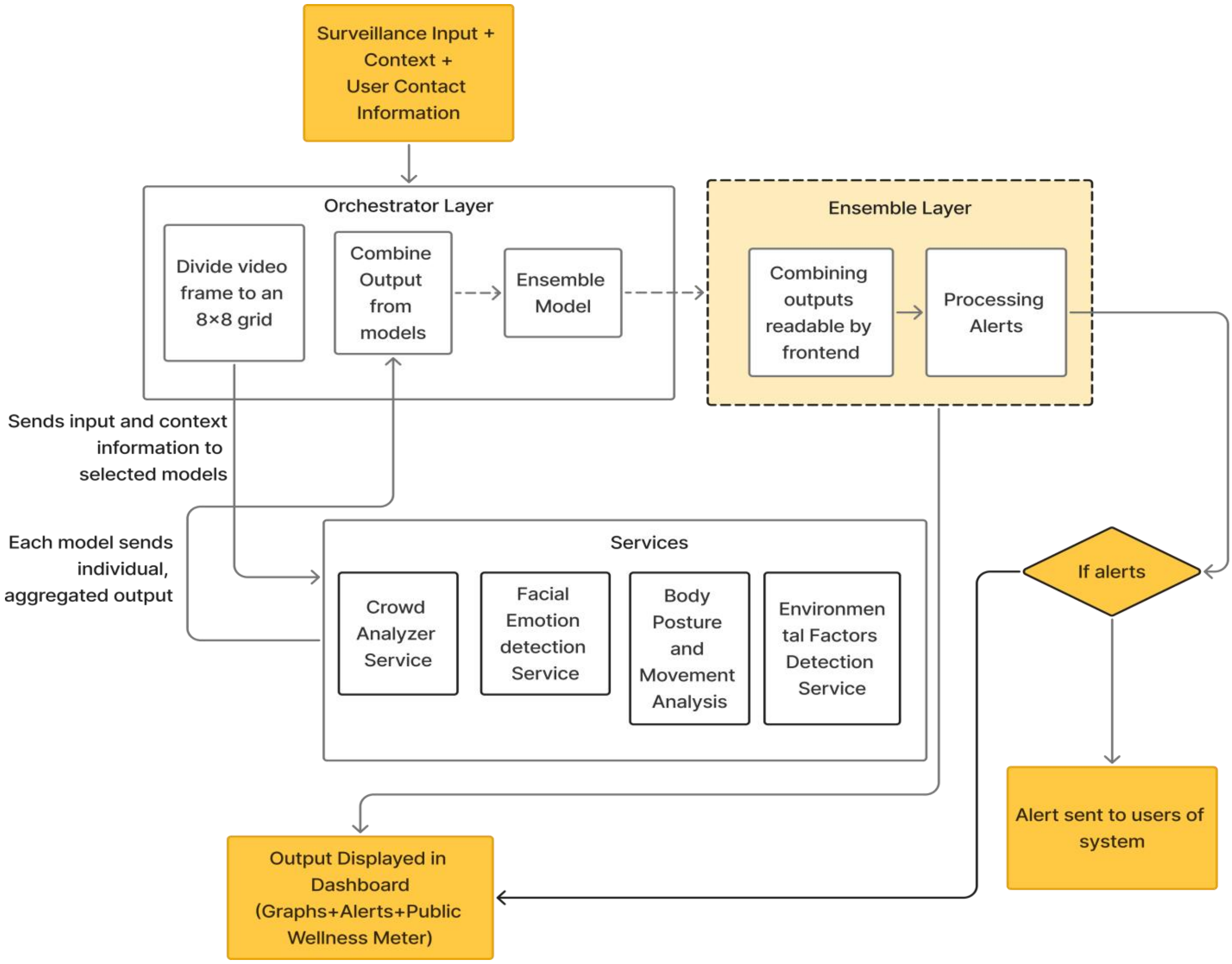

Fig. 1. System Architecture Diagram of the Proposed Methodology

Including a high priority “alerts” section to the data given to the ensemble layer, if required

The Microservice Layer consists of all the core services that are responsible for processing different features of the system’s input. This layer consists of 4 ML services – Crowd Facial Emotion Detection Service, Crowd Behavior Detection Service, Body Posture and Language Service and Environmental Factors Analysis Service. Each of these services follow a microservice architecture and contain their own database, which is accessed the system’s main database. They are explained in depth in the following sections. Fig. 2 is the basic pipeline that’s universally followed by the services in the microservice layer:

The Ensemble Layer is responsible for making the overall output interpretable and ready for use by the system’s user.

It contains utils responsible for explaining the JSON output and creating graphs in real-time. It is also responsible for identifying and sending alerts to the concerned user depending on the nature of the alert.

C. *Process pipeline*

The system takes video input from public surveillance sources. The Orchestrator Layer first analyzes the metadata and context of the video to decide which ML services to invoke. The selected frames are preprocessed and distributed to the microservices for parallel analysis. Each service outputs structured data (JSON format) containing extracted features such as emotion probabilities, crowd density, posture confidence scores, and environmental indicators. These are aggregated by the Ensemble Layer, which produces an interpretable summary along with graphical representations.

D. *Backend Service description*

1) *Crowd Facial Emotion Detection Service***:** This part of the system uses a trained ML model to look at people’s faces in a video through CCTV and finds out what emotions they are showing. It first detects every face in the crowd using a face finding tool in OpenCV. Then the emotion shown in each of the detected face is classified using the trained ML model, and the aggregated percentage of each emotion is calculated and returned in JSON format. First, it uses MTCNN to detect and crop the faces from the video frame. After that, each face is sent to a CNN model that was trained earlier using emotion datasets like FER2013. For every face, the model gives an emotion and a confidence score. These results are saved in simple JSON file. Later, all emotions from each frame are combined to understand the mood of the whole crowd, and this information is sent to the next part of the system for display and analysis.

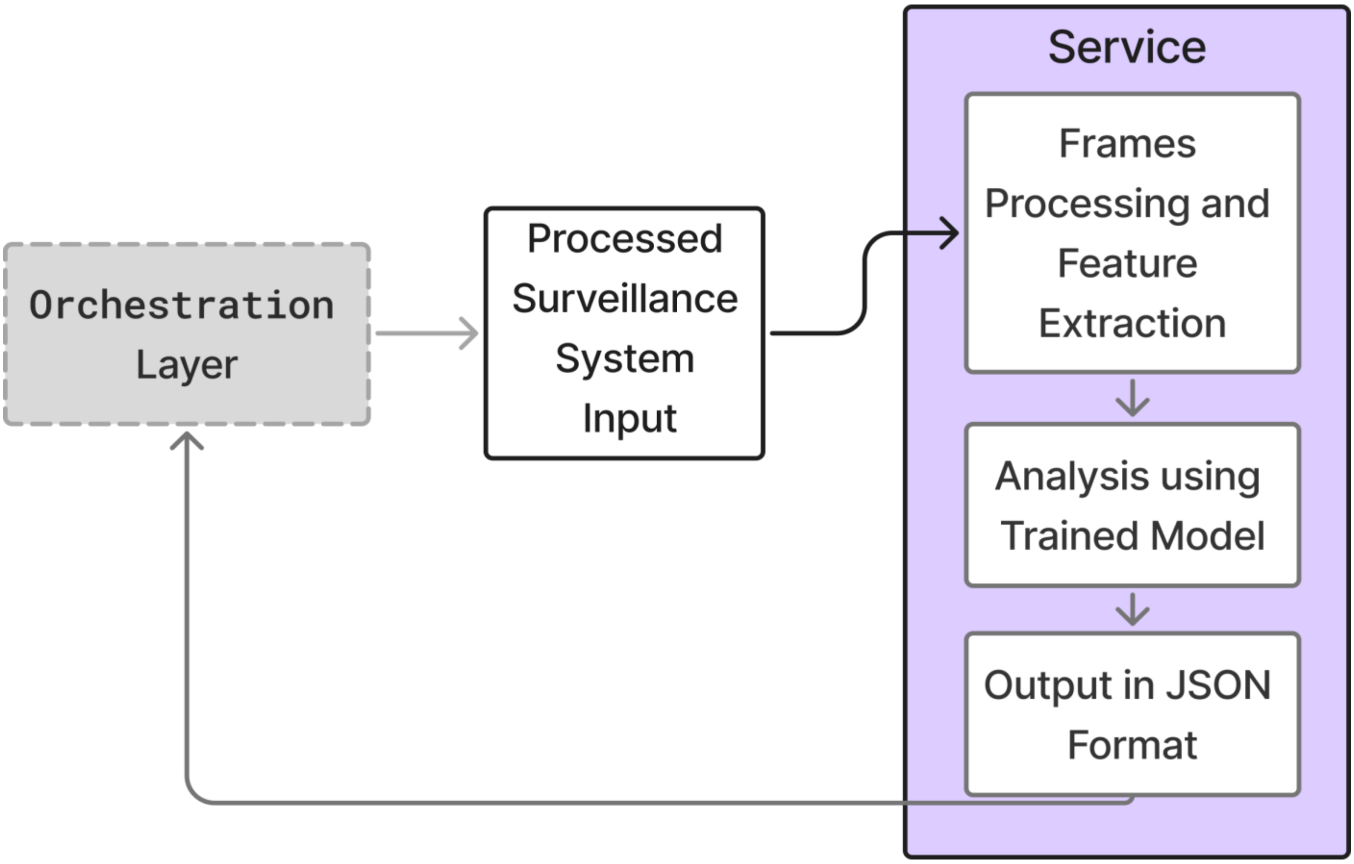

Fig. 2. Pipeline Diagram of Services in the service layer.

2) *Crowd Patterns and Behavior Service*: This service is responsible for providing insights on the crowd behavior, and to keep track of this behavior overtime in order to identify anomalies in the crowd. This service contains a model that is trained using various datasets such as ShanghaiTech, Avenue dataset, Mall dataset and MOT20 datasets. The MOT20 dataset is used to fine tune the Yolov8 model to be able to detect people more accurately, and the other mentioned datasets are used to train a model to classify different zones in a crowd image based on crowd movement (calm, chaotic, etc.), identify clusters and calculate the number of people and density of each zone. These analyzed details are saved in the service's database and are tracked in real time in order to identify any anomalies, and identify if the anomaly needs to be dealt with high priority.

3) *Body Posture and Language Service*: This This service comprises of two trained ML models – the body posture analysis model and body language analysis model. The body posture model was trained using a dataset created using images from COCO and MPII datasets, since the main purpose of this service is to classify different classes of body posture and body language. COCO and MPII contain an assortment of human images from both controlled and real-time environments, which is perfect to extract pose key-points and train a model that concerns human body analysis or any kind. It included five posture categories, namely, bent forward, crouching, slight lean, slouching, and upright, representing common sitting and standing postures. For the body language recognition model, four behavioral classes are considered, namely, normal, gesturing, aggressive, and defensive, to classify expressive human behavior. Each of the aforementioned classes contained around 100–150 labeled images, resulting in a dataset of about 800–1000 images. Diversity was introduced by including both indoor and outdoor samples in the dataset, and by including scenarios with different lightings and camera angles. All images were preprocessed using the YOLOv8 Pose model to extract pose key points.

4) *Environmental Factors Analysis Service*: Environmental factor analysis detection: This service is responsible for analyzing images to extract key environmental factors and to monitor these conditions. The images are preprocessed and augmented to improve model robustness, and a multi-head deep learning model is trained using a pre-trained ResNet-18 backbone. Throughout training, the model learns to identify weather (sunny, rainy, snowy, cloudy), lighting conditions (day, night, dim, bright), location types (indoor, outdoor), and cleanliness levels (clean, messy). After training, the service containing this trained model is able to take an input image, such as a video frame, and predict environmental conditions of classes as mentioned above, and return these predictions in a JSON format output, which is used by other system components downstream. These predictions can then be tracked over time to identify unusual or unexpected environmental patterns and decide whether any detected condition needs immediate attention.

E. Implementation details

The system backend is implemented using Python, with each ML model used in the system containerized as an independent microservice. The models leverage libraries such as Ultralytics, TensorFlow, MediaPipe, OpenCV, and Scikit-learn. These ML models are wrapped with FastAPI, and communicate with the Orchestrator and Ensemble Layer via REST APIs.

The frontend visualization is a Next.js module and uses Tailwind CSS for styling. This frontend displays real-time graphs generated with Matplotlib or Plotly through the backend API, and the explained, interpretable output produced by Gemini API is also displayed in a format that is easily consumable by the user.

F. Future enhancement

The following are the future plans for the smart surveillance for public mental health system:

- **Enhanced alerts system:** The future system can send concise and detailed alerts, sending the concerned frame(s) for further context as well.
- **Increased Accuracy for Backend Models:** The four current backend models can be fine-tuned further to increase accuracy of the entire system. More models such as object detection can also be added in order to extract more features and thereby provide more context and information based on the environment.
- **Extension of usability to other contexts**: Over time, the system is expected to help policy management, crowd management, event monitoring, hospital settings and more diverse contexts, and can effectively provide a better alternative solution to existing, traditional surveillance systems.

IV. Results

Each of the underlying models were evaluated for accuracy in performance with validation using a validation set from the dataset used to train the model and an unseen dataset. An exception to this is the crowd analysis model, which consists of a crowd counting component, which is evaluated using Mean Absolute Error (MAE) as the primary evaluation metrics.

A. Performance Evaluation

Performance for proposed solution was evaluated using performance metrics namely accuracy and MAE. However, to determine the overall accuracy of the entire system will be calculated by aggregating and finding the average accuracy of the models used in the backend of this system.

Table I shows the performance evaluation of each of the services used for this system. Accuracy for each model were calculated using the following formula:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (1)$$

Here, TP – True Positives, TN – True Negatives, FP – False Positives and FN – False Negatives.

These accuracy values demonstrate that the models are able to classify emotions, high-risk crowd patterns and anomalies to a considerable level of accuracy, and work well to provide in-depth and accurate overall analysis of the surveillance footage.

B. Qualitative Results

The system produces the following observed outcomes:

- The dashboard of the system accurately visualized trends observed in the surveillance footage overtime.

TABLE I. PERFORMANCE EVALUATION OF BACKEND MODELS

Models	Performance Details		
	Training Dataset	***Evaluation Dataset***	***Accuracy***
Crowd Facial Emotion Detection	FER-2013 (Augmented)	FER-2013 (Augmented)	79%
Body Posture and Language Analysis	COCO+MPII	COCO+MPII	85%
Crowd Behavior Analysis	ShanghaiTech+ UCSD	ShanghaiTech+ UCSD	83%
Environmental Factor Analysis	Weather Image Recognition	Weather Image Recognition	85.7%

- Well-being meter fluctuates according to the overall mental state of the crowd.
- The system successfully sends and displays alert messages in cases of distress/anomalous behavior.

The well-being meter fluctuates based on the Public Well-being Index (PWI) that's calculated as follows (This index is used to calculate value for the stress trend when facial emotions are detected, otherwise, the system uses outputs from other working models to update graphs in the dashboard):

$$PWI = 1 - (wf * Pf + wa * Pa + ws * Ps) \quad (2)$$

Where, Pf – proportion of fear expressions, Pa – proportion of angry expressions, Ps – proportion of sad expressions, wf, wa and ws – weight factors for fear, anger and sadness. A PWI score close to 1 indicates a crowd with positive crowd well-being, while a score close to 0 indicates distress.

C. System Level Evaluation

The complete application was evaluated using real-world surveillance footages and videos obtained from various sources, including YouTube and Human Events Dataset, and each of these videos represented various public environments such as classrooms, roads, cross-roads and malls, and each were selected to simulate realistic human behaviors with varying emotional and crowd states.

The videos were processed through the deployed application interface, which includes components for real-time analysis of surveillance footage represented with stress and people count graphs, and alert generation for anomalies or distress indicators. The system level evaluation criteria focused on two factors – Response Time and Detection Accuracy, which is defined by the correctness of identified emotions, anomalies and other classes focused by the system and is calculated using (1).

Fig. 3 contains frames from the videos used for evaluation of the system. As the figure shows, the videos used were from different contexts such as a mall, railway station and a classroom.

The proposed system achieved consistent performance across all the test cases, with an overall detection accuracy of 84.51% and an average response time of 6 seconds during system start-up and less than 2 seconds after system start-up, causing around 7-8 seconds of delay in updates. In cases of extreme occlusion, facial emotions remain undetected and crowd counting accuracy drops.

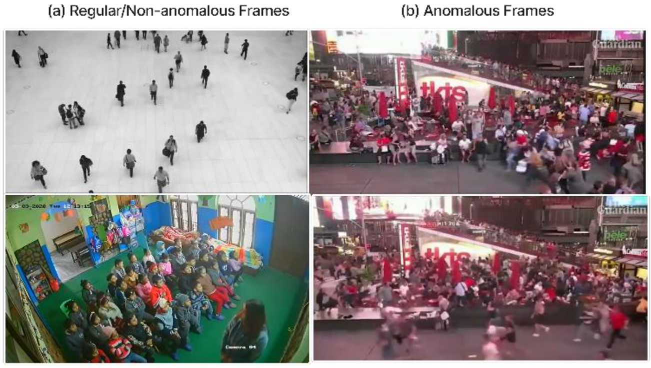

Fig. 3. Examples of regular (a) and anomalous (b) frames from videos used for evaulation of the model.

However, the system successfully maintained stable operation without significant performance degradation, especially since when one feature, such as facial emotion is not detectable, other features still aid analysis of the surveillance footage

These experimental results show that the proposed smart surveillance system effectively integrates emotion recognition and crowd behavior analysis for monitoring public mental well-being. Despite environmental and context variations, the system maintains a consistent and highly accurate performance with real-time responsiveness. Further improvements can focus on enhancing performance under extreme occlusions and distant cameras, leveraging data from more modalities such as audio and textual data, and with smarter decision making leveraging advanced, emerging technology such as agentic AI or Reinforcement Learning.

Fig. 4, 5 and 6 contain the outputs of the system for a calm classroom and a panicked crowd, respectively. These outputs contain a Stress Trend, People Count Trend, Public Wellness Meter and Environmental Trends.

Table. II provides a summary of the system level evaluation results. Detection Accuracy is obtained by finding the aggregated accuracy of detection under every class that the system focuses on.

Fig. 4. Output screenshots from the system for a panicked crowd.

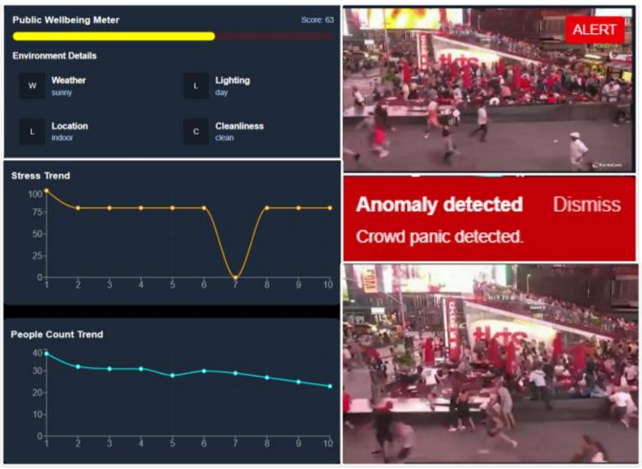

Fig. 5. Output screenshots from the system for a panicked crowd.

Fig. 6. Output screenshots from the system for a Jail Fight footage

TABLE II. SYSTEM LEVEL PERFORMANCE EVALUATION

Test Scenario	Performance Details		
	Detection Accuracy (%)	***Average Response Time (s)***	***Observations***
Classroom	94%	~ 2	Accurately detects posture anomalies when present.
New York Station	100%	~ 1	Facial emotion module failed due to severe occlusion, but the output was correct due to the other model's work.
Mall	76.92	~ 3	System detected 3 false anomalies in 13 seconds. Possibly due to distant camera position
Cross-road	70.86	~ 3	4 False positives. Reason same as above.
Jail Fight	80.75	~ 3	Minor fluctuation in crowd count and 4 out of 20 false negatives were observed.

ACKNOWLEDGMENT

We would like to acknowledge and give our sincere thanks to our project supervisor, Ms. Lavanya Santhosh, Assistant Professor, Dept. of CSE, Dr. AIT, for continuous and invaluable guidance and constructive feedback throughout the development of this work.

We would also like to acknowledge the creators and contributors of the datasets used for training the models used in this system. In particular, we thank the developers of the dataset FER 2013, COCO, MPII, ShanghaiTech, UCSD and the weather image recognition datasets. These datasets played an important role in building and validating the backend components of the system.

We would also like to extend our gratitude to the content creators on YouTube whose publicly available videos were used for testing and evaluating the complete proposed system.

Finally, we are immensely grateful to anyone who provided support, directly or indirectly, during the development and completion of this project.

REFERENCES

[1] Y. Zhang, H. Lyu, J. Luo, and F. Chen, "Monitoring depression trends on Twitter during the COVID-19 pandemic: Observational study," JMIR Infodemiology, vol. 1, no. 1, p. e26769, 2021. doi: 10.2196/26769.

[2] D. García-Retuerta et al., "An efficient management platform for developing smart cities: Solution for real-time and future crowd detection," Electronics, vol. 10, no. 18, p. 2295, 2021. doi: 10.3390/electronics10182295.

[3] R. M. Favaretto, P. Knob, S. R. Musse, F. Vilanova, and A. B. Costa, "Detecting personality and emotion traits in crowds from video sequences," arXiv preprint, 2021. [Online]. Available: https://arxiv.org/abs/2102.05856

[4] J. J. Titla-Tlatelpa, O. Camacho-Nieto, A. P. López-Monroy, and M. Montes-y-Gómez, "A profile-based sentiment-aware approach for depression detection on social media," EPJ Data Science, vol. 10, no. 1, p. 5, 2021. doi: 10.1140/epjds/s13688-021-00260-0.

[5] N. Wang et al., "Learning models for suicide prediction from social media posts," in Proc. 7th Workshop on Computational Linguistics and Clinical Psychology (CLPsych), 2021. doi: 10.18653/v1/2021.clpsych-1.13.

[6] L. Ren, X. Kang, D. Xu, G. Chen, and J. Cao, "Depression detection on Reddit with an emotion-based attention network," JMIR Medical Informatics, vol. 9, no. 9, p. e26892, 2021. doi: 10.2196/26892.

[7] R. Safa, A. Bayat, and A. Bagheri, "Automatic detection of depression symptoms in Twitter," J. Biomedical Informatics, vol. 120, p. 103846, 2021. doi: 10.1016/j.jbi.2021.103846.

[8] W. Xie, L. Liang, Y. Lu, H. Luo, and X. Liu, "Deep 3D-CNN for depression diagnosis with facial video recording of self-rating depression scale questionnaire," arXiv preprint, 2021. [Online]. Available: https://arxiv.org/abs/2108.07205

[9] A. Gamoran, H. Zhang, M. Chen, and P. Resnik, "Using psychologically-informed priors for suicide risk prediction," in Proc. CLPsych 2021. doi: 10.18653/v1/2021.clpsych-1.14.

[10] S. Renjith et al., "An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms," arXiv preprint, 2021. [Online]. Available: https://arxiv.org/abs/2112.10827

[11] W. Guo et al., "Deep neural networks for depression recognition based on 2D and 3D facial expressions under emotional stimulus tasks," Frontiers in Neuroscience, vol. 15, 609760, 2021. doi: 10.3389/fnins.2021.609760.

[12] Jin N, Ye R, Li P., "Diagnosis of depression based on facial multimodal data: E-DAIC study," IEEE Access / PubMed, 2021. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/39935533/

[13] A. Hassan and S. M. Islam, "Predicting depression risk from facial video-derived heart rate estimates," Revue d'Intelligence Artificielle, vol. 37, no. 4, pp. 1003–1011, 2021. doi: 10.18280/ria.370421.

[14] T. Alanzi et al., "AI-powered mental health virtual assistants acceptance: An empirical study on influence factors," Cureus, 2022. [Online]. Available: https://www.cureus.com/

[15] C. F. Hinton et al., "Developing a public health-tracking system for follow-up of newborn screening metabolic conditions," Genetics in Medicine, 2013. doi: 10.1038/gim.2013.177.

[16] M. F. Kamgang, S. Lafleur, M. M. Monti, and A. Salame-Alfie, "Temporal variation in indoor radon concentrations using environmental public health tracking data," Health Physics, 2023. [Online]. Available: https://journals.lww.com/health-physics/fulltext/2023/04000/

[17] C. Fontes, E. Hohma, C. C. Corrigan, and C. Lütge, "AI-powered public surveillance systems: Why we (might) need them," Technology in Society, 2022. doi: 10.1016/j.techsoc.2022.101993.

[18] M. V. Heinz, S. Bhattacharya, and N. C. Jacobson, "Testing domain knowledge and risk of bias of a large-scale AI model in mental health," Digital Health, 2023. doi: 10.1177/20552076231170499.

[19] D. Solans Noguero, D. Ramirez-Cifuentes, E. A. Rissola, and A. Freire, “Gender bias in AI for anorexia assessment on social media,” J. Med. Internet Res., vol. 25, p. e45184, 2023. doi: 10.2196/45184.

[20] P. Daily, N. P. Hitt, D. R. Smith, and C. D. Snyder, “Experimental and environmental factors affect spurious detection of ecological thresholds,” Ecological Applications, vol. 22, no. 5, pp. 1772–1782, 2012. doi: 10.1890/11-0516.1.

[21] J. Reubens et al., “Environmental factors influence the detection probability in acoustic telemetry,” Hydrobiologia, vol. 845, pp. 95–109, 2017. doi: 10.1007/s10750-017-3478-7.

[22] M. Quiroz, R. Patiño, J. Diaz-Amado, and Y. Cardinale, “Group emotion detection based on social robot perception,” Sensors, vol. 22, no. 10, p. 3749, 2022. doi: 10.3390/s22103749.

[23] M. A. Mansoor and K. Ansari, “Advancing early detection of major depressive disorder: A prospective observational study,” medRxiv, 2024. [Online]. Available: https://www.medrxiv.org/content/10.1101/2024.08.13.24311933

[24] A. Serra, F. Carrara, M. Tesconi, and F. Falchi, “The emotion of the crowd: Learning sentiments via cross-modal distillation,” arXiv preprint, 2023. [Online]. Available: https://arxiv.org/abs/2304.14942

[25] S. Bhardwaj et al., “Machine learning-based crowd behaviour analysis and forecasting,” ResearchGate, 2023. [Online].

[26] M. Islam, A. S. Dukyil, S. Alyahya, and S. Habib, “An IoT enabled anomaly detection system for smart city surveillance,” Sensors, vol. 23, no. 4, p. 2358, 2023. doi: 10.3390/s23042358.

[27] S. Shah et al., “An ensemble-learning-based technique for bimodal sentiment analysis,” Big Data and Cognitive Computing, vol. 7, no. 2, p. 85, 2023. doi: 10.3390/bdcc7020085.

[28] S. Singh et al., “Real-time surveillance system for women’s safety and crime detection,” in Proc. ICRT, pp. 1–6, 2023. doi: 10.1109/ICRT57042.2023.10146679.

[29] S. N. Bhagat, D. B. Hanchate, and S. S. Bere, “Big data enabled real-time crowd surveillance and threat detection,” IJCRT, vol. 12, no. 5, pp. 55–62, 2024.

[30] N. Alruwais and M. Zakariah, “Student-engagement detection in classroom using ML,” Electronics, vol. 12, no. 3, p. 731, 2023. doi: 10.3390/electronics12030731.

[31] R. Venkatesan et al., “Human emotion detection using DeepFace,” Engineering Proceedings, vol. 59, no. 1, p. 37, 2023. doi: 10.3390/engproc2023059037.

[32]] D. Godse et al., “Automated video and audio-based stress detection,” IJRITCC, vol. 11, no. 11S, pp. 487–495, 2023. doi: 10.17762/ijritcc.v11i11s.8178.

[33] D. Turcian and V. Stoicu-Tivadar, “Real-time detection of emotions based on facial expression,” in Telehealth Ecosystems in Practice, IOS Press, 2023, pp. 272–278. doi: 10.3233/SHTI230795.

[34] H. Yao, X. Yang, D. Chen, Z. Wang, and Y. Tian, “Facial expression recognition via channel–spatial attention transformer,” Sensors, vol. 23, no. 15, p. 6799, 2023. doi: 10.3390/s23156799.

[35] F. Abdullah and A. Jalal, “Semantic segmentation-based crowd tracking and anomaly detection,” Arabian J. Sci. Eng., vol. 48, pp. 2173–2190, 2023. doi: 10.1007/s13369-022-07092-x.

[36] E. M. C. L. Ekanayake, Y. Lei, and C. Li, “Crowd density and anomaly detection using MCMS-BCNN-Attention,” Applied Sciences, vol. 13, no. 1, p. 248, 2023. doi: 10.3390/app13010248.

[37] T. Alafif et al., “Hybrid classifiers for spatio-temporal abnormal behavior detection in Hajj crowds,” Electronics, vol. 12, no. 5, p. 1165, 2023. doi: 10.3390/electronics12051165.

[38] S. Goyal et al., “MindLift: AI-powered mental health assessment for students,” Neuroscience Informatics, p. 100208, 2025. doi: 10.1016/j.neuri.2025.100208.

[39] S. Vijayalakshmi, D. A. Sandhya, and P. N. Deve, “AI-powered holistic mental health monitoring,” in Proc. ICDSAAI, 2025. [Online].

[40] M. Qaraqe et al., “PublicVision: A secure smart surveillance system for crowd behavior recognition,” IEEE Access, vol. 12, pp. 23456–23468, 2024. doi: 10.1109/ACCESS.2024.10438436.

[41] C. I.-O. Gabriel, R. Kumar, and K. Prasad, “Enhancing deep boosting for panic disorder detection,” IEEE Trans. Artif. Intell., 2025. doi: 10.1109/TAAI.2025.11134432.

[42] Samsinas et al., “Early intervention strategies for self-harm prevention using AI-driven behavior tracking,” Int. J. Res. Counseling, 2025. [Online].

[43] N. Irfan et al., “AI predictive analytics with naturopathic interventions for maternal mental health,” Scientific Reports, vol. 15, 12345, 2025. doi: 10.1038/s41598-025-07885-8.

[44] J. Zulu and P. Munyenyembe, “AI-driven mental health support system for teenagers,” IJASRE, vol. 11, no. 4, pp. 15–28, 2025.

[45] N.-D. Mai, H.-T. Nguyen, and W.-Y. Chung, “Deep learning wearable ear-EEG emotion recognition,” IEEE Sensors Journal, vol. 24, no. 5, pp. 6789–6801, 2024. doi: 10.1109/JSEN.2024.10456557.

[46] A. Alzahrani, “Bioinspired image processing enabled FER using hybrid deep learning,” IEEE Access, vol. 12, pp. 45678–45691, 2024. doi: 10.1109/ACCESS.2024.10415427.

[47] M. Jallal et al., “Facial emotion recognition using Xception architecture,” in Proc. ICCI, 2024, pp. 1–6. doi: 10.1109/ICCI.2024.11073104.

[48] S. Arya, Anju, and N. A. Ramli, “Predicting stress level of students using ML & ANN,” Indian Journal of Engineering, vol. 21, pp. 45–56, 2024.

[49] S. R. Alotaibi et al., “Explainable AI with deep learning for crowd density estimation,” IEEE Access, vol. 13, pp. 56789–56801, 2025. doi: 10.1109/ACCESS.2025.10843209.

[50] H. Zhang, H. Liu, and C. Kim, “Semantic and instance segmentation for coastal urban perception,” Sustainability, vol. 16, no. 2, p. 833, 2024. doi: 10.3390/su16020833.

[51] B. Y. Yashas, L. Santhosh, K. N. Asha, V. Potdar, R. Shamanth, and G. K. Teja, “Gait based behaviometric identification using CASIA-B dataset and gait energy images,” in 2025 International Conference on Knowledge Engineering and Communication Systems (ICKECS), Chickballapur, India, pp. 1–5, 2025. doi: 10.1109/ICKECS65700.2025.11034813.

[52] P. S. Calabe, R. Prabha, and V. Potdar, “Cardiac ailment recognition using ML techniques in E-healthcare,” *World Journal of Advanced Research and Reviews*, vol. 17, no. 1, pp. 302–307, 2023, doi: 10.30574/wjarr.2023.17.1.0010.

[53] V. Potdar, L. Santhosh, and C. Y. Yashu Raj Gowda, “A survey on stroke disease classification and prediction using machine learning algorithms,”*Int. J. Eng. Res. Technol. (IJERT)*, vol. 10, no. 8, Aug. 2021.

[54] V. Potdar, L. Santhosh, and L. Jadhav, “Coronary heart disease prediction using machine learning,” *J. Emerging Technol. Innov. Res.*, vol. 9, no. 12, pp. e390–e396, 2022.