Chapter 1

Introduction

Machine Learning is a branch of Artificial Intelligence that is being used in various fields since it has a vast potential due to it's 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. Mental health is one of the various domains where Machine Learning has found it's use in.

AI Driven Public Mental Health Tracker aims to create a preventative solution for a group of people, rather than focusing on an individual. With the help of emerging Machine Learning technologies, this project aims to detect and accurately analyse the aggregated mental health of the public, and provide the possible cause of this mental state.

1.1 Existing System

There has been a rise in machine learning solutions to diagnose or detect mental health issues at early stages. Most of these solutions analyze large amounts of patient data to identify patterns, and create personalized treatment plans according to a patient's medical situation. There are also models that detect and recognize emotions and personality traits from videos, and solutions that leverage those models and methodologies to detect violence in crowd videos using CCTV sourced footage and the integration of online social media. Most of these solutions rely heavily on facial expressions, and focus less on other indicators such as body posture, environmental factors, crowd movement speed, fidgeting, etc.

1.2 Proposed System

Focusing on factors such as body posture, crowd movement and patterns, environmental factors, fidgeting, etc. can not only help with detecting the overall mental state of a group of individuals, but also help identify causes of positive and negative mental states of a group. This could also help with more accurate mental health forecasting, thereby helping more people before they reach a crisis. Most solutions also mainly rely on the use of transformer models, but combining various transformer models with the use of reinforcement learning, explainable AI, and LSTMs can help improve accuracy and efficiency of the system.

1.2.1 Scope of the Project

Public areas can include schools, workplace, restaurants, bars, concerts, neighbourhoods, etc. We plan to work upwards from public areas with more organized crowds with less background noise, and then explore the application of this system in bigger crowds and more unpredictable environments such as concerts, neighbourhoods, etc. Private areas such as houses are excluded.

The signs that the system can use for detection include:

- Facial Expressions: Emotion detection.
- Body Language: Detection of slouching, fidgeting, disengagement, etc.
- Crowd Dynamics: Sudden movement changes, restlessness, etc.
- Environmental factors: Noise levels, congestion, weather, etc.
- Anonymized Verbal Analysis

And the system will not include:

- Individual diagnosis, since the system is only used to identity trends of mental health, not to be used as a medical tool)
- Collection of personally identifiable information (PII) storage.

The system is expected to have a real-time dashboard showing public mental health trends, causes of these trends, predict areas that need more surveillance, and explain the analysis that the AI model performed in order to reach this conclusion. The system is also expected to have limitations; The system will not be 100% accurate and may struggle with occluded faces or extreme crowd density.

Challenges that need to be addressed for this system are accuracy of detection, real-time processing, and privacy.

1.2.2 Problem Statement of the Project

According to the WHO, in 2019, 970 million people globally were living with a mental disorder, with anxiety and depression being the most common. Mental health conditions can cause difficulties in all aspects of a person’s life, and impacts people around them too. Most of these issues result from or lead to problems in work and school. Early detection of such problems in public areas, and identification of the factors that lead to adverse mental health effects can help reduce the number of people affected by these disorders significantly.

Chapter 2

LITERATURE SURVEY

There have been various applications of AI and Computer Vision in the domain of mental health detection and diagnosis. These systems use various concepts, parameters and methods that determine the effectiveness and accuracy of the detected mental health state of an individual or a group.

This literature survey (Table 1) is an explorative view on the past advances in the use of Computer Vision, AI and ML in the domain of mental health and the methods and concepts used in these systems, their limitations and how it could be integrated in the novel solution presented in this article.

The inclusion criteria are that the article must be written in English, after the year of 2021 and published in a reputed literature database. It must also be closely related to the topic of “AI-driven Public Mental Health Tracker”, that is, it will consider literatures that are about the application of AI in the mental health domain.

Exclusion criteria include literature written before the year 2021, and not written in English. It also includes literature that proposes a method using AI and/or computer vision for individual mental health diagnosis rather than gauging the mental health of a group of people, and systems or methods that solely focus on the analysis of textual data.

Table 1: Table with Literature survey summaries.

Sl. No.	Authors, Year	Paper Title and Publisher	Assumptions	Summary
1.	Rodolfo Migon Favaretto, Paulo Knob, Soraia Raupp Musse, Felipe Vilanova, Angelo	Detecting Personality and Emotion Traits in Crowds from Video Sequences, Arxiv [1].	This paper connects personality and emotions of a person using the following analogy: “Personality is to emotion as climate	**Objective:** Using crowd analysis to detect personality and crowd emotions using the OCEAN dimensions through video sequences. **Methods used:** Obtaining people trajectories using a tracker and a homography transformation process, mapping crowd features with OCEAN dimensions with the use of NEO PI-R followed by emotion detection using information on the personality. Refer Fig 4.1.

	Brandelli Costa (2021)		is to weather". It hence assumes that a person's emotion is completely dependent on their personality	**Limitations:** This technology was not validated with real life experiments due to the challenges that comes with it, so the performance was evaluated using available information regarding OCEAN. Emotion detection is also completely dependent on personality detection, which is limited to 5 personality traits, which does not fully describe human nature and could cause bias and inaccuracy in mental health detection. **Conclusion:** Using this method alone may cause inaccuracies, however the system could be improved by adding context-based analysis, using information about the environment, making it so that the analysis of emotion is not only based on a person's personality can improve accuracy of the system.
2.	David Gimeno-Gomez, Ana-Maria Bucur, Adrian Cosma, Carlos-David, Martinez Hinarejos	Multi-Modal Depression Detection in Videos from Non-Verbal Cues, Arxiv [2].	It is assumed that the mental health information in the video doesn't depend on time, i.e., any random sample taken from any part of the video will lead to the same conclusion about the person's mental health.	**Objective:** This research explores the possibility of detecting depression from "in-the-wild" videos, with the main challenge being the significant amount of noise that will be present in such videos. It aims to provide a diagnostic method to diagnose individuals through real-life videos. **Methods used:** Any text-based data is not used by this model to avoid bias due to conversational topics. Using window sampling, random samples from the videos are considered, and modalities are extracted and encoded from these samples. Then, relevant features and nonverbal cues are extracted. They use Face Alignment Network model and MediaPipe toolkit to extract face and body landmarks respectively, and ETH-XGaze gaze estimator and InstBlink models to identify gaze and blinking patterns respectively. Refer Fig. 4.2.

				Limitations: This experiment highlights some very useful methods that can be used to detect depression and other mental health issues for the given individual. It makes use of non-verbal cues and facial expressions to determine its output, but does extract and make use of any information regarding the environment that could influence the results, and focuses purely on the subject, and removes everything else as noise. **Conclusion:** This paper highlights an efficient method to extract features and nonverbal cues, and this method can be used with additional information about the environment and the mental health information of others in a given area can help provide an accurate result for the aggregated mental health status of people in the given area.
3.	Ghulam Gilanie, Mahmood ul Hassan, Mutyyba Asghar, Ali Mustafa Qamar, Hafeez Ullah, Rehan Ullah Khan, Nida , Irfan	An Automated and Real-time Approach of Depression Detection from Facial Micro-expressions	The study assumes that specific facial micro-expressions, particularly those associated with sadness, disgust, and contempt, are indicative of depressive states. It posits that these micro-expressions can be effectively	**Objective**: To develop an automated, real-time system capable of detecting depression by analysing facial micro-expressions captured in video sequences. **Methods**: The system employs the Facial Action Coding System (FACS) to extract Action Units (AUs) corresponding to micro-expressions linked to depression. These features are then analysed using Support Vector Machine (SVM) and Convolutional Neural Network (CNN) models to classify individuals as depressed or non-depressed. **Limitations**: The study's dataset is limited to a specific demographic, which may affect the generalizability of the

	Ullah Khan (2022)	(Tech Science Press) [3].	captured and analysed using video sequences to detect depression.	results. Additionally, the reliance on facial micro-expressions alone may not capture the full spectrum of depressive symptoms, potentially impacting the system's accuracy across diverse populations. **Conclusion**: This study shows that by carefully analysing facial expressions, it's possible to detect signs of depression with high accuracy. However, since it only looks at facial indications and was tested on a specific group of people, more research is needed to ensure it works well for everyone.
4.	Jetli Chung, Jason Teo (2022)	Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges (Wiley) [4].	Assumes that machine learning techniques can be systematically categorized and applied to predict various mental health conditions, and that understanding these applications can highlight challenges and future research directions.	**Objective:** To provide a comprehensive review of machine learning approaches in predicting mental health issues, categorizing them by techniques and targeted disorders. **Methods used:** Conducted a systematic literature review adhering to PRISMA guidelines, analysing 30 selected studies focusing on different machine learning methods applied to mental health prediction. **Limitations:** Identified challenges include data scarcity, lack of standardized datasets, ethical concerns regarding data privacy, and the need for explainable AI models. **Conclusion:** Emphasizes the potential of machine learning in mental health prediction while highlighting the necessity for high-quality data, ethical considerations, and the development of interpretable models for clinical application.

5	Tabil Ahammed, Sudipto Ghosh, Silva Deena J (2022)	Real-time based Violence Detection from CCTV Camera using Machine Learning Method (ScienceDirect) [5].	The study assumes that violent behaviour in crowds can be detected by analysing facial expressions and body movements captured through CCTV footage. It posits that specific patterns in facial expressions and body language are indicative of violent actions.	**Objective:** To develop a real-time system capable of detecting violence in crowds by analysing facial expressions and body movements using CCTV footage. **Methods Used:** The researchers employed machine learning techniques, including Convolutional Neural Networks (CNNs), to process video data. The system analyses facial expressions and body movements to identify violent behaviour. **Results:** The proposed method demonstrated high accuracy in detecting violence in real-time scenarios, showcasing its potential for practical surveillance applications. **Limitations:** The study acknowledges challenges such as varying lighting conditions, occlusions, and the need for extensive training data to improve model robustness. **Conclusion:** This study shows that by analysing facial expressions and body movements from CCTV footage, it's possible to detect violent behaviours in real-time. While the system performs well, it faces challenges like different lighting conditions and the need for more training data to handle various scenarios effectively.
6.	Kyunguen Min,Migyeong kang,Daeun	Detecting depression on video logs	The study assumes that individuals with depression	**Objective**: This study aims to develop an effective depression detection model that utilizes both audio and visual features By analysing depression-related and non-

	lee,Eunil park,Jinyoung Han(2023)	using audiovisual features [6] .	exhibit detectable differences in their audio and visual behaviour in their vlogs, and also the vloggers who self identify as depressed and discuss their symptoms in videos and reliable indicators of depression. It also assumes that the selected audio and visuals are sufficient between depressed and non-depressed individuals.	depression-related vlogs, the model seeks to identify key indicators—such as voice intensity and facial expressions—that distinguish individuals experiencing depression. **Methods**: They collect the data from YouTube sources like depression vlogs and non-depression vlogs and they use YouTube as a tool and in depression vlogs must show that the person is speaking directly facing to the camera about their current depression state and non-depression vlogs are distinguished by not changing the pattern day by day and they must follow the same pattern and it has the feature extension like audio extension, visual extension like their facial expression etc and they have used algorithms like extreme gradient boosting which is one of the most popular algorithm. Harmonics-to-Noise Ratio (HNR).The HNR quantifies the amount of additive noise in the voice signal. The prior work has shown that the HNR has a negative relation with depression Jitter & Shimmer. The Jitter, a physiological-related feature, has a positive relationship with anxiety Jitter value can be obtained from the patients who suffer from severe depression with high suicide risk. **Limitations**: This model only helps in detecting depression either by its facial expression or audio using audio visuals. instead of that the model should be developed in such a way that it should identify the person's emotions by its voice only. **Conclusion**. We can include some additional features like multimodal emotion detection where it's has the features to integrate and provide the result like heart rate, skin temperature to provide a holistic assessment of an individual's mental health and also Incorporate

				environmental cues background noises, lighting, setting to provide better context when assessing emotional states. Track shifts in speech tone, language choice etc.
7.	Mohamed Baklola, Mohamed Terra (2024)	AI-Powered predictive analysis for crowd health management .(ScienceDirect) [7]	This article assumes that integrating AI models with live data streams can help public health officials to proactively identify and mitigate potential hazards such as heat exhaustion or respiratory infections. This system attempts to access the overall health of a group of individuals.	**Objective**: To study a powerful AI tool that helps manage crowd health, focusing on health risks at large gatherings through real-time analysis of environmental factors and crowd behaviour. **Methods**: An AI predictive model integrated with a live data stream is created to foresee health risks in crowds. Zhang et al. (2023) for instance, created a reinforcement learning-based model that allows agents to gather 3D environmental data, improving crowd dynamics. This system could also make the use of IoT devices to improve the precision of predictions and enables timely interventions to manage health risks at large events. **Limitations**: This model considers very few factors into consideration, since it focuses on the general health of the public, and not mental health specifically. They focus only on environmental factors and crowd behaviour, and ignore body languages, facial expressions and other factors that could provide additional information. **Conclusion:** A similar predictive model to this can be used to predict and detect crowd mental health crisis and its causes, however to achieve this, it is necessary to create a model that considers more factors, thereby providing a precise result.
8.	Elizabeth B. Varghese and	Towards the cognitive and	This study assumes that if	**Objective**: The main goal of this study is to dive deeper into the science of mentality of the masses in order to be able to monitor and manage these masses and possibly

	Sabu M. Thampi (2021)	psychological perspectives of crowd behaviour: a vision-based analysis (Taylor and Francis) [8]	we understand the mechanism, motives and mentality of a group, then it is easy to manage, control and monitor the crowd. It also assumes the presence of psychological theories in literature that largely help the determination of crowd behaviours, and attempts to understand and analyse these concepts and systems.	mitigate non-adaptive crowd behaviours, which have been a great concern for authorities in the past. **Methods**: Various psychological theories of crowd behaviour and approaches of analysing crowd behaviour is explored in this study. Some of the proposed systems are crowd simulation system, visual analysis of crowd video sequences by considering relevant factors in detecting crowd behaviour from video data such as emotions, with the help of an arousal-valence emotional model created from motion patterns of the crowd, and thereby identify the mood of the crowd. Other methods proposed include the usage of OCEAN personality model and the PEN (Psychoticism Extraversion Neuroticism) model. **Limitations**: The analysis models suggested in this article are all viewed individually, and it doesn't explore the combination of multiple means of analysis of crowd behaviour, such as combining the OCEAN personality model with a machine learning model. **Conclusion**: The psychological models proposed in this article can be used along with other approaches to analyse crowd behaviour such as Support Vector Machines (SVM) for multiclass classifications to create a powerful machine learning or deep learning model that produces accurate and efficient results.

9.	Fumiyasu Makinoshima & Yusuke Oishi (2022)	Crowd flow forecasting via agent-based simulations with sequential latent parameter estimation from aggregate observation(nature.com) [9]	The article is based on the assumption that even when individual-level data (such as exact trajectories of people in a crowd) is not available, it is still possible to forecast crowd movement patterns by using aggregate observations like overall density maps. It presumes that individual behaviors in a crowd can be	**Objectives**: The main objective of the study is to improve real-time crowd flow forecasting by combining ABMs with dynamic parameter estimation from limited observational data. The authors aim to bridge the gap between highly detailed simulation models, which typically require exact input data, and real-world applications where only aggregate or partial data is available, such as camera-based crowd density estimates. Their goal is not just to predict future positions of the crowd but also to adaptively tune the underlying behavioral assumptions in the model as more observational data is received over time. **Methods**: To achieve this, the researchers employ a methodology that integrates agent-based crowd simulation with a sequential estimation framework using particle filters. These filters allow the system to update its beliefs about both the crowd's physical state and the hidden parameters that drive agent behavior. The model is tested through numerical experiments simulating large-scale evacuation scenarios, involving thousands of agents, to demonstrate its predictive capabilities under realistic, data-limited conditions. **Limitations**: One of the major constraints is that the model is evaluated only on synthetic data generated by simulations, rather than on real-world observational datasets from actual crowds. This raises questions about how well the method would perform in noisy, unpredictable real environments. Moreover, while particle filters are powerful, they are computationally intensive, and the scalability of this approach to even larger, more complex urban settings is not fully explored. The model

			realistically simulated using agent-based models (ABMs), and that the parameters driving these behaviors—such as speed, path preference, or responsiveness to others—can change dynamically over time and should therefore be estimated sequentially. The authors also assume that particle filters, a	also heavily depends on the quality and resolution of the aggregate observation data; in scenarios where data is sparse or unreliable, forecasting accuracy may degrade significantly. **Conclusion**: This article proposes a novel and adaptive approach to crowd forecasting that is theoretically robust and promising for applications such as emergency planning and public safety. However, further validation with real-world data and improvements in computational efficiency are necessary before it can be reliably applied in practical, high-stakes scenarios.

			type of sequential Bayesian estimation technique, are suitable for estimating both the evolving state of the crowd and these latent behavioral parameters simultaneously.	
10.	Thomas Kopalidis, Vassilios Solachidis, Nicolas Vretos, Petros Daras. (2024)	Advances in Facial Expression Recognition: A Survey of Methods, Benchmarks, Models and Datasets. (mdpi) [10]	This article declares the fact that not all facial expressions can be accurately mapped to a human emotion, however, it assumes that there are six fundamental emotions – anger,	**Objective:** Facial Expression Recognition (FER) is the process of computers identifying and categorizing facial expressions to determine a person's emotional state in an image or video. This article takes an explorative view on FER, methods used to implement it, and algorithms that can be used to offer insights into intrinsic problems that come with FER. **Methods**: Various models are evaluated for performance in this article. Some of the most noteworthy models mentioned are FER-former, FER+, FER-VT, and TransFER. FER-former emerges as one of the best models, and uses three different datasets. It achieves an accuracy of

			fear, happiness, disgust, surprise and sadness – that are universal and can be represented in the same manner. Another universal emotion of contempt was added to this list, and these emotions are used in the FER system to measure valence and arousal circumplexes.	90.96% on FER+, which is the second-best model with 90.25%. FER-VT is a state-of-the-art model with the best accuracy of 100% and uses the CK+ dataset, and doesn't need another training dataset. It introduces a novel CNN-based FER framework that incorporates two attention mechanism at both the low-level learning and high-level semantic representation stages. It focuses on important facial regions and enhances the model's ability to capture discriminative features and extract meaningful representations. **Limitations**: "While multimodal systems have demonstrated strong performances in controlled laboratory settings, they still face challenges in achieving ecological validity when applied to real-world "in-the-wild" data." – This highlighted line is from the discussions section of the article. It indicates that these FER models that were talked about still face challenges when it comes to real-world applications and "in-the-wild" data. **Conclusion**: This article not only provides the appropriate methods to implement FER but also has a list of good quality datasets that are often used in FER systems, AffectNet being the most popular, as of the year 2023. The shortcomings of these systems when it comes to real-world "in-the-wild" data could be overcome using advanced computer vision and image processing to clean the noise present in real-world data. One of the articles reviewed above (Multi-Modal Depression Detection in Videos from Non-Verbal Cues) talks about methods to tackle issues highlighted in this article.
11.	You Wu, Qingwei Mi,	A comprehensi	This method assumes that the	**Objective:** This article explores Multimodal Emotion Recognition (MER) systems, which is a process that

	Tianhan Gao. (2025)	ve review of Multimodal Emotion Recognition: Techniques, Challenges, and Future Directions. (mdpi) [11]	multimodal data received is cleaned and correctly formatted as per each of their data types. This research also acknowledges the heavy reliance of MERs on publicly available datasets that provide the data needed to train and evaluate emotion recognition models. The most widely dataset being GEMEP corpus which provides audio, visual and physiological information for a range of expressions of emotions. Hence, this system is assumed to have all of this data at its disposal.	integrates various data types or modalities such as audio from speech, visual and textual data to identify human emotions. This system has three main components : feature extraction, multimodal information fusion and emotion classification. **Methods**: The research delves into two main types of multimodal information fusion for MER – Bimodal Emotion Recognition and Trimodal Emotion Recognition. Bimodal Emotion Recognition, as the name suggests, involves combining two different modalities. The two types that are often combined are speech and visual data. Bimodal approach captures complementary features from both modes of data, thereby enhancing the system's ability to recognize emotions more effectively than an unimodal system. This method goes into detail about the three fusion strategies in bimodal MER – early fusion, late fusion and hybrid fusion. Out of these fusion methods, late fusion is the most flexible, since early fusion has the requirement of having to match feature dimensions and the need for precise synchronization of data. Most systems such as Wear-BioNet use late fusion for their bimodal MER models. Hybrid fusion attempts to leverage the advantages of early and late fusion. Deep learning techniques used here are CNN, RNN and/or LSTMs. CNNs are used for extracting visual features, while RNNs or LSTMs capture temporal dependencies in speech data. Other techniques used are attention mechanisms, transfer learning between modalities, cross-model learning, canonical correlational analysis.

				Limitations: The systems discussed in this article are rather heavy weight and do not include explainable deep models. The article acknowledges the fact that this is crucial for making MER systems more accessible and interpretable. Computational costs of Deep Learning models are very high, which limits the accessibility and deployment to edge devices with limited resources. MERs also perform poorly across different datasets, referred to as the cross-corpus problem in the article. **Conclusion**: Despite the presented limitations of the MER system, there are actionable solutions to these problems such as using an XAI layer to add explainability, using adversarial training targets to deal with the gaps between features of the source and target datasets and more, as suggested in the article. We could also train MERs with a diverse dataset, using data augmentation to diversify our datasets so that the cross-corpus problem is minimized to an acceptable degree.
12.	Rawad Abdulghafor, Abdelrahman Abdelmohsen, Sherzod Turaev, Mohammed A.H. Ali and Sharyar Wani. (2022)	An Analysis of Body Language of Patients Using Artificial Intelligence. (mdpi) [12]	This paper makes the assumption that the interpretation of a person's body language such as their arms and legs movements, facial recognition and body postures are sufficient information to aid in an illness	**Objectives:** This article is a literature review that takes an explorative look at past papers that utilized AI to analyse body language and perform full-body tracking or facial expression detection for various tasks such as fall detection and COVID-19 detection, and looks at their significance and results. **Observations:** The article places significant emphasis on the importance of body language in healthcare. It is a crucial form of non-verbal communication, however, body language can have different interpretations and meanings in different parts of the world and different cultures, so it is important to keep that in mind. While making a body

			diagnosis, specifically during a pandemic or an endemic.	language analysis system, it is important to specify the boundaries and limitations of the working of the system. Some methods used for body language analysis, considering only the body and not facial features, are gesture recognition systems using 2D/3D CNNs, RNN/LSTM models for temporal patterns and motion based features to improve recognition, fall detection using wearable device data and LSTM models integrated to detect a fall, Sitting Posture Monitoring System (SMPS) to detect poor posture and sitting habits specifically in hospital beds to prevent ulcers, optical flow and blob tracking to identify risky behaviour in psychiatric wards. **Challenges**: Some of the biggest challenges to consider is the cultural bias in body language interpretation, as mentioned earlier. Diversified datasets need to be used to prevent this from happening as much as possible. Other concerns with these systems are privacy concerns and the limited scalability due to most studies being tested only on small patient groups. **Conclusion:** Body language analysis has proven to have some benefits and accurate analysis in past systems when it comes to detecting physical signs of a patient, thereby reducing the need for constant monitoring from human beings, who may miss crucial signs. These methods can be scaled to detect aggregate emotional trends in public spaces with applied anonymization to preserve the privacy of people in public.

Fig 2.1: System architecture diagram to detect personality and emotion traits using OCEAN attributes.

Chapter 3

Requirement Specifications

3.1 Functional Requirements

- The system must be able to analyse various indicators of mental health, such as facial emotions, body language, crowd behaviour, etc.
- Each ML model that makes up this system must provide the most accurate results possible so that the final result is accurate.
- All concerned users of the system must be able to use this system to improve their collective mental health.
- It should support anonymous user profiles and data logging.
- The system should send alerts or suggestions when signs of deteriorating mental health are detected.
- The system must provide the cause of the mental health result provided by it, regardless of if the results are positive or negative.

3.1.1 Hardware Requirements

- Windows 10 or above with at least 8 GB RAM.
- Computer or Smartphone Camera for testing.

3.1.2 Software Requirements

- Python 3.10 or above
- AI/ML libraries: TensorFlow, PyTorch, scikit-learn, NLTK, spaCy
- Backend: Django or Flask
- Frontend: HTML, CSS, JS
- Database: PostgreSQL
- Operating System: Windows/Linux/Android/iOS

- API Integration: Mental Health resources (WHO, NGOs)
- Visual Studio Code for the development environment.
- GitHub for version control.

3.2 Non-Functional Requirements

- Easy and intuitive user interface for users and administrators of diverse backgrounds.
- Ensure privacy and security of sensitive and personally identifiable information (PII).
- Ensure the system has high availability and reliability, and works with the highest accuracy possible.
- Must be scalable to handle multiple users and to be able to analyse data from various areas simultaneously.
- Must be able to store data needed for analysis and use it when the system is offline.

Chapter 4

Application

- Governments and NGOs can track population-level mental health trends in real-time. This can enable them to make informed developmental decisions to improve the overall mental health of the public.

- Can be used in educational environments to determine the overall mental state of students in an institution, and identify the causes of their mental states, thereby allowing educational institutions to make informed decisions to provide the best learning environment possible.

- Corporate businesses and other workspaces can track the mental wellness of their employees, and provide easy to understand and accurate results, thereby letting them make necessary changes to improve the productivity and wellbeing of the employees

- Rehabilitation Centres can use this application to track the mental health progress of individuals undergoing rehabilitation of any sorts, such as substance abuse recovery or behavioural therapy.

- Prison authorities can use the system to monitor mental health trends among inmates.

- Families, care providers or local authorities can use this system to monitor the emotional health of elderly individuals, especially those living alone.

Conclusion

The **AI-Driven Public Mental Health Tracker** represents a transformative leap in how mental health is approached, monitored, and managed across diverse populations. As mental health concerns continue to rise globally—fuelled by increasing stress, social isolation, and limited access to care—there is an urgent need for scalable, intelligent, and accessible solutions. This system addresses that need by leveraging the power of artificial intelligence, natural language processing, and emotion recognition technologies to detect early signs of mental distress, depression, anxiety, and other psychological challenges.

By enabling users to input data through simple interactions like journaling, answering questionnaires, voice notes, or even facial expressions (optional), the tracker uses AI models to interpret emotional states, behavioural patterns, and potential red flags. These insights are then presented through a personalized dashboard that visualizes trends over time, provides feedback, and offers scientifically-backed mental wellness recommendations. Moreover, the platform can notify designated caregivers, therapists, or crisis intervention services if it detects a critical deterioration in the user's mental state, all while maintaining user privacy and consent.

What sets this system apart is not just its technical capability, but its focus on **preventive care and early intervention**. Rather than waiting for individuals to seek help when symptoms become severe, the tracker fosters continuous, low-friction engagement—allowing problems to be identified and addressed sooner. Additionally, by integrating educational resources, coping strategies, and links to professional help, the system encourages users to build mental resilience and self-awareness.

In conclusion, this AI-Driven Public Mental Health Tracker offers a comprehensive, ethical, and scalable approach to one of the most pressing health issues of our time. It exemplifies how technology can be used not just to analyse data, but to deeply support and enhance human well-being—paving the way for a future where mental health care is proactive, personalized, and universally accessible.

References

1. Favaretto, R., Knob, P., Musse, S., Vilanova, F., & Brandelli Costa, A. (2021, April). *Detecting personality and emotion traits in crowds from video sequences*. https://doi.org/10.48550/arXiv.2104.12927
2. Gimeno-Gómez, D., Bucur, A.-M., Cosma, A., Martínez-Hinarejos, C.-D., & Rosso, P. (2024, January 5). *Reading between the frames: Multi-modal depression detection in videos from non-verbal cues* (arXiv:2401.02746v1 [cs.CV]). arXiv. https://doi.org/10.48550/arXiv.2401.02746
3. Gilanie, G., Hassan, M.U., Asghar, M., Qamar, A.M., Ullah, H. et al. (2022). An Automated and Real-time Approach of Depression Detection from Facial Micro-expressions. *Computers, Materials & Continua*, *73(2)*, 2513–2528. https://doi.org/10.32604/cmc.2022.028229
4. Chung, J., & Teo, J. (2022). Mental health prediction using machine learning: Taxonomy, applications, and challenges. *Applied Computational Intelligence and Soft Computing, 2022*(1), Article 9970363. https://doi.org/10.1155/2022/9970363
5. Sreenu, G., & Saleem Durai, M. A. (2024). Violence detection in crowd videos using nuanced facial expression analysis. *Systems and Soft Computing, 6*, 200104. https://doi.org/10.1016/j.sasc.2024.200104
6. Min, K., Yoon, J., Kang, M., Lee, D., Park, E., & Han, J. (2023). Detecting depression on video logs using audiovisual features. *Humanities and Social Sciences Communications, 10*(1), 788. https://doi.org/10.1057/s41599-023-02313-6
7. Alqahtani, S. A., Alasmari, A. M., Farooqi, N. S., & Alotaibi, Y. A. (2024). AI-powered predictive analytics for crowd health management. Mass Gathering Medicine, 2(1–12), Article 100009. https://doi.org/10.1016/j.mgmed.2024.100009
8. Kopalidis, T., Solachidis, V., Vretos, N., & Daras, P. (2024). Advances in Facial Expression Recognition: A Survey of Methods, Benchmarks, Models, and Datasets. *Information*, *15*(3), 135. https://doi.org/10.3390/info15030135
9. Varghese, E. B., & Thampi, S. M. (2020). Towards the cognitive and psychological perspectives of crowd behaviour: A vision-based analysis. Connection Science, 33(1), 1–26.

https://doi.org/10.1080/09540091.2020.1772723

10. Makinoshima, F., & Oishi, Y. (2022). Crowd flow forecasting via agent-based simulations with sequential latent-parameter estimation from aggregate observation. Scientific Reports, 12, Article 11168. https://doi.org/10.1038/s41598-022-14646-4
11. Wu, Y., Mi, Q., & Gao, T. (2025). A Comprehensive Review of Multimodal Emotion Recognition: Techniques, Challenges, and Future Directions. *Biomimetics*, *10*(7), 418. https://doi.org/10.3390/biomimetics10070418
12. Abdulghafor, R., Abdelmohsen, A., Turaev, S., Ali, M. A. H., & Wani, S. (2022). An Analysis of Body Language of Patients Using Artificial Intelligence. *Healthcare*, *10*(12), 2504. https://doi.org/10.3390/healthcare10122504
13. Konar, A.; Chakraborty, A. *Emotion Recognition: A Pattern Analysis Approach*; John Wiley & Sons: Hoboken, USA, 2015.
14. **Tánczos, Á. L.** (2022, August). *Detecting body language from interaction videos* (Master's thesis, Escuela Técnica Superior de Ingenieros Informáticos, Universidad Politécnica de Madrid). UPM Institutional Repository.
15. Qaraqe, M., Alotaibi, R., Qaraqe, K., Hussain, M., Muhammad, K., Lloret, J., & de Albuquerque, V. H. C. (2024). PublicVision: A secure smart surveillance system for crowd behavior recognition. *IEEE Access, 12*, 26474–26491. https://doi.org/10.1109/ACCESS.2024.3366693
16.