Master Thesis

Integrating Human Abnormal Behavior Detection with Air Quality Constraints for Safety in Manufacturing Environments

The Graduate School

Of the University of Ulsan

Department of Industrial Management Engineering

Soukaina R'bigui

Integrating Human Abnormal Behavior Detection with Air Quality Constraints for Safety in Manufacturing Environments

Supervisor: Professor Chiwoon Cho

A Thesis

Submitted to

The Graduate School of the University of Ulsan

In partial Fulfillment of the Requirements

For the Degree of

Master of Science

By

SOUKAINA R'BIGUI

Department of Industrial Management Engineering

University of Ulsan, Republic of Korea

December 2023

Integrating Human Abnormal Behavior Detection with Air Quality Constraints for Safety in Manufacturing Environments

This certifies that the thesis of Soukaina R’bigui is approved

Kihyo Jung

Committee Chair Professor Jung kihyo (정기효 교수)

Chiwoon Cho

Committee Member Professor Chiwoon Cho (조지운 교수)

Committee Member Professor Sudong Lee (이수동 교수)

Department of Industrial Management Engineering

University of Ulsan, Republic of Korea

December 2023

Acknowledgement

A master degree is a challenging and life-changing experience that would not be possible without the support and guidance from many people. I thank almighty Allah for his guidance and blessings.

First, I would like to express my sincere gratitude to my academic advisor, Professor Chiwoon Cho, for the immense knowledge, guidance, and enthusiasm throughout my study. He has taught me, both consciously and unconsciously, how good scientific research is done. Besides academic study, he helped me to grow as a professional with good work ethics. I would also like to thank all my thesis committee members, Prof. Kihyo Jung and Prof. Sudong Lee for their insightful comments, suggestions, and valuable questions during my thesis defense, which encouraged me to develop the outcome of my research work.

I would like to thank my family members, my grandmother Zaid Khadija, my sister Hind R'bigui , my brother Younes R'bigui, my brother in law Mohammed Abdulhakim Al-Absi and Professor Ahmed Abdulhakim Al-Absi, I thank them all for their enormous mentoring , guidance, sacrifices and blessings .

Last but not the least; I would like to thank all my family members and friends, for encouraging me to carry on.

Soukaina R'bigui

Ulsan, South Korea

December 2023

ABSTRACT

Integrating Human Abnormal Behavior Detection with Air Quality Constraints for Safety in Manufacturing Environments

In an era of rapid technological advancement, the manufacturing sector is undergoing a transformative shift toward Industry 4.0, embracing cutting-edge technologies to boost productivity and operational efficiency. However, with progress, new challenges comes, the most important is ensuring the safety, security, and air quality in manufacturing environments. This thesis aims to address these issues by designing novel Artificial Intelligence (AI) solutions that use Long Short-Term Memory (LSTM) networks to detect suspicious behaviors based on skeletal data.

The primary goal of this study is to design AI algorithms that improve safety and security by detecting suspicious behaviors based on skeletal data, with LSTM serving as a powerful tool for sequential data analysis. This algorithm is intended to detect behaviors such as excessive sneezing, frequent body scratching, or fainting, all of which may serve as early warning signs of health issues exacerbated by poor air quality.

Concurrently, the thesis delves into air quality measurement, employing a mathematical model based on short path distance analysis. This model, in addition to the behavioral detection aspect, provides a quantifiable framework for monitoring and assessing environmental conditions within manufacturing facilities. The combination of skeletal data analysis and mathematical air quality modeling represents a game-changing step toward increased vigilance, worker well-being, and efficient manufacturing operations.

Finally, the research aims to improve safety, security, and air quality monitoring, fostering a safer and healthier working environment for manufacturing personnel and optimizing manufacturing processes. This endeavor represents a synthesis of AI, quality measurement, and human activities in manufacturing, with the goal of ushering in a new era in which technology and well-being merge for the benefit of the manufacturing sector.

Keywords: Skeleton-data, artificial intelligence, LSTM, deep learning, safety management, mathematical model, short-path distance.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1.1 background

In the midst of Industry 4.0, where cutting-edge technologies transform manufacturing, we confront a pair of pressing challenges: ensuring worker well-being and maintaining pristine air quality in the workplace. This thesis embarks on an extensive exploration of these interlinked challenges, driven by technology and an unwavering commitment to excellence.

Our journey begins in a manufacturing environment, often susceptible to disease outbreaks and health risks. It is within this context, and especially during the times of disease outbreaks, that we find the urgency to prioritize worker safety and well-being, reinforced by the necessity to maintain high air quality standards (Figure 1). The foundation of our research is laid with a thorough review of established behavior detection algorithms, evaluated for their adaptability within the manufacturing landscape.

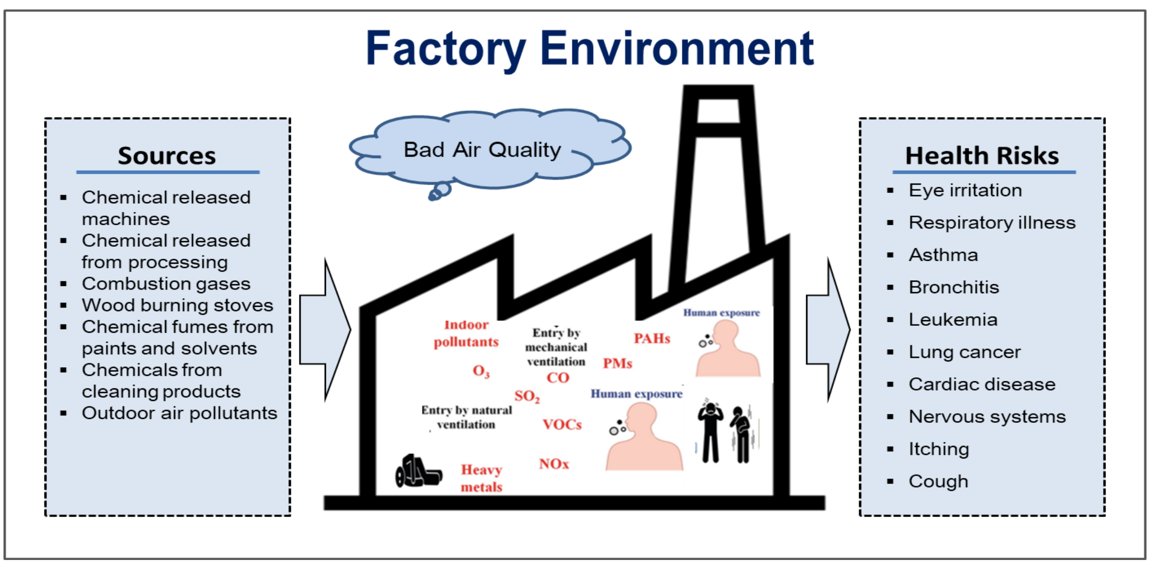

Figure 1. Factory Environment

These algorithms are pivotal as they become the vanguard in recognizing a wide spectrum of behaviors that may serve as early indicators of health risks. In this context, consider the sudden increases in coughing, frequent body scratching, unexplained fainting each bearing the potential to signal health concerns, particularly when air quality falters.

Our approach is underpinned by advanced AI, with a specific focus on Long Short-Term Memory (LSTM) networks renowned for their proficiency in sequential data analysis.

LSTM empowers us to swiftly and accurately identify these behaviors, allowing for immediate responses and effective risk mitigation.

Transitioning to the second phase of our exploration, our focus shifts to the dynamic realm of air quality, a global concern with profound implications for both the environment and human health. Here, the prediction of air pollution dispersion assumes paramount importance, given the inherent complexity and the unpredictability of meteorological conditions.

Our research endeavors are concentrated on the development of an Air Diffusion Model, fueled by the potential of AI and data. Named the Short Path Distance based Lagrangian Trajectory Model (SPD-LTM), this model adopts a trajectory-based approach and wind-field distance optimization to enhance predictions of air dispersion, leveraging insights from pollutant sensing data.

Within this framework, Particulate Matter (PM2.5) is our focal point. Our aim is to model its fluctuations and forecast its concentration, all while considering the subtleties of short path distance and time dependencies. Augmenting our understanding is the Hybrid Single-Particulate Lagrangian Integrated Trajectory algorithm (HYSPLIT), complemented by the Dijkstra algorithm for short path distance optimization. Results from this model underscore its superiority over conventional methods, enhancing predictive accuracy and deepening our comprehension of air quality dynamics.

In the fusion of these two phases, our thesis endeavors to chart a course toward manufacturing environments fortified by increased vigilance, worker well-being, and rapid responses to multifaceted challenges presented by compromised air quality and the emergence of abnormal behaviors. Our ultimate objective is a future defined by safety, health, and operational excellence, where technology and human welfare coalesce for the betterment of all.

The core of our research lies in the recognition of suspicious behaviors, embedded in the context of a manufacturing environment prone to disease outbreaks. At the same time, we ground our pursuit in the development of AI algorithms for abnormal behavior detection based on air quality requirements, striving for an exhaustive comprehension of these algorithms to further enhance worker safety and well-being.

1.2 Research Objectives

This dissertation is dedicated to investigating Innovative AI solutions for skeleton-based safety in manufacturing, and structured around the following objectives:

a) Develop Innovative AI Algorithms : Create advanced AI algorithms capable of detecting suspicious behavior based on skeleton data in manufacturing environments, with a specific focus on behaviors like excessive sneezing, frequent body scratching, and fainting.

b) Enhance Safety and Security: Utilize the developed AI algorithms to enhance safety and security within manufacturing settings, promptly identifying behaviors that may indicate health risks, particularly in the presence of compromised air quality.

c) Optimize Behavior Detection: Optimize the AI-based behavior detection algorithms to ensure their efficiency in recognizing ad categorizing unusual behaviors.

d) Air quality modelling: Develop an Air Diffusion Model that leverages AI and data to enhance the prediction accuracy of air pollution dispersion. Specifically, consider Particulate Matter (PM2.5) and it's concentration as a case study.
e) Safety and Health in Manufacturing: Ensure that technology and AI solutions contribute to safety, health and operational excellence within manufacturing environments, where the workplace is prone to diseases and health risks.

1.3 Dissertation Outlines

This dissertation is organized to investigate innovative AI solutions for improving manufacturing safety and security, particularly in the context of disease-prone environments and air quality concerns. The following is an outline of this research work:

1) Chapter 1 discusses the research background and the research objective will be addressed with appropriate references at the beginning of the chapter.
2) Chapter 2 discusses the implementation of skeleton data and LSTM to detect suspicious behaviors, indicating possible health concerns.
3) Chapter 3 reviews the innovations in wireless sensor networks that have transformed air quality monitoring, delivering real-time data for informed policy decisions.
4) Chapter 4 presents a new framework for atmospheric dispersion to forecast air quality.
5) Chapter 5 discusses the developments of the AI framework based skeleton data and indoor air quality conditions to study the abnormal behavior of worker in a manufacturing workplace prone to disease dispersion.
6) Chapter 6 discusses the summary of our work and directions to carry on this work in future.

Chapter 2

LSTM-Enhanced Suspicious Behavior Detection based on Skeleton Data

2.1 Introduction

In the fast-paced world of modern manufacturing, where operational efficiency is king, the need to ensure worker safety and security is critical. Manufacturing environments, which are characterized by intense industrial activity, face a dual challenge: protecting employee well-being while also addressing potential health hazards attributed to air quality. In the midst of these multifaceted challenges, the pursuit of novel solutions becomes essential, with a particular emphasis on skeleton-based behavior detection[1-4].

Safety and security are the central pillars of manufacturing operational excellence. Incidents, accidents, and health-related risks not only endanger workers' physical and psychological health, but they also have a significant impact on the overall operational efficiency and fiscal viability of manufacturing facilities. While significant progress has been made in strengthening safety protocols, critical gaps remain [5-9]. This chapter begins a thorough examination of AI-enhanced solutions designed to detect suspicious behavior in this context.

The need for innovative solutions to improve safety and security takes center stage in this chapter, with a focus on skeleton-based behavior detection. While traditional solutions have paved the way for increased safety, they frequently fall short of addressing the complex interplay of safety, security, and health. The creation of innovative algorithms based on artificial intelligence (AI) and utilizing skeleton data appears to be a promising avenue for addressing these multifaceted challenges. By meticulously analyzing workers' skeletal movements, this research aims to achieve a dual goal: preventing accidents while also monitoring air quality to mitigate health risks.

This pursuit of innovation goes far beyond theoretical speculation. In the following sections, we will look at specific manufacturing environments that are prone to air quality-related diseases. The implementation of AI-enhanced suspicious behavior detection systems, as well as their seamless integration with air quality monitoring mechanisms, will be investigated as part of a comprehensive approach to ensuring the well-being and security of the workforce.

This chapter marks the beginning of a transformative endeavor to improve manufacturing safety and security paradigms. As we progress through the chapters that follow, the implications of these efforts will become clear, potentially affecting not only the manufacturing sector but also occupational safety across industries. The imperative to ensure the safety and security of the workforce assumes paramount importance in the dynamic realm of modern manufacturing, where operational efficiency is a driving force. Manufacturing environments, which are characterized by intense industrial activity, face a dual challenge: protecting employee well-being while also addressing potential health hazards attributed to air quality. In the midst of these multifaceted challenges, the pursuit of innovative solutions becomes imperative, with a particular emphasis on skeleton-based behavior detection.

2.2 Kinect Camera

The Kinect camera, developed by Microsoft, represents a pioneering advancement in human-computer interaction and depth-sensing technology. Originally introduced as a peripheral for the Xbox gaming console, the Kinect utilizes a combination of infrared sensors, RGB cameras, and a depth sensor to capture and interpret the user's movements in three dimensions. This sophisticated system enables the Kinect to track the skeletal movements of individuals in real-time, translating their actions into on-screen interactions. Beyond gaming, the Kinect has found applications in a myriad of fields, including healthcare, education, robotics, and virtual reality. The depth-sensing capabilities of the Kinect create a detailed 3D point cloud of the surrounding environment, allowing for accurate spatial mapping and object recognition. Developers have harnessed the Kinect's potential for gesture-based controls, immersive augmented reality experiences, and even in rehabilitation therapies. While subsequent iterations of the Kinect have seen variations in design and capabilities, its impact on interactive technologies and its role in fostering innovation remain significant, making it a key subject of study and exploration in various research and application domains [10]. Figure 2 shown the standard parameters for Kinect V1 in outer space applications.

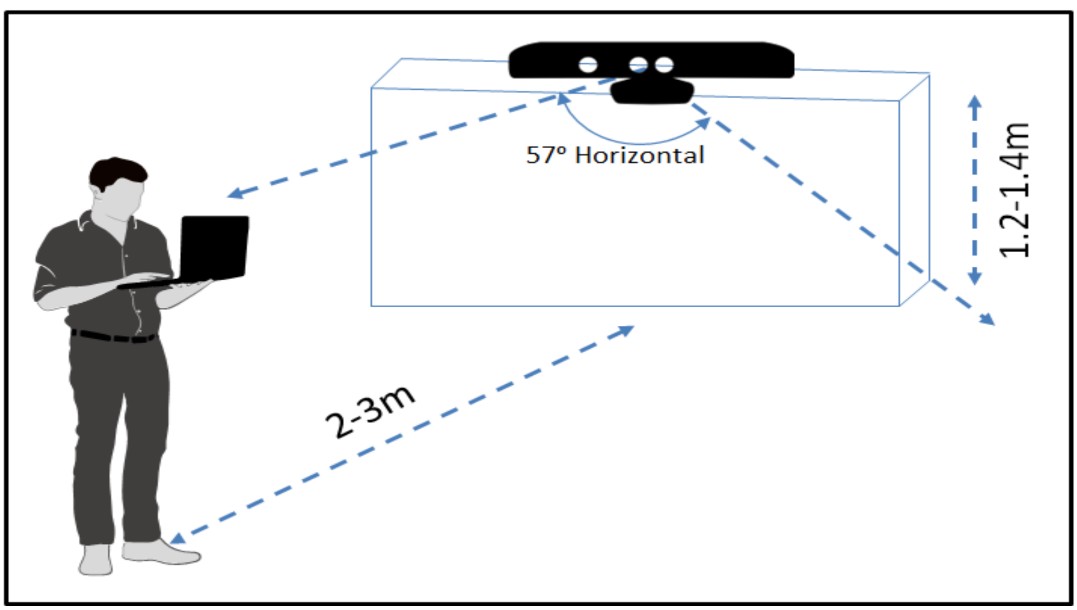

Figure 2. Standard Parameters for Kinect in Outer Space Applications

Kinect v1 has a recommended depth range of approximately 1.2 to 3.5 meters (3.9 to 11.5 feet). It is optimized for tracking users within this distance range. The horizontal field of view is around 57 degrees, and the vertical field of view is approximately 43 degrees. This determines the spatial coverage of the Kinect camera. Kinect v1 captures depth frames at a frame rate of 30 frames per second (fps). The color camera captures frames at a resolution of 640x480 pixels with a frame rate of 30 fps. Kinect v1 employs skeletal tracking technology, allowing it to recognize and track the movements of multiple users simultaneously. It can identify key joints and provide a real-time skeletal model.

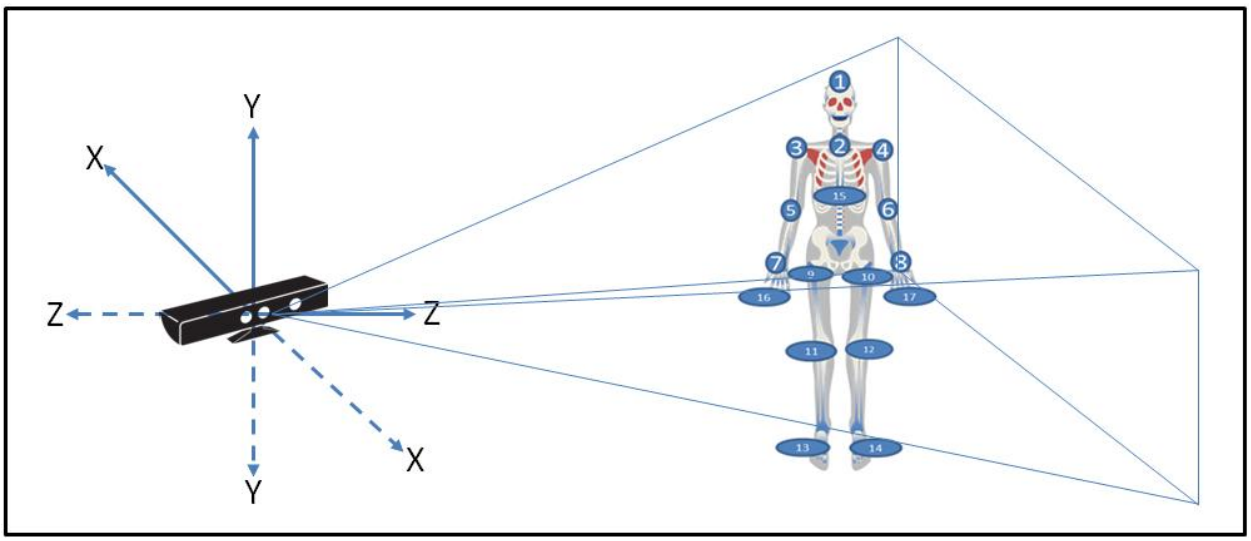

Figure 3. The 3D coordinate system of Kinect camera.

Figure 3 shows the 3D coordinate system of the Kinect V1 camera is a crucial aspect of its functionality, enabling precise spatial mapping and motion tracking. The system employs a Cartesian coordinate framework, where each point in the three-dimensional space is defined by its X, Y, and Z coordinates. The origin typically corresponds to the camera's position, with the X-axis extending horizontally, the Y-axis vertically, and the Z-axis capturing depth information. Kinect utilizes advanced depth-sensing technology, such as time-of-flight or structured light, to measure distances accurately and generate a detailed 3D point cloud of the scene. This coordinate system facilitates the mapping of real-world objects and movements into a digital space, making Kinect a valuable tool in various applications, including gaming, augmented reality, and motion analysis. Understanding the intricacies of the 3D coordinate system is fundamental for developers and researchers harnessing the Kinect's capabilities for innovative and immersive experiences.

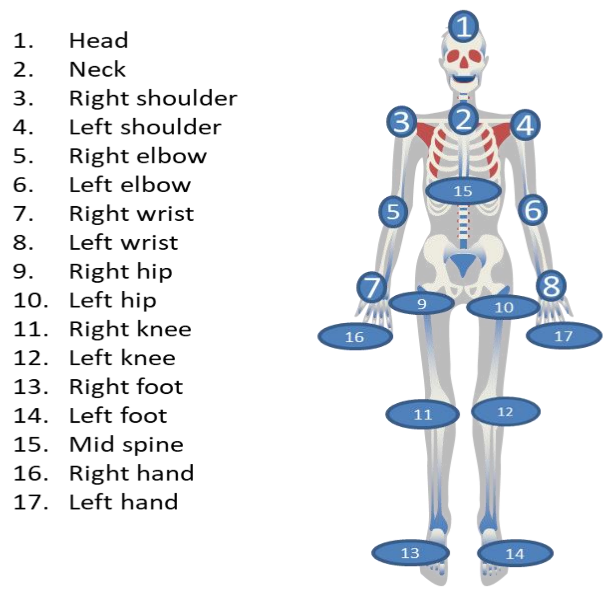

Figure 4. Structure of the human skeleton with 17 key joints.

2.3 Proposed methodology

Figure 4 is an illustrative representation of the human skeleton, focusing on 17 critical joints. These identified key points are anticipated to encompass specific bone structures for the human skeleton. Those 17 joints are head, neck, right shoulder, left shoulder, right elbow, left elbow, right wrist, left wrist, right hip, left hip, right knee, left knee, right foot, left foot, mid spine, right hand and left hand.

In this chapter we focused on 13 specific joints since not all joints are useful for our study. Those specific key joints are head, right shoulder, left shoulder, right elbow, left elbow, right hip, left hip, right knee, left knee, right foot, left foot, right hand and left hand.

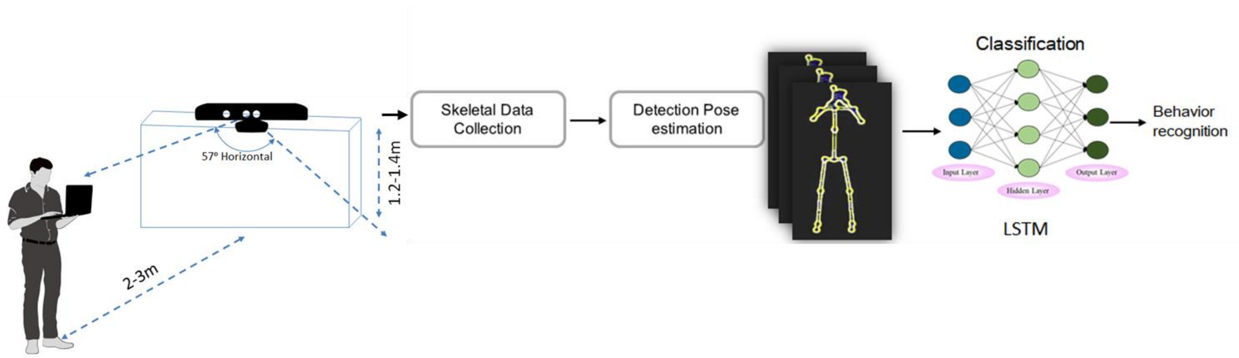

Figure 5. Behavior recognition methodology

The depicted Figure 5 is a visual representation or diagram illustrating the structure of a Long Short-Term Memory (LSTM) neural network. It is anticipated to provide a visual breakdown of the LSTM architecture, highlighting its key components and connections. Such diagrams serve as invaluable educational tools within the field of deep learning and artificial intelligence, aiding in the comprehensive understanding of the LSTM network's design and functionality for academic and professional purposes.

2.4 Experimental results

The behaviors under consideration are categorized into "Fainting" and "Non Fainting (Table1). The dataset consists of 1200 samples used for training the system and 600 samples for testing its accuracy. This partitioning of data is a common practice to assess how well the algorithm generalizes to unseen instances. The LSTM algorithm was employed to recognize and classify the behaviors. The system's recognition accuracy is presented in terms of both training and testing data. In 1200 training samples, the system correctly recognized 976 fall states and it is the 81.8% accuracy.

Table 1: Recognition results for fainting and non-fainting behavior

Worker Behavior	Type	Sample	Recognized Results		Accuracy %
			Fainting	Non Fainting	
Fainting	Training Data	1200	976	224	81.8%
	Testing Data	600	491	106	81.3%
Non Fainting	Training Data	1200	996	204	83.5%
	Testing Data	600	501	99	83%

2.5 Conclusion

In the manufacturing environment, our approach involves continuous monitoring of workers, focusing on the extraction of skeletal data features using a structured representation of the human skeleton with 13 key points. This detailed skeletal data serves as the foundation for our innovative solution. The Long Short-Term Memory (LSTM) network is employed as a pivotal component for the detection of suspicious behaviors exhibited by workers during their tasks.

The term "suspicious behavior" is assigned to actions that deviate from established norms and expectations. The anomalies we are particularly interested in encompass instances where a worker exhibits behavior such as excessive body scratching, fainting, or frequent and prolonged bouts of coughing. These behaviors are labeled as "suspicious" because they fall outside the realm of what is considered ordinary during routine work activities. It is important to emphasize that while these behaviors are deemed suspicious, they do not, at this stage, indicate the presence of a specific disease. Rather, they are indicative of actions that require further investigation and attention. The ultimate aim of our system is to serve as an early warning mechanism, alerting to deviations that may signal underlying health concerns. This preventive approach allows for timely intervention and the safeguarding of worker well-being in manufacturing environments.

Chapter 3

Real-time systems for air quality forecasting: A review of sensor networks, data fusion, and modeling approaches

An important environmental concern that has an impact on people's health and quality of life is air quality. For early detection of possible air quality issues and the creation of efficient mitigation plans, accurate air quality forecasting is crucial. To offer timely updates on the state of the air and increase the precision of air quality forecasts, real-time systems can be employed. These systems include sensor networks, data fusion, and modeling techniques. With an emphasis on these three strategies, this chapter examines the state of art of research on real-time systems for air quality forecasting. Sensor networks are groups of sensors placed all over an area to gather information on variables affecting air quality, such as temperature, humidity, and pollution concentrations. This work covers recent studies on the application of sensor networks for air quality forecasting going through the difficulties and advantages of this strategy. Data fusion consists of combining information from several sources, including as sensor networks and satellite data, to produce an accurate picture of the current state of the air. In this study, we cover recent research on data fusion methods for predicting air quality and highlight the advantages and drawbacks of this strategy. Using modeling techniques, it is possible to simulate how various environmental elements affect air quality and receive real-time updates on forecasts. This chapter presents a review on recent research on modeling methods for predicting air quality and outlines the difficulties and advantages of this method. Finally the main research gaps and areas of future research in real-time systems for air quality forecasting are emphasized in this chapter.

3.1 Introduction

Millions of people around the world's health and well-being are impacted by air pollution, a serious environmental problem. For early detection of possible air quality issues and the creation of efficient mitigation plans, accurate air quality forecasting is crucial. To offer timely updates on the state of the air and increase the precision of air quality forecasts, real-time systems can be employed. These systems include sensor networks, data fusion, and modeling techniques.

In order to gather information on air quality characteristics including temperature, humidity, and pollutant concentrations, sensor networks entail the deployment of several sensors throughout a region. These sensors can be linked to a central server or cloud computing platform, which can evaluate the data in real-time and offer air quality updates. Applications for sensor networks include tracking air quality in cities [11], tracking it inside buildings [12], and tracking it while moving [13].

Data fusion includes combining information from several sources, including sensor networks and satellite data, to produce an accurate picture of the current state of the air [14,15]. Data fusion can assist in addressing some of the drawbacks of separate data sources, such as sensor network data, which may have low temporal or spatial precision. Numerous applications, including monitoring of wildfire smoke [16] and urban air quality [17], have made use of data fusion approaches. Using modeling techniques, it is possible to simulate how various environmental elements affect air quality and receive real-time updates on forecasts [18,19,20]. These models can be used to mimic how changes in the weather, emission sources, and other environmental factors affect the quality of the air. Numerous applications have made use of modeling techniques, such as forecasting of wildfire smoke [21,22] and urban air quality [23]. The rest of this paper is conducted as follow: Section 2 discusses the advancements in wireless sensor networks for air quality monitoring. Section 3 highlights data fusion techniques for air quality monitoring. Section 4 presents real-time modelling approaches for air quality management. Section 5 summarizes and concludes the study.

3.2. Sensor network: Advancements in Wireless Sensor Networks for Air Quality Monitoring

Recent developments in wireless sensor networks have made it easier to employ them for a variety of purposes, such as environmental monitoring. In recent years, there has been a lot of interest in using wireless sensor networks to monitor air quality in particular. The ability of wireless sensor networks to deliver real-time information on air quality factors including temperature, humidity, and pollutant concentrations has been highlighted in a number of studies. A case study of a low-cost wireless sensor network for air quality monitoring in Northern Italy is shown in the paper by [24]. The study showed that using wireless sensor networks can produce trustworthy and accurate data on air quality, which can assist guide policies meant to improve air quality. A similar review of current advancements in wireless sensor networks for environmental monitoring, including air quality monitoring, is provided in [25]. The authors emphasize how wireless sensor networks can deliver real-time information on air quality that may be used to create efficient air quality management plans. Additionally, authors in [26] evaluate Internet of Things (IoT) and wireless sensor network-based air quality monitoring and early warning systems. The authors contend that combining wireless sensor networks and IoT can improve air quality management by increasing the precision and effectiveness of air quality monitoring systems.

3.3 Data Fusion Techniques for Air Quality Monitoring

The potential of data fusion techniques for enhancing air quality monitoring systems has been highlighted by recent studies. Data fusion is the process of merging information from various sources to get a more precise and comprehensive picture of a certain phenomenon. Data fusion can assist in overcoming some of the drawbacks of individual sensors in the context of air quality monitoring, such as their restricted coverage or sensitivity to particular contaminants. In order to predict fine particulate matter (PM2.5) concentrations using data from various sensors, [27] suggests a data fusion strategy based on a deep neural network. The authors show that the proposed method performs better in terms of accuracy and resilience than individual sensors and conventional data fusion techniques. Likewise, [28] uses a data fusion strategy to estimate China's PM2.5 concentrations by combining information from satellites and ground-based sensors. Comparing the suggested approach to individual sensors or satellite data alone, the authors demonstrate how it can deliver more accurate and thorough information on air quality. Furthermore, [29] suggests a data fusion method to predict and forecast air quality in Beijing, China, by combining data from many sources, such as sensors, meteorological data, and land use data. Authors present evidence that the suggested method can produce reliable and accurate air quality forecasts, which can be used to guide the development of policies aimed at lowering urban air pollution.

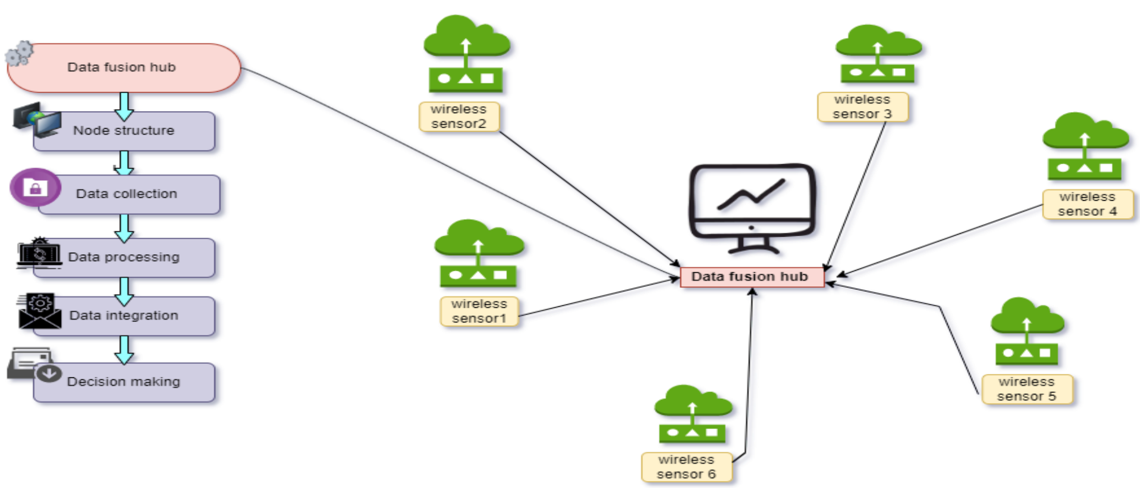

Figure 6. Data Fusion Process for Real-Time Air Quality Monitoring using network sensors

The procedure of monitoring air quality in a city using a wireless sensor network is shown in Figure 6. The sensors are positioned in different places to track various aspects of air quality, and the information gathered from each sensor is relayed to a central hub for data fusion. The phases of the data fusion process, including data gathering, processing, and integration, are also depicted in Figure 6. Data fusion process is described as follows:

1) Node Structure: In order to monitor various air quality metrics, the initial stage is establishing a wireless sensor network in the metropolitan area. Sensors are installed at various points across the network.
2) Data collection: The sensors gather data on the air quality indicators such as particulate matter, ozone, nitrogen dioxide, etc.
3) Data processing: The initial phase in the data processing process is pre-processing, which focuses on normalizing, transforming, and filtering raw data to get it ready for integration. The second stage includes data analysis and quality control. In the last phase, fusion techniques such kalman filters, neural networks, and Bayesian networks are used.
4) Data Integration: this step generates reports and alerts based on the integrated data and Combines data from different sources and creates a comprehensive picture of air quality.
5) Decision-making: Afterward, choices on the management of air quality and public health are made using the combined data.

3.4 Real-time modelling approaches for air quality management

The application of modeling approaches for air quality control has been made easier by recent developments in real-time systems and machine learning techniques. The potential of machine learning-based modeling techniques for air quality forecasting and management has been examined in a number of researches [30,31,32]. These modeling techniques produce precise estimates of air quality using real-time data from air quality monitoring stations, satellite data, and meteorological data. A deep learning-based air quality prediction model, for instance, was put forth and combines traffic, weather, and air quality data to forecast PM2.5 concentrations in [30]. The suggested approach beats conventional machine learning models in terms of prediction accuracy, according to the authors' research. Similar to this, [31] developed a hybrid model to forecast air quality that combines a support vector machine (SVM) model and a long short-term memory (LSTM) neural network model. According to the study, the proposed hybrid model has higher prediction accuracy than conventional models. A few research have also looked at the application of sophisticated modeling techniques, such as the hybrid model of deep learning and artificial neural networks (ANN) for forecasting air quality [33,34]. In order to forecast PM2.5 concentrations, for instance, [33] suggested a hybrid model that combines a convolutional neural network (CNN) with a gated recurrent unit (GRU) model. According to the study, the proposed hybrid model performs better in terms of prediction accuracy than conventional machine learning models. [34] Introduced an ANN-based air quality prediction model that makes use of numerous meteorological variables in a different study. The research demonstrated that the suggested model does a good job of forecasting the amounts of several air pollutants, such as PM2.5, NO2, and SO2. Generally, machine learning, ensemble methods, and time series analysis are different approaches that can be used for air quality prediction modeling, as depicted in Figure 7.

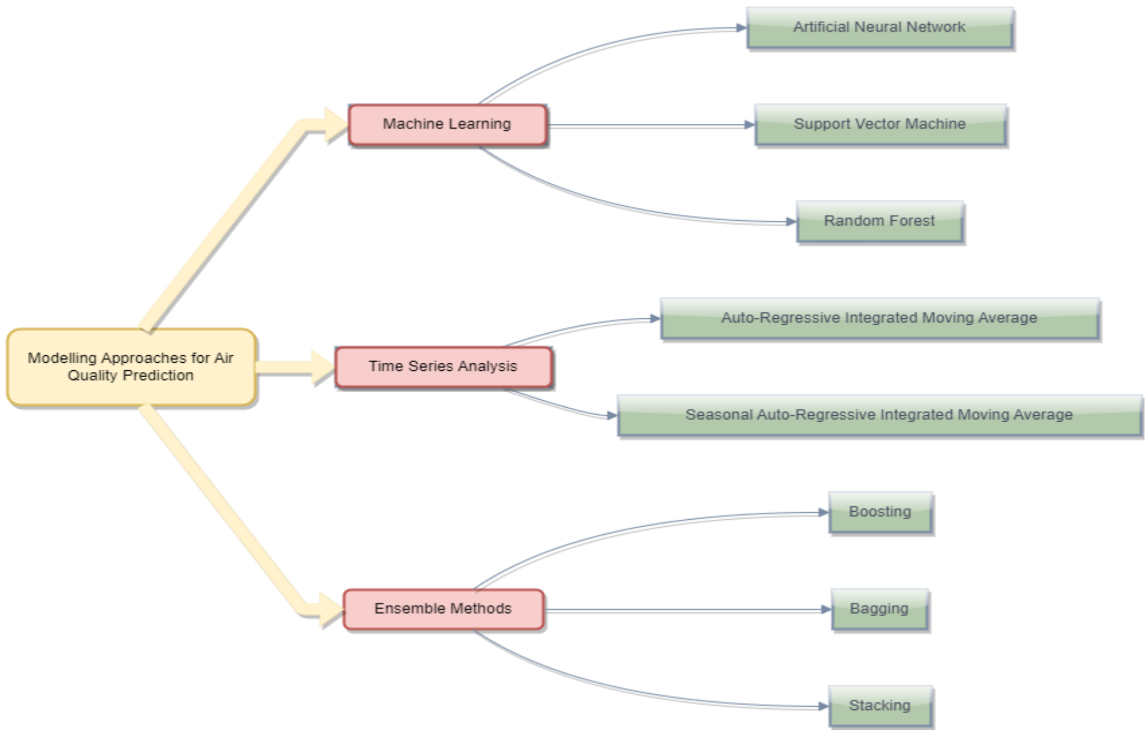

Figure 7. Example of best modelling approaches for air quality prediction

These modeling techniques, which include machine learning (Random Forest, Artificial Intelligence, and SVM), time series analysis (ARIMA and SARIMA), and ensemble methods (Boosting, Bagging, and Stacking) , help to increase the accuracy, timeliness, and reliability of air quality predictions. They make it possible to analyze sensor data in real-time, capture complicated correlations, and offer useful information for managing and making decisions about air quality.

In Machine learning, random Forest may be used to instantly examine huge amounts of data gathered from sensor networks. It is capable of accurately capturing intricate correlations between pollutant concentrations, meteorological information, and air quality metrics. Real-time air quality monitoring and forecasting can benefit from Random Forest's capacity to handle high-dimensional data and produce reliable forecasts [35].

Real-time air quality forecasting has showed potential for artificial intelligence approaches like neural networks and deep learning. They are able to make accurate forecasts on time, adapt to shifting trends, and learn from prior data. Real-time processing of sensor data by neural networks and deep learning models, enable proactive decision-making and prompt interventions, as well as their ability to recognize non-linear relationships and produce precise forecasts [36].

By utilizing SVM's capacity for classification and regression tasks, real-time air quality predictions can be accomplished. SVM models have the ability to interpret real-time data from sensor networks and forecast air quality based on past trends. SVM is suited for real-time analysis and forecasting because to its capacity to recognize non-linear relationships and manage high-dimensional data [37].

Furthermore, the utilization Auto-Regressive Integrated Moving Average (ARIMA) models of time series Analysis as shown in figure 2, is considered crucial for capturing temporal dependencies and trends in real-time air quality forecasting. ARIMA models can make short-term projections of air quality by examining historical data, including pollution concentrations and climatic variables. ARIMA is suitable for real-time monitoring and forecasting since it can take into account historical observations and recognize seasonality trends [38].

In addition to that, Seasonal Auto-Regressive Integrated Moving Average (SARIMA): It is frequently necessary to take seasonal changes and cyclic patterns into account when monitoring real-time air quality. SARIMA models are useful for real-time forecasting because they can efficiently capture these trends. SARIMA models enable precise forecasts in real-time by taking into account seasonal components in the data, allowing stakeholders to respond promptly to address air quality concerns [39].

In ensemble methods, by merging the results of various models and adjusting to shifting trends, boosting algorithms can improve real-time air quality forecasting. Boosting algorithms target enhancing model accuracy on difficult data examples by iteratively modifying the weights of training samples. Since of their versatility, boosting techniques are useful for real-time monitoring and forecasting since they allow for precise forecasts in circumstances with changing air quality [40].

Furthermore, by lowering variance and boosting stability, bagging techniques can enhance real-time air quality forecasts. Bagging reduces the danger of overfitting and produces reliable predictions by assembling a set of models trained on bootstrapped samples. In real-time systems, when precise and stable forecasts are essential for efficient decision-making, bagging techniques are very helpful [41].

Another aspect to consider is stacking. To take advantage of each model's advantages and enhance real-time air quality forecasting, stacking mixes the forecasts of various models. Stacking can identify various patterns and relationships in the data by using a meta-learner to integrate the predictions. Real-time systems can profit from stacking by incorporating many modeling approaches and utilizing their combined strength to produce precise and trustworthy projections [42].

3.5 Conclusion

The forecasting and monitoring of air quality might be greatly enhanced by real-time technologies. Each approach sensor networks, data fusion, and modeling has its own advantages and disadvantages, and the success of each depends on the application for which it is used and the data sources available. The development of new real-time systems that incorporate many methodologies and data sources should be the main emphasis of future research in order to produce air quality forecasts that are more precise and thorough. Such systems can be employed to educate public health professionals, legislators, and the general public about potential air quality problems and to create efficient mitigation plans.

Chapter 4

Short Path Wind-Field Distance based Lagrangian Trajectory Model for Enhancing Atmospheric Dispersion Prediction Accuracy

Air pollution is a major global issue that not only threatens the safety of our planet but also poses risks to global health. Weather plays a crucial role in the rapid dispersion of air pollution. Various models have been used to predict air pollution; however, atmospheric pollution dispersion remains unpredictable, especially in relation to meteorological conditions. Our research scope in this chapter focuses on developing an Air Diffusion Model using Future Wind and Pollutant sensing data to enhance prediction accuracy. In this paper, we present a new approach based on a mathematical model named the Short Path Distance based Lagrangian Trajectory Model (SPD-LTM). This model utilizes a trajectory approach and short path wind-field distance optimization to predict future air dispersion using pollutant sensing data. The framework developed in this work aims to model changes in Particulate Matter (PM2.5) and predict its concentration based on short path distance and time dependencies. The Lagrangian trajectory and concentration calculations are performed using the Hybrid Single-Particulate Lagrangian Integrated Trajectory algorithm (HYSPLIT). Then, we apply the short path distance algorithm using the Dijkstra algorithm. The obtained results demonstrate that the SPD-LTM outperforms the usual LTM and provides better accuracy to our predictive model.

4.1 Introduction

The evolutional wave of the fast industrialization has been leading us to face a serious climate changes and health issues. This global issue refers mainly to the toxic elements existing in the air we consume [43-44]. One of the most harmful contaminants in the air is PM2.5, which is a type of a small particulate matter that has a diameter of less than 2.5 micrometers [45-46]. PM2.5 particles are incredibly lightweight and small, which allows them to remain suspended in the atmosphere for a considerable amount of time [47]. They are produced by a variety of sources, including motor vehicles, power plants, factories, and wildfires, and can be dispersed over enormous area by wind and atmospheric conditions [48]. When breathed, these small particles may spread deep into the lungs and even in the bloodstream, causing many health diseases [49]. The concentrations of PM2.5 in the atmosphere are measured in micrograms per cubic meter (μg/m³) and are considered as a critical indicator of air quality. In many cities throughout the world, PM2.5 concentrations often surpass the Worth Health Organization's recommended limit of 10g/m3, reaching levels that are toxic to human health [50]. Atmospheric dispersion is a fundamental phenomenon that impacts the movements and distribution of air pollutants, especially PM2.5, in the atmosphere. It refers to the transport of pollutants through the air, driven by wind field, temperature gradients, and other meteorological conditions. Atmospheric dispersion models are used to forecast the movement and behavior of pollutants in the atmosphere, to help develop effective strategies for lowering air pollution and improving air quality, especially in industrial areas [51-57]. One way to examine the movement of air pollutants, such as PM2.5, is through Lagrangian trajectory models.

These models follow the movement of the air particles and can determine as well the source and path of air pollution. Indeed, lagrangian trajectory models use mathematical algorithms to simulate the transport of air pollutants over time and space. These models can take into consideration the impacts of wind, turbulence, and other meteorological conditions that affect the movement of air particles. By inputting data on the location and timing of air pollution emissions, as well as meteorological data such as wind speed and direction, these models may simulate the movement of air pollutants and provide insights into their dispersion patterns [58-70]. One of the most extensively used lagrangian trajectory models is the HYSPLIT. It employs both Lagrangian and Eulerian techniques to mimic the movement of air contaminants over huge distances and timescales. It has been used in a wide range of applications, including tracking the movement of pollutants from industrial and transportation sources, studying the dispersion of pollutants from wildfires and volcanic eruptions, and assessing the impact of long-range transport of air pollutants on human health and the environment [71-74]. Additionally, air pollution dispersion can be enhanced through the use of Short Path Distance (SPD) models. Considering that SPD models take into account the distance that air particles move during a short period of time, often several hours, in order to anticipate more correctly their dispersion patterns [75-77]. The goal of this study is to implement the lagrangian trajectory approach and the short path distance approach, by presenting a new method. Comparisons are conducted between real time concentration data of the PM2.5, the LTM concentration results and the SPD-LTM concentration to observe the prediction improvement achieved.

4.2 Related work

Models of atmospheric dispersion are crucial tools for determining how air pollutants affect the environment and human health. These models are used to mimic the behavior of air pollutants in the atmosphere, including their transport, diffusion, and chemical changes, according to [78]. The use of these models to forecast the concentration and dispersion of fine PM2.5 in the atmosphere has attracted increasing interest in recent years. For instance, [79] predicted South Korean PM2.5 concentrations using the Community Multiscale Air Quality (CMAQ) model. Operational Multiscale Environment Model with Grid Adaptivity (OMEGA) is one sort of atmospheric dispersion model frequently used for PM2.5. OMEGA is a cutting-edge air quality model that combines a variety of physical and chemical processes that affect the transport and fate of air pollutants in the atmosphere, according to [80-81]. It simulates air quality at numerous scales, from regional to local. The CMAQ model, a complete air quality modeling system created by the EPA and other research institutes, is another frequently used atmospheric dispersion model for PM2.5. The CMAQ model, which has been widely used for regulatory and research reasons, comprises accurate models of atmospheric chemistry, emissions, and meteorology according to [82-83]. The Weather Research and Forecasting model with Chemistry (WRF-Chem) [84], the Comprehensive Air Quality Model with Extensions (CAMx) [85], and the HYSPLIT model [85] are a few additional atmospheric dispersion models that have been used for PM2.5. A class of atmospheric transport models called lagrangian dispersion models simulates the flow of air contaminants by following the path of individual particles or air parcels. Lagrangian models, which can take into account the impacts of complicated meteorological circumstances such atmospheric turbulence and convection, are particularly well-suited for modelling long-range transport of pollutants, according to [86]. The HYSPLIT model, created by NOAA [87], is a typical Lagrangian dispersion model. HYSPLIT is a popular model for atmospheric transport and dispersion modeling, according to [88], and has been used for a range of environmental applications, such as air pollution, radioactive discharges, and volcanic ash dispersion. The Flexible Particle (FLEXPART) model, created by the Norwegian Institute for Air Research, is another Lagrangian model that has gained prominence recently [89]. FLEXPART has reportedly been used extensively in air quality and climate research, and has been applied to a wide range of atmospheric transport and dispersion investigations, according to [90]. Other Lagrangian dispersion models, such as the Particle and Dispersion Model (PDM) [91] and the Lagrangian Particle Dispersion Model (LPDM) [92], have been used in atmospheric transport and dispersion modeling in addition to HYSPLIT and FLEXPART. Overall, it has been found that Lagrangian dispersion models are useful tools for modelling the transport and dispersion of air contaminants in the atmosphere. On the other hand, short path distance models, a subset of atmospheric dispersion models, mimic the movement and dispersion of air pollutants over short distances, often of the order of a few kilometers or less. They are helpful for forecasting the amount of air pollution in the vicinity of emission sources, including factories or traffic. The Atmospheric Dispersion Modeling System (AERMOD) [93] and the California Puff (CALPUFF) model [94] are examples of frequently used short path distance models. These models have been used in a variety of regulatory and research situations, including environmental impact assessment [95] and air pollution modeling [96-98]. In general, atmospheric dispersion models are now a crucial tool for determining how PM2.5 affects both air quality and human health. These models are being continuously improved and updated to increase their precision and application, and it is anticipated that they will play a crucial part in determining future air quality laws and regulations [99]

4.3 MATERIALS AND METHODS

4.3.1 PM2.5 STUDY AREA

The Onsan industrial zone in Ulsan City is one potential study location that could concentrate on PM2.5 and weather factors. Ulsan city is renowned for having high levels of air pollution, notably PM2.5, which has been associated to harmful health impacts, according to prior studies [100]. About 21 monitoring stations were set up by researchers in Ulsan for the collection of both PM2.5 and meteorological data [101]. The Korea National Institute of Environmental Research (KNIER) station near the Onsan industrial complex is one of the areas where they placed air quality monitoring equipment for PM2.5 [102]. Basically Ulsan city has 5 districts: ulju-gun, buk-gu, jung-gu, namgu, and dong-gu. Onsan is located in the neighboring area of ulju-gun district as shown in Figure 8. The station is able to measure 6 metrics: SO_2, CO, O_3, NO_2, PM10, and PM2.5. The measuring station has been installed in 1993 and operated by the Ulsan Institute of Health & Environment agency [103]. The monitoring station location is shown in Figure 9 [104]. As indicated in Table 2, The Onsan station had a Latitude of 35.5388o N and a longitude of 129.1269° E.

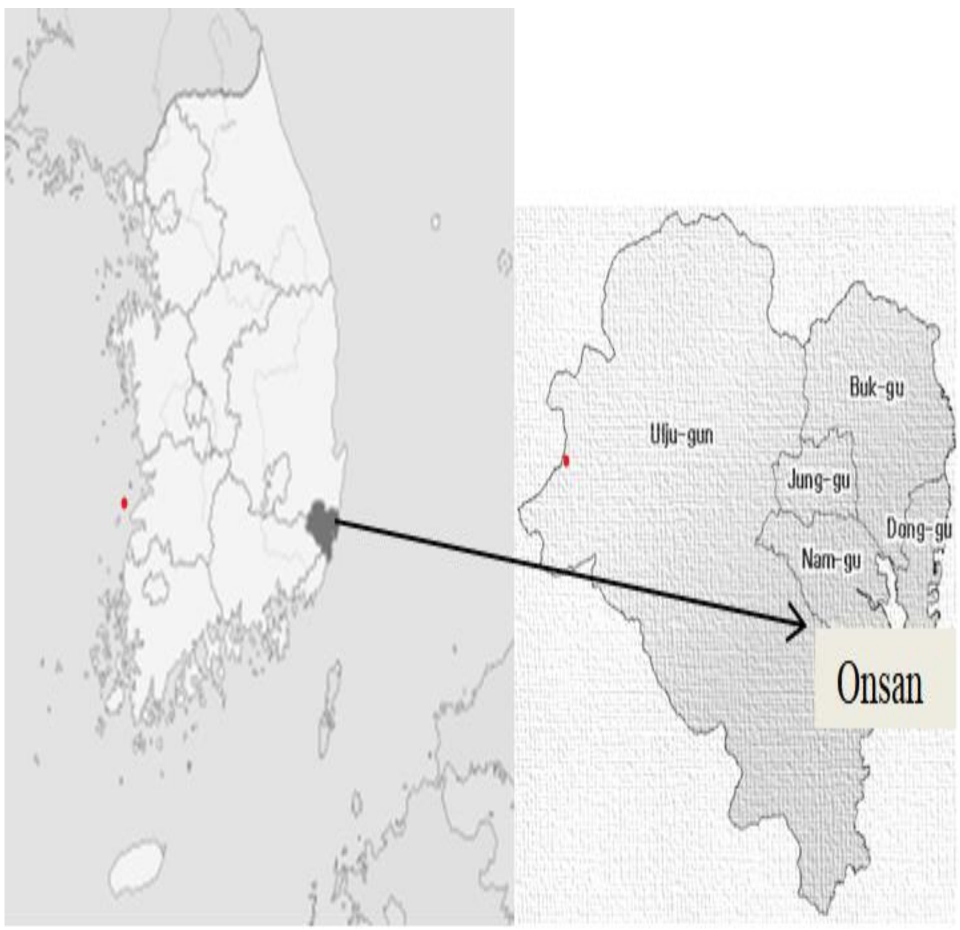

Figure 8. The study area Ulsan city

TABLE 2

station Specification

Air Quality Station Area	Onsan (Ulsan- South Korea)
Station Name	Deoksin-ro
Data Capturing Date	01/01/2022-12/31/2022
(Time span)	(Hourly Data)
Latitude	35.5388° N
Longitude	129.1269° E

4.3.2 DATA COLLECTION

The observed hourly meteorological data of Ulsan city used in this study was obtained from the Korea Meteorological Administration (KMA) which is the responsible agency for the meteorological observations in South Korea. The detailed observed weather data was provided by the Automated Weather System (AWS) , which is an automatic weather station network operated by the KMA [105]. PM2.5 data was retrieved from AirKorea which is a real-time air quality information disclosure system (www.airkorea.or.kr). The collected hourly data was from 1 January 2022 to 31 December 2022. As per the meteorological parameters, we retrieved wind speed, wind direction, temperature and humidity. Table 3 shows the related variables of the retrieved data at the monitored location of our study area. However, those units are commonly used in environmental monitoring, meteorology, and various scientific and data collection contexts to describe and quantify different parameters and variables. Year units represent time in years, typically used to indicate the calendar year when the data was collected. Month unit represents time in months, often used to specify the month within a year when the data was recorded. The Time is measured in hours, indicating the specific hours within a day when the data was taken (this is typically on a 24-hour clock (0~23). The longitude is measured in degrees and represents the east-west positon of a location on the Earth's surface (The ranges from -180° (West) ~ +180° (East)) with pri. Latitude is also measured in degrees and represents the north-south position of a location on the Earth's surface. It ranges from -90° (South) to +90° (North), with 0° at the Equator. Temperature is measured in degrees Celsius (°C), which is a common unit for indicating temperature in the metric system. It measures the warmth or coldness of the air or a substance. Humidity is expressed as a percentage (%), representing the relative humidity of the air. It measures the amount of moisture present in the atmosphere relative to the maximum amount it can hold at a given temperature. Wind speed is measured in meters per second (m/s) and represents the rate at which air is moving horizontally past a point. It's commonly used to describe the intensity of wind. Wind direction is measured in degrees and indicates the direction from which the wind is blowing. It's typically measured in degrees clockwise from north (0°). PM2.5 concentration is measured in micrograms per cubic meter ($\mu g/m^3$). It quantifies the concentration of fine particulate matter in the air with a diameter of 2.5 micrometers or smaller, which can have adverse effects on health when inhaled.

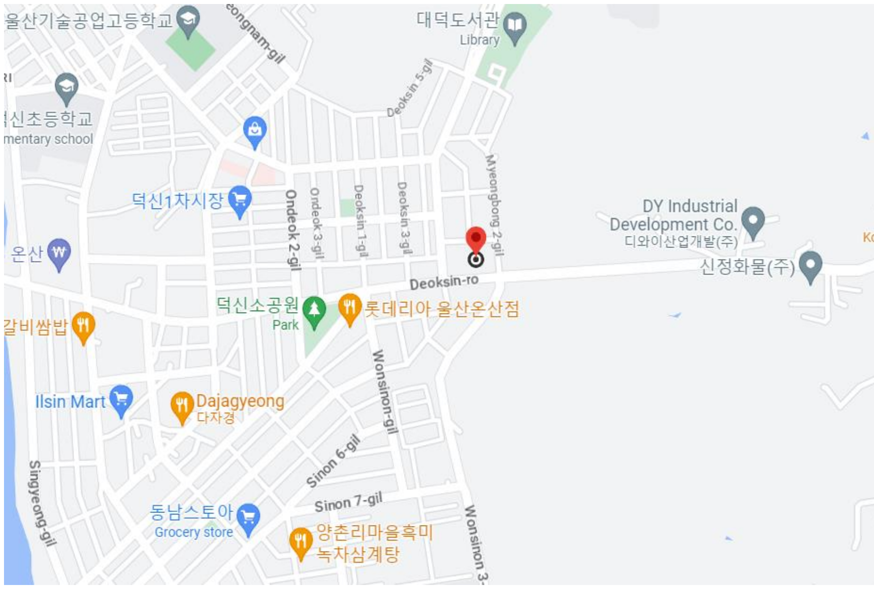

Figure 9. Monitoring station location of PM2.5

TABLE 3

Data Related variable

Related Variable	Unit
Year	year
Month	month
Hour	hour
Longitude (o)	degree
Latitude(o)	degree
Temperature	°C
Humidity	%
Windspeed	m/s
Winddirection	degree
Pm2.5	μg/m³

4.3.3 DATA PREPROCESSING

In order to analyze the PM2.5 data and meteorological data, a data preprocessing approach is necessary. Firstly, the two datasets need to be merged and mapped together to establish a comprehensive dataset for analysis. This can be achieved by aligning the timestamps or common identifiers present in both datasets. We used common timestamps during data integration to merge the PM2.5 data and meteorological data into a single dataset. Z-score standardization method has been applied to the related variables of the meteorological data and the pm2.5 data. However the geological data variables such as latitude and longitude didn’t require any standardization. Then, we ensured that both datasets have consistent and compatible formats. Interpolation techniques were used to fill in any missing values in the dataset. By estimating values based on the nearby accessible data points, interpolation aids in filling in the gaps in the data. We used spine interpolation, which uses smooth curves to approximatively fill in missing data, for handling missing values.

The equation for each cubic polynomial segment between two neighboring data points for the cubic spline interpolation is written as follows:

Taking into account two neighboring data sets (x_i, z_i) and(x_{i+1}, z_{i+1}), where x_i and x_{i+1} are x-coordinates, while z_i and z_{i+1} are y-coordinates. The cubic spline equation for the segment between these two points is as follows:

$$S(x) = a \times (x_{i+1} - x)^3 + b \times (x - x_i)^3 + c \times (x_{i+1} - x)^3 + d \times (x - x_i)^3 \qquad (1)$$

Where:

- The cubic spline function is denoted by S(x).
- The input value x is what we want to use to estimate the z-value.

The cubic spline must satisfy the following condition in order to compute the coefficients $a, b, c,$ and d. These conditions include:

1) Interpolation Condition: The spline passes through each data point, meaning $S(x_i)= z_i$ and $S(x_{i+1})= z_{i+1}$.

2) Continuity Condition: The first and second derivatives of adjacent polynomials are equal at each data point to ensure a smooth transition.

In order to meet the requirements of the integration of HYSPLIT algorithm, we made sure that our data comprised regular time intervals with constant temporal resolution, geographical coordinates (latitude, longitude), and relevant meteorological variables (wind speed, wind direction, and temperature). These aspects need to be considered in order to properly comprehend air transport and dispersion. To visually assess the data on wind direction, we used a wind rose in our preprocessing part. The wind rose shows the frequency and distribution of measured wind directions, providing information about the dominant winds and wind resources. By collecting and aggregating wind direction data into directional sectors, we create the wind rose in Figure 10, which offers a thorough picture of the wind condition in our study region. The observation of pm2.5 in correlation with the other features is shown in Figure 12. After the observation of each feature in Figures 11, 12, 13 and 14 in correlation with the measured pm2.5 retrieved from the measuring station of the study area, we conclude that temperature and humidity had no significant impact on the pm2.5 concentration as observed in the analyzed on Figure 15. In consequence, we are taking in consideration wind speed and wind direction features. The observation and analysis of these features are depicted in Figures 16.

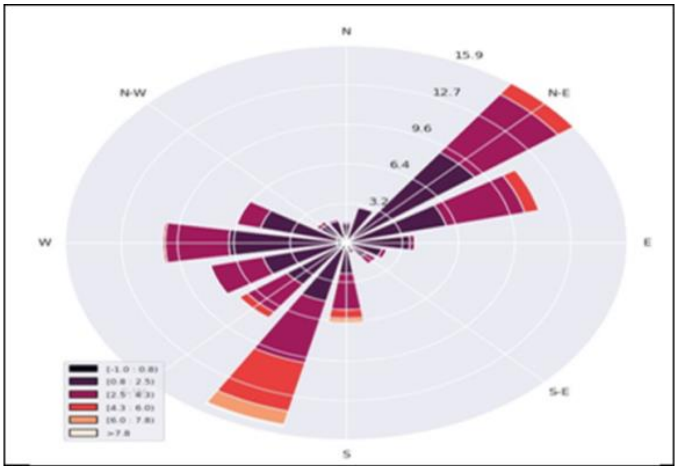

Figure 10. Wind rose of the wind speed and direction

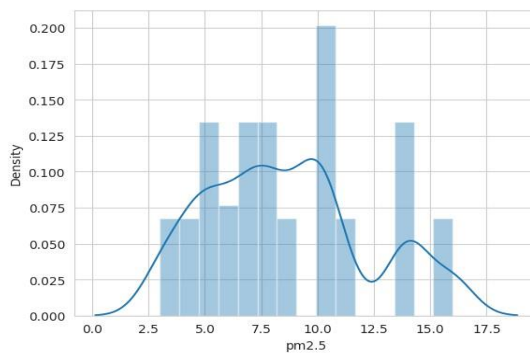

Figure 11. Measured pm2.5 distribution in the dataset

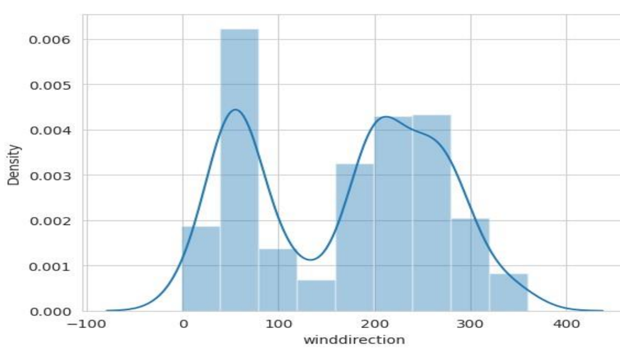

Figure 12. Winddirection distribution in the dataset

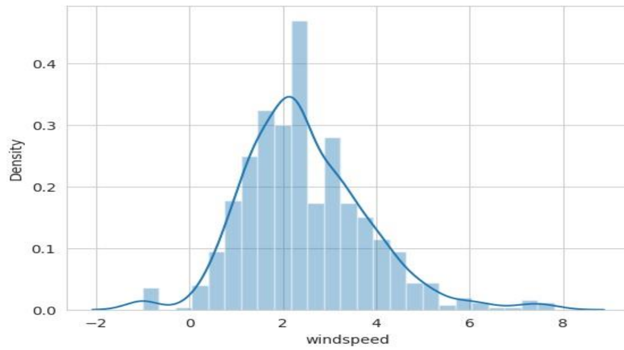

Figure 13. Windspeed distribution in the dataset

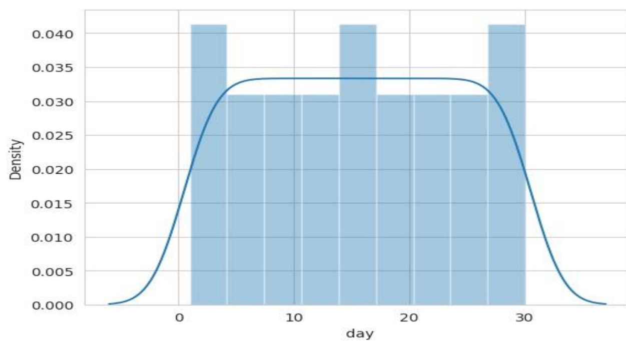

Figure 14. Time distribution in the dataset

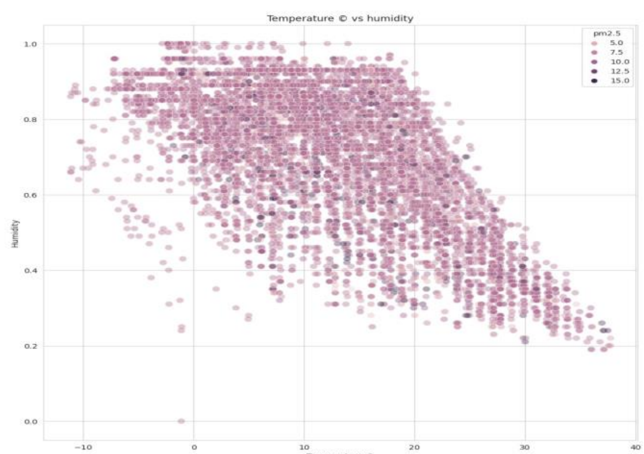

Figure 15. Temperature and Humidity by measured pm2.5

Figure 16. Windspeed and winddirection by measured pm2.5

Table 4 present the detailed data information statistics for five different variables: Temperature, Humidity, Windspeed, Winddirection, and Pm2.5 measured. Each row in the table represents a statistical measure for these variables, and the columns provide specific information for each measure. For instance, count shows the number of data points available for each variable (There are 8760 data points for each variable). Mean, represents the mean (average) value of each variable. Std represents the standard deviation of each variable (Bigger number means more spread). Min shows the minimum value observed for each variable. First Quartile (25%) shows the value at the 25th percentile of the data. Max gives the maximum (largest) value observed for each variable.

Table 4. Data Information Statistics

Index	Temperature	Humidity	Windspeed	Winddirection	Pm2.5 measured
Count	8760.0	8760.0	8760.0	8760.0	8760.0
Mean	12.17133	0.70149	11.30131	193.82842	13.10273
Std	9.18279	0.18450	7.01055	103.90505	12.28054
Min	-11.12778	0.0	0.0	0.0	0.0
(First Quartile) 25%	5.58333	0.57	6.2307	129.0	2.0
(Second Quartile) 50%	11.57778	0.74	10.6421	190.0	11.0
(Third Quartile)75%	18.68472	0.85	14.4417	291.0	20.0
Max	37.75556	1.0	55.9314	359.0	75.0

4.3.4 METHODOLOGY

In this part, we are discussing the attempted methodology used in this study for the pm2.5 concentration prediction. The process of this study is shown in Figure 17.The trajectories and concentrations of the pm2.5 were calculated using Hysplit algorithm. Equation (2) and equation (4) are the basic particle trajectory and particle concentration equations respectively [65]:

1. Particle trajectory:

$$p'(t + \Delta t) = p(t) + V(p, t). \Delta t \qquad (2)$$

The advection of the particle is calculated as the average of the three-dimensional velocity vectors $V(p, t)$ at the initial point $p(t)$ and the first-estimated position $p'(t + \Delta t)$ where t represents time. The final position is:

$$p(t + \Delta t) = p(t) + \frac{1}{2}[V(p, t) + V(p', t + \Delta t)]\Delta t \qquad (3)$$

To find the final position of the particle at time$t + \Delta t$ $(p(t + \Delta t))$, we update the initial position $p(t)$ by adding an average of the velocities at both time points$(t \text{ and } t + \Delta t)$.

The velocity vector at the first-estimated position $p'(t + \Delta t)$ is considered alongside the current velocity vector $V(p, t)$ at time t. The averaging of velocities helps to account for the potential changes in the particle's speed or direction over the time interval Δt.
The term $1/2\ [V(p, t) + V(p', t + \Delta t)]$ represents the average velocity vector over the time interval Δt, and when multiplied by Δt, it gives the displacement of .the particle over that time step. Adding this displacement to the initial position $p(t)$ gives the final position of the particle at time $t + \Delta t$. In both space and time, the velocity vectors are linearly interpolated.

1. Particle concentration :

$$\Delta C = m\left(2\pi\sigma_h^2 \Delta z\right)^{-1} \exp\left(\frac{-x^2}{2\sigma^2 h}\right) \qquad (4)$$

Where ΔC represents the change in particle concentration at a specific point in the three-dimensional space and m representing the mass of the particles. It scales the overall concentration level. σ_h is the standard deviation of the Gaussian distribution in the horizontal plane. It controls the spread or dispersion of the particle concentration in the horizontal direction, Δz is the variation or difference in height or altitude in the vertical (z) direction, x is the distance from the center point (where the concentration is at its peak). It is usually measured in the horizontal plane. σ is the standard deviation of the Gaussian distribution in the x-direction. It controls the spread of the concentration along the x-axis, and $h$is a constant parameter, which can influence the vertical spread of the concentration. Therefore, the spread of the distribution in both the horizontal and vertical directions is controlled by σ_h and Δz, respectively.
The cost function is used to quantify how well the predicted PM2.5 concentration aligns with the observed values while also considering the influence of the short path distance. By minimizing the cost function, the model aims to improve its predictive accuracy and effectively account for the relationship between PM2.5 concentration and the distance traveled along the short path. The cost function equation is as follows:

$$\text{cost} = W_1 \times (PM_{\text{predicted}} - PM_{\text{observed}})^2 + W_2 \times SD \qquad (5)$$

The squared difference between the predicted and observed pm2.5 concentrations is combined in the cost function. By squaring the difference, the necessity of an accurate forecast is highlighted and greater errors are amplified. SD variable refers to the short path distance component, is a measurement of the length of the trajectory connected to the concentration of pm2.5. The distance of the trajectory has been calculated in this phase using the Dijkstra method. Our model carefully examines the effect of the trajectory distance on the prediction of pm2.5 concentration while implementing the short path distance in this cost function. W_1 and W_2 are weighting factors that determine the relative importance of the prediction error component and the short path distance component in the cost function. W_1 controls the influence of the prediction error component, while W_2 controls the influence of the short path distance component. The notations and explanations used in this cost function as well as the notations used in this study, are summarized in Table 5.

Table 5. Notation and Explanations

Notation	Explanation
$PM_{predicted}$	The predicted pm2.5 concentration
$PM_{observed}$	The observed pm2.5 concentration
W_1	Weight factor related to the prediction error component
W_2	Weight factor related to the short path distance component
SD	Variable refers to the short path distance, a measurement of the length of the trajectory connected to pm2.5 concentration
$p(t)$	Initial point of the particle at time t
$\mathrm{p}'(\mathrm{t}+\Delta\mathrm{t})$	first-estimated position of the particle $(t+\Delta t)$

4.3.5 EVALUATION METRICS

We compared the observed concentrations of pm2.5, the predicted concentrations without the short path distance, and the predicted concentrations with the short path distance to assess the performance of our model and determine how much efficiency the short path distance would add to our prediction. The statistical evaluation metrics listed in Equations 6, 7 and 8 were employed for this evaluation. $\mathrm{y_i}$ is the observed pm2.5 concentration, $\hat{\mathrm{y}}_\mathrm{i}$ is the predicted pm2.5 concentrations, $\bar{\mathrm{y}}_\mathrm{i}$ is the mean of observed values and n is the length of the test set. Measures of the difference between the observed value and the predicted value include Root Mean Square Eroor (RMSE), Mean Absolute Error (MAE) and Means squared error (MSE). The RMSE and MAE metrics show how sensitive and resistant the model is to greater errors, respectively. The prediction impact is better when the two values are lower. MSE evaluates how well the accuracy predicted results match the observed data. The lower MSE value, the better the effect can be predicted.

We calculated the MAE by taking the absolute difference between each observed $\mathrm{y_i}$ and predicted $\hat{\mathrm{y}}_\mathrm{i}$ value and then averaging those absolute differences over all data points n. The formula is as follows:

$$\mathrm{MAE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i-1}}^{\mathrm{n}} |\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i}| \qquad (6)$$

This equation (6) calculates the absolute difference between each observed value $\mathrm{y_i}$ and its corresponding predicted value $\hat{\mathrm{y}}_\mathrm{i}$ for all data points ($\mathrm{i} = 1$ to n). Then, it takes the sum $\sum$ of all these absolute differences. Finally, it divides the sum by the total number of data points n to obtain the mean, which gives the average absolute difference between the observed and predicted values.

We have calculated the MSE by taking the squared difference between each observed $\mathrm{y_i}$ and predicted $\hat{\mathrm{y}}_\mathrm{i}$ value and then averaging those squared differences over all data points n. The formula is as follows:

$$\mathrm{MSE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i-1}}^{\mathrm{n}} (\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i})^2 \qquad (7)$$

Where the equation (7) calculates the squared difference between each observed value $\mathrm{y_i}$ and its corresponding predicted value $\hat{\mathrm{y}}_\mathrm{i}$ for all data points. Then, it takes the sum $\sum$

of all these squared differences. Finally, it calculates the average squared difference between observed and predicted values by dividing the sum by the total number of data points n.

We obtained the RMSE by taking the square root of the MSE. It provides a measure of the typical error magnitude between the observed and predicted values. The formula is as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i-1}^{n}(y_i - \hat{y}_i)^2} \qquad (8)$$

The equation (8) calculates the average difference between the observed and predicted values, squares it, takes the mean of all the squared differences, and then takes the square root of that mean. The resulting value represents the typical difference between the observed and predicted values, with lower values indicating better model performance.

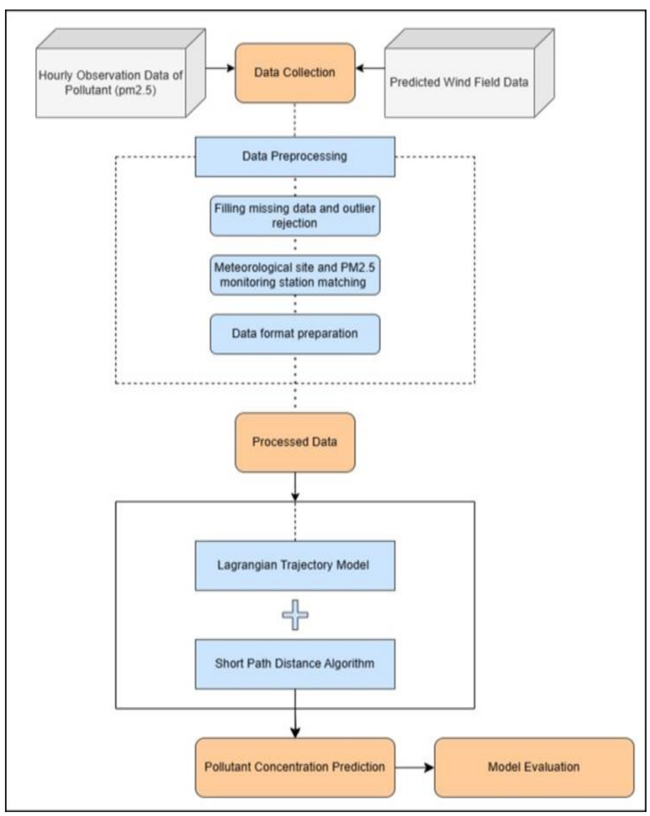

Figure 17.Overview of the study design

Table 6. Notations and explanations

Notation	Explanation
$PM_{predicted}$	The predicted pm2.5 concentration
$PM_{observed}$	The observed pm2.5 concentration
W_1	Weight factor related to the prediction error component
W_2	Weight factor related to the short path distance component
SD	Variable refers to the short path distance, a measurement of the length of the trajectory connected to pm2.5 concentration
$p(t)$	Initial point of the particle at time t
$\mathrm{p}'(\mathrm{t}+\Delta\mathrm{t})$	first-estimated position of the particle $(t+\Delta t)$ time
$\mathrm{p}(\mathrm{t}+\Delta\mathrm{t})$	Final position of the particle at $(t+\Delta t)$ time of the
$V(p,t)$	Velocity vector of the point at time t of the particle
Δt, t	Time
ΔC	Particle concentration
Δz	A certain height about the source of emission
σ	Vertical dispersion coefficient
σ_h	Horizontal dispersion coefficient
h	Reference level at which the concentration is being computed
m	Emission rate
X	Horizontal distance from the emission source
$S(x)$	The cubic spline function for the interpolation
(x_i, z_i)	two neighboring data sets of the cubic spline function
a, b, c, d	Interpolation coefficients of cubic spline function
MAE	Mean Absolute Error
MSE	Mean Squared Error
$RMSE$	Root Mean Squared Error
yi	observed pm2.5 concentration
$\hat{y}_i$	predicted pm2.5 concentration
$\bar{y}_i$	the mean of observed values
n	the length of the test set

4.4 RESULTS & DISCUSSION

Our model's evaluation showed useful predictions of PM2.5 values. The MAE between the predicted and observed values for our SPD-LTM model was determined as 1.1400. Additionally, our model's MSE, which measures the typical squared difference between predicted and observed values, was 1.2400. Additionally, the RMSE, which measures the residuals' standard deviation, was discovered to be 1.4800. Compared to the short path distance model, the Lagrangian trajectory model had somewhat larger errors, as seen by the MAE of 1.2600, MSE of 1.3000, and RMSE of 1.6700. When short path distance is not taken into account, these measurements show a bigger average absolute difference, squared difference, and standard deviation of the residuals. The better model performance, as shown by the reduced MAE, MSE, and RMSE values, emphasizes the important role of the short path distance parameter in improving the precision of PM2.5 concentration forecasts. We were able to consider the particular transport pathways and fully comprehend the spatiotemporal distribution of PM2.5 concentrations by including short path distance. Figure 18 shows the pm2.5 concentration monthly results. We conclude that the average concentration varies from 0 to 20($\mu g/m^3$) which is considered as good quality average. Contrary to the usual LTM results, the concentration of pm2.5 is not accurate as depicted in Figure 19. As a summary of our model performance, Figure 20 presents a comparison between the observed pm2.5 concentration, SPD-LTM predicted concentration and the LTM predicted concentration. The short path distance parameter's extra data enables more accurate calculations and depicts the intricate dynamics of pollution movement more clearly. These findings highlight the significance of coupling Lagrangian trajectory model with the short path distance, as well as the crucial impact of our proposed cost function in the short path distance has in improving PM2.5 concentration prediction. The importance of this parameter is shown by the enhanced performance of our model over the usual Lagrangian trajectory model. The results of this study have significant impact for managing air quality and making decisions on public health, and they offer insightful information for comprehending and combating PM2.5 pollution. Table 6 and Figure 21 show performance of SPD-LTM and a comparison with LTM.

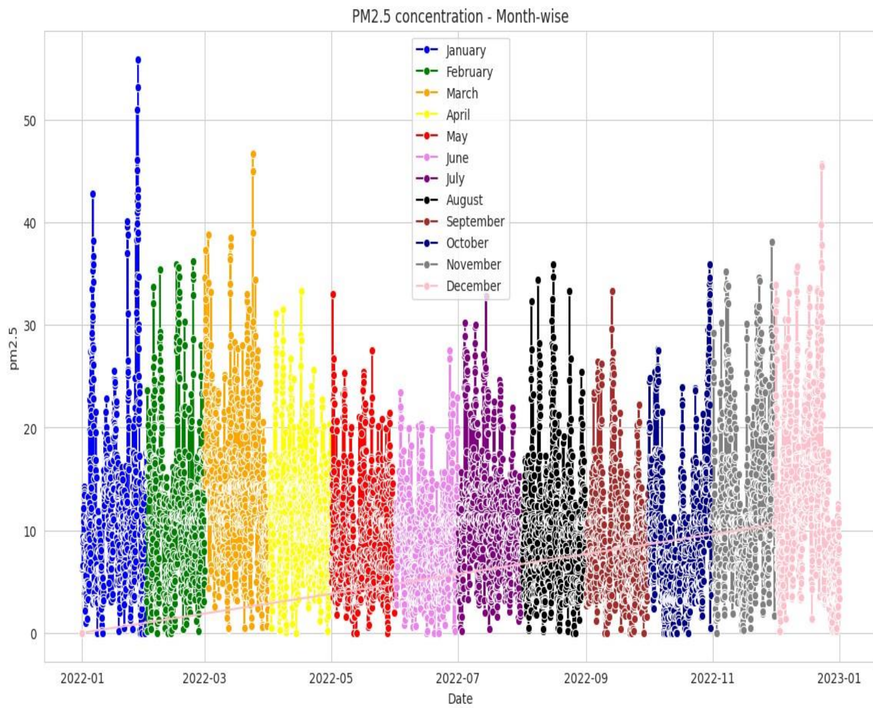

Figure 18. pm2.5 concentration monthly results for the SPD-LTM prediction

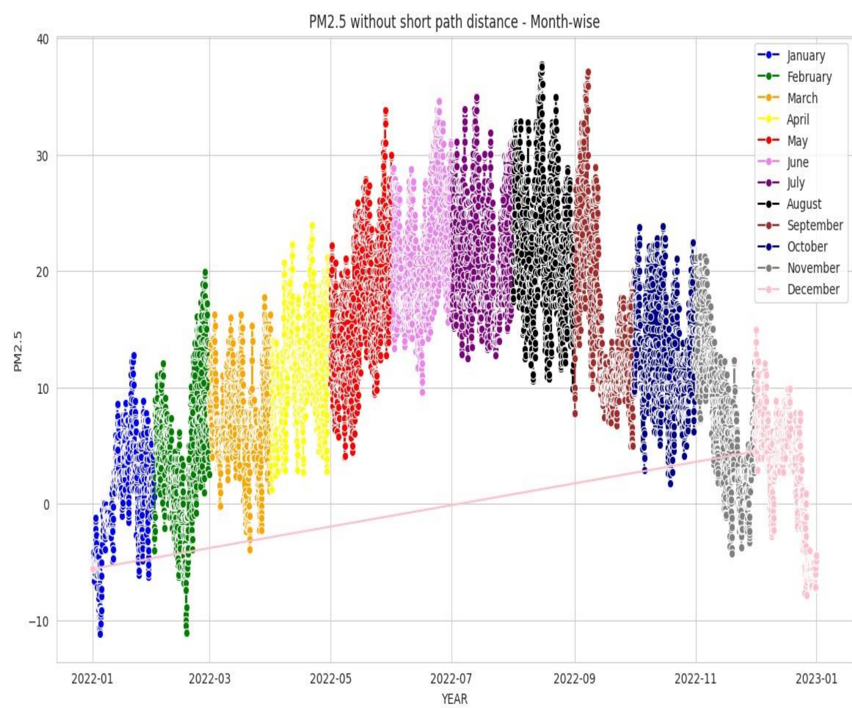

Figure 19. pm2.5 concentration monthly results for the LTM prediction

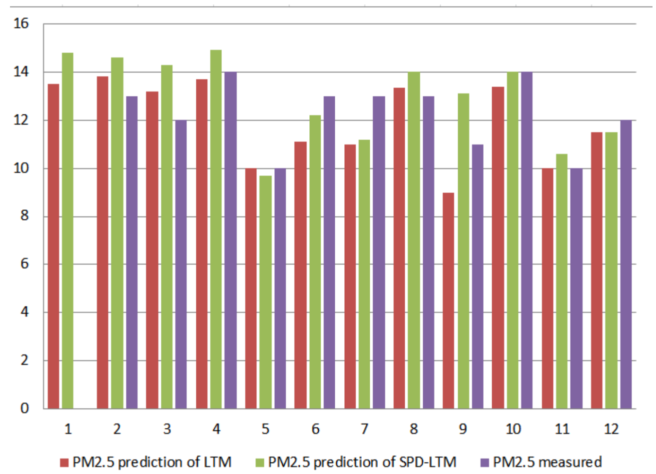

Figure 20. Comparison between the predicted pm2.5 of SPD LTM and LTM and the measured pm2.5

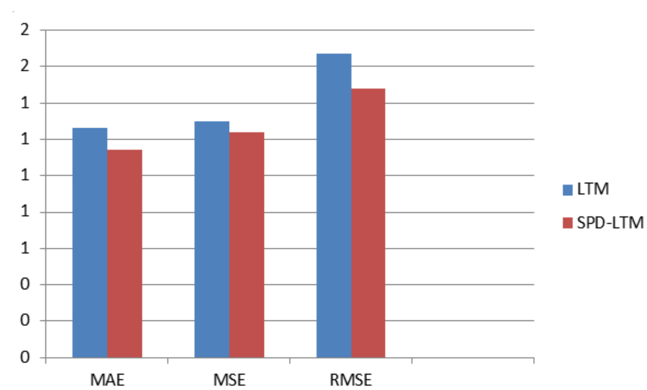

Figure 21. Model prediction performance comparison

Table 7. Model prediction performance at the study location

Method	LTM	SPD-LTM
MAE	1.2600	1.1400
MSE	1.3000	1.2400
RMSE	1.6700	1.4800

4.5 CONCLUSION

In this study, a new framework for pm2.5 propagation was presented. This study shows how a Lagrangian trajectory model with short path distance can be used to estimate PM2.5 concentrations by introducing a new mathematical method. The reduced values of MAE, MSE, and RMSE show that the model outperforms both the real-time data and the traditional Lagrangian trajectory model. The inclusion of short path lengths greatly increases the accuracy and precision of the predictions and provides a more accurate understanding of the dynamics of pollutant transport. The results highlight the importance of considering specific transport pathways and their effects on PM2.5 concentrations. This method holds great promise for use in treatments related to public health, environmental planning, and air quality management. Exploring the full potential of this technology and incorporating it into decision support systems should be the focus of future studies. Future work should focus on several areas to further enhance the application of this methodology. Firstly, the model's performance can be assessed in different geographical regions and under varying meteorological conditions to evaluate its generalizability. Additionally, the integration of other relevant parameters such as land use patterns, emission sources, and meteorological factors could contribute to a more comprehensive predictive model. Furthermore, the incorporation of artificial intelligence, machine learning techniques and advanced data analytics methods may enhance the model's predictive capabilities.

Chapter 5

LSTM & SPD-LTM based Abnormal Behavior Detection Model

5.1 Introduction

This thesis introduces a comprehensive framework designed for the continuous monitoring and assessment of human behavior and air quality within manufacturing environments in Figure 22. The framework employs advanced technologies, including real-time behavioral analysis and air quality sensing, to dynamically evaluate potential safety hazards. By integrating these assessments, the system aims to proactively identify abnormal conditions and trigger timely interventions, thereby enhancing overall safety in manufacturing settings. This generic framework is flexible, allowing adaptation to diverse manufacturing scenarios while providing a foundation for improved occupational safety practices. In the context of manufacturing safety, the proposed framework addresses the need for a dynamic and integrated approach to assess both human behavior and air quality. By utilizing cutting-edge technologies, the system aims to contribute to proactive safety measures, mitigating risks associated with diverse manufacturing processes. The framework begins with a robust system architecture that encompasses real-time monitoring of human behavior and air quality. Sensors and data acquisition devices are strategically deployed to capture relevant information, forming the basis for subsequent analyses.

A key component involves the continuous monitoring of human behavior on the manufacturing floor. Utilizing technologies such as skeleton data extraction and machine learning algorithms, the system captures, analyzes, and interprets worker actions in real-time. This enables the identification of abnormal behaviors that may pose safety risks. A pivotal feature of the framework lies in the integration of human behavior and air quality assessments. By correlating these datasets, the system distinguishes between normal and abnormal behaviors in context with the prevailing air quality conditions. This integrated assessment allows for nuanced responses tailored to specific safety concerns.

Upon identifying abnormal conditions, the framework triggers an early warning system. Immediate alarms prompt timely interventions, ensuring that potential safety hazards are addressed promptly. This proactive approach aligns with the overarching goal of preventing accidents and injuries in manufacturing environments. Recognizing the diversity of manufacturing processes, the framework is designed to be adaptable and scalable. Its generic nature allows for customization based on the specific needs and nuances of different manufacturing settings, promoting widespread applicability.

In this chapter, we introduce a dynamic system for LSTM-enhanced abnormal behavior detection based on air quality data. The system's operation is contingent on the interplay between air quality conditions and observed worker behavior.

Figure 23 shows Flowchart of the proposed manufacturing safety system. When a worker displays a suspicious behavior, such as fainting, the system's initial response is to assess the concurrent air quality conditions. Specifically, the system evaluates factors like temperature, gas concentrations, and contaminants in the environment. The predefined standard conditions for air quality serve as a baseline for normalcy. If the air quality conditions align with this norm, the system issues an alarm, signifying the potential gravity of the worker's fainting incident.

Conversely, when the suspicious behavior exhibited by the worker is characterized by behaviors such as coughing or body scratching, the system proceeds to the second tier of analysis. Again, it scrutinizes the air quality parameters. However, if the air quality remains within the normal range, the system refrains from issuing an alarm. This distinction underscores the differential treatment of worker behaviors, prioritizing those that may pose a more immediate health risk.

In essence, this chapter delves into the intricate workings of our AI-driven system, emphasizing its ability to dynamically assess and respond to both worker behavior and air quality conditions. This approach empowers the system to discriminate between behaviors that warrant immediate attention and those that, when aligned with favorable air quality conditions, may be benign. The system's capacity to issue alarms when necessary and withhold them when not required underscores its role in enhancing worker safety and well-being in manufacturing environments.

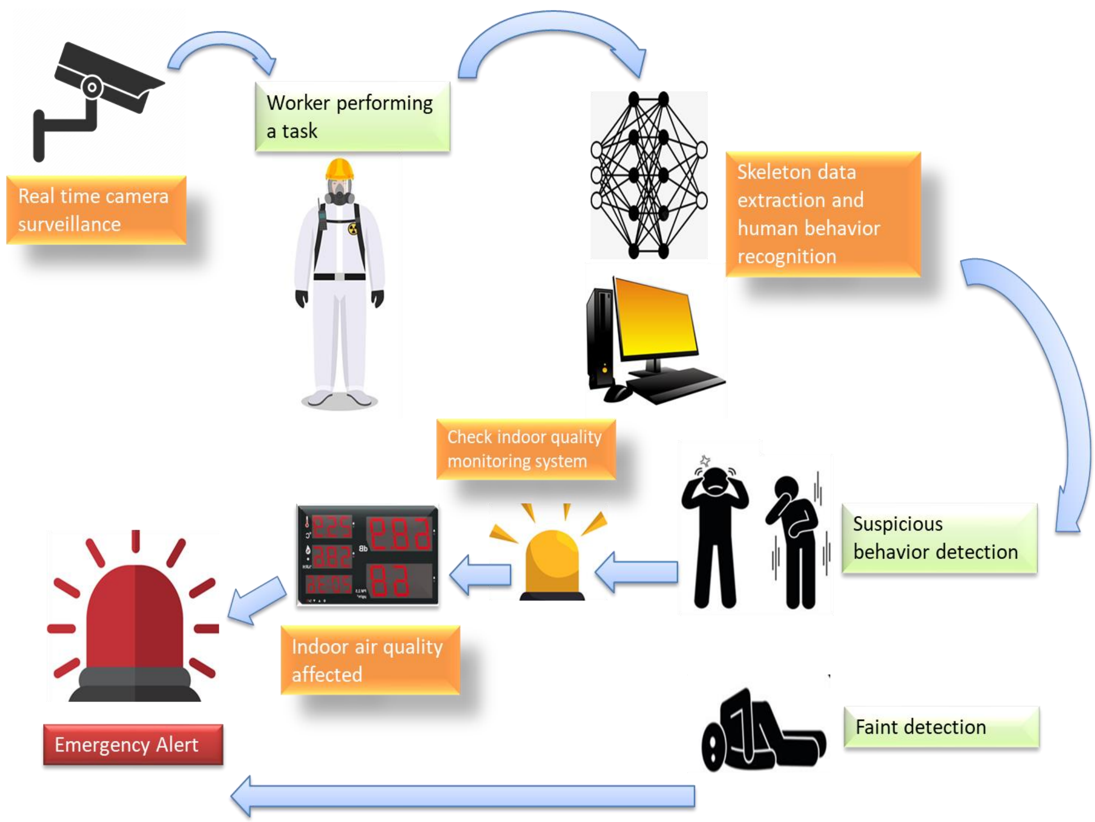

Figure 22. Generic of the dynamic human behavior and air quality assessment for safety in manufacturing

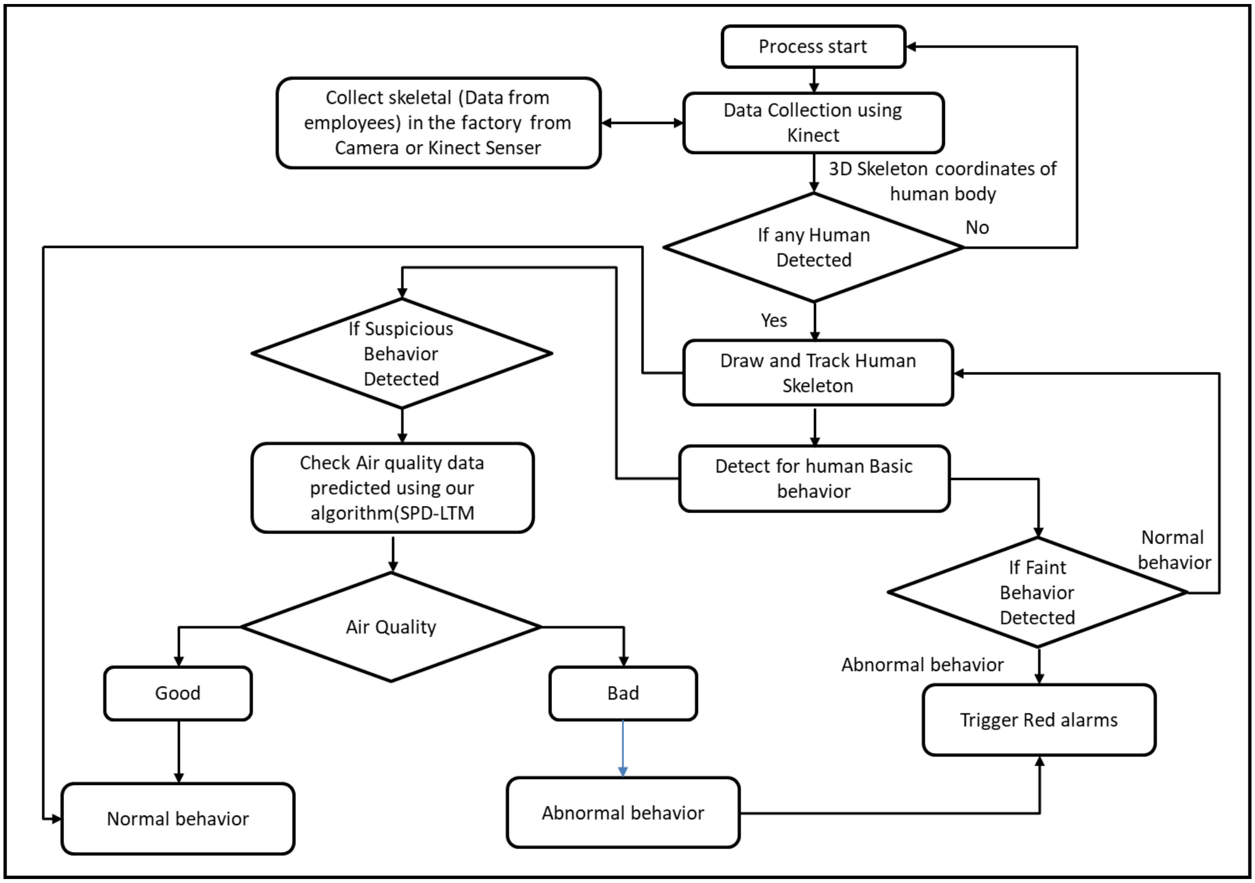

Figure 23. Flowchart of the proposed manufacturing safety system

5.2 Dynamic human behavior and air quality assessment for safety in manufacturing

In Table 8, our algorithm outlines the decision-making process for assessing worker behavior and air quality conditions to determine when to issue alarms for immediate action. It distinguishes between different behaviors and considers the state of air quality to prioritize worker safety in manufacturing environments.

This algorithm will presents a robust system for real-time monitoring of worker behavior with a focus on identifying potential health and safety concerns in the occupational environment. Leveraging advanced techniques such as skeleton data extraction and LSTM model learning, the system continuously assesses worker actions and responds promptly to abnormal behaviors. This thesis elucidates the key steps involved in the algorithm, its application in assessing behaviors like coughing, body scratching, and fainting, as well as its integration with air quality conditions for a comprehensive safety evaluation. The algorithm begins with the initialization phase, where the monitoring system is configured for real-time assessment. This includes setting up data recording mechanisms for observed behaviors and implementing skeleton data extraction techniques to facilitate ongoing LSTM model learning.

Table 8. Dynamic Worker Behavior Monitoring and Alarm System for Occupational Safety

Algorithm 1: Dynamic Worker Behavior Monitoring and Alarm System for Occupational Safety
Step 1. Initialize the monitoring system for real-time worker behavior assessment.
Step 2. Set up data recording mechanisms for observed behaviors.
Step 3. Implement skeleton data extraction techniques for subsequent LSTM model learning.
Step 4. Begin the continuous loop for real-time monitoring of worker behavior then
Step 5. Observe and record the worker's behavior then
Step 6. Extract skeleton data for ongoing LSTM model learning.
Step 7. For each observed behavior, If the behavior is identified as Coughing or body scratching do
Step 8. Evaluate the current air quality conditions.
Step 9. If air quality conditions are abnormal (within predefined standards do
Step 10. Mark the behavior as suspicious and abnormal then
Step 11. Issue an immediate alarm for intervention
Step 12. .End If
Step 13. End For
Step 14. End If
Step 15. Else (if air quality conditions are normal)
Step 16. Mark the behavior as suspicious but normal.
Step 17. End If
Step 18. End For
Step 19. If the behavior is identified as fainting do
Step 20. Issue an immediate alarm for intervention.
Step 21. End If
Step 22. Update

A continuous monitoring loop is established to observe and record worker behavior. This iterative process ensures that the system adapts to dynamic changes in the work environment, providing a real-time assessment of worker actions. Concurrently, skeleton data extraction occurs to enhance the learning capabilities of the LSTM model. Each observed behavior is systematically analyzed. In the case of coughing or body scratching, the algorithm evaluates the current air quality conditions. If abnormalities are detected within predefined standards, the behavior is flagged as suspicious and abnormal, triggering an immediate alarm for intervention. Conversely, if air quality conditions are deemed normal, the behavior is marked as suspicious but normal. For fainting behaviors, an immediate alarm is issued. The monitoring loop concludes upon the completion of the behavior analysis. The algorithm ensures a thorough examination of recorded data, providing a comprehensive overview of worker actions and associated safety implications.

Chapter 6

Conclusion and Future work

6.1 Conclusion

In this thesis, we have explored innovative approaches to enhance safety and security in manufacturing environments. We introduced a multi-faceted system that combines skeleton-based behavior detection and air quality monitoring to detect and respond to abnormal worker behaviors. Our research addressed the need for proactive measures to safeguard worker well-being and the potential risks associated with air quality concerns.

The utilization of skeleton data and Long Short-Term Memory (LSTM) networks for behavior detection has proven to be a valuable asset in our safety-enhancing system. We demonstrated that by continuously monitoring workers and assessing their movements, we can identify suspicious behaviors such as excessive body scratching, fainting, and persistent coughing. This real-time detection allows for prompt responses to potential health issues, contributing to a safer working environment.

Additionally, our dynamic system integrates air quality assessment into the decision-making process. By evaluating air quality conditions and contrasting them with observed behaviors, the system can issue alarms when necessary, prioritizing the immediate response to fainting incidents or abnormal air quality conditions.

This research underscores the significance of proactive safety measures, bridging the gap between behavior detection and air quality monitoring. By focusing on early detection and response, we aim to reduce the potential health risks faced by workers in manufacturing environments.

6.2 Future Work:

As we move forward, there are several avenues for future research and improvements:

1. **Disease Detection and Prediction:** Expand the system's capabilities to not only detect suspicious behaviors but also predict potential diseases based on behavioral patterns and air quality conditions. This proactive approach can contribute to early disease diagnosis.
2. **Integration of Environmental Data:** Incorporate additional environmental data, such as particulate matter (PM) concentrations, into air quality assessments for more comprehensive monitoring.
3. **Real-Time Notifications:** Develop a real-time notification system that can alert workers and supervisors to potential risks and abnormalities, ensuring immediate actions can be taken.
4. **Machine Learning Enhancements:** Explore advanced machine learning techniques to improve the system's accuracy and predictive capabilities for both behavior detection and air quality assessment.
5. **Validation and Testing:** Conduct extensive validation and testing in real manufacturing environments to fine-tune the system's performance and adapt it to different scenarios.
6. **Cost-Effective Sensor Integration:** Investigate cost-effective sensor solutions that can be easily integrated into manufacturing facilities, making this technology accessible to a wider range of industries.

In conclusion, the journey of enhancing safety and security in manufacturing environments is ongoing. Our work has laid the foundation for a more proactive and responsive approach, and future research and advancements will contribute to the well-being of workers and the overall efficiency of industrial operations.

6.3 List of Papers

International journals:

1) R’bigui, S.; R’bigui, H.; and Chiwoon. “Short Path Wind-Field Distance based Lagrangian Trajectory Model for Enhancing Atmospheric Dispersion prediction Accuracy” IEEE Access, South Korea, published on 28 September 2023.

International Conferences :

2) R’bigui, S.; R’bigui, H.; and Chiwoon. “Real-time systems for air quality forecasting: A review of sensor networks, data fusion, and modeling approaches” International Conference on Smart Computing and Cyber Security, Lecture Notes in Networks and Systems, South Korea, Accepted and presented on the 29th of June, 2023, considered to be published in Springer proceeding of Lecture Notes in Networks and Systems by December 2023.

References

[1] Nale, R., Sawarbandhe, M., Chegogoju, N., & Satpute, V. (2021, September). Suspicious human activity detection using pose estimation and lstm. In 2021 International symposium of Asian control Association on Intelligent Robotics and Industrial Automation (IRIA) (pp. 197-202). IEEE.

[2] Miao, Y., Yang, J., Alzahrani, B., Lv, G., Alafif, T., Barnawi, A., & Chen, M. (2022). Abnormal Behavior Learning Based on Edge Computing toward a Crowd Monitoring System. IEEE Network, 36(3), 90-96.

[3] Zhao, Q., Gu, Q., Gao, Z., Shao, S., & Zhang, X. (2023). Building Indoor Dangerous Behavior Recognition Based on LSTM-GCN with Attention Mechanism. CMES-Computer Modeling in Engineering & Sciences, 137(2).

[4] Chen, Z., Wu, L., He, H., Jiao, Z., & Wu, L. (2022). Vision-based skeleton motion phase to evaluate working behavior: case study of ladder climbing safety. Human-centric Computing and Information Sciences, 12.

[5] Li, Z., Zhang, A., Han, F., Zhu, J., & Wang, Y. (2023). Worker Abnormal Behavior Recognition Based on Spatio-Temporal Graph Convolution and Attention Model. Electronics, 12(13), 2915.

[6] Jebur, S. A., Hussein, K. A., Hoomod, H. K., Alzubaidi, L., & Santamaría, J. (2022). Review on deep learning approaches for anomaly event detection in video surveillance. Electronics, 12(1), 29.

[7] Bae, H. J., Jang, G. J., Kim, Y. H., & Kim, J. P. (2021). LSTM (long short-term memory)-based abnormal behavior recognition using AlphaPose. KIPS Transactions on Software and Data Engineering, 10(5), 187-194.

[8] Park, J. H., Salim, M. M., Jo, J. H., Sicato, J. C. S., Rathore, S., & Park, J. H. (2019). CIoT-Net: a scalable cognitive IoT based smart city network architecture. Human-centric Computing and Information Sciences, 9(1), 1-20.

[9] Shah, R., & Chircu, A. (2018). IoT and AI in healthcare: A systematic literature review. Issues in Information Systems, 19(3).

[10] Han, J., Shao, L., Xu, D., & Shotton, J. (2013). Enhanced computer vision with microsoft kinect sensor: A review. IEEE transactions on cybernetics, 43(5), 1318-1334.

[11] Kingsy Grace, R., Manju, S. A Comprehensive Review of Wireless Sensor Networks Based Air Pollution Monitoring Systems. Wireless Pers Commun 108, 2499–2515 (2019). https://doi.org/10.1007/s11277-019-06535-3.

[12] Maduranga, M. W. P., Kosgahakumbura, K. N. M. H. H., & Karunarathna, G. H. C. J. (2020). Design of an IoT based indoor air quality monitoring system.

[13] Q. P. Ha, S. Metia and M. D. Phung, "Sensing Data Fusion for Enhanced Indoor Air Quality Monitoring," in IEEE Sensors Journal, vol. 20, no. 8, pp. 4430-4441, 15 April15, 2020, doi: 10.1109/JSEN.2020.2964396.

[14] Zhao, L., Yang, Y., & Wu, Z. (2022). Review of Communication Technology in Indoor Air Quality Monitoring System and Challenges. Electronics, 11(18), 2926.

[15] Zhao, Z., Wang, J., Fu, C., Liu, Z., Liu, D., & Li, B. (2018). Design of a smart sensor network system for real-time air quality monitoring on green roof. Journal of Sensors, 2018.

[16] Zervas, E., Mpimpoudis, A., Anagnostopoulos, C., Sekkas, O., & Hadjiefthymiades, S. (2011). Multisensor data fusion for fire detection. Information Fusion, 12(3), 150-159.

[17] More, N., Nikam, V. B., & Banerjee, B. (2020). Machine learning on high performance computing for urban greenspace change detection: satellite image data fusion approach. International Journal of Image and Data Fusion, 11(3), 218-232.

[18] Lin, Y. C., Chi, W. J., & Lin, Y. Q. (2020). The improvement of spatial-temporal resolution of PM2. 5 estimation based on micro-air quality sensors by using data fusion technique. Environment international, 134, 105305.

[19] Criado, A., Armengol, J. M., Petetin, H., Rodriguez-Rey, D., Benavides, J., Guevara, M., ... & Jorba, O. (2023). Data fusion uncertainty-enabled methods to map street-scale hourly NO 2 in Barcelona: a case study with CALIOPE-Urban v1. 0. Geoscientific Model Development, 16(8), 2193-2213.

[20] Gressent, A., Malherbe, L., Colette, A., Rollin, H., & Scimia, R. (2020). Data fusion for air quality mapping using low-cost sensor observations: Feasibility and addedvalue. Environment international, 143, 105965.

[21] Hu, K., Guo, X., Gong, X., Wang, X., Liang, J., & Li, D. (2022). Air quality prediction using spatio-temporal deep learning. Atmospheric Pollution Research, 13(10), 101543.

[22] Zou, Y., O'Neill, S. M., Larkin, N. K., Alvarado, E. C., Solomon, R., Mass, C., ... & Shen, H. (2019). Machine learning-based integration of high-resolution wildfire smoke simulations and observations for regional health impact assessment. International journal of environmental research and public health, 16(12), 2137.

[23] Shi, G., Leung, Y., Zhang, J. S., Fung, T., Du, F., & Zhou, Y. (2021). A novel method for identifying hotspots and forecasting air quality through an adaptive utilization of spatiotemporal information of multiple factors. Science of The Total Environment, 759, 143513.

[24] Ragazzi, M., Rada, E. C., Chisté, A., Schiavon, M., Ciolli, M., Zambelli, P., ... & Chelodi, M. (2012). A contribution to the development of a public–private integrated network of wireless sensors for an enhanced management of air quality. Sustainable Technology for Environmental Protection, 1-4.

[25] Ye, D., Gong, D., & Wang, W. (2009, December). Application of wireless sensor networks in environmental monitoring. In 2009 2nd international conference on power electronics and intelligent transportation system (PEITS) (Vol. 1, pp. 205-208). IEEE.

[26] Jo, B. W., & Khan, R. M. A. (2017). An event reporting and early-warning safety system based on the internet of things for underground coal mines: A case study. Applied Sciences, 7(9), 925.

[27] Mirzaei, A., Bagheri, H., & Sattari, M. (2023). Data level and decision level fusion of satellite multi-sensor AOD retrievals for improving PM2. 5 estimations, a study on Tehran. Earth Science Informatics, 1-19.

[28] Wang, S., Wang, P., Qi, Q., Wang, S., Meng, X., Kan, H., ... & Zhang, H. (2023). Improved estimation of particulate matter in China based on multisource data fusion. Science of The Total Environment, 161552.

[29] Cai, J., Dai, X., Hong, L., Gao, Z., & Qiu, Z. (2020). An air quality prediction model based on a noise reduction self-coding deep network. Mathematical Problems in Engineering, 2020, 1-12.

[30] Zhang, Y., Wang, Y., Gao, M., Ma, Q., Zhao, J., Zhang, R., ... & Huang, L. (2019). A predictive data feature exploration-based air quality prediction approach. IEEE Access, 7, 30732-30743.

[31] Wu, Q., & Lin, H. (2019). A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors. Science of the Total Environment, 683, 808-821.

[32] Liu, H., Yin, S., Chen, C., & Duan, Z. (2020). Data multi-scale decomposition strategies for air pollution forecasting: A comprehensive review. Journal of Cleaner Production, 277, 124023.

[33] Zhang, Z., Zeng, Y., & Yan, K. (2021). A hybrid deep learning technology for PM 2.5 air quality forecasting. Environmental Science and Pollution Research, 28, 39409-39422.

[34] Xu, Y., Huang, Y., & Ma, G. (2020). A beetle antennae search improved BP neural network model for predicting multi-factor-based gas explosion pressures. Journal of Loss Prevention in the Process Industries, 65, 104117.
[35] Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.
[36] Russell, S. (2016). Artificial Intelligence: A Modern Approach, eBook, Global Edition. Pearson Education, Limited.
[37] Hao, P. Y., Chiang, J. H., & Chen, Y. D. (2022). Possibilistic classification by support vector networks. Neural Networks, 149, 40-56.
[38] Provenza, M. M., Luna, A. S., & Xavier, V. L. (2022). Evaluating How the Social Restriction, the Government Response, the Health, and Economic Indices Affected the Prediction of the Number of Deaths Provoked by COVID-19 in Brazil Using Classical Statistical and Machine Learning Models. Brazilian Archives of Biology and Technology, 66.
[39] Brockwell, P. J., & Davis, R. A. (Eds.). (2002). Introduction to time series and forecasting. New York, NY: Springer New York.
[40] Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences, 55(1), 119-139.
[41] Breiman, L. (1996). Bagging predictors. Machine learning, 24, 123-140.
[42] Wolpert, D. H. (1992). Stacked generalization. Neural networks, 5(2), 241-259.
[43] Park, J., & Chang, S. (2021). A Particulate Matter Concentration Prediction Model Based on Long Short-Term Memory and an Artificial Neural Network. International *Journal of Environmental Research and Public Health*, 18(13), 6801. https://doi.org/10.3390/ijerph18136801.
[44] Chuluunsaikhan, T., Heak, M., Nasridinov, A., & Choi, S. (2021). Comparative Analysis of Predictive Models for Fine Particulate Matter in Daejeon, South Korea. Atmosphere, 12(10), 1295.
[45] Zhang, F., Cheng, H., Wang, Z., Lv, X. R., Zhu, Z., Zhang, G., & Wang, X. F. (2014). Fine particles (PM2.5) at a CAWNET background site in Central China: Chemical compositions, seasonal variations and regional pollution events. Atmospheric Environment, 86, 193–202. https://doi.org/10.1016/j.atmosenv.2013.12.008.
[46] Yang, Q., Yuan, Q., Li, T., Shen, H., & Zhang, L. (2017). The Relationships between PM2.5 and Meteorological Factors in China: Seasonal and Regional Variations. International Journal of Environmental Research and Public Health, 14(12), 1510. https://doi.org/10.3390/ijerph14121510.
[47] Bei, W., & Li, F. (2022b). PM2.5 concentration prediction using deep learning in internet of things air monitoring system. Environmental Engineering Research, 28(1), 210456–0. https://doi.org/10.4491/eer.2021.456
World Health Organization: Global Health Observatory (GHO) Data for Ambient Air Pollution. Available online: www.who.int/gho/phe/outdoor_air_pollution/en/ (accessed on 12 January 2020).
[48] Kasznia-Kocot, J., Kowalska, M., Górny, R. L., Niesler, A., & Wypych-Slusarska, A. (2010). Environmental risk factors for respiratory symptoms and childhood asthma. Annals of Agricultural and Environmental Medicine, 17(2), 221-229.
[49] Qadeer, K., Rehman, W. U., Sheri, A. M., Park, I., Kim, H. K., & Jeon, M. (2020). A long short-term memory (LSTM) network for hourly estimation of PM2. 5 concentration in two cities of South Korea. Applied Sciences, 10(11), 3984.
[50] Li, J., Li, X., Wang, K. L., & Cui, G. (2021). Atmospheric PM2.5 concentration prediction and noise estimation based on adaptive unscented Kalman filtering. Measurement & Control, 54(3–4), 292–302. https://doi.org/10.1177/0020294021997491
[51] Kavousi-Fard, A., Samet, H., & Marzbani, F. (2014). A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting. Expert systems with applications, 41(13), 6047-6056.
[52] Wen, C.; Liu, S.; Yao, X.; Peng, L.; Li, X.; Hu, Y.; Chi, T. A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. Sci. Total Environ. 2019, 654, 1091–1099.
[53] Wang, J.; Du, P.; Hao, Y.; Ma, X.; Niu, T.; Yang, W. An innovative hybrid model based on outlier detection and correction algorithm and heuristic intelligent optimization algorithm for daily air quality

index forecasting. J. Environ. Manag. 2020, 225, 109855.
[54] Cha, J.; Kim, J. Development of Data Mining Algorithm for Implementation of Fine Dust Numerical Prediction Model. KIICE 2018, 22, 595–601
[55] Oh, B. D., Song, H. J., Kim, J. D., Park, C. Y., & Kim, Y. S. (2019). Predicting Concentration of PM10 Using Optimal Parameters of Deep Neural Network. Intelligent Automation & Soft Computing, 25(2).
[56] Zhao, Y. (2013). Machine learning algorithms for predicting roadside fine particulate matter concentration level in Hong Kong Central. Computational Ecology and Software, 3(3), 61.
[57] Ulfah, S., Awalludin, S. A., & Wahidin. (2018). Advection-diffusion model for the simulation of air pollution distribution from a point source emission. Journal of Physics, 948, 012067. https://doi.org/10.1088/1742-6596/948/1/012067
[58] Lakshminarayanachari, K., Pai, K. S., Prasad, M. S., & Pandurangappa, C. (2013). Advection-diffusion numerical model of air pollutants emitted from an urban area source with removal mechanisms by considering point source on the boundary. International Journal of Application or Innovation in Engineering & Management, 2, 251-268.
[59] Arystanbekova, N. K. (2004). Application of Gaussian plume models for air pollution simulation at instantaneous emissions. Mathematics and Computers in Simulation, 67(4-5), 451-458.
[60] Zhao, C., & Song, G. (2017). Application of data mining to the analysis of meteorological data for air quality prediction: A case study in Shenyang. IOP Conference Series, 81, 012097. https://doi.org/10.1088/1755-1315/81/1/012097
[61] Jeong, J. I., Park, R. J., Woo, J. H., Han, Y. J., & Yi, S. M. (2011). Source contributions to carbonaceous aerosol concentrations in Korea. Atmospheric environment, 45(5), 1116-1125.
[62] Huilier, D. (2021). An Overview of the Lagrangian Dispersion Modeling of Heavy Particles in Homogeneous Isotropic Turbulence and Considerations on Related LES Simulations. Fluids, 6(4), 145. https://doi.org/10.3390/fluids6040145
[63] Wilson, J.D.; Sawford, B.L. Review of Lagrangian stochastic models for trajectories in the turbulent atmosphere. Bound. Layer Meteorol. 1996, 78, 191–210.
[64] Wang, L.-P.; Stock, D.E. Stochastic trajectory models for turbulent diffusion: Monte Carlo process versus Markov chains. Atmospheric Environment. Part A. Gen. Top. 1992, 26, 1599–1607.
[65] Shuen, J.-S.; Chen, L.-D.; Faeth, G.M. Evaluation of a stochastic model of particle dispersion in a turbulent round jet. AIChE J. 1983, 29, 167–170.
[66] Reynolds, A.M. On the Formulation of Lagrangian Stochastic Models for Heavy-Particle Trajectories. J. Colloid Interface Sci. 2000, 232, 260–268.
[67] Launay, K.; Huilier, D.; Burnage, H. An improved Lagrangian method for predicting the long-time turbulent dispersion in gas-particle flows. In Proceedings of the ASME Summer Fluids Engineering Meeting, FEDSM98-5012, Washington, DC, USA, 21–25 June 1998.
[68] Toschi, F.; Bodenschatz, E. Lagrangian Properties of Particles in Turbulence. Annu. Rev. Fluid Mech. 2009, 41, 375–404.
[69] Stockie, J. M. (2011). The mathematics of atmospheric dispersion modeling. Siam Review, 53(2), 349-372.
[70] Li, Y., Tong, D. Q., Ngan, F., Cohen, M. D., Stein, A. F., Kondragunta, S., ... & Kahn, R. A. (2020). Ensemble PM2. 5 forecasting during the 2018 camp fire event using the HYSPLIT transport and dispersion model. Journal of Geophysical Research: Atmospheres, 125(15), e2020JD032768.
[71] Fedorova, N., Levit, V., Da Silva, A. O., & dos Santos, D. M. B. (2013). Low visibility formation and forecasting on the northern coast of Brazil. Pure and Applied Geophysics, 170, 689-709.
[72] Guang, J., Xue, Y., Mei, L., Li, Y., Wang, Y., Xu, H., & Guo, J. (2010, July). Forecasting air quality by integration of satellite data and hysplit trajectory model. In 2010 IEEE International Geoscience and Remote Sensing Symposium (pp. 2773-2776). Ieee.
[73] Donnelly, A., Naughton, O., Broderick, B., & Misstear, B. (2017). Short-term forecasting of nitrogen dioxide (NO 2) levels using a hybrid statistical and air mass history modelling approach. Environmental Modeling & Assessment, 22, 231-241.
[74] Li, L., Gong, J., & Zhou, J. (2014). Spatial interpolation of fine particulate matter concentrations using the shortest wind-field path distance. PloS one, 9(5), e96111.

[75] Consortini, A., Innocenti, C., & Paoli, G. (2002). Estimate method for outer scale of atmospheric turbulence. Optics communications, 214(1-6), 9-14.
[76] Moon, E. B., Jeon, T. I., & Grischkowsky, D. R. (2015). Long-path THz-TDS atmospheric measurements between buildings. IEEE Transactions on Terahertz Science and Technology, 5(5), 742-750.
[77] Ye, Q., Upshur, M. A., Robinson, E. S., Geiger, F. M., Sullivan, R. C., Thomson, R. J., & Donahue, N. M. (2018). Following particle-particle mixing in atmospheric secondary organic aerosols by using isotopically labeled terpenes. Chem, 4(2), 318-333.
[78] Park, S., Shin, M., Im, J., Song, C. K., Choi, M., Kim, J., ... & Kim, S. K. (2019). Estimation of ground-level particulate matter concentrations through the synergistic use of satellite observations and process-based models over South Korea. Atmospheric Chemistry and Physics, 19(2), 1097-1113.
[79] Dunn, T., Bacon, D. P., & Boris, P. (1994). Development of the Operational Multiscale Environment Model with Grid Adaptivity (OMEGA) and its Aerosol Transport and Diffusion Model (ATDM) (No. CONF-9405167-). Air and Waste Management Association, Pittsburgh, PA (United States).
[80] Dunn, T. J., Bacon, D. P., Lee, P. C., Hall, M. S., & Sarma, A. Parallelization Development and Testing of the Operational Multiscale Environment Model With Grid Adaptivity (OMEGA). APPLICATION BRIEFS 2002, 808, 14.
[81] Appel, K. W., Napelenok, S. L., Foley, K. M., Pye, H. O., Hogrefe, C., Luecken, D. J., ... & Young, J. O. (2017). Description and evaluation of the Community Multiscale Air Quality (CMAQ) modeling system version 5.1. Geoscientific model development, 10(4), 1703-1732.
[82] Wyat Appel, K., Napelenok, S., Hogrefe, C., Pouliot, G., Foley, K. M., Roselle, S. J., ... & Mathur, R. (2018). Overview and evaluation of the community multiscale air quality (CMAQ) modeling system version 5.2. In Air Pollution Modeling and its Application XXV 35 (pp. 69-73). Springer International Publishing.
[83] Berg, L. K., Shrivastava, M., Easter, R. C., Fast, J. D., Chapman, E. G., Liu, Y., & Ferrare, R. A. (2015). A new WRF-Chem treatment for studying regional-scale impacts of cloud processes on aerosol and trace gases in parameterized cumuli. Geoscientific Model Development, 8(2), 409-429.
[84] Chang, X., Wang, S., Zhao, B., Xing, J., Liu, X., Wei, L., ... & Zheng, M. (2019). Contributions of inter-city and regional transport to PM2. 5 concentrations in the Beijing-Tianjin-Hebei region and its implications on regional joint air pollution control. Science of the Total Environment, 660, 1191-1200.
[85] Chen, B., Stein, A. F., Castell, N., de la Rosa, J. D., de la Campa, A. M. S., Gonzalez-Castanedo, Y., & Draxler, R. R. (2012). Modeling and surface observations of arsenic dispersion from a large Cu-smelter in southwestern Europe. Atmospheric Environment, 49, 114-122.
[86] Stohl, A., Forster, C., Frank, A., Seibert, P., & Wotawa, G. (2005). The Lagrangian particle dispersion model FLEXPART version 6.2. Atmospheric Chemistry and Physics, 5(9), 2461-2474.
[87] Draxler, R. R., & Hess, G. D. (1998). An overview of the HYSPLIT_4 modelling system for trajectories. Australian meteorological magazine, 47(4), 295-308.
[88] Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J., Cohen, M. D., & Ngan, F. (2015). NOAA's HYSPLIT atmospheric transport and dispersion modeling system. Bulletin of the American Meteorological Society, 96(12), 2059-2077.
[89] Stohl, A., Forster, C., Frank, A., Seibert, P., & Wotawa, G. (2005). The Lagrangian particle dispersion model FLEXPART version 6.2. Atmospheric Chemistry and Physics, 5(9), 2461-2474.
[90] Pisso, I., Sollum, E., Grythe, H., Kristiansen, N. I., Cassiani, M., Eckhardt, S., ... & Stohl, A. (2019). The Lagrangian particle dispersion model FLEXPART version 10.4. Geoscientific Model Development, 12(12), 4955-4997.
[91] Park, I., & Seo, I. W. (2018). Modeling non-Fickian pollutant mixing in open channel flows using two-dimensional particle dispersion model. Advances in water resources, 111, 105-120.
[92] Fast, J. D., & Easter, R. C. (2006, June). A Lagrangian particle dispersion model compatible with WRF. In 7th Annual WRF User's Workshop (pp. 19-22).
[93] Arunachalam, S., Baek, B. H., Holland, A., Adelman, Z., Binkowski, F. S., Hanna, A., ... & Soucacos, P. (2008, January). An improved method to represent aviation emissions in air quality modeling systems and their impacts on air quality. In Proceedings of the 13 Conference on Aviation, Range and Aerospace Meteorology, New Orleans, LA (Vol. 135626).

[94] Scire, J. S., Strimaitis, D. G., & Yamartino, R. J. (2000). A user's guide for the CALPUFF dispersion model. Earth Tech, Inc, 521, 1-521.
[95] Gulia, S., Kumar, A., & Khare, M. (2015). Performance evaluation of CALPUFF and AERMOD dispersion models for air quality assessment of an industrial complex.
[96] Im, U., Bianconi, R., Solazzo, E., Kioutsioukis, I., Badia, A., Balzarini, A., ... & Galmarini, S. (2015). Evaluation of operational on-line-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part I: Ozone. Atmospheric environment, 115, 404-420.
[97] Sinha, V., Kumar, V., & Sarkar, C. (2014). Chemical composition of pre-monsoon air in the Indo-Gangetic Plain measured using a new air quality facility and PTR-MS: high surface ozone and strong influence of biomass burning. Atmospheric Chemistry and Physics, 14(12), 5921-5941.
[98] Gowardhan, A., Spoon, H., Riechers, D. A., González-Alfonso, E., Farrah, D., Fischer, J., ... & Bizzocchi, L. (2018). The dual role of starbursts and active galactic nuclei in driving extreme molecular outflows. The Astrophysical Journal, 859(1), 35.
[99] Seinfeld, J. H., & Pandis, S. N. (2016). Atmospheric chemistry and physics: from air pollution to climate change. John Wiley & Sons.
[100] Lee, S. J., Lee, H. Y., Kim, S. J., Kang, H. J., Kim, H., Seo, Y. K., ... & Choi, S. D. (2023). Pollution characteristics of PM2. 5 during high concentration periods in summer and winter in Ulsan, the largest industrial city in South Korea. Atmospheric Environment, 292, 119418.
[101] Ulsan Metropolitan City. (n.d.). Real-time air quality monitoring data. Retrieved from http://www.ulsan.go.kr/s/uihe/realtime/area/index.ulsan?area=238121&mId=001003005001002000&datatype=data&group=0&term=30
[102] Choi, W. J., Jung, B., Lee, D., Kang, H., Kim, H., & Hong, H. (2021). An Investigation into the Effect of Emissions from Industrial Complexes on Air Quality in the Ulsan Metropolitan City Utilizing Trace Components in PM2. 5. Applied Sciences, 11(21), 10003.
[103] Air Korea. (n.d.). Station information. Retrieved April 27,2023,from https://www.airkorea.or.kr/eng/stationInformation?pMENU_NO=158
[104] https://www.airkorea.or.kr/web/images/station_map/238120.gif
[105] Korea Meteorological Administration. (n.d.). Automated Weather System (AWS). Retrieved May 4, 2023, from https://web.kma.go.kr/eng/weather/aws.jsp.