ABSTRACT

Machine Learning models and Deep Learning models have been widely used to predict the air quality. Monitoring air quality involves both regulatory measures and public awareness campaigns to reduce emissions from various sources such as vehicles, industrial activities, agriculture, and household combustion. Air pollution predict is very useful for informing about the pollution level that allow policy makers to adopt measures for reducing its impact.

Over the past few decades, due to human activities, industrialization, and urbanization, air quality condition has become a life-threatening factor in many countries around the world. It causes various illnesses such as respiratory tract and cardiovascular diseases. Hence, it is necessary to accurately predict the PM2.5 concentrations in order to prevent the citizens from the dangerous impact of air pollution. Air pollution refers to the presence of harmful or excessive quantities of substances in the air we breathe, which can be detrimental to human health, the environment and ecosystems. These substances, known as pollutants, can come from various sources, including industrial activities, vehicle emissions, agricultural practices, and natural phenomena. Common air pollutants include particulate matter (PM), nitrogen oxides (NO_x), sulfur dioxide (SO_2), carbon monoxide (CO), volatile organic compounds (VOCs), and ozone (O_3). The air quality prediction is used to predict the future state of air quality in a particular location based on the existing data, such as historical air quality data.

In the first phase of the research work, an Improved Sparse Auto Encoder-Deep Learning Algorithm (ISAE-DL) is used to predict the air quality system and the feed forward neural network is utilized as a sparse auto encoder. The combined k-Nearest Neighbor Euclidean Distance (kNN-ED) and kNN- Dynamic Time Warping Distance (kNN-DTWD) is used to acquire the particulate matter and the meteorological data. In addition, Artificial Neural Network (ANN) and Long Short-Term Memory (LSTM) are used to acquire the relative information and the classification model is generated with training data.

In the second phase of research work, Voronoi Clustering Sparse Auto Encoder- Deep Learning (VCSAE-DL) is developed to handle the long time delay based locations for better air quality prediction. Then, the temporal and spatial features are identified to retrieve the most important features for air quality prediction. The formation of clusters is continued with different centers and the clustering process is stopped until all the data are covered. The

clustered data and the terrain information are given as input to the Neural Network layer such as Artificial Neural Network (ANN), Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) and their results are combined and transferred to Sparse Auto encoder for the prediction of air quality. This method efficiently reduces the long-term delay issues, but this method can also suffer to learn from the long-term dependencies of air pollutant concentrations.

In the third phase of research work, a Transferred Stacked Bidirectional and Unidirectional Long Short-Term memory (T-SBU-LSTM) is proposed to minimize the long term dependencies for LSTM for air quality prediction. Then, the Transferred Stacked Bidirectional and Unidirectional LSTM (T-SBU-LSTM) was adopted in learning from long-term PM2.5 dependencies, and it uses Transfer learning to transfer knowledge from smaller temporal resolutions to higher temporal resolutions. Transfer learning is used to improve prediction accuracy at higher temporal resolutions which identifies the similarities between two separate datasets, tasks, or models to transmit data from the source to the new domain. This combined architecture enhances the feature learning from the large-scale spatial-temporal time series data by learning both forwards and backward dependencies. This phase of research expands the air quality prediction from a specific location to several adjacent locations varying small period to long period time delays.

In the fourth phase of work, Wasserstein Distance - Deep Transfer Learning (WD-DTL) is proposed to reduce the learning time of Transfer Learning. Then, the Wasserstein distance based Deep Transfer Learning (WD-DTL) is constructed to learn invariant features between source and target domains. Initially, a base LSTM model is trained with sufficient data in source domain.

Finally, the developed approaches like Improved Sparse Auto Encoder Using Deep Learning (ISAE-DL), Voronoi Clustering Sparse Auto Encoder Using Deep Learning (VCSAE-DL), Transferred Stacked Bidirectional and Unidirectional Using Long Short Term Memory Algorithm (T-SBU-LSTM) and Wasserstein distance using Deep Transfer Learning (WD-DTL) based air quality prediction system were compared using the performance metrics, Accuracy, Precision, Specificity, Sensitivity, AUC, MCC and MAER. The experimental results proved that WD-DTL based air quality prediction system accomplishes better than the other prediction algorithms.