CHAPTER 2

LITERATURE SURVEY

2.1 INTRODUCTION

Several works have been proposed for the Air quality Prediction. Different algorithms have been developed overtime, and the Machine learning Techniques, Deep Learning Techniques, Transfer learning are discussed in this chapter.

2.2. MACHINE LEARNING METHODS

Cosma et al., (2018) proposed Machine Learning methods such as Support Vector Machine (SVM) classifier using a human-centric control system. The method used skin temperature as the sole input and demonstrated good predictive power in identifying fixed heat needs. Although using just one skin temperature accurately predicted 80% of heat demand, it was not applicable to time-series data analysis and exhibited poor accuracy. The approach was not designed to handle time-series data, limiting its ability to account for dynamic changes in heat demand over time, such as daily or seasonal variations

Suleiman et al., (2019) proposed Machine Learning techniques such as Support Vector Machine (SVM), Boosted Trees Regression (BTR) and Artificial Neural Network **(**ANN) to reduce PM10 and PM2.5 pollution levels in a reduction scenario. Data was collected from the AQM site, which included polluting gases and temperature. They employed the principal component analysis method to pre-process the data for feature selection. Various metrics, including RMSE, Coefficient of Efficiency and Coefficient of Correlation were used to assess the project. Both the testing and training datasets were examined, and the results were compared to identify the best outcomes. Researchers also studied the traffic composition of each vehicle type, such as electric cars, diesel cars, buses and coaches. In contrast to the BRT and ANN models, which predicted an increase in PM10 statistics at the majority of sites, the SVM models exhibited distinct prediction behavior. However, they performed similarly in the case of PM2.5. The reason was that the models have overfitted the training set and thus failed to generalize the performance.

Masih et al., (2019) proposed Machine Learning Techniques, including Linear Regression (LR), Neural Networks (NN) and Support Vector Machines (SVM), as well as Ensemble learning algorithms The authors divided the work into two main categories: estimation and forecasting of air pollutants in air quality. The research work indicated that linear regression was suited for pollution estimation, while other algorithms, such as SVM-based approaches were suitable for predicting the air quality index. Forecasting was largely restricted to specific models (NN and SVM) and air pollutants (AQI, PM10, and PM2.5), despite machine learning algorithms registering some of the highest correlation coefficient values.

L.Ma et al., (2020) proposed Machine Learning techniques to determine the concentration of air contaminants. The purpose of their study was to minimize model complexity by reducing the number of model parameters and to increase efficiency and performance by employing a structured regularizer. The researchers also developed an improved methodology for calculating hourly air pollution concentrations using meteorological data from the previous day. By using various regularization strategies, the works presented in this research enabled the identification of an exactly suited model. The authors offered a beneficial regularization approach that required sequential hour prediction models. They demonstrated that these regularizations achieved better performance compared to existing standard regression models and other regularizations.

Liang et al., (2020) proposed multiple linear regression model to analyze the factors affecting air quality. SPSS method was used to associate a phenomenon with multiple factors through the optimal combination of multiple independent variables. It was more effective and realistic to predict or estimate the dependent variables than to predict or estimate only one independent variable. The factors influencing air quality index such as were PM2.5, PM10, SO2, NO2, CO and O3 is used for analysis. Through regression analysis of one year's data samples, the prediction model was obtained and it has been proved that the prediction method was worth popularizing.

Mishra et al., (2020) proposed the Adaboost Algorithm, which modified the data of each training sample (xi, yi) by applying weights w1, w2, …, wN. Initially, each observation received equal consideration from the fundamental learner. Once weights were assigned to each observation, the weak learner could be utilized for prediction. In this method, predictions made by the base learner after the weak learner were more likely to be accurate. This process was carried out iteratively until the tth iteration, after which the limit of the Ti base learning algorithm was reached. The outputs of weak learners could be combined to generate more robust learners, enhancing the ability to predict outcomes.

Li. Ma et al., (2022) proposed Machine Learning Techniques to determine the concentration of air contaminants. The purpose of their study was to minimize the model complexity by reducing the number of model parameters, as well as to increase efficiency or performance by employing a structured regularizer. The researchers developed an improved methodology for calculating hourly air pollution concentrations using previous day meteorological data. Using various regularization strategies, the works offered in this research enabled finding a model that suited the exact requirements. The authors offered a beneficial regularization that required sequential hour prediction models. They showed in the work that the regularizations achieved good performance with respect to existing standard regression models and existing regularization techniques.

Gokulan et al., (2023) proposed Machine Learning Techniques for predicting the Air Quality Index (AQI) and were compared to conventional methods. The AQI for the city of Visakhapatnam, Andhra Pradesh, India, focusing on 12 contaminants and 10 meteorological parameters from July 2017 to September 2022, was evaluated using several Machine learning models, including Light GBM, Random Forest, Cat boost, Adaboost, and XGBoost. The results showed that the Catboost model outperformed other models with an R2 correlation coefficient of 0.9998, a mean absolute error (MAE) of 0.60, a mean square error (MSE) of 0.58, and a root mean square error (RMSE) of 0.76. The Adaboost model was the least effective prediction, with an R2 correlation coefficient of 0.9753. In summary, machine learning predicted AQI with Catboost being the best-performing model for AQI prediction. Training advanced models like Catboost and XGBoost requires substantial computational resources, which might limit their application in resource-constrained environments.

2.3 DEEP LEARNING METHODS

Zheng et al., (2015) proposed data-driven approach that incorporated current and forecasted meteorological data, weather forecasts, and air quality data from one station to other stations within a few hundred kilometers. This model of prediction was relied on four factors: The first method, a linear regression-based temporal predictor, was used to model local factors that influenced air quality, and the second method, a Neural Network- based spatial predictor was used to model global factors. The third method, a meteorological data-driven dynamic aggregator, combined the predictions of the spatial and temporal predictors, and the fourth factor an inflection predictor, was used to detect sudden changes in air quality. By incorporating traffic emission data, meteorological attributes, and PM2.5 concentrations, the model accounts for diverse factors influencing air quality, leading to more robust and accurate predictions.

Tamas et al., (2016) proposed a novel method for detecting pollution peaks by combining ANNs with clustering. Models for predicting the $O3$, $NO2$, and PM10 concentrations were developed using machine learning techniques. Initially, Multilayer Perceptron (MLP) was used alone, then it was hybridized successively with hierarchical clustering and with a combination of self-organizing map and k-means. The dataset was segmented using clustering methods, and then Multilayer Perceptron was trained independently on each subset. Two cities on the western Mediterranean island of Corsica were the focus of this research. An examination of sensitivity was carried out using Receiver Operating Characteristic curves (ROC curves). The models face challenges in handling highly dynamic and nonlinear pollutant variations that are difficult to capture with clustering techniques.

Li, X. et.al., (2016) suggested Spatial Temporal Deep Learning (STDL) to predict air quality for temporal and spatial correlations The extraction of intrinsic air quality features was accomplished by training a Stack Auto Encoder (SAE) model. Unlike traditional time series prediction models, this approach simultaneously and accurately estimated the air quality at all sites, regardless of the season. In terms of air quality prediction, the suggested method was also found to outperform the Spatial Temporal Artificial Neural Network (STANN) and SVR models. Models with an excessively complicated structure were found to be prone to over-fitting, and deeper structures with more than four layers did not provide additional advantages.

Li et al., (2017) proposed a unique LSTM Neural Network Extended (LSTME) model and used for air pollutant concentration prediction which takes into the process of spatiotemporal correlations by design. Historical air pollution data were processed through Long Short-Term Memory (LSTM) layers to extract valuable features, while ancillary data including meteorological data and time stamp data were combined with the proposed model to boost its performance. The technique was also evaluated using Spatial Temporal Deep Learning (STDL) model, Time Delay Neural Network (TDNN) model, Autoregressive Moving Average (ARMA) model, Support Vector Regression (SVR) model, and the conventional NN model, where the LSTME model had a superior performance.

Mehdiyev et al., (2017) suggested a framework based on stacked LSTM Autoencoder Networks Algorithm. After the Time-series data was compressed by LSTM Auto encoders, then it was fed into a Deep Feed Forward Neural Network for classification. The framework was used to analyze sensor time series data in the process sector to assess intermediate product quality and predict the air quality in the next stage. Finer-grained results and a more comprehensive implementation of the data were gained from the model.

Freeman et. al., (2018) proposed a Deep Learning method for the time series air pollution using weather data from a Kuwaiti air monitoring station. A recurrent neural network equipped with LSTM was trained using this method, which was then used to make O_3 predictions at 8 different time points. First-order imputation was used to fill in missing and censored data, taking into consideration seasonal effects for bigger gaps and the sequential influence of previous readings for smaller ones. Pollution exceedances were categorized based on the input parameters, and a decision tree was used to rank the most important features for training. While decision trees are interpretable, integrating them with an LSTM model adds complexity. The combined framework can make it harder to understand how specific predictions are made.

Yang et al., (2018) developed an enhanced model for hourly PM2.5 value prediction and utilizing extended Support Vector Regression. The study region was partitioned into several smaller sections using spatial clustering and spatial variation. A novel Gauss vector weight function technique was employed to compute and choose spatial autocorrelation variables as input features. Additionally, each neighborhood interconnections were mapped using the time-honored SVR technique. Experimental data on PM2.5 concentrations in Beijing were used to demonstrate the proposed method's superiority over competing approaches and its high forecast accuracy and reliability. Only five meteorological criteria were used to anticipate PM2.5 concentrations, and none of the included techniques exhibited high accuracy.

Wang et al., (2018) developed a data-driven strategy for predict air quality that makes use of both past air quality and weather records. A complex Spatial-Temporal Ensemble (STE) model was proposed using three factors. The first factor was a weather- based partitioning technique for an ensemble method. It worked by training a number of separate models and then combining them in real time. The second factor was to generate spatial data as relative stations and relative areas by an analysis of Granger causalities between stations. The third factor was a deep LSTM based temporal predictor that could learn both long term and short-term correlations in air quality. Relative stations and relative areas were employed to add spatial information into the model, and Granger causality was used to determine the spatial correlation between stations. However, when atmospheric pressure was particularly high, air flow was restricted.

Zhu et al., (2018) proposed the Multilayer Perceptron method for predicting hourly air pollution levels using weather data from the day before. This allowed for the utilization of several other regularization techniques in order to find an appropriate model. Various popular regularizations for Multilayer perceptron, including the standard Frobenius norm regularization, the nuclear norm regularization, were compared to propose the prediction models of consecutive hours to be close to each other.

Han et. al., (2018) presented an analysis of NO_2 and O_3 levels and their forecasting using machine learning methods. Data collected from monitoring stations was pre-processed and using Exploratory Data Analysis (EDA), the study identifies trends, seasonal patterns, and correlations of NO_2 and O_3 with factors like traffic, temperature, and humidity. Feature engineering incorporates time-based and meteorological factors to improve the predictive power of the models. The Machine learning algorithm such as Random Forest, Support

Vector Machine, and LSTM are trained and evaluated using metrics like MAE and RMSE. Results indicate that certain models perform well in predicting pollutant levels, with implications for air quality management and health risk mitigation.

Jiang et al., (2019) suggested Hybrid Learning Strategy using pigeon inspired optimization and particle swarm optimization. The original AQI data was decomposed into lower-frequency subseries using Wavelet Packet Decomposition (WPD). The optimization of Extreme Learning Machine (ELM) weights and thresholds was carried out using the Improved Pigeon-Inspired Optimization (IPIO) method and the predictions of subseries were made using the Modified ELM (MELM) approach. Moreover, the anticipated results were partitioned into high, medium-high, medium-low and low-frequency subseries utilizing multidimensional scaling and K-means clustering methods. The final data was obtained by combining the subseries using the ensemble method MELM. Therefore, more precise predictions were necessary for optimal results.

Zhang et al., (2019) constructed a Lite GBM model that incorporated prediction data as one of the data sources for air quality predict. This was aimed at improving prediction accuracy by delving deeper into the predictive data feature and maximizing the utilization of the current spatial data. The sliding window technique was suggested for deep mining of the high-dimensional temporal characteristics, allowing for an increase in training dimensions to millions despite the lack of data. While these approaches were found to be useful, they were deemed inadequate for mining temporal and statistical data characteristics related to air quality. The challenge of processing high- dimensional, large- scale data was addressed by the researchers.

Wen et al., (2019) proposed Spatial Temporal Convolutional Long Short-Term Memory Neural Network extended (C-LSTME) to predict air quality. The model's reliability and accuracy were attributed to several factors, including the incorporation of PM2.5 data from highly linked neighboring stations, the utilization of historical PM2.5 concentration data, meteorological data, and aerosol data, which had been ignored by earlier models but are more connected to PM2.5. Finally, using the same dataset, it was found that the C-LSTME model had higher prediction accuracy than other state-of-the-art models, as measured by the RMSE, MAE, and MAPE. On the other hand, substantially lower standard deviations in Root-Mean-Squared Error (RMSE) were achieved by the C- LSTME model compared to the LSTME model, indicating that the C-LSTME model was considered more accurate and stable over the long run.

Du et al., (2019) proposed a novel deep learning model for air quality (mostly PM2.5) prediction, and a Deep Learning architecture was used to learn the spatial- temporal correlation characteristics and dependencies of multivariate air quality related time series data. The nonlinearity and variability of the underlying multivariate air quality time series data were accounted for by the basic components, One-Dimensional Convolutional Neural Networks (1D-CNNs) and Bi-LSTM. The latter was utilized by the former to glean insights about local trends and geographical correlations to better understand spatial-temporal dependencies. Together, a one-dimensional convolutional neural network (CNN) and a bias-inducing long short-term memory (Bi-LSTM) network formed a jointly hybrid deep learning architecture that could be used to learn shared representation features from multivariate air quality time series data.

Schürholz et al., (2020) introduced a Cutting-Edge context prediction model that combined the user's health status and environmental factors by using a Long Short-Term Memory Deep Neural Network. This method can accurately predict when and where air pollution will occur. Following successful implementation and evaluation in a real- world use case, this model was included into a product called Air Quality Index(AQL).

Lin et al., (2020) built a model for estimating indoor air quality using neuro fuzzy networks. Future air pollutant concentrations and environmental parameters were predicted using a neuro-fuzzy network, which comprised a set of fuzzy rules based on historical time series data. The forecasting technique had fuzzy components due to the unpredictability of the impact factors. The initial step involved the division of the training data into fuzzy clusters and the definition of membership functions based on the estimated means and variances. The recovered fuzzy rules from these fuzzy clusters were then utilized to construct a four-layer fuzzy neural network. Subsequently, it was trained using genetic, particle swarm optimization and steepest descent back propagation methods. The AQI indicated that the air quality was worse and more detrimental to human health.

Xayasouk et al., (2020) suggested that many countries worldwide have poor air quality due to the emission of particulate matter (i.e., PM_{10} and $PM_{2.5}$), which has led to concerns about human health impacts in urban areas. In this study, we developed models to predict fine PM concentrations using long short-term memory (LSTM) and deep autoencoder (DAE) methods, and compared the model results in terms of root mean square error (RMSE). The models were applied to hourly air quality data from 25 stations in Seoul, South Korea, for the period from 1 January 2015, to 31 December 2018. Fine PM concentrations were predicted for the 10 days following this period, at an optimal learning rate of 0.01 for 100 epochs with batch sizes of 32 for LSTM model, and DAEs model performed best with batch size 64. The proposed models effectively predicted fine PM concentrations, with the LSTM model showing slightly better performance.

Ma et. al., (2020) proposed Multivariate analysis and identified the most important factors affecting air quality. Environmental, demographical, economic, climatic, and energy factors were among the 171 aspects collected and analyzed to better understand the dynamics at play. To address this big data challenge, the relationship was modeled and the significance of the variables was quantified using a non-linear machine learning approach called Extreme Gradient Boosting (XG Boost). The performance of XG Boost's algorithm was then compared to the competing models, and its parameters were fine- tuned via Bayesian Optimization.

Fong et al., (2020) proposed a LSTM and RNNs been utilized in the future APS concentration in Macau. APS concentration data and weather data were also utilized. Certain forms of APS had fewer observed data reported by some Macau AQMSs, and the amount of observed data was lower in some AQMSs. The AQMSs with little observational data relied on transfer learning and pre-trained neural networks to expedite the development of a neural network capable of producing accurate predictions.

Samal et al., (2020) developed an information-driven SVR Autoencoder model for real-time trend analysis and dynamic data modeling. Additionally, the multiple imputation technique known as Multi modal Imputation by Chained Equations (MICE) was created, to fill in PM2.5 data gaps using meteorological attributes. The weight was also dynamically adjusted using an LSTM-based Stacked Denoising Autoencoder layer to produce reliable predictions in real-time. However, the incorporation of traffic emission data and meteorological elements was found to anticipate PM2.5 and boost the accuracy of predictions.

Eslami et.al., (2020) proposed Deep CNNs which recommended to use NOx, ozone, and weather data from the previous day. A full year was devoted for testing the model by the researchers. In modeling the interplay of local meteorological and species concentrations in an urban setting, it was demonstrated by this study that the deep learning approach could accurately estimate hourly concentrations. The CNN model accurately anticipated the monthly and daily changes in ozone levels in Seoul.

Cai et al., (2020) proposed a Denoising Autoencoder Deep Network (DAEDN) for current air pollution prediction, as existing methods using an LSTM-based model often suffer from low predictive accuracy. The original monitoring data underwent processing with a noise reduction Autoencoder within an LSTM structure, enhancing the accuracy of predictions regarding future air quality. To address the latency issue inherent in one-way LSTM prediction and to generate more precise predictions, the DAEDN model exclusively utilized LSTM networks with a bidirectional topology. The combination of Denoising Autoencoders and bidirectional LSTM networks significantly increases computational complexity, requiring more resources for both training and deployment.

Mengara et al., (2020) proposed Convolutional particle concentration prediction using a bidirectional LSTM autoencoder model which was trained on a distributed architecture (PM2.5 and PM10). Because of the suggested deep learning approach, the inherent association between contaminants in various areas were understood automatically. Additionally, the model's effectiveness was boosted by factoring in the local weather and pollutant gas data. Layers of 1D-CNNs were utilized to extract local spatial information, and a stacked Bi-LSTM layer was employed to learn the spatiotemporal correlation of air quality particles.

Kaya, et al., (2020) proposed the Deep Flexible Sequential Model to accurately predict the onset of air pollution 4, 12, and 24 hours in advance. The results were achieved by combining the capabilities of a Dropout layer with those of a Convolutional Neural Network and a Long Short-Term Memory. CNN was able to efficiently show off the data's quirks. LSTM excelled at separating causal relationships from correlations in time series data. In the end, the Dropout layer acted as a counterweight while doing sequential modeling. However, during network training, RNN's drawback became obvious due to the Vanishing Gradient or Explosive Gradient problem caused by multiplicative decreases or increase in back-propagated gradients.

Seng et al., (2021) suggested Multi-Output Multi-Index Supervised Learning and to construct a full-fledged prediction model (MMSL). This model improved upon previous attempts to anticipate the presence of airborne particulate matter, gaseous pollutants and meteorological conditions by considering the interaction between pollutants. Hourly concentration data of each station was used to train the model. Determining the origin and temporal distribution of bio-aerosols in the atmosphere was a complex task that was heavily influenced by other air pollutants. The MMSL framework is computationally intensive, requiring significant resources for training and implementation, especially with large datasets.

Zhang et al., (2021) suggested that the PM2.5 concentrations be estimated using a semi supervised model. The method was employing Empirical Mode Decomposition (EMD) and recurrent neural networks trained with bidirectional LSTM (BiLSTM) data. The model was only accepting signal data in the form of PM2.5 time series. The frequency and amplitude information were extracted using EMD, an unsupervised feature learning method. This improved the ability to foresee short-term trends, particularly for unexpected changes. During the training stage with supervised data, BiLSTM was used. Hourly and daily PM2.5 datasets from Beijing were used to test the proposed model's prediction capacity. Additionally, errors tended to accumulate when making predictions across multiple steps with a typical LSTM-based model.

Lin et al., (2021) proposed Two Ensemble Learning Models to accurately predict air pollution levels. These models integrated deep learning algorithms with a variety of features. They initially proposed a Multiple Linear Regression based GRU (MLEGRU) Ensemble learning forecasting model using a Gated Recurrent Unit (GRU) Network architecture to combine different deep learning prediction models (GRU13d,GRUAW13d, GRUAW14d, GRUSS13d, and GRUST14d, for example). To analyze MLEGRU's efficiency, they evaluated real forecasted findings from 67 EPA (Environmental Protection Agency) monitoring stations in Taiwan and data acquired between the years 2013 and 2019. The results are compared with other ensemble methods by using various factors, such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Absolute Error less than 3(AEL3).The results reveal that the proposed MLEGRU model provides a better forecasting accuracy than other ensemble learning methods.

Dairi et al. (2021) proposed a deep learning-driven model to predict air pollutant levels with high accuracy and flexibility. The investigation focused on developing a predictive modeling strategy using the state-of-the-art Integrated Multiple Directed Attention (IMDA) DL architecture. The application was validated through an experiment using air pollution data from four U.S. states, demonstrating the effectiveness of the proposed method for accurately predicting air quality levels.

Heydari et al., (2021) proposed Long Short-Term Memory (LSTM) and a Multi-Verse Optimization Method (MVO). As a forecaster engine, LSTM was utilized to estimate the NO_2 and SO_2 emissions from the Combined Cycle Power Station. The prediction error was reduced by optimizing the LSTM parameters with the MVO technique. In addition, real data from a Combined Cycle Power Station in Kerman, Iran was used to evaluate the performance of the proposed model. During the course of five months, wind speed, air temperature, nitrogen dioxide, and sulfur dioxide levels were tracked every three hours. The model was also evaluated with two sets of input parameters: (1) a set that contained lagged values of the output variables (NO_2 and SO_2) in addition to wind speed and air temperature, and (2) a set that had only lagged values of the output variables (NO2 and SO2). Nonetheless, people's openness to adopting energy-efficient technologies expanded alongside their environmental consciousness and concern.

Mao et al., (2021) proposed a Deep Learning framework utilizing a neural network equipped with a Temporal Sliding LSTM Extended Model (TS-LSTME). Sliding prediction with the appropriate time lag was achieved with the assistance of hourly historical PM2.5 concentrations, meteorological data, and temporal data, using multi-layer Bi-LSTM. The proposed model was used to forecast the average PM2.5 concentration in China, the world's most polluted country, for the following day. With a higher correlation coefficient R2, the proposed model was stable and effective than the MLR, SVR, and LSTME models However, due to the sophisticated pollutant spread mechanism, limitations was identified, including problems such as the large amount of computational work involved, the complexity of the processes involved, and the uncertainty of the parameters.

Janarthanan et al., (2021) proposed The Grey Level Co-occurrence Matrix (GLCM) technique to extract characteristics such as the mean, standard deviation, and so on. An optimized set of retrieved features was obtained through optimization of the features initially extracted. Classification was implemented using the deep learning method of SVR with LSTM model, allowing for the prediction of the AQI at the desired location.

Ditsuhi et al., (2023) proposed grid-based (Bidirectional Convolutional Long Short-Term Memory) and graph-based (Attention Temporal Graph Convolutional Network) algorithms to predict air quality. The methods were implemented on a spatiotemporal combination of air quality, meteorological, and traffic data of the city of Madrid. It was exposed that the two methods could be reused for prediction in other scenarios and different air quality phenomena. The combination of Bi-Conv LSTM and ATGCN requires significant computational resources for training and deployment, which may limit its use in resource-constrained settings.

2.4 TRANSFER LEARNING METHODS

Fang et al., (2013) proposed a Hybrid Deep Migration Learning Strategy based on long and short-term memory (LSTM) And Domain Adversarial Neural Networks (DANN), where the temporal features of the source and target buildings were extracted by LSTM, and DANN was used to find the domain invariant features between the source and target buildings through domain adaptation.

Ma et al., (2019) proposed Transferred Bi-Directional Long Short-Term Memory (TL-BLSTM) model and to predict environmental conditions. The framework's methodology involved the utilization of the bidirectional LSTM model and transfer learning to learn from PM2.5's long-term dependencies. With an RMSE reduction of 36. 85% at a daily resolution and an RMSE reduction of 42.59% at a weekly resolution, it was determined that transfer learning significantly reduced the prediction error of BLSTM for PM2.5 at higher temporal resolutions. As less data is required for training with transfer learning, it was anticipated that it performed better at finer granularities. However, as the error indications show that self-transfer learning-based BLSTM model was unable to outperform the baseline BLSTM model.

Lv et al., (2019) proposed Multiview Transfer Semi-Supervised Learning Estimation (MTSAE) for air quality. The topographical characteristics were utilized to distinguish between built-up and rural areas. Secondly, a transfer regression approach was employed to train the first models with labeled data from other cities. Lastly, to further enhance the current models, a semi-supervised regression strategy was applied to an unlabeled dataset from the target city. The combined transfer and semi-supervised learning methods as well as those that employed spatial information to distinguish between urban and non-urban regions, were shown in numerous trials for predicting air quality. A high computational complexity was result of the repeatedly analyzing the models at each iteration.

Ma et al., (2020) proposed a Transfer Learning-based Bidirectional Long Short-Term Memory (TL-BiLSTM) network to predict the air quality of new stations lacking data. This method transferred the knowledge learned from the existing air quality monitoring stations to the new monitoring stations to improve the prediction accuracy of the new stations.

Fong et al. (2020) proposed a Transfer Learning Model combining LSTM and RNN to predict the concentration of air pollutants. The method inputs the data of all source domain sites into the model for pretraining, then adds the number of network layers to input the data of the target domain to train and predict the air quality of the target domain.

Ma et al., (2020) suggested a stacked Bidirectional Long Short-Term Memory (TLS-BLSTM) network trained with transfer learning and to predict air quality at unmeasured sites. The strategy incorporated cutting-edge deep learning techniques with transfer learning statics to enhance forecasting. The Rolling window technique was used to predict air quality forecasting and build time series samples. Transfer learning improved the performance of the SBLSTM model for air quality prediction at newly constructed monitoring stations that were struggling with a lack of data.

Ma et al. (2020) introduced a Transfer Learning-based Stacked BLSTM (TL-SBLSTM) model. The TL-SBLSTM model, which initially trained the data from an existing station and then fine-tuned the layers from a new station. By freezing the initial layers of the pretrained model, spatial features learned from the existing station were retained while the remaining layers adapt to the new station's conditions. This method effectively transferred spatial knowledge between stations, enabling accurate air quality forecasts with limited data from the new location which improved prediction accuracy.

Zhang et al., (2021) proposed a spatiotemporal model that combined LSTM and Graph Attention Network (GAT) for predicting air quality and improving transfer learning. GAT-LSTM outperformed the conventional air quality prediction models because it more accurately modeled the temporal and geographical correlation of pollutants at all monitoring stations. In the absence of sufficient training data in the target city, the proposed meta-learning technique for GAT-LSTM was able to transfer knowledge efficiently from the cities with sufficient data, resulting in the joint training of an acceptable prediction model.

Dhole et al., (2021) proposed an Ensemble Technique, for Multi-Source Transfer Learning to address the issue of limited data. This method was aimed at offering a cumulative forecast by integrating the knowledge gained from multiple source stations into an individual target station, which was expected to improve the performance of the prediction process. Regrettably, insufficient information resulted in diminished efficiency.

Nishant.et al., (2023) proposed an Unsupervised Transfer Learning approach called Deep Air quality method and demonstrated it for the cities of Accra in Ghana and Africa. The Deep AQ model performed two steps: first, a DL model was trained to estimate annual mean levels at 200m resolution over cities with sufficient training data. Los Angeles and New York City (NYC) were chosen as two candidate cities because of: 1) the availability of labeled data and 2) a wide distribution of levels and associated patterns. In the second step, the Deep AQ model was transferred to Accra in a completely unsupervised setting, i.e. labeled data from Accra was not required during model training. For performance validation, the AQ data collected by our team across 10 fixed sites in Accra for one year. The sites were selected to represent a range of land uses and sources such as road traffic, commercial, industrial, and residential areas, and various neighborhoods with diverse socio-economic classes. The entire Deep AQ model training and transfer learning happens in a single end-to-end regime. The focus on annual mean air quality levels, while useful for long-term assessment, may not capture short-term variations or pollution events that are crucial for real-time decision-making.

2.5 CHAPTER SUMMARY

In this chapter, the various research in the area of the air quality prediction were presented. Several authors discussed the various air quality prediction techniques using Machine learning techniques like, Supervised learning, and Unsupervised learning, etc. Several authors discussed the air quality prediction techniques using deep learning algorithms such as Convolutional neural network, Recurrent Neural Network, Long Short-Term Memory etc. Several authors discussed the various air quality prediction techniques in Transfer learning such as Transfer Learning-Based Bidirectional Long Short-Term Memory (TL-BiLSTM) etc. This chapter provides the contributions of the various works as different air pollution or air quality prediction systems that are used to predict and monitor the air quality prediction.