CHAPTER 5

TRANSFERRED STACKED BIDIRECTIONAL AND UNIDIRECTIONAL LONG SHORT TERM MEMORY ALGORITHM

5.1 TRANSFER LEARNING

A Transfer learning (TL) uses parallels between two different kinds of tasks, data, or models to move knowledge from one domain to another. More accurate predictions over longer time intervals are possible with the proposed Transferred Stacked Bidirectional and Unidirectional Long Short-Term Memory (T-SBU-LSTM) model. This model uses transfer learning, which transfers knowledge from lower to higher temporal resolutions, to learn from long-term dependencies. This hybrid architecture enhances feature learning from large spatiotemporal time series data by learning both forward and backward dependencies. The structure of model transfer approach is given in Figure 5.1.

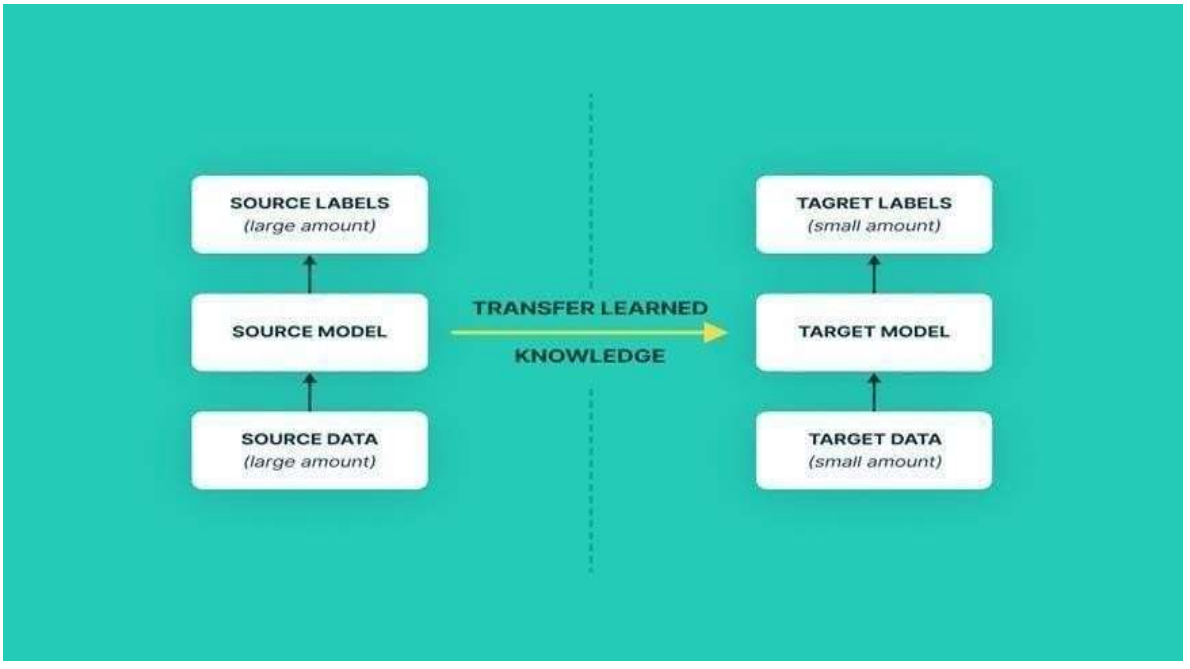

Figure 5.1 Structure of Model Transfer Approach

5.2 PROPOSED METHODOLOGY

5.2.1 Predictive model -Transferred Stacked Bidirectional and unidirectional Long Short-Term Memory(T-SBU-LSTM)

The spatial-temporal analysis is the most significant interconnections between location points investigating sequence latencies and links between locations by observing historical temporal patterns and location feature patterns as potential causes. The neighboring areas or areas with similar temporal patterns, especially if there are strong relationships between them. The associated sites are found with the use of the ISAE-DL and VCSAE-DL connection extractors once the processed data has been fed into the system, and a training dataset is then developed. The LSTM model is used to gather historical time series data from the temporal relationships extractor (TRE)methods.

The proposed predictive model frame work is made up of four primary components. They are target location time, associated location, spatial-temporal data, and terrain data which are incorporated into the proposed T-SBU-LSTM model. The data stream includes patterns in pollutants, weather, and target characteristics over the last few hours, as well as other statistical data from the target and nearby places. The LSTM and Adaptive temporal extractors (ASE), and ANN were trained using these records as inputs. A grid of 121 squares, or 11 X 11 coordinate points spaced every 500 meters was used to store information about the topography, with the central square standing in our actual location. As a result, AQI estimates for 120 previously undiscovered locations are available through the deep adversarial transfer learning framework that makes use of Wasserstein distances. The relative height is concentrated with the undiscovered point AQIs and fed into the CNN to give impact of AQIs at higher altitudes. The inputs to a CNN are used to improve its performance.

In ISAE-DL defined l_{min} Y=6 hour as the minimal time interval for weather forecasts. The ISAE-DL and VCSAE-DL found similar regions, and those pollutants, weather conditions, and target feature(s) are TRE and VCSAE-DL. The three separate methods of Terrain relation extractors (TRE), Spatial relation extractors (SRE), and Terrain extractors (TE) were integrated using an FFNN with two layers. The Predict error isthe difference between the target feature value q and a fixed time t_q $were$ t_q $< fi_t$.

The Proposed model framework of T-SBU- LSTM is shown in Figure 5.2.

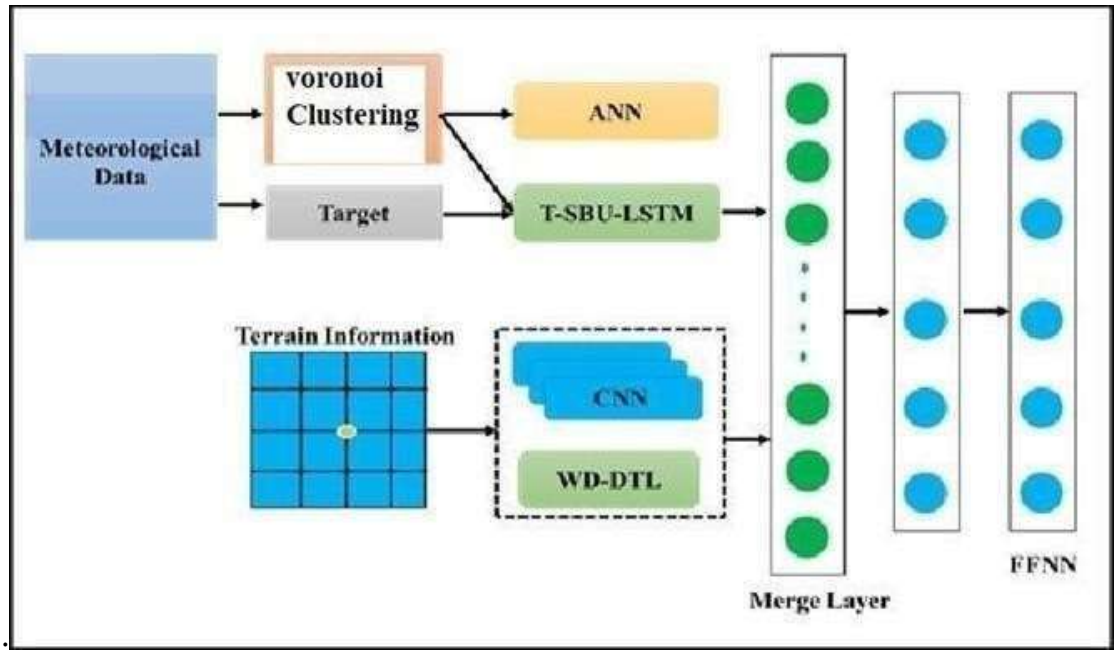

Figure 5.2 Proposed model framework of T-SBU-LSTM

Air quality Information and weather data are transmitted to VCSAE-DL and ASE target variable, while CNN is Privy to information on the landscape. Specifically, humans learn models for each time period by associating inputs with the cyclical changes in the target feature. The uniformity of the input sizes throughout space and time is an advantage of this layout. Air quality characteristic data is extracted from linked sites using neural network methods such as ANN model in the Spatial and Temporal Relationships Extractor (SRE). The Terrain Relation Extractor (TRE) uses an LSTM model to analyze the records from the preceding several hours to extrapolate information about the air quality prediction. The three categories are temporal relation, Spatial Relations, Terrain Extractor.

Temporal Relations: Data on target locations that has already been gathered is fed into the Terrain Relation Extractor. Due to their continuity and coherence, time series input data can be divided into low frequency bandwidth and high bandwidth categories. While the VCSAE-DL behaves as though it were in the past, the T-SBU-LSTM is designed to acquire target location time series patterns; however, because it only uses recent data, the ANN is prone to sudden changes.

Spatial Relations

The interconnectedness of local and regional air quality is referred to as "contaminant diffusion". SRE incorporates historical spatial-temporal features of the surrounding areas as inputs to account for the impact of regional and local pollutants on the data on air quality. Because of this, SRE is built to extrapolate local air quality from a collection of nearby weather stations and AQIs. With the use of circles with different radii and between two points, a large area was divided into smaller sections. The spatial- temporal neighborhood time series properties are stable and consistent, just like the target location.

Terrain Extractor

The barriers of mountains and rivers cause relationships between locations to differ. Therefore, the primary use of terrain data is to enhance location correlations (Seng et al. 2021).The terrain of the area was mapped using a 121-square-section matrix, with 500 meters separating each of its 11*11 coordinate lines. To ascertain whether relationships exist between terrain and time series data, Qin et al. (2015) used a method in which the elevation of each site is normalized as a function of the mean elevation of the surrounding area.

$$El_s = \frac{ele - ele_{st}}{ele_{st}} \quad (5.1)$$

$$ele_{rel} = \frac{1}{e^{El_s}} \quad (5.2)$$

where El_s is Standardized altitude and it mitigates the impact of greater altitudes, even if the distribution is highly sensitive to elevation.

An improved sparse auto encoder collects the outputs of the TE, TRE and TRE feeds them into the FFNN. Layer mixing and outlier identification are both completed by the improved sparse auto encoder. Effective and systematic input data management has been achieved. There is a blending of continuous and categorical data for the prediction of air quality system. This method uses a perceptron-based feed-forward artificial neural network to generate a value and then multiplied by the input data (Lin et al. 2020).

Data in an activation function, linear function, log-sigmoid, hard limit, hyperbolic tangent, and most likely a saturated input are the inputs used by this technique. Value of the performance is determined by:

$$B = (wp + a) \qquad (5.3)$$

where wp is predefined weights

In order to make predictions, perceptron typically use a primary function, and its calculated by,

$$(p) = \frac{1}{(1 + e^{-x})} \qquad (5.4)$$

where e^{-x} is the weight reducing error in the output function

Accurately predicting the weight reduces errors in the output, and the value assumed is determined by the data used for training. The input for the following layer in the hierarchy is the output of one perceptron layer. Complex linear separable classification problems are no match for this multilayer network. The input data is spread using the starting weights (Kurt and Oktay, 2010). Variation in the output of feed forward propagation can be used as a proxy for the error value. Each new iteration of the algorithm is tested thoroughly before the weight is changed. The weights are optimized using the median elevation value, after which the weights are re-probed on each input and the output data is classified accordingly

5.2.2 ARCHITECTURE OF T-SBU-LSTM

In order to facilitate learning from long-term dependencies, the suggested methods use transfer learning to move information from lower to higher temporal resolutions. Long-term dependencies are often handled using the T-SBU-LSTM technique. The output of one hidden layer in a deep LSTM architecture function as the input for the hidden layer after it. The T-SBU-LSTM architecture is shown in Figure 5.3.

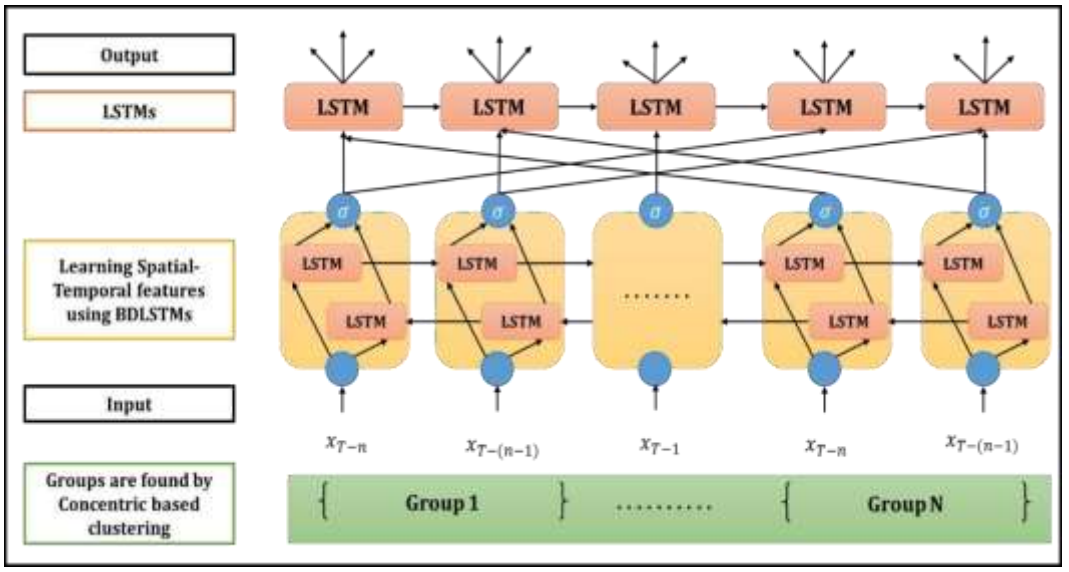

Figure 5.3 Architecture of T-SBU-LSTM

Prediction accuracy at higher temporal resolutions can be enhanced by transfer learning. Using the similarities between two sets of data, transfer learning allows one to one acquired expertise to a new datasets, tasks, or models. The stacked-layers method, used in this study, has been found to increase the effectiveness of neural networks.

By observing both forward and backward sequences of time series data, a boosted RNN termed a bi-directional LSTM (BDLSTM) may discover long-term associations. Predicting air quality using a model created by stacking LSTM and BDLSTM yields SBU-LSTM. The proposed method is well-suited to both long-term and short-term dependencies. Air quality estimation at several nearby locations with variable delay durations is the focus of the next round of research. The integrated architecture learns both forward and backward links to enhance feature learning from massive spatiotemporal time series data.

In transfer learning it is used to learn data from lower to higher resolutions data. The SBU-LSTM employs an LSTM layer as the last layer and a BDLSTM layer for feature learning at the beginning. The inclusion of one or more LSTM/BDLSTM layers in the middle of the model allows the T-SBU-LSTM to extract detailed and intricate features from the input data. The spatial time series data is input layer and T-SBU-LSTM predicts values for forward and backward feature learning of data. The SBU-LSTM predicts for a large number of subsequent time steps based on historical data.

Time sequence features that contain spatial information are used to make predictions about the target area after it selected the regions with the strongest spatial- temporal linkages to it. Hence, the places are always the same, but the time sequences are always different. Hence, the prediction model considers temporal and spatial associations and the interconnected spatial and temporal connection features are used for feature extraction and prediction.

Assuming a set of areas: A={a_1,a_2,…a_n} and set of features: FS = {fs_1,fs_2,…..fs_m}, although latitude and longitude are included in the description of each region, to use the term area coordinate(AC) instead.

$$ACi = (ai, li, mi) \in A \quad (5.5)$$

where Area ai has coordinates in the range of $[li, mi]$, where l_i and m_i are the latitude and longitude, respectively. The separation of two locations in terms of the number of degrees between them, on the assumption that the presence of linking geographical features can enhance forecast accuracy.

$$Dp, q = \text{dist area}\,(ACp, ACq) = \text{dist area}\,((ap, lpmp), (aq, lqmq)), ap, aq \in A\ p\ G\ q \quad (5.6)$$

Finding the most related areas in the Spatial Relationships Sequence (SRSS).

$$SRSS = *D1,2, D1,3 \ldots \ldots . Dn - 1, +, Di, = 0, 0 < i < n + 1 \quad (5.7)$$

where n is represented in the total number of observations, $D_{i,}$ is the zero-based diagonal, and the matrix elements are calculated using equation (5.7). The closest regions are those that make up the SRSS (a,) which is a group of x regions that are closest to ai in terms of spatial distance. Observing the spatial features of these critical spots is important to improve future event prediction. The characteristics of the Feature Sequence Interval (FSI) are

$$(a_i, fs_j, t_{vt,} = \{(a_i, fs_j, t_{vt}), (a_i, fs_i, t_{vt+1}), \ldots . . b(a_i, fs_i, t_{vt}), a_i \in A, fs_j \in FS, v_t < s_t\}) \quad (5.8)$$

where the feature asi of ai varies from beginning to end (vt to st) time ($vt < st$); and the measured value $obbsi$ at is reflected in (a_i, fs_j, t_{vt}). The distance between any two sites' feature sequences can be written as,

$$DSp = distseq((ap, fstarget, tvt, st), F(aq, fstarget, tvt, st)), ap, aq \in A, p \neq q \quad (5.9)$$

Prediction model accuracy can be enhanced by employing transfer learning. The set of n_t samples $Aq_t = \{(a^t,), \ldots \ldots, (a^t, b^t)\}$ selected from some distribution Di, and a source training set of samples, $Aq_s = \{(a^t, b^t), \ldots \ldots, (a^t, b^t)\}$.

Feature vectors from both the original and the final training sets exist in the same underlying feature space and label space r $a_i \in E_n$, and the associated class label $b_i \in \{C1,...,CL\}$.

Due to the variations in the time series, the models trained from the training set will be unable to accurately identify the (target) test date. The training set Aqs does not come from the same distribution as Aqt. Unfortunately, Aqt is frequently too small to train a good classifier on the test data. Using the knowledge from Aqs to acquire the knowledge required to predict targets in Aqt, where *Aqs* is the start time and *Aqt* is the end time, is known as transfer learning.

To begin the metric, a target prediction sequence, denoted by the feature fs target, must be selected. In this case, PM2.5 was used, but other targets are possible. In this way, use equation to calculate the set of temporal relation sequences (TRSS) It is given by equation (5.10).

$$TRSSt\ vt,st = \{DS1,2,tvt,st, \ldots \ldots \ldots DSn-1,n,tvt,st\} \quad (5.10)$$

Then, as TRSS_cand (ai,x), and select the x-point cluster that's most unlike to i. In order to construct the spatial-temporal relations (STR) set, it must take into both linkages at the same time.

The set of points x that are the least similar to point i is then $TRSS\ (ai, x)$,. It considered between the connection in the spatial and temporal features. It is calculated by

$$(ai,x) = (ai,) \cup TRSS \quad (5.11)$$

The union of SRS_cand(li ,k) and T RS_cand(li , k) is used rather than the intersection, to provide a larger number of relationships for the model to learn and since some location behaviors may differ from adjacent locations, the intersection would have fewer candidates (or none), resulting in the loss of useful target features. $(ai,)$ is the formula for the Spatial-Temporal Predictor (STP).

$$(STR_{cand(a,x)})[t_{t,tq}] = (a_i, fs_{target}, tvtF,), t_l < b_q < vt' \leq st' \quad (5.12)$$

sequence set from a predicted target feature sequence, S, containing the target features for the time interval t_{vt} to, which is used to build the system. By comparing t_l and t_q, where ti is the lookback time, that may determine which time series is more comparable and so yield F.

5.3.3 Pseudo code for T-SBU-LSTM

Input: Target Station T; Set of location's coordinate LC, where, number of candidates n

Output: Set of locations L

Step 1: Start the process

Step 2: Assume a set of areas: A={a_1,a_2,…a_n}

Step 3: Assume set of features: FS = {fs_1,fs_2,…..fs_m

Step 4: Compute area coordinate (AC)

$AC_i = (a_i, l_i, m_i)$ *//Let us consider latitude and longitude of area a_i are l_i and m_i*

Step 5: Calculate the distance between two locations

$$D_{p,q} = dist_{area}(AC_p, AC_q) = dist_{area}.(a_p, l_p m_p), (a_q, l_q m_q)/, a_p, a_q \in A$$

//Apply transfer learning for T-SBU-LSTM

Step 6: Initialize air quality data $A\tilde{q}_s = \emptyset$.

Step 7: for l = 1 to L do

7.1 Initialize a multilayer autoencoder MAE^l (W,b)

7.2 Select location based air quality $Aq_t^c{}_1$ from Aq_t

7.3 Train AE^l(W,b) using $Aq_t^c{}_1$

7.4 Select location based air quality $Aq_s^c{}_1$ from Aq_s

7.5 Reconstruct data $Aq^{cl} = \tilde{M}AE^l(Aq^{cl})_s$

7.6 Update the reconstructed data $A\tilde{q}_s \cup Aq_s$

Step 8: end for

Step 9: Find spatial relationships sequence

SRSS ={$D_{1,2}$, $D_{1,3}$, ….. $D_{n-1,n}$}, *//D_{ii}= 0*

// n is the number of spots, and the diagonal values are $D_{ii,}$ are Zero

Step 10: Compute Feature Sequence Interval (FSI)

$$(a_i, fs_j, t_{vt,} = *(a_i, fs_j, t_{vt}), (a_i, fs_i, t_{vt+1}), \ldots\ldots (a_i, fs_i, t_{vt}), a_i \in A, fs_j \in FS, v_t < s_t+)$$

// a_i has feature fs_i that varies from start to finish (vt to st) time (vt<st); and $b(a_i,fs_j,t_x)$ represents the measured value of fs_i at t_x.

Step 11: Find distance between feature sequences

$$DS_{p,,vt,st} = dist_{seq}.F(a_p, fs_{target}, t_{v_t,s_t}), F(a_q, fs_{target}, t_{v_t,s_t})/, a_p, a_q \in A$$

Step 12: Compute temporal relations sequences (TRSS)

$$TRSS_{t_{vt,}} = *DS_{1,2,t_{vt,st}}, \ldots\ldots \ldots DS_{n-1,n,t_{vt,st}}$$

// The set of x locations with the minimum difference from location i is then chosen as TRSS_cand (a_i,x).

Step 13: Evaluate spatial-temporal relations (STR)

$$STR\_(a_i, x) = SRSS\_cand(a_i, x) \cup TRSS\_cand(a_i, x)$$

Step 14: Each location is normalized as

$$El_s = \frac{ele - ele_{st}}{ele_{st}}$$

Step 15: Transferred to equivalent altitude is expressed as

$$ele_{rel} = \frac{1}{e^{El_s}}$$

Step 16: Calculate spatial-temporal predictor (STP)

$$(STR\_(a_i, x)), t_{t_l,tq} - = F(a_i, fs_{target}, t_{vtF,stF})$$

Step 17: ASE receive data on air quality and meteorological conditions, whereas CNN receives data on terrain

Step 18: Concatenate each other and the parameters are transferred on to the next layer

Step 19: Perform a feedforward pass using FFNN

Step 20: Computing the activations for layers L_2, L_3, and so on up to the output layer L_n

Step 21: End the process

5.4 Experimental Results

The Proposed Algorithm T-SBU-LSTM is compared with existing methods Transferred Bidirectional and Long Short-Term Memory (Tl-BiLSTM) and Denoising Autoencoder method (DAEDN) for various performance metrics.

5.4.1 Accuracy

The accuracy of the proposed T-SBU-LSTM method is compared to the existing TL-BiLSTM and DAEDN techniques across different iterations. The experimental results are presented in Figure 5.4.

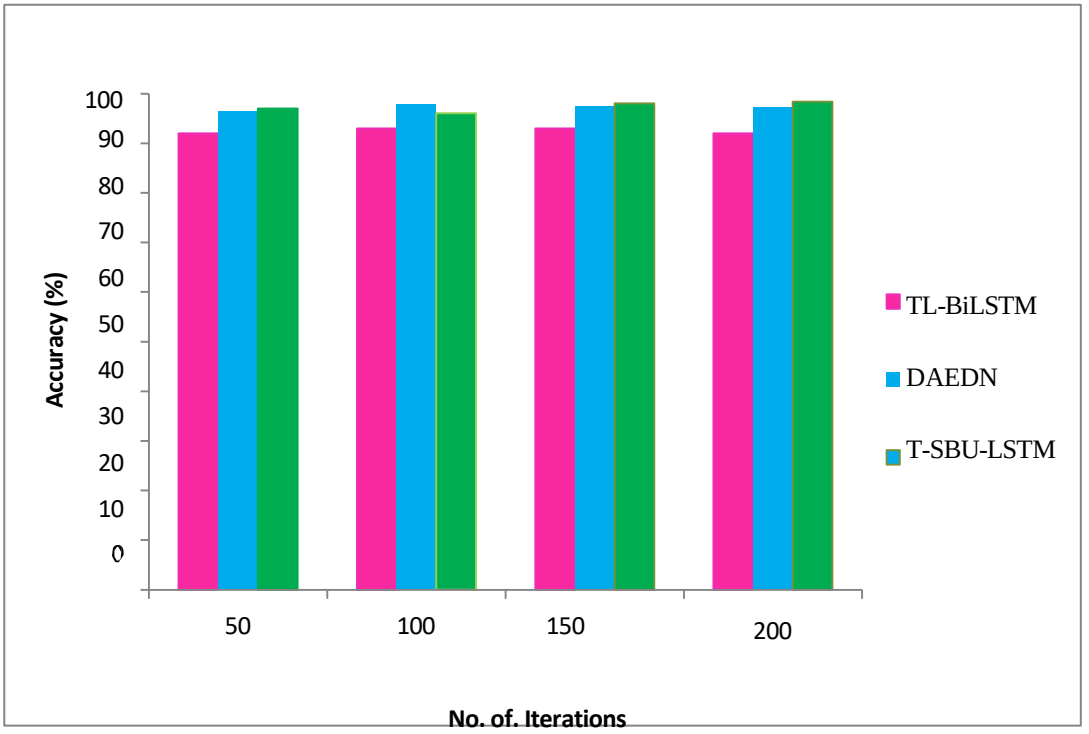

Figure 5.4 Comparison of accuracy

The results in Figure 5.4 indicate that the accuracy values of TL-BiLSTM is 92%, DAEDN is 98% and T-SBU-LSTM is 98.4%. The accuracy of T-SBU-LSTM improved by 0.01% and 0.3% compared to TL-BiLSTM and DAEDN, respectively. TL-BiLSTM and DAEDN had lower accuracy values, while T-SBU-LSTM demonstrated higher accuracy values. The Transferred T-SBU-LSTM provided the highest accuracy due to the combination of transfer learning, where the model benefits from pretrained knowledge on similar tasks, and the

sparse bidirectional LSTM structure, which helped in effectively capturing both past and future temporal dependencies. The bidirectional feature enables the model to better handle both short-term and long-term trends, increasing forecast accuracy because air quality data frequently displays complicated temporal patterns.

5.4.2 Precision

The precision of the proposed T-SBU-LSTM method is compared to the existing TL-BiLSTM and DAEDN techniques across different iterations. The experimental results are presented in Figure 5.5.

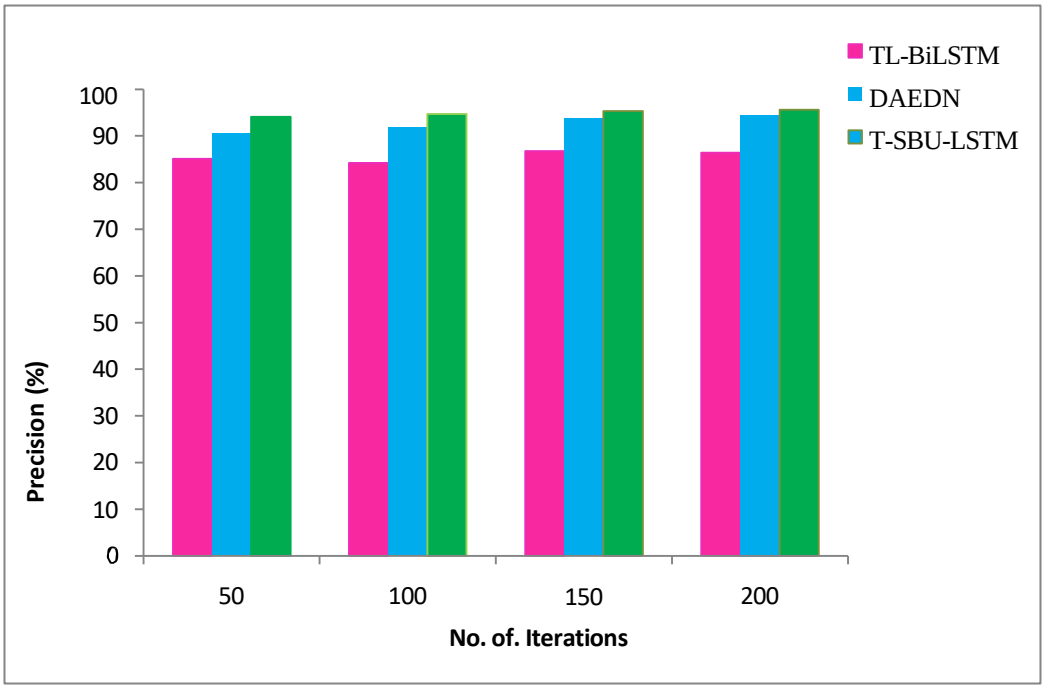

Figure 5.5 Comparison of Precision

The experimental results reveal that the precision of TL-BiLSTM is 86%, DAEDN is 94.5%, and the proposed T-SBU-LSTM method is 95.6%. In the air quality prediction system, T-SBU-LSTM improved precision by 1.1% compared to DAEDN and by 8.1% compared to TL-BiLSTM. TL-BiLSTM and DAEDN had lower precision values, whereas Transferred T-SBU-LSTM outperforms both TL-BiLSTM and DAEDN .The sparsity in T-SBU-LSTM helped to reduce false positives, making it highly precise in predicting pollution levels.

5.4.3 Specificity

The Specificity of the proposed T-SBU-LSTM method is compared to the existing TL-BiLSTM and DAEDN techniques across different iterations. The experimental results are presented in Figure 5.6.

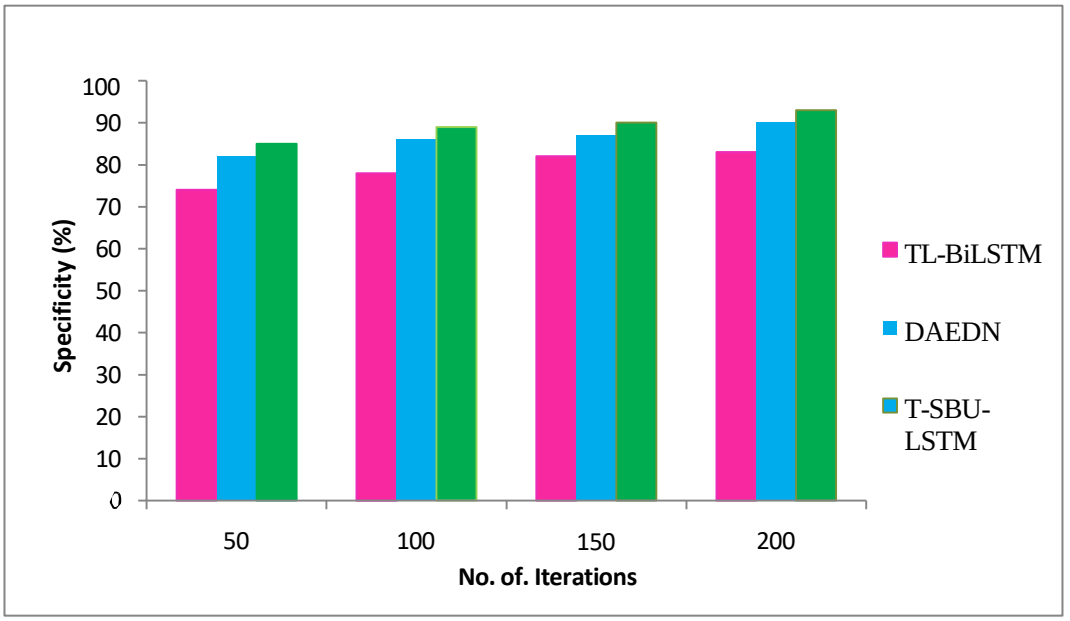

Figure 5.6 Comparison of Specificity

Figure 5.6 illustrates the specificity of T-SBU-LSTM, TL-BiLSTM, and DAEDN across various methods and different iterations. It is noted that the specificity of TL-BiLSTM is 83%, while DAEDN achieves 90%, whereas the proposed method T-SBU-LSTM attains 93%. The specificity of T-SBU-LSTM has improved by 0.3% compared to TL-BiLSTM and by 0.7% compared to DAEDN. T-SBU-LSTM exhibits higher specificity values, whereas TL-BiLSTM and DAEDN demonstrate lower specificity values, Transferred T-SBU-LSTM is the most effective, due to the sparse representation, which helps it focus on key features of the data, reducing false positives

5.4.4 Sensitivity

In Figure 5.7, the sensitivity of the proposed method T-SBU-LSTM is compared with TL-BiLSTM and DAEDN.

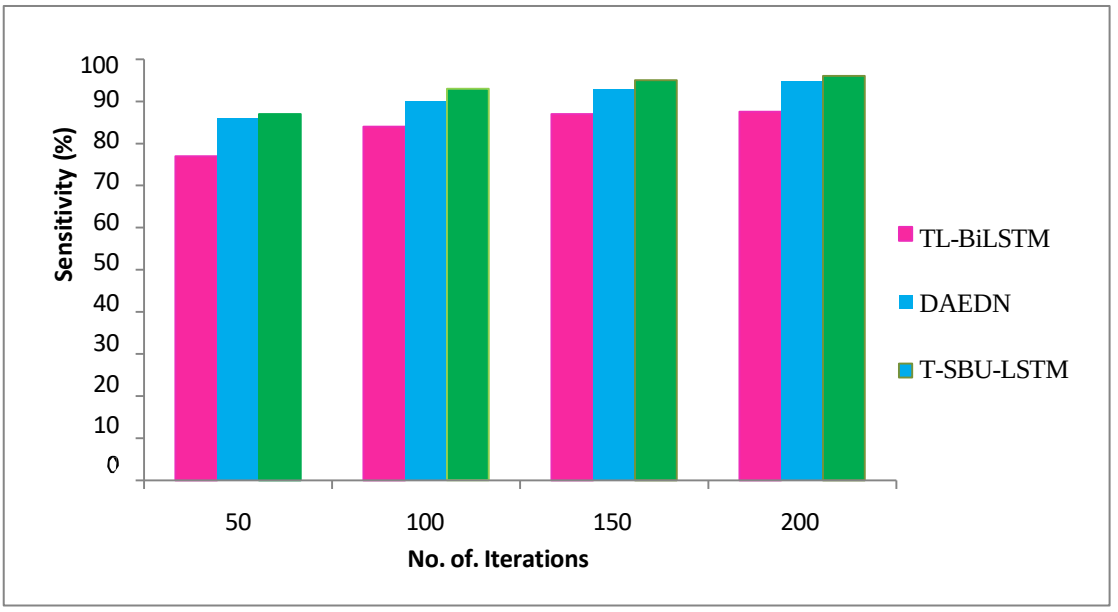

Figure 5.7 Comparison of Sensitivity

From the Figure 5.7, it is evident that the sensitivity of TL-BiLSTM is 87.5 % and DAEDN is 94 % and T-SBU-LSTM is 96%. The sensitivity of T-SBU-LSTM has improved by 0.2 % and 7.3% respectively. T-SBU-LSTM exhibits higher sensitivity values, whereas TL-BiLSTM and DAEDN demonstrate lower sensitivity values SBU-LSTM excels in its performance due to its bidirectional architecture, which helps capture both historical and future contextual information. This is crucial in air quality prediction, where recognizing spikes in pollution (such as during specific times of the day or in response to weather changes) is important.

5.4.5 Area under Curve (AUC)

Figure 5.8 compares the Area under Curve (AUC) of existing techniques TL-BiLSTM and DAEDN with the proposed method T-SBU-LSTM**.**

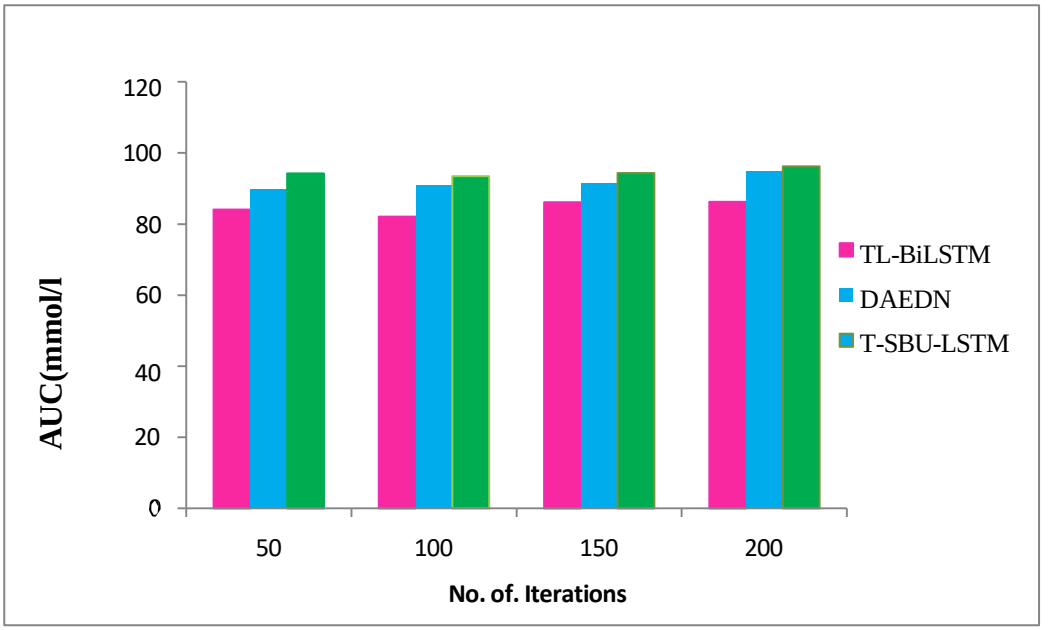

Figure 5.8 Comparison of AUC

Figure 5.8 shows that comparatively, The T-SBU-LSTM achieves 94.8% and it outperforms other algorithms. The AUC of T-SBU-LSTM has improved by 0.8 % and 8.6% respectively. Compared to TL-BiLSTM and DAEDN suggests low performance. Because the model struggles to differentiate between the positive and negative classes. T-SBU-LSTM indicates perfect discrimination, meaning the model perfectly distinguishes between positive and negative classes. T-SBU-LSTM shows the best performance as the combination of bidirectional LSTM and sparse representation helps it more effectively discriminate between varying levels of pollution, leading to a higher AUC.

5.4.6 Matthew's Correlation Coefficient (MCC)

Figure 5.9 presents a comparison of the Matthew's Correlation Coefficient (MCC) among existing techniques TL-BiLSTM and DAEDN, and the proposed method T-SBU-LSTM.

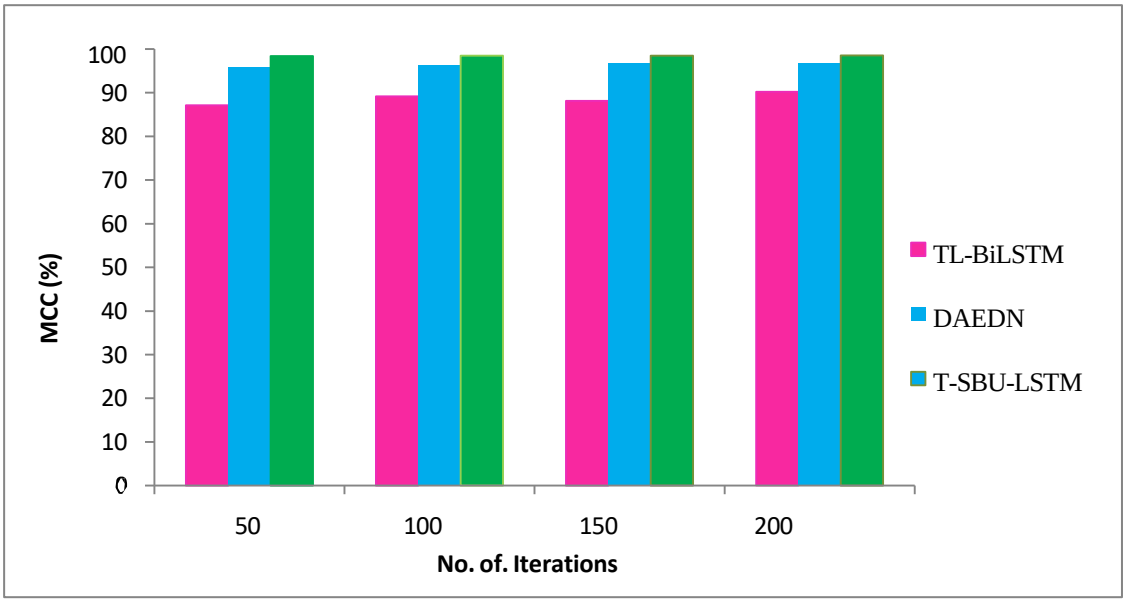

Figure 5.9 Comparison of MCC

The analysis reveals that the MCC of TL-BiLSTM and DAEDN is 90% and 96%, respectively, while T-SBU-LSTM achieves 98.5%. In the air quality prediction system, the MCC of T-SBU-LSTM has improved by 8.5% compared to TL- BiLSTM and 2.5% compared to DAEDN demonstrates higher MCC values, whereas and DAEDN exhibit lower MCC values. The ability to leverage transfer learning and bidirectional LSTM architecture ensures that the model not only correctly classifies air quality data but also performs well in distinguishing between different pollution levels, thus improving the MCC.

5.4.7 Mean Absolute Error Rate (MAER)

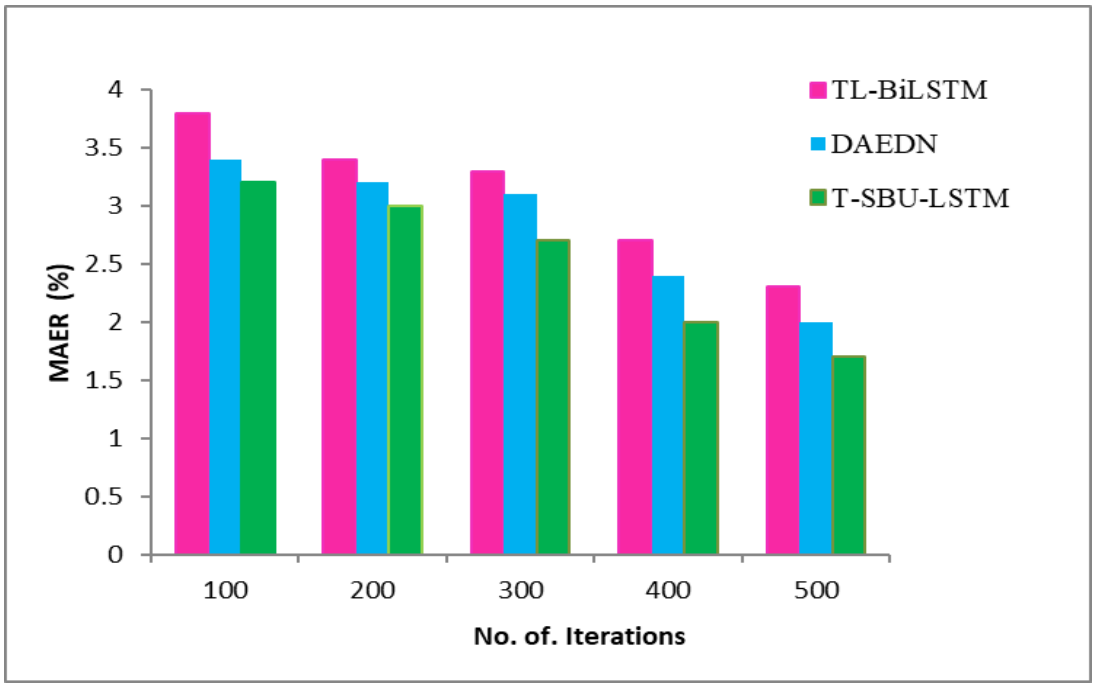

Figure 5.10 Comparison of MAER

In Figure 5.10, the Mean Absolute Error Rate (MAER) of the proposed method T-SBU- LSTM is compared with the existing techniques TL-BiLSTM and DAEDN. The results indicate that T-SBU-LSTM exhibits a lower error rate than TL-BiLSTM by 26.08% and DAEDN by 15%. In TL-BiLSTM and DAEDN demonstrate higher error rate values compared to T-SBU-LSTM.

5.4.8 F -Measure

Figure 5.11 presents a comparison of F-measure among existing techniques TL-BiLSTM and DAEDN, and the proposed method T-SBU- LSTM.

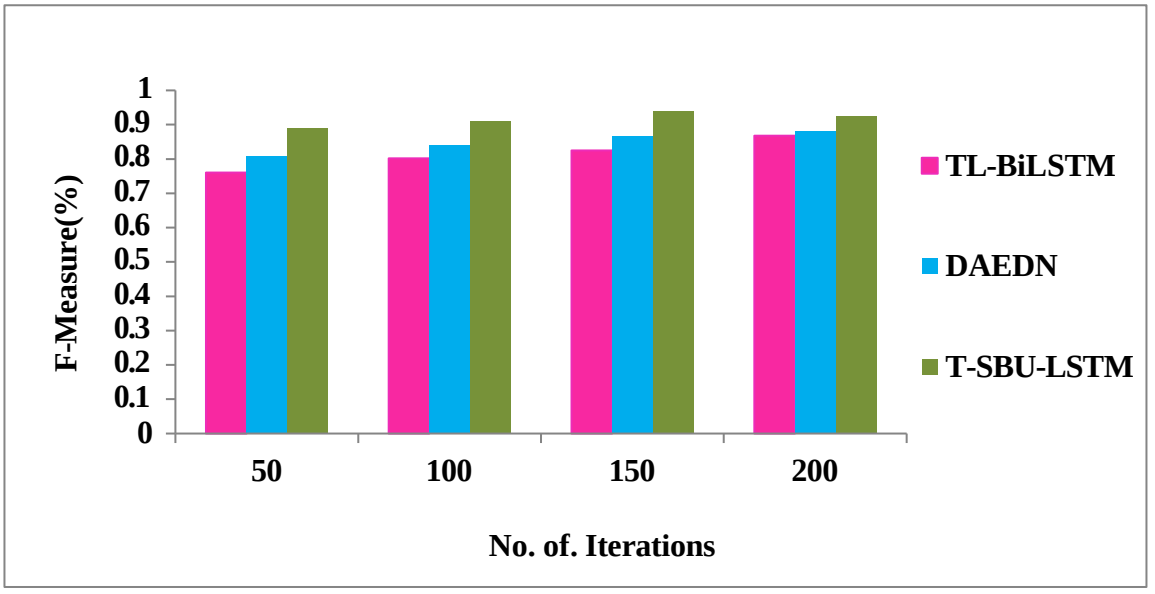

Figure 5.11 Comparison of F-Measure

The analysis reveals that the F-measure of TL-BiLSTM and DAEDN is 90% and 96%, respectively, while T-SBU-LSTM achieves 98.5%. In the air quality prediction system, the F-measure of T-SBU-LSTM has improved by 6.79% compared to TL- BiLSTM and 4.98% compared to DAEDN. Transferred T-SBU-LSTM achieved highest F-measure compared to TL-BILSTM and DAEDN in air quality prediction due to its superior handling of both temporal dependencies and sparsity.