Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 1, February 2025, pp. 625~637
ISSN: 2302-9285, DOI: 10.11591/eei.v14i1.7968

Leveraging pretrained transformers for enhanced air quality index prediction model

Santhana Lakshmi Velusamy, Vijaya Madhaya Shanmugam
Department of Computer Science, PSGR Krishnammal College for Women, Coimbatore, India

Article Info

Article history:

Received Nov 30, 2023
Revised Sep 13, 2024
Accepted Sep 28, 2024

Keywords:

Air quality index
Machine learning
Prediction
Pre-trained model
Temporal fusion transformer
Transfer learning

ABSTRACT

Air pollution mitigation is essential to ensure sustainable development, as it directly affects climate change, economic productivity, and social well-being. Despite the availability of numerous prediction techniques, machine learning (ML) remains the optimal solution for forecasting air pollution. Constructing a prediction model for a region with limited data poses a challenge. This study presents a novel technique that combines temporal fusion transformer (TFT) with transfer learning to create an inventive air quality index (AQI) prediction model, effectively utilizing temporal insights and prior knowledge. The TFT is an advanced deep neural architecture engineered to enhance time series forecasting through the fusion of sequence modelling and global temporal patterns. By fusing TFT with transfer learning, the research pioneers a fresh approach to AQI prediction for region with data scarcity issue, capitalizing on cross-domain knowledge transfer. Utilizing meteorological and pollutant data from the Cochin region, a hybrid AQI prediction model is constructed through TFT and transfer learning. Employing a preexisting TFT model trained on Trivandrum data, transfer learning technique is utilized to adapt the model for predicting AQI in the Cochin region. The study demonstrates that integrating TFT with transfer learning yields superior accuracy compared to an exclusive TFT-based approach.

Corresponding Author:

Santhana Lakshmi Velusamy
Department of Computer Science, PSGR Krishnammal College for Women
Peelamedu, Coimbatore, India
Email: sanlakmphil@gmail.com

1. INTRODUCTION

Air pollution refers to the presence of harmful substances in the atmosphere, posing risks to human health and the environment. Originating from both human activities and natural sources, air pollutants encompass diverse particles and gases, including particulate matter, nitrogen dioxide, sulfur dioxide, volatile organic compounds, and ozone. Their cumulative effect contributes to air quality deterioration and consequent health and environmental consequences [1]. Accurate prediction of pollution levels is essential for proactively implementing measures to protect public health and maintain ecological equilibrium. This paper aims to develop an air quality index (AQI) prediction model for the Cochin region, which faces challenges due to limited data availability. Several methods are available for predicting AQI levels. Statistical models involve analyzing historical air quality data alongside meteorological variables to establish regression-based predictions that account for correlations between different parameters and AQI outcomes [2]. Chemical transport models simulate the dispersion and chemical interactions of pollutants in the atmosphere, considering emission sources, atmospheric conditions, and transport mechanisms to predict AQI levels [3]. Satellite observations provide valuable insights by monitoring pollutant concentrations and atmospheric conditions on a broader geographical scale, facilitating the estimation of AQI over larger areas. Additionally, sensor networks comprising distributed air quality sensors offer real-time data collection for immediate AQI calculations and localized pollution monitoring, enhancing our ability to predict and respond to air quality fluctuations effectively [4].

Machine learning (ML) techniques, such as neural networks (NN), random forests (RF), and support vector machines (SVM), offer the capability to discern intricate relationships within vast datasets, enabling accurate AQI forecasting. These methods can leverage historical air quality data and other relevant features to build robust predictive models [5]. Deep learning (DL) algorithms are widely utilized for AQI prediction. DL excels in capturing complex patterns and nonlinear relationships within air quality data, enabling more accurate predictions and adaptability to changing environmental conditions. To address this, a review of research articles utilizing ML and DL algorithms was conducted and summarized [6]. Das *et al.* [7] have used ML and DL to build an accurately predicting AQI prediction models. They compared the performance of ML and DL algorithms. The study utilized a comprehensive dataset comprising historical air quality measurements and meteorological parameters from multiple monitoring stations in a metropolitan area over several years. Employing a comparative approach, the authors implemented various algorithms including RF, SVM, recurrent neural network (RNN), and long short-term memory (LSTM) networks. Results revealed that the LSTM network, a DL model, outperformed traditional ML algorithms, particularly in capturing temporal dependencies within the data.

Gupta *et al.* [8] aimed to establish a robust predictive model for AQI through the utilization of ML methods. The study focused on assessing the effectiveness of various algorithms in accurately forecasting AQI levels by leveraging historical air quality and meteorological data. The dataset employed encompassed air quality measurements including PM2.5, PM10, NO_2, SO_2, CO, and O_3 concentrations, alongside meteorological variables such as temperature, humidity, and wind speed. By applying ML techniques like decision trees (DT), RF, and gradient boosting (GB), the authors constructed predictive models. The results indicated that the GB ensemble technique outperformed other algorithms, effectively capturing non-linear interactions within the data and yielding heightened predictive accuracy. This study contributes to the existing body of knowledge by spotlighting the potency of ensemble ML approaches, particularly GB, in refining AQI predictions.

Shankar and Arasu [9] aimed to explore the utilization of DL techniques for accurately predicting AQI levels. The research focused on assessing the potential of DL models in capturing intricate patterns within air quality and meteorological data. The dataset employed encompassed historical air quality measurements, including concentrations of PM2.5, PM10, CO, NO_2, SO_2, and O_3, as well as meteorological features like temperature, humidity, and wind speed, collected from urban monitoring stations over an extensive time frame. The study utilized various DL models, including convolutional neural networks (CNN) and LSTM networks, to predict AQI levels. Notably, the LSTM network emerged as a standout performer, effectively capturing temporal dependencies and correlations within the data. The LSTM model's ability to retain information over time contributed to its success in accurately predicting AQI levels.

Sarkar *et al.* [10] aimed to develop and assess hybrid ML models for predicting AQI levels. The research focused on leveraging the strengths of different algorithms to enhance the accuracy of AQI forecasts. The dataset employed encompassed historical air quality measurements including concentrations of PM2.5, PM10, SO_2, NO_2, CO, and O_3, alongside pertinent meteorological parameters such as temperature, humidity, and wind speed, gathered from monitoring stations spanning an urban area. Hybrid ML models were constructed by integrating multiple algorithms, including SVM, RF, and artificial neural networks (ANN). The results revealed that the hybrid models outperformed the performance of individual algorithms in predicting AQI levels. By harnessing the collective predictive power of different techniques, the hybrid models demonstrated improved accuracy. Halsana [11] used data from various web sources like central pollution control board (CPCB) and University of California, Irvine (UCI) repository to construct an AQI prediction model. The model was built employing supervised learning algorithms such as multiple linear regression, RF regression, DT regression, and SV regression. Improved accuracy was achieved through the utilization of RF regression. The performance of the algorithms was assessed using mean squared error (MSE) and mean absolute error (MAE).

The conducted literature review has provided valuable insights into the contemporary methods and technologies employed in constructing AQI models. The authors of various studies have utilized both ML and DL techniques to develop models for predicting AQI values. Additionally, it is evident that all of them have used ample amounts of data for building the predictive model, highlighting the significance of robust datasets in achieving accurate predictions. A notable gap exists in addressing the challenge of developing an AQI prediction model for a region with limited dataset size. Hence, in this paper, a transfer learning approach is proposed, aiming to leverage knowledge gained from models trained on datasets from regions with ample data to improve AQI prediction accuracy in regions with limited dataset sizes. This paper also presents an innovative methodology that combines temporal fusion transformer (TFT) with transfer learning to develop an enhanced and accurate model for predicting AQI. Given the limited data instances for the Cochin region, a strategic approach is adopted. The AQI prediction model built already using Trivandrum data with TFT is employed as a pretrained model. This pretrained knowledge is then utilized to construct a specialized TFT-based AQI model tailored to the Cochin area. The results of transfer learning-based models are compared with AQI prediction models created without transfer learning. The architecture of TFT, method of implementing transfer learning and experimental results are explained in detail.

2. BASE AND PRETRAINED AQI PREDICTION MODEL WITH TEMPORAL FUSION TRANSFORMER AND TRANSFER LEARNING

Constructing an accurate AQI prediction model for the regions with limited data availability is a significant challenge. However, leveraging transfer learning offers a promising solution to this dilemma. By harnessing the knowledge acquired from a region with abundant data, transfer learning enables the adaptation of pre-existing models to suit the characteristics of a data-scarce region. In this context, a specialized transformer-based architecture known as TFT emerges as a crucial component. TFT, with its ability to capture complex temporal dependencies, serves as the foundation for the AQI prediction model. Central to the implementation of transfer learning is the utilization of pretrained models trained on data-rich regions. These pretrained models encapsulate valuable insights that can be transferred to enhance the performance of our AQI prediction model in regions with limited data. In this section, an exploration into the architecture of TFT, the concept of transfer learning, and the pivotal role of pretrained models is undertaken. The focus primarily lies on presenting the foundational framework of the base model, serving as the cornerstone for subsequent comparative analyses against transfer learning-enhanced models, as detailed in the experimental results section.

2.1. Temporal fusion transformer

TFT is an advanced sequence-to-sequence forecasting model that effectively captures temporal dependencies and interactions within time-series data, enabling accurate predictions by incorporating both historical and contextual information [12]. TFT utilizes autoregressive components alongside the attention mechanisms, which facilitates the incorporation of different contextual information from variable time windows. This adaptability is particularly useful for time series data with varying context lengths, making TFT well-suited for capturing both short-term and long-term dependencies, enhancing its predictive capabilities. The architecture of TFT is given in Figure 1.

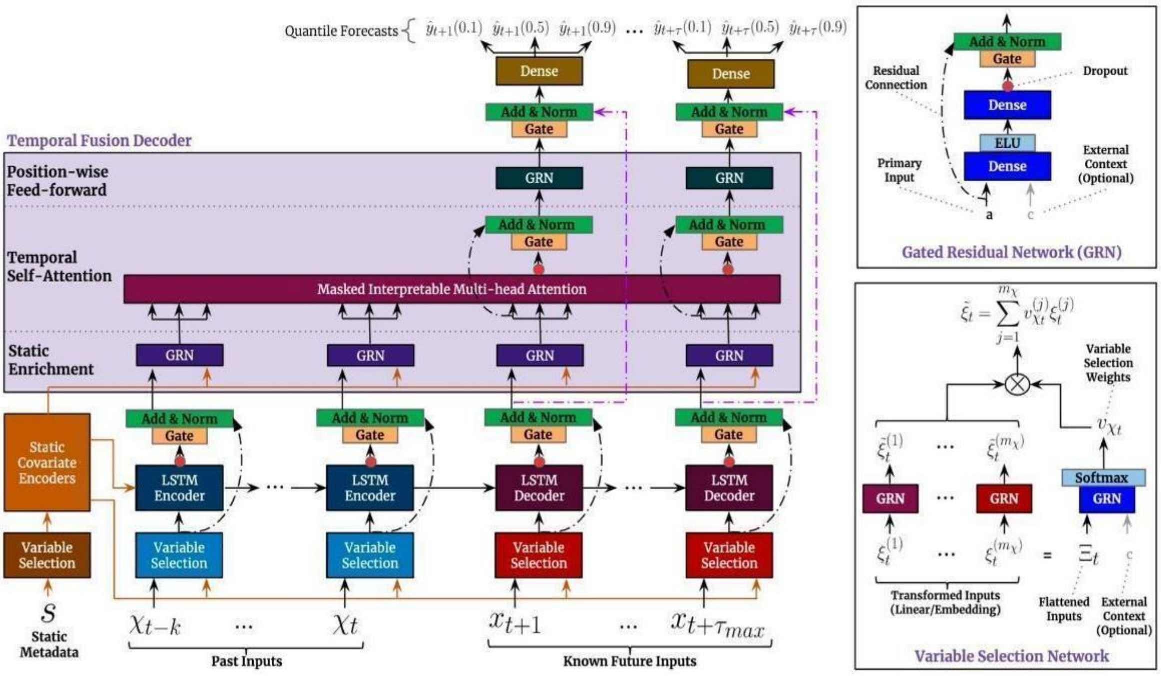

Figure 1. Architecture of temporal fusion transformer

TFT's architecture is composed of encoder and decoder components. The encoder takes in the historical observations, embedding them into a higher-dimensional space where self-attention mechanisms operate. This enables the model to understand the relationships between different time steps and extract meaningful features from the input sequence. The contextual embeddings generated by the encoder are then fed into the decoder, which employs autoregressive components to predict future values. TFT's architecture is designed to automatically learn relevant features from the raw time series data that reduces the dependency on extensive manual feature engineering [13]. TFT's ability to capture complex patterns and dependencies within the data means that it can effectively handle various data formats, making it a versatile solution that requires less preprocessing effort while still delivering accurate predictions.

2.2. Base air quality index prediction model with temporal fusion transformer

In our previous work, a predictive model for air quality in the Cochin region was developed by integrating meteorological and pollutant features. Meteorological data obtained through the pollution control board's portal and pollutant data sourced from the visual crossing website were combined based on the date. The comprehensive dataset spanned a period of 7 months, starting from July 2020 and ending in January 2021, and was used in the study. Initial exploration of the data involved conducting exploratory data analysis (EDA), which facilitated a deeper understanding of the data's inherent patterns and characteristics. The insights gleaned from the EDA phase guided the determination of necessary preprocessing steps. Finally, after implementing the necessary preprocessing steps, the dataset AQI-PMF-Cochin is formed with 5161 instances and 17 attributes. The dataset was then input into the TFT model, leveraging its advanced architecture to enhance air quality predictions in the Cochin region.

As data enters the architecture, the encoder layer captures temporal relationships using self-attention mechanisms, understanding interdependencies among time steps. Contextual embeddings, forged in this phase, encapsulate intricate patterns. These embeddings transition to the decoder layer, where autoregressive elements exploit historical insights for forecasting AQI values. This step ensures TFT comprehends time series data's sequential nature, improving prediction accuracy. Passing through TFT's layers, blending self-attention and autoregressive mechanisms, yields predictions capturing air quality conditions. TFT effectively transforms time series air quality data, bridging history and future for enhanced AQI forecasting.

The TFT architecture's performance was enhanced by tuning hyperparameters [14]. The number of hidden units influences the model's complexity, while dropout prevents overfitting learning rate, window size, and sequence length affect convergence and temporal context. Activation functions like ReLU, sigmoid, and Tanh introduce non-linearity. Optimizing these parameters was crucial for accurate AQI predictions with TFT, involving iterative experimentation and validation.

A side from commonly tuned hyperparameters like learning rate, dropouts, and activation functions, several key parameters significantly influenced the performance tuning of the TFT-based AQI prediction model. The prediction time step dictates the model's forecast duration, while the deep neural networks (DDN) encoding layer controls the depth of temporal encoding. The "state size" and "dropout rate" in TFT impacted the model's representational capacity and regularization, respectively, crucial for capturing complex patterns and preventing overfitting.

Various NN including LSTM, bidirectional long short-term memory (BiLSTM), and gated recurrent unit (GRU) were simultaneously implemented with the same dataset and AQI models were developed. The dataset was split, utilizing 80% of records for training and the remaining 20% for testing purposes. To assess model efficacy, performance evaluation metrics such as MAE, MSE, root mean squared error (RMSE), and R2 were employed. The developed models were named as LSTM-AQI-PMF, BILSTM-AQI-PMF, GRU-AQI-PMF, and TFT-AQI-PMF. The outcomes of these assessments are presented in Table 1 and illustrated in Figure 2. The x axis represents different models used for prediction. The y axis represents scaled performance metrics ranging from 0 to 0.8.

Table 1. Performance of DL models for Cochin data

Models	MAE	MSE	RMSE	R2 score
LSTM-AQI-PMF-Cochin	0.5013	0.4270	0.6534	0.6543
BILSTM-AQI-PMF-Cochin	0.5123	0.4321	0.6723	0.6471
GRU-AQI-PMF-Cochin	0.5462	0.4815	0.6939	0.6214
TFT-AQI-PMF-Cochin	0.4285	0.3314	0.5756	0.7332

2.3. Pre-trained air quality index prediction model

In the previous work, an AQI prediction model was developed using Trivandrum data. This data encompassed meteorological data collected from the central pollution control board portal and pollutant data sourced from the visual crossing website spanning a timeframe of 3 years from 2017 to 2020. The data used includes 26,305 instances with 24 attributes.

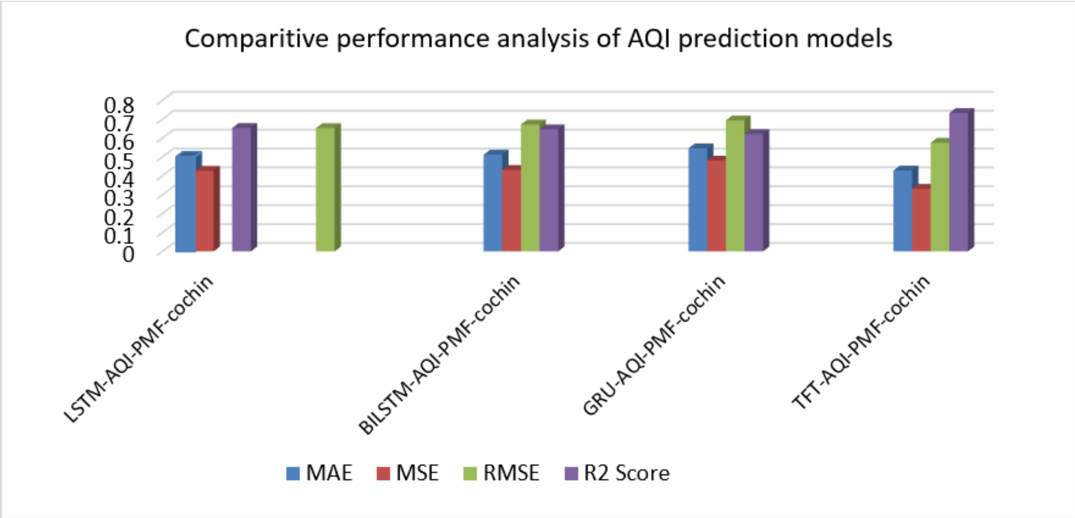

Figure 2. Comparative performance analysis of AQI prediction models

The research commenced with an EDA phase, employing tools like histograms, heatmaps, boxplots, bar charts, and pair plots to identify the insights and trends within the dataset [15]. Outliers were identified and preprocessing measures were executed to rectify this issue, ensuring data integrity [16]. To enhance predictive capabilities, feature engineering was conducted to identify the attributes crucial for AQI prediction. The target attribute, AQI, was calculated for each data point and appended as the final attribute. Finally, the dataset was formed with 26,305 instances and 17 attributes. The dataset was named as AQI-PMF. The attributes used were timestamp, temperature, relative humidity, dew, sea level pressure, cloud cover, atmospheric temperature, rainfall, wind speed, PM2.5, PM10, CO, SO_2, ozone, nox and NH_3 and AQI.

For model development, 80% of the records were allocated for training, while the remaining 20% were allocated for testing. Employing DL architectures, namely LSTM, BILSTM, GRU, and TFT, the prediction models were constructed and named it as LSTM_AQI_PMF, BILSTM_AQI_PMF, GRU_AQI_PMF and TFT_AQI_PMF. Performance of the models were evaluated based on MAE, MSE, RMSE, and R-squared value. The predictive results obtained are provided in Table 2 and illustrated in Figure 3. The developed models were named as LSTM-AQI-PMF, BILSTM-AQI-PMF, GRU-AQI-PMF and TFT-AQI-PMF. The outcomes of these assessments are presented in Table 1 and illustrated in Figure 2. The x axis represents different models used for prediction. The y axis represents scaled performance metrics ranging from 0 to 1.

Table 2. Performance of pretrained models

	LSTM_AQI_PMF	BILSTM_AQI_PMF	GRU_AQI_PMF	TFT_AQI_PMF
MAE	0.31	0.51	0.37	0.20
RMSE	0.42	0.60	0.49	0.26
MSE	0.17	0.36	0.24	0.09
R2	0.85	0.70	0.82	0.92

Out of the four models used to predict air quality, the TFT_AQI_PMF model seems to be the best. TFT_AQI_PMF has the lowest MAE of 0.20, RMSE value of 0.26, and MSE of 0.09, indicating that it has the smallest error on average. TFT_AQI_PMF has the highest R2 of 0.92, suggesting that it captures the variance in the data most effectively and provides a better fit to the actual values. When comparing the results of all the models, the model TFT-AQI-PMF shows superior performance in predicting the AQI value. When comparing the results of the AQI models built for Trivandrum region given in Table 2 with the performance of the model's LSTM-AQI-Cochin, BILSTM-AQI-Cochin, GRU-AQI-Cochin and TFT-AQI-Cochin built for Cochin given in Table 1, in all the aspects such as MSE, MAE, RMSE and R2 the models built for Trivandrum region shows superior performance. The two key features that contributed to the efficiency of the models are the ample availability of data instances used during the model's construction and the implementation of TFT for constructing the model.

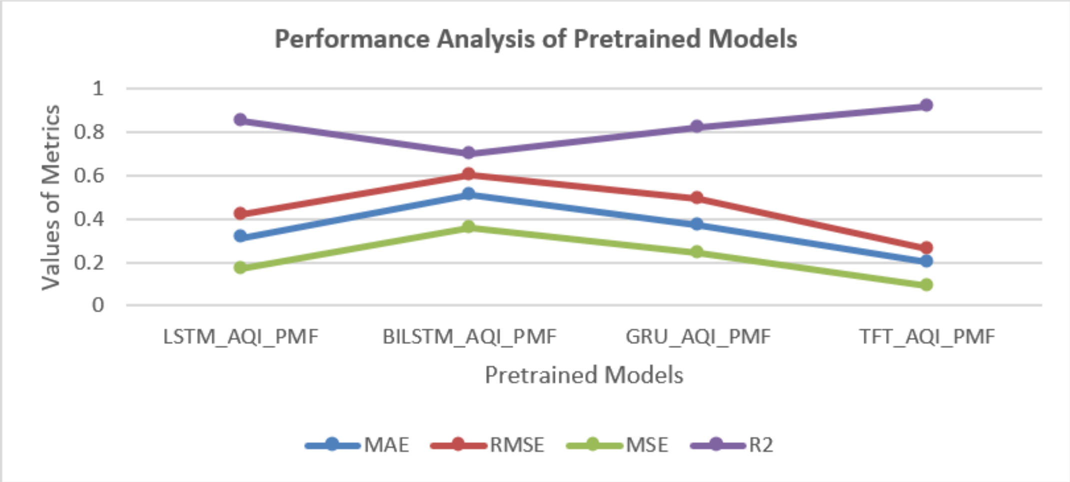

Figure 3. Performance analysis of pretrained models

Recognizing the success achieved with the AQI models created for Trivandrum, there is potential for further improvements in the efficiency of an AQI model designed for the Cochin region by leveraging the knowledge acquired from the Trivandrum dataset AQI-PMF. Therefore, the model TFT-AQI-PMF is utilized as the pretrained model for developing the AQI model specifically tailored to the Cochin region. This transfer learning approach ensures that, the valuable insights and patterns learned from Trivandrum's air quality data can be effectively applied to enhance the performance of the Cochin AQI model.

2.4. Transfer learning

Transfer learning is a potent ML approach that capitalizes on the insights gained from solving one task to enhance learning and performance on a different but related task [17]. This process involves training a model on a source task, often utilizing a vast and intricate dataset. The model learns to extract meaningful features and recognize complex patterns within the data during this pre-training phase. Subsequently, these learned features are extracted from specific layers of the model. They serve as a representation of the data's intrinsic characteristics, encapsulating valuable information. With the extracted features as a foundation, a new model is constructed for the target task. While the architecture might resemble the original pretrained model, the parameters are fine-tuned using a smaller dataset closely linked to the target task. This fine-tuning process allows the model to adapt its features to the specific nuances of the new task, resulting in enhanced performance.

Transfer learning is classified into two main types homogeneous transfer learning and heterogeneous transfer learning. Homogeneous transfer learning involves transferring knowledge between similar domains or tasks, while heterogeneous transfer learning deals with transferring knowledge between dissimilar domains or tasks [18]. Transfer learning encompasses two distinct approaches: the developer approach and the pretrained model approach. The pretrained model approach capitalizes on models previously trained on extensive datasets. These pretrained models encapsulate learned features and patterns [19]. By fine-tuning them with the specific dataset, they become adaptable to the new task. This strategy optimizes time and computational resources, as it leverages existing knowledge effectively. On the other hand, the developer approach involves crafting a custom model architecture tailored to the specific task [20]. This method demands domain expertise and manual feature engineering to extract pertinent insights from the limited dataset. In this work, homogeneous transfer learning is implemented. As the pretrained model is not available for AQI prediction, developer approach is followed. A model was built for Trivandrum in the previous work was used as a pretrained model. An inherent advantage of transfer learning lies in its ability to expedite training. By leveraging knowledge from a source task, models can avoid starting from scratch and benefit from pre-trained features, leading to quicker convergence during fine-tuning [21]. This acceleration is particularly valuable when faced with limited data or complex architectures, ultimately facilitating rapid model deployment and iterative development cycles.

2.4.1. Transfer learning mechanics

Transfer learning represents a DL technique enabling the utilization of a pre-trained model as a foundational starting point for a novel task. The process involves creating a pretrained model and subsequently fine-tuning it on task-specific data. An essential step in this process is freezing the final layer of the pretrained model when transferring knowledge. During the initial training phase on a source task, the model is exposed to a large-scale dataset, learning to extract valuable features and patterns. Once this training is complete, the model's layers become feature extractors, capturing higher-level representations that are potentially relevant to various tasks. When the focus shifts to the target task, the pretrained model is modified to align with the task's requirements. A new architecture like pretrained model is created and the weights are transferred from the pretrained model to new model [22]. The new model is then fine-tuned using the task-specific data. Keeping the weights obtained from the pretrained model as foundation, new weights are calculated. This makes the model retains its ability to recognize more universal features, and accommodate the nuances of the target task. This approach strikes a balance between leveraging existing knowledge and adapting to the specific context of the target task, making it especially effective when there's a scarcity of data for the target domain. The working of transfer learning is depicted in Figure 4.

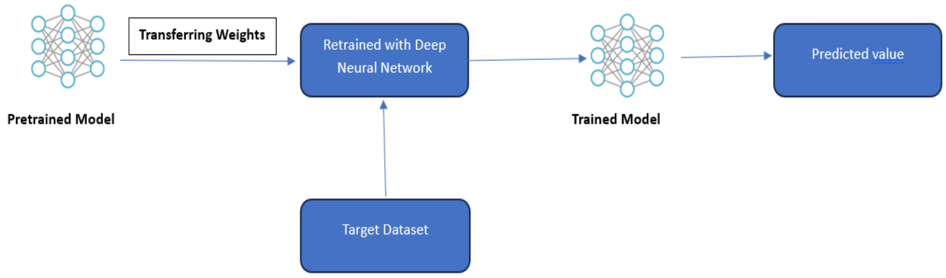

Figure 4. Working of transfer learning

3. PROPOSED AQI PREDICTION MODEL USING TFT AND HOMOGENEOUS TRANSFER LEARNING

This study introduces a novel approach for constructing an accurate AQI model by integrating transfer learning with the TFT framework. Addressing the challenge of limited prediction accuracy observed in AQI models designed for the Cochin region due to a scarcity of training samples, this research proposes the integration of transfer learning as a solution. The proposed architecture is given in Figure 5.

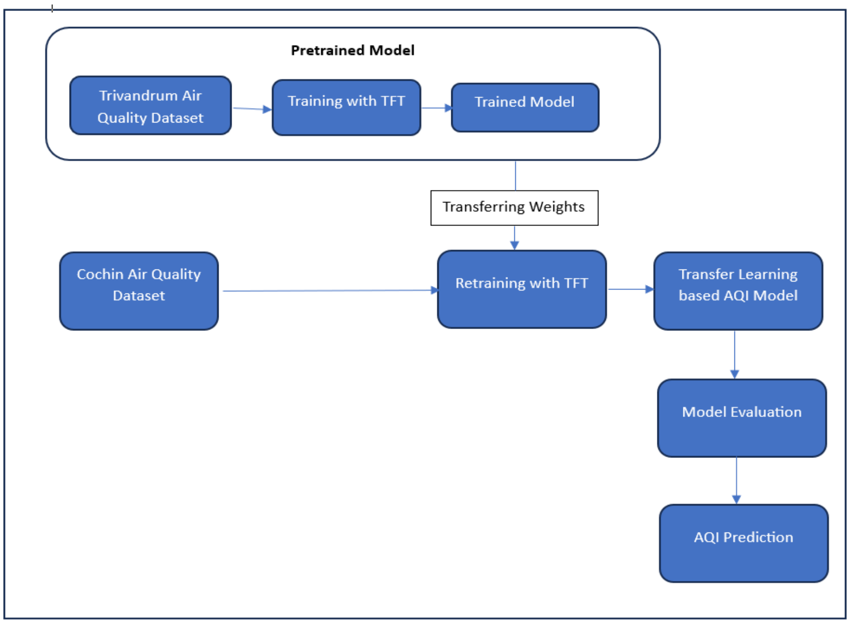

Figure 5. Proposed architecture of transfer learning based AQI prediction model

The method for constructing an AQI prediction model for the Cochin region entails several sequential steps: i) creation of a pretrained model, ii) transfer of knowledge, and iii) fine-tuning with hyperparameters.

- Creation of a pretrained model: a pretrained model for AQI prediction is currently unavailable. Therefore, the study relies on a model created in prior research as the pretrained AQI prediction model thus following developer approach. The pretrained model utilized in this work was developed using meteorological and pollutant features of Trivandrum region. Various architectures including LSTM, BILSTM, GRU, and TFT were employed to construct the models and are labelled as LSTM-AQI-PMF, BILSTM-AQI-PMF, GRU-AQI-PMF, and TFT-AQI-PMF. The detailed explanation about the dataset used, exploratory data analytics and preprocessing steps performed and the experimental results are explained in section as TFT consistently exhibits superior accuracy relative to other architectures, the TFT-AQI-PMF model was selected for use as the pretrained model.
- Transfer of knowledge: the pretrained TFT model, initially trained on the Trivandrum dataset, has been adept at capturing intricate temporal patterns and interrelationships present within the data. The features derived from the model's layers hold invaluable insights into the diverse influences of meteorological and pollutant attributes on AQI values. Serving as a foundational element for constructing the AQI model for Cochin data, the pretrained TFT model marks the outset of the process. Both the pretrained model and the Cochin dataset were trained with an identical set of feature attributes, thereby constituting a form of homogeneous transfer learning. Following this, a new TFT architecture resembling the structure of the pretrained model is constructed and initialized with the weights obtained from the pretrained model trained on the Trivandrum dataset. The newly created model is then trained with the Cochin dataset, leveraging the foundational weights from the pretrained model. This process facilitates the creation of new weights tailored specifically to the Cochin data. The knowledge acquired by the pretrained TFT model regarding general temporal patterns, correlations between attributes, and how they affect AQI is effectively transferred to the Cochin AQI model. However, as the model fine-tunes with Cochin data, it also learns to adapt to the specific intricacies of Cochin's air quality data.
- Finetuning with hyperparameters: the performance of the model is further enhanced by assigning the Hyperparameters such as number of hidden units, dropouts, learning rate, and activation functions. The special hyperparameters used to finetune the TFT model are DDN encoding layer, state size and loss functions. The number of hidden units in NN layers affects the model's capacity to capture intricate data patterns, while dropout regularization mitigates overfitting by randomly deactivating neurons. Setting an appropriate learning rate influences how quickly the model converges during training. The size and sequence length of the input window determine the temporal scope of information used for predictions, adapting to the data's time dependencies. Activation functions introduce non-linearity, impacting the model's capacity to grasp complex relationships within the data [23]. DDN encoding layer significantly influences the model's ability to represent AQI data. State size affects the model's capacity to capture sequential patterns. Finally, the choice of loss functions is instrumental in guiding the model towards minimizing prediction errors. Careful hyperparameter tuning, considering these factors, is essential to optimize the AQI prediction model's performance for a given dataset and use case [24]. The details about the values assigned for the hyperparameters are provided in Table 3.

Table 3. Special hyperparameters of TFT

No of time steps	No of DDN encoder layers	Number of batch sizes	State size	Learning rates	No of attention heads	Dropout rate	Loss function A	Loss function B	Loss function G
50	5	128	64	0.01	5	0.20, 0.30, 0.40	0.5	0.01	0.1

4. EXPERIMENTS AND RESULTS

To create an AQI prediction model using transfer learning, the process involves training the AQI-PMF-Cochin dataset while incorporating knowledge gained from the pretrained TFT-AQI-PMF model. By implementing the TFT architecture using python, a new AQI prediction model named TFT_AQI_PMF_TL is formulated. Eighty percentage of the records are used for training and twenty percentage of the records are used for testing. The performance of the models are evaluated using performance metrics MSE, MAE, RMSE, and R-squared value.

4.1. Results of air quality index prediction model with various activation functions

Activation functions introduce non-linearity, allowing NN to capture complex data patterns, mitigate gradient issues, and stabilize training. The choice of activation function creates an impact on the NN ability to model and approximate non-linear relationships in data. Using activation functions like ReLU, Leaky ReLU, and Tanh, experiments were conducted with epoch value as 50 to evaluate their impact on the performance of the model and the results are given in the Table 4. When comparing the model performance using activation functions ReLU, Leaky ReLU, and Tanh over 50 epochs, ReLU outperforms the others. It exhibits lower MAE, MSE, and RMSE, while Tanh results in higher errors. Additionally, ReLU yields a higher R2, indicating better fit. Thus, for further analysis, ReLU is recommended as the preferred activation function. The results are illustrated in Figure 6. The x axis represents different activation functions and y axis represents scaled performance metrics.

Table 4. Performance of the TFT_AQI_PMF_TL model with various activation functions

Performance metrics	Activation functions		
	TFT-ReLU	TFT-leaky ReLU	TFT-Tanh
MAE	0.5267	0.5693	0.5926
MSE	0.3637	0.4070	0.4603
RMSE	0.6031	0.6377	0.6783
R2	0.5443	0.4581	0.4192

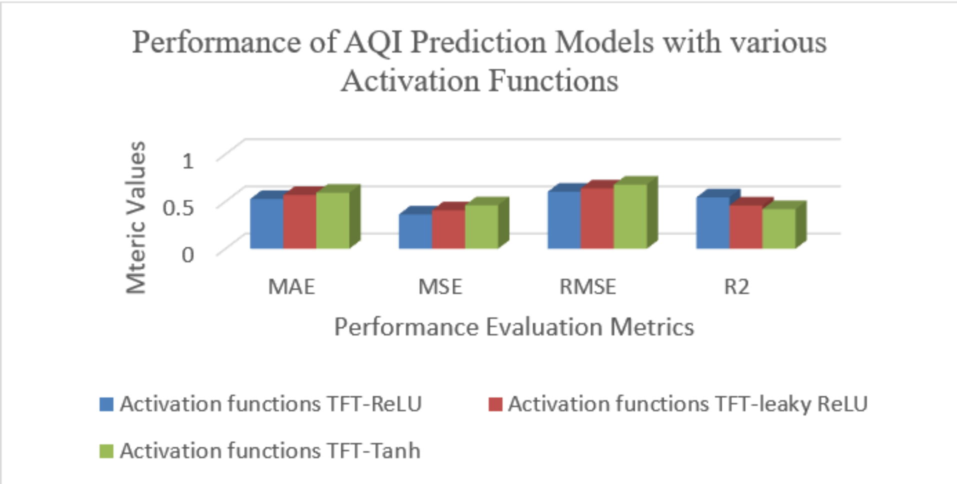

Figure 6. Performance analysis of AQI model with various activation functions

4.2. Results of AQI prediction model across the range of dropout rates

Dropout enhances model performance by mitigating overfitting, promoting feature independence, and improving robustness, preventing NN from relying too heavily on specific neurons [25]. Choosing the appropriate dropout rate is essential to avoid underfitting. Experiment is conducted by altering dropout values in conjunction with the ReLU activation function across 50 epochs and the results obtained are provided in Table 5. In assessing the impact of varying dropout rates (0.2, 0.3, and 0.4) with a ReLU activation function across 50 epochs, several performance metrics were considered. Notably, the MAE exhibited a decreasing trend as the dropout rate increased. Similarly, the RMSE displayed its best performance at a dropout rate of 0.3, implying improved precision in model predictions. The MSE is low at a dropout rate of 0.3. Lastly, the R2 score, which measures the goodness of fit, achieved its highest value at a dropout rate of 0.3, suggesting that this configuration resulted in the best overall model performance and ability to explain the variance in the data compared to the other dropout rates. The comparative results are provided in Figure 7. The x axis represents performance metrics. The y axis represents values of each metric corresponding to different dropout rates that is represented using coloured bars.

Table 5. Performance of the TFT_AQI_PMF_TL model with different drop out rates

Performance metrics	Dropouts		
	0.2	0.3	0.4
MAE	0.6423	0.6139	0.6922
RMSE	0.7482	0.7229	0.7985
MSE	0.5598	0.5226	0.6376
R2	0.5264	0.5946	0.4891

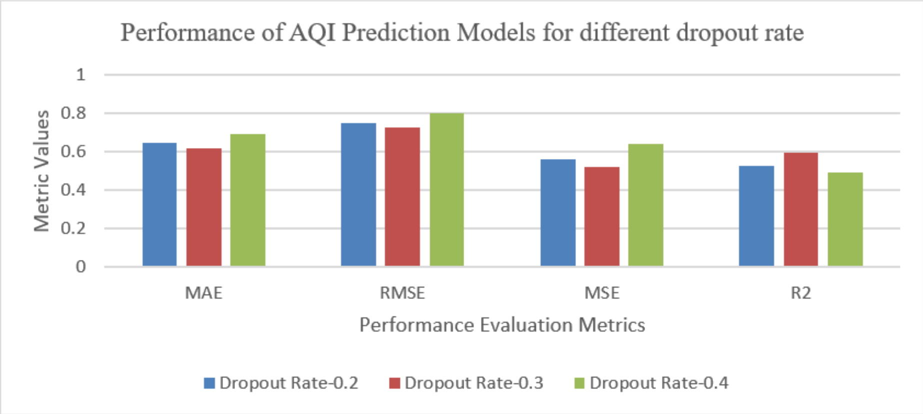

Figure 7. Performance analysis of AQI prediction model with different dropout rate

4.3. Result of AQI prediction model with different epochs

Epochs enhance model performance by allowing iterative learning, improving accuracy, and enabling the model to capture complex patterns. In the experiment, the ReLU activation function and a dropout rate of 0.3 were employed, while values were systematically adjusted over varying epochs and the results obtained are provided in Table 6. When comparing the results observed at the various epochs, the results observed at the epoch 100 seem to be better when compared to other epochs. Low MAE, RMSE, and MSE are obtained at the epoch 100. High R2 value of 0.7332 was obtained at the epoch 100 whereas low R2 value of 0.5946 was obtained for the epoch 50. The same is illustrated in Figure 8. The x axis represents performance metrics. The y axis represents values of each metric corresponding to different epochs that are represented using coloured lines.

Table 6. Result of TFT_AQI_PMF_TL models with various epochs

Performance metric	TFT_AQI_PMF_TL		
	Epoch – 50	Epoch-100	Epoch-150
MAE	0.6139	0.4285	0.6143
RMSE	0.7229	0.5756	0.7010
MSE	0.5226	0.3314	0.4915
R Squared	0.5946	0.7332	0.6239

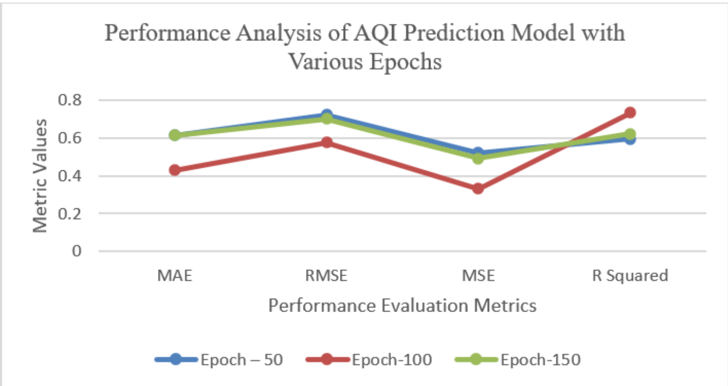

Figure 8. Performance analysis of AQI model with various epochs

4.4. Comparative analysis of base models vs transfer learning-based air quality index model

The effectiveness of the transfer learning-driven model, TFT_AQI_PMF_TL, is assessed by contrasting it with the foundational model, TFT_AQI_PMF_Cochin. The comparison, as depicted in Table 2, reveals that the transfer learning-based approach consistently outperforms the baseline models across all evaluated criteria. This is evidenced by the transfer learning models exhibiting lower values for MAE, RMSE, and MSE in comparison to the baseline models. The results observed for base model and transfer learning-based model is provided in Table 7 and illustrated in Figure 9. The x axis represents performance metrics. The y axis represents values of each metric corresponding to different models such as base model and transfer learning-based model.

Table 7. Performance of base model and transfer learning-based model

	MAE	RMSE	MSE	R2
TFT_AQI_PMF	0.5013	0.6534	0.4270	0.6543
TFT_AQI_PMF_TL	0.4285	0.5756	0.3314	0.7332

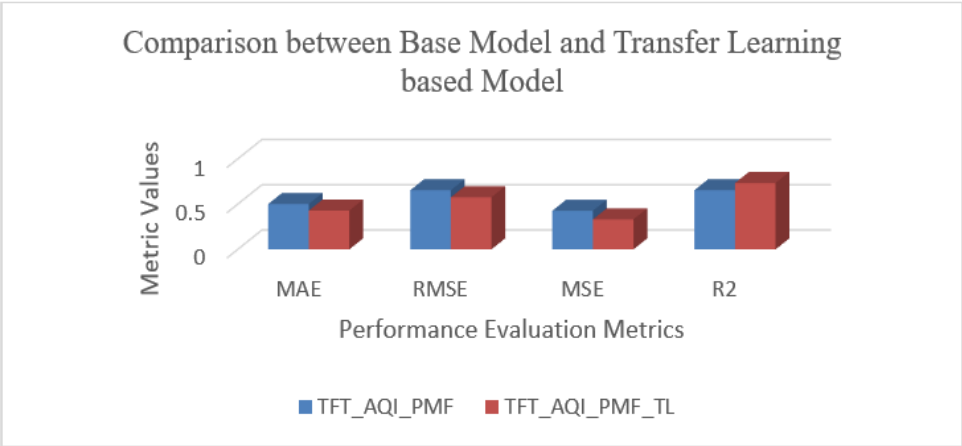

Figure 9. Comparative performance analysis between base model and transfer learning-based model

The transfer learning based TFT_AQI_PMF_TL model achieves the lowest MAE value of 0.4285, indicating superior predictive accuracy, in contrast to the relatively higher MAE observed with the baseline TFT_AQI_PMF_Cochin model. Furthermore, the transfer learning model demonstrates a high R2 of 0.7332, indicative of its better overall goodness-of-fit, while the base model, TFT_AQI_PMF_Cochin, yields a comparably lower R2 of 0.6543. The findings of this study highlight the successful development of an accurate AQI model, specifically tailored for regions with limited data instances. Employing TFT, a powerful DDN, the model is meticulously crafted. TFT's proficiency in capturing temporal patterns, handling diverse inputs including pollutant and meteorological data, and estimating uncertainties makes it a standout in AQI prediction. Due to the limited data available, the initial AQI model designed for the Cochin region did not perform well. However, by incorporating TFT alongside transfer learning, the model's performance was notably improved. Leveraging the pre-existing model trained on the Trivandrum region, which possessed a more substantial dataset, played a pivotal role in significantly enhancing the accuracy of AQI predictions for Cochin. Additionally, fine-tuning hyperparameters further contributed to optimize the model's efficiency.

5. CONCLUSION

This study aims to build an effective AQI prediction model for a region with data scarcity issue. The model is constructed through the integration of temporal fusion transfer and transfer learning technique. The dataset used in this research is obtained from the central pollution control board portal and the visual crossing website, encompassing both meteorological and pollutant data specific to the Cochin region. The prediction model is built using a special transformer-based architecture TFT. The performance of the model is further empowered with the application of transfer learning. A significant aspect of this approach is the utilization of knowledge from a pre-trained model developed using data from the Trivandrum region. By effectively applying transfer learning in conjunction with TFT, the model becomes more adaptable and precise, capitalizing on the insights gained from the Trivandrum model to enhance predictions in the Cochin region. To assess the model's performance, a comparison is made between the base model and the TFT-based transfer learning model. The results clearly demonstrate the superiority of the transfer learning-based model in all the aspects. Future work could extend this study by enriching the dataset with multimodal data and implementing heterogeneous transfer learning techniques. This approach would potentially enhance the model's ability to leverage diverse types of information sources, improving its overall performance and robustness across varied data modalities and domains.

ACKNOWLEDGEMENTS

We are grateful to PSGR Krishnammal College for Women for providing the academic environment and resources essential for this research.

REFERENCES

[1] R. J. Sram *et al.*, "Teplice Program: The Impact of Air Pollution on Human Health," *Environmental Health Perspectives*, vol. 104, p. 699, Aug. 1996, doi: 10.2307/3432707.
[2] S. S. Ganesh, S. H. Modali, S. R. Palreddy, and P. Arulmozhivarman, "Forecasting air quality index using regression models: A case study on Delhi and Houston," in *2017 International Conference on Trends in Electronics and Informatics (ICEI)*, IEEE, May 2017, pp. 248–254, doi: 10.1109/ICOEI.2017.8300926.
[3] Á. Leelőssy, F. Molnár, F. Izsák, Á. Havasi, I. Lagzi, and R. Mészáros, "Dispersion modeling of air pollutants in the atmosphere: a review," *Open Geosciences*, vol. 6, no. 3, Jan. 2014, doi: 10.2478/s13533-012-0188-6.
[4] A. Rowley and O. Karakuş, "Predicting air quality via multimodal AI and satellite imagery," *Remote Sensing of Environment*, vol. 293, p. 113609, Aug. 2023, doi: 10.1016/j.rse.2023.113609.
[5] S. Singh and A. Singh, "Enhancing Air Quality Prediction through Linear Regression," *International Research Journal of Modernization in Engineering Technology and Science*, Sep. 2023, doi: 10.56726/IRJMETS44874.
[6] Q. Liao, M. Zhu, L. Wu, X. Pan, X. Tang, and Z. Wang, "Deep Learning for Air Quality Forecasts: a Review," *Current Pollution Reports*, vol. 6, no. 4, pp. 399–409, Dec. 2020, doi: 10.1007/s40726-020-00159-z.
[7] B. Das, Ö. O. Dursun, and S. Toraman, "Prediction of air pollutants for air quality using deep learning methods in a metropolitan city," *Urban Climate*, vol. 46, p. 101291, Dec. 2022, doi: 10.1016/j.uclim.2022.101291.
[8] N. S. Gupta, Y. Mohta, K. Heda, R. Armaan, B. Valarmathi, and G. Arulkumaran, "Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis," *Journal of Environmental and Public Health*, vol. 2023, pp. 1–26, Jan. 2023, doi: 10.1155/2023/4916267.
[9] L. Shankar and K. Arasu, "Deep Learning Techniques for Air Quality Prediction: A Focus on PM2.5 and Periodicity," *Migration Letters*, vol. 20, no. S13, pp. 468–484, Dec. 2023, doi: 10.59670/ml.v20iS13.6477.
[10] N. Sarkar, R. Gupta, P. K. Keserwani, and M. C. Govil, "Air Quality Index prediction using an effective hybrid deep learning model," *Environmental Pollution*, vol. 315, p. 120404, Dec. 2022, doi: 10.1016/j.envpol.2022.120404.
[11] S. Halsana, "Air Quality Prediction Model Using Supervised Machine Learning Algorithms," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, pp. 190–201, Jul. 2020, doi: 10.32628/CSEIT206435.
[12] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal Fusion Transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, Oct. 2021, doi: 10.1016/j.ijforecast.2021.03.012.
[13] B. Ayhan, E. P. Vargo, and H. Tang, "On the exploration of temporal fusion transformers for anomaly detection with multivariate aviation time-series data," *Aerospace*, vol. 11, no. 8, p. 646, doi: 10.3390/aerospace11080646.
[14] S. F. Stefenon, L. O. Seman, L. S. A. da Silva, V. C. Mariani, and L. dos S. Coelho, "Hypertuned temporal fusion transformer for multi-horizon time series forecasting of dam level in hydroelectric power plants," *International Journal of Electrical Power & Energy Systems*, vol. 157, p. 109876, Jun. 2024, doi: 10.1016/j.ijepes.2024.109876.
[15] K. Sahoo, A. K. Samal, J. Pramanik, and S. K. Pani, "Exploratory Data Analysis using Python," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 12, pp. 4727–4735, Oct. 2019, doi: 10.35940/ijitee.L3591.1081219
[16] V. Hodge and J. Austin, "A Survey of Outlier Detection Methodologies," *Artificial Intelligence Review*, vol. 22, no. 2, pp. 85–126, Oct. 2004, doi: 10.1023/B:AIRE.0000045502.10941.a9.
[17] M. Weber, M. Auch, C. Doblander, P. Mandl, and H.-A. Jacobsen, "Transfer Learning With Time Series Data: A Systematic Mapping Study," *IEEE Access*, vol. 9, pp. 165409–165432, 2021, doi: 10.1109/ACCESS.2021.3134628.
[18] P. Mignone, G. Pio, S. Džeroski, and M. Ceci, "Multi-task learning for the simultaneous reconstruction of the human and mouse gene regulatory networks," *Scientific Reports*, vol. 10, no. 1, p. 22295, Dec. 2020, doi: 10.1038/s41598-020-78033-7.
[19] S. T. Krishna and H. K. Kalluri, "Deep Learning and Transfer Learning Approaches for Image Classification," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 7, no. 5S4, pp. 427–432, 2019.
[20] G. Vrbancic and V. Podgorelec, "Transfer Learning With Adaptive Fine-Tuning," *IEEE Access*, vol. 8, pp. 196197–196211, 2020, doi: 10.1109/ACCESS.2020.3034343.
[21] K. R. Weiss and T. M. Khoshgoftaar, "An Investigation of Transfer Learning and Traditional Machine Learning Algorithms," in *2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)*, IEEE, Nov. 2016, pp. 283–290, doi: 10.1109/ICTAI.2016.0051.
[22] A. Hosna, E. Merry, J. Gyalmo, Z. Alom, Z. Aung, and M. A. Azim, "Transfer learning: a friendly introduction," *Journal of Big Data*, vol. 9, no. 1, p. 102, Oct. 2022,doi: 10.1186/s40537-022-00652-w.
[23] M. R. Hossain and D. Timmer, "Machine Learning Model Optimization with Hyper Parameter Tuning Approach," *Global Journal of Computer Science and Technology*, vol. 21, no. 2, 2021.
[24] K. He, L. Zheng, Q. Yang, C. Wu, Y. Yu, and Y. Zou, "Crude oil price prediction using temporal fusion transformer model," *Procedia Computer Science*, vol. 221, pp. 927–932, 2023, doi: 10.1016/j.procs.2023.08.070.
[25] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural

Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

BIOGRAPHIES OF AUTHORS

Santhana Lakshmi Velusamy is a Research Scholar in the Department of Computer Science, PSGR Krishnammal College for Women. She completed her post graduation M.Sc. Information Technology. Her area of expertise includes machine learning, data mining, Python, and Pyspark. She can be contacted at email: sanlakmphil@gmail.com.

Vijaya Madhaya Shanmugam is an Associate Professor and currently serves as the Head of the Computer Science Department at PSGR Krishnammal College for Women. She completed her master's degree at PSG College of Technology and earned her Ph.D. from Amrita University, Coimbatore. Her areas of expertise include data mining, machine learning, support vector machines, and pattern recognition. She has successfully guided numerous M.Phil. Research Scholars and Ph.D. Scholar, with several other Ph.D. candidates having submitted their theses and awaiting viva voce. She can be contacted at email: msvijaya@psgrkcw.ac.in.