Enhancing Urban Cities Air Quality Prediction Using Deep Learning Techniques

1st Ms.Kovuru Lourd Victoria
Computer Science and Engineering
Karunya Institute of Technology and Sciences
Coimbatore, India
kovurulourd@karunya.edu.in

3rd Dr.G.Naveen Sundar
Computer Science and Engineering
Karunya Institute of Technology and Sciences
Coimbatore, India
naveensundar@karunya.edu

2ndDr.D.Narmadha Naveen
Computer Science and Engineering
Karunya Institute of Technology and Sciences
Coimbatore, India
narmadha@karunya.edu

***Abstract*—The environment has been greatly impacted in recent years by significant amounts of industrial modulation. Air quality is a significant ecological factor that has an enormous effect on human health, ecosystems, and the climate. Clear air is a crucial component for the survival of a being. With growing concern over air pollution and its detrimental effects, The need for precise and quick air quality prediction models is increasing. As air quality is dynamic and complex, predicting it is a challenging task for atmospheric conditions and pollutants, and for implementing effective pollution control measures, issuing timely health advisories, and promoting public health and environmental well-being. Various additionals, such as meteorological variables, geographical features emission sources, and chemical reactions, influence air quality. Accurate air quality prediction has numerous practical applications. It allows governments and environmental agencies to implement effective pollution control measures and issue timely health advisories to protect public health. The general public can also benefit from real-time air quality forecasts by making informed decisions, such as adjusting outdoor activities or using air purifiers during poor air quality days. Here the data has been gathered from the places, in India. As technology advances and more data becomes available, the potential for improved air quality prediction continues to grow, contributing to a healthier and more sustainable future for all. Deep learning algorithms most importantly LSTM, and BiLSTM contribute a larger scope for the application. The proposed models for air quality prediction are acquired through the BiLSTM model, which produces a lower NMSE of 0.002 compared with LSTM of 0.008.**

***Index Terms*—LSTM, BiLSTM, DNN, CNN, DaBiLSTM, NMSE, RMSE, MAPE, Nitrogen dioxide, sulfur dioxide, particular matter, Carbon Monoxide, Volatile Organic Compounds, ozone, PM2.5, NCC, COTS, SMA, KAMA, LSTM-FCN, VA, Hyperparameters.**

I. Introduction

Air quality is a vital aspect of our environment, profoundly impacting human health, quality of life, and the overall well-being of ecosystems. With urbanization, industrialization, and increasing vehicular emissions, concerns over air pollution have become more pressing than ever before. However, the quality of the air we breathe has been compromised by a wide range of pollutants, including particulate matter (PM), nitrogen dioxide (NO2), sulfur dioxide (SO2), carbon monoxide (CO), volatile organic compounds (VOCs), and ground-level ozone (O3). Continuous contact with these irritants may trigger respiratory diseases, as well as heart-related difficulties, and even early mortality. Furthermore, air pollution adversely affects vegetation, wildlife, and the integrity of ecosystems. Delhi, India's capital city, is renowned for its prosperous heritage, bustling streets, and vivacious atmosphere. However, in recent years, it has also gained notoriety for a pressing issue that poses a severe threat to its residents' health and well-being: air pollution. Delhi's air quality crisis has garnered global attention due to its severe and persistent nature These pollutants, also include fine particulate matter (PM2.5) that have grave consequences for public health. considering this a critical issue in executing a fine-tuned deep-learning model that greatly satisfies the application needs.

II. Related Work

Deep learning techniques have been extensively employed in air quality prediction due to air quality data's ability to expose complicated temporal and geographic patterns. A range of hybrid architectures and LSTM have been utilized for air quality prediction. Air quality prediction with deep learning is attainable via Deep Air. Researchers have used LSTM and BiLSTM architectures in Deep Air to model the temporal dependencies in air quality data. This approach has been applied to various cities for short and long terms for air quality predicting. ([1] DeepAir). LSTM and BiLSTM models are performed in short-term traffic data-driven models. It was proposed to predict parameters such as travel time, speed, volume, and density. The model includes ARIMA to help with traffic problems [2]Certainly, floods are among the most common and devastating natural disasters, and they represent a significant danger to our communities. Having practical and efficient methods for predicting floods is incredibly important because it allows us to respond quickly to prevent and mitigate the disaster's impact. This, in turn, helps save lives, protect property, and maintain the overall stability of our society. To address these various analysis methods are used such as signal processing [3] Wavelet transform, and probabilistic approach used to analyze the data related to timing behavior.LSTM is a specialized type of neural network.[4] recurrent neural networks (RNNs). It was developed by researchers to address challenges in capturing and remembering important information over long periods in data sequences.

Research on estimating uncertainty in air quality predictions using LSTM and BiLSTM models has also progressed. Uncertainty quantification is crucial for decision-making and risk assessment. [8]Beijing, China, hourly PM2.5 concentration forecast on a spatiotemporal causal convolutional network. The use of multimodal LSTM models that incorporate data from multiple sources, such as meteorological data, traffic data, and satellite imagery, in addition to air quality data. These models aim to improve prediction accuracy by considering a wider range of factors influencing air quality. [6]Zhang, L., Lin, W., Kong, X. (2018). Multimodal LSTM-Based Air Quality Prediction Model.[7]Based on the wide range of performance of neural networks produces patterns that were a better stand for predicting the FOREX Rate within a 60-minute duration, the proposed approach determines the market's EUR/USD exchange rate. Through a trained neural network model inserting 4 currency pairs as an input. experimental metrics includes MAE,MSE,NCC(NORMALIZED CORRELATION COEFFICIENT),RMSE.[8]Crude oil Time series LSTM network-based Model prediction with optimization on chaotic Henry gas solubility.To develop a highly accurate forecasting tool that can maintain sustainable crude oil prices by analysis of chaotic COTS and non-linear dynamic behavior. Along with several momentum oscillators, the EMA, SMA, and KAMA indicators are used.[14]Time-series prediction utilizing transductive LSTM: A weather forecasting application, Building an LSTM model from the dataset gathered at the weather underground website with real measurements determining the best model by performance metrics MSE, MAE with the developed parameters of LSTM as tanh and sigmoid. At the end an LSTM model for temperature prediction.[10]Multivariate classification, LSTM-FCN is used for classifying univariate time series. Applied to other variant classification problems. The proposed model converts univariate to multivariate variants. state-of-the-art algorithms are compared with the proposed model.[16]] Integrating Two-Task Deep LSTM Networks for Combined Learning of RUL Prediction and Degradation Assessment for Aeroengines, Implementing Diagno-sis and Prognostics Together learning model Firstly, the difference between diag-nosis and prognostics is observed. Secondly, reduces the complex feature design and humanless experience. Metrics include RMSE and classification accuracy.[15]Bidirectional layer formation through matrices using lowercase and uppercase letters then Temporal Attention-Augmented Bilinear Layer is to predict se-quence-to-sequence learning.[17] High-frequency trading forecasting is challenging in trading. A technique for automated inference could assist in speed and accuracy. The proposed model incorporates bilinear projection along with an attention mechanism.[18] Inability to get ideal settings using useful tools for MA. MA optimization was used on the historical dataset. MA bias detection was included for multiple MA procedures.[19]It is stated as the most powerful approach for time-series processes. Deep Belief Echo State Network addresses the slow convergence in DBN.[20]Weerakody, P., et al. (2021).Time-series data are increased due to multisensor systems and are unstructured. Conventional methods are used for handling such data.[21]Zhiwei Xue (2020)Discussed the inaccurate predictions on the battery life of lithium-ion batteries using an AUKF method and GA-SVR.[22]states that the forecasting model is required for many applications.LSTF (lONG SHORT TERM FORECASTING. The probSparse self-attention mechanism was used to process the forecasting.[23]Elman, J. L. (1990). Representation of time is most important. Implicitly representing the time series is far better than explicitly representing it.[24]Gallicchio, C. (2011).Determines the limitations of the ESN model that are used in RNNs over the Architectural model. [25]Determines the analysis of the model-generated time series reservoir computing using non-linear VA.[26] Gil-Alana, L. A. (2004) Temperatures in Australia are measured based on time using fractional integration techniques.[27] Gonon, L. (2019).Stochastic discrete-time semi-infinite input reservoir computers. Lp criteria were used.[28]Usage of deep neural networks that were prosperous in both industry and research. Inquiries on the topic of knowledge distillation from the viewpoints of knowledge categories[29] For simpler inference algorithms, CP is used to eliminate the explaining-away effects like greedy can learn faster are defined.[30] Hochreiter, S., Schmidhuber, J. (1997) Limitation of backpropagation learning that it is time-consuming.LSTM corresponds with both space and time, and it can manage severe artificial long-time lag challenges.

Bidirectional LSTM -For sequential data processing, particularly time series analysis, speech recogni-tion, and language processing, the BiLSTM model is an example of RNN architec-ture. It is an extension of the traditional LSTM (Long Short-Term Memory) model that can capture dependencies in data both in the past and future, as opposed to just one direction (past to future) in the case of traditional LSTMs. In a BiLSTM, The input sequence is simultaneously induced in two ways: from the start to the end (forward) and from the end to the start (backward). This means that at each time step, there are one of the two separate LSTM layers takes the sequence start-ing at the beginning. and the other processing it from the end.

Table 1 Data represents the positions of the cities and their period over location using Latitude and Longitude accordingly. This helps us to maintain and track the prediction of the components more accuracte.The above model flow begins from collecting the dataset. Then pre-process the data corresponding

City	Latitude	Longitude	Time- period
Amaravati	16.5167° N	80.6167° E	24-11-2017
Amritsar	31.6330° N	74.8656° E	27-02-2017
Chandigarh	30.7333° N	76.7794° E	02-09-2019
Delhi	28.6139° N	77.2090° E	01-01-2015
”Gurugram	28.4595° N	77.0266° E	27-11-2015
Hyderabad	7.3850° N	78.4867° E	04-01-2015

TABLE I
DATASET

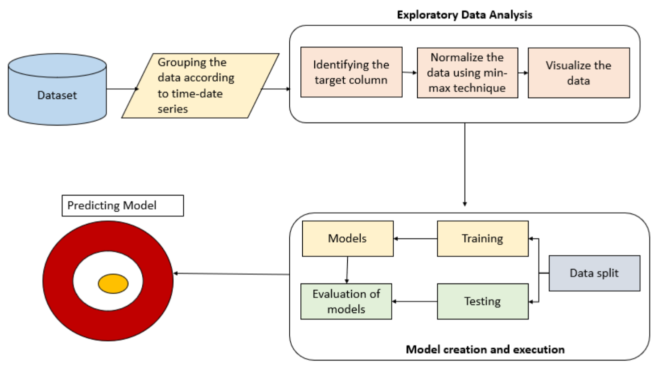

Fig. 1. Model Flow

to the variables. Pre-processed data undergoes Exploratory Data Analysis that includes techniques like normalization. Now split the data accordingly, then train the model with the training dataset thereby evaluating it using test data.

A. Exploratory Data Analysis

The initial processing starts by identifying the targeted column from the whole dataset that has been collected and loaded. Following normalizing the data, normalization can be handled either of the ways from min-max or z-score for this model min-max normalization was used to normalize the values. Getting more insights into data visualization is the best way that make it easier and simpler to understand. This process as a whole is termed EDA (EXPLORATORY DATA ANALYSIS).

B. Model Creation

This step includes the creation of the model starting from splitting the data as 80 and 20 as the training and testing phase.As the model is initialized with all required hyperparameters here LSTM and BiLSTM models are initialized. They are evaluated using the test data to find the best model using performance metrics RMSE, and NMSE.

C. Predicting Model

Finally the model with the least value of the metrics is considered to be the better predictive approach method for a good performance in representing the quality of air.

Applications-Bi LSTM's are commonly used in tasks such as POS tagging, NER, sentiment analysis, and MT. In NLP, they are effective at capturing context and dependencies in language. Fig.2 represents the components that are held inside the single layer of the LSTM model.

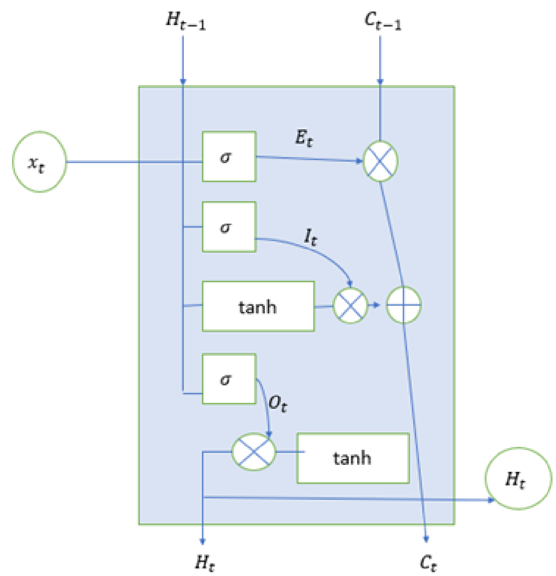

Fig. 2. Layer of LSTM

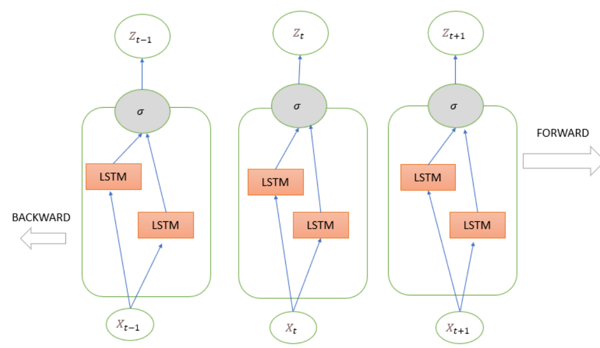

Fig. 3. Layers of BiLSTM

Strengths of Bi LSTMs-They can Capture prior and future understanding, which is particularly useful when the context for a prediction comes from both directions in a sequence. They are robust in handling long-range dependencies and preserving information over many time steps. They can help in improving the accuracy of various sequence-to-sequence tasks.

D. Abbreviations and Acronyms

Long Short-Term Memory, or LSTM NMSE means Normalized Mean Squared Error; MSE means Mean Squared; MAE refers for Mean Absolute Error; VA stands for Vector Autoregression; and MAPE refers for Mean Absolute Percentage Error. Fig 2 Represents the single hidden layer of the LSTM model it contains components such as sigma, and tan to perform mathematical calculations with the formulated equations.

Fig 3 As the BiLSTM model resembles the formation of multiple LSTM layers this is an illustration of these hidden layers over the developed BiLSTM model.

E. Performance Matrix

- **NMSE =MSE/Variance of the Actual Data**
  NMSE is a variation of MSE that is normalized to provide a more interpretable measure of the degree of compatibility between a model's predictions and the actual data. It is often used to compare different models or datasets
- **MSE= (1 / n) * $\sum$(actual - predicted)**2**
  The difference between the real data and the predictions of a model is determined as MSE. Getting more intensely aware of larger errors than smaller ones makes it prone to outliers.
- **MAPE = (1 / n) * $\sum$((Actual - Predicted) / Actual)*100**
  MAPE is a commonly used metric in statistics and data analytics for as-sessing the accuracy of forecasts or predictions, specifically with the topic of forecasting and time series analysis. MAPE evaluates the cor-rectness of a model by computing the average percentage difference be-tween the actual and anticipated values. A forecasting model's error size may be evaluated by expressing this as a percentage.

F. Equations

Suppose yt and Ht-1 These hidden layers are measured using these equations, given the ordered input data and the output state at time t.

$$at = \sigma(xf[bt-1, Wt] + cf) \tag{1}$$

$$It = \sigma(viyi + xiHt - 1 + bi) \tag{2}$$

$$Et = \sigma(vfyt + xfHt - 1 + bf) \tag{3}$$

$$Ot = \sigma(v0yt + x0Ht - 1 + b0) \tag{4}$$

$$Bt = \tanh(vcyt + xcHt - 1 + bc) \tag{5}$$

$$Ft = FtxorCt - 1 + ItxorBt \tag{6}$$

$$Ht = Otxortanh(Bt) \tag{7}$$

xi, xf, xo, and xc, which symbolize the recurrent xi, xf, xo, and xc, which symbolize the recurrent Weight matrices are assigned to four different gates: input, forgetting, output, and memory cell gates. They are represented by vi, vf, vo, and vc, respectively. These matrices are integral components within a neural network, playing pivotal roles in controlling information flow, retention, and processing at various stages of computation.

TABLE II
NMSE

Model	Layers		nmse	rmse
	HIDDEN	NO.OF.LAYERS		
LSTM	50	1	0.008	1.7
BiLSTM	64,32	1	0.0002	0.29
DaBiLSTM	64	1	2.8	0.09
DNN	64	1	0.049	4.08
CNN	64	10	5.7	43.9

TABLE III
MAPE

Model	dense layers	mape
LSTM	1	0.12
	25	0.1424
BILSTM	1	10.44
	25	14.83

TABLE IV
HYPER-PARAMETERS

LOSS	BATCHSIZE	RMSE
0.0074	1	0.39
0.0066	1	0.39
0.01	2	1.94
0.0112	3	4
0.0139	5	3.5

G. Analysis

Table I consists of the models that are being trained according to the data. It has parameters Layers that define the number of layers along with hidden layers are included to build, NMSE score determines the error rate, and RMSE helps in determining the variation between the actual value and the projected value. The model that "fits" the data performs better the lower the NMSE and RMSE values.

Table II has a score calculated for MAPE that determines the average percentage of variation between the expected and actual numbers to determine how accurate the model is then finally helps to select the model that acts as more accurate for the given series.

Figure fig.4 determines the relation between the Number of layers of created models and the performance matrix NMSE. Certainly, for the model, BiLSTM produces 0.0002 which is considered the most suitable model for the prediction of air quality.

The pictorial representation of the scores in Fig. 5 is calculated from the tuned model and is compared among dense layers. Here 1,2 on the horizontal plane represents the dense layers as 1 and 25 on the vertical plane is the range determining the score for performance measures rmse,nmse,mae, and mape for the LSTM model.

The following picture Fig. 6 interprets similar observations but through the developed BiLSTM model. Further, the model can also be tuned using parameters like dense, and number of hidden layers over the developed BiLSTM model. Table III contains the hyper-parameters that are used to tune the model. Various parameters help to uphold the model that can be deployed easily. LOSS and BATCH SIZE are represented

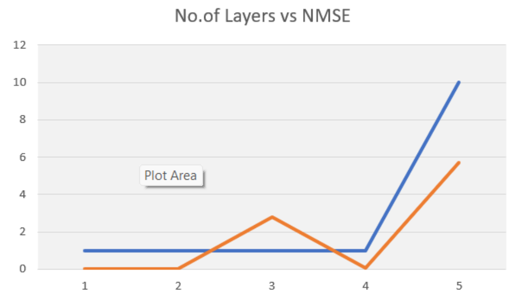

Fig. 4. NMSE VS LAYERS

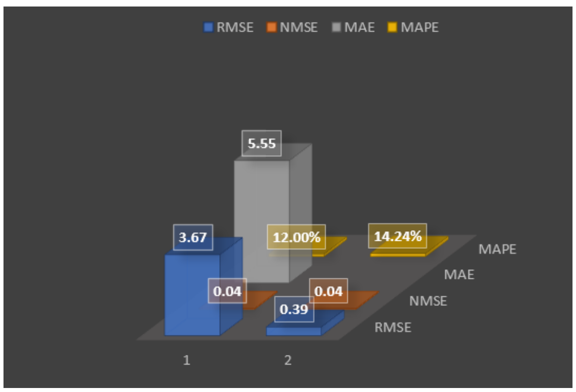

Fig. 5. LSTM

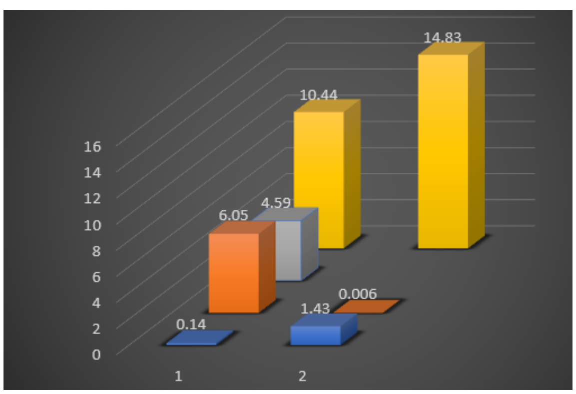

Fig. 6. bilstm

in a relation. The Loss described can be MSE and MAE but most developed models follow MSE as a retained parameter for loss. It varies(slightly) per Batch size. RMSE performance metrics have also shown a variation in the Batch size of the model.
The following Table IV gathers the values of hidden layers along with the NMSE score that varies accordingly with hidden layers. This is one of the hyperparameters that allows the model to work much more accurately.
Fig.7 determines the relation between the date and the component present in the air, PM2.5.The line plot with the axis as the year competing with the concentration value of components.

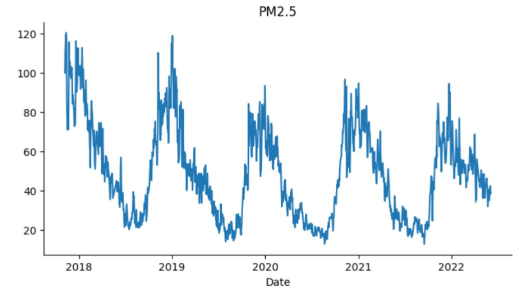

Fig. 7. Time vs PM2.5

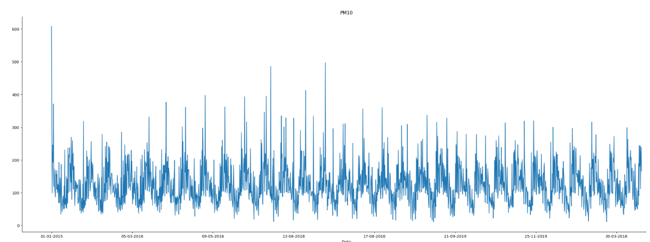

Fig. 8. Time vs PM10

In following Fig.8 discusses the relationship between the year and concentration of PM10.
Fig.9 contains the dimensional space divided as the batch size versus the RMSE score. The score batch size increases from 1,3,5 which makes a huge difference in the RMSE score. This hyperparameter is also important in training the model to be tuned properly to produce a minimum error and loss and determine the accurate model. Fig.10 represents the mean of all the components that are considered to be crucial pollutants of air as legends. The axis represents the place plotted over the concentration of each component.

III. Conclusion

The proposed model for Enhancing Urban Cities' Air Quality Prediction using deep learning techniques determines the BiLSTM model to be the best among the conventional models and the DaBiLSTM, LSTM model generated.BiLSTM model produces an RMSE value of 0.29 and NMSE of 0.0002 with the parameters included as Hidden layers and the number of LSTM layers built. Hyperparameter tuning includes dense layers and batch size optimizers these make the model much more optimized and produce effective values while predicting. To enhance the performance of the model, DAFA-BiLSTM (DEEP AUTOGRESSION FEATURE- AUGMENTED BiLSTM) is introduced which makes the prediction more efficient

TABLE V
Hidden Layers vs NMSE

HIDDEN	NMSE
50,50	0.004
64,64	0.011
32,64	0.048
64,64	0.03

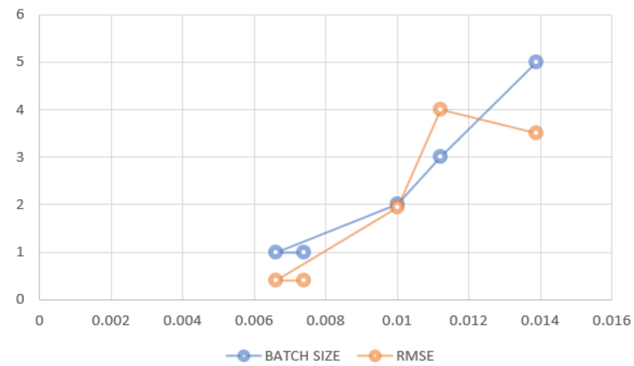

Fig. 9. comparision graph

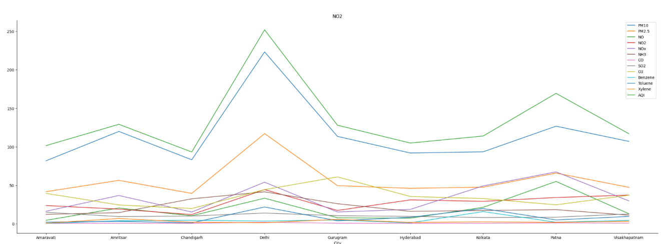

Fig. 10. City vs components

and accurate. It is an advanced deep-learning model. It produces an RMSE value of 0.096.

References

[1] Aytaç Altan a, Seçkin Karasu a, Enrico Zio b,c, "On a hybrid model for wind speed forecasting"

[2] JAnghinoni, Zhao on Time series trend detection and forecasting using complex network topology analysis.

[3] Bengio, Yoshuaa, b; Simard, Patricec; Frasconi, Paolo "Learning Long-Term Dependencies with Gradient Descent is Difficult.

[4] Steve A. Billings, X. Hong ."Dual-orthogonal radial basis function networks for nonlinear time series prediction ".

[5] Yiwei Cheng a, Chao Wang b, Jun Wu b, Haiping Zhu c, C.K.M. Lee d , "Multi-dimensional recurrent neural network for remaining useful life prediction under variable operating conditions and multiple fault modes" .

[6] J.B. Butcher a, D. Verstraeten b, B. Schrauwen b, C.R. Day a, P.W. Haycock a ,"Reservoir computing and extreme learning machines for non-linear time-series data analysis ".

[7] Zhiyong Cui, Ruimin Ke, Ziyuan Pu, Yinhai Wang , Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values .

[8] Yukai Ding a, Yuelong Zhu a, Jun Feng a, Pengcheng Zhang a, Zirun Cheng b , Interpretable spatio-temporal attention LSTM model for flood forecasting .

[9] Shoubo Feng a, Weijie Ren a, Min Han a, Yen Wei Chen b , Robust manifold broad learning system for large-scale noisy chaotic time series prediction: A perturbation perspective .

[10] Fazle Karim, Somshubra Majumdar, Houshang Darabi, Samuel Harford - Multivariate LSTM-FCNs for time series classification.LSTM-FCN(Long Short Term Memory Fully Convolutional Network) used for classifying univariate time series .

[11] Geoffrey E. Hinton, Simon Osindero, Yee-Whye Teh, A Fast learning algorithm for Deep belief Nets .

[12] Robert Jarusek, Eva Volna, Martin Kotyrba ,FOREX rate prediction improved by Elliott waves patterns based on neural networks .

[13] Seçkin Karasu, Aytaç Altan ,Crude oil time series prediction model based on LSTM network with chaotic Henry gas solubility optimization.

[14] Zahra Karevan, Johan A.K. Suykens , Transductive LSTM for time-series prediction: An application to weather forecasting .

[15] Dat Thanh Tran; Alexandros Iosifidis; Juho Kan- niainen; Moncef Gabbouj - Temporal Attention-Augmented Bilinear Network for Financial Time-Series .

[16] Huihui Miao; Bing Li; Chuang Sun; Jie Liu - Joint Learning of Degradation Assessment and RUL Prediction for Aero-engines via Dual-Task Deep LSTM Networks.

[17] Tran, D. T., Iosifidis, A., Kanniainen, J., et al. (2018). Temporal attention,augmented bilinear network for financial time-series data analysis. IEEE transactions on Neural Networks and Learning Systems.

[18] van Rossum, H. H., Kemperman, H. (2017). Optimization and validation of moving average quality control procedures using bias detection curves and moving average validation charts. Clinical Chemistry and Laboratory Medicine,55(2), 218–224

[19] Wang, H., Wu, Q. M. J., Xin, J., et al. (2020). Optimizing deep belief echo state network with a sensitivity analysis input scaling auto-encoder algorithm.Knowledge-Based Systems, 191, Article 105257.

[20] Weerakody, P. B., Wong, K. W., Wang, G., et al. (2021). A review of irregular time series data handling with gated recurrent neural networks. Neurocomputing.

[21] Xue, Z., Zhang, Y., Cheng, C., et al. (2020). Remaining useful life prediction of lithium-ion batteries with adaptive unscented kalman filter and optimized support vector regression. Neurocomputing.

[22] Zhou, H., Zhang, S., Peng, J., et al. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting.

[23] Elman, J. L. (1990). Finding structure in time. Cognitive Science. Feng, S., Ren, W., Han, M., et al. (2019). Robust manifold broad learning system for large scale noisy chaotic time series prediction: A perturbation perspective. Neural Networks.

[24] Gallicchio, C., Micheli, A. (2011). Architectural and Markovian factors of echo state networks. Neural Networks.

[25] Gauthier, D. J., Bollt, E., Griffith, A., et al. (2021). Next generation reservoir computing. Nature Communications.

[26] Gil-Alana, L. A. (2004). Long memory behavior in the daily maximum and minimum temperatures in Melbourne, Australia. Meteorol. Appl.

[27] Gonon, L.,Ortega, J. P. (2019). Reservoir computing universality with stochastic inputs. IEEE Transactions on Neural Networks and Learning Systems.

[28] Gou, J., Yu, B., Maybank, S. J., et al. (2021). Knowledge distillation: A survey. International Journal of Computer Vision, 129(6), 1789–1819

[29] Hinton, G. E., Osindero, S., Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. Neural Computation,

[30] Hochreiter, S.,Schmidhuber, J. (1997). Long short-term memory. Neural Computation.