Exogenous Regressors (SARIMAX), Simple Exponential Smoothing (SES), Holt Winter's Exponential Smoothing (HWES).

Many solutions were proposed to predict the future value of the O3 layer in the coming years, but these classical models were not so progressive as to be able to work on multivariate data to determine the level of the O3 layer that depends on other multiple environmental air quality factors [7].

In this study, we presented three approaches for multivariate and multi-step time series forecasting, Vector Auto Regression (VAR), Multi-layer perceptron (MLP) and Long Short Term Memory (LSTM) and compared their performance to analyze air pollution data in multivariate time series [13-18]. These forecast models not only incorporate the current data, but also what they have previously recognized in time to generate new O3 forecasts.

2. Material and Methodology

2.1 Data Source

The presented methodology was tested on air quality data from the open source machine learning repository of the UCI [8]. The dataset contains 9358 instances of average hourly responses from a range of 5 metal oxide chemical sensors embedded in an air quality chemical multi- sensor. The device was in a highly polluted area on the ground, on the road, in an Italian city. Data from March 2004 to February 2005(one year) representing the longest available records of the response of chemical sensor devices deployed in the field.

2.2 Data Preparation

In a data set of 9358 instances, missing values were found in some records, which were replaced by the mean attribute value. The challenge in pre-processing was to turn the multivariate time domain problem into a supervised learning problem where dependent and independent variables can be targeted. We had to perform the multivariate, multi time step forecasting, so that future forecasting could not only be done on current input but also on what occurred at previous state; that is the concept of multi-step forecasting. Since our goal was to predict the Ozone O3 level, we have defined O3 explicitly as our target variable [9-12].

The data is divided into a training and test cohort with a ratio of 60:40 respectively to train the model using the training data set and validate it with the test data set. Unlike normal machine learning models, this split is carried out in time without the instances being shuffled.

2.3. Forecasting Machine Learning Models

2.3.1. Vector Auto Regression (VAR)

It is a multivariate linear time series model design to capture the join dynamic of multiple time series. For forecasting purpose reduced form VAR's is sufficient. VAR model is a multi-equation system where all the variables are treated as dependent. The vector auto regression (VAR) model is one of the most successful, flexible, and easy to use models for the analysis of multivariate time series. It is a natural extension of the univariate autoregressive model. The VAR model is useful for describing the dynamic behavior of multivariate time series and for forecasting. The superior forecasts to those from univariate time series models and elaborate theory-based simultaneous equations models can be provided by using VAR models. Forecasting is quite flexible since they can be made conditional on the potential future paths of specified variables in the model [25].

2.3.2. Multi-Layer Perceptron (MLP)

A Multi-layer perceptron **(**MLP) is a deep artificial neural network. It is composed of more than one perceptron. They are consist of input layer to receive signal, an output layer that makes a decision to predict about the input and in between those two arbitrary number of hidden layer. If we have multiple hidden layer with nonlinear activation function then it gives better prediction. The perceptron consist of weight which is the summation processor and an activation function. The input values are presented to perceptron if the output which we predicted as same as the desire output then the performance is good and no changes to the weight are made. However if the output does not fulfilled the output which we desire. Then the weight need to be change to reduction of error.

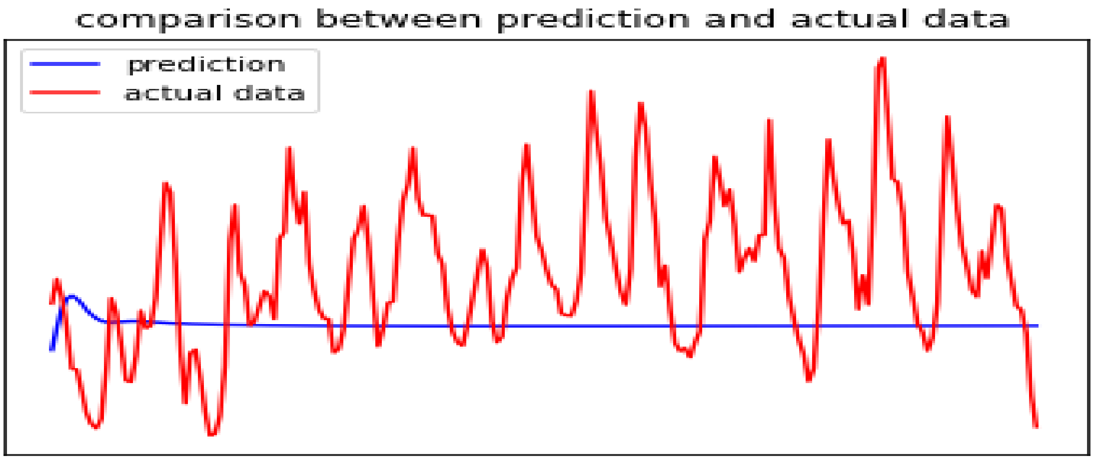
2.3.3. Long Short Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a specific recurrent neural network (RNN) architecture that was designed to model temporal sequences and their long-range dependencies more accurately than conventional RNNs. The LSTM contains special units called memory blocks in the recurrent hidden layer [23]. LSTM have chain like structure. From *forget gate* operation we take input from current time step and previous time step and concatenate them. We pass this value through a *sigmoid* function. Which give output between 0 and 1 through *update gate* operation we sum the value from current time step and also the previous time step. Then pass this value through a *tanh* function. We produce candidate value and by passing it through a *sigmoid* function. We choose values to be selected from candidate. From *output gate* operation we summed up the value from current time step and also from previous time step and pass it through a *sigmoid* function. To choose which value we use as output, we take the cell state and applying a *tanh* function which let only selected output.

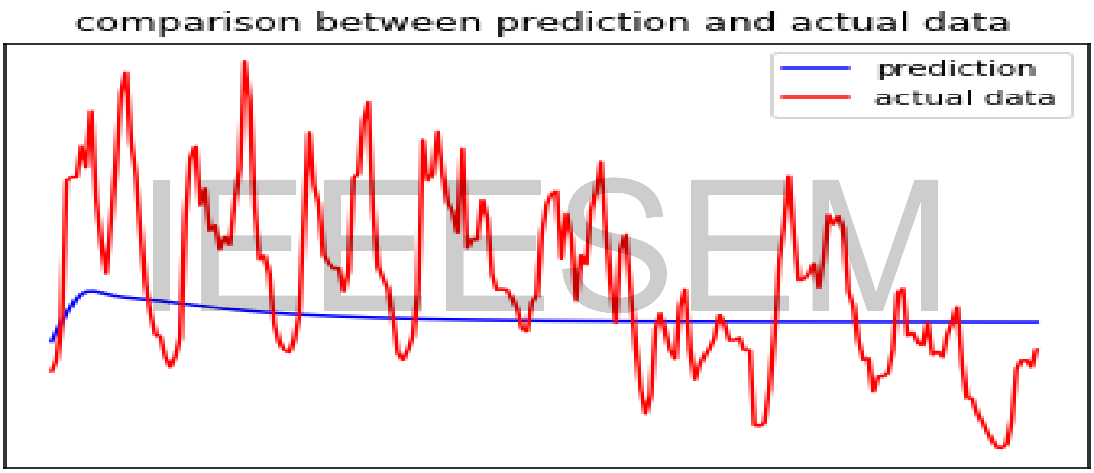
3. Experiment and Discussion

The forecast models used in this study to deal with multivariate and multi-step time series are Vector Auto Regression (VAR), Multi-layer perceptron (MLP) and Long Short Term Memory (LSTM). Training of the forecasting model is carried out on the data set of training [19-21]. Once our models have been trained, we evaluated these models further with unseen test data to validate our proposed models.

The comparison graphs between prediction and actual data on training and testing data reveals that how the forecasting models predicted the time series vs the actual time series on train and test cohort respectively.



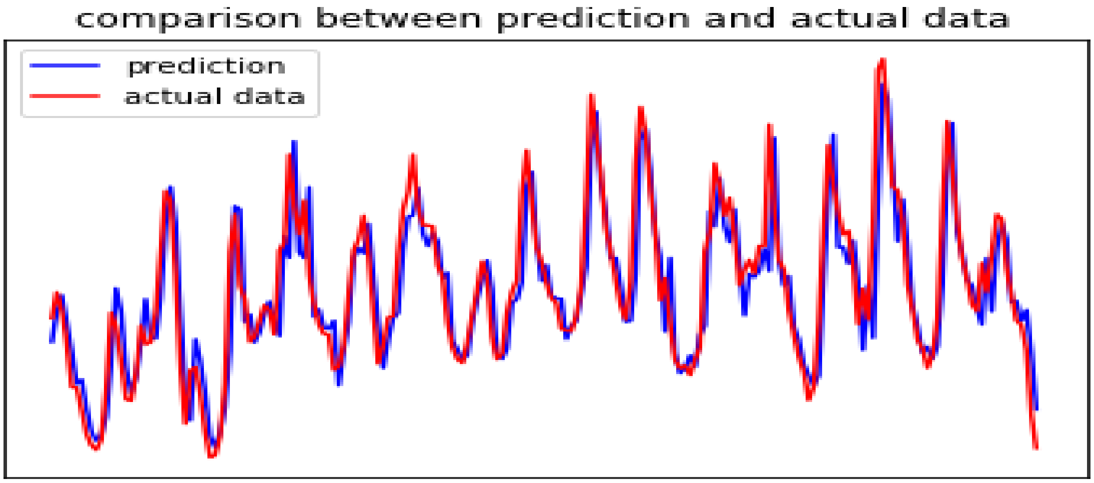
(a)



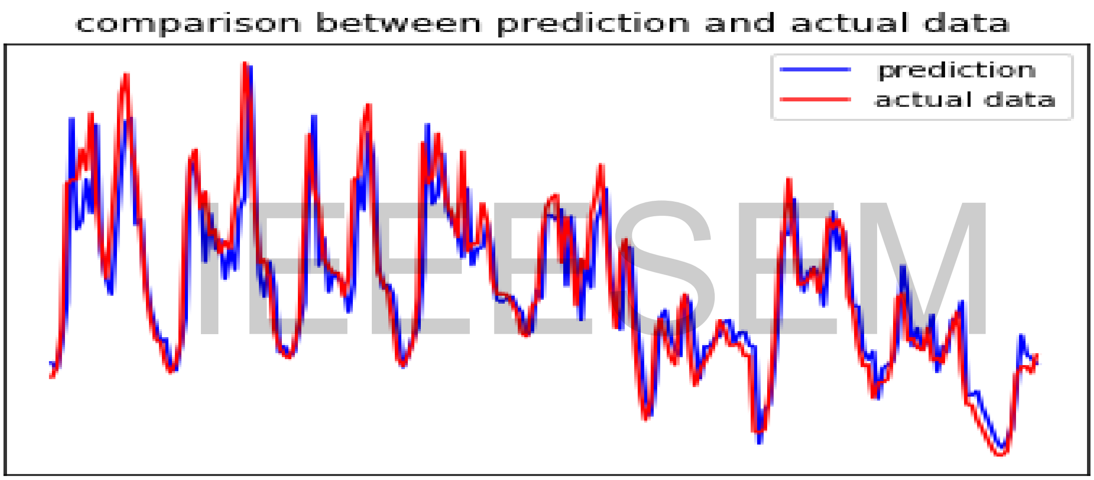
(b)

Fig. 1. (a) Shows the comparison between actual and predicted forecasting on train cohort using VAR, (b) presents the comparison between actual and predicted forecasting on test cohort using VAR

As shown above, the results of forecasting using VAR are not reliable and satisfactory for multivariate and multi-step time series forecasting to predict Ozone (O3) in our case.

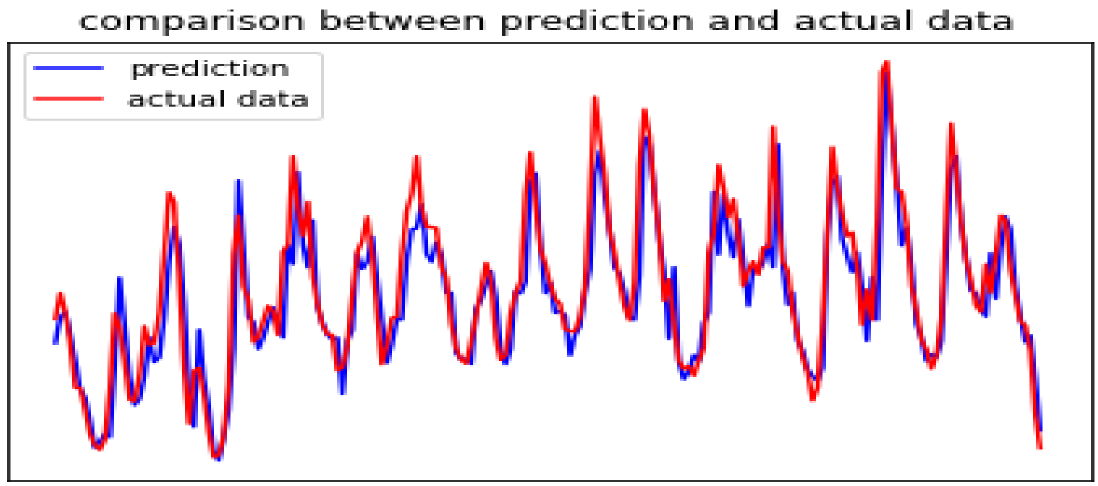


(a)

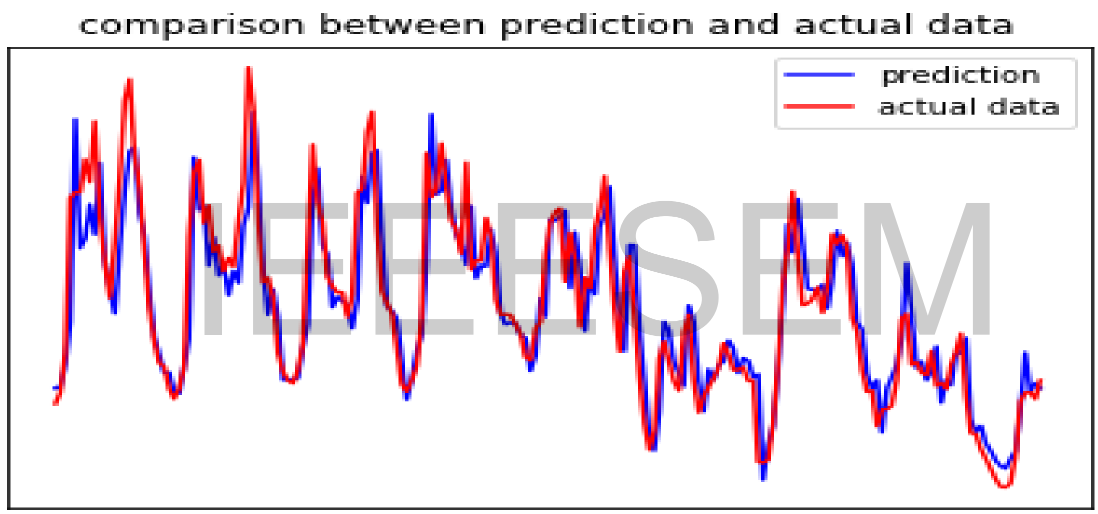


(b)

Fig. 2. (a) Shows the comparison between actual and predicted forecasting on train cohort using MLP, (b) presents the comparison between actual and predicted forecasting on test cohort using MLP



(a)



(b)

Fig. 3. (a) Shows the comparison between actual and predicted forecasting on train cohort using LSTM, (b) presents the comparison between actual and predicted forecasting on test cohort using LSTM

The comparison graphs between actual and predicted forecasting seems much better using MLP (Fig.2) and LSTM (Fig.3) models in comparison with VAR (Fig.1) to predict Ozone (O3). It is clear from the comparison graph that deep neural networks performed better on multivariate and multi-step time series to better manage multiple parameters of the time series forecasting with large window of dependencies to predict the future value of targeted variable, in our case, Ozone (O3).

4. Results

The performance metrics used to evaluate the quality of forecast time series models are Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). The results of the forecast metrics for predicting the O3 layer level are shown in Table 3 and Table 4. On the training dataset, it was seen that the forecasting model generated by Long Short Term Memory (LSTM) and Multi-layer perceptron (MLP) had the lowest errors rate, MAPE (< 14%) in contrast to Vector Auto regression (VAR) which had MAPE (>33%). However, when these trained multivariate and multi- stage forecast models were validated using test data to predict the O3, the LSTM [22-23] forecast model outperformed to predict the future value of O3 based on its previous input.

Table 1: Comparison of forecasting models on training data

Evaluation Matrices	VAR	MLP	LSTM
MAE	271.622	116.355	107.717
MAPE	33.202	12.715	11.266
RMSE	342.448	158.612	150.343

Table 2: Comparison of forecasting models on testing data

Evaluation Matrices	VAR	MLP	LSTM
MAE	346.21	116.16	116.01
MAPE	36.27	11.733	11.223
RMSE	448.03	159.79	159.39

In conclusion, the LSTM model has the lowest error rate (Fig.4, 5 and 6) in our case to predict the O3, because the LSTM model is more context- based and capable of learning long- term dependencies than using MLP, which is a key forecast.

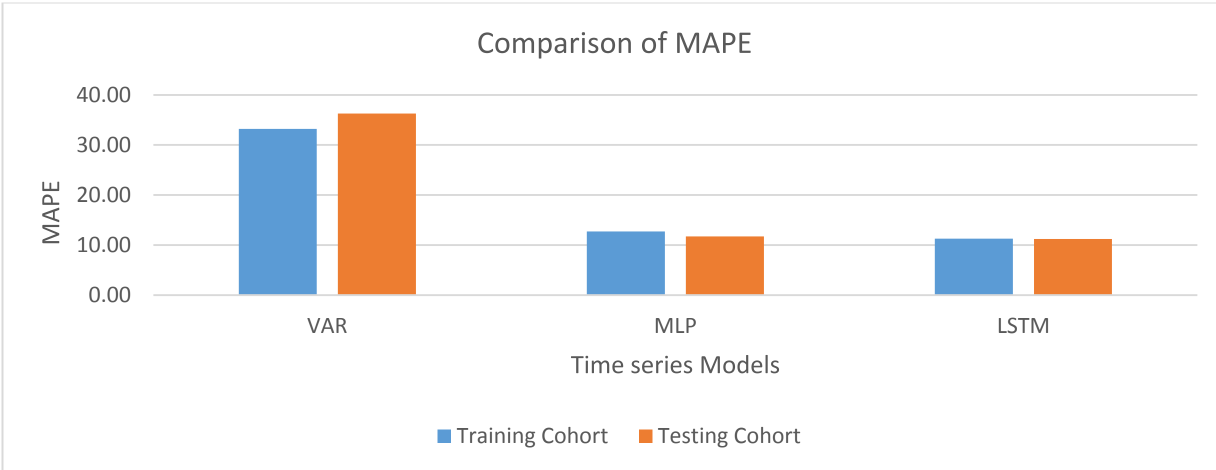


Figure 4: Comparison of MAPE

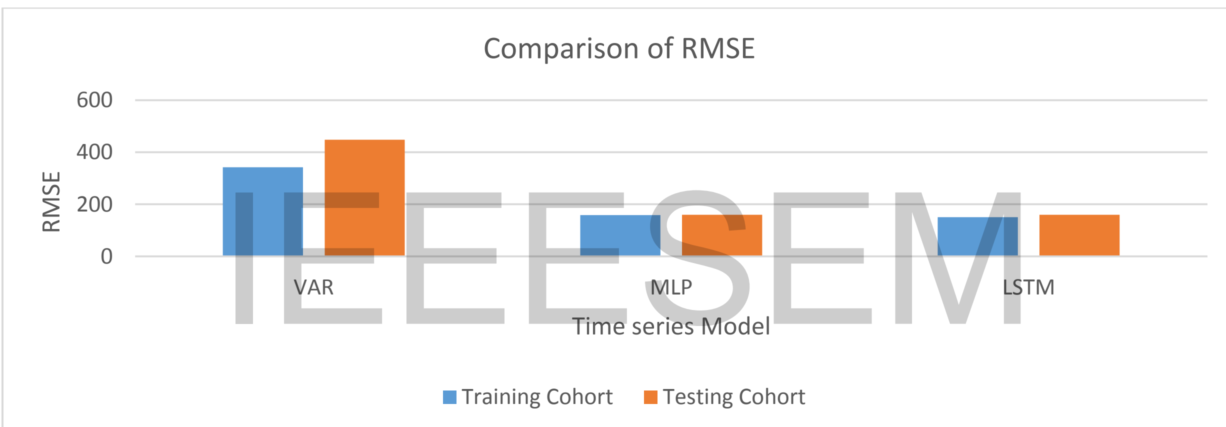


Figure 5: Comparison of RMSE

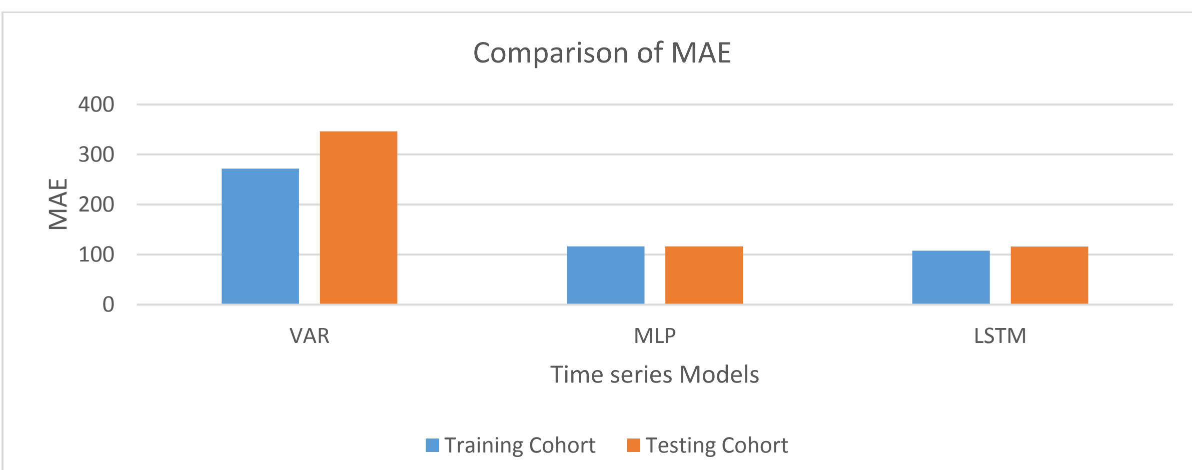


Figure 6: Comparison of MAE

Conclusion

In this study, we investigated recent advances in the prediction of ozone levels (O3) using multivariate and multi-step time series forecast models. We compared the VAR, MLP and LSTM models and as a result, the performance matrices (MAE, MAPE, and RMSE) using LSTM are 10%-20% more accurate then VAR and MLP models for prediction of Ozone (O3) future values based on previous N sequential, measurement record. We observed that the LSTM model is better able to learn long- term dependencies than to use MLP for Ozone (O3). This proposed approach is therefore used not only to forecast the multivariate time series, but can also handle multiple step time series with less amount of error ingesting real time inputs and generating future prediction for O3 layer.

Conflict of Interest: The author declares that they have no conflict of interest.

References

1. Sivasakthivel.T, K.K.Siva Kumar., 2011. Ozone Layer Depletion and Its Effects: A Review. International Journal of Environmental Science and Development, Vol.2.
2. Zheng, Y.; Liu, F.; Hsieh, H.-P. U-Air: When urban air quality inference meets big data. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11–14 August 2013.
3. Kalapanidas, E.; Avouris, N. Short-term air quality prediction using a case-based classifier. Environ. Model. Softw. 2001, 16, 263–272.
4. Kurt, A.; Oktay, A.B. Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. Expert Syst. Appl.2010, 37, 7986–7992.
5. Kleine Deters, J.; Zalakeviciute, R.; Gonzalez, M.; Rybarczyk, Y. Modeling PM2.5 urban pollution using machine learning and selected meteorological parameters. J. Electr. Comput. Eng. 2017, 2017, 5106045.
6. Bougoudis, I.; Demertzis, K.; Iliadis, L.; Anezakis, V.-D.; Papaleonidas, A. FuSSFFra, a fuzzy semi-supervised forecasting framework: The case of the air pollution in Athens. In Neural Computing and Applications; Springer: Berlin, Germany, 2017; pp. 1–14.
7. Tien-Cuong Bui, Van-Duc Le, Sang K. Cha., 2018. A Deep Learning Approach for Forecasting Air Pollution in South Korea Using LSTM
8. Repository UCI. [ENEA - National Agency for New Technologies, Energy and Sustainable Economic Development]. Available from: http://archive.ics.uci.edu/ml/datasets/air+quality
9. Alemdar, H., Caldwell, N., Leroy, V., Prost-Boucle, A., Petrot, F., 2016. Ternary ´ neural networks for resource-efficient AI applications. CoRR abs/1609.00222.
10. Durichen, R.; Pimentel, M.; Clifton, L.; Schweikard, A.; and Clifton,D. 2014. Multi-task Gaussian processes for multivariate physiological time-series analysis.
11. Artificial Neural Networks and Machine Learning – ICANN 2013. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 451–458.
12. Bengio, Y., 2009. Learning deep architectures for ai. Foundations and Trends in Machine Learning 2 (1), 1–127.
13. Sze, V., Chen, Y., Yang, T., Emer, J. S., 2017. Efficient processing of deep neural networks: A tutorial and survey. CoRR abs/1703.09039.
14. Romeu, P., Zamora-Mart´ınez, F., Botella-Rocamora, P., Pardo, J., 2013. Time-series forecasting of indoor temperature using pre-trained deep neural networks. In: Mladenov, V., Koprinkova-Hristova, P., Palm, G., Villa, A. E. P., Appollini, B., Kasabov, N.Canziani, A.,

Paszke, A., Culurciello, E., 2016. An analysis of deep neural network models for practical applications.

15. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature , vol. 521, no. 7553, pp. 436–444, 2015.
16. Gers, F. A., Eck, D., Schmidhuber, J., 2002. Applying lstm to time series predictable through time-window approaches. In: Tagliaferri, R., Marinaro, M. (Eds.), Neural Nets WIRN Vietri-01. Springer London, London, pp. 193–200.
17. Courbariaux, M., Bengio, Y., 2016. Binarynet: Training deep neural networks with weights and activations constrained to +1 or -1. CoRR abs/1602.02830.
18. P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long Short Term Memory networks for anomaly detection in time series," in Proceedings of the 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN '15), pp. 89–94, April 2015.
19. R. S. Tsay, Multivariate Time Series Analysis: with R And Financial Applications, John Wiley and Sons, New Jersey, NJ,USA, 2014
20. Wilson, J.H. and Keating, B. (1990). Business Forecasting (Homewood, III: Richard D. Irwin,)
21. P. J. Brockwell and R. A. Davis, Time Series:Theory and Methods, Springer, New York, NY, USA, 2ND edition, 2006.
22. Kompella, R., 2017. Using lstms to forecast time-series.
23. Has¸im Sak, Andrew Senior, Franc¸oise Beaufays. Long Short Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling. Google, USA
24. Hsu, D., 2017. Time series forecasting based on augmented long short-term memory. CoRR abs/1707.00666.
25. Selcuk Bayraci and Yakup Ari and Yavuz Yildirim., April 2011. A Vector Auto-Regressıve (VAR) Model for the Turkish Financial Markets. Yeditepe University