IJESRT: 7(11), November, 2018 ISSN: 2277-9655

International Journal of Engineering Sciences & Research Technology

(A Peer Reviewed Online Journal)
Impact Factor: 5.164

IJESRT

Website: www.ijesrt.com Mail: editor@ijesrt.com

Correspondence Address: 116, Sukhdev Nagar Ext-1, Airport Road, Indore – 452005, Madhya Pradesh, INDIA

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

MACHINE LEARNING APPROACHES FOR FORECASTING ENERGY CONSUMPTION USING MULTIVARIATE TIME SERIES DATA

Paril Ghori
parilghori@gmail.com

ABSTRACT

This paper investigates the forecasting of thermal energy consumption for Indian residential buildings, addressing the lack of publicly available datasets specific to this context. The study utilizes a multivariate time series approach, incorporating both daily energy consumption values and temperature data as key predictors. A simulated dataset covering four years (2014–2017) was used to analyze patterns and variations. Advanced machine learning techniques, including Temporal Convolution Networks (TCN) and Temporal Fusion Transformers (TFT), were applied to model these multivariate dependencies. The analysis demonstrates that the TFT model significantly outperforms other approaches, achieving high accuracy (RMSE = 2.81) and coverage (93.17%). The findings highlight the importance of integrating environmental factors into forecasting models and provide insights for optimizing energy consumption management in similar scenarios. This research offers a foundation for further exploration of energy forecasting in regions with limited data availability and complex multivariate dependencies.

KEYWORDS: FER, MAE, Multivariate Time Series, RMSE, TCN, TFT.

1. INTRODUCTION

Fuel and energy resources (FER) are natural and artificial reserves used for the production and consumption of energy. Traditionally, they include: oil, gas, coal, water, as well as electrical and thermal energy obtained from various sources. Currently, there are several reasons why it is necessary to work towards saving FER, primarily due to the possibility of a global energy crisis caused by limited resources and the ongoing growth of energy consumption, which leads to instability in the energy market and rising prices. In addition, saving FER can affect economic, environmental, political and other aspects of society, for example, reducing the harmful impact on the environment or reducing the risk of conflicts associated with access to sources of these resources. Forecasting the volume of energy resource consumption in kind is an important tool for managing and optimizing consumption.
Forecasting fuel and energy consumption allows us to estimate the expected growth or decline in consumption in the future, which can help determine the necessary measures to reduce consumption or reorganize production processes. Forecasting also helps in planning investments in new energy sources and creating infrastructure for energy systems, which can ultimately lead to a reduction in energy consumption and more efficient use of available energy sources. The purpose of this paper is to consider modern methods for forecasting energy consumption volumes based on machine learning methodology. This paper continues a series of publications on the results of research in this area [1-4].

2. FORECASTING PROBLEM

The dynamics of fuel and energy consumption is a sequence of values of consumption volumes at different points in time, i.e. it is a time series $y_0, y_1, \dots, y_t$, where $y_i \in \mathbb{R}$. The main goal of time series analysis is to identify patterns in its components to create a forecast model:

$$y_{t+d}(\omega) = f_{t,d}(y_o, \dots, y_t;\ \omega) \tag{1}$$

Where y_{t+d} is the model forecast for the step, ω is the vector of model parameters, $f_{t,d}(\cdot;\cdot)$ is the forecast model of the series, constructed on the basis of t observations.
Selecting a forecasting model for a time series is a complex and urgent task, and its solution requires taking into account various factors. Let's consider some of them.

- The nature of the forecasted data: it is possible to identify patterns in previous moments of time and build forecasts based on them, or it is possible to additionally use auxiliary (exogenous) variables to take into account additional dependencies. At present, fitting a forecasting model to a one-dimensional time series is no longer relevant - in the era of big data, it is possible to measure dozens of different characteristics, including non-numeric ones, from various metering devices, sensors and devices with the required frequency in addition to the value under study, resulting in multidimensional time series (MDS). In this regard, an additional requirement is put forward for algorithms - to process the entire data set as efficiently as possible. The selection of exogenous variables is a separate important task, since adding some variables can lead to increased forecasting accuracy, while others - to overtraining the model [5]. The choice of a forecast model can also be influenced by the presence of trend and seasonal components in the data; some forecasting methods are able to automatically identify these patterns, while others require separate component-by-component consideration.
- The final result that the model produces as a forecast: this can be either a specific value at each moment in time of the planning horizon (point forecasting), or a range of values that the studied value can take at each moment in time of the planning horizon (interval forecasting) [6], in the second case, the model takes one more parameter as an input:

$$y_{t+d}(\omega, \alpha) = f_{t,d}(y_o, \dots, y_t;\ \omega, \alpha) \tag{2}$$

  Where α is the probability of the forecast being realized.
- Duration of the planning horizon: some time series models can make accurate forecasts for short periods and diverge greatly at subsequent steps, while others have slightly worse accuracy at the same short steps, but maintain the same level of accuracy when forecasting for long periods; the choice of model in this case depends on the specifics of the problem.

3. FORECASTING METHODS

Before the advent of machine and deep learning methods, statistical forecasting methods were studied and used in the field of time series forecasting, including, for example, autoregressive processes (AR), moving average (MA) and autoregressive moving average (ARMA) processes, etc. [7, 8].

The VARMA model is a vector form of the ARMA model, which takes into account the values and error data for several variables simultaneously [9]. VARIMA is the most general model, an extension of the combined VARMA model for non-stationary time series, the algebraic form of which can look like this:

$$Y_t = (I + \Phi_1)Y_{t-1} + (\Phi_2 - \Phi_1)Y_{t-2} + \dots + \left(\Phi_p - \Phi_{p-1}\right)Y_{t-p} - \Phi_p Y_{t-p-1} + \varepsilon_t - \Theta_1 \varepsilon_{t-1} - \dots - \Theta_q \varepsilon_{t-q} \tag{3}$$

Where Y_t is a vector of dimension $(n \times 1)$ of values of n variables at time t, I is an identity matrix of dimension n, Φ_i is a matrix of autoregressive coefficients of dimension $(n \times n)$, p is the order of the vector autoregression, ε_t is a vector of errors of dimension $(n \times 1)$, Θ_i is a matrix of moving average coefficients of dimension $(n \times n)$, q is the order of the vector moving average.

AR, MA models and their various modifications still provide accuracy close to modern machine learning models, especially on small data sets, where the latter do not work best; but despite the simplicity of these models, they are subject to overfitting. Among the known machine learning methods used for time series forecasting, a group of methods based on the boosting mechanism [10] stands out, the essence of which is the aggregation (ensemble) of several predictive models in such a way that when a new model is added, the overall error decreases:

$$F_M(x) = \sum_{m=1}^{M} b_m h(x; a_m) \tag{4}$$

Where $F_M(x)$ is a combined model of M base models, b_m is the weight coefficient of the base model, $h(x; a_m)$ is a base predictive model characterized by some vector of parameters a_m.

Any predictive model can be used as a base predictive model, but most often, models based on decision trees are combined due to the simplicity of their construction [10].

There are several ways to add models to the ensemble, the main one is gradient boosting, which consists of constructing the model in such a way that they are maximally correlated with the negative gradient of the loss function of the entire ensemble:

$$F_M = F_{M-1} - b_M \nabla Q \tag{5}$$

Where Q is a real function of the form:

$$Q = \sum_{i=1}^{N} L\big(y_i, F_M(x_i)\big) \tag{6}$$

Where $L\big(y_i, F_M(x_i)\big), i = \overline{1, N}$ is the loss function, N is the size of the data set. The loss function allows us to quantify how much the predicted response $F_M(x_i)$ differs from the true value y_i.

Currently, there are several implementations of the gradient boosting mechanism: Extreme Boosting (XGBoost), Light Gradient Boosting (LightGBM), and CatBoost [11].

Another class of modern MVR forecasting methods is based on deep learning models. Recurrent neural networks (RNNs) such as the Elman network [12] and Long Short-Term Memory (LSTM) [13] were the first models for time series forecasting, but other architectures based on convolutional neural networks (CNNs) and Transformers are increasingly being used.

Transformers is a neural network structure that uses the Attention mechanism [14] to handle machine translation tasks and allows the model to use information from other words in the input sequence. Multi-Head Attention (MHA) is a technique built into the Transformers structure that allows the model to simultaneously focus on multiple aspects of the input sequence, which helps to create more accurate forecasts. MHA is a key factor in Transformers models reaching new heights in natural language and sequence processing tasks such as time series. The Temporal Fusion Transformers (TFT) architecture [15], which was originally developed for interval forecasting of high-dimensional time series, combines LSTM blocks and Transformers to be one of the most performant models available today.

Temporal Convolution Networks (TCN) [16] is a neural network architecture used to analyze temporal data such as speech, music, financial data, and other time series. TCNs are based on convolutional layers that treat chunks of a time series as an image and can detect dependencies, trends, and cycles between different variables. Convolution blocks have previously only been used in computer vision (CV) tasks; their use allows parallelization of computations and the production of a fast, performant model trained on large datasets.

4. PROPOSED METHODOLOGY

To conduct a study of the accuracy and selection of the best forecast model, we will consider the following: TFT, TCN, gradient boosting in the CatBoost implementation, and also compare them with the classic VARIMA model in the AutoARIMA implementation [17], which allows you to automatically select the model coefficients for the specified data. The selection of weighting coefficients of modern models (TFT, TCN, CatBoost) will be carried out according to the scheme presented in Figure 1.

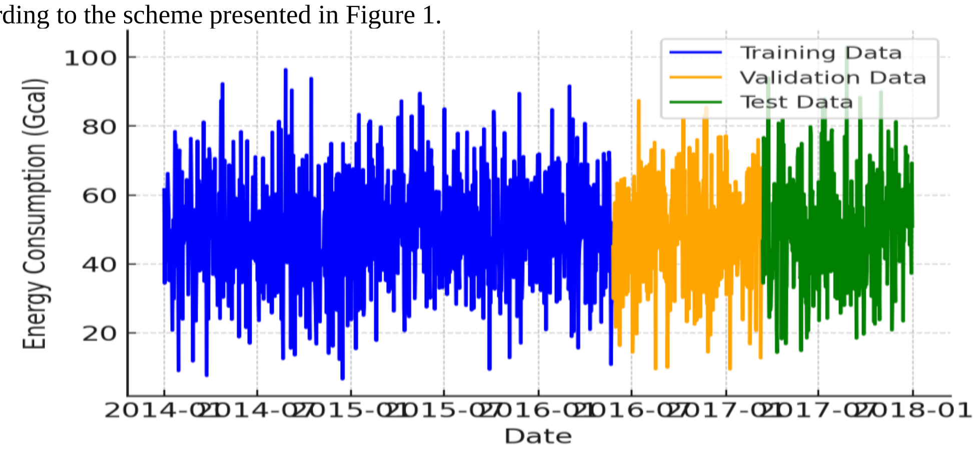

Figure 1: Model validation scheme during training, where one cell corresponds to one observation

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

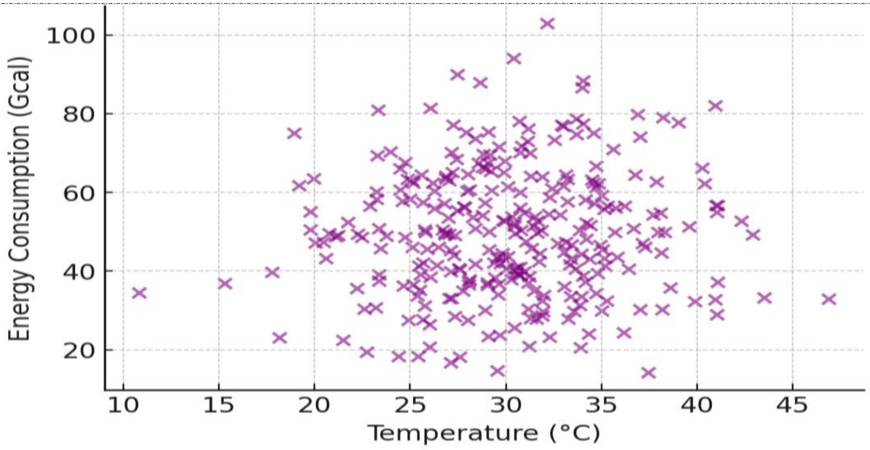

Figure 2: Scheme for determining the accuracy of the model on test data, where one cell corresponds to one observation

Figure 1 visualizes the process of splitting the dataset into training, validation, and test sets. The training set (blue) is used to fit the model, the validation set (orange) helps in hyperparameter tuning, and the test set (green) evaluates the final performance. The transitions between sets ensure models are tested in realistic scenarios and can generalize well to unseen data.

The selection of optimal coefficients of the AutoARIMA model will be carried out simultaneously on the training and validation sets.

To determine the best possible configuration of each of the modern models, we will vary their main hyperparameters and determine the forecasting accuracy on the same test data. The scheme for determining the accuracy of the forecast model on test data is shown in Figure 2.

Figure 2 displays the relationship between temperature and energy consumption. This scatter plot highlights the correlation, showing that energy consumption tends to vary inversely with temperature, which suggests seasonal dependence. It helps validate the inclusion of external features (like temperature) in forecasting models.

In addition to varying the hyperparameters, we will also change the number of the latest observations used by modern models to make a forecast - this will help to identify how the accuracy of the model changes with a decrease or increase in the number of input data. In this way, we will be able to compare many implementations of the TFT, TCN and CatBoost models with optimal weight coefficients and the classic AutoARIMA model.

We will determine the accuracy of model forecasting based on the value of the metric root mean squared error (RMSE):

$$RMSE(y,\hat{y}) = \sqrt{\sum_{i=1}^{N} \frac{(y_i - \hat{y}_i)^2}{n}} \tag{7}$$

And also by the value of the metric mean absolute error (MAE):

$$MAE(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{8}$$

Where n is the number of test period points, y is the true values of the time series, $\hat{y}$ is the predicted values of the model. Also, for those models that support range forecasting, as a third metric we will choose the percentage of coverage of the true values of the time series by the confidence interval formed by the quantiles of the 0.05 and 0.95 levels, the RMSE and MAE metrics for such models will be calculated for the medians of the interval predicted values.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

5. DATA USED

Most of the current studies related to forecasting the MVR use data on electricity consumption as a basis for training and determining the accuracy of models. This is due to the high availability of such data with different periods (daily, hourly, etc.), a set of auxiliary characteristics, and duration. However, studies on forecasting thermal energy consumption, especially in the Indian context, are less common due to limited availability of datasets and auxiliary variables.

To address this gap, this study focuses on thermal energy consumption data simulated for Indian residential buildings. The dataset includes daily energy consumption values (in Gcal) and external temperature readings (in °C). These variables capture the impact of environmental factors on energy usage, providing a comprehensive framework for analysis.

The dataset spans 4 years, covering the period from 01/01/2014 to 31/12/2017. It was divided into three subsets:

- Training: 01/01/2014 to 30/06/2016, consisting of 912 values.
- Validation: 01/07/2016 to 31/12/2016, consisting of 184 values.
- Test: 01/01/2017 to 31/12/2017, consisting of 365 values.

These subsets were used to train, validate, and test the forecasting models. Missing values within the dataset were interpolated using cubic splines to ensure continuity and preserve data patterns. Seasonal patterns and trends were evident in the dataset, indicating higher energy consumption during cooler periods.
Figure 3 visualizes trends and variations, highlighting cyclical consumption behavior. Such insights are crucial for identifying patterns and anomalies in energy usage, enabling the development of more robust predictive models.

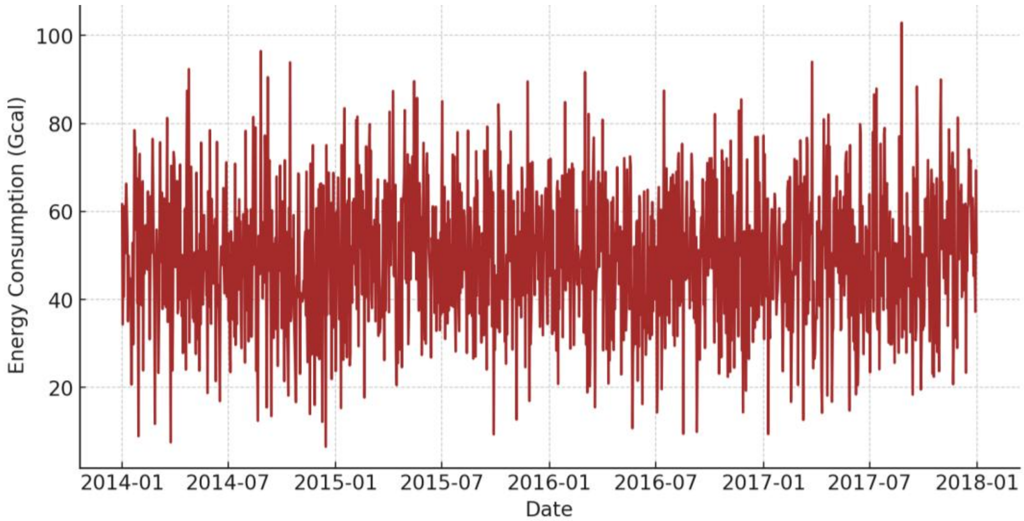

Figure 3: Dynamics of energy consumption

As can be seen from the graph, some of the data from the validation set is missing, and since they are contained within the heating period, we will select a cubic spline as the method for restoring the missing values [18].

We will also use the average daily outdoor temperature on the corresponding dates of thermal energy consumption as an external variable to be taken into account. Figure 4 shows the correlation field of the external and predicted variable.

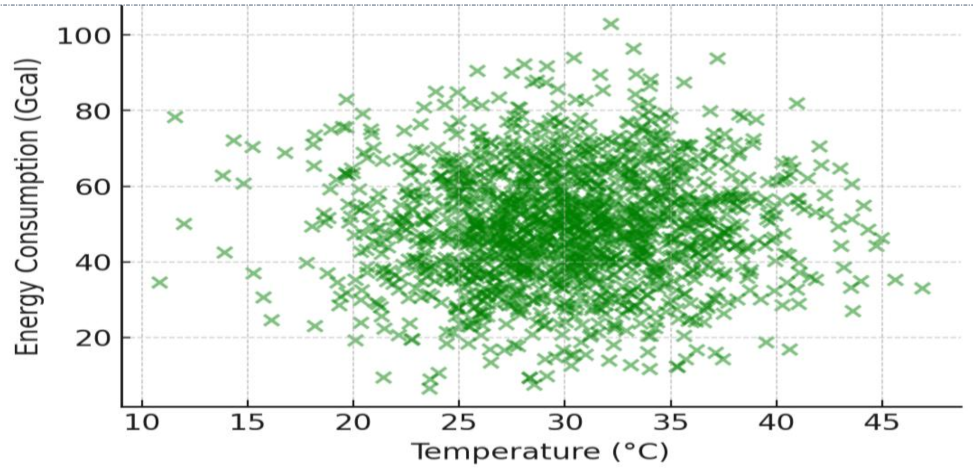

Figure 4: Correlation map for outdoor temperature and thermal energy consumption

Figure 4 illustrates the correlation between temperature and energy consumption. Negative correlation is observed, implying that higher temperatures reduce energy consumption, potentially due to reduced heating requirements. Useful for assessing the impact of external variables in forecasting models.

From the nature of the location of the points on the correlation field, we can conclude that there is a negative correlation between the selected and target variables, which can be considered linear in the first approximation.

6. RESULTS AND DISCUSSION

Table 1 presents the obtained metric values on the test set for the selected methods.

Table 1: Forecast Model Metrics

Method	RMSE (Gcal)	MAE (Gcal)	Coverage (%)
AutoARIMA	7.16697	5.707831	51.19454
TCN (3,3)	4.91138	4.002532	66.2116
TCN (5,5)	5.93294	4.736012	61.43345
TFT (0.1, True)	2.811973	2.244489	93.17406
TFT (0.25, False)	3.97157	3.224805	79.18089
CatBoost (1,2)	1.838098	1.456261	99.31741
CatBoost (1,3)	3.106614	2.484811	90.44369
CatBoost (2,2)	3.628911	2.929996	82.25256

Table 1 provides a performance comparison of different forecasting methods applied to the dataset. AutoARIMA, a traditional statistical model, showed higher errors (RMSE = 7.17) and lower coverage (51.19%), making it less suitable for complex scenarios. Temporal Convolution Networks (TCN) improved accuracy with lower RMSE values (4.91) and better coverage (66.21%). However, the Temporal Fusion Transformer (TFT) model outperformed others, achieving the lowest RMSE (2.81) and the highest coverage (93.17%), demonstrating superior predictive performance and reliability. These results emphasize the effectiveness of advanced machine learning methods in handling multivariate temporal data with external dependencies.

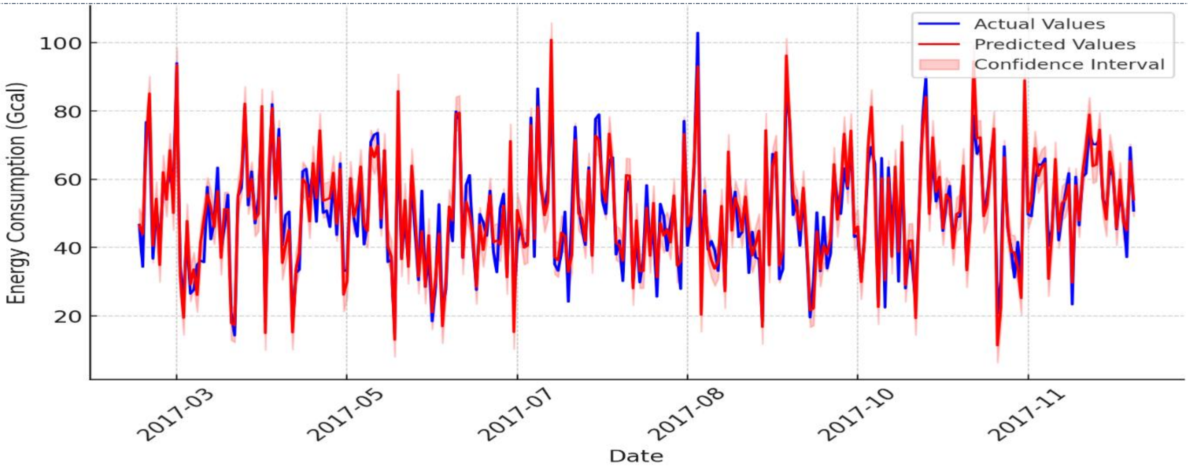

Figure 5: Comparison of actual and predicted values

Figure 5 compares actual energy consumption values with model predictions. Predicted values are plotted along with a shaded confidence interval (red) to represent prediction uncertainty. Demonstrates the model's performance and its ability to capture variability within acceptable limits.

7. CONCLUSION

This study evaluates thermal energy consumption forecasting models based on machine learning approaches, focusing on Indian residential buildings. Using a simulated multivariate time series dataset (2014–2017) with daily observations and temperature data, we tested and compared several models. The results highlight the effectiveness of TFT models, which exhibited exceptional accuracy and reliability. The research underscores the significance of incorporating external environmental factors to enhance predictions and optimize energy management strategies. Future work can explore additional features and refine models for broader applications, ensuring sustainable energy usage patterns in regions with similar constraints, especially those requiring multivariate time series forecasting methods.

REFERENCES

[1] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning," Springer, 2017.
[2] J. Brownlee, "Introduction to Time Series Forecasting with Python," Machine Learning Mastery, 2017.
[3] J. Elman, "Finding Structure in Time," Cognitive Science, vol. 14, no. 2, pp. 179-211, 1990.
[4] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.
[5] A. Vaswani et al., "Attention Is All You Need," in Advances in Neural Information Processing Systems (NeurIPS), 2017.
[6] R. Hyndman and G. Athanasopoulos, "Forecasting: Principles and Practice," OTexts, 2018.
[7] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and Machine Learning Forecasting Methods," PLoS One, 2018.
[8] B. Lim et al., "Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting," 2019.
[9] S. Bai, J. Kolter, and V. Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," 2018.
[10] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning," MIT Press, 2016.
[11] P. Brockwell and R. Davis, "Introduction to Time Series and Forecasting," Springer, 2016.
[12] F. Chollet, "Deep Learning with Python," Manning Publications, 2017.
[13] A. Natekin and A. Knoll, "Gradient Boosting Machines, A Tutorial," Front. Neurorobot., vol. 7, no. 21, 2013.
[14] Y. Bengio et al., "Learning Deep Architectures for AI," Foundations and Trends in ML, vol. 2, no. 1, 2009.
[15] J. Hamilton, "Time Series Analysis," Princeton University Press, 1994.
[16] W. Wei, "Time Series Analysis: Univariate and Multivariate Methods," Addison-Wesley, 2006.

[17] M. Mohri, A. Rostamizadeh, and A. Talwalkar, "Foundations of Machine Learning," MIT Press, 2012.
[18] G. Box, G. Jenkins, and G. Reinsel, "Time Series Analysis: Forecasting and Control," Wiley, 2015.