Electric Power Components and Systems

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uemp20

Smart Short-Term Load Forecasting through Coordination of LSTM-Based Models and Feature Engineering Methods during the COVID-19 Pandemic

Seyed Mohammad Shobeiry, Sasan Azad & Mohammad Taghi Ameli

Electric Power Components and Systems, 51(2): 171–187, 2023

ISSN: 1532-5008 print / 1532-5016 online
DOI: 10.1080/15325008.2023.2168092

Smart Short-Term Load Forecasting through Coordination of LSTM-Based Models and Feature Engineering Methods during the COVID-19 Pandemic

Seyed Mohammad Shobeiry, Sasan Azad, and Mohammad Taghi Ameli
Department of Power & Control, Faculty of Electrical Engineering, Shahid Beheshti University, Tehran, Iran

CONTENTS

Abstract—Short-term load forecasting is essential for power companies because it is necessary to ensure sufficient capacity. This article proposes a smart load forecasting scheme to forecast the short-term load for an actual sample network in the presence of uncertainties such as weather and the COVID-19 epidemic. The studied electric load data with hourly resolution from the beginning of 2020 to the first seven days of 2021 for the New York Independent Operator is the basis for the modeling. The new components used in this article include the coordination of stacked long short-term memory-based models and feature engineering methods. Also, more accurate and realistic modeling of the problem has been implemented according to the existing conditions through COVID-19 epidemic data. The influential variables for short-term load forecasting through various feature engineering methods have contributed to the problem. The achievements of this research include increasing the accuracy and speed of short-term electric load forecasting, reducing the probability of overfitting during model training, and providing an analytical comparison between different feature engineering methods. Through an analytical comparison between different feature engineering methods, the findings of this article show an increase in the accuracy and speed of short-term load forecasting. The results indicate that combining the stacked long short-term memory model and feature engineering methods based on extra-trees and principal component analysis performs well. The RMSE index for day-ahead load forecasting in the best engineering method for the proposed stacked long short-term memory model is 0.1071.

Keywords: short-term load forecasting, machine learning, deep learning, uncertainty, feature engineering, long short-term memory, COVID-19
Received 31 May 2022; accepted 7 January 2023
Address correspondence to Mohammad Taghi Ameli, Faculty of Electrical Engineering, Shahid Beheshti University, Tehran, Iran. E-mail: m_ameli@sbu.ac

1. INTRODUCTION

1.1. Motivation

Short-Term Load Forecasting (STLF) is essential for the power company's operation and grid operators because it is necessary to ensure adequate capacity and proper power generation arrangement; this will affect operating efficiency and short-term decisions. Meanwhile, the COVID-19 epidemic as a nonlinear factor will be effective in STLF, and based on previous solutions; load forecasting (LF) may not be accurate. A nonlinear and complex relationship between the factors affecting the LF problem explains the need to use intelligent methods such as machine learning (ML). According to the epidemic of Corona disease, there are support measures related to the electricity industry based on global experience [1]. Studies show that the measures taken in other parts of the world for the electricity industry during the Corona crisis can be divided into the following three categories: (1) monitoring the conditions of network operation, (2) concern about the health of human resources in the electricity industry [2], and (3) payment of electricity price; Therefore, in line with the first category of these measures, i.e., monitoring the conditions of network operation, LF in the future and its short-term trend is essential for power companies and network operators. Effective operation in power companies becomes necessary with the advancement of technology. The amount of electricity generation and consumption should be equal because there is no advanced system for storing or sudden generation of electrical energy. An imbalance between electricity generation and demand leads to economic, technical, and social losses. Therefore, power companies need to maintain a balance between generation and demand. LF is one of the necessary techniques to improve the balance between electricity generation and demand. But at the heart of this issue, many nonlinear factors, such as weather, temperature, and seasonal change, must be considered. At the same time, we are now facing another influential factor, the COVID-19 pandemic.

1.2. Background

LF is an essential process in smart grid operations. LF is an approach to estimating the future electric load for a given landscape based on current system information [3]. LF is classified into short-term, mid-term, and long-term forecasting according to periods. Studying the load trend in daily operation plans such as energy transfer planning and demand-side management in STLF is considered [4]. The influence of external factors such as time, weather conditions, and random elements on the load and dynamic uncertainties in the smart grid are the main obstacles to the accuracy and speed of LF. Improving LF performance, including accuracy, speed, and stability, has become a turning point in LF research. From a modeling point of view, the proper selection of input parameters, models, and algorithms significantly impacts LF performance. A Deep Neural Network (DNN) is a set of intelligent computational algorithms that provide a comprehensive solution for modeling a complex nonlinear relationship between input and output [5]. According to the literature review, using recurrent neural networks (RNN), especially long short-term memory (LSTM), to solve the STLF problem will have advantages like maintaining long-term time dependencies. Conversely, this method has been suggested in many kinds of research, so the focus in reviewing the literature is on studies that have used methods based on LSTM to get enough knowledge about the different approaches to using this method.

In Ref. [6], the load, temperature, cooling load, and gas consumption data have been used as input features in the last five days. Convolutional Neural Network (CNN) has extracted valuable features. The LSTM then forecasts the load for the next hour along with the BiLSTM layers. The results show that the proposed method has better forecasting performance than CNN-BiLSTM, CNN-LSTM, BiLSTM, LSTM, Back-propagation Neural Network (BPNN), and Random Forest Regression (RFR) and Support Vector Regression (SVR). The Ref. [7] presents an integrated network structure consisting of self-organized mapping (SOM), irregular time series, intelligent optimization algorithm, and LSTM. The purpose of this structure is to extend the LF interval, reduce artificial debugging and improve the accuracy of STLF. In Ref. [8], input features are divided into historic and forecast data. Historical data is entered into the LSTM layer to model the relationships between previously observed data (feature extraction). Fully connected layer outputs with forecast data (weather information for the forecast day) are combined with LSTM layer outputs. In Ref. [9], the combined CNN-LSTM model, including CNN layers, extracts features from input data, and LSTM layers are used to learn sequences. The proposed model has higher accuracy compared to the LSTM-based model. The clustering analysis based on energy consumption behavior and representative model training can also improve forecasting accuracy. Reference [10] presents different algorithms and the proposed parallel LSTM-CNN network or PLCNet for STLF. The results show that DNN models, especially PLCNet, are good candidates for STLF. In Ref. [11], the combined approach of CNN and LSTM network is proposed. The effectiveness of the proposed method is confirmed by comparing the forecasting errors with existing techniques such as the LSTM network, radial performance network, and severe gradient amplification algorithm.

Reference [12] presents the LSTM neural network (LSTM-NN) model for STLF. A series of rolling time indices, including an index of the daytime, a holiday flag, and a day of the week index, are embedded as class feature vectors to increase forecast accuracy. The proposed model in Ref. [13] is the modified support vector regression (SVR) and LSTM model called SVR-LSTM. Factors affecting microgrid loads, such as household and commercial entities, have been selected as input variables. Identifying the behavioral patterns of input variables and modeling their behavior over short periods is one of the significant capabilities of the SVR-LSTM model. The results show that the SVR-LSTM model with the highest correlation coefficient has better results than SVR and LSTM. In Ref. [14], an STLF framework using stacked multilayer short-term memory is presented. Forward computation and reverse computation are designed to solve the one-sidedness of the memory process during the training process. Also, a Deep Learning (DL) process is introduced to extract the correlation between historical loads to improve the algorithm's convergence. Comparing the proposed method's performance results on actual data in different ways shows that the proposed method can extract dynamic features from the data. In Ref. [15], deep bidirectional long short-term memory-based sequence to sequence (Bi-LSTM S2S) is introduced to forecast the "day-ahead" peak demand. Also, the proposed deep Bi-LSTM network learned from the future and past values of the input parameters for forecasting. For performance comparison, Bi-LSTM S2S, Shallow LSTM S2S, Deep LSTM S2S, Levenberg–Marquardt Back-propagation Artificial Neural Networks (LMBP-ANN), and Mean Gaussian Support Vector Regression (MG-SVR) forecasting models are also developed and tested. The models were trained, validated, and tested using daily peak demand data. In this research, outliers and missing values in the historical load data have also been investigated because the electricity demand reading has typing and human error due to the unavailability of smart meters. The results indicate that the proposed model performs better than all other models on public holidays and regular days.

1.3. Research Gaps and Contributions

For a more appropriate summary of recent research and the contribution of this article, the objective, model, and input data of the reviewed papers and this article are presented in Table 1.

Recent research uses a method combination approach to achieve higher accuracy [6–13]. The fixed part of this combination is mainly the LSTM method. Other available methods have been combined to implement the side steps of the LF, such as extraction or selection of valuable features. Therefore, from examining the subject literature, it can be seen that the LSTM method should be done to solve the LF problem. In addition to this primary method, other auxiliary tools can be used to improve the accuracy and speed of LF. This article uses the optimal combination of the forecasting elements, namely Feature Engineering (FE) and modeling of the Stacked LSTM (S-LSTM) network. The contribution of this article to the literature can be summarized in a few cases:

1. Using the S-LSTM-based model to preserve relatively long-time dependencies to increase the accuracy of DL to solve the LF problem.
2. More realistic modeling of conditions for LF, considering the COVID-19 pandemic data and its impact on the short-term electric load curve, compared to Refs. [14, 15] that only used electric load and weather data.
3. The use of simple algorithms based on LSTM along with FE and, as a result, the high speed of solving the LF problem compared to the complex CNN-LSTM algorithms used in Refs. [6, 9–11].
4. Using different FE methods to make it possible to compare their performance and coordinate them with the forecasting model to improve model training. In contrast, most recent research has only emphasized enhancing and adding more complexity to the forecasting model [6, 10, 11, 14, 15].

The rest of the article structure is organized: The second section introduces the LSTM model. In the third section, the modeling of proposed STLF schemes is presented. The fourth section includes simulation and analysis results, and at the end, the fifth section presents conclusions and suggestions for future research.

2. LSTM ARCHITECTURE

A particular RNN network known as an LSTM can store past data in its memory unit. It is very effective for time series data forecasting [16]. An LSTM network consists of four essential components; Cell, input gate, output gate, and forget gate. Information is transmitted by the cell at random intervals. The gates track the flow of input and output data from the cell. Gradient disappearance and explosion problems obtained from RNN can be reduced using the LSTM network [17, 18]. The basic configuration of an LSTM network is shown in Figure 1. The node outputs of an LSTM network are calculated as follows:

$$I_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_I) \tag{1}$$

$$F_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_F\right) \tag{2}$$

Reference	Publication year	Objective	Model	Problem input data
[6]	2021	STLF in an integrated energy system with different energy generation sources	Combination of CNN, LSTM, and Bi-LSTM	Electric load data, temperature data, cooling load data, and gas consumption data
[7]	2020	Extend LF range, reduce artificial debugging and improve STLF accuracy	LSTM, in combination with self-organized mapping, irregular time series, and intelligent optimization algorithm	Hourly electric load data and weather data
[8]	2020	Increase the accuracy of STLF by synchronizing the type of DNN according to the nature of the data	DNN with LSTM layer	Hourly electric load data and hourly temperature data
[9]	2020	household STLF	CNN-LSTM hybrid model	Hourly electric load data
[10]	2021	Comparison of different models for STLF	CNN-LSTM hybrid model	Hourly electric load data
[11]	2021	Increase the accuracy of STLF and compare the proposed model with the base models	CNN-LSTM hybrid model	Electric load data
[12]	2020	Increase the accuracy of STLF and develop forecast horizons	LSTM-NN hybrid model	Hourly electric load data and weather data (temperature and humidity)
[13]	2020	STLF in microgrids by identifying patterns and modeling the behavior of input variables in the short run	LSTM-SVR hybrid model	Different types of households data and commercial entities' data
[14]	2021	Extract dynamic features from the data	Multilayer stacked bidirectional LSTM	Power load profile with a sampling time of 15 min
[15]	2021	Forecast the "day-ahead" peak demand, investigating outliers and missing values, Proper forecasting of the model on holidays	Bidirectional LSTM-based sequence-to-sequence (Bi-LSTM S2S) regression	Daily peak electricity demand data for 30 months and weather data (max temperature and max dew point temperature)
This article	2022	Increasing the accuracy and speed of STLF, comparing feature engineering methods, and considering the COVID-19 pandemic variables	LSTM-based models and compatibility with feature engineering tools	Hourly electric load data, weather data, social distance data, COVID-19 epidemic data

TABLE 1. Summary of objective, model, and input data of the reviewed papers and this article.

$$\tilde{C}_t = \tan h(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{3}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{4}$$

$$O_t = \sigma(W_O \cdot [h_{t-1}, x_t] + b_O) \tag{5}$$

$$h_t = O_t * \tan h(c_t) \tag{6}$$

The input variable at time t is displayed with x_t. Also W_I, W_F, W_c, W_O are weight matrices. I_t refers to the input gates, F_t, O_t, and C_t represent the forget gate, the output gate, and the cell output, respectively. The sigmoid activation function is defined by σ, and h_t is assigned to compute the hidden state outputs at time t in the vector structure. b_I, b_F, b_c, and b_O refer to the bias values of different gates.

3. PROPOSED SCHEME

As mentioned in previous chapters, improving the accuracy and speed of LF has become a significant turning point in recent research. From a modeling point of view, the factors that affect the performance of LF include the appropriate selection of input parameters, models, and algorithms. Therefore, it is necessary to improve the performance of LF, especially in the field of LF models and parameters. Thus, this article presents a new LF approach using FE and a stacked LF model based on the LSTM algorithm. The proposed LSTM-based model is implemented to further improve the accuracy and speed of the conventional LSTM model by increasing the parameters and the level of architectural complexity and can maintain longer time dependencies of the data on the LF problem.

In addition to load and weather data, COVID-19 epidemic data have improved DNN training and LF accuracy. COVID-19 epidemic data accounts for the uncertainty caused by this event in the LF problem, bringing the problem modeling closer to reality. Also, to improve the LF performance, different Feature Selection (FS) and Feature Extraction (FEx) methods have been used.

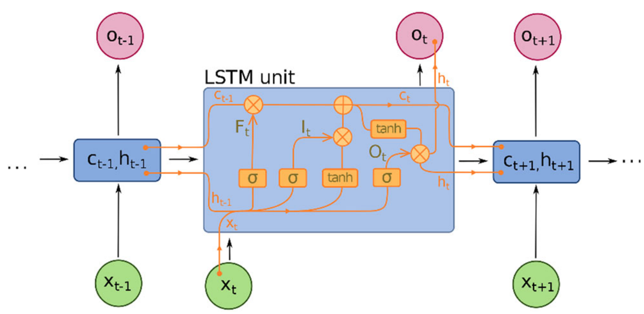

FIGURE 1. LSTM structure [19].

By applying FS and FEx methods to the input data of the LSTM-based model, the accuracy of the LSTM-based forecasting model increases due to the selection and extraction of appropriate and correlated features. Also, the amount of input data is reduced, and naturally, the speed of training the model increases. Another side purpose of this article to achieve the main objectives, namely to increase the accuracy and speed of LF, is to compare the efficiency of different FE methods.

In this study, the New York Independent Operator's electric load data, meteorological data, electric load sequence data, COVID-19 epidemic data, and social distance data for 2020 are used to model training. Then, for testing and evaluation, the model predicts a weekday load in the first week of 2021. The data analysis of this article will be discussed in detail in Subsection 3.3. The LF problem will be modeled in the following, and the FE methods and DL algorithms will be discussed. Figure 2 illustrates the STLF process presented in this article.

3.1. FE Methods

FE methods are often divided into FS and FEx, each with its characteristics. FE methods deal with the high dimensions of a problem with many features. FS methods for

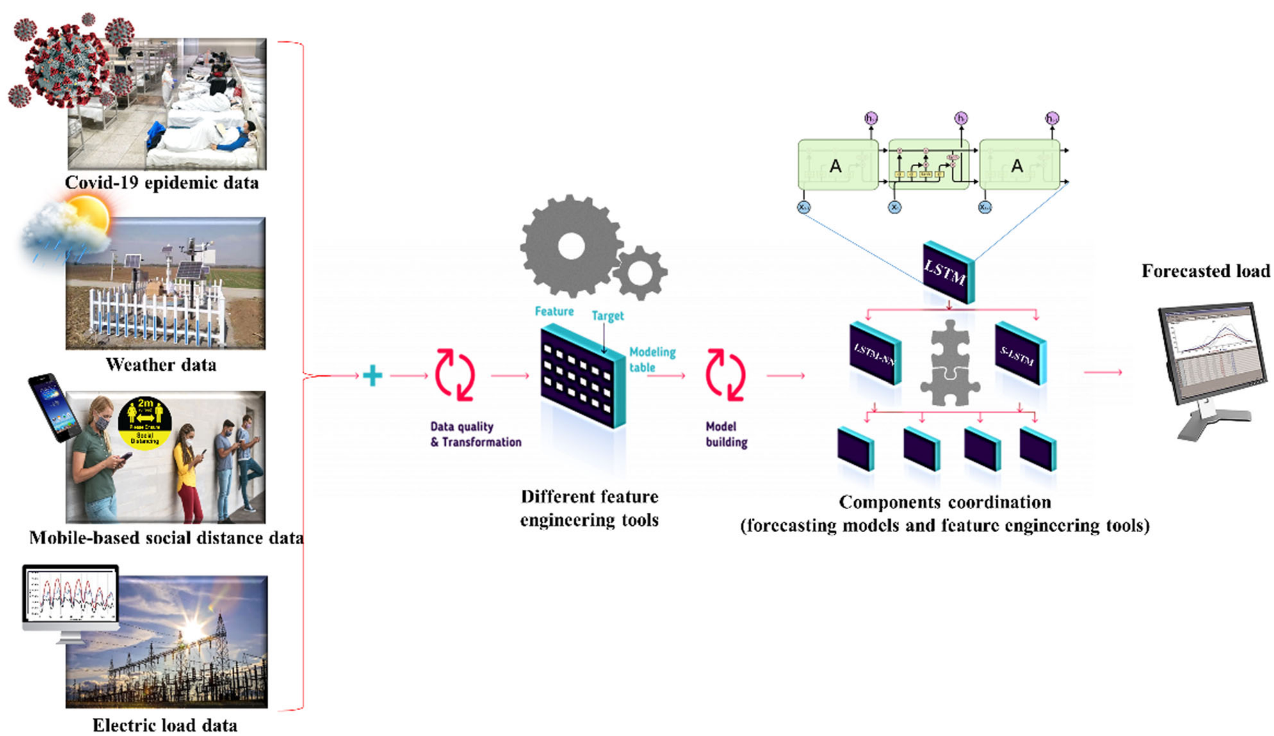

FIGURE 2. The STLF process presented in this article during the COVID-19 epidemic.

dealing with high-dimensional data have become integral to learning. Selecting a suitable feature can improve the inductive learner, including learning speed, generalization capacity, and model simplicity. The term "Big Dimensionality" refers to this problem and is constructed compared to "Big Data," which deals with a large sample size.

On the one hand, FEx methods reduce dimensions by combining key features. Hence, they can build a set of new features that are usually more compact and distinctive. These methods are preferred in applications such as image analysis, image processing, and information retrieval because, in these cases, the model's accuracy is more important than its interpretability. FS is beneficial for applications where the main features are essential for model interpretation and knowledge extraction because the main features of the data set are preserved during this process. FS can be defined as identifying related features and removing unrelated and duplicate features to view a subset of features that describe the problem well. On the other hand, FS reduces dimensions by eliminating irrelevant and identical features.

We are applying FS and FEx methods to the input data of the LSTM-based model introduces appropriate and sufficient features to the problem. Also, the speed of model training increases due to the reduced volume of input data and model calculations. Conversely, due to the selection and extraction of accurate and correlated features, the accuracy of the LSTM-based forecasting model increases.

3.1.1. Pearson Correlation Coefficient. The correlation coefficient is a statistical tool to determine the type and degree of relationship between a quantitative variable and another quantitative variable. The correlation coefficient is one of the criteria used to determine the correlation between two variables. The correlation coefficient indicates the intensity of the relationship and the type of relationship (direct or inverse) [20]. This coefficient is between 1 and -1, and if there is no relationship between the two variables, it equals zero. The correlation between two random variables, X and Y, is defined as follows:

$$\mathrm{corr}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \tag{7}$$

E is the expected value operator, cov means covariance, corr is the usual symbol for Pearson correlation, and σ is the standard deviation symbol. Pearson correlation coefficient is a parametric method used for data with normal distribution or large amounts of data. Pearson correlation coefficient varies between -1 and 1.

3.1.2. Extra-Trees. Random decision forests are a combined learning method for classification and regression, which work on the training time and output of classes (category) based on a structure consisting of many decision trees. Extra trees are similar to typical random forests [21–23], but there are two significant differences:

1. Each tree is taught using the whole learning example (instead of the bootstrap example),
2. Top-down division learns random trees. An arbitrary cutoff point is selected instead of calculating the desired local cutoff point for each feature under study. This value is chosen from the uniform distribution in the feature range (in the tree training set). Then, the slot with the highest score is selected to split the node from all the random divisions created.

3.1.3. Chi-Square. This FS method measures the significant difference between the features' observed and expected frequency. The null hypothesis in this method states no correlation between these features. This method is also known as the chi-square test of independence. [24] The calculation of the score in this method is presented in Eq. (8).

$$X^2 = \frac{(\text{Observed Frequency} - \text{Expected Frequency})^2}{\text{Expected Frequency}} \tag{8}$$

3.1.4. Multiple Linear Regression. One of the standard methods in multivariate analysis is the multiple linear regression (MLR) technique. A linear relationship is established between the output variable and one or more descriptive variables based on regression analysis. Sometimes, the output variable is the dependent variable, and the explanatory variables are also called independent variables. In MLR, the linear model parameters are estimated using an objective function and the value of variables. The considered model is a linear relationship based on the model parameters in linear regression. [25]. Thus, if there are n observations of the independent variable X with p dimension and the goal is to establish a linear relationship with the output variable y, the following linear regression model can be used:

$$y_i = \beta_0 1 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i, \ i = 1, \ldots, \ n \tag{9}$$

3.1.5. Principal Component Analysis. The principal component analysis (PCA) is a simple and efficient linear transformation technique. The primary purpose of PCA analysis is to identify patterns in the data; PCA aims to identify the correlations between variables. Reducing the dimensions will be significant if there is a strong correlation between the variables. PCA generally finds the direction of maximum variance in high-dimensional data and designs it in a smaller sub-space to store complete information [26].

3.1.6. Singular Value Decomposition. Singular value decomposition (SVD) is a method that decomposes one matrix into three other matrices. For example, matrix A can be written as follows:

$$A = U \cdot S \cdot V^T \tag{10}$$

The matrix U and V are symmetric, and S is a diagonal matrix. The main matrix (matrix A) multiples U, S, and V transposed. This process is called "Factorization," which means that matrix A can consist of three matrices, U, S, and V. The values in S can help in the feature reduction operation. Using matrix analysis with SVD makes it possible to determine which matrix columns have more information [27].

3.2. DL Models Based on LSTM

According to the literature study, one of the most appropriate methods for STLF is LSTM. LSTM is a type of RNN with many benefits, such as maintaining long-term time dependencies in time series. Electric load is also a time series that has relatively long time dependencies. Therefore, two developed DNN models based on the LSTM network are proposed and implemented in this article.

The proposed LSTM-based models, while comparing different LSTM structures, have been implemented to further improve the accuracy and speed of the typical LSTM model by increasing the parameters and the level of architectural complexity. In the following, these models are introduced.

3.2.1. A Hybrid Model of LSTM and Fully Connected Decreased NN. A hybrid DNN consisting of an LSTM layer to maintain time dependencies and fully connected layers to establish model accuracy is implemented. In the first layer of this network, the LSTM architecture is used to keep the time dependencies of the data. A fully connected architecture with decreased neurons in each layer toward the output is used in the following layers to increase the model's accuracy. In the first part (first layer), the goal is to establish appropriate time dependencies of the data. The second part (second layer onwards) aims to verify accuracy based on these dependencies. Some effective parameters in promoting DNN training include the number of layers, the number of neurons in each layer, and the activation function. These parameters were adjusted to the best values throughout the study to achieve the minimum forecasting error in the test data. The architecture of the second layer to the network's output is also presented in the best values to establish forecasting accuracy.

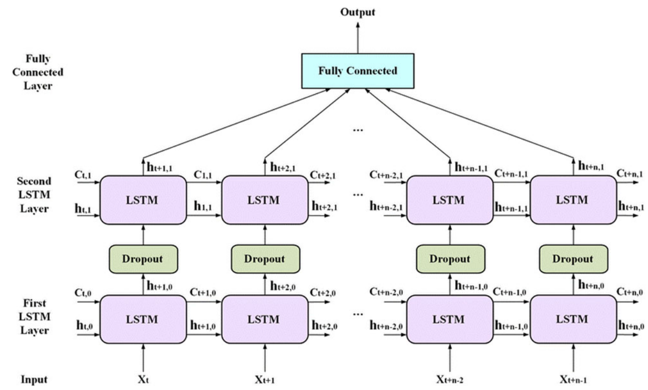

FIGURE 3. The S-LSTM general structure [29].

3.2.2. S-LSTM Networks. In this article, a 3-layer RNN based on LSTM is used. This model is known as the stacked LSTM (S-LSTM) network. The original LSTM model consists of a hidden LSTM layer, followed by a standard feed-forward output layer. S-LSTM is a development for this model with several hidden LSTM layers. LSTM layering deepens the model and more accurately captures information. Because LSTMs are based on sequence data, adding layers increases the levels of abstraction of input observations over time.

Stacked LSTMs or deep LSTMs were introduced by Graves *et al*. [28]. These LSTMs are used in speech recognition and benchmark extraction of a challenging problem. RNNs are inherently deep because their hidden state functions all previous hidden states. S-LSTMs are now a sustainable way to challenge sequencing forecasting problems. S-LSTM architecture can be defined as an LSTM model consisting of multiple LSTM layers. The previous LSTM layer provides an output sequence instead of an output value in the next LSTM layer. The structure of this model is presented in Figure 3. Dropout blocks have been used between the LSTM layers.

3.3. Data Analysis

Hourly electric load data were collected from the beginning of 2020 to the first seven days of 2021 for the New York

Independent Operator and were used as the basis for modeling. In addition to electrical load data, weather data, COVID-19 epidemic data, and social distance data as factors influencing short-term load behavior were also collected during the period. The data were extracted and prepared using an interdomain database to investigate the COVID-19 epidemic's impact on US electricity markets [30–32]. The data used is processed and so-called clean. There are no blank, unregistered records in the data used.

Therefore, the focus is on preparing the data structure. Weather data include temperature, wind speed, relative humidity, and dew point temperature with hourly resolution. COVID-19 epidemic data have accumulated confirmed case numbers, newly confirmed cases, infection ratio calculated, accumulated confirmed deth numbers recorded, newly established death numbers, and fatal rates calculated. Social distance data include the percentage of devices that stay at home 24 hr out of all devices, the median proportion of home dwell time in one day, the percentage of devices that go to workplaces for 3 to 6 hr out of all devices, the percentage of devices that go to workplaces for more than 6 hr out of all devices, the count of devices that stay at home 24 hr, the total count of devices, count of devices that go to workplaces for 3 to 6 hr, and count of devices that go to workplaces for more than 6 hr. The term "device" refers to mobile devices representing social distance behavior modeling well. Using pandemic and social distance data in the problem leads to a more realistic modeling of the conditions for STLF. In other words, the intervention data in the issue under the current epidemic conditions has been used to improve the model training and achieve the main goals, such as increasing accuracy.

Electric load data have time dependencies in different sequences. Various sequences were created as input features for appropriate model training. Depending on factors such as the volume of input data and the interval studied, different sequences of electrical load can be created. According to the hourly records and the LF researchers' recommendation [33, 34], the following sequences have been used in this article:

1. The previous day's load at the same time,
2. The previous day of the week's load at the same time,
3. The previous day's load at 1 hr ago,
4. The previous day's load at 2 hr ago.

All data used is prepared with the day-ahead LF arrangement. Therefore, four weather features related to the previous day, six features of the COVID-19 epidemic associated with the previous day, eight social distance features related to the last day, and electrical load features with the mentioned sequence will be used for day-ahead LF.

3.4. Case Study Scenarios

As previously discussed in recent research, the LSTM model has been highly recommended for STLF due to advantages such as maintaining time dependencies in data, so in this article, three designs using the LSTM algorithm were developed, which include the following:

1. Proposed stacked LSTM (S-LSTM),
2. Combined LSTM with fully connected decreased NN (LSTM-NN), and
3. Common LSTM (LSTM)

The first design has stacked LSTM blocks, which makes the performance of maintaining relatively longer time dependencies more prominent. The second design has an initial LSTM layer to keep time dependencies and a fully connected decreased NN to stabilize forecast performance. The third scheme also has an LSTM layer that only preserves time dependencies on the data. The reason for selecting these three designs is to compare the different behaviors of the models based on the LSTM algorithm. Through this comparison, the most compatible models with FE methods are identified to improve the training of the STLF model. Therefore, the mentioned designs have been compared in different FE cases. Simulations have been performed for three scenarios or, in other words, to evaluate three different models. Each scenario has seven FE modes with various methods introduced in the previous section, and one scenario is without FE. Day-ahead LF for each scenario is performed; the speed and accuracy of the model are recorded for each mode. STLF curves are obtained using LSTM, LSTM-NN, proposed S-LSTM algorithms, and various FE tools, and the results are compared. The purpose of defining the above cases is to review and compare the results for all three models when using different FE methods to identify the most consistent issues to achieve this article's main objectives, namely to increase the accuracy and speed of LF. The various stimulation modes for each scenario, along with the abbreviation used, are:

1. Complete entry of features without engineering and reduction of dimensions to the model (Full Feature)
2. Using Pearson correlation to select compelling features (Cor)
3. Using MLR to select practical features (Reg)
4. Using the chi-square score method to select compelling features (K2)

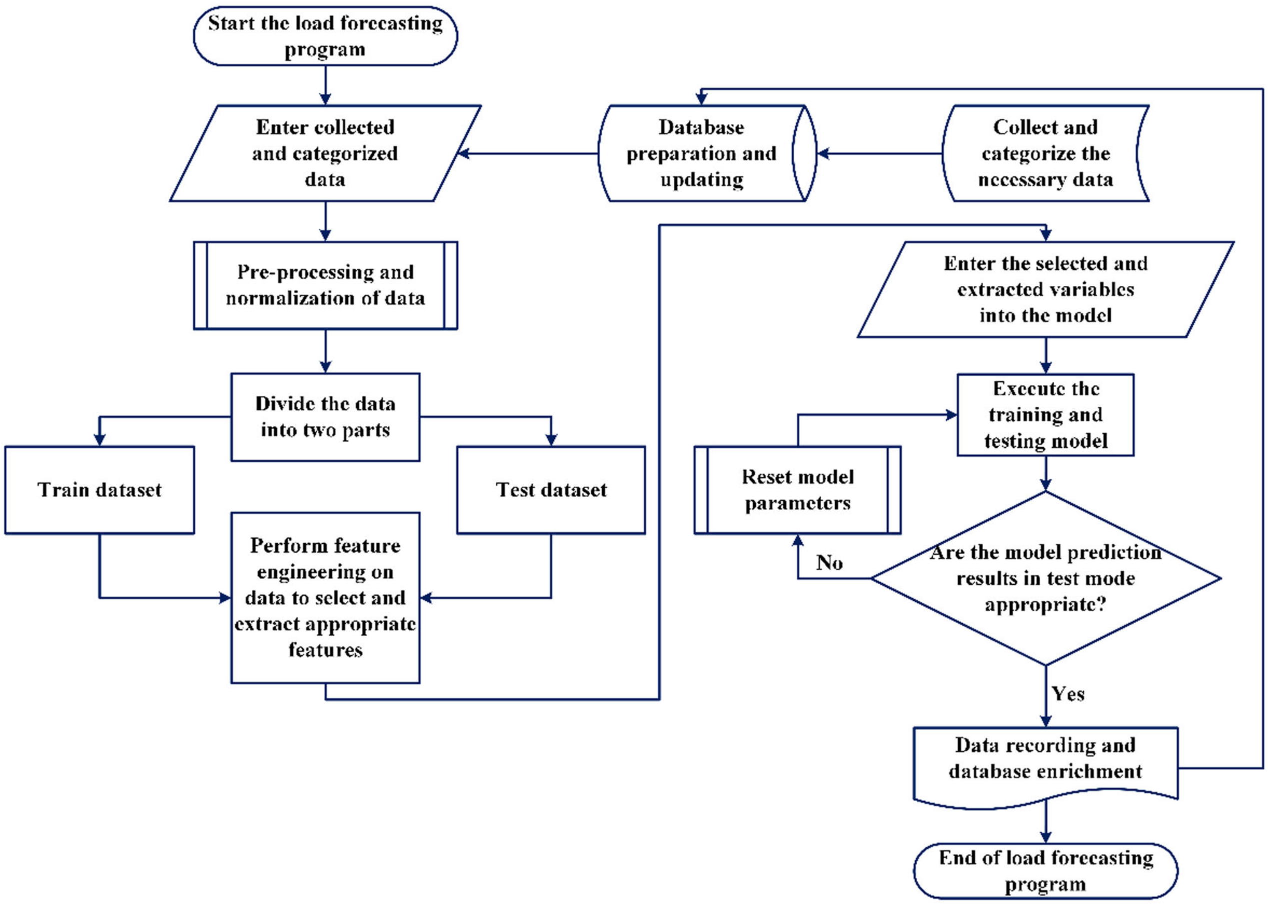

FIGURE 4. Study flowchart.

5. Using the extra-trees method to select practical features (ET)
6. Using Pearson correlation combination to select effective features and SVD to extract less critical features (SVD)
7. Using Pearson correlation combination to select useful features and PCA to extract less essential features (CorPCA)
8. Use pure PCA to implement dimensional reduction on all features (PCA and PCA2)

3.5. Study Algorithm

Section 3 discusses the proposed STLF model and its benefits for increasing LF's speed and accuracy. Also, the modeling of DNN used in the LF problem was introduced. LSTM-based methods have been introduced in new LF research due to their superiority in maintaining relatively long-term dependencies. This article develops two DL models based on LSTM: a stacked LSTM network and a collaborative LSTM fully connected decreased NN. Also, this article uses the usual LSTM network to validate the proposed methods. In addition to presenting these models, various FE tools are used. Finally, the proposed models' performance and the FE ancillary tools of the features are compared according to the predetermined main goals, i.e., increasing the accuracy and speed of LF. The proposed algorithm flowchart for STLF is presented in Figure 4.

4. RESULTS AND DISCUSSION

This section studies the proposed FE design and simulation of DNNs. As mentioned earlier, the primary purpose of this study is to improve the training of DNNs and optimize their performance to increase the accuracy and speed of STLF. To achieve the main objectives, dimensional reduction, including FEx and FS, is done simultaneously with the optimal modeling of the LSTM network, using two stacked and combined algorithms.

It can be said that using the FE and selecting the most practical features to enter the model reduces its computational complexity and increases the accuracy and speed of the LF. This project's effectiveness and the use of three

LSTM-based DL models to forecast the load of an actual sample network have been demonstrated. The simulation is implemented in the Jupyter Notebook software environment based on the Python programming language.

4.1. FE Results

4.1.1. Pearson Correlation Coefficient. The initial statistical analysis uses Pearson correlation between each category's features and the desired output. According to the concepts of correlation, a subset of features is a suitable subset with the following characteristics: have a high correlation with the target feature and have a low correlation with each other. Therefore, according to Table 2, variables B, D, and E are selected in the category of features related to the COVID-19 epidemic. Also, variable H is chosen in the category of variables related to social distance, and in weather data, variables P and R are chosen. Finally, variable S is selected from the four variables associated with the electric load sequence.

4.1.2. FS using MLR. Two modes were implemented by the MLR method to select practical features. The first mode involves MLR on each feature category with output, and the second mode involves applying MLR on all features and output. The results of FS analysis include the combination of the effects of these two modes. Features with a p-value below 5% at a 95% confidence interval were selected. According to the analysis, the variables F, G, H, I, J, O, P, and R are finally selected.

4.1.3. FS using the Chi-Square Method. As mentioned, this FS method measures the significant difference between the features' observed and expected frequency. In this method, the null hypothesis corresponds to the lack of correlation between these features. The results of assigning points to features through this method are presented in Figure 5. Thus, the features S, U, V, T, R, P, F, and D are selected in this method.

4.1.4. FS using the Extra-Trees Method. Different results are obtained each time this method is performed due to its random nature. Therefore, this method was implemented ten times, the most common features in all process implementations, including the features O, P, Q, R, S, T, U, V, and, less importantly, the features I, J, M, and N.

4.1.5. FS using Correlation and FEx using SVD. Dimension reduction or feature extraction increases the accuracy and speed of the model. The feature must be included in the FEx process that strongly correlates with each other. FEx must be applied to each feature category separately to maintain the effect of each type on the electric load. Therefore, features B and E are selected, and dimension reduction is performed on features D and F. Also, features G, H, and K are strongly correlated, and dimension reduction is achieved. The feature P is selected, the dimension reduction is performed on the features O and R, and finally, the dimension reduction is performed on the features S, T, U, and V.

4.1.6. FS using Correlation and FEx using PCA. The process of FS and candidate features for FEx is the same as the previous method, and only in this section the FEx method is PCA. A pure PCA execution mode is also considered, meaning that only one FEx is applied to all features. Figure 6 shows the PCA and variance values. According to this figure, selecting two dominant and seven important features is suitable for training this model. Therefore, a noncorrelated PCA dimension reduction for two and seven features is also implemented.

Finally, as a summary of the FE (FS and FEx) section, the proposed features of each method are presented in Table 3. Features that have been suggested twice or more are highlighted in green.

4.2. STLF based on DL

After the FE process significantly improves the DL model training and thus increases its accuracy and speed, it is time to solve the LF problem. Three models based on the LSTM algorithm have been used, including the proposed stacked LSTM (S-LSTM), combined LSTM with a fully connected decreased NN (LSTM-NN), and common LSTM (LSTM). According to studies, the training parameters of the models presented in Table 4 have been adjusted to the best possible values to achieve high accuracy. Also, an optimization and training process is used in coordination for all three proposed models, including using a "rmsprop" optimizer and batch size equal to 512 with 200 epochs. Each modeling scenario's different modes of using the FE tools are compared. In the following, the results of each model are reviewed and analyzed. To evaluate the models' accuracy, the Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) criteria have been used [35], which are presented in (11) and (12), respectively. L_{Actual} is the actual load, $L_{\mathrm{Forecasted}}$ is the forecasted load, and N is the number of time steps.

Degree of correlation with output	Desired feature	Letter code	Degree of correlation with output	Desired feature	Letter code
0.4072	Dew point temperature	O	0.0679	Accumulated confirmed case number recorded	A
–0.1238	Relative humidity	P	–0.3529	Newly confirmed case number	B
0.0033	Wind speed	Q	0.0679	Infection ratio calculated	C
0.5266	Temperature	R	0.1797	Accumulated confirmed deth number recorded	D
0.9244	The previous day's load at the same time	S	–0.2004	Newly confirmed death number	E
0.8681	The previous day of the week's load at the same time	T	0.1558	Fatal rate calculated	F
0.9097	The previous day's load at 1 hr ago	U	–0.2001	Percentage of devices that stay at home 24 hr out of all devices	G
0.8731	The previous day's load at 2 hr ago	V	–0.3202	The median proportion of home dwell time in one day	H
0.0996	Total count of devices	L	0.0899	Percentage of devices that go to workplaces for 3 to 6 hr out of all devices	I
0.0957	Count of devices that go to workplaces for 3 to 6 hr	M	0.0516	Percentage of devices that go to workplaces for more than 6 hr out of all devices	J
0.0613	Count of devices that go to workplaces for more than 6 hr	N	–0.1593	Count of devices that stay at home 24 hr	K

TABLE 2. Correlation results in input data.
Features of the same group are presented with similar colors. Green color means high correlation.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} \left(L_{\mathrm{Forecasted}} - L_{\mathrm{Actual}}\right)^2}{N}} \tag{11}$$

$$\mathrm{MAPE} = \sum_{i=1}^{N} \left| \frac{L_{\mathrm{Actual}} - L_{\mathrm{Forecasted}}}{L_{\mathrm{Actual}}} \right| \times \frac{100\%}{N} \tag{12}$$

4.2.1. LSTM Model. A network with a typical LSTM layer was implemented as one of the models discussed in this article to validate and compare the proposed method with a standard model. As previously explained, this network has no hidden or stacked layers, so it is not expected to perform well in maintaining long-term time dependencies. Figure 7 shows the actual and forecasted load curves for the day-ahead forecast period.

As shown in Figure 7, the benefits of using FE are well-defined. Except for the chi-square method, most dimensional reduction methods make low-limit forecasting than the actual load. The results of multiple regression methods, PCA with two features, PCA with seven features, and correlation, are very close. Generally, the most significant forecasting error occurs in sudden changes in the curve behavior, such as peak points and the curve floor. It can be said that most methods in LF for the last hours of the day have performed better than the initial and middle hours due to the relatively more consistent consumption behavior at night than during the day. Also, during the day's mid-hours from 10:00 AM to 5:00 PM, when consumption is almost constant, the results of most methods are close to each other.

4.2.2. LSTM-NN Model. This model combines the LSTM network and a fully connected decreased neural network.

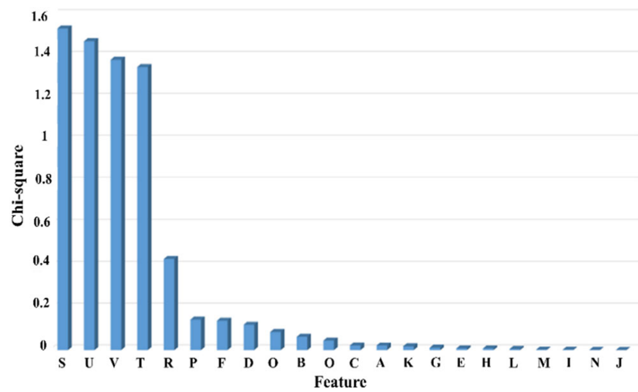

FIGURE 5. Feature score by implementing the chi-square method.

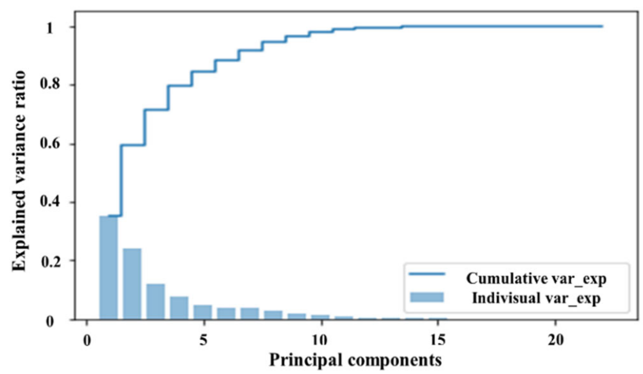

FIGURE 6. PCA and variance values.

The time dependencies are preserved in the first layer, and the model accuracy is established in the following layers. Figure 8 shows the actual and forecasted load curves for the day-ahead forecast period. As shown in Figure 8, the benefits of using FE are clear. The correlation, multiple regression, PCA with seven features, and SVD are close. Also, the LSTM-NN model performs better than the LSTM model in forecasting peak points. In this model, sudden upward behavioral changes in the curve are well forecasted. Also, this model has performed better in the day's middle hours than in the final hours.

4.2.3. Proposed S-LSTM Model. In this model, time dependencies are strongly preserved due to stacking the LSTM blocks. Due to the high accumulation of layers and the greater probability of over-fitting, 20% of neuron dropout is used. Dropout of neurons improves the model training process and speeds up training. Finally, the training details of this model have been adjusted as in the previous cases so that a proper comparison can be made between this proposed model and the last two models. Figure 9 shows the actual and forecasted load curves for the day-ahead forecast period. As shown in Figure 9, the proposed model's advantages are clear. So that without performing FE on the input data, a good accuracy for LF has occurred. The results of multiple regression methods, correlation, correlation and PCA, correlation and SVD, PCA with seven features, and PCA with two features are very close. In this model, the results of most FE methods converged, while in the previous models, the results were scattered. As shown in Figure 9, using the S-LSTM model and at the same time, if no FE methods are used, no particular problem occurs in the LF process, and the accuracy is acceptable. In this case, the model focuses on the properties of time dependencies rather than on quantitative structure in data.

4.2.4. Comparing Results. According to the presented results, the proposed S-LSTM model brings the output of different FE methods closer together. According to the previous explanations is due to the presence of S-LSTM layers and maintaining relatively long time dependencies in the data. Figure 10 shows the training times for each model and FE method. According to Figure 10, the proposed S-LSTM model has a relatively higher training time than the other two models. However, using PCA-based FEx methods have significantly reduced training time. All models' best FE performance for increasing the day-ahead LF speed is related to the combined correlation with the PCA method and the pure PCA with seven features method. Another important conclusion that can be inferred from the figure is about the correct and coordinated choice of the FE method with the forecasting model. In the case of the nonengineered features, the training time is less than in some FE cases, such as extra trees.

Figure 11 compares the LF accuracy of the three models in different modes of FE. This figure shows that the proposed S-LSTM model has more minor errors than the other two conventional models. In the proposed model, in the case that all input data enter the model without engineering, contrary to the results of different models, relatively good accuracy occurs. In other words, if the proposed S-LSTM model is used, FE can be ignored, and reasonably good accuracy can be achieved. The best FE performance to increase the accuracy of the day-ahead LF in all models is related to the extra-tree method and then to the combined correlation with the PCA method and pure PCA with seven features. Only the FS with the chi-square case has a relatively large error in the proposed model. The chi-square method inherently maintains the time dependencies of the data due to the periodicity. In this case, repeated use of the

Selected features												FS method
						R	**P**	**H**	**E**	**D**	**B**	Correlation
				R	**P**	**O**	**J**	**I**	**H**	**G**	**F**	MLR
				V	**U**	**T**	**S**	**R**	**P**	**F**	**D**	Chi-square
J	**I**	**N**	**M**	**V**	**U**	**T**	**S**	**R**	**Q**	**P**	**O**	Extra-tree

TABLE 3. Proposed features in each FS method.
Darker green means more feature suggestions.

S-LSTM	LSTM-NN	LSTM	
50, 50 fully LSTM blocks, respectively, with a dropout of 20% of neurons	50 and 25 fully connected neurons, respectively, with Relu activation	has not	Network hidden layers
From 1.22 min to 3.5 min	From 43 sec to 1.5 min		Network training time
	100 LSTM blocks		Network first layer
	rmsprop		Optimization algorithm
	Mean squared error		Loss function
	512		Batch size
	200		Epoch
	intel core i5-7200 u 3.1 GHz		Computing information (CPU based)

TABLE 4. LSTM-based DNN information.

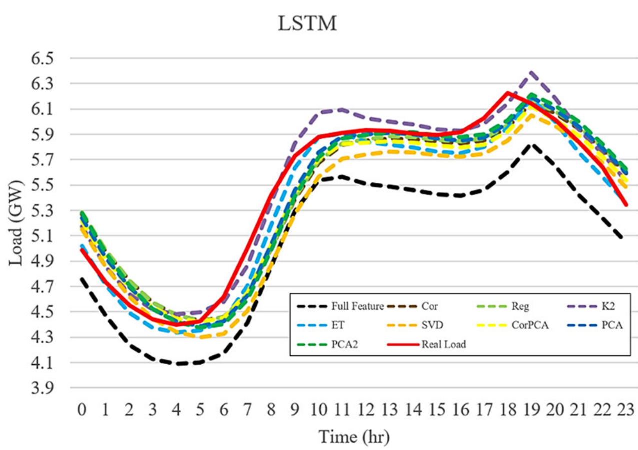

FIGURE 7. The actual and forecasted day–ahead load curves for the LSTM model.

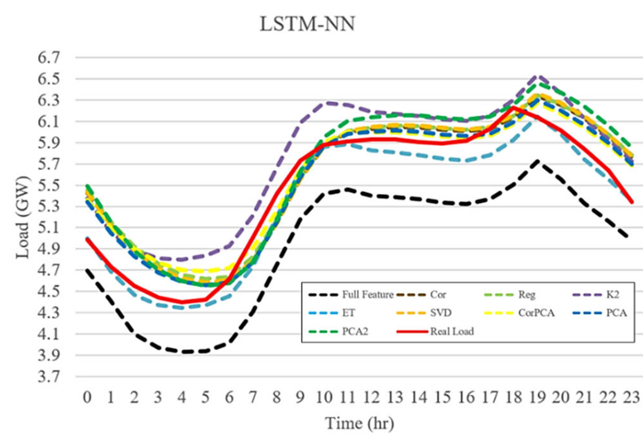

FIGURE 8. The actual and forecasted day–ahead load curves for the LSTM-NN model.

advantage of preserving time dependencies has resulted in poor performance of the proposed model. Also, the lowest LF error in the proposed model is related to the FE model through the extra-tree algorithm. More exclusive features are selected in the extra-trees method than in other ways.

By summarizing all the items that promote DNN training, finally, to perform STLF with reasonable accuracy and speed, the S-LSTM model and the pure PCA with seven features or a combination of correlation and PCA as the FE method are proposed.

4.2.5. Validation of Results. The first distinguishing feature of this study from previous close research [8–12] is the investigation of the COVID-19 epidemic uncertainty variables in the LF problem. Considering these variables leads to more accurate and realistic modeling and ultimately increases the accuracy of the LF model. Before the

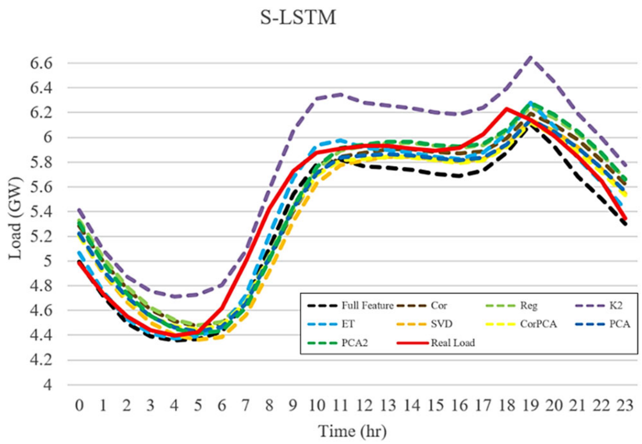

FIGURE 9. The actual and forecasted day–ahead load curves for the S-LSTM model.

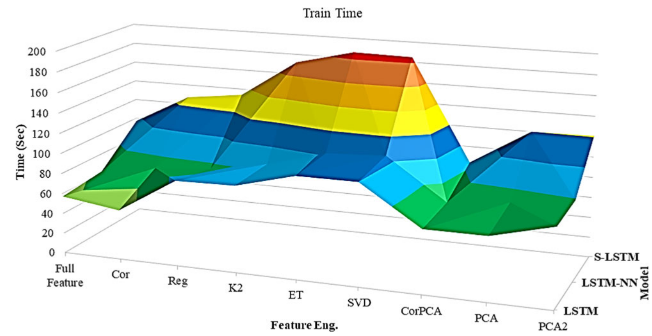

FIGURE 10. Training time for each model and FE methods.

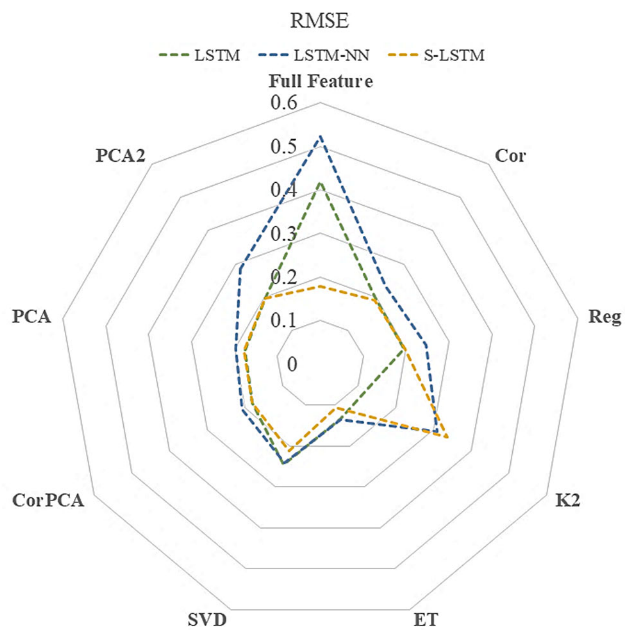

FIGURE 11. Comparison of LF accuracy of three models in different modes of FE.

forecasting process, the most practical features are selected from the features of the problem. Therefore, if the epidemic is over, this plan will work again according to the same process. The second distinguishing feature is the implementation of several FE methods to extract the most appropriate and coordinated approach with the forecasting model. Finally, an S-LSTM model is used for LF that previous research has not used. Although these differences and data and the study period make it difficult to accurately and adequately compare this study with previous research, it is necessary to refer to the reference results closer to this research for validation.

In Ref. [10], different algorithms and the new model of combined DL, LSTM networks, and the CNN model to perform STLF are provided. In Ref. [10], the LSTM-CNN model performed well on Malaysian data. But the problem with this hybrid model is its relatively high execution time. The LSTM model also performed well on German load data. But this model also has a rather long run time of 431 sec. The S-LSTM model in this research, along with FE methods compatible with the proposed model, has a shorter execution time than the LSTM and LSTM-CNN hybrid models presented in [10], as a 2.5-fold to sixfold increase in speed is observed in the model used in this study. Also, the accuracy of the proposed model is more appropriate than the LSTM and LSTM-CNN hybrid models presented in Ref. [10] on German data. The training process of models with Malaysian data is complete due to the higher number of data records, so this case is more accurate than the proposed model.

To better present the strengths of this study compared to other methods of DL, Figure 12 provides a comparison between the results of this study and the Ref. [10] for German load data. As shown in Figure 12, only the LSTM-CNN hybrid model can compete with the model proposed in this study, provided that large data volumes are used for training. The proposed model's two modes, the extra-tree modes and the correlation with PCA modes, are the best in terms of accuracy and speed for forecasting the day-ahead load, respectively. As shown in Figure 12, the proposed S-LSTM model with correlation and PCA as FE has better performance in terms of accuracy and speed than the combined LSTM-CNN model. As shown in Figure 12, the DNN model performed very well in speed but has the worst accuracy performance among other DL models. In Ref. [10], the LSTM-CNN hybrid model is faster than the LSTM model due to using CNN to extract features before forecasting with the LSTM model.

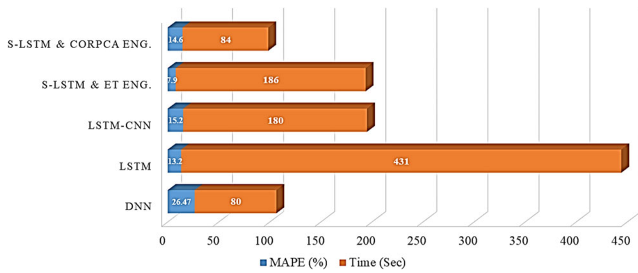

FIGURE 12. Comparison between the results of this article and the Ref. [10] for German load data.

5. CONCLUSION

The behavior of electric load during the COVID-19 pandemic is affected by countermeasures taken by governments. In this situation, the epidemic of COVID-19 is considered a nonlinear factor in the behavior of electric load. The previous solutions for STLF do not bring good results. This article presents a method for improving DNN training to forecast the short-term load of an actual sample network in the presence of uncertainties such as weather conditions and the COVID-19 pandemic. For this purpose, various FE methods and the S-LSTM network were used to increase forecasting performance and maintain time dependencies in data. Therefore, coordination between the forecasting model and the FE method was established to forecast the short-term load of an actual sample network with acceptable accuracy. For this purpose, the accumulation of LSTM network blocks was used to preserve as many temporal dependencies as possible in the input data, as well as different methods of dimensionality reduction (selecting and extracting features). In other words, the separate construction of existing components in an electric LF problem was used to achieve optimal coordination and efficiency between the forecasting model and engineering methods. The process of training and optimization of the proposed model and the usually compared models were carried out similarly by the adaptive optimizer algorithm.

The day-ahead LF curve for the S-LSTM model was compared with the day-ahead LF curve for the LSTM monolayer model and the LSTM-NN model. Although the training presented in this article is 84 sec away from achieving a real-time implementation, the training speed of the S-LSTM model used in this article is approximately fivefold and twofold higher than that of the LSTM and LSTM-CNN models, respectively. While in the LF process, accuracy is a higher priority than speed. In terms of the probability of not overfitting and the closeness of the results of different modes, the proposed S-LSTM model has performed well compared to the other two typical models. Also, a forecast error index, including RMSE and MAPE, was used to show the proposed scheme's effectiveness compared to other methods. The RMSE index for day-ahead LF in the best engineering methods for the LSTM, LSTM-NN scheme, and the proposed S-LSTM scheme is equal to 0.1215, 0.1354, and 0.1071, respectively, which indicates an increase in the accuracy of STLF in the proposed scheme case. According to the results, the proposed S-LSTM model had a relatively higher training time than the other two models. But the training time was significantly reduced using PCA-based feature extraction methods. Also, choosing the correct and coordinated FE method with the forecasting model is essential. The proposed S-LSTM model also centralized various FE methods due to S-LSTM layers and the maintenance of relatively long-term time dependencies in the data rather than focusing on their quantitative state.

The best accuracy obtained for forecasting the next day's load is related to the proposed S-LSTM model and the feature selection method of extra trees. As mentioned, more features are selected in the extra-trees process than in other ways. In the proposed model, relatively good accuracy occurs when all the input data are entered into the model without engineering, in contrast to the results of other models. In addition, if all features are included in this model, unlike the other two models, overfitting does not occur. The best FE performance to increase the accuracy of the day-ahead LF in all models is related to the FS method with extra trees and then related to the combined FS with correlation and FEx with PCA and only FEx with PCA. The best FE performance to increase the speed of day-ahead LF in all models was related to the combined FS methods with correlation with FEx with PCA and only FEx with PCA. In the case of using the Chi2 score feature selection method, a relatively large error has occurred in the proposed model because the repeated use of the advantage of maintaining time dependencies in the model and FE has caused the poor performance of the proposed model. The following can also be suggested for future research:

Use a new variable or input index to enhance model training at peak load points and consider electricity price variables, demand-side management concepts, and types of energy generation sources for more realistic modeling of conditions.

REFERENCES

[1] R. Madurai Elavarasan *et al.*, "COVID-19: Impact analysis and recommendations for power sector operation," *Appl. Energy*, vol. 279, pp. 115739, 2020. DOI: 10.1016/j.apenergy.2020.115739.

[2] K. Otoufat-Shamsi, S. Ghane, R. Ghaemmaghami, H. Hedayati, and A. Abdalisousan, "Effect of Corona virus on electricity consumption in Iran and the world," *J. Intell. Procedures Elect. Technol.*, vol. 12, no. 45, pp. 49–62, 2021.
[3] M. H. Asgharinejad Keisami, S. Azad, R. Mohammadi Chabanloo, M. Nazari-Heris and S. Asadi, "Machine learning and power system planning: Opportunities and challenges," in *Application of Machine Learning and Deep Learning Methods to Power System Problems. Power Systems*, M. Nazari-Heris, S. Asadi, B. Mohammadi-Ivatloo, M. Abdar, H. Jebelli, and M. Sadat-Mohammadi, Eds. Cham: Springer. DOI: 10.1007/978-3-030-77696-1_3.
[4] A. Y. Alani and I. O. Osunmakinde, "Short-term multiple forecasting of electric energy loads for sustainable demand planning in smart grids for smart homes," *Sustainability*, vol. 9, no. 11, pp. 1972, 2017. DOI: 10.3390/su9111972.
[5] W. Xuewei, C. Ruofa, Y. Hongchao, and L. Caiyan, "Support vector machine short term load forecasting based on similar days," *Comput. Measur. Contrl.*, vol. 26, no. 01, pp. 47–50, 2018.
[6] K. Wu *et al.*, "An attention-based CNN-LSTM-BiLSTM model for short-term electric load forecasting in integrated energy system," *Int. Trans. Electr. Energ. Syst.*, vol. 31, no. 1, pp. e12637, 2021. DOI: 10.1002/2050-7038.12637.
[7] W. Zhang *et al.*, "Short-term power load forecasting using integrated methods based on long short-term memory," *Sci. China Technol. Sci.*, vol. 63, no. 4, pp. 614–624, 2020. DOI: 10.1007/s11431-019-9547-4.
[8] B. S. Kwon, R. J. Park, and K. B. Song, "Short-term load forecasting based on deep neural networks using LSTM layer," *J. Electr. Eng. Technol.*, vol. 15, no. 4, pp. 1501–1509, 2020. DOI: 10.1007/s42835-020-00424-7.
[9] M. Alhussein, K. Aurangzeb, and S. I. Haider, "Hybrid CNN-LSTM model for short-term individual household load forecasting," *IEEE Access*, vol. 8, pp. 180544–180557, 2020. DOI: 10.1109/ACCESS.2020.3028281.
[10] B. Farsi, M. Amayri, N. Bouguila, and U. Eicker, "On short-term load forecasting using machine learning techniques and a novel parallel deep LSTM-CNN approach," *IEEE Access*, vol. 9, pp. 31191–31212, 2021. DOI: 10.1109/ACCESS.2021.3060290.
[11] S. H. Rafi, S. R. Deeba, E. Hossain, and Nahid-Al-Masood, "A short-term load forecasting method using integrated CNN and LSTM network," *IEEE Access*, vol. 9, pp. 32436–32448, 2021. DOI: 10.1109/ACCESS.2021.3060654.
[12] M. S. Hossain and H. Mahmood, "Short-term load forecasting using an LSTM neural network," 2020 IEEE Power and Energy Conference at Illinois (PECI), Champaign, IL, USA, 2020, pp. 1–6. DOI: 10.1109/PECI48348.2020.9064654.
[13] A. Moradzadeh, S. Zakeri, M. Shoaran, B. Mohammadi-Ivatloo, and F. Mohammadi, "Short-term load forecasting of microgrid via hybrid support vector regression and long short-term memory algorithms," *Sustainability*, vol. 12, no. 17, pp. 7076, 2020. DOI: 10.3390/su12177076.
[14] C. Cai, Y. Tao, T. Zhu, and Z. Deng, "Short-term load forecasting based on deep learning bidirectional LSTM neural network," *Appl. Sci.*, vol. 11, no. 17, pp. 8129, 2021. DOI: 10.3390/app11178129.
[15] N. Mughees, S. A. Mohsin, A. Mughees, and A. Mughees, "Deep sequence to sequence Bi-LSTM neural networks for day-ahead peak load forecasting," *Exp. Syst. Appl.*, vol. 175, pp. 114844, 2021. DOI: 10.1016/j.eswa.2021.114844.
[16] Y. Wang *et al.*, "Probabilistic individual load forecasting using pinball loss guided LSTM," *Appl. Energy*, vol. 235, pp. 10–20, 2019. DOI: 10.1016/j.apenergy.2018.10.078.
[17] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, 1994. DOI: 10.1109/72.279181.
[18] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, 2000. DOI: 10.1162/089976600300015015.
[19] A. Karpathy and L. Fei-Fei, 2015. "Deep visual-semantic alignments for generating image descriptions," Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3128–3137.
[20] J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Noise Reduction in Speech Processing*, vol. 2. Springer Topics in Signal Processing, Berlin Heidelberg: Springer, 2009, pp. 115–120.
[21] S. M. Piryonesi and T. E. El-Diraby, "Data analytics in asset management: Cost-effective prediction of the pavement condition index," *J. Infrastruct. Syst.*, vol. 26, no. 1, pp. 04019036, 2020. DOI: 10.1061/(ASCE)IS.1943-555X.0000512.
[22] S. M. Piryonesi and T. E. El-Diraby, "Role of data analytics in infrastructure asset management: Overcoming data size and quality problems," *J. Transp. Eng. B: Pavements*, vol. 146, no. 2, pp. 04020022, 2020.
[23] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*, vol. 1, no. 10. New York: Springer Series in Statistics, 2001.
[24] M. L. McHugh, "The chi-square test of independence," *Biochem. Med. (Zagreb)*, vol. 23, no. 2, pp. 143–149, 2013. DOI: 10.11613/bm.2013.018.
[25] D. J. Olive, "Multiple linear regression," in *Linear Regression*. Cham: Springer, 2017, pp. 17–83.
[26] H. Abdi and L. J. Williams, "Principal component analysis," *WIREs Comput. Stat.*, vol. 2, no. 4, pp. 433–459, 2010. DOI: 10.1002/wics.101.
[27] M. E. Wall, A. Rechtsteiner, and L. M. Rocha, "Singular value decomposition and principal component analysis," in *A practical Approach to Microarray Data Analysis.* Boston, MA: Springer, 2003, pp. 91–109.
[28] A. Graves, A. Mohamed, and G. E. Hinton. "Speech recognition with deep recurrent neural networks," Presented at the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 6645–6649.
[29] K. Zhang, Y. Jiang, D. Liu, and H. Song, "Spatio-Temporal Data Mining for Aviation Delay Prediction," Presented at the 2020 IEEE 39th International Performance Computing and Communications Conference (IPCCC), Austin, TX, USA, 2020, pp. 1–7. DOI: 10.1109/IPCCC50635.2020.9391561.
[30] G. Ruan *et al.*, "A cross-domain approach to analyzing the short-run impact of COVID-19 on the US electricity sector," *Joule*, vol. 4, no. 11, pp. 2322–2337, 2020. DOI: 10.1016/j.joule.2020.08.017.

[31] G. Ruan, J. Wu, H. Zhong, Q. Xia, and L. Xie, "Quantitative assessment of US bulk power systems and market operations during the COVID-19 pandemic," *Appl. Energy*, vol. 286, pp. 116354, 2021. DOI: 10.1016/j.apenergy.2020.116354.

[32] H. Zhong, Z. Tan, Y. He, L. Xie, and C. Kang, "Implications of COVID-19 for the electricity industry: A comprehensive review," *CSEE J. Power Energy Syst.*, vol. 6, no. 3, pp. 489–495, 2020.

[33] P. P. Phyo and C. Jeenanunta, "Daily load forecasting based on a combination of classification and regression tree and deep belief network," *IEEE Access*, vol. 9, pp. 152226–152242, 2021. DOI: 10.1109/ACCESS.2021.3127211.

[34] P. P. Phyo, C. Jeenanunta, and K. Hashimoto, "Electricity load forecasting in Thailand using deep learning models," *Int. J. Elect. Electron. Eng. Telecommun.*, vol. 8, no. 4, pp. 221–225, 2019.

[35] R. Jiao, T. Zhang, Y. Jiang, and H. He, "Short-term non-residential load forecasting based on multiple sequences LSTM recurrent neural network," *IEEE Access*, vol. 6, pp. 59438–59448, 2018. DOI: 10.1109/ACCESS.2018.2873712.

BIOGRAPHIES

Seyed Mohammad Shobeiry was born in 1997 in Tehran, Iran. He received the B. Sc. degree in electrical engineering from Semnan University, Semnan, Iran, in 2019. He received the M. Sc. degree in electrical engineering from Shahid Beheshti University, Tehran, Iran, in 2021. At present, he is an electrical engineering Ph. D. student at Shahid Beheshti University, Tehran, Iran. He is also a researcher at Electricity Network Research Institute, Shahid Beheshti University, Tehran, Iran. His research interests include power systems operation and dispatching, power systems restructuring, load forecasting and state estimation.

Sasan Azad is a Ph.D. student in the Faculty of Electrical Engineering and a researcher at the Electrical Networks Institute of the Shahid Beheshti University. He obtained his BSc degree from the Razi University of Kermanshah and his MSc degree from the Shahid Beheshti University. His main areas of interest are the security and voltage stability of power systems, smart grids, and electric vehicles.

Mohammad Taghi Ameli graduated with a PhD in Electrical Engineering from the Technical University of Berlin, Germany, in 1997. He is currently a Professor of Department of Electrical Engineering and Head of Electrical Networks Institute at Shahid Beheshti University. He specializes in power systems operation and control, micro grid and smart grid.