Expert Systems With Applications 235 (2024) 121207

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A new framework for electricity price forecasting via multi-head self-attention and CNN-based techniques in the competitive electricity market

Alireza Pourdaryaei [a], Mohammad Mohammadi [b], Hamza Mubarak [c], Abdallah Abdellatif [d], Mazaher Karimi [e,*], Elena Gryazina [f], Vladimir Terzija [f]

[a] Substations and Relays Repairs and Maintenance Affairs, Operation and Protection Office, Hormozgan Regional Electric Company, Bandar Abbas 7916795599, Iran
[b] Department of Power and Control, School of Electrical and Computer Engineering, Shiraz University, Shiraz 7194684334, Iran
[c] School of Engineering and Built Environment, Griffith University, Southport, QLD 4222, Australia
[d] Department of Electrical Engineering, Faculty of Engineering, Universiti Malaya, Kuala Lumpur 50603, Malaysia
[e] School of Technology and Innovations, University of Vaasa, Wolffintie 34, 65200 Vaasa, Finland
[f] Center for Energy Science and Technology (CEST), Skolkovo Institute of Science and Technology (Skoltech), Bolshoy Boulevard 30, bld. 1, Moscow 121205, Russia

A R T I C L E I N F O

Keywords:
Convolutional neural networks
Electricity price forecasting
Feature selection
Electricity market
Multi-head attention
1D-CNN

A B S T R A C T

Due to recent technical improvements, the smart grid has become a feasible platform for electricity market participants to successfully regulate their bidding process based on demand-side management (DSM) perspectives. At this level, practical design, implementation, and assessment of numerous demand response mechanisms and robust short-term price forecasting development in day-ahead transactions are all critical. The accuracy and effectiveness of the day-ahead price forecasting process are crucial concerns in a deregulated market. In this market, the reason for low accuracy is the limitation of electricity generation compared to the electricity demand variations. Hence, this study proposes a suitable technique for forecasting electricity prices using a multi-head self-attention and Convolutional Neural networks (CNN) based approach. Further, this study develops a feature selection technique using mutual information (MI) and neural networks (NN) to choose suitable input variable subsets significantly affecting electricity price predictions simultaneously. The combination of MI and NN reduces the number of input features used in the model, thereby decreasing the computational complexity of the NN. The actual data sets from the Ontario electricity market in 2020 are acquired to verify the simulation results. Finally, the simulation results proved the efficiency of the proposed method by demonstrating increased accuracy by attaining the lowest average value for MAPE and RMSE with a value of 1.75% and 0.0085, respectively, and compared to results obtained by recent computational intelligence approaches. By attaining accurate electricity price results, the significance of this study can be summed up as aiding the electricity industry's operators in administering effective energy management, efficient resource allocation, and informed decision-making.

1. Introduction

Modern power grids are held to stringent quality standards to ensure a steady and reliable supply that can keep up with increasingly diverse consumer needs. These complex issues are what motivate the ongoing improvement and refinement of smart grid technologies. Complex obstacles, such as optimizing the capacity of distributed generators (DGs), transmission and distribution (T&D) systems, and effective energy storage technologies, stand in the way of the full realization of smart grid technologies (Mohseni, Brent, Kelly, & Browne, 2022). These difficulties necessitate extensive research and significant funding. Consequently, the operation of the smart grid and today's sophisticated electricity market both heavily rely on electricity price forecasts. Each generator can choose the best bidding layout with the aid of these predictions. Besides, price forecasting has a significant impact on joint agreements and investments in new generation facilities over the long

* Corresponding author.
E-mail addresses: m.mohammadi@shirazu.ac.ir (M. Mohammadi), Mazaher.karimi@uwasa.fi (M. Karimi), E.Gryazina@skoltech.ru (E. Gryazina), V.Terzija@skoltech.ru (V. Terzija).

https://doi.org/10.1016/j.eswa.2023.121207
Received 17 March 2023; Received in revised form 10 August 2023; Accepted 13 August 2023
Available online 18 August 2023

term (Wolak, 2022). It is essential to forecast the price of electricity for generation companies or the Independent System Operator (ISO), as well as for investors and customers of varying levels. Essentially, in a competitive electricity market, various bidders require future electricity prices to maximize their profit margins. Price forecasting has become more difficult than in the past due to the modern energy markets' high degree of deregulation and nonlinearity. For the matter, the accuracy of the price prediction has decreased because of this system's nonlinearity and instability (Wang, Zhang, & Li, 2022). Additionally, it impacts the bidding policies, which results in a volatile electricity market.

Furthermore power market participants are paying close attention to adapting to the variation in electricity prices as new electricity reforms are imposed in the electric power industry. Forecasting electricity prices is essential information for managers and participants in the power market. In order to meet Demand-Side Management (DSM) models, the smart grid is evolving into a platform that enables power suppliers to modify their bidding methods to meet the requirements (Abdellatif, et al., 2022). Even though many approaches for electricity price forecasting have been proposed, which primarily differ in the data preparation, model selection, calibration, and testing phases, the literature on price forecasting is consistent. The main reason behind these restrictions is that the majority of electricity suppliers were founded on monopoly systems (Panapakidis & Dagoumas, 2016), resulting in the market being devoid of minimal competition. However, the electrical market has become more competitive in recent years. Therefore, price forecasting is becoming increasingly important, and many experts are starting to consider this element. Indeed, the electricity market is becoming accessible to all and competitive. Hence, the majority of electricity suppliers have devoted considerable time and effort to developing new techniques to find suitable methods for price forecasting (Abdellatif, et al., 2023; Mubarak, Ahmad, et al., 2023). Furthermore, in this deregulated and competitive market, electricity price forecasting (EPF) is a highly valued tool for the majority of electricity participants. By modifying their production schedules and selecting the appropriate bidding tactic, power providers and customers can use prediction approaches to maximize their advantages and reduce their electricity costs (Mubarak, Mansor, et al., 2021; Mubarak, Mokhlis, et al., 2021; Mubarak et al., 2022).

A robust, steady, and well-organized operation of the electricity market is ensured by price forecasting (Pourdaryaei, Mokhlis, Illias, Kaboli, & Ahmad, 2019). The EPF can help the operators to give an overview for future investments and developments if the implemented method is of high quality and accuracy. Nevertheless, various unique features are needed to consider, such as non-linearity, multiple seasonality, mean reversion, high frequency and volatility, price spikes, and the calendar effect to perform the high accuracy forecasting since it is a relatively tricky endeavor (Yang, Ce, & Lian, 2017). In addition, system load data are influenced by the status of the weather and the prices in that season, where both are controlled by a wider and more diverse collection of criteria (Panapakidis & Dagoumas, 2016).

Some of the strategies that have been offered in the past for accurately forecasting the electricity day-ahead price are outlined in (Aggarwal, Saini, & Kumar, 2009; Cerjan, Krželj, Vidak, & Delimar, 2013; Panapakidis & Dagoumas, 2016). Using a k-factor GIGARCH algorithm, for example, the German electricity pricing market was able to forecast its price in (Diongue, Guegan, & Vignal, 2009). As mentioned in (Aggarwal, Saini, & Kumar, 2008), ANN was utilized to forecast electricity demand and pricing. Using a combinatorial neural network-based forecasting engine, researchers in (Abedinia, Amjady, Shafie-Khah, & Catalão, 2015) predicted future pricing data values using a neural network, while in (Amjady & Daraeepour, 2009b), an iterative neural network was utilized to predict both load and price. In (Hahn, Meyer-Nieberg, & Pickl, 2009), the Support Vector Machine was used in the load forecasting domain. An SVM model was used to approximate the price forecasting in (Shiri, Afshar, Rahimi-Kian, & Maham, 2015), where it assesses the uncertainty by forecasting the extents of stated quantities.

Besides employing the aforementioned data-driven algorithms separately, additional feature selection techniques have been integrated with machine learning methods to improve the accuracy of the electricity price predictions. For instance, in (Amjady & Daraeepour, 2009a), the electricity prices were forecasted using CNN combined with the data mining approach. In (Amjady & Keynia, 2008), a feature selection technique and a cascaded neuro-evolutionary algorithm (CNEA) were combined to predict electricity prices. The work in (Amjady & Keynia, 2009) incorporates neural networks (NN) in addition to a stage selection feature system (SSFS) to forecast prices. In mainland Spain and New York, a technique based on discrete cosine transforms and a cascade forward NN was employed to categorize the electricity market price prediction. The adaptive neuro-fuzzy inference system (ANFIS) approach has been utilized to forecast electricity prices in this work (Pousinho, Mendes, & Catalão, 2012) by merging a fuzzy system and an ANN model (Neuro-Fuzzy). In (Catalão, Pousinho, & Mendes, 2010), multiple strategies have been merged into a hybrid model to increase prediction accuracy and non-linearity of the anticipated electricity price. The forecasters' ANN, ANFIS, and ARIMA are employed in (Catalão, Mariano, Mendes, & Ferreira, 2007) to anticipate the price of electricity in the Spanish electricity system using a Kalman filter. Out of three models, a modified ordered weighted average technique was chosen for price prediction (Catalão, Mariano, Mendes, & Ferreira, 2007). In (Catalão, Pousinho, & Mendes, 2009), a hybrid system for predicting electricity prices in the Spanish market is introduced, in which a mutual information technique picks the inputs and feeds them into the ANFIS is presented. A hybrid strategy utilizing a mix of evolutionary particle swarm optimization and ANFIS has been proposed for electricity price forecasting (Catalão, et al., 2010). Recent progress in price prediction has been investigated via developed computational intelligence approaches in (Pourdaryaei, et al., 2021) and (Huang, Shen, Chen, & Chen, 2021). Where a new hybrid model called SEPNet comprised of CNN, variational mode decomposition and gated recurrent unit for one hour ahead EPF (Huang, et al., 2021). The results showed that the proposed model outperforms the other models, and it was revealed that the MAPE and RMSE for all diverse seasons might be lowered by 84% and 81%, respectively, when variational mode decomposition is employed. For instance, long short-term memory (LSTM) was employed for electricity price forecasting. In (Zhou, Zhou, Mao, Tai, & Wan, 2019), the proposed LSTM managed to attain the lowest RMSE with a value of 1.72, whereas in (Torres, Martínez-Álvarez, & Troncoso, 2022), the proposed LSTM achieved the minimum error with less than 1.5% for one day-ahead electricity price forecasting. Further, the CNN-LSTM model was used widely for electricity price forecasting (Heidarpanah, Hooshyaripor, & Fazeli, 2022; Kuo & Huang, 2018; Lehna, Scheller, & Herwartz, 2022). In (Kuo & Huang, 2018), the one-hour-ahead forecasting horizon was applied to the PJM dataset, whereas the short-term electricity price was done for one day ahead using the German electricity market in (Lehna, et al., 2022), where the results demonstrated that the LSTM yields the best average prediction performance, followed closely by the two-stage VAR, which excels for shorter forecast horizons. In (Qiao & Yang, 2020), a different hybrid model comprised of WT, sparse autoencoder (SAE), and LSTM model has been employed for EPF. Further, the suggested study determines the optimal orders and layers of WT for predicting US power prices, offering a valuable reference to implement WT in other forecast scenarios and for electric market participants. Further, this work (Pourdaryaei, Mokhlis, Illias, Kaboli, & Ahmad, 2019) presented a hybrid approach for EPF based on ANN and artificial cooperative search algorithm (ACS), where the results showed that the proposed ANN-ACS attained mean absolute percentage error (MAPE) values of 4.58%, 2.62% and 3.79 % in winter, summer, and autumn, respectively. Moreover, a hybrid technique for EPF consisting of a two-stage selection algorithm and an optimized ANFIS method as a prediction engine is presented in (Pourdaryaei, Mokhlis, Illias, Kaboli, & Ahmad, 2019) for the purpose to accurately determine the electricity price. Both (Pourdaryaei, Mokhlis, Illias,

Table 1
A comprehensive summary of recent approaches employed for electricity price forecasting.

Ref.	Model	Inputs	forecasting horizon	Datasets	Feature selection	Verification test	Sensitivity indeces
(Kuo & Huang, 2018)	CNN-LSTM	Record of electricity price of the past 24 h	Short-term (Hour ahead)	PJM Regulation Zone Preliminary Billing Data	×	×	×
(Lehna, et al., 2022)		24, 168, and 720 hourly lagged EP values	Short-term (Day, week, and month ahead)	German electricity spot price			
(Heidarpanah, et al., 2022)		168 hourly lagged EP values	Short-term (Day ahead)	Iranian electricity market			
(Chang, Zhang, & Chen, 2019)	WT-Adam-LSTM	168 hourly lagged EP values	Short-term (Hour ahead, day ahead)	New South Wales of Australia and French			
(Qiao & Yang, 2020)	WT-SAE-LSTM	Record of electricity price of the past 24 h	Short-term (Hour ahead)	U.S. Energy Information Administration (EIA)			
(C. J. Huang, et al., 2021)	SEPNet			Electricity price data for New York City in the United States from			
(Alkawaz, Abdellatif, Kanesan, Khairuddin, & Gheni, 2022)	ARD-ETR	168 hourly lagged EP values	Short-term (Day ahead)	Nord Pool market			
(Alamaniotis, Bargiotas, Bourbakis, & Tsoukalas, 2015)	RVMs-LR			New England Electricity market			
(Zhou, et al., 2019)	LSTM	24, 168, hourly lagged EP values	Short-term (Hour and day ahead)	PJM			
(Torres, et al., 2022)		Record of electricity price for the past 96 h	Four hours ahead	Spanish electricity dataset			
(T. Zhang, Tang, Wu, Du, & Chen, 2022)	MD-Res.-EEMD-DE-ELM-DE-EL	Record of electricity price for the past 24, 48, 96, and 144 h	1, 2, 4, and 6 h ahead	Spanish and Australian electricity dataset			
(Pourdaryaei, Mokhlis, Illias, Kaboli, & Ahmad, 2019)	MOBBSA-ANFIS	168 hourly lagged EP values	Short-term (Day ahead)	Ontario market	MOBBSA		
Proposed	1D-CNN with self-attention				MI + NN	Two step-verification tests	PPC, SRC, KTC

Kaboli, & Ahmad, 2019; Pourdaryaei, Mokhlis, Illias, Kaboli, & Ahmad, 2019) consider the near-term horizon and verify their findings utilizing Ontario electricity markets. Compared to the proposed method, Table 1 provides a summary of new models for forecasting electricity prices that take into account various methods and numerous parameters.

According to the previous work discussed above, many didn't consider feature selection in their proposed models even though feature selection in electricity price forecasting provides advantages such as enhanced model performance, improved interpretability, reduced over fitting, and faster training and inference as well as scalability and adaptability. However, few works consider it, yet some approaches might yield a better result. With the non-linearity of the price signal, it has been noted that selecting a powerful feature selection strategy for forecasting electricity prices is quite challenging. Hence, a feature selection strategy combining mutual information (MI) and Neural Network (NN) is proposed to improve the model effectiveness with reasonable computational time. Further, many works employed the LSTM and CNN models to forecast the electricity price; adapting a self-attention mechanism in electricity price forecast can yield better results. The self-attention excels in capturing long-term dependencies and is efficient in modelling complex temporal patterns present in electricity price data. In this regard, a novel methodology for electricity price forecasting depending on a multi-head self-attention and 1D-CNN-based technique is provided in this work to address the above difficulties. In addition, most of the previous work overlooked taking into consideration the verification test and sensitivity indices to validate the effectiveness of their proposed model. Finally, to fill the previous work gaps, the research contribution is formulated as follows:

The following is the list of the research work's contributions:

- The study's key contribution is to offer a better price forecasting technique based on a multi-head self-attention, and Convolutional Neural networks approach to supersede other price forecasting methods in achieving less error according to the performance metrics such as MSE, RMSE, and MAE. An attention function was used to find the best appropriate CNN weights values for obtaining the least error, resulting in enhanced predicting precision. Attention has significantly improved natural language processing in translation and language generation. It was also extended to other fields, such as computer vision in tasks like image classification. Despite its simple structure, CNN has been frequently employed in numerous numerical problems because of its efficacy in solving multidimensional functions.
- A second contribution has been carried out concerning the feature selection problem related to these issues. A robust hybrid feature selection method has been created to address this problem, which utilizes a combination of NN and mutual information techniques. In this case, NN was utilized to select the optimal subset of features, whereas MI was used to extract input variables with the least redundancy in addition to the most relevance. When the network training procedure is completed, redundant network connections can be separated from those relevant by utilizing a penalty term to the error function of the network.
- The effectiveness and accuracy of the proposed hybrid forecasting approach are examined by comparing the obtained results with those obtained by recent development techniques for price prediction via enhanced deep learning approaches.

The next sections are structured in the following ways: A quick overview of the structure and evolution of CNN and multi-head

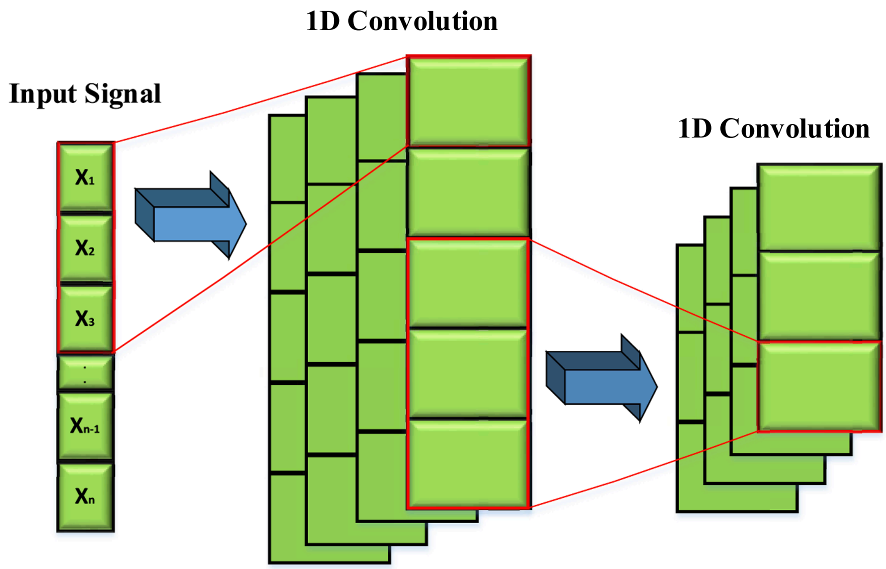

Fig. 1. The convolution process for 1D-CNN.

attention along with model development, respectively, is provided in section 2. The process of developing short-term electricity price forecasting models is discussed in Section 3. Additionally, it covers the efficient feature selection technique for obtaining the maximum effective merits for short-term EP forecasting by first filtering the input variables utilizing MI and creating the NN method in the next phase. Section 4 contains a detailed explanation and results, which demonstrates that the proposed approach is highly suited and practical in a deregulated and competitive electricity market for forecasting future electricity prices. Finally, section 5 brings the work to a conclusion.

2. Model architecture

Convolutional neural networks (CNNs) are extensively used in image processing due to their superior classification and regression abilities (Abo-Tabik, Costen, Darby, & Benn, 2020; Bai, et al., 2020; Krizhevsky, Sutskever, & Hinton, 2017). Essentially, digital filters can be used to extract features before employing completely linked layers for categorization or regression-based prediction based on original data. Gradually, CNNs have been classed in various ways, including one-dimensional convolutional neural networks (1D-CNN) and two-dimensional convolutional neural networks (2D-CNN), so on and so forth. The data to be analyzed and forecasted in this study is a one-dimensional time series for the price of electricity. As a result, 1D-CNN is deployed to extract the hourly electricity price time-domain feature for the Ontario mainland in due course. The convolution processes for 1D-CNN are depicted in Fig. 1. The numbers on the left and right with the same tone represent the input and output after convolution, respectively as illustrated in Fig. 1. Notably, the filtering direction is unidirectional, i.e., vertical, in 1D-CNN, in contrast to 2D-CNN, which can be calculated based on Eq. (1).

$$y_{f,CNN} = \sum_{m=3f-2}^{3f} \sum_{n=1}^{6} x_{mn} w_{cn} \tag{1}$$

where, $f \in \{1,2\}$ and $c = m - 3 \times (f - 1)$

It is a routine trend to include a pooling layer between the convolutional layers of a CNN to facilitate down-sampling, model size reduction, and calculation boosting and increase the resilience of the features retrieved. Time-domain feature extraction is completed when the filter's convolution operation is performed on the original data. This layer acts as a fully connected layer for the purpose of performing the final forecasting. In a deep learning model, the activation function is equally crucial. The activation functions in this paper are the Rectified Linear Unit (ReLU) and the Scaled Exponential Linear Unit (SeLU). The ReLU function is written in the following form in equation (2).

$$\begin{gathered} ReLU_{CNN}(x) = x \\ for\ x\ greater\ than\ Zero \end{gathered} \tag{2}$$

ReLU activation is clearly a piecewise linear function, as can be seen. It transforms all minus numbers to zero while leaving positive numbers stagnant. Unilateral suppression is the term used to describe this procedure. At the same time, the sparse activity of the neurons in a CNN is caused by unilateral suppression. Therefore, the model can mine-related features and fit the training data more effectively. ReLU benefits from the stagnant gradient of its non-negative portion, preventing the vanishing gradient problem, such as the sigmoid, in comparison to other activation functions. In 2017, Klambauer et al. presented the SeLU (Zhang, Xing, Bai, Sun, & Meng, 2020). Self-normalization of the SeLU has been deployed to characterize it. With the Banach fixed-point theorem, it is possible to demonstrate that activations propagating through numerous network layers around zero mean and unit variance will converge to zero mean and unit variance even in the face of noise and perturbations. A piecewise linear function is an activation function of SeLU. It can be stated in the following manner in equation (3).

$$\begin{gathered} SeLU_{CNN}(x) = \lambda\alpha(e^{x} - 1) for\ x \leqslant 0 \\ SeLU_{CNN}(x) = \lambda x\ for\ x > 0 \\ where\ \lambda = 1.0507 \\ where\ \alpha = 1.6733 \end{gathered} \tag{3}$$

This work proposes a new approach based on a self-attention mechanism combined with a one-Dimensional Convolutional Neural Network (1D-CNN) for more electrical price forecasting accuracy. The self-attention tool is commonly employed in deep learning in order to solve problems with small inputs and outputs. The main concept of the attention mechanism is to implement the multi-headed self-attention instead of the recurrent layers, which is most frequently utilized in encoder-decoder architectures (Vaswani, et al., 2017). The self-attention mechanism has many features, but the most significant merit is its ability to learn long-range dependencies. The self-attention has proven superiority over other models with difficulties learning long-range dependencies and only relies on the short paths between the input and output sequences. The function of attention can be represented as a function that maps a query and collection of key-value pairs to an

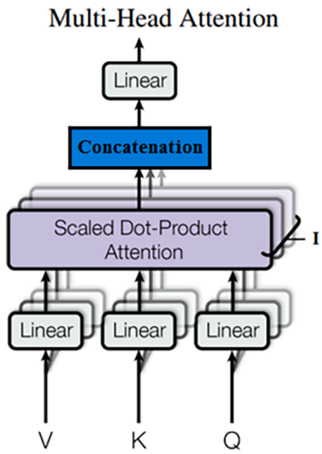

Fig. 2. Multi-Head Attention.

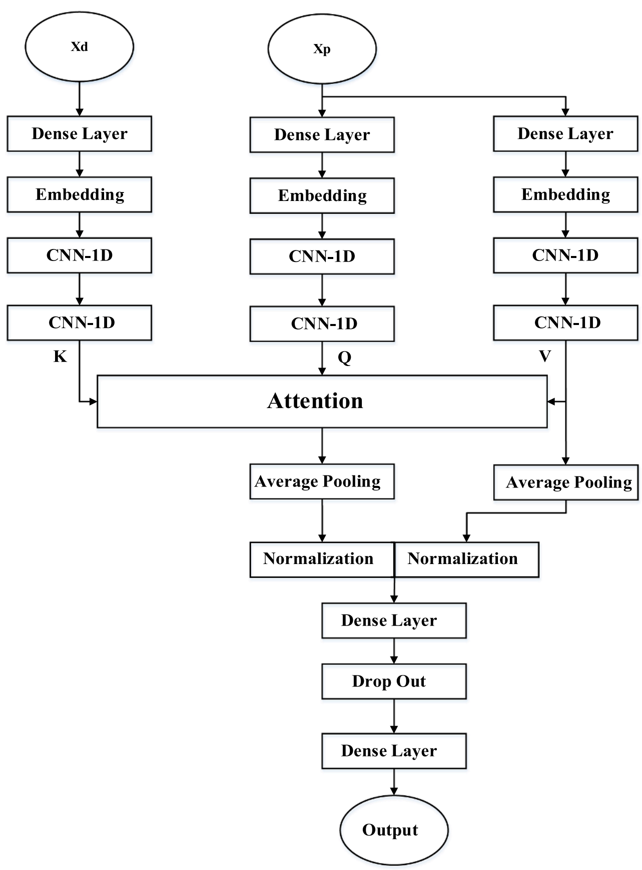

Fig. 3. The process of implementing the proposed 1D-CNN and self-attention in forecasting electricity price.

output, where parameters such as keys, output, query, and values, where each one of them is in vector form. The result is generated as a weighted sum of the values, with a compatibility function of the query calculating the weight allocated to every value concerning the compatible key (Vaswani, et al., 2017). There are two types of attention; the first one is scaled dot-product attention, while the second one is multi-head attention. In this work, we will consider the multi-head attention and elaborate more on it. In multi-head attention, the layers are running in parallel in contrast to the scaled dot-product where it's executed once. The model in multi-head attention is permitted to simultaneously access the information from various representation subspaces at diverse positions. On the other hand, the averaging prevents that in the case of the single attention head. We discovered that it was more effective to linearly project the queries, keys, and values h times using various, learned linear projections to d_k, d_k, and d_v dimensions, respectively. On each of these projected versions of queries, keys, and values, then we parallelly apply the attention function to generate the d_v-dimensional output values. The equations (4–6) demonstrate how the attention function works, and Fig. 2 illustrates the multi-head attention process where I represent the multi-head.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4}$$

$$\text{MultiHead}(Q, K, V) = \text{Concat}(head_1, \cdots\cdots, head_h)W^O \tag{5}$$

$$head_i = \text{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right) \tag{6}$$

On the other hand, 1D-CNN is mainly utilized in applications with time-series data, where the input–output data in the one-dimensional CNN is 2D. The combination between the self-attention and 1D-CNN is used to maximize the performance of the relation extraction (Huang & Du, 2019). Consequently, using the proposed approach helps to enhance the electrical price forecasting accuracy by finding a better relation extraction between the long-range dependencies. 1D-CNN with self-attention is applied in this study. It features a basic structure with excellent efficiency in solving multimodal functions that readily adapt to varied numerical optimization issues. Moreover, 1D-CNN with self-attention has been improved to cope with the weaknesses of meta-heuristic methods. For example, metaheuristic methods have various control parameters and over-sensitivity to the initial value of these parameters, premature convergence, and time-consuming computation (Civicioglu, 2013)). The 1D-CNN and self-attention are more straightforward since they have only one control parameter, and it is not overly sensitive to the initial parameter value. The flowchart of the proposed 1D-CNN and self-attention is demonstrated in Fig. 3, where X_d and X_p refer to the electrical demand and electricity price, respectively.

3. Electricity price forecasting enrichment

It displays a time series divided into per-hour intervals, which can be observed in a competitive electricity market according to the relationship between electricity price and demand circumstances. The price of electricity is determined by its current and the previous values of electricity prices and demand. It is written as the following:

$$HEP(t)=F\begin{pmatrix}HEP(t-1),HEP(t-2),HEP(t-3),\ldots,HEP(t-NL_{HEP}),\\HED(t),HED(t-1),HED(t-2),HED(t-3),\ldots,HED(t-NL_{HED})\end{pmatrix} \tag{7}$$

In deregulated and competitive markets, electricity prices and demand at instants time *t* and considering them as a time series are represented by *HEP (t)* and *HED (t)*, respectively. In the Eq.7, NL_{HEP} presents the number of lag order for the electricity price and similarly NL_{HED} denotes electricity demand lag order.

This work aims to implement a new method for forecasting hourly Ontario electricity prices (HEP). The 2020 input HED and HEP historical data sets were obtained from (IESO, 2018-11-17). Since historical data has a wide range, the independent and dependent variables are normalized using Eq. (8). It is common practice to calibrate the data gathered on different scales to estimate a common scale prior to the data processing to normalize the data. Where the already normalized data is denoted by $\overline{Z}_n$ the data required to be normalized is denoted by Z_n, and finally, the interval per hour is represented by t (IESO, 2018-11-17).

$$\overline{Z_n}(t) = \frac{Z_n(t) - \min(Z_n)}{\max(Z_n) - \min(Z_n)} \tag{8}$$

In this analysis, one week of exogenous variables ($NL_{HEP} = NL_{HED} =$

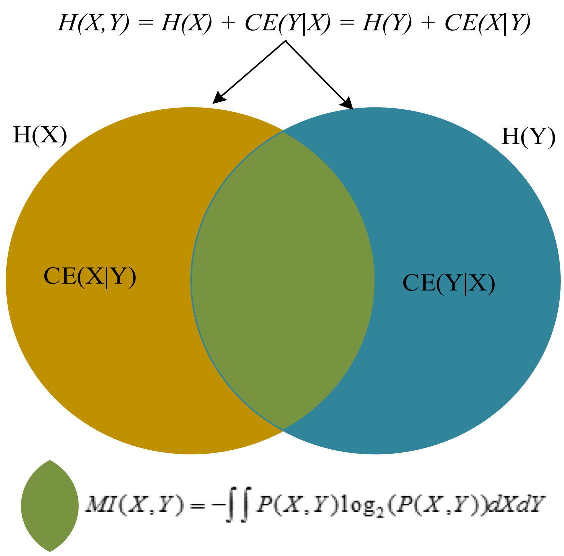

Fig. 4. The MI between each input and output characteristic.

168) with hourly lagged values is considered, with the availability of a total of 336 exogenous variables delayed values. Machine learning algorithms which overburdened with too many features have caused the training data to be overfitting, a slowdown learning process, and underperformance. SAs a result, machine learning algorithms should be attributed to only those features that significantly impact the output (for the process of forecasting electricity price). In statistics and machine learning, feature (variables or predictors) selection, also known as attribute selection, variable selection, or variable subset selection is a technique aimed to construct models by picking a subset of relevant features. Utilizing feature selection approaches has a three-fold purpose (Renani, Elias, & Rahim, 2016):

1. Reducing overfitting in the predictors to improve their prediction performance (formally, variance reduction).
2. Expediting and reducing the cost of the model-building procedure (facilitate learning procedures).
3. Present the streamlined model, which is simpler to comprehend (enhancing generalization capability).

The mutual information (MI) techniques have been extensively used in (Amjady & Hemmati, 2006) for electricity market price forecasting. Nevertheless, there are issues with this technique because of the lagged numbers provided by the electrical market in terms of price, load demand, and other variables. Because of this, it is obtaining the individual and combined probability distributions of the candidate input is challenging. In addition, the price of electricity is a signal that changes over time. As a result, a long history of the candidate input is insignificant to use because market conditions change constantly. Consequently, it can potentially mislead or provide erroneous price forecasting processes due to a lack of values of data (Amjady & Daraeepour, 2009a).

The primary objective of the MI is to acquire the mutual relationship among two arbitrary variables X and Y. In contrast, in this technique, the information amount is attained according to another variable that is random in nature by using one variable. As a result, if variable X does not hold any information concerning variable Y, the mutual information is zero, and vice versa. Therefore, these two random variables are distinct. If variable X is a deterministic function of variable Y and variable Y is a deterministic function of variable X, high mutual information is gained (Cover & Thomas, 2006).

Fig. 4 depicts the correlation between MI and conditional entropy (CE). When MI is huge, it is clear that X and Y are inextricably linked and reliant on one another. Aside from entropy, CE is also monitored. Following the second random variable, it is a measure of the first random variable's average uncertainty. As illustrated in Fig. 4, the joint probability distribution of PXY(X,Y) is implemented to accomplish the MI between X and Y and MI(X,Y) random variables because random variable entropy is inextricably linked to the MI topic.

Let $X=\{HEP(t-1), HEP(t-2), HEP(t-3),\ldots, HEP(t-NL_{HEP}), HED(t), HED(t-1), HED(t-2), HED(t-3),\ldots, HED(t-NL_{HED})\}$ is the input feature vector, and $Y = HEP(t)$ is the output feature goal. In the electrical market, *HEP (t)* represents the price while *HED (t)* represents demand. The MI between each input and output characteristic is depicted in Fig. 4. For example, the MI price and demand of electricity between the first lagged values are generated by *MI {HEP (t-1); HED (t)}*. Based on MI, all input features are then ranked discerningly. A higher value of MI indicates a stronger relationship between input and output variables. When the MI value is lower than the threshold *TH*, it is discarded because it has less impact on the output, and the subset, $TX \subset X$ is formed with the remaining input features.

The neural network approach detects and removes redundant features in the second step once redundant features are eliminated. A three-layer feed-forward neural network is shown in Fig. 5 to demonstrate its usage. For the redundant input, the input to the hidden layer and the input from the hidden to the output layer are both sluggish. As a result, it will be removed because it has a minor impact on network precision. Typically, the error function which is assigned throughout the training process contains:

$$F_{NN} = \frac{1}{N}\sum_{j=1}^{N}\left(t_j - y_j\right)^2 \tag{9}$$

where N denotes observation number, t denotes network output, and the

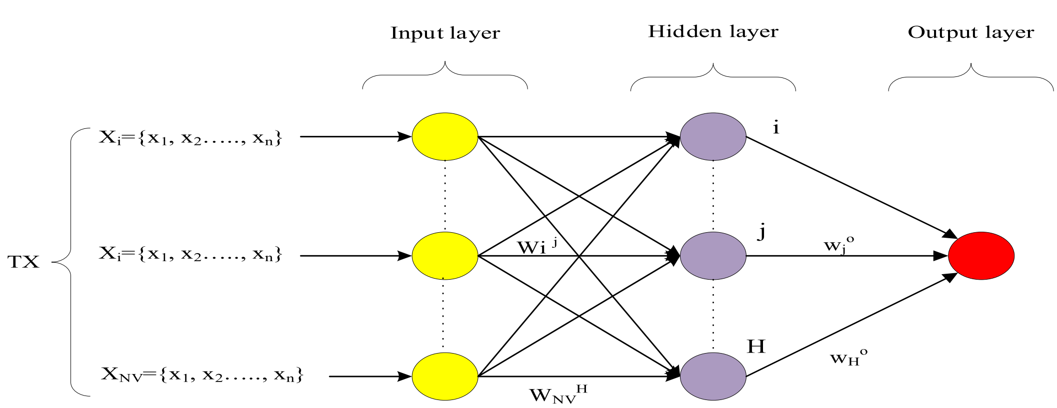

Fig. 5. The 2nd stage of feature selection is according to the NN model.

real value is represented by y.

An additional Penalty Function (PF) is combined with the error function by Eq. (10) to discover the unnecessary and features redundancy, where the impact of the PF is controlled by coefficients, $\alpha_{PF,1}$, $\alpha_{PF,2}$, and β_{PF}. The number of hidden units is denoted by H. The number of features (Variables) chosen in the 1st stage is indicated by NV, in which the weight is linking from the *i-th* attribute to *j-th* hidden unit, and the weight is linking from the *j-th* hidden unit to the output of the network.

$$P.F(W) = \alpha_{PF,1}\left(\sum_{j=1}^{H}\sum_{i=1}^{NV}\frac{\beta_{PF}\left(W_i^j\right)^2}{1+\beta_{PF}\left(W_i^j\right)^2} + \sum_{j=1}^{H}\frac{\beta_{PF}\left(W_j^o\right)^2}{1+\beta_{PF}\left(W_j^o\right)^2}\right) + \alpha_{PF,2}\left(\sum_{j=1}^{H}\sum_{i=1}^{NV}\left(W_i^j\right)^2 + \sum_{j=1}^{H}\left(W_j^o\right)^2\right) \tag{10}$$

Utilizing the set of input features $TX=\{Xi,\ldots,X_{NV}\}$, $TX \subset X$, $NV < (NL_{HEP} + NL_{HED})$, the NN model will initially assess the network's accuracy. The number of features is then lowered successively to generate a new set of input features, and the accuracy of that network is then evaluated, N_P where $P=\{1,2,\ldots,NV\}$. The network's accuracy will be calculated to estimate the number of features that can be removed altogether. The following are the steps involved in applying feature selection:

1. For a given input vector $TX=\{Xi,\ldots,X_{NV}\}$, the dual data set is divided from $TX \subset X$, which are training set, TX_{tr} and testing set TX_{ts}. Both TX_{tr} and TX_{ts} accuracy are calculated using the network N that has been trained. According to (Setiono & Liu, 1997), using this technique, the value of $\alpha_{PF,1}$, $\alpha_{PF,2}$, and β_{PF} are set to 10^{-1}, 10^{-4} and 0.03, respectively.
2. N_P does not provide several NV features for $TX=\{Xi,\ldots,X_{NV}\}$ denotes as the network input features. For example, input features of network N_3 is represented by $TX=\{4,5,\ldots,NV\}$. The network N_3 is trained, and the precision of the training and testing sets, PTX_{tr} and PTX_{ts}, respectively, are computed.
3. Based on the accuracy of the testing set $PTX_{ts}^1 \geqslant PTX_{ts}^2 \geqslant \ldots \geqslant PTX_{ts}^{NV}$, the network N_k has been ranked. The average value of this accuracy PTX_{ts}^{avg} is then calculated.
4. By updating the penalty parameter, the step is continued. For any, the network's accuracy N_P, represented by PTX_{ts}^{NV} less than PTX_{ts}^{avg}, the multiplication between the weight values W_i^j and W_j^o and 1.1 is carried out. Otherwise, the values will be divided by 1.1. Thus, after the network has been retrained, this permits significant input to have a greater connection magnitude. It's worth noting that this algorithm will discard the input feature with a poor connection magnitude.

The application of feature selection techniques is necessary when narrowing the list of potential input variables. The best NN model's final predictor variables (HED and HEP in the preceding hours) were evaluated by selecting a feature. These variables were discovered after multiple models with different input variables combinations were developed and controlled. It is necessary to use a hybrid feature selection to shorten the overall running time. TH = 0.46 was chosen as the relevancy level for filtering out redundant features in the first stage of hybrid feature selection, and then 67 relevant features were identified following the filtering stage. Even though the threshold TH was computed by an experiment, a larger value of TH leads to missing a lot of information. While, a lower value of TH may include too many features, either relevant or irrelevant result in a large computational burden. Therefore, several thresholds were experimented as mentioned in price forecasting related area. The obtained value is a suitable and the best value in designing feature selection process in first stage of filtering is selected. In the second stage, this study uses (NN) to choose input variable subsets that significantly impact electricity price prediction. As a result of hybrid MI and NN in this stage, the most relevant and different attributes from the previously selected 67 candidates were chosen and used as input for the predicting procedure. Table 2 is shown the input variable subsets selected by (MI + NN). The details process of feature selection has been presented in Table 3 as a general Pseudo code.

Table 2
Selected features by (MI + NN) for forecasting electricity price of Ontario mainland.

Selected features by MI + NN	No.	Selected features by MI + NN	No.
HED(t)	1	*HED(t-73)*	16
HEP(t-1)	2	*HEP(t-96)*	17
HED(t-1)	3	*HEP(t-97)*	18
HEP(t-2)	4	*HED(t-97)*	19
HED(t-2)	5	*HEP(t-121)*	20
HEP(t-3)	6	*HED(t-121)*	21
HEP(t-23)	7	*HEP(t-144)*	22
HEP(t-24)	8	*HED(t-144)*	23
HED(t-24)	9	*HEP(t-145)*	24
HEP(t-25)	10	*HEP(t-168)*	25
HED(t-25)	11	*HED(t-168)*	26
HEP(t-48)	12	*HEP(t-169)*	27
HEP(t-49)	13	*HEP(t-192)*	28
HEP(t-72)	14	*HEP(t-193)*	29
HEP(t-73)	15	*HEP(t-337)*	30

Three groups of feature selection methods are formed by how a search strategy is integrated with a learning process for model development, notably wrappers and filtering methods. Wrapper approaches evaluate feature subsets using a predictive model (Kohavi & John, 1997). Wrapper approaches include training a model that is then evaluated on a holdout set using each potential feature subset. By assessing the model's error rate on the testing set, the score for each candidate subset is generated. Wrapper approaches frequently provide the highest-performing feature set for that particular type of model at the expense of computationally intensive chores since they train a new predictive model for each candidate subset. In filter (information gain) approaches, a proxy measure is utilized to grade a candidate feature subset rather than the error rate. In order to enable quick computation for capturing the efficacy of the feature set, proxy measures such as pointwise mutual information, Pearson product-moment correlation coefficient, and mutual information are selected. The computing burden of the wrapper techniques is significantly larger than that of the filter methods; nevertheless, the wrapper methods supply a subset of features, and a particular sort of learning algorithm evaluates the performance of these features. Due to the absence of a learning algorithm in filters, a feature set derived from filter methods is typically more general and provides inferior prediction performance compared to a feature set derived from wrapper methods. The filter functions as a dimensionality reduction technique in this hybrid approach, allowing the wrapper technique to be used to choose the most important features from bigger data sets. Multiple combinations of filter and wrapper techniques are implemented in Table 4 to demonstrate the superiority of the proposed method. The table shows the proposed MI + NN achieved the lowest computational time at 96.65 s.

4. Experimental setup and results

4.1. The experimental setup of the proposed methodology

Due to its single-settlement nature, the electricity market is considered a primarily inconstant market globally. The multi-head attention and 1D-CNN are deployed in this study to enhance the accuracy of HEP's electricity price forecasting. The suggested feature selection approach (MI-NN) selects the most influential input variables for electricity price prediction. Additionally, to prove the efficacy of ANFIS-BSA short-term

Table 3
The Pseudo code for feature selection process.

```
Input: x, y, Maximum epoch, Number of Neurons in Hidden layers along with hidden layers number,
Analyzing transfer function algorithm
The initial weight is allocated arbitrarily
for every epoch value ← 1 to Maximum epoch insert
    for creating each pattern in the training phase consider
        display the network pattern
        for each layer at the network insert
            for each node in every layer insert
                Calculate the summation weight for inputs out going from the node
                The threshold is added to the summation weight
                Compute for every node the activation function
            end
        end
        for every node inside the output layer insert
            The error signal is calculated in this step
        end
        for the whole hidden layers consider
            for every neuron inside layer consider
                Compute the signal of every error
                Find new value for every node's weight at network
            end
        end
        Compute the final amount of the Error Function
    end
    Data: TX-DP, P, N, V
    P=P+1 ← do
    Train the neural network using the pseudo code [#number]
end
While P<NV, choose the best features
Computation output: ŷ_NN
```

EPF, the acquired findings are compared to the following techniques such as Attention-LSTM, CNN1D, LSTM, and ANN. Fig. 6 shows the selection of different months in each season along with the forecasting process, where in each selected month, the first three weeks were used for training while the last week of the month was employed for testing, as demonstrated in Fig. 7. There are two challenging concerns for selecting the train/test split: with less training data, the model estimates have greater variance. With less testing data, the model performance statistic will have greater variance. Generally, should be concerned with splitting the train/test data such that neither variance is too high, which relates more to the absolute number of instances in each group than the percentage. In this study, we select a 70:30 ratio for data split. Fig. 8 illustrates the structure development for the suggested methodology for short-term EPF. The following procedures are considered in general when executing the short-term EPF according to optimum computational intelligence models.

A. The electricity price is calculated using hourly intervals, which take into account the previous hours for the electricity price and demand. According to the suggested feature-selection technique, the Ontario electricity market in 2020 (HEP and HED) are chosen as independent variables, whereas HEP is calculated as a dependent variable.
B. One month from each season is used to evaluate the performance of applicable methodologies for forecasting the electricity price in different seasons as the pattern of electricity demand seasonally changes. Variables are classified into dependent and independent. They are separated into two sections; the first is the training, where the first three weeks of hourly data are utilized for design phase training, while the second is testing, the last week of each month's hourly data is utilized for the testing phase of models created from the analysis.

C. The training phase is where the learning process takes place. The computer programs, which are derivative as a learning process, relate the independent factors to the dependent variables. It can be made faster by normalizing the input and output, according to (8). Although this stage of testing has no bearing on the development of the models, it is employed to assess the performance of AI-based methods. The model's performance in predicting price is illustrated by taking into account $\mathbf{MAPE_{EP}}$ (mean absolute percentage error) and $\mathbf{RMSE_{EP}}$ (root mean square error). A similar sort of index is $\mathbf{MAE_{EP}}$ (mean absolute error) and $\mathbf{MSE_{EP}}$ (mean squared error), which are computed additionally to indicate the efficacy of prediction models (Mubarak, Hammoudeh, et al., 2023). These indexes' mathematical equations are as follows (Mubarak, Hammoudeh, et al., 2023):

$$MAPE_{EP}\% = \frac{1}{N}\sum_{t=1}^{N}\frac{\left|\left(HEP(t)_{real} - HEP(t)_{predicted}\right)\right|}{HEP(t)_{real}} \times 100 \tag{11}$$

$$RMSE_{EP} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(HEP(t)_{real} - HEP(t)_{predicted}\right)^2} \tag{12}$$

Table 4
Comparisons between feature selection techniques.

MI + NN	MI + MOPSO + ANFIS	MI + NSGAII + ANFIS	MI + NSGAIII + ANFIS	MI + MOPSO + ANN	MI + NSGAII + ANN	MI + NSGAIII + ANN	MI + MI
HEP(t-1)	HEP(t-1)	HEP(t-1)	HEP(t-1)	HEP(t-1)	HEP(t-1)	HEP(t-1)	HEP(t-1)
HEP(t-2)	HEP(t-2)	HEP(t-2)	HEP(t-2)	HEP(t-2)	HEP(t-2)	HEP(t-2)	HEP(t-3)
HEP(t-3)	HEP(t-3)	HEP(t-3)	HEP(t-3)	HEP(t-3)	HEP(t-3)	HEP(t-6)	HEP(t-23)
HEP(t-23)	HEP(t-23)	HEP(t-23)	HEP(t-23)	HEP(t-23)	HEP(t-23)	HEP(t-23)	HEP(t-24)
HEP(t-24)	HEP(t-25)	HEP(t-47)	HEP(t-24)	HEP(t-24)	HEP(t-73)	HEP(t-24)	HEP(t-25)
HEP(t-25)	HEP(t-47)	HEP(t-48)	HEP(t-48)	HEP(t-25)	HEPt-96)	HEP(t-96)	HEP(t-97)
HEP(t-48)	HEP(t-48)	HEP(t-49)	HEP(t-49)	HEP(t-47)	HEP(t-97)	HEP(t-120)	HEP(t-120)
HEP(t-49)	HEP(t-49)	HEP(t-71)	HEP(t-72)	HEP(t-49)	HEP(t-120)	HEP(t-121)	HEP(t-144)
HEP(t-72)	HEP(t-71)	HEP(t-72)	HEP(t-73)	HEP(t-73)	HEP(t-121)	HEP(t-144)	HEP(t-168)
HEP(t-73)	HEP(t-72)	HEP(t-73)	HEP(t-96)	HEP(t-94)	HEP(t-144)	HEP(t-167)	HEP(t-169)
HEP(t-96)	HEP(t-96)	HEP(t-96)	HEP(t-97)	HEP(t-120)	HEP(t-145)	HEP(t-168)	HEP(t-192)
HEP(t-97)	HEP (t-144)	HEP(t-119)	HEP(t-120)	HEP(t-167)	HEP(t-167)	HEP(t-169)	HEP(t-193)
HEP(t-121)	HEP(t-168)	HEP(t-120)	HEP(t-121)	HEP(t-168)	HEP(t-168)	HEP(t-192)	HEP(t-335)
HEP(t-144)	HEP(t-169)	HEP(t-144)	HEP(t-144)	HEP(t-169)	HEP(t-169)	HEP(t-336)	HEP(t-337)
HEP(t-145)	HEP(t-192)	HEP(t-145)	HEP(t-145)	HEP(t-191)	HEP(t-335)	HEP(t-337)	HEP(t-503)
HEP(t-168)	HEP(t-193)	HEP(t-168)	HEP(t-168)	HEP(t-192)	HEP(t-336)	HEPt-504)	HEP(t-504)
HEP(t-169)	HEP(t-335)	HEP(t-192)	HEP(t-169)	HEP(t-336)	HEP(t-337)	HEP(t-505)	HED(t)
HEP(t-192)	HEP(t-337)	HEP(t-334)	HEP(t-192)	HEP(t-504)	HEP(t-504)	HED(t)	HED(t-1)
HEP(t-193)	HED(t)	HEP(t-335)	HEP(t-335)	HED(t)	HEP(t-505)	HED(t-1)	HED(t-4)
HEP(t-337)	HED(t-1)	HED(t)	HED(t)	HED(t-1)	HED(t)	HED(t-3)	HED(t-12)
HED(t)	HED(t-2)	HED(t-1)	HED(t-1)	HED(t-2)	HED(t-1)	HED(t-24)	HED(t-24)
HED(t-1)	HED(t-24)	HED(t-2)	HED(t-2)	HED(t-24)	HED(t-3)	HED(t-25)	HED(t-25)
HED(t-2)	HED(t-25)	HED(t-24)	HED(t-24)	HED(t-25)	HED(t-24)	HED(t-71)	HED(t-48)
HED(t-24)	HED(t-72)	HED(t-25)	HED(t-72)	HED(t-72)	HED(t-72)	HED(t-72)	HED(t-72)
HED(t-25)	HED(t-73)	HED(t-73)	HED(t-73)	HED(t-73)	HED(t-73)	HED(t-73)	HED(t-73)
HED(t-73)	HED(t-121)	HED(t-96)	HED(t-96)	HED(t-96)	HED(t-96)	HED(t-96)	HED(t-96)
HED(t-97)	HED(t-144)	HED(t-121)	HED(t-120)	HED(t-97)	HED(t-97)	HED(t-97)	HED(t-97)
HED(t-121)	HED(t-168)	HED(t-144)	HED(t-121)	HED(t-144)	HED(t-168)	HED(t-168)	HED(t-168)
HED(t-144)	HED(t-169)	HED(t-168)	HED(t-168)	HED(t-168)	HED(t-169)	HED(t-192)	HED(t-192)
HED(t-168)	HED(t-192)	HED(t-335)	HED(t-169)	HED(t-169)	HED(t-193)	HED(t-336)	HED(t-335)
	HED(t-336)	HED(t-336)	HED(t-335)	HED(t-192)	HED(t-335)	HED(t-335)	HED(t-336)
			HED(t-336)	HED(t-336)	HED(t-336)	HED(t-336)	HED(t-503)
Computational Time							
96.6523	113.3563	115.4775	116.5256	125.4375	130.4467	140.6275	155.3658

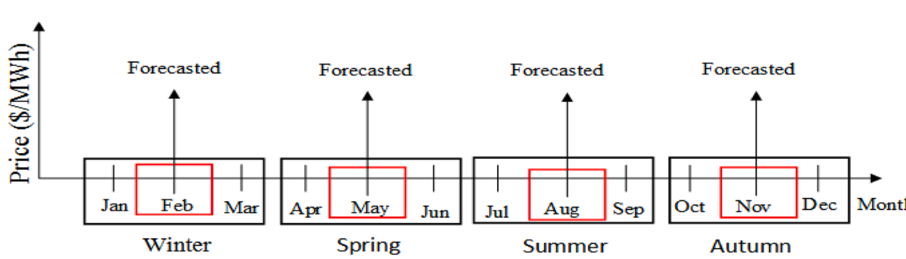

Fig. 6. Time horizon with seasonality selection for electricity price prediction.

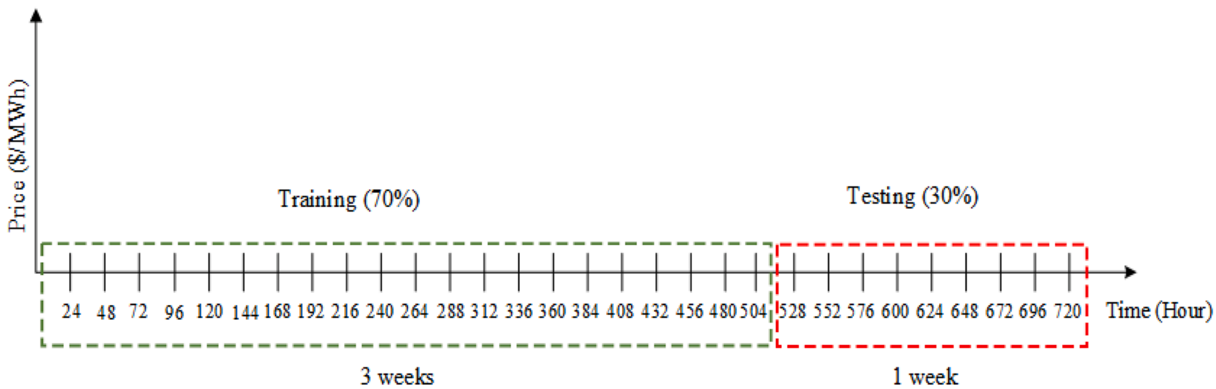

Fig. 7. Data splitting for training and testing phases for each season.

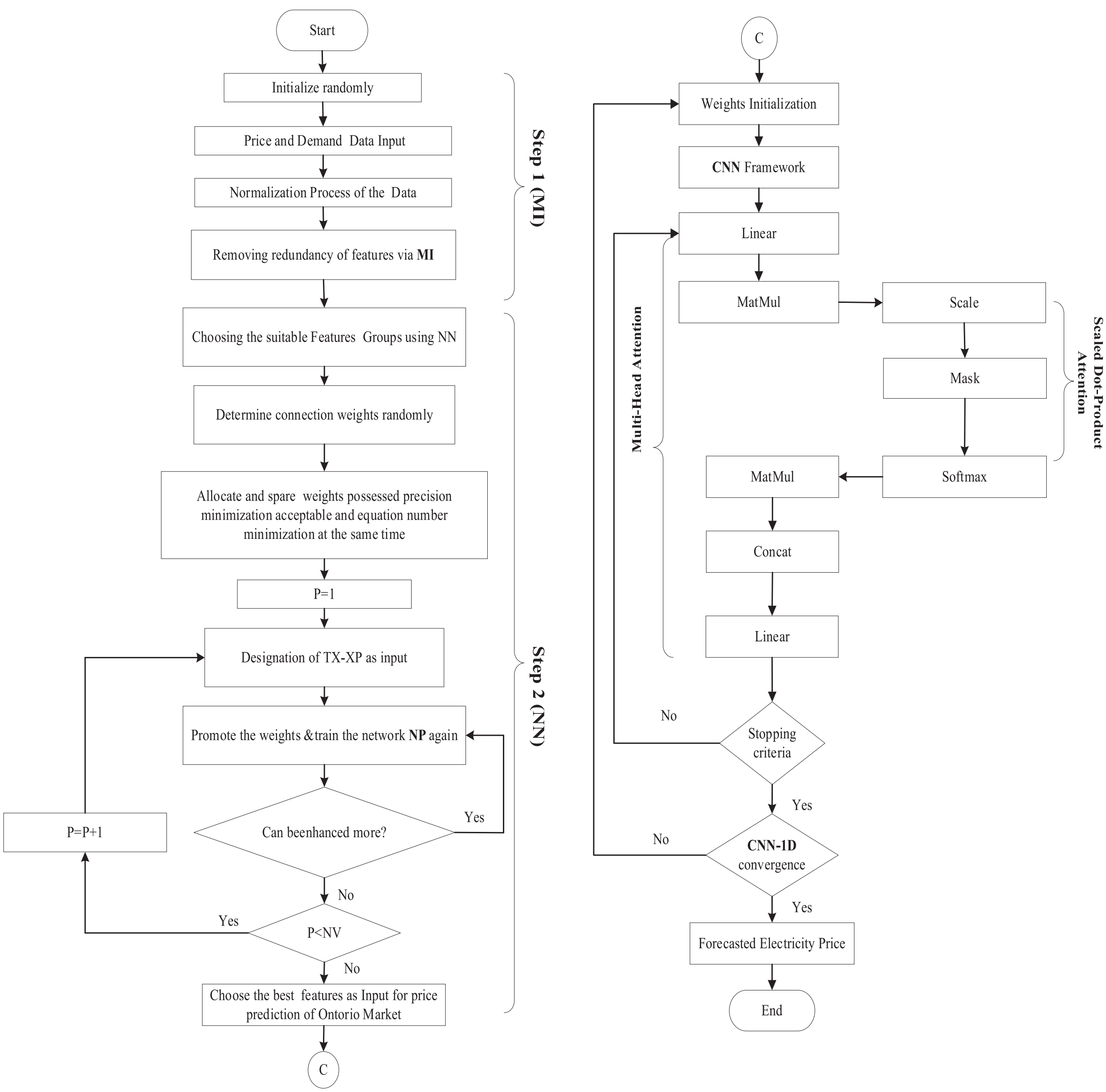

Fig. 8. The comprehensive outlook of the development of feature selection method along with forecasting technique for price prediction.

Table 5
User-defined parameter for applied methods: all models trained for 100 epochs with early stopping using Adam optimizer and momentum of 0.1 and l1 regularization.

Method	Parameters
CNN (1D)	Five layers, 30 filters
LSTM	Five layers, 30 units
ANN	Five layers, 30 neurons
Attention	Two attention heads

$$MSE_{EP} = \frac{1}{N}\sum_{t=1}^{N}\left(HEP(t)_{real} - HEP(t)_{predicted}\right)^2 \tag{13}$$

$$MAE_{EP} = \frac{1}{N}\sum_{t=1}^{N}\left|HEP(t)_{real} - HEP(t)_{predicted}\right| \tag{14}$$

Due to no unanimity in the optimal values of the AI-based technique parameters, control parameters of chosen methods are usually determined based on similar methodologies that have been effectively implemented for electricity price or demand forecasts in the literature. Table 5 summarizes all parameter settings for the methods used.

Table 6
Statistical analysis between forecasting error of applied techniques for Ontario market in winter 2020.

Performance index		Techniques				
		ANN	LSTM	1D-CNN	Attention-LSTM	Attention-1D-CNN
$RMSE_{EP}$	Training	0.011082	0.010161	0.009754	0.009222	0.008214
	Testing	0.011374	0.010517	0.0101	0.009512	0.00852
MAE_{EP}	Training	0.008804	0.007916	0.007903	0.007944	0.006829
	Testing	0.009482	0.008351	0.008341	0.00854	0.007216
MSE_{EP}	Training	0.000123	0.000103	9.51E-05	8.5E-05	6.75E-05
	Testing	0.000129	0.000111	0.000102	9.05E-05	7.26E-05
$MAPE_{EP}$	Training	2.337524	2.143411	2.133759	1.901571	1.701711
	Testing	2.567859	2.344382	2.255633	2.034482	1.796719

Table 7
Statistical analysis between forecasting error of applied techniques for Ontario market in spring 2020.

Performance index		Techniques				
		ANN	LSTM	1D-CNN	Attention-LSTM	Attention-1D-CNN
$RMSE_{EP}$	Training	0.010875	0.009965	0.009761	0.009401	0.008479
	Testing	0.011269	0.010458	0.010101	0.009676	0.008708
MAE_{EP}	Training	0.008663	0.007511	0.007537	0.007606	0.006605
	Testing	0.009488	0.008276	0.008167	0.00834	0.007287
MSE_{EP}	Training	0.000118	9.93E-05	9.53E-05	8.84E-05	7.19E-05
	Testing	0.000127	0.000109	0.000102	9.36E-05	7.58E-05
$MAPE_{EP}$	Training	2.397207	2.261776	2.020379	1.885354	1.578617
	Testing	2.628551%	2.395201%	2.233926%	2.032756%	1.746826

Table 8
Statistical analysis between forecasting error of applied techniques for Ontario market in summer 2020.

Performance index		Techniques				
		ANN	LSTM	1D-CNN	Attention-LSTM	Attention-1D-CNN
$RMSE_{EP}$	Training	0.01083	0.009916	0.009714	0.009315	0.008264
	Testing	0.011407	0.010446	0.010104	0.009617	0.008519
MAE_{EP}	Training	0.008803	0.007651	0.007331	0.008004	0.006735
	Testing	0.009424	0.008493	0.008102	0.008433	0.007352
MSE_{EP}	Training	0.007352	9.83E-05	9.44E-05	8.68E-05	6.83E-05
	Testing	0.00013	0.000109	0.000102	9.25E-05	7.26E-05
$MAPE_{EP}$	Training	2.467448	2.259005	2.147628	1.861794	1.62315
	Testing	2.626638	2.423371	2.277373	2.026645	1.759079

Table 9
Statistical analysis between forecasting error of applied techniques for Ontario market Ontario in fall 2020.

Performance index		Techniques				
		ANN	LSTM	1D-CNN	Attention-LSTM	Attention-1D-CNN
$RMSE_{EP}$	Training	0.010825	0.010246	0.009554	0.009351	0.008339
	Testing	0.011402	0.010601	0.010043	0.009735	0.008642
MAE_{EP}	Training	0.008224	0.007825	0.007639	0.007715	0.006831
	Testing	0.009076	0.008372	0.008487	0.008164	0.007359
MSE_{EP}	Training	0.000117	0.000105	9.13E-05	8.74E-05	6.95E-05
	Testing	0.00013	0.000112	0.000101	9.48E-05	7.47E-05
$MAPE_{EP}$	Training	2.374374	2.201555	2.044104	1.837457	1.548672
	Testing	2.545843	2.35098	2.24329	1.980703	1.720309

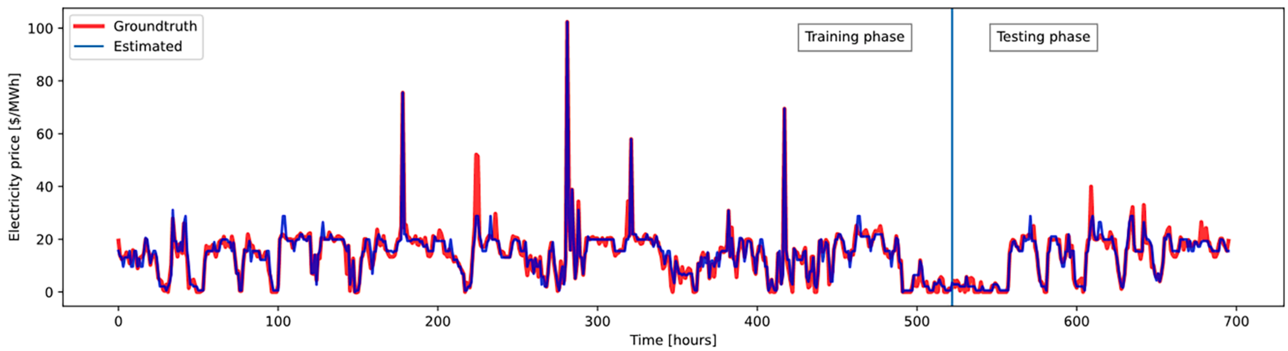

Fig. 9. The performance of multi-head attention-CNN during winter season 2020: actual prices, red line, together with the forecasted prices, blue line, in dollar per megawatt-hour.

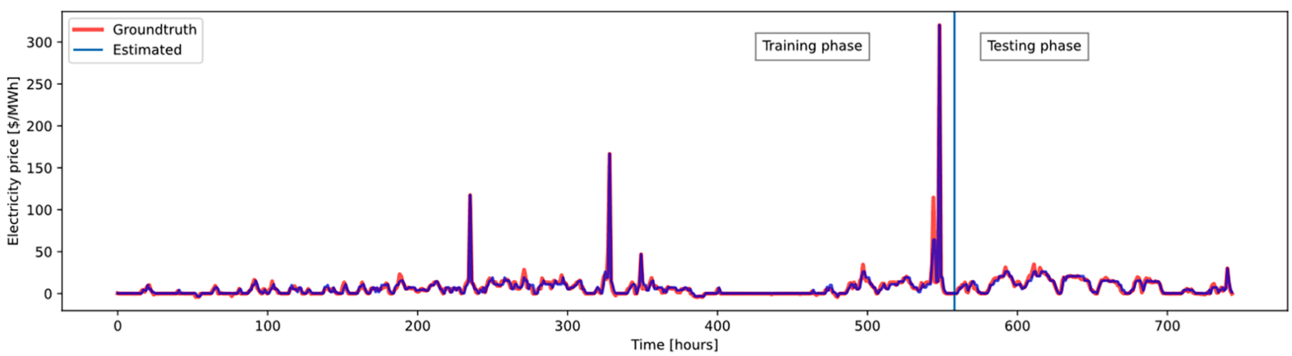

Fig. 10. The performance of multi-head attention-CNN during spring season 2020: actual prices, red line, together with the forecasted prices, blue line, in dollar per megawatt-hour.

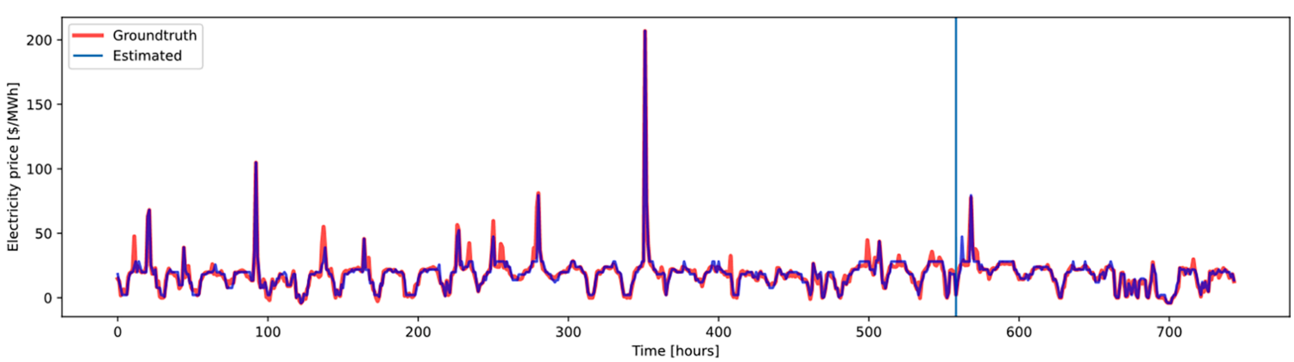

Fig. 11. The performance of multi-head attention-CNN during summer season 2020: actual prices, red line, together with the forecasted prices, blue line, in dollar per megawatt-hour.

For EPF of Ontario in winter, spring, summer, and fall in 2020, the deep learning methods are detailed in Tables 6, 7, 8, and 9, respectively. Based on these data, the forecasting methodologies for the Ontario electricity market according to various criteria that were used for judgments are listed as follows: Attention-CNN1D > Attention-LSTM > CNN1D > LSTM > ANN. It can be shown in Table 6-9; that the proposed approach outperforms the rest of the research methodologies.

4.2. Results and discussion

Furthermore, multi-head attention-CNN was determined to be the most effective algorithm for forecasting the electricity price of Ontario mainland, whereby the proposed technique achieved the optimal $MAPE_{EP}$ in percentage, MAE_{EP}, $RMSE_{EP}$, and $MSE_{EP,}$ which are significantly smaller than the values obtained by other algorithms for the whole seasons plainly. Figs. 9-12 demonstrate the performance of modeled Attention-CNN1D with two input data sets pictorially where it shows the well trend of EP forecasting data to the corresponding actual data in different months of each season. Each figure shows the actual prices, red line, together with the forecasted prices, blue line. As shown from figures, during working hour, electricity demand slightly changes, which affects electricity price but during pick hour, electricity demand gradually change and providing electricity price for such demand is a challenging task. Therefore, there is high competition for electricity prices when electricity demand is so high, and the generation is limited. In other words, with the inherent correlation between electricity price and demand, prediction in a smart grid environment such as the Ontario electricity market is more complex than the conventional power systems, and the proposed method is in this market to provide highly accurate forecasts. Besides, the presence of renewable energy is reduced

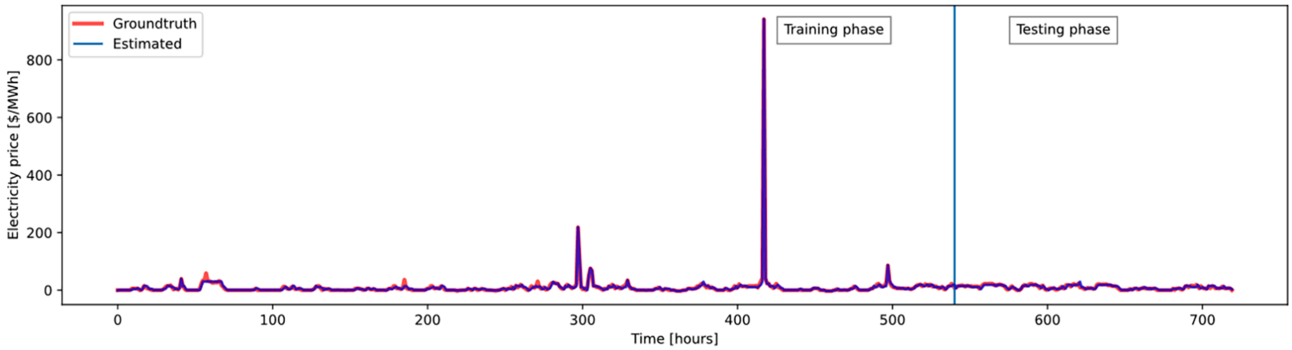

Fig. 12. The performance of multi-head attention-CNN during fall season 2020: actual prices, red line, together with the forecasted prices, blue line, in dollar per megawatt-hour.

due to environmental or unpredicted conditions, which will influence electricity of generation and electricity prices. From the standpoint of demand management, the negative price may be regulated by giving an incentive for utilizing power at a specific time to cut consumption correspondingly when demand is low to stabilize the frequent power market change with the power station to create new business opportunities.

4.2.1. The sensitivity analysis

The sensitivity analysis assigns a "sensitivity index" to every model input. There are numerous methods to calculate these indices. These indices can be determined if the model is known mathematically and is straightforward. In most instances, however, the paradigm is more complex. These sensitivity indicators can be numerically determined through statistical techniques by manipulating the model's inputs and evaluating their effects on the model's output. Regarding sensitivity analysis, many papers have been published (Trabelsi, Abid, Taktak, Fakhfakh, & Haddar, 2017). Based on the available literature, these analyses are founded on three methods: screening, local, and global sensitivity. These later subsections investigate the impact of the variability of inputs on the output by identifying which portion of output variance is attributable to a given entry and set of inputs.

4.2.1.1. The sensitivity indices. Numerous sensitivity measures exist, including the Partial Correlation Coefficient (PCC), Standardized Regression Coefficient (SRC), and Kendall's tau coefficient (KTC). These indices are provided in the following section.

1- Partial Correlation Coefficient (PCC)

A third variable may explain the relationship between Y and X. The Partial Correlation Coefficient (PCC) index is designed to eradicate the influence of other variables. The calculation is as follows:

$$PCC = \left(\frac{\mathrm{cov}(Y, X_i | X_{\setminus i})}{\sqrt{\mathrm{V}(Y|X_{\setminus i}).V(X_i|X_{\setminus i})}} \right) \tag{15}$$

where $X_{\setminus i}$ is the X vector withour its i^{th} componet. This coefficient measures the intensity of any possible association among variables. It always falls between -1 and 1. This provides an indication of the relationship's strength: A PCC value near to 1 means X can account for the majority of Y's uncertainty. On the contrary, a PCC value near 0 shows a linear relationship between the input and output and explains little of the output's uncertainty (Bouazizi, et al., 2019).

2- Standardized Regression Coefficient (SRC).

Using a linear regression model, the Standardized Regression Coefficient (SRC) provides the intensity of the correlation between an output

Table 10
The sensitivity analysis for the proposed models across different seasons.

Sensitivity Indices	ANN	LSTM	1D-CNN	Attention-LSTM	Attention-1D-CNN
Winter					
PCC	0.955	0.944	0.959	0.965	**0.978**
SRC	0.953	0.919	0.93	0.952	**0.958**
KTC	0.821	0.752	0.773	0.82	**0.842**
Spring					
PCC	0.943	0.969	0.98	0.979	**0.982**
SRC	0.922	0.946	0.948	**0.957**	0.953
KTC	0.777	0.81	0.818	0.829	**0.833**
Summer					
PCC	0.949	0.959	0.97	0.977	**0.986**
SRC	0.898	0.951	0.959	0.973	**0.987**
KTC	0.73	0.829	0.837	0.873	**0.911**
Fall					
PCC	0.966	0.934	0.92	0.960	**0.967**
SRC	0.932	0.934	0.94	0.947	**0.954**
KTC	0.789	0.789	0.8	0.806	**0.828**

Y and a given input X_i The R^2 of the linear model has a significant impact on the dependability of the SRC (R^2 is the correlation coefficient) (Yang et al., 2016). The linear model is represented as follows:

$$Y = \beta_o + \sum_{i=1}^{K} \beta_i X_i \tag{16}$$

The variables X_i are assumed to be independent, the varance of Y is:

$$V(Y) = \sum_{i=1}^{K} \beta_i{}^2 V(X_i) \tag{17}$$

The SRC sensitivity index of the variable X_i can be determined by the following equation:

$$SRC_i = \frac{\beta_i{}^2 V(X_i)}{V(Y)} \tag{18}$$

Where the $\beta_i{}^2 V(X_i)$ refer to the impact of the variable X_i in the variance of Y. Therefore, the SRC index states the variance of Y reaction due to variable X_i.

3- Kendall's tau coefficient (KTC)

The Kendall's coefficient measures the numerical relationship among two measured quantities. Depending on the coefficient, the test is a non-parametric hypothesis test for statistical dependence. It measures rank correlation, which is the similarity between the data orderings when ranked by each quantity.

4.2.1.2. The sensitivity results for the proposed models. To begin, the

Table 11
The results of the Friedman rank and Iman-Davenport tests for each model.

Model	Friedman Rank	ImanDavenport p-value
ANN	400	0.003
LSTM	169	
1D-CNN	100	
Attention-LSTM	169	
Attention-1D-CNN	16	

Table 12
The comparative results of all models compared to the Attention-1D-CNN for Friedman post hoc test.

Comparison	Post hoc p-value
Attention-1D-CNN VS ANN	1.04638E-06
Attention-1D-CNNVS LSTM	1.33514E-06
Attention-1D-CNN VS 1D-CNN	4.59142E-05
Attention-1D-CNN VS Attention-LSTM	0.000173029198624161

results will be elaborated on one season (Winter) since the other seasons follow the same pattern. Table 10 shows the results of the sensitivity analysis using different sensitivity indices. For instance, in the case of the winter season, The PPC calculates the linear correlation between two variables. It ranges from −1 to 1, where −1 represents an ideal negative linear relationship, 1 represents a perfect positive linear relationship, and 0 represents the absence of any linear relationship. In this instance, PCC is 0.9788, indicating a very robust positive linear relationship between the actual and predicted values. This implies that theAttention-1D-CNN model's predictions are quite precise and that the model is performing admirably. However, the SRC rank correlation coefficient measures the monotonic relationship between two variables without presuming a linear relationship. It also ranges from −1 to 1, with −1 representing a perfect negative monotonic relationship, 1 representing a perfect positive monotonic relationship, and 0 representing the absence of a monotonic relationship. In this case, the SRC coefficient is 0.9588, indicating a robust positive monotonic relationship between the ground truth and predicted values. This indicates that as one variable increases, so does the other, and the model's performance in this regard is satisfactory. Finally, KTC is another rank-based correlation measure that contrasts the number of concordant to discordant pairs in the data. As with the other coefficients, its range is between −1 and 1, where −1 indicates a perfect negative association, 1 indicates a perfect positive association, and 0 indicates no association. In the current case, the KTC for the Attention-1D-CNN model is 0.8425, implying a strong positive correlation between the actual and estimated values. This further demonstrates that the proposed model's predictions are accurate and that it is performing excellently. In conclusion, all three coefficients indicate that the model's predictions closely match the actual values. Since the coefficients are near 1, it suggests that the Attention-1D-CNN model predicts the target variable accurately.

4.2.2. Statistical analysis using the Friedman test

Two-step statistical tests were conducted for method validation to show the considerable difference between the proposed Attention-1D-CNN model and the other single and hybrid DL models across multiple seasons. In compliance with (Demšar, 2006) suggestions, a comprehensive test utilizing the Friedman rank is administered first. If variations in regressor performance are identified, the Friedman post hoc test is conducted. The Friedman test then investigates the hierarchy of benchmarked models, while the Iman-Davenport test determines whether or not at least one model has a significant advantage over others. Then, for multiple comparisons, a pair-wise test employing Friedman post hoc with the corresponding p-value is conducted following the detection of the difference.

The Friedman post hoc test takes into consideration a comparison with the reference (Attention-1D-CNN). The Attention-1D-CNN is chosen as a comparison standard for other models. A p-value less than the 0.05 threshold can assess a variance's significance level. The average rank and p-value of Friedman's Iman-Davenport test are shown in Table 11. It should be noted that a model's ranking falls as its excellence rises. According to the data in Table 11, the Attention-1D-CNN technique is the best model because it has the lowest rank. The p-value is 0.003, indicating a statistically significant difference (p-value less than 0.05) between at least two benchmarked approaches, allowing us to reject the null hypothesis that all models have comparable performance. In addition, following the rejection of the null hypothesis, the Friedman post hoc test is used to compare the performance of each pair. Table 12 displays the outcomes of a statistical comparison of the pairs. Amazingly, the performance variations between the proposed model and current models are statistically significant (p-value 0.05) for all existing

Table 13
A comparison table for electricity price forecasting.

Reference	Year	Model	Dataset	MAPE	RMSE	MAE	MSE
Proposed	Present	Attention-1D-CNN	Ontario	1.72%	0.008214	0.0072	0.0001
(Jahangir, et al., 2019)	2020	Dimension reduction strategy and rough ANN		5.64%	1.17	1.59	1.37
(Pourdaryaei, Mokhlis, Illias, Kaboli, & Ahmad, 2019)	2019	ANFIS-BSA		0.87%	0.02	–	0.0004
(Pourdaryaei, Mokhlis, Illias, Kaboli, & Ahmad, 2019)		Hybrid ANN-artificial cooperative search algorithm		1.2%	0.02	–	0.0004
(Razak, et al., 2016)	2016	SVM + GA		9.22%	–	–	–
(Shrivastava & Panigrahi, 2014)	2014	WELM		4.68%	3.04	2.29	9.24
(Macaš & Lhotská, 2013)	2013	FTD NN + Wrapper model		12.74%	–	–	–
(Sharma & Srinivasan, 2013)		Recurrent NN + excitable dynamics		12.47%	–	–	–
(Shayeghi & Ghasemi, 2013)		LSSVM + CGSA		15%	–	–	–
(Amjady, Daraeepour, & Keynia, 2010)	2010	Numerical sensitivity analysis + HNN		11.97%	–	–	–
		Correlation analysis + HNN		11.16%	–	–	–
		Original Relief + HNN		9.92%	–	–	–
		Modified Relief + Hybrid NN (HNN)		9.23%	–	–	–
	Different Dataset						
(Alkawaz, et al., 2022)	2022	ARD + ETR	Nord pool	–	3.09	2.03	9.55
(Zhang, et al., 2022)		SSA-DELM		–	4.7	3.8	22.1
(Sun, Li, Liu, & Saeedi, 2021)	2021	SDR-MASES-SPSDAE	Australia	–	4.2	4.98	17.6
(Zhang, Tan, & Wei, 2020)	2020	VMD-DBN	Spanish, Australia, and PJM	–	3.28	0.18	10.76

Table 14
Statistical factors of the proposed technique for EPF of Ontario in various seasons 2020.

Steps	Formulation	Criteria	Winter	Spring	Summer	Autumn
1	R	$0.8 < R_O$	0.9991	0.9967	0.9955	0.9994
2	$K = \frac{\sum_{i=1}^{n}(h_i \times t_i)}{\sum_{i=1}^{n} h_i^2}$	$0.85 < k < 1.15$	0.9987	0.9981	0.9949	0.9983
3	$K' = \frac{\sum_{i=1}^{n}(h_i \times t_i)}{\sum_{i=1}^{n} t_i^2}$	$0.85 < k' < 1.15$	1.0010	1.0004	1.0047	1.0015
4	$m = \frac{R^2 - R_o^2}{R^2}$	$\lvert m \rvert < 0.1$	−0.0013	−0.0055	−0.0066	−0.0017
5	$n = \frac{R^2 - R_O^2}{R^2}$	$\lvert n \rvert < 0.1$	−0.0014	−0.0065	−0.0067	−0.0014
6	$R_m = R^2 \times \left(1 - \sqrt{\lvert R^2 - R_o^2 \rvert}\right)$	$0.5 < R_m$	0.9960	0.9920	0.9771	0.9987
Where	$R_o^2 = 1 - \frac{\sum_{i=1}^{n}(t_i - h_i^o)^2}{\sum_{i=1}^{n}(t_i - \overline{t_i})^2}, h_i^o \times t_i$	$0.8 < R_0^2 < 1$	1.0000	1.0000	0.9985	0.9988
	$R_{o'}^2 = 1 - \frac{\sum_{i=1}^{n}(h_i - t_i^o)^2}{\sum_{i=1}^{n}(h_i - \overline{h_i})^2}, t_i^o$	$0.8 < R_{o'}^2 < 1$	1.0000	1.0000	0.9988	0.9989

models. Consequently, these results demonstrated that the Attention-1D-CNN model is more capable of learning than the other models.

4.2.3. Comparative analysis

The most recently developed techniques in the literature that have been applied for electricity price forecasting are summarized in Table 13. The results show that our model achieved better performance compared to other models that have been used on different datasets. Based on Table 13, the MAPE values were selected according to the lowest value from the four different seasons. However, the MAPE values of the proposed model from the training set were 1.796719, 1.746826, 1.759079, and 1.720309 for the Winter, Spring, Summer, and Fall, respectively, with an average value of 1.75% which is less than the average values of the previous papers on Ontario dataset. For instance, in this reference (Pourdaryaei, Mokhlis, Illias, Kaboli, & Ahmad, 2019), the MAPE values were 2.79%, 0.87%, 1.7%, and 3.17% in February, May, August, and November, respectively, with an average of 2.1325%. Further, the MAPE values in reference (Pourdaryaei, Mokhlis, Illias, Kaboli, & Ahmad, 2019) were 4.58%, 1.2%, 2.62%, and 3.79 % in winter, spring, summer, and autumn, respectively with an average of 3.0475% which indicate the proposed model achieved lowest average MAPE from the four seasons compared to the previous work. In addition, the RMSE results of the proposed work for the all seasons almost perform the same where the values were 0.00852, 0.008708, 0.008519, and 0.008519 for the Winter, Spring, Summer, and Fall, respectively, with an average value of 0.00859 which indicate that the model forecast is stable regardless on the changes in the seasons in contrast to the other work in the literature which varies depending on the season.

4.2.4. Validating mathematical models by developing statistical factors as an external verification

A thorough analysis using another statistical technique was conducted to demonstrate the effectiveness of the proposed model. The following criteria were employed to assess the performance of the model by comparing the predicted and actual values (Mousavi, et al., 2014; Roy & Roy, 2008).

When analyzing data, if the absolute value of the correlation coefficient $|R|$ is above 0.8, it suggests a strong correlation exists within the model. If $|R|$ falls between 0.2 and 0.8, there is still a correlation present. However, if $|R|$ is below 0.2, it indicates a weak correlation or no correlation at all within the model. R has been computed as presented below.

$$R = \frac{\left(\sum_{i=1}^{n}(h_i - \overline{h_i})(t_i - \overline{t_i})\right)}{\sqrt{\sum_{i=1}^{n}(h_i - \overline{h_i})^2 \sum_{i=1}^{n}(t_i - \overline{t_i})^2}} \tag{19}$$

In this equation, h_i and t_i represent the actual and predicted output values, respectively. $\overline{h_i}$ is the average of actual output, and $\overline{t_i}$ is the average of predicted output. The relationship between actual and predicted output is defined by the R equation, where$h_i = k\, t_i$ and$t_i = k'h_i$ in which k and k' are the slopes of the regressions.

Table 14 displays the statistical factors of the proposed model for Ontario during various seasons of 2020. As shown in the table, the developed models meet all necessary criteria. The validation stage confirms that the combination of multi-head self-attention and CNN-based techniques is a promising and optimistic approach for predicting future electricity prices in any competitive and deregulated electricity market.

In order to explain the practical significance of our electricity price forecasting study, we have summarized our findings in the conclusion section. This includes discussing potential applications of our model in the industry, its benefits, and acknowledging its limitations. Our study can help electricity producers optimize power plant operations and reduce greenhouse gas emissions by providing a better understanding of price fluctuations. Accurate forecasts are also necessary for effective demand-side management, which can reduce uncertainties related to pricing dynamics. Policymakers can use our projections to make well-informed decisions about energy consumption and develop sustainable energy policies for the future.

5. Conclusions

This study developed and validated an interpretable multi-head self-attention and CNN-based framework for the deregulated and competitive electricity market with hourly electricity demand and real price data from the Ontario region in 2020.

- In machine learning algorithms, performing suitable feature selection is a key component of the process. To obtain high predicted accuracy, selecting a suitable subset of features from a pool of 67 features is necessary, as choosing substantial data attributes from a large pool of irrelevant and redundant properties is necessary. This paper proposes a hybrid feature selection methodology integrating ANN and mutual information techniques for selecting the best subset of features to be used as input for a direct prediction method. The proposed technique's robustness is demonstrated by efficiently selecting the most appropriate features by removing irrelevant and superfluous qualities.
- Compared to other computational intelligence models, the proposed multi-head self-attention and CNN forecasting approach offered improved accuracy in forecasting electricity prices with the lowest intricacy, achieving MAPE values of 1.8 percent, 1.6 percent, 1.9 percent, and 1.5 percent in winter, spring, summer, and fall, respectively.

Table A1
Table of abbreviations.

$\overline{Z_n}$	Normalized data	NV	Number of features
Z_n	Un normalized data	t	Hourly interval
H	Hidden layer	H(X)	Entropy of random variable X
TH	Threshold	P(X)	Probability function of X
t	The network output	N	Observation number
PF	Penalty function	H(X,Y)	Joint probability distribution of variable X and Y
y	The real value	$H(X\|Y)$	Conditional entropy
$\alpha_{PF,1}, \alpha_{PF,2} and \beta_{PF}$	Coefficient of penalty function	PTX_{tr}	Precision of training
$HEP\ (t)_{real}$	Real electricity price at time t	PTX_{ts}	Precision of testing
$HEP(t)_{predicted}$	Predicted electricity price at time t	$\$/MW.h$	dollar per Megawatts hour
MAE_{EP}	Mean absolute error	MSE_{EP}	Mean square error
w_H^O	The weight connecting from H-th hidden unit to network output	w_{NV}^H	The weight connecting from NV-th features to H-th hidden unit
NN	Neural network	CNEA	Cascaded neuro-evolutionary algorithm
ReLU	Rectified linear unit	MI	Mutual Information
CNN	Convolutional neural network	ANN	Artificial neural network
ANFIS	Adaptive neuro-fuzzy inference system	ARIMA	Auto regressive integrated moving average
LSTM	Long short-term memory	GA	Genetic algorithm
DSM	Demand-side management	SeLU	Scale exponential linear unit
HED(t)	Electricity demand at time t	MW	Megawatts
HEP(t)	Electricity price at time t	EPF	Electricity Price Forecasting
PSO	Particle swarm optimization	$RMSE_{EP}$	Root mean square error
CE	Conditional entropy	SVM	Support vector machine
SSFS	Stage selection feature system	NL_{HED}	Number of lag order for electricity demand
NL_{HEP}	Number of lag order for electricity price	MOPSO	Multi-objective particle swarm optimization

- Based on the findings, the proposed technique to EP forecasting is unavoidable in the development of a sustainable smart grid in the coming days. As a result, the presented synthesis might be a powerful scheme to design energy plans for the electricity participants in bidding as well as for the academics who are studying EP forecasting. In comparison to well-known based models, the results show that the proposed method is more accurate and reliable. Based on the positive outcomes, it can be deduced that recent advances in AI-based methodologies, such as those used in this study, could lead to improved accuracy for energy price forecasting with the least amount of complexity. Self-producers, traditional generation businesses, suppliers/retailers, and aggregators could all benefit from using the proposed method, as its results show that it is a strong and useful forecast engine to their actual needs. Market participants may benefit from these contributions, which could lead to higher bids, more efficient operations, and more profits for the organization as a whole.

The limitation of this work can be highlighted by considering only one dataset (Ontario 2020 for four seasons) to validate the proposed work, which limits the generalizability of our findings to other electricity market regions and periods. In order to address the generalizability issue, the extension of this work will concentrate on testing the proposed method on different datasets for different regions and deploying the model to an end-to-end system. Hence, this deployment will evaluate the proposed method's effectiveness in various contexts and allow continuous improvement and refinement based on real-time data. The implications of this study can be summed up as the accurate electricity price can aid the electricity industry's operators in administering effective energy management, efficient resource allocation, and informed decision-making.

CRediT authorship contribution statement

Alireza Pourdaryaei: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing. **Mohammad Mohammadi:** Methodology, Investigation, Writing – review & editing, Supervision. **Hamza Mubarak:** Conceptualization, Software, Validation, Writing – original draft. **Abdallah Abdellatif:** Software, Formal analysis. **Mazaher Karimi:** Conceptualization, Validation, Investigation, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Elena Gryazina:** Visualization. **Vladimir Terzija:** Writing – original draft, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

This research has been supported by the Ditartis project, "Network of Excellence in Digital Technologies and AI Solutions for Electromechanical and Power Systems Applications" with Grant Agreement number: 101079242 – HORIZON-WIDERA-2021-ACCESS-03. The work of E. Gryazina and V. Terzija was carried out as a part of the AMPaC Megagrant project supported by the Ministry of Science and Higher Education of the Russian Federation, Grant Agreement No 075-10-2021-067, Grant identification code 000000S707521QJX0002.

Appendix

Table A1

References

(IESO), T. I. E. S. O. (2018-11-17). http://www.ieso.ca/.
Abdellatif, A., Mubarak, H., Ahmad, S., Ahmed, T., Shafiullah, G., Hammoudeh, A., … Gheni, H. M. (2022). Forecasting photovoltaic power generation with a stacking ensemble model. *Sustainability, 14*, 11083.
Abdellatif, A., Mubarak, H., Ahmad, S., Mekhilef, S., Abdellatef, H., Mokhlis, H., & Kanesan, J. (2023). In *Electricity Price Forecasting One Day Ahead by Employing Hybrid Deep Learning Model* (pp. 1–5). IEEE.
Abedinia, O., Amjady, N., Shafie-Khah, M., & Catalão, J. P. (2015). Electricity price forecast using combinatorial neural network trained by a new stochastic search method. *Energy Conversion and Management, 105*, 642–654.
Abo-Tabik, M., Costen, N., Darby, J., & Benn, Y. (2020). Towards a smart smoking cessation app: A 1D-CNN model predicting smoking events. *Sensors, 20*, 1099.
Aggarwal, S. K., Saini, L. M., & Kumar, A. (2008). Electricity price forecasting in Ontario electricity market using wavelet transform in artificial neural network based model. *International Journal of Control, Automation, and Systems, 6*, 639–650.

Aggarwal, S. K., Saini, L. M., & Kumar, A. (2009). Electricity price forecasting in deregulated markets: A review and evaluation. *International Journal of Electrical Power & Energy Systems, 31*, 13–22.

Alamaniotis, M., Bargiotas, D., Bourbakis, N. G., & Tsoukalas, L. H. (2015). Genetic optimal regression of relevance vector machines for electricity pricing signal forecasting in smart grids. *IEEE Transactions on Smart Grid, 6*, 2997–3005.

Alkawaz, A. N., Abdellatif, A., Kanesan, J., Khairuddin, A. S. M., & Gheni, H. M. (2022). Day-Ahead Electricity Price Forecasting Based on Hybrid Regression Model. *IEEE Access, 10*, 108021–108033.

Amjady, N., & Daraeepour, A. (2009a). Design of input vector for day-ahead price forecasting of electricity markets. *Expert Systems with Applications, 36*, 12281–12294.

Amjady, N., & Daraeepour, A. (2009b). Mixed price and load forecasting of electricity markets by a new iterative prediction method. *Electric Power Systems Research, 79*, 1329–1336.

Amjady, N., Daraeepour, A., & Keynia, F. (2010). Day-ahead electricity price forecasting by modified relief algorithm and hybrid neural network. *IET Generation, Transmission & Distribution, 4*, 432–444.

Amjady, N., & Hemmati, M. (2006). Energy price forecasting-problems and proposals for such predictions. *IEEE Power and Energy Magazine, 4*, 20–29.

Amjady, N., & Keynia, F. (2008). Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm. *IEEE Transactions on Power Systems, 24*, 306–318.

Amjady, N., & Keynia, F. (2009). Day-ahead price forecasting of electricity markets by a new feature selection algorithm and cascaded neural network technique. *Energy Conversion and Management, 50*, 2976–2982.

Bai, T., Pang, Y., Wang, J., Han, K., Luo, J., Wang, H., … Zhang, H. (2020). An Optimized faster R-CNN method based on drnet and roi align for building detection in remote sensing images. *Remote Sensing, 12*, 762.

Bouazizi, L., Trabelsi, H., Othmani, C., Taktak, M., Chaabane, M., & Haddar, M. (2019). Uncertainty and sensitivity analysis of porous materials acoustic behavior. *Applied Acoustics, 144*, 64–70.

Catalão, J., Mariano, S., Mendes, V., & Ferreira, L. (2007). An artificial neural network approach for short-term electricity prices forecasting. In *2007 International Conference on Intelligent Systems Applications to Power Systems* (pp. 1-6): IEEE.

Catalão, J., Pousinho, H., & Mendes, V. (2009). Neural networks and wavelet transform for short-term electricity prices forecasting. In *2009 15th International Conference on Intelligent System Applications to Power Systems* (pp. 1-5): IEEE.

Catalão, J. P.d. S., Mariano, S. J. P. S., Mendes, V., & Ferreira, L. (2007). Short-term electricity prices forecasting in a competitive market: A neural network approach. *Electric Power Systems Research, 77*, 1297–1304.

Catalão, J. P.d. S., Pousinho, H. M. I., & Mendes, V. M. F. (2010). Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting. *IEEE Transactions on Power Systems, 26*, 137–144.

Cerjan, M., Krželj, I., Vidak, M., & Delimar, M. (2013). A literature review with statistical analysis of electricity price forecasting methods. In *Eurocon 2013* (pp. 756–763). IEEE.

Chang, Z., Zhang, Y., & Chen, W. (2019). Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform. *Energy, 187*, Article 115804.

Civicioglu, P. (2013). Backtracking search optimization algorithm for numerical optimization problems. *Applied Mathematics and Computation, 219*, 8121–8144.

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd edition). NJ: Willey-Interscience.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine learning research, 7*, 1–30.

Diongue, A. K., Guegan, D., & Vignal, B. (2009). Forecasting electricity spot market prices with a k-factor GIGARCH process. *Applied Energy, 86*, 505–510.

Hahn, H., Meyer-Nieberg, S., & Pickl, S. (2009). Electric load forecasting methods: Tools for decision making. *European Journal of Operational Research, 199*, 902–907.

Heidarpanah, M., Hooshyaripor, F., & Fazeli, M. (2022). Daily electricity price forecasting using artificial intelligence models in the Iranian electricity market. *Energy, 126011*.

Huang, C. J., Shen, Y., Chen, Y. H., & Chen, H. C. (2021). A novel hybrid deep neural network model for short-term electricity price forecasting. *International Journal of Energy Research, 45*, 2511–2532.

Huang, Y., & Du, J. (2019). Self-attention enhanced CNNs and collaborative curriculum learning for distantly supervised relation extraction. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)* (pp. 389-398).

Jahangir, H., Tayarani, H., Baghali, S., Ahmadian, A., Elkamel, A., Golkar, M. A., & Castilla, M. (2019). A novel electricity price forecasting approach based on dimension reduction strategy and rough artificial neural networks. *IEEE Transactions on Industrial Informatics, 16*, 2369–2381.

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence, 97*, 273–324.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60*, 84–90.

Kuo, P.-H., & Huang, C.-J. (2018). An electricity price forecasting model by hybrid structured deep neural networks. *Sustainability, 10*, 1280.

Lehna, M., Scheller, F., & Herwartz, H. (2022). Forecasting day-ahead electricity prices: A comparison of time series and neural network models taking external regressors into account. *Energy Economics, 106*, Article 105742.

Macaš, M., & Lhotská, L. (2013). Wrapper feature selection significantly improves nonlinear prediction of electricity spot prices. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on* (pp. 1171–1174). IEEE.

Mohseni, S., Brent, A. C., Kelly, S., & Browne, W. N. (2022). Demand response-integrated investment and operational planning of renewable and sustainable energy systems considering forecast uncertainties: A systematic review. *Renewable and Sustainable Energy Reviews, 158*, Article 112095.

Mousavi, S. M., Mostafavi, E. S., & Hosseinpour, F. (2014). Gene expression programming as a basis for new generation of electricity demand prediction models. *Computers & Industrial Engineering, 74*, 120–128.

Mubarak, H., Ahmad, S., Hossain, A. A., Horan, B., Abdellatif, A., Mekhilef, S., … Kanesan, J. (2023). Short-term Electricity Price Forecasting Using Interpretable Hybrid Machine Learning Models. In *2023 IEEE IAS Global Conference on Renewable Energy and Hydrogen Technologies (GlobConHT)* (pp. 1–6). IEEE.

Mubarak, H., Hammoudeh, A., Ahmad, S., Abdellatif, A., Mekhilef, S., Mokhlis, H., & Dupont, S. (2023). A hybrid machine learning method with explicit time encoding for improved Malaysian photovoltaic power prediction. *Journal of Cleaner Production, 382*, Article 134979.

Mubarak, H., Mansor, N. N., Mokhlis, H., Mohamad, M., Mohamad, H., Muhammad, M. A., … Afzal, S. (2021). Optimum distribution system expansion planning incorporating DG based on N-1 criterion for sustainable system. *Sustainability, 13*, 6708.

Mubarak, H., Mokhlis, H., Mansor, N. N., Mohamad, M., Khairuddin, A. S. M., & Afzal, S. (2021). Optimal Distribution Networks Expansion Planning with DG for Power Losses Reduction. In *2021 Innovations in Power and Advanced Computing Technologies (i-PACT)* (pp. 1–5). IEEE.

Mubarak, H., Muhammad, M. A., Mansor, N. N., Mokhlis, H., Ahmad, S., Ahmed, T., & Sufyan, M. (2022). Operational cost minimization of electrical distribution network during switching for sustainable operation. *Sustainability, 14*, 4196.

Panapakidis, I. P., & Dagoumas, A. S. (2016). Day-ahead electricity price forecasting via the application of artificial neural network based models. *Applied Energy, 172*, 132–151.

Pourdaryaei, A., Mohammadi, M., Karimi, M., Mokhlis, H., Illias, H. A., Kaboli, S. H. A., & Ahmad, S. (2021). Recent Development in Electricity Price Forecasting Based on Computational Intelligence Techniques in Deregulated Power Market. *Energies, 14*, 6104.

Pourdaryaei, A., Mokhlis, H., Illias, H. A., Kaboli, S. H. A., & Ahmad, S. (2019). Short-term electricity price forecasting via hybrid backtracking search algorithm and ANFIS approach. *IEEE Access, 7*, 77674–77691.

Pourdaryaei, A., Mokhlis, H., Illias, H. A., Kaboli, S. H. A., Ahmad, S., & Ang, S. P. (2019). Hybrid ANN and artificial cooperative search algorithm to forecast short-term electricity price in de-regulated electricity market. *IEEE Access, 7*, 125369–125386.

Pousinho, H. M. I., Mendes, V., & Catalão, J. P.d. S. (2012). Short-term electricity prices forecasting in a competitive market by a hybrid PSO–ANFIS approach. *International Journal of Electrical Power & Energy Systems, 39*, 29–35.

Qiao, W., & Yang, Z. (2020). Forecast the electricity price of US using a wavelet transform-based hybrid model. *Energy, 193*, Article 116704.

Razak, I. A. W. A., Abidin, I. Z., Yap, K. S., Abidin, A. A. Z., Rahman, T. K. A., & Nasir, M. N. M. (2016). A novel hybrid method of LSSVM-GA with multiple stage optimization for electricity price forecasting. In *In 2016 IEEE International Conference on Power and Energy (PECon)* (pp. 390–395). IEEE.

Renani, E. T., Elias, M. F. M., & Rahim, N. A. (2016). Using data-driven approach for wind power prediction: A comparative study. *Energy Conversion and Management, 118*, 193–203.

Roy, P. P., & Roy, K. (2008). On some aspects of variable selection for partial least squares regression models. *QSAR & Combinatorial Science, 27*(3), 302–313.

Setiono, R., & Liu, H. (1997). Neural-network feature selector. *IEEE Transactions on Neural Networks, 8*, 654–662.

Sharma, V., & Srinivasan, D. (2013). A hybrid intelligent model based on recurrent neural networks and excitable dynamics for price prediction in deregulated electricity market. *Engineering Applications of Artificial Intelligence, 26*, 1562–1574.

Shayeghi, H., & Ghasemi, A. (2013). Day-ahead electricity prices forecasting by a modified CGSA technique and hybrid WT in LSSVM based scheme. *Energy Conversion and Management, 74*, 482–491.

Shiri, A., Afshar, M., Rahimi-Kian, A., & Maham, B. (2015). Electricity price forecasting using Support Vector Machines by considering oil and natural gas price impacts. In *In 2015 IEEE International Conference on Smart Energy Grid Engineering (SEGE)* (pp. 1–5). IEEE.

Shrivastava, N. A., & Panigrahi, B. K. (2014). A hybrid wavelet-ELM based short term price forecasting for electricity markets. *International Journal of Electrical Power & Energy Systems, 55*, 41–50.

Sun, L., Li, L., Liu, B., & Saeedi, S. (2021). A novel Day-ahead electricity price forecasting using multi-modal combined integration via stacked pruning sparse denoising auto encoder. *Energy Reports, 7*, 2201–2213.

Torres, J., Martínez-Álvarez, F., & Troncoso, A. (2022). A deep LSTM network for the Spanish electricity consumption forecasting. *Neural Computing and Applications, 34*, 10533–10545.

Trabelsi, H., Abid, M., Taktak, M., Fakhfakh, T., & Haddar, M. (2017). Effect of the aerodynamic force modeling on the tonal noise prediction model for axial fan: Sensitivity and uncertainty analysis. *Applied Acoustics, 117*, 61–65.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998-6008).

Wang, J., Zhang, L., & Li, Z. (2022). Interval forecasting system for electricity load based on data pre-processing strategy and multi-objective optimization algorithm. *Applied Energy, 305*, Article 117911.

Wolak, F. A. (2022). Long-term resource adequacy in wholesale electricity markets with significant intermittent renewables. *Environmental and Energy Policy and the Economy, 3*, 155–220.

Yang, S., Tian, W., Cubi, E., Meng, Q., Liu, Y., & Wei, L. (2016). Comparison of sensitivity analysis methods in building energy assessment. *Procedia Engineering, 146*, 174–181.

Yang, Z., Ce, L., & Lian, L. (2017). Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods. *Applied Energy, 190*, 291–305.

Zhang, J., Tan, Z., & Wei, Y. (2020). An adaptive hybrid model for short term electricity price forecasting. *Applied Energy, 258*, Article 114087.

Zhang, T., Tang, Z., Wu, J., Du, X., & Chen, K. (2022). Short term electricity price forecasting using a new hybrid model based on two-layer decomposition technique and ensemble learning. *Electric Power Systems Research, 205*, Article 107762.

Zhang, Y., Tao, P., Wu, X., Yang, C., Han, G., Zhou, H., & Hu, Y. (2022). Hourly electricity price prediction for electricity market with high proportion of wind and solar power. *Energies, 15*, 1345.

Zhang, Y., Xing, K., Bai, R., Sun, D., & Meng, Z. (2020). An enhanced convolutional neural network for bearing fault diagnosis based on time–frequency image. *Measurement, 157*, Article 107667.

Zhou, S., Zhou, L., Mao, M., Tai, H.-M., & Wan, Y. (2019). An optimized heterogeneous structure LSTM network for electricity price forecasting. *IEEE Access, 7*, 108161–108173.