A Synchronous Prediction Model Based on Multi-Channel CNN with Moving Window for Coal and Electricity Consumption in Cement Calcination Process

Xin Shi, Gaolu Huang, Xiaochen Hao *, Yue Yang and Ze Li

School of Electrical Engineering, Yanshan University, 438 Hebei Avenue, Qinhuangdao 066004, China; shixin0329@163.com (X.S.); huanggaolu2018@sina.com (G.H.); 18712737027@163.com (Y.Y.); lize_1995@163.com (Z.L.)
* Correspondence: haoxiaochen@ysu.edu.cn

Abstract: The precision and reliability of the synchronous prediction of multi energy consumption indicators such as electricity and coal consumption are important for the production optimization of industrial processes (e.g., in the cement industry) due to the deficiency of the coupling relationship of the two indicators while forecasting separately. However, the time lags, coupling, and uncertainties of production variables lead to the difficulty of multi-indicator synchronous prediction. In this paper, a data driven forecast approach combining moving window and multi-channel convolutional neural networks (MWMC-CNN) was proposed to predict electricity and coal consumption synchronously, in which the moving window was designed to extract the time-varying delay feature of the time series data to overcome its impact on energy consumption prediction, and the multi-channel structure was designed to reduce the impact of the redundant parameters between weakly correlated variables of energy prediction. The experimental results implemented by the actual raw data of the cement plant demonstrate that the proposed MWMC-CNN structure has a better performance than without the combination structure of the moving window multi-channel with convolutional neural network.

Keywords: energy consumption prediction; moving window; multi-channel convolutional neural networks

1. Introduction

The cement industry is considered as an energy intensive sector [1]. The energy consumption of the entire cement manufacturing process depends largely on the electricity and coal consumption [2]. Accurate and timely energy consumption prediction is of great significance for reasonable energy scheduling, energy saving, and a reduction in production costs [3,4]. As one of the most important processes of cement production, the cement calcination process is to calcine ground raw material into cement clinker. Traditionally, electricity and coal consumption are measured mainly by sensors and weighing machines, respectively, however, the changing trend and coupling relationship of electricity and coal consumption in cement production cannot be detected, which result in the faultiness in providing guidance to energy scheduling and production optimization. Due to the lack of a coupling relationship between electricity and coal consumption, the optimal solution for the optimal control of the production process may not be in the solution space, which hinders the monitoring and optimization of the cement manufacturing process and the formulation of a process control strategy. Consequently, it is necessary to investigate a method that can realize the synchronous prediction of electricity and coal consumption.

Accurate and reliable synchronous prediction of multi energy consumption indicators is complicated by the fact that the cement calcination process consists of multiple production process variables. The relationships between cement production process variables and energy consumption indicators have the characteristics of time-varying delay, uncertainty,
and nonlinearity. For example, as the amount of raw materials increases, the electricity and coal consumption required for the calcination process increase. It takes 50–60 min, varying with the different conditions, for the cement raw material to be converted into calcined clinker. As a result, the electricity and coal consumption required for the calcination of raw materials in the rotary kiln cannot be calculated using the current amount of raw materials, and can only be manually adjusted by the operators based on historical experience. Using the current number of process variables to predict the energy consumption indexes will result in data misalignment and low accuracy. Therefore, the impact of the time-varying delay on the prediction of electricity and coal consumption of the cement calcination process needs to be considered. In production, a large amount of production process data are recorded, which makes it possible to synchronously predict electricity and coal consumption by data driven methods.

On the basis of the collection of numerous industrial data and the development of data driven methods, various data driven forecasting approaches have been proposed for prediction or diagnosis in the industrial field. For example, fault detection using statistical regression [5], electricity system load forecasting [6], and tendency prediction of the blast furnace hearth thermal state [7] using support vector machine (SVM) and electricity consumption prediction [8]. Furthermore, in the cement manufacturing industry, power consumption is predicted by the multiple non-linear regression algorithm [9] and empirical mode decomposition based hybrid ensemble model [10].

However, the above investigations all focused on a single index prediction strategy, which is insufficient for coal and electricity synchronous prediction in the cement calcination process because the coupling relationship of electricity and coal consumption cannot be obtained. The optimal control solution of the cement calcination process obtained from single index prediction model may not be in the actual solution space, which is invalid as it cannot be used to optimize the cement calcination process.

Although many multi-task process studies have been done in many fields such as the forecasting of the maximum connections in wireless communication [11], the prediction of solar radiation with multi-time scale and multi-component [12], quality monitoring of wind turbine blade icing processes [13], multi-task-oriented production layout in manufacturing factories [14], and multi-objective task scheduling in cloud computing [15], there are still obstacles caused by time-varying delay and parameter redundancy. Time-varying delay between variables is considered to be one of the characteristics of process industry. It is difficult to build accurate prediction models without considering the time delay problem. The operating conditions are constantly changing in actual cement production, which causes the time delay between the process variables and the target variables, which vary with the process time. Thus, the timing matching of process variables and target predictors loses its effect due to the time-varying delay. The parameter redundancy is another feature. Synchronous prediction models have some advantages that make them indispensable. However, in the simultaneous prediction of multi-energy indicators of the cement calcination process, there are weak correlations between some input process variables, which have impacts on the prediction accuracy when simultaneously used as input data.

According to the above analyses, a multi-indicator prediction model was established to simultaneously predict the coal and electricity consumption in the cement calcination process. On one hand, this reduces the modeling burden in the practical application of industrial systems, which contributes to the efficient use of industrial system resources. On the other hand, it provides a precise prediction of coal and electricity consumption in the cement calcination process, which contributes to the reasonable production scheduling and energy planning of the cement calcination process. In addition, the multi-output prediction model can well explain the strong coupling relationship between the prediction indicators, which is consistent with the process mechanism of the cement calcination process. Therefore, research into this issue is of great significance to promote the development of the cement industry.
The proposed model combines moving window with convolutional neural networks (CNN) to reduce the negative impact of time-varying delay on the prediction of cement energy consumption. The moving window integrates time-varying delay information hidden in time series data into the input layer of the CNN. As a result of powerful feature extraction capability, CNN is not only used in image processing [16], but also in manufacturing industry [17], which is used to extract the data characteristics of variables in the cement calcination process.

The multi-channel structure of CNN was designed to reduce the negative impact of the redundant parameters of weakly correlated variables on cement energy consumption prediction in which the single index prediction strategy was first used to predict the electricity and coal consumption separately, with the purpose of reducing the coupling relationship, and then, the coupling relationship was rebuilt before the final output in order to fuse the variables’ characteristics. The design of a multi-channel structure in this paper not only preserves the practical application advantages of multi-output models, but also reduces the negative impact of parameter redundancy on the energy consumption prediction.

The combination of a moving window, multi-channel structure with CNN contributes the two characteristics of this paper:

1. A method combining moving window and multi-channel convolutional neural networks is proposed to accommodate time-varying delay information implied in the time series data according to the analysis of the mechanism of cement calcination. The time-varying delay is a kind of unique phenomenon in process industries such as the cement industry. The variable data in the previous period are accommodated by moving window, which is more efficient for feature extraction by CNN with a one-dimensional convolution kernel. The adoption of CNN effectively reduces the excessive parameters brought by moving windows and avoids complex timing matching.

2. The multi-channel structure of MWMC-CNN was designed to predict coal and electricity consumption simultaneously, which effectively addresses the parameter redundancy problem caused by correlated variables of the cement calcination process. In addition, the adverse impact on energy consumption indicators of the synchronous prediction caused by the coupling relationship of the cement production process variables is also eliminated. Finally, the coupling relationship of electricity and coal consumption was rebuilt to provide references for the optimization of the cement calcination process.

The actual raw data of the cement plant were used to train the model and the experimental results demonstrate its superiority.

The rest of this paper is organized as follows. In Section 2, we introduce the related works about the existing literature related to data driven forecasting. Section 3 describes the technological process of a cement rotary kiln and the selection of prediction variables. The proposed MWMC-CNN is detailed in Section 4. The models with different structures were simulated, and the prediction results compared with least square support vector machine (LSSVM), CNN, and long short term memory (LSTM) in Section 5. Our conclusions and future research directions are given in Section 6.

2. Related Work

Due to the extensive usage of sensors, a large number of industrial data are recorded, which creates the basis for the optimization and operational control of the system by data driven methods. There have been many data driven prediction investigations for energy consumption based on statistical regression, SVM, tree-based method, ANN, and LSTM.

Statistical regression is a direct method for data driven, which is a predictive model technique that studies the relationship between independent variables and dependent variables, and is widely used for short-term predictions in many industries. Bianco applied linear regression to predict urban power consumption [18] and Catalina applied it to forecast the thermal energy consumption of buildings [19]. The advantage of linear regression is that the structure is simple, but its nonlinear ability is limited and it is easily interfered...
by outliers. Compared with linear regression, principal component analysis and partial least squares regression analysis have faster speed, better nonlinear ability, and a good effect in the short-term load forecasting of the power system [20]. Although the methods of statistical regression are simple and easy to implement, they are susceptible to outliers, however, while dealing with complex problems, they are also prone to overfitting. In addition, statistical regression methods cannot eliminate the negative impact of time-varying delays on prediction accuracy.

SVM is a kind of typical machine learning method that has strong nonlinear ability, and the computational complexity depends on the number of support vectors, so it still has better performance even if the sample space has a higher dimension [21]. These advantages show that it can be used to conduct energy analysis by continuous time series data in industrial fields. Based on SVM, Wang achieved the prediction of hydropower consumption [22] and Assouline estimated the solar photovoltaic potential [23]. Furthermore, some investigations were implemented to optimize SVM, where the optimal parameters of each kernel function [24] and optimization algorithms were obtained [25]. Although SVM can deal with the nonlinear regression problem and its parameter can be adjusted by the corresponding adaptive algorithm, the performance of SVM is still limited when the amount of training data is large [26]. The scale of time series data with time-varying delays recorded from the production line is usually very large, thus support vector machine cannot perform well when dealing with these data.

Tree-based methods divide the predictor space into sub-regions, and then the predicted results based on statistical indicators are obtained, which can be used not only for classification but also regression. Tree-based models are widely used to predict critical bus voltage and load in PV buses [27], the heat load of residential buildings [28], and the hourly performance forecasting of ground source heat pump systems [29]. The use of tree-based methods in industry have also been investigated by scholars such as the directed edge weight prediction of industrial Internet of Things [30] and the analysis of industrial data in cognitive manufacturing [31]. Regression trees split the training dataset into distinct and non-overlapping regions, which is inappropriate for energy consumption forecasting in cement manufacturing because of the strong coupling in cement production data.

ANN is a data-driven method of artificial intelligence that has a complex structure and is used for information processing problems that cannot be solved by theoretical analysis [32]. It is applied to industrial process analysis because of the advantages of adaptive learning, fast optimization, and strong nonlinear ability [33]. ANN has been used to forecast electricity consumption [34,35] annually or household electricity consumption daily and hourly [36] as well as day-ahead electricity price [37]. Moreover, it is also used for the prediction of Polish gas consumption [38]. Although the above studies fully utilized the advantages of artificial neural networks, appropriately selected the variables associated with the target variables, established a model for the corresponding actual situation, and achieved certain results, ANN takes a set of handcrafted features as the input and it is difficult to obtain the most useful features relevant to the estimation task [39].

LSTM is good at processing time series prediction because of its special structure, which is relevant to time factor. Thus, LSTM is applied to predict energy consumption, for example, gas consumption [40]. Besides, it is widely utilized for electricity consumption prediction such as the energy consumption of housing [41] and commercial buildings [42], medium- and long-term power forecasting [43], and electricity load forecasting in the electric power system [44]. Although all of the above investigations achieved the prediction of electricity consumption and are used as references for electric departments or companies for decision-making in power production and dispatching, there is only one forecasting index that is discrepant with the purpose of optimizing the industrial production process.

Deep learning not only shows excellent performance in image analysis [45], speech recognition [46], and text understanding [47], but it is also applied in the industrial field [48], which inspires us to apply deep learning to predict electricity and coal consumption synchronously.
CNN is a popular structure of deep learning, which was first proposed by Y. LeCun et al. (1998) [49]. The characteristics of weight sharing, local connection, and pooling reduction parameters give it computational advantages in the processing of big data [50] and additional feature engineering research on data is unnecessary [51], thus some scholars have applied CNN to industrial analysis such as big data classification of the Internet of Things [52] and civil infrastructure crack defect detection [53]. Since the inputs of the above two studies are two-dimensional data, which is different to one-dimensional time series data, the difference may cause the loss of the advantages of one-dimensional time series data [54]. In response to this situation, convolutional neural networks with one-dimensional convolution kernels are used in industrial fields such as industrial machine health status detection [55] and fault diagnosis for the condition monitoring of gearbox [56]. In the above two studies, the CNN with one-dimensional convolution kernel is more effective than the traditional structure in the extraction of the characteristics of time series data, which inspired this paper because of the nonlinear and non-stationary cement prediction energy synchronous production.

3. Process Analysis and Variable Selection of Cement Industry

The cement calcination system is the main energy consumption subsystem in the cement production process. Aiming at this subsystem, the relevant variables reflecting energy consumption and the time range of the moving window were selected for the proposed method. This section describes the mechanism of the cement calcination process and the reasons for selecting the parameters required for the model. The cement calcination process can be roughly divided into three parts: preheating and pre-decomposition, calcination, and cooling, during which the cement raw material is converted into calcined clinker. The cement calcination process is shown in Figure 1.

![Calcination process of cement clinker](image)

**Figure 1.** Calcination process of cement clinker.

First, the cement raw material is sent into the cyclone preheater to preheating. The cement raw material in the preheater is in suspension due to the action of the ID fan, which increases the contact area between the gas and raw material. The high temperature gas is discharged from the rotary kiln and the decomposition furnace exchanges heat with the cement raw material sufficiently, and the larger contact area contributes to the decomposition of carbonate in the raw material. Second, the cement raw material is sent into the rotary kiln that rotates at a constant speed for calcination, which contributes to the
rapid decomposition of the carbonate of the cement raw material and the progress of other physical and chemical reactions in the rotary kiln. Finally, the high temperature clinker from the kiln is cooled by the grate cooler until the temperature can be withstood by a subsequent process. It takes about 50–60 min for the cement raw material to be converted into calcined clinker. Therefore, the electricity consumption of per ton cement production (ECPC) and the coal consumption of per ton cement production (CCPC) of the current batch are reflected by the production process variables in the past hour.

As shown in Figure 1, the main energy consumption of the cement calcination process consists of coal consumption and electricity consumption; different energy consumptions are marked by different color fonts. A large amount of coal is consumed to maintain the high temperature in the entire cement rotary kiln and the preheater of the decomposition furnace. The cement rotary kiln flips a lot of raw materials and the ID fan transports the high temperature gas, so these processes consume a lot of electric energy. Furthermore, the decomposition furnace and the rotary kiln both consume a large amount of coal to maintain the temperature. Therefore, some production process variables affect both the consumption of electricity and coal, which causes an uncertain coupling relationship between electricity consumption and coal consumption. In addition, the energy consumption of cement calcination systems in the future is related to the working conditions of the past, which reflects the energy consumption of the past period, affecting that in the future.

From the above analysis, we can see that the electricity and coal consumption of the cement calcination process are affected by the amount of cement raw material and the energy consumption of the past period. Therefore, the amount of cement raw material, the ECPC, and CCPC at historical moments were selected for the prediction of coal and electricity consumption. In addition, the coal consumption is impacted by the temperature in the preheater, decomposition furnace, and rotary kiln, so the process variables selected for coal consumption prediction were as follows: primary cylinder temperature, decomposition furnace coal consumption, decomposition furnace temperature, kiln temperature, secondary air temperature, and kiln head coal consumption. Electricity consumption is influenced by process variables such as kiln current and fan speed, so the ID fan speed, EP fan speed, Kiln average current were selected for the prediction of electricity consumption.

4. The Establishment of the MWMC-CNN Model

According to the above analysis, time-varying delay, parameter redundancy, and coupling are the main barriers in the forecasting of coal and electricity consumption simultaneously, thus we propose a MWMC-CNN structure that employs a multi-channel CNN model and moving window technique to solve the time-varying delay and data redundancy problem of energy consumption prediction in the cement calcination process. The moving window technique was employed to establish the input layer of the MWMC-CNN model to solve the time-varying delay problem. The multi-channel structure was adopted in the proposed model instead of the traditional structure to solve the parameter redundancy problem. This section describes the components of the proposed MWMC-CNN model and the specific steps of the energy prediction algorithm for the cement calcination process.

4.1. The Structure of the MWMC-CNN Model

The structure of the proposed model can be divided into three parts. The first part is a time-varying delay data input layer, where the cement calcination process variables are processed in time series. The second part is the data feature extraction layers in which different cement calcination process variables are convolved and pooled separately. The third part is a regression prediction layer that integrates the output data of two independent channels and finally outputs the predicted energy consumption value. Then, the Adam algorithm propagates back errors that come from the output values of MWMC-CNN and the training tags to layers of MWMC-CNN, and then modifies the weights until the errors
are less than the expected values. The structure of the MWMC-CNN model is shown in Figure 2.

![MWMC-CNN model structure](image)

Figure 2. MWMC-CNN model structure.

4.1.1. The Time-Varying Delay Input Layer Structure of MWMC-CNN

The moving window technique was employed to establish the input layer of the MWMC-CNN model to solve the time-varying delay problem. The establishment can be divided into three parts: variables selection, window parameter selection, and time series moving window construction.

The first step in this section is to select the key variables that affect the energy consumption of the cement calcination process and normalize the selected variables data $\hat{X}$ as follows:

$$X = \frac{\hat{X} - \hat{X}_{\min}}{\hat{X}_{\max} - \hat{X}_{\min}}$$

(1)

According to the analysis of the energy consumption of the cement calcination process in Section 3, the amount of raw material ($X_1$), the ECPC at historical moments ($X_2$), and the CCPC at historical moments ($X_3$) affect the coal and electricity consumption in future. Therefore, they were selected as input variables in both channels.

The selected six variables for the A channel were as follows: ID fan speed ($X_{A1}$), kiln average current ($X_{A2}$), EP fan speed ($X_{A3}$), the amount of raw material ($X_1$), ECPC at historical moments ($X_2$), and CCPC at historical moments ($X_3$). As for the B channel, the selected nine variables were as follows: primary cylinder temperature ($X_{B1}$), decomposition furnace coal consumption ($X_{B2}$), decomposition furnace temperature ($X_{B3}$), kiln temperature ($X_{B4}$), secondary air temperature ($X_{B5}$), kiln head coal consumption ($X_{B6}$), the amount of raw material ($X_1$), ECPC at historical moments ($X_2$), and CCPC at historical moments ($X_3$).

Second, the time intervals of the data selection window are selected. As shown in Figure 3, the time interval of selecting data was selected to be $s$, which means that the time series data of the cement production process variables that contain the time-varying delay characteristic are sent to the input layer. The variable data of past $s$ time stamps correspond to the predicted value of energy consumption indicators after $p$ time stamps, which means that the $t-s-1$ period variable data in the $\hat{X}$ are adopted, to predict the energy consumption indicators at $t + p$ time stamps. The size of the moving window should be larger than the interval of time-varying delay, which means that the time delay information contained in the time series data can be extracted by the model to eliminate its influence on prediction accuracy.
Finally, the data selection input layer structure of the A channel of MWMC-CNN is shown in Figure 3. Different rows represent different variables of the A input channel. During the model training process, the cement calcination process variables data are continuously selected by the rolling data selection window and there are always m groups of data, which are arranged into a matrix in the input sequence. For instance, the amount of raw material ($X_1$) selected by the moving window is expressed as

$$x_1 = (X_1(t-s+1), X_1(t-s+2), \cdots, X_1(t))$$  \hspace{1cm} (2)

The other variables selected are in the same form as Equation (2). The data selected as the A input channel by the time series moving window are as follows:

$$x^{(s \times 6)} = (x_{A_1}, x_{A_2}, x_{A_3}, x_1, x_2, x_3)^T$$  \hspace{1cm} (3)

where $x^{(s \times 6)}$ represents the input matrix of A channel with $s$ rows and 6 columns, which means that the input sequence contains 6 variables in the past $s$ sampling times.

The data selected as the B input channel by the time series moving window are as follows:

$$x^{(s \times 9)} = (x_{B_1}, x_{B_2}, x_{B_3}, x_{B_4}, x_{B_5}, x_{B_6}, x_1, x_2, x_3)^T$$  \hspace{1cm} (4)

where $x^{(s \times 9)}$ represents the input matrix of B channel with $s$ rows and 9 columns, which means that the input sequence contains 9 variables in the past $s$ sampling times.

4.1.2. The Structure of the MWMC-CNN Data Feature Extraction Layer

Weak correlation exists among some input variables, which reduces the prediction accuracy. The multi-channel structure of CNN was designed in the data feature extraction layers of the proposed model to solve the parameter redundancy problem between weakly correlated variables.

As shown in Figure 4, the variable data of the cement calcination process selected by the moving window become the input time series data. Then, the time series data enter the data feature extraction layer of the proposed model, which performs convolution and pooling operations on the selected time series data through two independent channels. The A channel is used to extract the time series information about electricity consumption during the cement calcination process, and the B channel is used for the coal consumption.
Convolution on time series data using one-dimensional convolution kernels is considered to be efficient. The one-dimensional convolution kernel $w_u$, rather than square, is used by the proposed model to extract the time delay information implied in the time series data of the cement calcination process. The size of the one-dimensional convolution kernel is $h \times 1$, and the convolution stride is 1, which means that the kernel with length $h$ moves by one step every time. For the A channel, the input time series data are convoluted by $n$ one-dimensional convolution kernels as follows:

$$a_i^{(s-h+1)\times 6} = \sigma_{\text{relu}}(w_i x^{(s\times 6)} + b_i), i = 1, 2, \cdots n$$  \hspace{1cm} (5)$$

where $w_i$ represents the convolution kernel; the biases of $w_i$ is $b_i \cdot x^{(s\times 6)}$, which represents the input time series data with dimension of $s \times 6$, meaning that the input data has $s$ rows and 6 columns. $a_i^{(s-h+1)\times 6}$ represents the data after they are convolved and activated. $\sigma_{\text{relu}}(\cdot)$ represents the relu activation function, which is defined as follows:

$$\sigma_{\text{relu}}(x) = \max(0, x)$$ \hspace{1cm} (6)$$

The input time series data are activated by the relu function after being convolved by $n$ convolution kernels, and $n$ feature maps are output as the input of the pooling layer. Every feature map is equivalent to a data matrix. To compress the characteristics of feature maps, the one-dimensional pooling kernels were used as follows:

$$q_i^{((s-h+1)/k)\times 6} = f_{\text{avg-pooling}}(a_i^{(s-h+1)\times 6}), i = 1, 2, \cdots n$$  \hspace{1cm} (7)$$

where $a_i^{(s-h+1)\times 6}$ represents the output of convolution layer. $q_i^{((s-h+1)/k)\times 6}$ represents the data after pooled by the one-dimensional pooling kernels. The proposed model adopts one-dimensional average pooling as follows:

$$f_{\text{avg-pooling}}(x) = \frac{1}{k} \sum_{\tau=1}^{k} x_{\tau}$$ \hspace{1cm} (8)$$

One-dimensional average pooling takes an average for every $k$ value in the column and the pooling stride is $k$. In the proposed method, the dimension of input feature maps is $(s-h+1) \times 6$ and the dimension of output feature maps is $((s-h+1)/k) \times 6$, which means that the time series data are compressed vertically by the one-dimensional pooling kernel with a size of $k \times 1$ and stride of $k$. The length of feature map matrices is reduced while the number of which is unchanged after the one-dimensional pooling.

The above processes are one convolution layer and one pooling layer of the A channel, which is constructed by repeating the processes multiple times, and the structure of the B channel is similar to A except for the size of the input data matrix.
4.1.3. The Regression Prediction Layer of MWMC-CNN

Although the interference problem of weakly correlated variables does not exist in the regression prediction layer, it exists in the process of convolution and pooling, thus the multi-channel structure was adopted by the proposed model to solve this problem. The fully connected layer integrates the time-varying delay features extracted from the two CNN channels, rebuilding the coupling relationship of coal and electricity consumption. Therefore, the time-series data that have been convolved and pooled multiple times in each channel are integrated into a column as the input of the fully connected layer. As shown in Figure 2, the A channel outputs \( T_1 \) neurons, the B channel outputs \( T_2 \) neurons, thus the fully connected layer is inputted \( T_1 + T_2 \) neurons. Each input data of the fully connected layer is connected to the \( T \) neurons, which are also the output units of the fully connected layer, thus every neuron of the fully connected layer has \( T_1 + T_2 \) weights. The outputs of the full connection layer are as follows:

\[
y_z = \sigma_{relu}(w_z \rho_u + b_z), z = 1, 2, \cdots, T, u = 1, 2, \cdots, T_1 + T_2
\]

where \( \rho_u \) is the input data of fully connected layer; \( w_z \) represents the weight of the fully connected layer neuron; \( b_z \) represents the bias corresponding to \( w_z \); and \( y_z \) represents the output of the fully connected layer neurons.

The output data of fully connected layer are linearly weighted and summed to obtain the outputs of the proposed model as follows:

\[
y_a = \sum_{z=1}^{T} (y_z), \alpha = 1, 2
\]

where \( y_a \) represents the predicted values of the proposed model; \( y_1 \) represents the ECPC of the cement calcination process; \( y_2 \) represents the CCPC of the cement calcination process; and \( y_z \) represents the output data of neurons of the fully connected layer.

4.1.4. Parameter Adjustment Algorithm of MWMC-CNN

Traditional CNN usually uses the stochastic gradient descent method for backpropagation, but it is prone to overfitting and local optimal solutions. To prevent these disadvantages, it can be replaced by the Adam algorithm [57], which iteratively updates the neural network weights based on training data. Adam designs independent adaptive learning rates for parameters \( w \) and \( b \) by calculating the first moment estimation and second moment estimation of the gradient. The Adam algorithm is used to perform backpropagation to adjust the parameters by inversely adjusting the weights \( w \) and the biases \( b \) of the multi-channel convolutional neural network model proposed in this paper. For the regression prediction problem, it is convenient to use the mean square error as the objective function to measure the gradient increases or decreases with error, which contribute to the convergence of the model. The study of this paper is the synchronous regression prediction problem of coal and electricity consumption in the cement calcination process. The \( \text{MSE} \) (mean squared error) is performed as the objective function of the proposed model, which is implemented as follows:

\[
\text{Minimize : } J(w, b; x) = \sum_{i=1}^{n} (y_a(w, b; x) - \hat{y})^2
\]

\[
x = \begin{bmatrix}
X_{A_1}(t - s + 1) & X_{A_1}(t - s + 2) & \cdots & X_{A_1}(t) \\
X_{A_2}(t - a + 1) & X_{A_2}(t - a + 2) & \cdots & X_{A_2}(t) \\
\vdots & \vdots & \ddots & \vdots \\
X_{B_b}(t - a + 1) & X_{B_b}(t - a + 2) & \cdots & X_{B_b}(t)
\end{bmatrix}^T
\]

where \( \hat{y} \) represents the true value of energy consumption in cement calcination process; \( y_a(w, b; x) \) represents the predicted energy consumption of the proposed model; and \( x \)
represents the time series data of the cement calcination process, which are input into the A channel and B channel of the proposed model to predict ECPC and CCPC, respectively.

The following process is the update process of weights \( w \) and bias \( b \) by the Adam algorithm. The default parameters of the Adam algorithm are preserved, which include the learning rate \( r \) (\( r = 0.001 \)), the exponential decay rate of the moment estimate \( \beta_1 \) and \( \beta_2 \) (\( \beta_1 = 0.9 \), \( \beta_2 = 0.999 \)); and the constant \( \delta \) (\( \delta = 10^{-\delta} \)).

First and second moment variables and time steps are initialized to be zero (\( m_0 = 0 \), \( v_0 = 0 \), \( t = 0 \)). The error converges to a certain value \( e (e = 0.001) \), which is regarded as the stopping criterion of the proposed model. In the training process, time steps increase step by step (\( t = t + 1 \)) and in every step, \( L \) sets of data are randomly selected from the cement calcination process variable training dataset as follows:

\[
x_{random}^L = \left\{ x^{(1)}, x^{(2)}, \ldots, x^{(l)}, \ldots, x^{(L)} \right\}
\]

(13)

where \( x^{(i)} \) represents a randomly selected set of training samples. The average gradient \( g_w \) and \( g_b \) at time step \( t \) of the \( L \) samples are calculated according to the adopted objective function as follows:

\[
g_w = \frac{1}{L} \sum_{i=1}^{L} J(w_{l-1}; x^{(l)})
\]

(14)

\[
g_b = \frac{1}{L} \sum_{i=1}^{L} J(b_{l-1}; x^{(l)})
\]

(15)

The biased first moment estimates based on the gradient are updated as follows:

\[
m_{w_l} = \beta_1 \cdot m_{w_{l-1}} + (1 - \beta_1) \cdot g_w
\]

(16)

\[
m_{b_l} = \beta_1 \cdot m_{b_{l-1}} + (1 - \beta_1) \cdot g_b
\]

(17)

The biased second moment estimates based on the gradient are updated as follows:

\[
v_{w_l} = \beta_2 \cdot v_{w_{l-1}} + (1 - \beta_2) \cdot g_w
\]

(18)

\[
v_{b_l} = \beta_2 \cdot v_{b_{l-1}} + (1 - \beta_2) \cdot g_b
\]

(19)

The deviations of the first moment estimates based on the biased first moment estimate are corrected as follows:

\[
\delta_{w_l} = m_{w_l} / (1 - \beta_1^l)
\]

(20)

\[
\delta_{b_l} = m_{b_l} / (1 - \beta_1^l)
\]

(21)

The deviations of the second moment estimates based on the biased second moment estimates are corrected as follows:

\[
\hat{\delta}_{w_l} = v_{w_l} / (1 - \beta_2^l)
\]

(22)

\[
\hat{\delta}_{b_l} = v_{b_l} / (1 - \beta_2^l)
\]

(23)

where \( \beta_1^l \) and \( \beta_2^l \) represent the values of \( \beta_1 \) and \( \beta_2 \) at time step \( t \), respectively.

The correction values of the parameter \( \Delta w_t \) and \( \Delta b_t \) are calculated based on the above deviations as follows:

\[
\Delta w_t = -r \cdot \delta_{w_t} / (\sqrt{\delta_{w_t}} + \delta)
\]

(24)

\[
\Delta b_t = -r \cdot \delta_{b_t} / (\sqrt{\delta_{b_t}} + \delta)
\]

(25)

\[
w_t = w_{l-1} + \Delta w_t
\]

(26)

\[
b_t = b_{l-1} + \Delta b_t
\]

(27)
The parameter update process of the Adam algorithm acts on every element, thus the weight of every neuron of the proposed model is adjusted separately. For the training of the proposed model, the parameters were continuously updated by the Adam algorithm until the objective function error was less than convergence error ε.

4.2. Research on MWMC-CNN Algorithm

The above analysis is the forward propagation process of the proposed MWMC-CNN model, which can be divided into three parts: the moving window input part, the convolution and pooling part, and the fully connected output part. In the MWMC-CNN training process, the Adam algorithm is used to fine-tune parameters. This section gives a full description of the pseudo code of the MWMC-CNN model Algorithms 1 and 2. The details of the formulas in the algorithm are in Section 4.1.

**Algorithm 1 Energy consumption prediction algorithm of MWMC-CNN**

**Input A:**
ID fan speed (X_{A_1}), EP fan speed (X_{A_2}), kiln average current (X_{A_3}), the amount of raw material (X_1), the ECPC at historical moments (X_2), and the CCPC at historical moments (X_3)

**Input B:**
Primary cylinder temperature (X_{B_1}), decomposition furnace coal consumption (X_{B_2}), decomposition furnace temperature (X_{B_3}), kiln temperature (X_{B_4}), secondary air temperature (X_{B_5}), kiln head coal consumption (X_{B_6}), the amount of raw material (X_1), the ECPC at historical moments (X_2), and the CCPC at historical moments (X_3).

**Output:**
The ECPC at future moments (y_1), the CCPC at future moments (y_2).

**(Step 1) Initialization:**
**(Step 1.1) Normalization:**
X ← \frac{X-\bar{X}}{\sigma_{X}}

**(Step 2.1) Convolution and activation:**
A channel: \( a_{i}^{(n-h+1)\times 6} \leftarrow \sigma_{relu}(w_{x}x^{(n-h+1)\times 6} + b_{i}), i = 1, 2, \cdots n \)
B channel: \( a_{i}^{(n-h+1)\times 9} \leftarrow \sigma_{relu}(w_{y}y^{(n-h+1)\times 9} + b_{i}), i = 1, 2, \cdots n \)

**(Step 2.2) Pooling:**
A channel: \( q_{i}^{(n-h+1)\times 6} \leftarrow f_{\text{avg-pooling}}(a_{i}^{(n-h+1)\times 6}), i = 1, 2, \cdots n \)
B channel: \( q_{i}^{(n-h+1)\times 9} \leftarrow f_{\text{avg-pooling}}(a_{i}^{(n-h+1)\times 9}), i = 1, 2, \cdots n \)
Repeat Step 2.1–2.2 several times in each channel.

**(Step 3) Full connection and output:**
**(Step 3.1) Full connection:**
y_{z} \leftarrow \sum_{u=1}^{T_{1} + T_{2}} \sigma_{relu}(w_{z}a_{u} + b_{z}), z = 1, 2, \cdots T

**(Step 3.2) Output:**
y_{a} \leftarrow \sum_{z=1}^{T} (y_{z}), a = 1, 2
12,500 values. A total of 11,500 groups of all the data were used for model training, while the remaining 1000 groups were used as the test set.

The experimental results were compared and analyzed to verify the superiority of the experiment, in which every variable in every group has one value, so every variable has 48 hidden units to predict the electricity consumption and coal consumption of the cement calcination process.

In this section, we used a MWMC-CNN model with two channels where each channel contained three convolutional layers and three pooling layers; a CNN model with three convolution layers and three pooling layers; a LSSVM model in which $p = 0.03$ and $g = 0.01$; and a LSTM model with two LSTM layers where each layer contains 48 hidden units to predict the electricity consumption and coal consumption of the cement calcination process. The experimental results were compared and analyzed to verify the superiority of the proposed model. In total, 12,500 datasets containing the electricity and coal consumption and cement calcination process variables stated in Section 4.1.1 sampled by sensors on the production line of a cement manufacturing enterprises in China were selected for the experiment, in which every variable in every group has one value, so every variable has 12,500 values. A total of 11,500 groups of all the data were used for model training, while the remaining 1000 groups were used as the test set.

### Algorithm 2 Parameter adjustment algorithm of MWMC-CNN

The parameter update process of Adam algorithm for every element.

**Objective function:**

$$Minimize : J(w, b; x) = \sum_{i=1}^{l} (y_{sa}(w, b; x) - \hat{y})^2$$

$$x \leftarrow \begin{bmatrix} A_1(t-s+1) & A_1(t-s+2) & \cdots & A_1(t) \\ A_2(t-a+1) & A_2(t-a+2) & \cdots & A_2(t) \\ \vdots & \vdots & \ddots & \vdots \\ B(t-a+1) & B(t-a+2) & \cdots & B(t) \end{bmatrix}^T$$

**Parameter Initialization:**

- $r \leftarrow 0.001$ (Learning rate)
- $\beta_1 \leftarrow 0.9, \beta_2 \leftarrow 0.999$ (Exponential decay rates of moment estimates)
- $\delta \leftarrow 10^{-8}$ (Constant)
- $m_0 \leftarrow 0$ (Initialize 1st moment vector)
- $r \leftarrow 0$ (Initialize 2nd moment vector)
- $t \leftarrow 0$ (Initialize time step)

**while $\varepsilon$ not converged do:**

1. $t \leftarrow t + 1$
2. $x_{\text{random}}^{l} \leftarrow \{x(1), x(2), \cdots, x(l), \cdots, x(L)\}$ (Randomly select $L$ sets of data)
3. $g_{w} = \frac{1}{L} \sum_{l=1}^{L} \nabla_{w} J(w_{t-1}; x(l))$ (Calculate the gradient of $w_l$)
4. $g_{b} = \frac{1}{L} \sum_{l=1}^{L} \nabla_{b} J(b_{t-1}; x(l))$ (Calculate the gradient of $b_l$)
5. $m_{w} \leftarrow \beta_1 \cdot m_{w_{t-1}} + (1 - \beta_1) \cdot g_{w}$ (Update biased 1st moment estimate of $w_l$)
6. $m_{b} \leftarrow \beta_1 \cdot m_{b_{t-1}} + (1 - \beta_1) \cdot g_{b}$ (Update biased 1st moment estimate of $b_l$)
7. $v_{w} \leftarrow \beta_2 \cdot v_{w_{t-1}} + (1 - \beta_2) \cdot \hat{g}_{w}$ (Update biased 2nd moment estimate of $w_l$)
8. $v_{b} \leftarrow \beta_2 \cdot v_{b_{t-1}} + (1 - \beta_2) \cdot \hat{g}_{b}$ (Update biased 2nd moment estimate of $b_l$)
9. $\hat{m}_{w} \leftarrow m_{w} / (1 - \beta_1)$ (Compute bias-corrected 1st moment estimate of $w_l$)
10. $\hat{m}_{b} \leftarrow m_{b} / (1 - \beta_1)$ (Compute bias-corrected 1st moment estimate of $b_l$)
11. $\hat{v}_{w} \leftarrow v_{w} / (1 - \beta_2)$ (Compute bias-corrected 2nd moment estimate of $w_l$)
12. $\hat{v}_{b} \leftarrow v_{b} / (1 - \beta_2)$ (Compute bias-corrected 2nd moment estimate of $b_l$)
13. $\Delta w_{l} \leftarrow -r \cdot \hat{m}_{w} / (\sqrt{\hat{v}_{w} + \delta})$ (Calculate parameter update value of $w_l$)
14. $\Delta b_{l} \leftarrow -r \cdot \hat{m}_{b} / (\sqrt{\hat{v}_{b} + \delta})$ (Calculate parameter update value of $b_l$)
15. $w_{l} \leftarrow w_{l-1} + \Delta w_{l}$ (Update parameter $w_l$)
16. $b_{l} \leftarrow b_{l-1} + \Delta b_{l}$ (Update parameter $b_l$)

**end while**

**return** $w_l, b_l$
The root mean square error (RMSE), the mean relative error (MRE), and the mean absolute error (MAE) are used as the indicators of predictive performance, whose formulae are respectively shown as:

\[
RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2}
\]  
(28)

\[
MRE = \frac{1}{m} \sum_{i=1}^{m} \left| \frac{\hat{y}_i - y_i}{y_i} \right|
\]  
(29)

\[
MAE = \frac{1}{m} \sum_{i=1}^{m} |\hat{y}_i - y_i|
\]  
(30)

where \(\hat{y}_i\) is the sample value and \(y_i\) is the predictive value. \(m\) is the size of the forecasting sample.

5.1. Parameter and Structure Adjustment Experiment of MWMC-CNN

The depth of the CNN and the size of the convolution kernel affect the performance of the proposed model. In order to find the parameters that make the proposed model perform better, we conducted a parameter adjustment experiment to compare the errors of the MWMC-CNN with different depths and different convolution kernel sizes under the actual cement calcination process data. The experimental results are shown in Figure 5.

We performed ten trials on MWMC-CNN with different sizes of the convolution kernel. All the above comparison experiments were optimized by the same optimization algorithm with the same parameters except for the depth and kernel size. The mean value of prediction errors of MWMC-CNN with six layers are represented by the dark-colored rectangles, and the line segment at the top of rectangle represents the variance of the experimental prediction error. The corresponding mean values and variances of the experimental prediction errors of MWMC-CNN with four layers are represented by light-colored rectangles and line segments above them, respectively.

As shown in Figure 5, the mean value and variance of experimental prediction error of MWMC-CNN with six convolution layers and a kernel size of 10-1 were the smallest. The above experimental results show that MWMC-CNN with six convolution layers is more suitable for feature extraction of the time series data of the cement calcination process. Using a convolution kernel size of 10-1 is beneficial to the improvement in the prediction performance of the proposed model.

5.2. Experiments Comparison of MWMC-CNN and Other Models

In order to show the intuitiveness of the experimental results, the MWMC-CNN, CNN, LSSVM, and LSTM were trained and tested under the same datasets, respectively. As a kind of highly competitive shallow network, LSSVM is widely used in industry field, so it is reasonable to choose LSSVM for the comparison. As a kind of time-dependent model, LSTM is designed to extract time series information and simulate long-term dependencies and short-term dependencies. The performance of time series information extraction and hidden feature mining of the proposed model can be verified by a comparison with LSTM and CNN, respectively. The experimental results are shown in Figures 6–10. Figures 6 and 7 indicate that all the models learned the characteristics of coal and electricity consumption synchronous prediction, which was the basis of the test. Figure 9 shows the loss curves of the models. Figures 8 and 9 are the forecasting results, which demonstrate the superiority of the proposed model.
Figure 5. Error comparison of different structure MWMC-CNN. (a) RMSE comparison of different structure MWMC-CNN; (b) MRE comparison of different structure MWMC-CNN; and (c) MAE comparison of different structure MWMC-CNN.
Figure 6. Electricity consumption training result curves of LSSVM, CNN, LSTM and MWMC-CNN.

Figure 7. Coal consumption training result curves of LSSVM, CNN, LSTM and MWMC-CNN.

Figure 8. The loss curves of LSTM, CNN and MWMC-CNN.
In order to evaluate whether LSSVM, CNN, LSTM, and MWMC-CNN learned the characteristics of training data, we carried out prediction experiments using the training dataset. The more overlapped the two lines, the more characteristics have been learned. The training result curves of the electricity and coal consumption of the four models are shown in Figures 6 and 7, respectively, in which the blue line represents the original data.
samples while the red line represents the training output values of the corresponding model. The same termination condition of training for the four models was to reach a set error. It can be seen from the figures that the performances of the four models in the training stage were similar. The experiments indicate that all four prediction models grasped the characteristics of the training data.

Figure 8 shows the loss curves of CNN, LSTM, and MWMC-CNN. Due to the learning mechanism of hyperplane mapping, we did not draw the loss curve of LSSVM. It can be seen that all the losses converged to around 0. The fluctuation range of LSTM was the largest. The loss of MWMC-CNN was very close to 0 and lower than that of CNN. In addition, the convergence speed of MWMC-CNN was faster than CNN.

To evaluate the actual prediction abilities of the four prediction models, we tested the models using the test dataset, which was different with training dataset. The test result curves of the electricity and coal consumption of the four models are shown in Figures 9 and 10, respectively.

In Figures 9 and 10, the blue line represents the original data samples while the red line represents the test output values of the corresponding model. Below every subgraph is the error curve of the corresponding model for the prediction index. It can be seen that the prediction curves of both the electricity and coal consumption of LSSVM were the worst and CNN was better than LSSVM, but worse than LSTM, which could match the actual value of electricity and coal consumption essentially but the precision was lower than MWMC-CNN. Through the comprehensive analysis of power and coal consumption synchronous prediction error curves of the four models, it can be concluded that the error curves of MWMC-CNN were the most stable among the four models for power and coal prediction synchronously. The experimental results verify the validity of the proposed model in this paper and demonstrate that the MWMC-CNN model has better generalization ability and higher precision for multi-energy index synchronous forecasting. The test errors of the four models are shown in Table 1.

| Model   | Parameter | Prediction Error |
|---------|-----------|------------------|
|         |           | RMSE  | MRE    | MAE    |
| LSSVM   | $p = 0.03$ $g = 0.01$ | 0.0172 | 0.0340 | 0.0132 |
| CNN     | 4-1       | 0.0168 | 0.0284 | 0.0126 |
| LSTM    | 48-48     | 0.0142 | 0.0227 | 0.0101 |
| MWMC-CNN| 10-1      | 0.0108 | 0.0175 | 0.0079 |

As shown in Table 1, the minimum error was achieved when the convolution kernel size of MWMC-CNN with six convolution layers was 10-1. In all of the above results, deep networks such as CNN and LSTM performed better than the shallow network such as LSSVM in the time series prediction of the cement calcination process, which indicates that deep networks such as CNN are more suitable for the research in this paper. Compared with CNN and LSTM, MWMC-CNN had higher precision, the RMSE of MWMC-CNN was better than of 35.7% for CNN and 23.9% for LSTM; the MRE of MWMC-CNN was better than of 38.3% for CNN and 22.9% for LSTM; the MAE of MWMC-CNN was better than of 37.3% for CNN and 21.8% for LSTM. The results indicate that the proposed model has better time series feature extraction ability.

6. Conclusions

In this paper, we proposed a multi-channel convolutional neural network (MWMC-CNN) structure with a moving window based on the analysis of the mechanism of the cement calcination process and coupling relationship between the cement production process variables, which reduces the impact of time-varying delay in the cement calcination process on prediction accuracy while reducing the burden of industrial system modeling. The experimental results demonstrate the outstanding performance of the proposed model.
compared with the CNN model. In addition, according to different industrial mechanisms, the proposed model can be expanded to more output or input to deal with more complicated time series data, which provides a reference for the multi-output soft measurement modeling of the process industry and time series analysis of industrial big data. In our future studies, the proposed method will be expanded to predict the consumption of other kinds of energy or the energy consumption of other production processes, which provides a reference for energy scheduling and intelligent optimization control of the process industry.

Author Contributions: Conceptualization, X.S., G.H. and X.H.; Methodology, X.S.; Software, Y.Y.; Validation, X.S. and Y.Y.; Formal analysis, G.H. and X.H.; Investigation, X.S., X.H. and Y.Y.; Data curation, Y.Y. and Z.L.; Writing—original draft preparation, Z.L.; Writing—review and editing, X.S. and G.H.; Supervision, X.S. and X.H.; Funding acquisition, X.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 62073281); the Hebei Provincial Natural Science Foundation (Grant No. F2019203385); the Hebei Provincial Science and Technology Plan Project (Grant No. 19211602D); and the Second Batch of Youth Top-Notch Talent Support Program in Hebei Province (Grant No. 5040050).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Zhang, C.; Yu, B.; Chen, J.; Wei, Y. Green transition pathways for cement industry in China. *Resour. Conserv. Recyl.* 2021, 166, 1–16. [CrossRef]
2. Shen, W.; Liu, Y.; Yan, B.; Wang, J.; He, P.; Zhou, C.; Hao, X.; Zhang, W.; Xu, G.; Ding, Q. Cement industry of China: Driving force, environment impact and sustainable development. *Renew. Sustain. Energy Rev.* 2017, 75, 618–628. [CrossRef]
3. Jin, X.; Qi, F.; Wu, Q.; Mu, Y.; Jia, H.; Yu, X.; Li, Z. Integrated optimal scheduling and predictive control for energy management of an urban complex considering building thermal dynamics. *Int. J. Electr. Power Energy Syst.* 2020, 123, 106273. [CrossRef]
4. Banik, R.; Das, P.; Ray, S.; Biswas, A. Prediction of electrical energy consumption based on machine learning technique. *Electr. Eng.* 2021, 103, 909–920. [CrossRef]
5. Ma, M.D.; Wong, D.S.H.; Jang, S.S.; Tseng, S.T. Fault Detection Based on Statistical Multivariate Analysis and Microarray Visualization. *IEEE Trans. Ind. Inform.* 2010, 6, 18–24.
6. Ceperic, E.; Ceperic, V.; Baric, A. A Strategy for Short-Term Load Forecasting by Support Vector Regression Machines. *IEEE Trans. Power Syst.* 2013, 28, 4356–4364. [CrossRef]
7. Gao, C.H.; Jian, L.; Luo, S.H. Modeling of the Thermal State Change of Blast Furnace Hearth with Support Vector Machines. *IEEE Trans. Ind. Elect.* 2012, 59, 1134–1145. [CrossRef]
8. Fan, G.F.; Wei, X.; Li, Y.T.; Hong, W.C. Forecasting electricity consumption using a novel hybrid model. *Sustain. Cities Soc.* 2020, 62, 102320. [CrossRef]
9. Zhao, H.; Zhang, N.; Wang, H.J. Power Consumption Prediction Modeling of Cement Manufacturing Based on the Improved Multiple Non-linear Regression Algorithm. *Appl. Mech. Mater.* 2014, 687–691, 5185–5189. [CrossRef]
10. Liu, Z.; Wang, X.H.; Zhang, Q.; Huang, C.Y. Empirical mode decomposition based hybrid ensemble model for electrical energy consumption forecasting of the cement grinding process. *Measurement* 2019, 138, 314–324. [CrossRef]
11. Cao, K.; Hu, T.; Li, Z.; Zhao, G.; Qian, X. Deep multi-task learning model for time series prediction in wireless communication. *Phys. Commun.* 2021, 44, 101251. [CrossRef]
12. Zhou, Y.; Liu, Y.; Wang, D.; De, G.; Li, Y.; Liu, X.; Wang, Y. A novel combined multi-task learning and Gaussian process regression model for the prediction of multi-timescale and multi-component of solar radiation. *J. Clean. Prod.* 2021, 124710. [CrossRef]
13. Yan, S.; Yan, X. Joint monitoring of multiple quality-related indicators in nonlinear processes based on multi-task learning. *Measurement* 2020, 165, 108158. [CrossRef]
14. Sun, Z.; Liang, D.; Zhuang, Z.; Chen, L.; Ming, X. Multi-task processing oriented production layout based on evolutionary programming mechanism. *Appl. Soft Comput.* 2021, 98, 106896. [CrossRef]
15. Lavanya, M.; Shanthi, B.; Saravanan, S. Multi objective task scheduling algorithm based on SLA and processing time suitable for cloud environment. *Comput. Commun.* 2020, 151, 183–195. [CrossRef]
16. Hamer, A.M.; Simms, D.M.; Waine, T.W. Replacing huming interpretation of agricultural land in Afghanistan with a deep convolutional neural network. *Int. J. Remote Sens.* 2021, 42, 3017–3038. [CrossRef]
17. Youcef, D.; Gautam, S.; Jerry, C.L. Fast and Accurate Convolution Neural Network for Detecting Manufacturing Data. *IEEE Trans. Ind. Inform.* 2021, 17, 2947–2955.

18. Bianco, V.; Manca, O.; Nardini, S. Electricity consumption forecasting in Italy using linear regression models. *Energy* 2009, 34, 1413–1421. [CrossRef]

19. Catalina, T.; Iordache, V.; Caracaleanu, B. Multiple regression model for fast prediction of the heating energy demand. *Energy Build.* 2013, 57, 302–312. [CrossRef]

20. Dudek, G. Pattern-based local linear regression models for short-term load forecasting. *Electr. Power Syst. Res.* 2016, 130, 139–147. [CrossRef]

21. Drucker, H.; Burges, C.J.C.; Kaufman, L.; Smola, A.; Vapnik, V. Support Vector Regression Machines. *Adv. Neural Inf. Process. Syst.* NIPS 1996, 9, 155–161.

22. Wang, S.; Yu, L.; Tang, L.; Wang, S.Y. A novel seasonal decomposition based least squares support vector regression ensemble learning approach for hydropower consumption forecasting in China. *Energy* 2011, 36, 6542–6554. [CrossRef]

23. Assouline, D.; Mohajeri, N.; Scarlettini, J.L. Quantifying rooftop photovoltaic solar energy potential: A machine learning approach. *Solar Energy* 2017, 141, 278–296. [CrossRef]

24. Zhou, J.Y.; Shi, J.; Li, G. Fine tuning support vector machines for short-term wind speed forecasting. *Energy Convers. Manag.* 2011, 52, 1990–1998. [CrossRef]

25. Kavousi-Fard, A.; Samet, H.; Marzbani, F. A new hybrid Modified Firefly Algorithm and Support Vector Regression model for accurate Short Term Load Forecasting. *Expert Syst. Appl.* 2014, 41, 6047–6056. [CrossRef]

26. Yin, S.; Ding, S.X.; Xie, X.C.; Luo, H. A Review on Basic Data-Driven Approaches for Industrial Process Monitoring. *IEEE Trans. Indusrial Electron.* 2014, 61, 6418–6428. [CrossRef]

27. Fachini, F.; Fuly, B.I.L. A comparison of machine learning regression models for critical bus voltage and load mapping with regards to max reactive power in PV buses. *Electr. Power Syst. Res.* 2021, 191, 106883. [CrossRef]

28. Gong, M.; Wang, J.; Bai, Y.; Li, B.; Zhang, L. Heat load prediction of residential buildings based on discrete wavelet transform and tree-based ensemble learning. *J. Build. Eng.* 2020, 32, 101455. [CrossRef]

29. Lu, S.; Li, Q.; Bai, L.; Wang, R. Performance predictions of ground source heat pump system based random forest and back propagation neural network models. *Energy Convers. Manag.* 2019, 197, 111864. [CrossRef]

30. Qiu, T.; Zhang, M.; LIU, X.; Liu, J.; Chen, C.; Zhao, W. A Directed Edge Weight Prediction Model Using Decision Tree Ensembles in Industrial Internet of Things. *IEEE Trans. Ind. Inform.* 2021, 17, 2160–2168. [CrossRef]

31. Kumar, A.; Jaiswal, A. A Deep Swarm-Optimized Model for Leveraging Industrial Data Analytic in Cognitive Manufacturing. *IEEE Trans. Ind. Inform.* 2021, 17, 2938–2946. [CrossRef]

32. Marugan, A.P.; Marquez, F.P.G.; Perez, J.M.P.; Ruiz-Hernandez, D. A survey artificial neural network in wind energy systems. *Appl. Energy* 2018, 228, 1822–1836. [CrossRef]

33. Debnath, K.B.; Moursed, M. Forecasting methods in energy planning models. *Renew. Sustain. Energy Rev.* 2018, 88, 297–325. [CrossRef]

34. Azadeh, A.; Ghaderi, S.F.; Sohrabkhani, S. Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors. *Energy Convers. Manag.* 2008, 49, 2272–2278. [CrossRef]

35. Zeng, Y.R.; Zeng, Y.; Choi, B.; Lin, W. Multifactor-influenced energy consumption forecasting using enhanced back-propagation neural network. *Energy* 2017, 127, 381–396. [CrossRef]

36. Rodrigues, F.; Cardeira, C.; Calado, J.M.F. The daily and hourly energy consumption and load forecasting using artificial neural network method: A case study using a set of 93 households in Portugal. *Energy Procedia* 2014, 62, 220–229. [CrossRef]

37. Panapakidis, I.P.; Dagoumas, A.S. Day-ahead electricity price forecasting via the application of artificial neural network based models. *Appl. Energy* 2016, 172, 132–151. [CrossRef]

38. Szoplik, J. Forecasting of natural gas consumption with artificial neural networks. *Energy* 2015, 85, 208–220. [CrossRef]

39. Zhang, C.; Lim, P.; Qin, A.K.; Tan, C.K. Multiobjective Deep Belief Networks Ensemble for Remaining Useful Life Estimation in Prognostics. *IEEE Trans. Neural Netw. Learn. Syst.* 2017, 28, 2306–2318. [CrossRef] [PubMed]

40. Anagnostis, A.; Papageorgiou, E.; Bachtis, D. Application of Artificial Neural Networks for Natural Gas Consumption Forecasting. *Sustainability* 2020, 12, 6409. [CrossRef]

41. Kim, T.-Y.; Cho, S.-B. Predicting residential energy consumption using CNN-LSTM neural networks. *Energy* 2019, 182, 72–81. [CrossRef]

42. Hwang, J.; Suh, D.; Otto, M.O. Forecasting Electricity Consumption in Commercial Buildings Using a Machine Learning Approach. *Energies* 2020, 13, 5885. [CrossRef]

43. Son, N.; Yang, S.; Na, J. Deep Neural Network and Long Short-Term Memory for electric Power Load Forecasting. *Appl. Sci.* 2020, 10, 6489. [CrossRef]

44. Atef, S.; Eltawil, A.B. Assessment of stacked unidirectional and bidirectional long short-term memory networks for electricity load forecasting. *Electr. Power Syst. Res.* 2020, 187, 106489. [CrossRef]

45. Gu, J.X.; Wang, Z.H.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognit.* 2018, 77, 354–377. [CrossRef]

46. Liu, W.B.; Wang, Z.D.; Liu, X.H.; Zeng, N.; Liu, Y.; Alsaaedi, F.E. A survey of deep neural network architectures and their applications. *Neurocomputing* 2017, 234, 11–26. [CrossRef]
47. Zhang, Q.C.; Yang, L.T.; Chen, Z.K.; Li, P. A survey on deep learning for big data. *Inf. Fusion* **2018**, *42*, 146–157. [CrossRef]
48. Wang, J.J.; Ma, Y.L.; Zhang, L.B.; Gao, R.X.; Wu, D.Z. Deep learning for smart manufacturing: Methods and applications. *J. Manuf. Syst.* **2018**, *48*, 144–156. [CrossRef]
49. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
50. Chen, X.W.; Lin, X.T. Big data deep learning: Challenges and perspectives. *IEEE Access* **2014**, *2*, 514–525. [CrossRef]
51. Bengio, Y.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828. [CrossRef] [PubMed]
52. Li, P.; Chen, Z.K.; Yang, L.T.R.; Zhang, Q.C.; Deen, M.J. Deep Convolutional Computation Model for Feature Learning on Big Data in Internet of Things. *IEEE Trans. Ind. Inform.* **2018**, *14*, 790–798. [CrossRef]
53. Cha, Y.J.; Choi, W.; Buyukozturk, O. Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks. *Comput. Aided Civ. Infrastruct. Eng.* **2017**, *32*, 361–378. [CrossRef]
54. Zhang, W.; Li, C.H.; Peng, G.L.; Chen, Y.H.; Zhang, Z.J. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mech. Syst. Signal Process.* **2018**, *100*, 439–453. [CrossRef]
55. Zhao, R.; Yan, R.Q.; Wang, J.J.; Mao, K.Z. Learning to Monitor Machine Health with Convolutional Bi-Directional LSTM Networks. *Sensors* **2017**, *17*, 273. [CrossRef] [PubMed]
56. Jing, L.Y.; Zhao, M.; Li, P.; Xu, X.Q. A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement* **2017**, *111*, 1–10. [CrossRef]
57. Kingma, D.P.; Ba, J.L. Adam: A Method for Stochastic Optimization. In Proceedings of the International Conference on Learning Representations, ICLR, San Diego, CA, USA, 7–9 May 2015; pp. 1–15.