


Article

Forest Fire Prediction with Imbalanced Data Using a Deep Neural Network Method

Can Lai ¹, Shucai Zeng ², Wei Guo ¹, Xiaodong Liu ¹, Yongquan Li ¹ and Boyong Liao ^{1,*} 

¹ College of Horticulture and Landscape Architecture, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China; laican@zhku.edu.cn (C.L.); gwei717@163.com (W.G.); qi_yueliuhuo@163.com (X.L.); yongquanli@zhku.edu.cn (Y.L.)

² College of Forestry & Landscape Architecture, South China Agricultural University, Guangzhou 510642, China; sczeng@scau.edu.cn

* Correspondence: liaoby05@126.com

Abstract: Forests suffer from heavy losses due to the occurrence of fires. A prediction model based on environmental condition, such as meteorological and vegetation indexes, is considered a promising tool to control forest fires. The construction of prediction models can be challenging due to (i) the requirement of selection of features most relevant to the prediction task, and (ii) heavily imbalanced data distribution where the number of large-scale forest fires is much less than that of small-scale ones. In this paper, we propose a forest fire prediction method that employs a sparse autoencoder-based deep neural network and a novel data balancing procedure. The method was tested on a forest fire dataset collected from the Montesinho Natural Park of Portugal. Compared to the best prediction results of other state-of-the-art methods, the proposed method could predict large-scale forest fires more accurately, and reduces the mean absolute error by 3–19.3 and root mean squared error by 0.95–19.3. The proposed method can better benefit the management of wildland fires in advance and the prevention of serious fire accidents. It is expected that the prediction performance could be further improved if additional information and more data are available.

Keywords: fire prediction; imbalanced data; deep learning; sparse autoencoder; deep neural network



Citation: Lai, C.; Zeng, S.; Guo, W.; Liu, X.; Li, Y.; Liao, B. Forest Fire Prediction with Imbalanced Data Using a Deep Neural Network Method. *Forests* **2022**, *13*, 1129. <https://doi.org/10.3390/f13071129>

Received: 26 May 2022

Accepted: 16 July 2022

Published: 18 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Forest covers 31% of the global land surface and is of great importance in the wildland ecosystem. Forest fires are one of the major challenges for the preservation of forest, and cause great economic and ecological losses and even loss of human lives [1]. Even though much attention and expense have been paid to monitor and control forest fires [2], global annual burned forest hectares (ha) are in the millions [3]. The development of prediction models are expected to benefit fire management strategies in the ecosystem [4].

Traditionally a forest fire is monitored by watch-keepers on a watchtower, but it is not feasible to construct a lot of watchtowers scattered in extensive forests, and the watch-keeper can only detect fires already occurring instead of making predictions [5]. It is known that the occurrence of a forest fire is correlated with environmental conditions. For example, a forest fire is more likely to occur in hot and dry condition compared to old and humid ones [6]. In the modern world, since numerous meteorological stations are available, the collection of weather data is fast and cheap. In addition, with the help of satellite remote sensing technology, local area conditions, such as the state of crops and land surface temperature, can be computed based on satellite images [7]. This information can greatly benefit the construction of real-time and non-costly forest fire prediction methods.

Environmental data include many different variables, such as temperature, air pressure, humidity, wind speed, and vegetation index. These variables are usually recorded as numerical values and have valuable patterns correlated with the occurrence of forest fires. Since the dimensions of environmental data, i.e., the number of variables, are large, it

is difficult for human experts to analyze the complex patterns. Machine learning models have been employed for automatically learning the relationship between the environmental data and the occurrence of forest fires [8,9]. Logistic regression and random forest were compared for fire detection in Slovenian forests [10]. Cortez et al. tested five different machine learning models for burned area prediction in the Montesinho Natural Park of Portugal using meteorological data [5]. Artificial Neural Networks (ANNs) and logistic regression were applied for prediction of fire danger in Galicia [11]. Genetic programming was adopted for forest burned areas prediction based on meteorological data in [12]. A hybrid machine learning approach for predicting fire risk indices was proposed in [13]. West et al. predicted the occurrence of large wildfires based on multivariate regression and future climate information [14]. The performance of different methods, such as the cascade correlation network, ANN, polynomial neural network, radial basis function and support vector machine (SVM), were assessed for forest fires prediction in [15]. Sayad et al. predicted the occurrence of wildfires based on weather-related metrics using ANN and SVM [7]. Fuzzy logic models were utilized for forest fire forecasting in [16]. Meteorological and land-cover parameters were used for burned area prediction in [17]. A comprehensive precipitation index was built and used for predicting forest fire risk in central and northern China [18].

The limitations of these methods are (i) they are based on shallow machine learning models that require the selection of useful features as the input, and (ii) they do not consider the imbalanced problem in the historical data, e.g., the number of large-scale forest fire is much less than that of small-scale ones [19], making the prediction models neglect information of large-scale forest fires, which is, in fact, more important to prevent serious consequences [20].

To contribute to addressing these limitations, the use of deep learning (DL) methods for forest fire prediction is considered in this work. DL methods are able to automatically extract useful features based on neural networks with multiple layers [21]. Among the DL methods, convolutional neural networks and recurrent neural networks have been employed to process satellite images or videos for forest fire detection [22–24]. An autoencoder-based deep neural network (DNN) is the most suitable for the processing of numerical input values. The autoencoder is constructed by an “encoder” network and a “decoder” network whose structures are symmetrical to the encoder [25]. The encoder network automatically extracts features from the input data and the decoder reconstructs the data from the extracted features. A sparse autoencoder tends to extract more representative and discriminative features by adopting a sparsity penalty for the activation of network neurons [26]. A sparse autoencoder-based DNN can be built by taking the encoder of the sparse autoencoder and adding a regression/classification layer on top of it.

To deal with the imbalanced problem, a data balancing procedure is proposed, whose key idea is to generate synthetic data samples by over-sampling and introducing Gaussian noise to balance the distribution of the prediction targets. Under-sampling majority and over-sampling minority are commonly used strategies for tackling imbalanced problems [27]. Since the available dataset recording historical forest fire condition is typically small, under-sampling is not preferred since it discards a valuable proportion of data. Over-sampling generates new data samples by randomly copying minority samples. We introduced Gaussian noise to the copied samples to increase the diversity of generation. Since DNN is capable of extracting useful features from a large amount of data, the generation of synthetic samples allows improvement of its performance.

The proposed method is evaluated using a real-world dataset provided by [5], which contains environmental conditions of forest fires and the corresponding burned size collected from the Montesinho Natural Park of Portugal. Experimental results show that the prediction accuracy of the burned area of large-scale fires, which constitutes a minority in the dataset, was improved by adopting the proposed method. This is particularly important to prevent serious consequences of large-scale fires. To the best of our knowledge, a sparse

autoencoder-based DNN has not been used for forest fire prediction, and the imbalanced problem is seldom considered in this task.

The objective of this work was to develop a method able to predict a forest fire. We assume the availability of a set of historical data, composed of many records of small-scale forest fires and few records of large-scale ones, which is common considering the occurrence of small-scale forest fire is much more frequent. Specifically, we assumed the availability of N records of forest fires, which hereafter are called data samples, $(x_i, y_i)_{i=1, \dots, N}$, where x is a vector containing K numerical variables associated with the environmental condition of the forest fire, e.g., weather measurements and metrics that are computed based on satellite images, and y is the corresponding severity of the forest fires. The forest fire prediction method receives the test vector x^{TEST} as the input, containing the environmental variables collected at a certain location and time, and is required to provide a prediction y^{TEST} . The prediction is better when it is closer to the true value y^{TRUE} .

2. Materials and Methods

2.1. Benchmark Data for Forest Fire Prediction

We considered forest fire data collected during 2000–2003 from the Montesinho Natural Park of Portugal. The dataset contains 517 records of forest fires, whose environmental condition is described using 12 numerical variables, and severity is represented by the burned area measured in ha. Table 1 shows the numerical variables. The Natural Park is divided into 81 subareas using a 9×9 grid; therefore, the x and y coordinates indicate a certain subarea. Considering that 517 data samples are not enough to show the relevance of 81 subareas to the forest fire, the coordinates are not used for the prediction of burned area of the forest fire. The “month” and “day” are transformed to numerical data by denoting January to December as 1 to 12, and Monday to Sunday as 1 to 7, respectively. For variables, each is normalized into the scale $[0, 1]$, with $K = 10$ as the dimensional input vector x for the prediction of burned area. More details of the dataset can be found in [5].

Table 1. Variables describing the occurring condition of the forest fire.

Variable	Description
X	x -axis coordinate (1 to 9)
Y	y -axis coordinate (1 to 9)
month	Month of the year (January to December)
day	Day of the week (Monday to Sunday)
FFMC	Moisture content of surface litter (18.7 to 96.2)
DMC	Moisture content of shallow organic layer (1.1 to 291.3)
DC	Moisture content of deep organic layer (7.9 to 860.6)
ISI	Score correlating with fire velocity spread (0 to 56.1)
temp	Temperature in °C (2.2 to 33.3)
RH	Relative humidity in % (15 to 100)
wind	Wind speed in km/h (0.4 to 9.4)
rain	Rain in mm/m ² (0 to 6.4)

A histogram of burned area is shown in Figure 1a. Notice that the burned area is obviously imbalanced and that the number of small-scale fires is much greater than that of large-scale fires. In the dataset, there are 247 samples with a zero burned area indicating that the burned area is lower than 0.01 ha.

To ease the unbalanced problem, a logarithm transformation was applied to the original burned area values:

$$y = \frac{\ln(y_{raw,max} + 1) - \ln(y_{raw} + 1)}{\ln(y_{raw,max} + 1) - \ln(y_{raw,min} + 1)} \quad (1)$$

where y_{raw} represents the original burned area, $y_{raw,max}$ and $y_{raw,min}$ are maximum and minimum of the burned area, respectively, and y is the transformed burned area, which is

normalized into the scale $[0, 1]$ to be used as the output for the prediction task, as shown in Figure 1b.

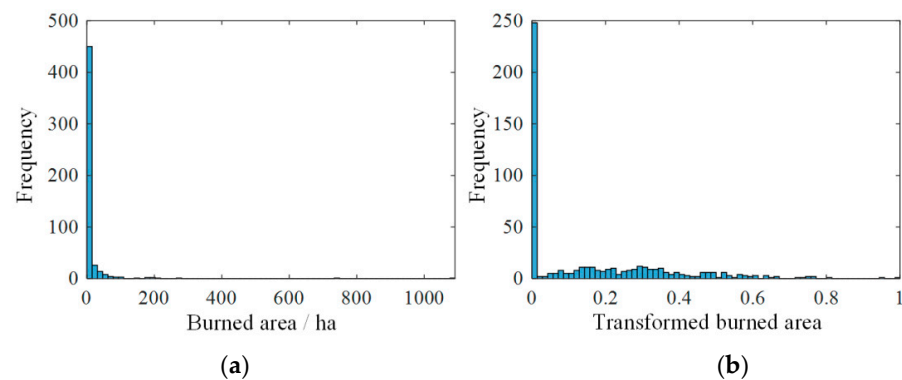


Figure 1. Histogram of (a) original burned area; (b) transformed burned area using logarithm transformation and normalization.

2.2. Sparse Autoencoder

The sparse autoencoder [26] aims at extracting useful features from the K -dimensional input of N training samples x_i , $i = 1, \dots, N$. As shown in Figure 2, the encoder extracts a feature vector $q_i = [q_{i,1}, \dots, q_{i,K_1}]$ from the input vector x_i as follows:

$$q_i = f_1(W_1 x_i + b_1) \quad (2)$$

where f_1 , W_1 , b_1 are the activation function, the weight matrix and the bias vector of encoder, respectively. Then, the decoder reconstructs x_i to \hat{x}_i based on q_i :

$$\hat{x}_i = f_2(W_2 q_i + b_2) \quad (3)$$

where f_2 , W_2 , b_2 are the activation function, the weight matrix and the bias vector of decoder, respectively. The training of sparse autoencoder aims at minimizing the following loss function to encourage the extraction of discriminative features:

$$L = T_{re} + \eta_1 T_{sparse} + \frac{\eta_2}{2} T_{L2} \quad (4)$$

where T_{re} , T_{sparse} and T_{L2} are terms with respect to reconstruction error, the sparsity regularization and the L_2 regularization, respectively, and η_1 and η_2 are coefficients. T_{re} measures how much the reconstruction \hat{x}_i is close to the input x_i :

$$T_{re} = \frac{1}{N} \sum_{i=1}^N \|x_i - \hat{x}_i\|^2 \quad (5)$$

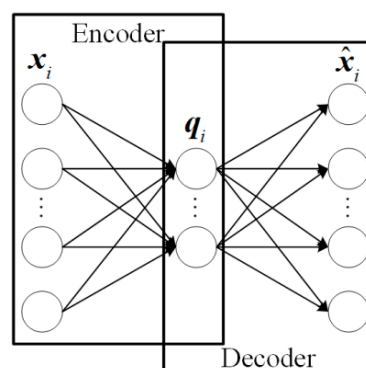


Figure 2. Sparse autoencoder.

T_{sparse} is used to constrain the hidden neurons to be inactive most of the time in order to extract discriminative features. Denote the mean activation of the j -th hidden neuron, $j = 1, \dots, K_1$, over all the samples x_i , $i = 1, \dots, N$ as:

$$\hat{p}_j = \frac{1}{N} \sum_{i=1}^N q_{i,j} \quad (6)$$

Ideally the expected value of \hat{p}_j should be a small value, e.g., 0.05, since the activation is required to be at zero for most of the samples. T_{sparse} is computed using the Kullback-Leibler (KL) divergence function to evaluate whether \hat{p}_j is close to an expected value p :

$$T_{sparse} = \sum_{j=1}^{K_1} \text{KL}(p \parallel \hat{p}_j) = \sum_{j=1}^{K_1} \left[p \log \frac{p}{\hat{p}_j} + (1-p) \log \frac{1-p}{1-\hat{p}_j} \right] \quad (7)$$

The above function reaches zero, the minimal value, when all \hat{p}_j are equal to p .

The T_{L2} is used to constrain the weight values to prevent the network from overfitting:

$$T_{L2} = \| \mathbf{W}_1 \|^2 + \| \mathbf{W}_2 \|^2 \quad (8)$$

2.3. Deep Neural Networks

The DNN aims at constructing an empirical mapping function from a K -dimensional input space $\mathbf{X} \subset \mathbb{R}^K$ to an output space $\mathbf{Y} \subset \mathbb{R}$ based on N input-output samples $(x_i, y_i)_{i=1, \dots, N} \in \mathbf{X} \times \mathbf{Y}$.

The DNN is composed of an encoder and a regression layer (Figure 3). The encoder extracts high-level features from x_i using multiple hidden layers, and the regression layer provides a prediction \hat{y}_i of y_i .

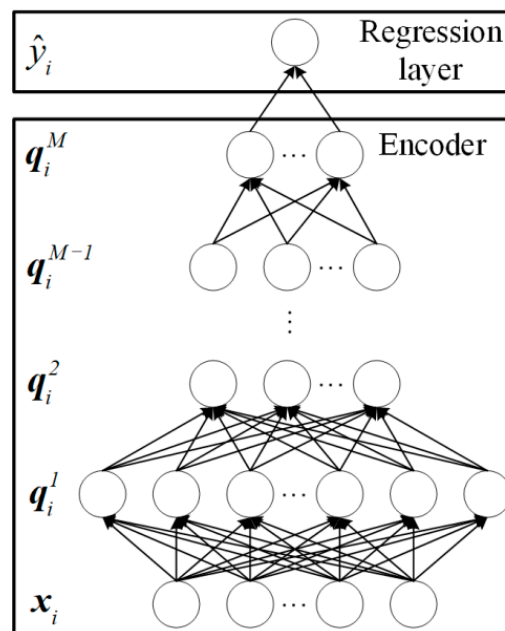


Figure 3. A deep neural network with M hidden layers.

Following the idea of the sparse autoencoder, T_{sparse} and T_{L2} defined in Equation (4) are applied to assist the encoder in extracting high-level features from x_i using M hidden layers. Denote the feature vectors progressively extracted from the hidden layers as q_i^1, \dots, q_i^M . Then, the dimension of q_i^1 , K_1 , is typically larger than the input dimension K to obtain a sparse-overcomplete feature vector, which was found to be capable of benefiting feature extraction of the following layers [28]. For the following hidden layers, the dimension

$K_m < K_{m-1}$, $m = 2, 3, \dots, M$, forces effective feature extraction. The ReLU activation function is adopted for all the layers of the encoder to allow the fast training of DNN [29].

The regression layer computes the prediction based on high-level feature q_i^M :

$$\hat{y}_i = f(w \cdot q_i^M + b) \quad (9)$$

where f is the sigmoid activation function, w and b are the weight matrix and bias of the regression layer, respectively. Then, the training of DNN is to minimize:

$$L_{DNN} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \eta_1 \sum_{m=1}^M T_{sparse}^m + \frac{\eta_2}{2} \|W\|^2 \quad (10)$$

where the first term is the mean squared prediction error, T_{sparse}^m is the sparsity regularization for the m -th hidden layer, the last term is the L_2 regularization.

2.4. Data Balancing Procedure

In order to deal with the imbalanced distribution of the output y , a data balancing procedure is proposed. It is described as following:

Step 1. Identify the range of the output $[y_{min}, y_{max}]$, and divide it equally into N_a non-overlapping intervals, i.e., $[y^{(0)} = y_{min}, y^{(1)}], [y^{(1)}, y^{(2)}], \dots, [y^{(N_a-1)}, y^{(N_a)} = y_{max}]$;

Step 2. Generate a random number y_{rand} from the uniform distribution $U_{[y_{min}, y_{max}]}$ and select the interval $[y^{(j-1)}, y^{(j)}]$ for synthetic sample generation, where $y^{(j-1)} < y_{rand} \leq y^{(j)}$, $j = 1, \dots, N_a$. Notice that each interval has the same possibility of selection due to random sampling from a uniform distribution, which helps to avoid over or under-estimating a specific interval;

Step 3. Randomly choose a sample (x_i, y_i) whose y_i is in the selected interval;

Step 4. Generate a synthetic sample $(x_i + n_x, y_i)$ by introducing the Gaussian noise n_x . There is a possibility of 10% that elements of n_x are all zeros, i.e., the synthetic sample is the same as the original sample, and a possibility of 90% that elements of n_x are randomly sampled from a Gaussian distribution $N(\mu=0, \sigma)$ to increase the diversity of the synthetic samples;

Step 5. Repeat steps 2–4 until N_{max} random samplings are done. Then, the obtained synthetic samples are used for training the DNN.

3. Results

The general step of applying the proposed method for forest fire prediction is shown in Figure 4.

Since the number of data samples is limited, a 10-fold cross-validation is adopted to evaluate the performance of the DNN. The dataset is randomly divided into 10 subsets, each containing approximately 10% of the samples. For each fold of the cross-validation, a subset is taken as the test set, and the other subsets are used as the training set for building the DNN model with the data balancing procedure. The predictions are obtained on the test set using the DNN and are inversely transformed to the original scale. The subsets are used as the test set in turn to obtain predictions on the whole dataset. The performance of the DNN is evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE):

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_{raw,i} - \hat{y}_{raw,i}| \quad (11)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{raw,i} - \hat{y}_{raw,i})^2} \quad (12)$$

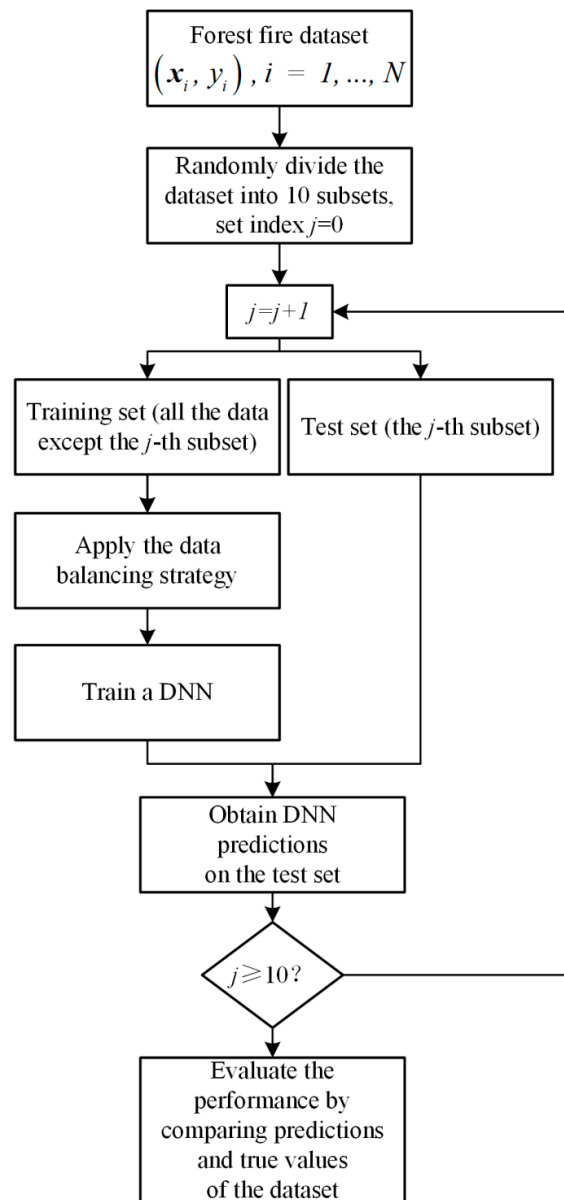


Figure 4. Procedure of the 10-fold cross-validation.

Lower values of these two metrics indicate better performance. Considering the random effect caused by dataset splitting, the 10-fold cross-validation is repeated 10 times. The average *MAE* and *RMSE* are computed as the final performance for DNN.

Figure 5 shows an example of applying the data balancing procedure (Section 2.4). The scaled range $[0, 1]$ of y is divided into $N_a = 100$ non-overlapping intervals with length 0.01. Figure 5a shows the number of samples with respect to the intervals, where the first interval contains much more samples than the others, i.e., the distribution of y is extremely imbalanced. Synthetic samples are generated by randomly choosing an interval, randomly choosing a sample from the interval, and adding Gaussian noise to the input of the sample. Since variables of x are scaled into $[0, 1]$, the standard deviation σ of noise distribution $N_{(\mu=0, \sigma)}$ is set to be 0.001 to slightly modify the original variables for the diversity of synthetic samples. $N_{max} = 20,000$ samplings are performed and approximately 200 synthetic samples are generated for each interval, as shown in Figure 5b. The number of synthetic samples is less than N_{max} since nothing is generated for the intervals which do not have any data sample.

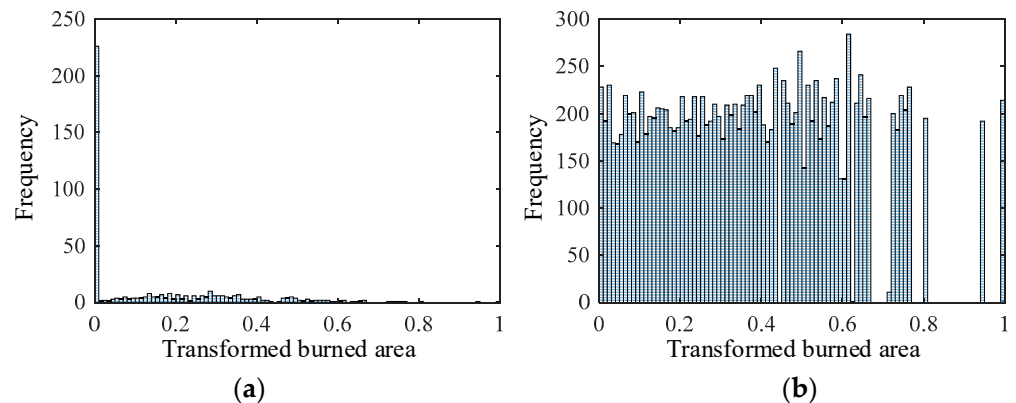


Figure 5. Histogram of targets of training set: (a) original; (b) after applying the data balancing procedure.

A DNN with $M = 3$ hidden layers is constructed. Its encoder structure is $K = 10$, $K_1 = 100$, $K_2 = 50$, $K_3 = 6$, which is set following the general principle discussed in Section 2.3. For the hyperparameters, $\eta_1 = 1$, $\eta_2 = 10^{-7}$ which is set by computing the magnitude ratio in the loss function to keep the balance, the expected sparsity p is chosen considering a possible set $\{0.01, 0.02, \dots, 0.2\}$ using an internal 10-fold grid search, i.e., for each fold of the cross-validation, first perform 10-fold cross-validations using only the training data for DNN with a given value of p , and then select the DNN with best-performing p on the training set to make predictions on the test set. The median value of selected p is 0.08. The obtained results are reported in Table 2 in terms of the mean and standard deviation of MAE and $RMSE$, which are computed based on 10 times of cross-validation.

Table 2. Segmented performance of different methods in form of MAE average and standard deviation ($RMSE$ average and standard deviation), the best performance in each interval is reported in **bold**.

Interval of y	Interval of y_{raw} (ha)	Methods			
		Proposed Method	ANN	SVM	RF
[0,0.1)	[0,1.01)	13.00 \pm 1.66 (24.52 \pm 7.75)	1.64 \pm 0.00 (1.81 \pm 0.00)	1.07 \pm 0.01 (1.12 \pm 0.01)	3.05 \pm 0.08 (5.17 \pm 0.40)
[0.1,0.2)	[1.01,3.05)	9.00 \pm 1.22 (16.55 \pm 4.03)	0.72 \pm 0.01 (0.86 \pm 0.01)	0.83 \pm 0.01 (0.98 \pm 0.01)	2.05 \pm 0.30 (3.95 \pm 1.00)
[0.2,0.3)	[3.05,7.16)	8.80 \pm 1.30 (15.69 \pm 2.93)	3.23 \pm 0.01 (3.58 \pm 0.01)	3.86 \pm 0.01 (4.07 \pm 0.01)	3.27 \pm 0.08 (3.69 \pm 0.08)
[0.3,0.4)	[7.16,15.42)	8.48 \pm 1.12 (12.67 \pm 2.40)	8.46 \pm 0.01 (8.74 \pm 0.01)	8.75 \pm 0.02 (9.09 \pm 0.01)	7.27 \pm 0.14 (7.86 \pm 0.11)
[0.4,0.5)	[15.42,32.04)	16.81 \pm 2.15 (20.55 \pm 6.69)	22.07 \pm 0.01 (22.77 \pm 0.01)	22.30 \pm 0.02 (23.01 \pm 0.01)	20.37 \pm 0.27 (21.50 \pm 0.27)
[0.5,0.6)	[32.04,65.51)	34.71 \pm 1.30 (36.69 \pm 1.02)	43.78 \pm 0.02 (44.72 \pm 0.02)	44.54 \pm 0.02 (45.47 \pm 0.02)	42.58 \pm 0.27 (43.42 \pm 0.32)
[0.6,0.7)	[65.51,132.88)	68.50 \pm 4.73 (71.58 \pm 4.19)	84.16 \pm 0.04 (85.15 \pm 0.04)	84.92 \pm 0.02 (85.89 \pm 0.02)	82.08 \pm 0.69 (82.95 \pm 0.74)
[0.7,0.8)	[132.88,268.48)	166.17 \pm 5.54 (169.45 \pm 4.64)	185.56 \pm 0.03 (186.53 \pm 0.02)	186.54 \pm 0.02 (187.49 \pm 0.02)	184.05 \pm 0.79 (185.04 \pm 0.82)
[0.8,0.9)	[268.48,541.43)	273.75 \pm 4.67 (273.75 \pm 4.67)	277.45 \pm 0.06 (277.45 \pm 0.06)	277.40 \pm 0.03 (277.40 \pm 0.03)	276.75 \pm 0.35 (276.75 \pm 0.35)
[0.9,1]	[541.43,1090.84)	724.80 \pm 10.83 (724.80 \pm 10.83)	744.10 \pm 0.05 (744.10 \pm 0.05)	745.16 \pm 0.10 (745.16 \pm 0.10)	744.30 \pm 0.68 (744.30 \pm 0.68)

4. Discussion

For comparison, the popular regression methods ANN, SVM and RF were used for forest fire prediction and their average *MAEs* and *RMSEs* of 10 times 10-fold cross-validation are computed. The ANN is typically a feedforward network with one hidden layer. The activation function of the hidden and output neurons is sigmoid. The number of hidden neurons, N_h , is selected by an internal 10-fold grid search considering the possible set $\{4, 6, \dots, 20\}$, the median value of selected N_h is 10. The SVM maps the input variables into a high dimensional space, and then finds the best linear hyperplane for regression with the support of a nonlinear kernel function. The regularization parameter, C , is selected by an internal 10-fold grid search considering the possible set $\{0.01, 0.1, 1, 10, 100\}$. The median value of selected C is 1. A Random Forest (RF) was constructed by averaging the outputs of multiple decision trees, each trained using training samples random selected by bootstrap technique. The number of trees, N_{tree} , was selected by an internal 10-fold grid search considering the possible set $\{50, 100, \dots, 500\}$, the median value of selected N_{tree} is 200. All experiments were conducted using a computer with Intel i7-8550U CPU, 8.00 GB RAM, Windows 10 OS. The DNN was built using the Keras 2.0 framework, and the other models for comparison were constructed using the scikit-learn 0.24.2 package. The obtained results are shown in Table 2 in terms of the mean and standard deviation of *MAE* and *RMSE*. The metrics were computed regarding 10 intervals of the burned area to investigate the prediction accuracy of the methods for different scales of forest fire.

In Table 2, the ANN, SVM and RF give better performance for small-scale forest fire whose burned area is less than 15.42 ha. The proposed method outperforms the other methods when the scale is larger indicating that the DNN successfully pays more attention to the large-scale forest fires with the support of data balancing procedure.

To better understand the performance of the methods, Figure 6 shows prediction results with respect to one fold of the cross-validation. Since the ANN, SVM and RF behave similarly, and to make the figure clear, only the SVM which has the best performance on the smallest interval is shown for comparison. Since in the dataset there are more than half of the samples with zero burned areas, the SVM actually learns always to provide a very small value no matter what the input is. To further investigate the behavior of different methods, Figure 7 shows the histogram of their predictions obtained from one fold of prediction. It is verified that SVM, ANN and RF always provide small predictions regardless of the input variables. This means that SVM, ANN and RF are not reliable; they ignore or extremely underestimate all large-scale fires.

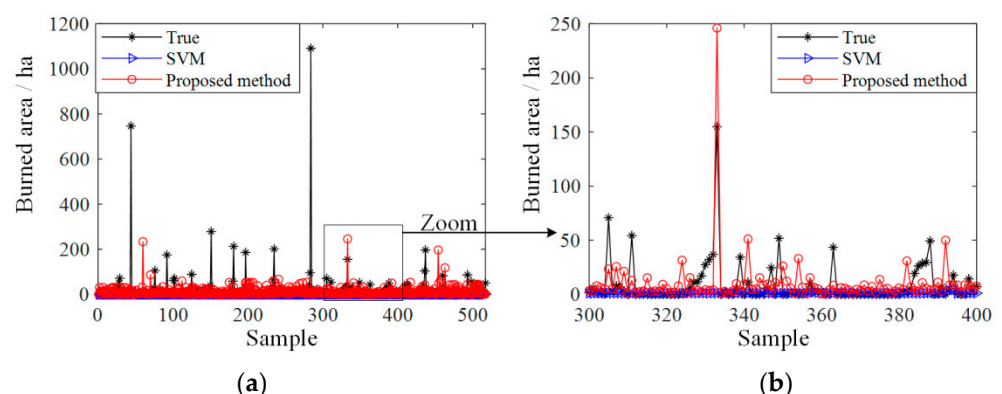


Figure 6. An example of prediction results: (a) obtained from one fold of cross-validation; (b) local zoom.

Different from classical methods that don't learn useful information, the proposed method pays attention to all fire scales. For most small-scale forest fires, the proposed method provides close predictions, and the relatively larger prediction error is mainly due to some occasional large predictions (Figure 6). For large-scale fires, the proposed

method attempts closer predictions though these are still not very accurate. The prediction histogram of the proposed method (Figure 7) is closer to the original data distribution (Figure 1a), indicating the proposed method facilitates the extraction of correlation from input variables to the output. However, the performance of the proposed method is limited by the lack of enough information, since synthetic data, generated by adding Gaussian noise to original data, can help the training of DNN but do not provide any new information different from the original data. Therefore, the proposed method cannot fully detect the causes of large-scale fires because the information is insufficient, leading to wrong over-estimation of some small-scale fires. As a result, it shows a trade-off between improving large-scale prediction accuracy and over-estimating small-scale fires (Table 2).

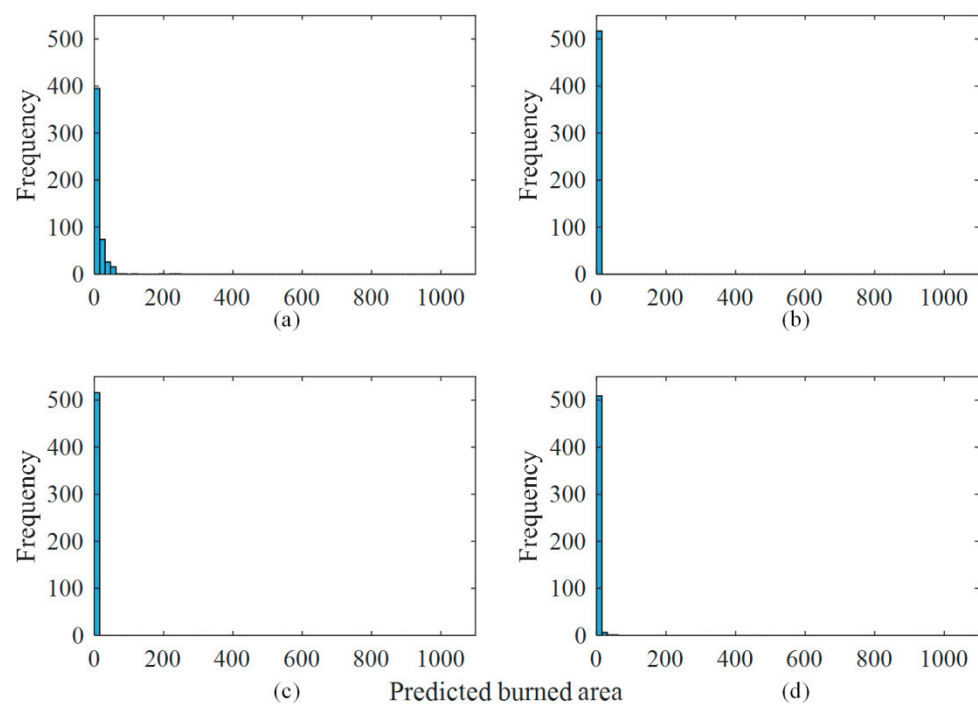


Figure 7. Histogram of predicted burned area provided by: (a) the proposed method; (b) SVM; (c) ANN; (d) RF, respectively.

The over-estimation of small-scale fires may cause overreaction and extra costs. However, considering that a large-scale fire can cause serious consequences and huge losses, its accurate prediction usually is more important. We expect that the performance of the proposed method can be further improved if more data containing more information are collected.

In Table 2, notice that performance variations of the proposed method are larger than those of SVM, ANN and RF. The variations of MAE and RMSE are computed based on the results obtained from 10 cross-validations. Heavily affected by the imbalanced data, ANN, SVM and RF always provide very small and stable values in all the 10 cross-validations, resulting in small variations. Performance variation of the proposed method among cross-validations is mainly due to the change in the training set (Figure 4). Since the dataset is relatively small (517 records), the random division of subsets makes the training set very different in cross-validations. When more data are collected, the effect of random data division can be weakened to reduce the variation of the proposed method. The large variations actually indicate that the proposed method tends to learn useful information in all the cross-validations.

To further investigate why it is difficult for SVM, ANN and RF to learn useful patterns for fire prediction, the trends of forest fires by input variables are shown in Figure 8. With respect to month, large-scale fires tend to occur in summer (July to September), which is

consistent with common sense. The day of the week seems to not strongly influence fire occurrence, and large-scale fires happened in all the days except Friday. By definition, FFMC, DMC, DC and ISI suggest more severe burning conditions with larger values [5]. From the collected data, FFMC and DC behave in accordance with the definition, and DMC and ISI are more likely to indicate severe burning conditions in the middle of their ranges. The remaining four variables are the most intuitive. Large-scale fires are likely to be driven by high temperature, low RH (relative humidity), wind with speed of about 2–6 km/h and a small amount of rain. However, notice that the variable values more likely leading to severe burning can also suggest small-scale fires, and variables correlating nonlinearly with the burning area and can have various value combinations. Thus, it is difficult to explicitly extract rules for fire prediction from the data. A method able to automatically extract useful features from the data is required. The ANN, SVM and RF rely on high-quality features strongly correlated with the output as the input, whereas the proposed method based on DNN can further extract useful features for prediction from the input variables.

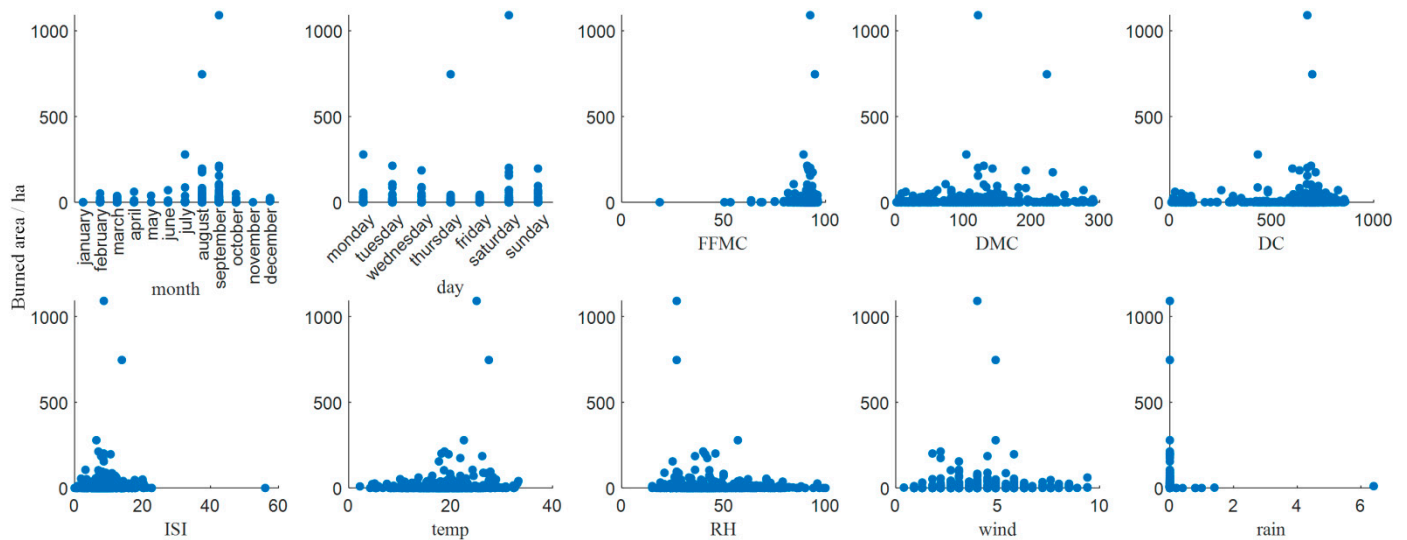


Figure 8. Trends of the burned area and input variables.

In summary, SVM, ANN and RF cannot make meaningful predictions due to the complexity of the mapping between input variables and the burned area. The proposed method performs relatively well in the prediction of large-scale forest fires. However, it sometimes over-estimates small-scale fires and its prediction variation is relatively large. If more data are available, the superiority of the proposed method based on DNN is expected to be more obvious and overcome its drawbacks. Fortunately, more and more forest monitoring data can be collected in the big data era, and the proposed method is a promising tool to be considered for forest fire prediction.

5. Conclusions

This work proposed a new method based on a sparse autoencoder-based DNN and data balancing procedure for forest fire prediction using numerical environmental variables. Its main contributions are: (i) it employs a DNN model with sparse regularization, which can automatically extract features from a large amount of data without requiring expert intervention for feature selection, and (ii) it develops a data balancing procedure to tackle the problem of imbalanced dataset.

The forest fire data collected from the Montesinho Natural Park of Portugal is used to investigate the performance of the proposed method. The dataset is seriously imbalanced when more than half of the sample outputs are zeros. The results show that the proposed method outperforms ANN, SVM and RF for the prediction of large-scale forest fires. The prediction error on small-scale forest fires is mainly due to some occasionally

large predictions. Prediction of forest fires is a challenging task. It is expected that the prediction performance could be further improved if additional information and more data are available, given the capability of DNN to extract useful information from a large amount of data.

Author Contributions: Conceptualization, C.L. and B.L.; methodology, C.L.; validation, W.G. and X.L.; formal analysis, S.Z.; investigation, B.L.; writing—original draft preparation, C.L.; writing—review and editing, S.Z. and B.L.; visualization, W.G.; funding acquisition, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Foundation for young talents in Zhongkai University of Agriculture and Engineering, grant number KA210319252; Key-Area Research and Development Program of Guangdong Province, grant number 2020B020215003.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found here: <http://archive.ics.uci.edu/ml/datasets/Forest+Fires> (accessed on 29 February 2008).

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

References

1. Resco de Dios, V.; Nolan, R.H. Some Challenges for Forest Fire Risk Predictions in the 21st Century. *Forests* **2021**, *12*, 469. [CrossRef]
2. Vásquez, F.; Cravero, A.; Castro, M.; Acevedo, P. Decision Support System Development of Wildland Fire: A Systematic Mapping. *Forests* **2021**, *12*, 943. [CrossRef]
3. Brushlinsky, N.N.; Ahrens, M.; Sokolov, S.V.; Wagner, P. World Fire Statistics. *Cent. Fire Stat.* **2016**, *10*.
4. Monjarás-Vega, N.A.; Briones-Herrera, C.I.; Vega-Nieva, D.J.; Calleros-Flores, E.; Corral-Rivas, J.J.; López-Serrano, P.M.; Pompa-García, M.; Rodríguez-Trejo, D.A.; Carrillo-Parra, A.; González-Cabán, A.; et al. Predicting Forest Fire Kernel Density at Multiple Scales with Geographically Weighted Regression in Mexico. *Sci. Total Environ.* **2020**, *718*, 137313. [CrossRef] [PubMed]
5. Cortez, P.; Morais, A. A Data Mining Approach to Predict Forest Fires Using Meteorological Data. In *New Trends in Artificial Intelligence, Proceedings of the 13th EPIA 2007—Portuguese Conference on Artificial Intelligence, Guimarães, Portugal, December 2007*; Neves, J., Santos, M.F., Machado, J., Eds.; pp. 512–523. APPIA, 2007; ISBN 13 978-989-95618-0-9. Available online: <http://www3.dsi.uminho.pt/pcortez/fires.pdf> (accessed on 26 May 2022).
6. Joseph, M.B.; Rossi, M.W.; Mielkiewicz, N.P.; Mahood, A.L.; Cattau, M.E.; St. Denis, L.A.; Nagy, R.C.; Iglesias, V.; Abatzoglou, J.T.; Balch, J.K. Spatiotemporal Prediction of Wildfire Size Extremes with Bayesian Finite Sample Maxima. *Ecol. Appl.* **2019**, *29*, e01898. [CrossRef]
7. Sayad, Y.O.; Mousannif, H.; Al Moatassime, H. Predictive Modeling of Wildfires: A New Dataset and Machine Learning Approach. *Fire Saf. J.* **2019**, *104*, 130–146. [CrossRef]
8. Jain, P.; Coogan, S.C.P.; Subramanian, S.G.; Crowley, M.; Taylor, S.; Flannigan, M.D. A Review of Machine Learning Applications in Wildfire Science and Management. *Environ. Rev.* **2020**, *28*, 478–505. [CrossRef]
9. Phelps, N.; Woolford, D.G. Comparing Calibrated Statistical and Machine Learning Methods for Wildland Fire Occurrence Prediction: A Case Study of Human-Caused Fires in Lac La Biche, Alberta, Canada. *Int. J. Wildl. Fire* **2021**, *30*, 850–870. [CrossRef]
10. Stojanova, D.; Panov, P.; Kobler, A.; Džeroski, S.; Taškova, K. Learning to Predict Forest Fires with Different Data Mining Techniques. In *Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD 2006)*, Ljubljana, Slovenia, 17 October 2006; pp. 255–258.
11. Bisquert, M.; Caselles, E.; Sánchez, J.M.; Caselles, V. Application of Artificial Neural Networks and Logistic Regression to the Prediction of Forest Fire Danger in Galicia Using MODIS Data. *Int. J. Wildl. Fire* **2012**, *21*, 1025–1029. [CrossRef]
12. Castelli, M.; Vanneschi, L.; Popović, A. Predicting Burned Areas of Forest Fires: An Artificial Intelligence Approach. *Fire Ecol.* **2015**, *11*, 106–118. [CrossRef]
13. Anezakis, V.-D.; Demertzis, K.; Iliadis, L.; Spartalis, S. A Hybrid Soft Computing Approach Producing Robust Forest Fire Risk Indices. In *Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations*, Thessaloniki, Greece, 16–18 September 2016; pp. 191–203.
14. West, A.M.; Kumar, S.; Jarnevich, C.S. Regional Modeling of Large Wildfires under Current and Potential Future Climates in Colorado and Wyoming, USA. *Clim. Chang.* **2016**, *134*, 565–577. [CrossRef]
15. Al Janabi, S.; Al Shourbaji, I.; Salman, M.A. Assessing the Suitability of Soft Computing Approaches for Forest Fires Prediction. *Appl. Comput. Inform.* **2018**, *14*, 214–224. [CrossRef]
16. Nebot, A.; Mugica, F. Forest Fire Forecasting Using Fuzzy Logic Models. *Forests* **2021**, *12*, 1005. [CrossRef]
17. Song, Y.; Wang, Y. Global Wildfire Outlook Forecast with Neural Networks. *Remote Sens.* **2020**, *12*, 2246. [CrossRef]

18. Chen, J.; Wang, X.; Yu, Y.; Yuan, X.; Quan, X.; Huang, H. Improved Prediction of Forest Fire Risk in Central and Northern China by a Time-Decaying Precipitation Model. *Forests* **2022**, *13*, 480. [[CrossRef](#)]
19. Hagmann, R.K.; Hessburg, P.F.; Salter, R.B.; Merschel, A.G.; Reilly, M.J. Contemporary Wildfires Further Degrade Resistance and Resilience of Fire-Excluded Forests. *For. Ecol. Manag.* **2022**, *506*, 119975. [[CrossRef](#)]
20. Diaz-Avalos, C.; Juan, P.; Serra-Saurina, L. Modeling Fire Size of Wildfires in Castellon (Spain), Using Spatiotemporal Marked Point Processes. *For. Ecol. Manag.* **2016**, *381*, 360–369. [[CrossRef](#)]
21. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
22. Zhang, Q.; Lin, G.; Zhang, Y.; Xu, G.; Wang, J. Wildland Forest Fire Smoke Detection Based on Faster R-CNN Using Synthetic Smoke Images. *Procedia Eng.* **2018**, *211*, 441–446. [[CrossRef](#)]
23. Saponara, S.; Elhanashi, A.; Gagliardi, A. Real-Time Video Fire/Smoke Detection Based on CNN in Antifire Surveillance Systems. *J. Real-Time Image Process.* **2021**, *18*, 889–900. [[CrossRef](#)]
24. Cao, Y.; Yang, F.; Tang, Q.; Lu, X. An Attention Enhanced Bidirectional LSTM for Early Forest Fire Smoke Recognition. *IEEE Access* **2019**, *7*, 154732–154742. [[CrossRef](#)]
25. Zhang, D.; Yin, J.; Zhu, X.; Zhang, C. Network Representation Learning: A Survey. *IEEE Trans. Big Data* **2018**, *6*, 3–28. [[CrossRef](#)]
26. Ng, A. Sparse Autoencoder. *CS294A Lect. Notes* **2011**, *72*, 1–19.
27. Mohammed, R.; Rawashdeh, J.; Abdullah, M. Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results. In Proceedings of the 2020 11th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 7–9 April 2020; pp. 243–248.
28. Ahmed, H.O.A.; Wong, M.L.D.; Nandi, A.K. Intelligent Condition Monitoring Method for Bearing Faults from Highly Compressed Measurements Using Sparse Over-Complete Features. *Mech. Syst. Signal Process.* **2018**, *99*, 459–477. [[CrossRef](#)]
29. Agarap, A.F. Deep Learning Using Rectified Linear Units (Relu). *arXiv* **2018**, arXiv:1803.08375.