

## Article

# A Well-Overflow Prediction Algorithm Based on Semi-Supervised Learning

Wei Liu <sup>1</sup>, Jiasheng Fu <sup>1</sup>, Yanchun Liang <sup>2,3</sup>, Mengchen Cao <sup>2</sup> and Xiaosong Han <sup>2,\*</sup> 

<sup>1</sup> CNPC Engineering Technology R&D Company Limited, Beijing 102206, China; liuweidri@cnpc.com.cn (W.L.); fujsdr@cnpc.com.cn (J.F.)

<sup>2</sup> Key Laboratory for Symbol Computation and Knowledge Engineering of National Education Ministry, College of Computer Science and Technology, Jilin University, Changchun 130012, China; ycliang@jlu.edu.cn (Y.L.); caomc21@mails.jlu.edu.cn (M.C.)

<sup>3</sup> Zhuhai Laboratory of Key Laboratory for Symbol Computation and Knowledge Engineering of Ministry of Education, Zhuhai College of Science and Technology, Zhuhai 519041, China

\* Correspondence: hanxiaosong@jlu.edu.cn

**Abstract:** Oil drilling is the core process of oil and natural gas resources exploitation. Well overflow is one of the biggest threats to safety drilling. Prediction of the overflow in advance can effectively avoid the occurrence of this kind of accident. However, the drilling history has unbalanced distribution, and labeling data is a time-consuming and laborious job. To address this issue, an overflow-prediction algorithm based on semi-supervised learning is designed in this paper, which can accurately predict overflow 10 min in advance when the labeled data are limited. Firstly, a three-step feature-selection algorithm is conducted to extract 22 features, and the time series samples are constructed through a 500-width sliding window with step size 1. Then, the Mean Teacher model with Jitter noise is employed to train the labeled and unlabeled data at the same time, in which a fused CNN-LSTM network is built for time-series prediction. Compared with supervised learning and other semi-supervised learning frameworks, the results show that the proposed model based on only 200 labeled samples is able to achieve the same effect as supervised learning method using 1000 labeled samples, and the prediction accuracy can reach 87.43% 10 min in advance. With the increase in the proportion of unlabeled samples, the performance of the model can sustain a rise within a certain range.

**Keywords:** oil drilling; overflow; time series prediction; deep learning; semi-supervised learning



**Citation:** Liu, W.; Fu, J.; Liang, Y.; Cao, M.; Han, X. A Well-Overflow Prediction Algorithm Based on Semi-Supervised Learning. *Energies* **2022**, *15*, 4324. <https://doi.org/10.3390/en15124324>

Academic Editor: Reza Rezaee

Received: 5 May 2022

Accepted: 11 June 2022

Published: 13 June 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

As the “blood” of modern industry, oil is important primary energy. It not only plays an important role in basic necessities but also works as an indispensable strategic resource for national survival and development that promotes the economy and safeguards security. Drilling is a key step in oil and gas exploitation, in which overflow is one of the greatest threats to the safety of the operation. If it is not handled properly, the overflow will evolve into a blowout, resulting in wellbore scrapping, which will not only cause great economic losses but also endanger the lives, property, and safety of drilling workers and surrounding people. The most effective prevention approach is the early detection of overflow. Therefore, predicting the occurrence of overflow based on real-time drilling data can strive for precious time control overflow, to reduce safety risks timely and effectively.

In the traditional oil drilling technique, overflow is usually judged by the drilling engineers on the ground with relevant instrument data, by analyzing the changes in drilling feature parameters, such as standpipe pressure, inlet and outlet flow difference, and so on. However, artificial judgment highly depends on the experience of engineers, and it brings great work pressure to the engineers. With the development of machine learning technology, more and more scholars construct machine-learning models to predict overflow risk. Hargreaves et al., (2001) analyzed deep-sea acoustic data to monitor overflow by

Bayesian model and calculated the probability of overflow [1]. Lian (2013) fused a rough set and support vector machine (RS-SVM) to monitor the occurrence of overflow [2]. Lind et al., (2014) proposed a radial basis function (RBF) neural network based on the k-means clustering algorithm to predict drilling risk [3]. Li et al., (2015) put forward a prediction method of the overflow based on the fuzzy expert system [4]. Liang et al., (2018) proposed a fuzzy multilevel algorithm based on Particle swarm optimization (PSO) to optimize Support vector regression machine (SVR), and realized real-time dynamic evaluation of drilling risk [5]. Liang et al., (2019) established a model for overflow diagnosis based on the monitoring of standpipe pressure and casing pressure in pressure wave transmission with the genetic algorithm and BP neural network (GA-BP). In this model, the genetic algorithm was used to accelerate the convergence speed of neural networks and avoid falling into the local extremum. The early diagnosis of drilling overflow was realized, and the misjudgment rate of drilling overflow was reduced [6]. In the same year, based on the correlation between overflow accidents and the trend of casing pressure, Liang et al., proposed an intelligent early warning method for drilling overflow accidents based on an improved DBSCAN clustering method. The early warning method used the idea of time-series scanning and hierarchical rule clustering to improve the speed and accuracy of clustering [7]. Zhu et al., (2019). collected data such as geological lithology, designed well structure, real-time drilling fluid performance, rock physical properties of backflow cuttings, and drilling engineering parameters to build an artificial neural network to predict the risk probability of stuck pipe [8]. Sergey Borozdin et al., (2020) used deep learning method and created a drilling simulator, which makes it possible to recreate a digital twin of a real well and simulate an almost unlimited number of complications of various kinds on it [9]. Mohammad Sabah et al., (2020) combined a number of heuristic search algorithms including genetic algorithm (GA), particle swarm size (PSO), and cuckoo search algorithm (COA), with multilayer perception (MLP) neural network and least square support vector machine (LSSVM) to present different hybrid algorithms in the prediction of lost circulation [10]. Liu et al., (2021) developed a dynamic Bayesian network to create a dynamic risk assessment model for evaluating the safety of deep-water drilling operations [11]. In the same year, Yin et al., applied a similar method to risk analysis of offshore blowout [12], and Liang et al., established a random forest overflow accident identification and classification model based on bat algorithm optimization [13]. Wang et al., (2022) proposed a drilling identification method based on optimized SVM [14].

According to the above literature, machine-learning and deep-learning models, such as support vector machine (SVM), artificial neural network, long-term and short-term memory network (LSTM), and so on, become the main steam to predict overflow. The accuracy of these supervised learning-based methods highly depends on a large number of labeled training data. In practice, drilling data produced by one well is massive, and labeling data manually is time-consuming and heavily dependent on the experience of engineers. Besides, overflow data is very rare. The generalization ability limits the application of the above models in drilling engineering. Therefore, to solve the problem of the small amount of labeled data and a large number of unlabeled data, a semi-supervised learning model is proposed that can predict overflow with limited label data.

## 2. Related Work

### 2.1. Semi-Supervised Learning

Semi-supervised learning (SSL) is a kind of learning method that combines supervised learning with unsupervised learning. SSL model is built with a small number of labeled samples and a large number of unlabeled samples. In practice, collecting labeled samples is often difficult, expensive, and time-consuming in practical conditions, while unlabeled samples are easy to obtain. In this case, SSL is more suitable for the application, since SSL can effectively utilize the unlabeled data to improve the model performance. The application of the SSL results from the hypothesis of the model. When the model hypothesis

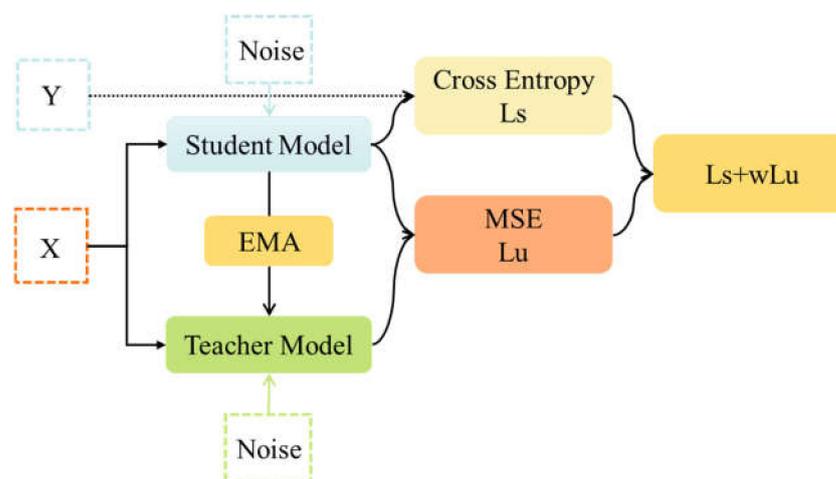
is established, unlabeled data can improve the learning performance of the model with a high probability and vice versa. The main SSL assumptions are as follows [15]:

1. Smoothing hypothesis.
2. When two samples are very close in the high-density data region, their class labels are likely to be the same; on the contrary, when the low-density data regions divide the two samples, they are likely to have different class labels.
3. Clustering hypothesis.
4. When two samples belong to the same cluster, their class labels are probably the same. This hypothesis is also named as low-density separation hypothesis, which means that the classification decision surface should be located in the low-density data region instead of the high-density data area. The decision surface should not divide the samples from the same high-density data area into both sides of the surface.
5. Manifold hypothesis.
6. On the one hand, in high-dimensional space, the data volume increases exponentially as the dimension increases, so it is difficult to estimate the real data distribution. On the other hand, if the input data is on some low-dimensional manifold, a low-dimensional representation could be found by unlabeled data, and then the simplified task will be fulfilled with labeled data. Therefore, the manifold hypothesis maps the high-dimensional data to the low-dimensional manifold, and if the two samples are located in the local neighborhood of the low-dimensional manifold, their class labels are likely to be the same.

It is worth noting that, when the semi-supervised model hypothesis is not valid, unlabeled data will actually degrade the learning performance of the model.

## 2.2. Mean Teacher Algorithm

The Mean Teacher [16] method is a kind of consistent regularization method based on the smoothing hypothesis. The main idea of the Mean Teacher model is to reduce the over-fitting problem of the neural network through the consistent regularization method of unlabeled data, that is, the model can be trained to consistently predict a given unlabeled data and its perturbed data. The structure of the Mean Teacher algorithm is shown in Figure 1.



**Figure 1.** Scheme 16.

As shown in Figure 1, Mean Teacher consists of a student model and a teacher model, which share the same framework based on supervised learning, but the parameters of both models are different. The parameters  $\theta$  represent the student model, and the parameters  $\theta'$  represent the teacher model.  $f_{\theta}$  represents the output of the student model, while  $f_{\theta'}$  represent the output of the teacher model. In each training iteration, the same sample

with different noise interference is input into the student model and the teacher model, the disturbance of the student model is  $\eta$ , and the disturbance of the teacher model is  $\eta'$ .

The purpose of calculating the classification cross-entropy loss ( $L_s$ ) between the prediction label of the student model output and the real label of the sample is to ensure the data-fitting of the labeled sample. The purpose of calculating the consistency loss ( $L_u$ ) between the prediction labels of the student and the teacher model is to preserve the similarity between the prediction labels of the student and the teacher model under different noise disturbances. The whole loss function is obtained by weighting the cross-entropy classification loss and consistency loss, and then the parameter weights of the student model are updated by back propagation. In the training phase, the classification cross-entropy loss and consistency loss are computed simultaneously for the labeled data, while the classification cross-entropy loss is not used for the unlabeled data. The teacher model is not trained by back propagation directly, but Exponential Moving Average (EMAs) as Equation (1) is conducted on the parameters of the student model to update the parameters of the teacher model as the parameter of the teacher model, where  $N$  is the periodic size. The overall loss function of the Mean Teacher model is constructed as Equation (2), where  $H(y, f_\theta(x))$  is the cross-entropy of the student model.

$$\theta'_t = a\theta'_{t-1} + (1-a)\theta_t, a = \frac{N-1}{N+1} \quad (1)$$

$$L = \frac{1}{|D_l|} \sum_{x,y \in D_l} H(y, f_\theta(x)) + w \frac{1}{|D_u|} \sum_{x \in D_u} d_{MSE}(f_\theta(x), f_{\theta'}(x)) \quad (2)$$

The specific training process of Mean Teacher is shown in Algorithm 1.

---

**Algorithm 1.** Mean Teacher learning algorithm.

---

1. The labeled data after disturbance  $\eta$  is input into the student model, and then the classification cross-entropy loss between the prediction label and the real label of the training set is calculated.
  2. All the data after perturbed  $\eta'$ , including labeled data and unlabeled data, are input into the student model and the teacher model, and the consistency loss between the prediction label of the student model and the teacher model is calculated.
  3. According to Equation (2), parameter weights of the student model are updated by back propagation.
  4. According to Equation (1), the exponential moving average of the parameters of the student model is taken as the parameter of the teacher model.
  5. Repeat the above process until the network converges.
- 

### 3. Dataset and Feature Selection

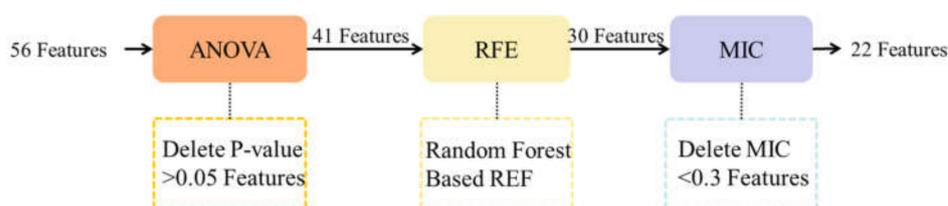
In this study, 10 overflow samples have been collected based on the historical drilling data from one real well in an oil field. Each sample consists of the Well Logging, Pressure While Drilling (PWD), and Managed Pressure Drilling (MPD) data around the overflow once per second, with a total of 56 features, as shown in Table 1.

In this paper, there are three steps to select features to obtain the best feature subset—Analysis of Variance (ANOVA), Recursive Feature Elimination (RFE), and Mutual Information Coefficient (MIC)—which can ensure the maximum classification accuracy in the subsequent process [17]. ANOVA is used to filter features with small variance. RFE is a greedy algorithm for selecting the best subset of features; the main principle is to build a machine-learning model constantly, delete the worst features based on the weight of the model, and iterate this process repeatedly until all features are traversed. The redundant relationship between features is not considered in the above two feature selection methods. MIC method is used to capture the relationship between each feature and label to further

screen features. The feature selection process is shown in Figure 2 and the final selected features are shown as Table 2.

**Table 1.** Descriptive statistics of data.

No.	Time	Total Number	Samples before Overflow	Overflow Samples
1	10.11 10:53–10.11 11:52	3187	2608	579
2	10.11 9:30–10.11 10:53	4392	3169	1223
3	10.11 10:53–10.11 11:52	3283	2645	638
4	10.29 6:14–10.29 7:55	5375	3221	2154
5	10.29 0:45–10.29 2:15	4800	3172	1628
6	10.28 16:39–10.28 18:14	4973	3867	1106
7	10.28 12:41–10.28 14:13	4834	2724	2110
8	10.24 1:56–10.24 2:25	1595	956	639
9	10.24 0:28–10.24 1:55	4663	3181	1482
10	10.22 23:49–10.23 0:50	6524	3180	3344



**Figure 2.** Flowchart of three-step feature selection.

**Table 2.** Data feature table after feature selection.

No.	Feature Name	No.	Feature Name
1	Drilling Time (min/m)	12	Inlet Temperature (°C)
2	Bit Pressure (KN)	13	Outlet Temperature (°C)
3	Hook Load (KN)	14	Total Hydrocarbons (%)
4	Torque (KN·m)	15	PWD Vertical Depth (m)
5	Hook Position (m)	16	PWD Annulus Pressure (MPa)
6	Hook Speed (m/s)	17	PWD Angle of Inclination (°)
7	Standpipe Pressure (MPa)	18	PWD Direction (°)
8	Totle Pump Stroke (SPM)	19	C2 (%)
9	Mud Tanks Volume (m <sup>3</sup> )	20	Wellhead Pressure (MPa)
10	Circulating Pressure Loss (MPa)	21	Outlet Flow (L/s)
11	Lag Time (min)	22	Inlet Flow (L/s)

The model constructed in this paper inputs 10 min of drilling data to predict whether overflow will occur in the next 10 min, and MPD device stores 50 pieces of data in 1 min, so the time series of 500 samples are used to predict whether overflow will occur in the next 500 samples. In view of the above intercepted data, a sliding time window is employed to construct time series samples, in which the window size is 500 timesteps and the sliding step size is 1. We collect the data of 500 timesteps as sample feature  $X$ , and collect whether overflow occurs after 500 timesteps as label  $y$ . For label  $y$ , if the feature data of 1–500 is taken as a sample feature  $X_1$  (the blue box in Figure 3), the corresponding label  $y_1$  is calibrated by finding whether overflow occurs in 501–1000 unit time  $Y_{501} - Y_{1000}$ , and the occurrence is 1, otherwise it is 0. Taking the feature data of 2–501 as the second sample feature  $X_2$  (see the red box in Figure 3), the corresponding label  $y_2$  is calibrated by judging whether overflow occurs in 502–1001 unit time  $Y_{502} - Y_{1001}$ , and the occurrence is 1, otherwise it is 0, and so on. Finally, in order to eliminate the influence of physical dimension, the Min-Max Normalization method is adopted for the original data, and the data range is reduced to [0, 1] to minimum training error.

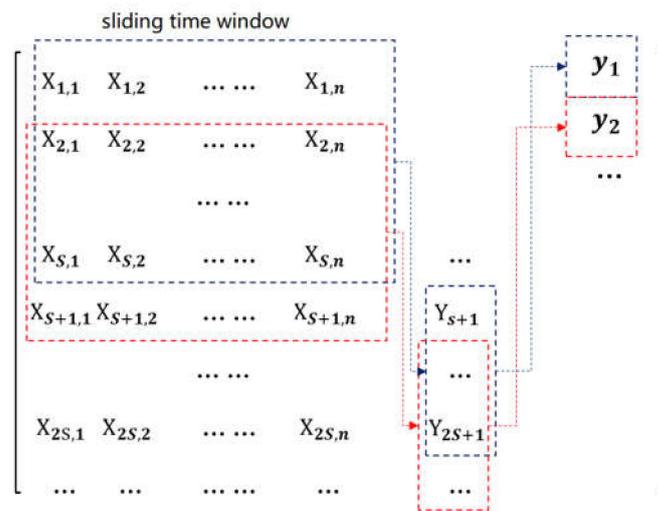


Figure 3. Sample label construction.

### 4. Methodology

#### 4.1. The Prediction Model

Overflow prediction is essentially an issue concerning the multi-variable and multi-step time-series prediction. The overflow condition can be determined according to the trend of the features of multiple drilling parameters, and the convolution neural network (CNN) can extract and map these parameters to produce higher and more extensive effective features. Regarding the long timestep of overflow prediction, the problem of gradient disappearance can be avoided by using the Long Short Term Memory (LSTM) model. In this paper, we build a CNN-LSTM fusion network for overflow prediction, using CNN and LSTM to merge a variety of effective features to capture the long-term dependence of time series and avoid the disappearance of gradients. The network structure is shown as Figure 4, and the detailed structure is shown as Supplementary Materials Table S1.

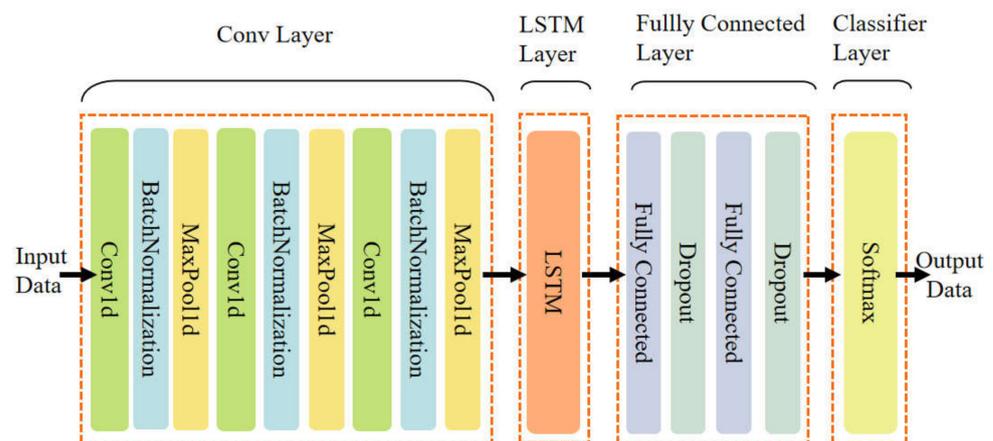


Figure 4. CNN-LSTM network architecture diagram.

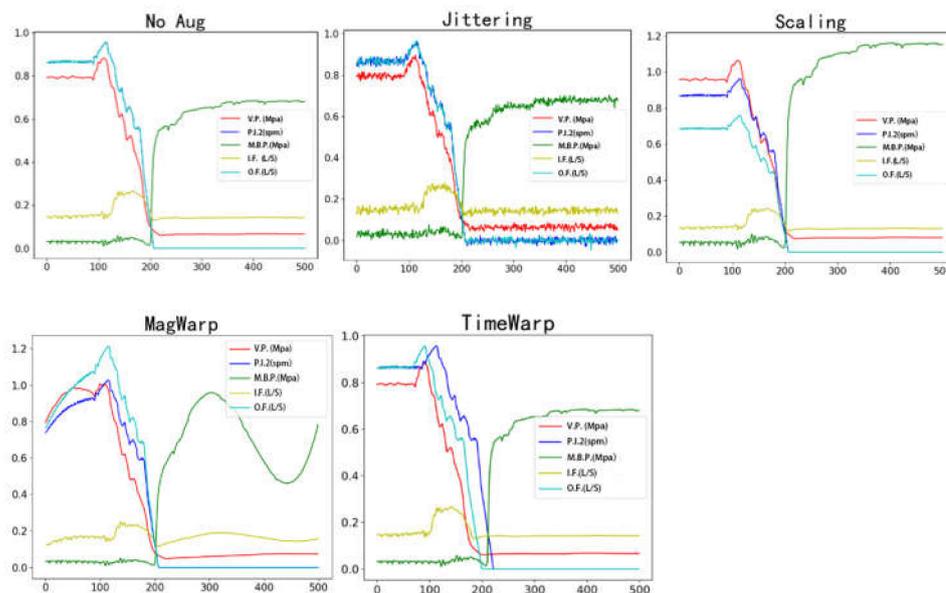
The CNN-LSTM model consists of a one-dimensional convolution network (1D-CNN), a LSTM layer, a three-layer full connection layer, and an output layer. Specifically, the input layer dimension is (500, 22). We use convolution layer for convolution on one-dimensional sequence and complete feature extraction through convolution operation. The one-dimensional convolution network built in this paper is composed of three sets of convolution and pooling layers. The first layer convolution has 64 convolution kernels, the kernel size is  $11 \times 22$ . There are 128 convolution kernels in the second layer, and the size of

the kernel is  $7 \times 64$ . The third layer has 128 convolution kernels, and the size is  $10 \times 128$ . The output size of the model is reduced through the pool layer, and finally the features extracted by 1D-CNN are output. And then LSTM layer, the full connection layer and the output layer follow the CNN part. In order to avoid the over-fitting of the neural network, we add the batch normalization layer after the convolution layer and the dropout layer after the full connection layer. Finally, the output layer is the SoftMax layer.

#### 4.2. Construction of Semi-Supervised Framework

In this paper, Mean Teacher algorithm is used to build a semi-supervised learning framework, and the input data needs to be enhanced according to the characteristics of Mean Teacher algorithm. After researching the characters of the drilling data, noise injection is more suitable for our data, which injects a small amount of noise/abnormal values into the time series without changing the corresponding label [18]. Terry et al., use methods such as Jitter, Scale, MagWarp, and TimeWarp to enhance wearable sensor data; appropriate enhancement can improve classification performance from 77.54% to 86.88% [19]. Jitter is usually to simulate additional sensor noise, and we use Gaussian noise in this paper. Scale resizes the data in the time window by multiplying a random scalar. MagWarp changes the size of each sample by converting the smooth curve of the data window. TimeWarp changes the time position of samples by smoothly distorting the time interval between samples. These data-enhancement methods can improve the robustness to multiplicative and additive noise.

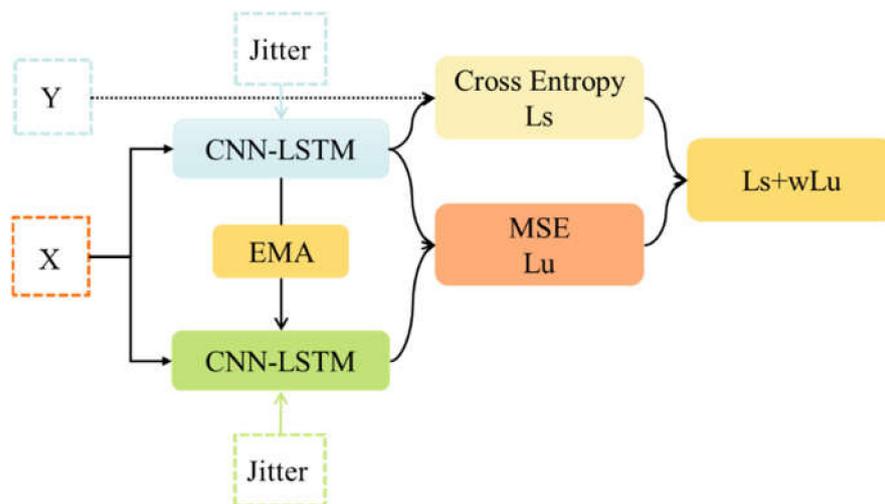
In order to explore the effect of the above four kinds of data enhancement, we randomly select a data sample with the size of (500, 22). Due to the significant changes in Stand Pipe Pressure (Mpa), Pump Impulse (spm), Wellhead Pressure (Mpa), Outlet Flow (L/S), and Inlet Flow (L/S) when overflow occurs, these five features are selected to enhance in four ways above. According to Figure 5 as follows, it can be seen that Jitter increases noise to time series data but does not change the trends of time series data. Therefore, Jitter is used to enhance the input data of Mean Teacher, that is, to increase Gaussian noise.



**Figure 5.** Enhancement effect of different time series. V.P. represents Stand Pipe pressure. P.I.2 represents Pump impulse. M.B.P. represents Wellhead pressure. I.F. represents Inlet flow. O.F. represents Outlet flow.

Finally, the overall framework of Mean Teacher is as follows. Firstly, the training dataset is obtained by pre-processing the original data. Then, the input data is enhanced by Jitter to effectively avoid the problem of over-fitting of the model. Finally, the CNN-LSTM

model [20] is used as the student model and teacher model of Mean Teacher for time-series prediction. The CNN-LSTM model is proved to be effective to predict overflow as the accuracy 89% in 10 min advance. The framework is shown in Figure 6.



**Figure 6.** Mean Teacher overall frame diagram.

## 5. Results

### 5.1. Model Training

In the experiment, 15% of the data are randomly selected as the verification set, that is, the proportion of training set and verification set is 17:3. In the training stage, to fully use the unlabeled data, the unlabeled data ratio  $\lambda_u$  is defined as Equation (3), where  $L$  is the number of labeled samples, while  $Lu$  is the number of unlabeled samples.

$$\lambda_u = \frac{Lu}{L} \quad (3)$$

Firstly,  $L$  pieces of data are randomly selected from the training set as labeled data sets. Secondly, the corresponding amount of data  $\lambda_u L$  is randomly selected from the remaining data at the ratio of  $\lambda_u$ , and the label is deleted as the unlabeled data set. At the same time, in each iteration, the label batch size is set to  $N$ , which means that the batch data of each training contains  $N$  labeled samples and  $\lambda_u N$  unlabeled samples. In order to verify the experimental results, Pseudo-label [21] semi-supervised framework and CNN-LSTM supervised learning model are built for comparing. In this experiment, almost the same hyper-parameters are used on Pseudo-label and Mean Teacher, Adam is used as optimizer,  $lr$  is learning rate, initial value is 0.001, weight decay is set as exponential decay, value is 0.0001, and  $\lambda_u$  is given a value of 5. Under different number of labeled training data, namely 50, 200, 1000 label samples, Mean Teacher, Pseudo-label, and supervised learning models are compared. The experimental parameters are shown as follows (Table 3).

**Table 3.** Experimental parameter settings.

Parameter	50 Labels	200 Labels	1000 Labels	All Labels
Batch_Size ( $N + \lambda_u N$ )	60	300	600	200
Label_Batch_Size ( $N$ )	10	50	100	200
Epoch	100	100	100	300

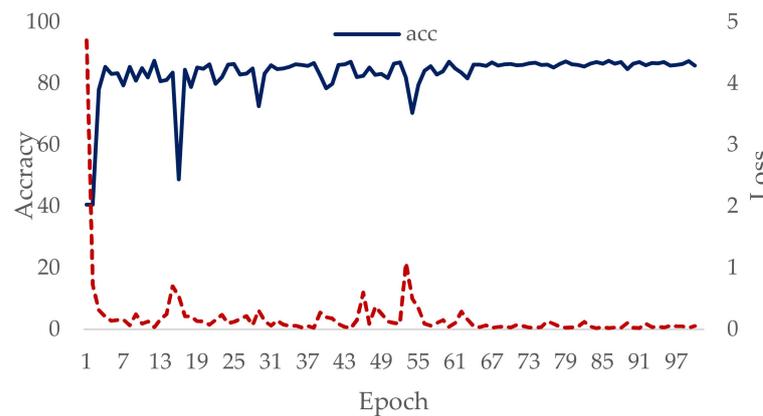
### 5.2. Model Results

The model is implemented by PyTorch, and the model is trained on the 32-core Tesla P40 calculation card. The accuracy is selected as the evaluation metric, and the results are

shown in Table 4. MeanTeacher+ represents MeanTeacher with Jitter data enhancements. Supervised represents a CNN-LSTM model that uses only labeled data. The training process is shown as Figure 7. It can be seen that the model is convergent after 70 epochs.

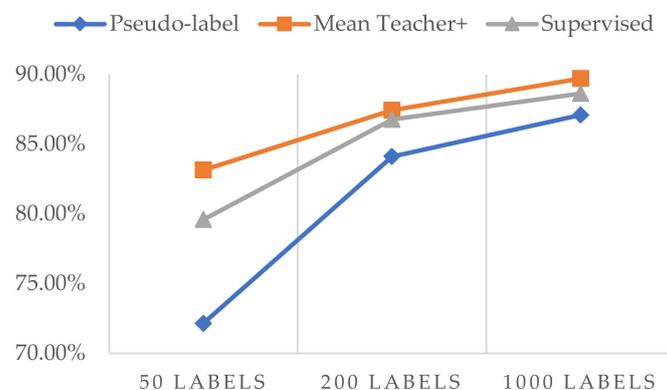
**Table 4.** Accuracy of various algorithms under different label samples.

Method	50 Labels	200 Labels	1000 Labels	All Labels
Pseudo-label	72.17%	84.13%	87.10%	-
Mean Teacher+	<b>83.15%</b>	<b>87.43%</b>	<b>89.70%</b>	-
Supervised	79.59%	86.77%	88.62%	89.90%



**Figure 7.** Loss and accuracy iteration curve.

According to Figure 8, the accuracy of MeanTeacher+ is higher than that of supervised learning in the case of 50, 200, 1000 label samples. This is because supervised learning only uses a small number of labeled samples for model training and the model is easy to over-fit, resulting in poor performance. While MeanTeacher+ inputs the unlabeled data into the model for training, the consistency loss provided by the unlabeled data can make the classification decision boundary fall in the low-density area. On the contrary, the performance of Pseudo-label is slightly worse. Pseudo-label uses the labeled data to train the model first, and then the trained model is used to predict the unlabeled data’s pseudo-labels. Obviously, due to the initial label data being less and the accuracy of the model being limited, pseudo-labels of unlabeled data are likely to be wrong. The large percentage of wrong pseudo-labels are input for training, which has a negative impact on the performance of the model.



**Figure 8.** Accuracy of various algorithms under different labeled samples.

When the ratio of unlabeled samples to labeled samples is 5, in the case of 50 labeled samples, MeanTeacher+ model is 3.56% better compared with supervised learning; in the

case of 200 labeled samples, the highest prediction accuracy of 10 min in advance is 87.43%, which is only 1% lower than that of supervised learning using 1000 labeled samples. The results suggest that a similar predictive result can be achieved with 200 labeled samples by MeanTeacher+ compared with supervised learning, which needs more than 1000 samples. In the case of 1000 labeled samples, the accuracy of the MeanTeacher+ is almost equal to that of supervised learning using all label samples.

To further verify the effect, ablation experiments are conducted to compare different machine-learning models and the effect of main component of MeanTeacher+. The machine-learning models such as SVM, Random Forest, LightGBM, XGBoost, and CNN-LSTM are trained in supervised mode under 200 label samples. MeanTeacher− means MeanTeacher+ without feature selection and Jitter noise, while MeanTeacher means MeanTeacher+ without Jitter noise. All the MeanTeacher models are trained with 200 labeled samples under  $\lambda_u = 5$ . The results are listed in Table 5.

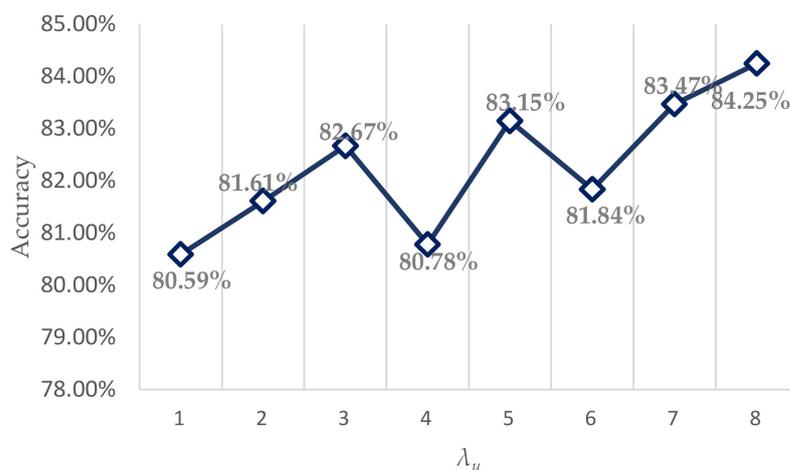
**Table 5.** Results of ablation experiments.

Method	Precision	Recall	F1
SVM	0.9679	0.6047	0.7443
Random Forest	0.9876	0.6220	0.7632
LightGBM	0.6246	0.6827	0.6523
XGBoost	0.9504	0.5039	0.6586
CNN-LSTM	0.7851	0.8064	0.7956
MeanTeacher−	0.8746	0.7527	0.8091
MeanTeacher	0.8804	0.7539	0.8123
MeanTeacher+	0.9154	0.7467	0.8225

From the table, it shows CNN-LSTM perform better than other supervised models, and feature selection and Jitter noise are able to improve the effect of MeanTeacher.

### 5.3. Sensitivity Analysis

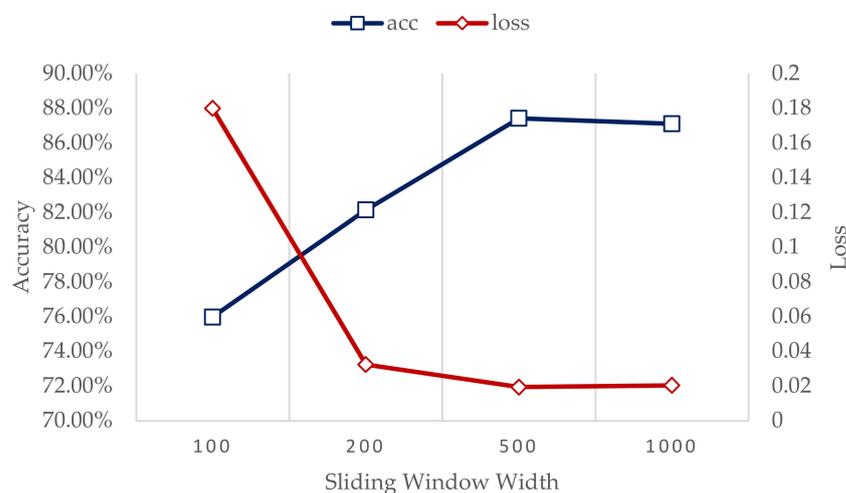
To verify the effect of the main parameter unlabeled data ratio  $\lambda_u$ , we change  $\lambda_u$  to compare the accuracies with the number of labeled data  $L$ . Figure 9 shows the accuracy of MeanTeacher with different  $\lambda_u$  in each batch training data given 50 labeled samples. It can be seen that, when  $\lambda_u$  is increased, the accuracy of the algorithm is improved. The result is valuable for real-world engineering application. Unlabeled samples are quite easy to obtain, so it is possible to improve the accuracy of the model by increasing the number of unlabeled samples.



**Figure 9.** Model accuracy with different  $\lambda_u$ .

The sliding window width is another key parameter for our model. The width represents the model's visual field. The bigger the width is, the more information the model

can obtain. More information will surely help the model to improve the effect, but it also brings expensive training costs and makes convergence hard. Hence, the width should be set properly. Several experiments are conducted to find the width, and the results are shown in Figure 10. It can be seen 500 is a proper width, and when the width is beyond 500, the accuracy begins to decrease.



**Figure 10.** Model accuracy and loss with different sliding window widths.

## 6. Conclusions

In order to predict drilling overflow, a three-step feature selection algorithm is conducted to extract 22 effective features from historical drilling data. The Mean Teacher semi-supervised learning framework is employed to train labeled data and unlabeled data at the same time, in which the CNN-LSTM fusion network works as the time-series prediction model, and Jitter noise is added to the time-series data as enhancement to prevent over-fitting. Compared with supervised learning and other semi-supervised learning frameworks, the results show that under the ratio of labeled to unlabeled data as 1:5, our model only needs 200 labeled samples to achieve the effect of the supervised learning method under 1000 samples, and the prediction accuracy can reach 87.43% 10 min in advance. Therefore, the drilling overflow prediction algorithm based on semi-supervised learning designed in this paper can predict drilling overflow accidents in advance to ensure drilling safety, even when the amount of labeled data is limited.

The success of the deep-learning model is based on the assumption that the distribution of training data and test data is consistent. Due to the differences in data distribution of different wells, the deep-learning model has poor prediction ability when facing new wells. Regarding the prediction of new wells, it is still necessary to label the new well data for model re-training and re-prediction. However, labeling manually is time-consuming and heavily depends on the experience of practitioners, and often only a small amount of labeled data can be obtained. The main advantage of this algorithm is to achieve high accuracy by a small amount of labeled data. It not only reduces the workload of labeling, but also predicts the overflow accurately.

According to the above process, when facing the prediction problem of a new well, the limitation of this algorithm is that it still needs a small amount of the labeled data of the new well, and the small amount of the labeled data must consist of normal and overflow data. However, it is quite difficult to acquire these data completely during the initial drilling process, and that means it is hard to apply the model in a short time for a new well. To address this issue, we will study transfer learning and realize the adaption for a semi-supervised domain with a small amount of labeled data from the new well and the previous well-source domain samples. Besides, the proposed model is a general model,

and we will transfer the model to predict other complex work conditions such as mud losses and pipe sticking.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/en15124324/s1>, Table S1: Detailed Network Framework.

**Author Contributions:** Data curation, M.C.; funding acquisition, W.L.; investigation, W.L.; methodology, W.L. and X.H.; project administration, J.F.; resources, J.F. and M.C.; software, J.F., M.C. and X.H.; visualization, Y.L.; writing—original draft, Y.L. and X.H.; writing—review & editing, Y.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors are grateful for the support of the National Key Research and Development Program of China (2019YFA0708304, 2021YFF1201203, 2021YFF1201205), the National Natural Science Foundation of China (61972174 and 62172187), the Key core technology research project of CNPC (2020B-4019), the Science and Technology Planning Project of Guangdong Province (2020A0505100018), Guangdong Universities' Innovation Team Project (2021KCXTD015) and Guangdong Key Disciplines Project (2021ZDJS138).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Not applicable.

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

## References

1. Hargreaves, D.; Jardine, S.; Jeffryes, B. Early Kick Detection for Deepwater Drilling: New Probabilistic Methods Applied in the Field. In Proceedings of the SPE Annual Technical Conference and Exhibition, New Orleans, LA, USA, 30 September–3 October 2001; OnePetro: Richardson, TX, USA, 2001.
2. Wang, X.L.; Lian, X.; Yao, L. Fault Diagnosis of Drilling Process Based on Rough Set and Support Vector Machine. *Adv. Mater. Res.* **2013**, *709*, 266–272. Available online: <https://www.scientific.net/AMR.709.266> (accessed on 4 May 2022).
3. Lind, Y.B.; Kabirova, A.R. Artificial Neural Networks in Drilling Troubles Prediction. In Proceedings of the SPE Russian Oil and Gas Exploration & Production Technical Conference and Exhibition, Moscow, Russia, 14–16 October 2014; OnePetro: Richardson, TX, USA, 2014.
4. Li, X.; Liang, H.; Wang, X. Overflow Prediction Method Using Fuzzy Expert System. *Int. Core J. Eng.* **2015**, *7*, 58–64.
5. Liang, H.; Zou, J.; Li, Z.; Khan, M.J.; Lu, Y. Dynamic evaluation of drilling leakage risk based on fuzzy theory and PSO-SVR algorithm. *Future Gener. Comput. Syst.* **2019**, *95*, 454–466. [[CrossRef](#)]
6. Liang, H.; Zou, J.; Liang, W. An early intelligent diagnosis model for drilling overflow based on GA–BP algorithm. *Clust. Comput.* **2019**, *22*, 10649–10668. [[CrossRef](#)]
7. Haibo, L.; Zhi, W. Application of an intelligent early-warning method based on DBSCAN clustering for drilling overflow accident. *Clust. Comput.* **2019**, *22*, 12599–12608. [[CrossRef](#)]
8. Zhu, Q.; Wang, Z.; Huang, J. Stuck Pipe Incidents Prediction Based On Data Analysis. In Proceedings of the SPE Gas & Oil Technology Showcase and Conference, Dubai, United Arab Emirates, 21–23 October 2019; OnePetro: Richardson, TX, USA, 2019.
9. Borozdin, S.; Dmitrievsky, A.; Eremin, N.; Arkhipov, A.; Sboev, A.; Chashchina-Semenova, O.; Fitzner, L.; Safarova, E. Drilling Problems Forecast System Based on Neural Network. In Proceedings of the SPE Annual Caspian Technical Conference, Online, 21–22 October 2020; OnePetro: Richardson, TX, USA, 2020.
10. Sabah, M.; Mehrad, M.; Ashrafi, S.B.; Wood, D.A.; Fathi, S. Hybrid machine learning algorithms to enhance lost-circulation prediction and management in the Marun oil field. *J. Pet. Sci. Eng.* **2021**, *198*, 108125. [[CrossRef](#)]
11. Liu, Z.; Ma, Q.; Cai, B.; Liu, Y.; Zheng, C. Risk assessment on deepwater drilling well control based on dynamic Bayesian network. *Process Saf. Environ. Prot.* **2021**, *149*, 643–654. [[CrossRef](#)]
12. Yin, B.; Li, B.; Liu, G.; Wang, Z.; Sun, B. Quantitative risk analysis of offshore well blowout using bayesian network. *Saf. Sci.* **2021**, *135*, 105080. [[CrossRef](#)]
13. Liang, H.; Han, H.; Ni, P.; Jiang, Y. Overflow warning and remote monitoring technology based on improved random forest. *Neural Comput. Appl.* **2021**, *33*, 4027–4040. [[CrossRef](#)]
14. Wang, K.; Liu, Y.; Li, P. Recognition method of drilling conditions based on support vector machine. In Proceedings of the 2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA), Shenyang, China, 21–23 January 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 233–237.
15. Ouali, Y.; Hudelot, C.; Tami, M. An Overview of Deep Semi-Supervised Learning. *arXiv* **2020**, arXiv:2006.05278.
16. Tarvainen, A.; Valpola, H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4 December 2017; Curran Associates, Inc.: New York, NY, USA, 2017; Volume 30.

17. Ge, R.; Zhou, M.; Luo, Y.; Meng, Q.; Mai, G.; Ma, D.; Wang, G.; Zhou, F. McTwo: A two-step feature selection algorithm based on maximal information coefficient. *BMC Bioinform.* **2016**, *17*, 142. [[CrossRef](#)] [[PubMed](#)]
18. Wen, Q.; Sun, L.; Yang, F.; Song, X.; Gao, J.; Wang, X.; Xu, H. Time Series Data Augmentation for Deep Learning: A Survey. *arXiv* **2021**, *5*, 4653–4660.
19. Um, T.T.; Pfister, F.M.J.; Pichler, D.; Endo, S.; Lang, M.; Hirche, S.; Fietzek, U.; Kulić, D. Data augmentation of wearable sensor data for parkinson’s disease monitoring using convolutional neural networks. In Proceedings of the 19th ACM International Conference on Multimodal Interaction, Glasgow, UK, 13–17 November 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 216–220.
20. Fu, J.; Liu, W.; Han, X.; Li, F. CNN-LSTM Fusion Network Based Deep Learning Method for Early Prediction of Overflow. *China Pet. Mach.* **2021**, *49*, 16–22. [[CrossRef](#)]
21. Lee, D.-H. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In Proceedings of the ICML 2013 Workshop: Challenges in Representation Learning, Atlanta, GA, USA, 21 June 2013; ICML: San Diego, CA, USA, 2013; pp. 1–6.