

Article

State Prediction Method for A-Class Insulation Board Production Line Based on Transfer Learning

Yong Wang^{1,2,3}, Hui Wang⁴, Xiaoqiang Guo⁴, Xinhua Liu⁴ and Xiaowen Liu^{1,5,*}¹ School of Electrical Engineering, China University of Mining and Technology, Xuzhou 221000, China² School of Information and Engineering, Xuzhou College of Industrial Technology, Xuzhou 221000, China³ IOT Perception Mine Research Center, China University of Mining and Technology, Xuzhou 221000, China⁴ School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou 221000, China⁵ School of Electrical and Power Engineering, China University of Mining and Technology, Xuzhou 221000, China

* Correspondence: xwliucumt@126.com

Abstract: It is essential to determine the running state of a production line to monitor the production status and make maintenance plans. In order to monitor the real-time running state of an A-class insulation board production line conveniently and accurately, a novel state prediction method based on deep learning and long short-term memory (LSTM) network is proposed. The multiple layers of the Res-block are introduced to fuse local features and improve hidden feature extraction. The transfer learning strategy is studied and the improved loss function is proposed, which makes the model training process fast and stable. The experimental results show that the proposed Res-LSTM model reached 98.9% prediction accuracy, and the average R^2 -score of the industrial experiments can reach 0.93. Compared with other mainstream algorithms, the proposed Res-LSTM model obtained excellent performance in prediction speed and accuracy, which meets the needs of industrial production.

Keywords: transfer learning; state prediction for production line; LSTM; domain adaptation**MSC:** 68T07

Citation: Wang, Y.; Wang, H.; Guo, X.; Liu, X.; Liu, X. State Prediction Method for A-Class Insulation Board Production Line Based on Transfer Learning. *Mathematics* **2022**, *10*, 3906. <https://doi.org/10.3390/math10203906>

Academic Editor: Ioannis G. Tsoulos

Received: 15 September 2022

Accepted: 19 October 2022

Published: 21 October 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

A-class thermal insulation board is widely used in the construction industry because of its good thermal insulation performance. However, the composite insulation board manufacturing technique is very complex. The thermal insulation board production line consists of six components: uncoiling machine, mold machine, high-pressure foaming machine, laminating conveyor, cutting machine and palletizing machine [1,2]. Considering the production complexity of the thermal insulation board and automation of the production line, the possibility of related failures in the whole system also increases. Therefore, the effective state prediction of the equipment on the production line can ensure safe and stable operation and make the production line run efficiently.

As the value of transfer learning [3,4] has been continuously explored, many scholars have also studied the application of transfer learning in the field of predictive maintenance of industrial equipment. At present, the research mainly focuses on the diagnosis prediction for equipment under complex working conditions, as well as prediction based on few-shot datasets. The main studies are concentrated in the petroleum, chemical, electric power, aerospace and machinery manufacturing industries. Large essential mechanical equipment, such as heavy-load gearboxes and reciprocating compressors, are usually complex in structure and run under harsh environments and multiple loading cases, resulting in variable fault features and difficulty in obtaining effective features. The modeling of intelligent fault diagnosis is difficult, and component failures are uncertain, which makes fault prediction difficult. Therefore, the fault diagnosis and state prediction of equipment

under variable working conditions has become a research hotspot in the field of predictive maintenance [5,6]. In the field of predictive maintenance, failure samples are sparse because the device is rarely in a faulty state and the fault status is difficult to record. Before data on failed devices are collected, the devices may be damaged and not able to run normally. As a result, the number of collected fault samples is far smaller than that of normal samples. In the meanwhile, other factors, such as differences between similar equipment or operating conditions, make it difficult to apply fault samples of certain equipment to other equipment. This leads to the scarcity of fault data in the industry. A few-shot dataset of fault samples will further lead to serious overfitting problems in traditional machine learning [7–9].

Based on the theory of transfer learning, samples with certain similarities in different target domains are transferred to the source domain for model training, or the source domain data related to the target can be effectively migrated to the target fault training, which assists the training process of fault classifiers. The feature space mapping can also be used to train semi-supervised or unsupervised models. Transfer learning theory can effectively address the problem of poor generalization caused by small sample conditions.

In this paper, a novel state prediction method based on LSTM and transfer learning theory is proposed. The Res-LSTM model predicts various faults that may occur in the production process of an A-class insulation board production line. The main contributions of this paper are detailed as follows:

(1) The improved LSTM network architecture is proposed in this paper to predict the state of the production line. The Res-block is introduced to fuse local information and extract features from time series data. The proposed Res-LSTM model has the feature extraction ability of a convolutional layer and can make predictions on sequence data.

(2) The transfer learning theory is introduced in the model training process, which addresses the problem of limited labeled data in the target domain. Thereby, the Res-LSTM model obtains better training stability and generalization ability.

(3) The improved loss function is proposed based on transfer learning and Kullback–Leibler (KL) divergence. The introduced adaptation term controls the training process and makes the training of the Res-LSTM model more stable.

The rest paper is organized as follows: Section 2 introduces related works on state prediction methods and the background of deep learning methods based on LSTM. The proposed state prediction method for an A-class thermal insulation panel production line is demonstrated in Section 3. Experiments and corresponding results are presented in Section 4. Finally, Section 5 shows the application of the detection system, and Section 6 concludes this paper and discusses future works.

2. Related Works

State prediction for a production line means monitoring the production equipment condition based on various sensors deployed and predicting the production status of the next state according to the current data of multiple sensors.

2.1. State Prediction Methods

In recent years, some experts and scholars have made great achievements based on intelligent sensors and proposed novel industrial intelligent prediction algorithms for production line running state monitoring systems. Sun et al. [10] designed a control system for a biomass solid molding fuel production line and realized the automatic operation and remote monitoring of the production line using PROFIBUS-DP fieldbus (S7-300PLC) and Monitor and Control Generated System (MCGS) software. Izonin [11] proposed a new method for solving the multiple linear regression task via a linear polynomial to identify coefficients of multiple polynomials at high speed, which can be used in multiple fields. Cui et al. [12] developed an automatic control system based on WinCC software and PLC to remotely monitor an assembly line of the spindle blade and worm gear of a spinning machine. Sun et al. [13] proposed a three-level network monitoring system, including host, master and slave nodes, which could monitor the running status of the production

line with the help of configuration software and an Access database. It can query the historical production information and fault information according to the database storage information. Chen et al. [14] designed a production shop monitoring system based on the wireless sensor network and radio frequency identification technology of the Industrial Internet of Things (IIOT), which can automatically collect the real-time status of the shop floor and provide a wealth of information for business decisions. Neuro-like structure and its advantages for achieving high-speed training and visualization in 2D or 3D formats were described in [15,16]. Vijayaraghavan et al. [17] proposed an automatic monitoring and management system for machine tools based on MTConnect and corresponding standard technologies. It realizes the function of real-time and quantitative monitoring and provides data analysis for effective energy consumption and evaluation of machine tools. Duro et al. [18] proposed a processing monitoring framework for NC machine tools based on the fusion of multi-sensor information.

2.2. State Prediction Methods Based on LSTM

With the development of deep learning, special network architectures have been designed to address different problems. Recurrent neural networks (RNNs) [19] are deep learning networks that focus on working with sequence data. Many improved network architectures, such as LSTM [20], bidirectional LSTM (BiLSTM) [21] and gate recurrent unit (GRU) [22], have better performance under special conditions. Compared with basic RNN networks, the improved ones have memory ability. The BiLSTM model consists of two flows: forward and backward, which can make predictions based on both past information and future information. GRU is also a kind of recurrent neural network for dealing with time series data, but it is a simplified version of LSTM. It contains two gates, namely a reset gate and an update gate, and costs less computation power. Ordonez and Roggen [23] proposed a generic deep framework for activity recognition based on convolutional and LSTM recurrent units. Kong et al. [24] proposed a prediction framework based on an LSTM recurrent neural network to perform short-term load forecasting for individual residential households. Sun et al. [25] proposed an additional model to mitigate the uncertainty of the tested PA when transforming phases and make a bridge between the memory effects of the nonlinear PAs and BiLSTM neural networks. Hua et al. [26] proposed an end-to-end network based on a bidirectional LSTM network. This BiLSTM model can explore the co-occurrence relationship of various classes and classify aerial images effectively. Che et al. [27] developed a GRU-based deep learning model to classify time series data. Two pattern representations were introduced to address the problem of missing data values. Chen et al. [28] proposed a remaining life prediction method to address the deep learning model problem of a complex system with multiple components and an extremely large number of parameters. The proposed GRU network described a very complex system with a compact model and required less training time.

2.3. Prediction Methods Based on Transfer Learning Theory

With the development of deep learning theory, many scholars have proposed state prediction methods based on deep learning and transfer learning theory. By preparing a large number of labeled samples for a deep network, a deep learning model can be obtained to complete complicated tasks after training [29–31]. Further, transfer learning theory is usually applied to deep learning models trained on datasets with few samples. Samples of different fields and tasks can be correctly classified and identified based on transfer learning theory, which makes few-shot learning possible. For example, the trained model performs well for samples from the same distribution as the training dataset. However, for samples from different distributions in the test dataset, the generalization performance of the model will be degraded. In the theory of transfer learning, different distributions of samples are usually called different domains. In the early studies, the generalization of the trained model is realized by matching the feature distribution of different domains, or the invariant features of the target domain can be obtained by minimizing the differences between

domains. The representative algorithms are DAN [32], DDC [33] and Deep-CORAL [34]. Compared with Deep-CORAL, DDC is a method to fine-tune the pre-trained model of AlexNet based on maximum mean discrepancy (MMD) and minimizing the distribution distance between the source domain and the target domain. The difference between these two algorithms is that the loss function of Deep-CORAL adopted the second-order statistics of the target domain and the source domain. It can be concluded that the early transfer learning algorithms improve the classification accuracy based on the loss function design. Saito et al. [35] proposed an ATDA transfer learning algorithm to address the problem of domain matching in the DDC algorithm. Manders et al. [36] proposed a CPUA transfer learning algorithm that greatly improved the accuracy of the target dataset. The distribution difference between the source and the target domain in the feature level is reduced, which forces the model to learn the invariant features between domains. Pei et al. [37] proposed a MADA transfer learning algorithm, which applies to binary classification problems and replaces one target classifier with multiple target classifiers. Fault data, as an important part of the dataset, are essential for predicting the production line state. However, they are only a small part of the state dataset in the production process. Based on the transfer learning theory, a model with strong generalization ability that monitors the state of the production line and predicts potential failure accurately can be obtained based on transfer learning theory and the limited fault data of a production line.

2.4. Discussion

At present, the models established by deep learning networks exhibit excellent performance, however, they need balanced datasets with a large quantity of labeled data. The transfer learning theory addresses the problem of few-shot learning, and the trained model can obtain relatively excellent performance. Therefore, to address the above problems, a state prediction method based on transfer learning theory is proposed to address the problem of small datasets. The proposed Res-LSTM model makes full use of multi-source data and gains sufficient feature extraction ability.

3. The Proposed State Prediction Method for an A-Class Insulation Panel Production Line

3.1. The Proposed Network Architecture Based on LSTM

In this paper, a state prediction method for an A-class insulation board production line based on transfer learning is proposed. The proposed model combines the convolutional residual module and LSTM network. Firstly, the convolutional neural network is used to capture the spatial characteristics of the local range production line data. Then the residual modules are applied to deepen the number of network layers, and the long short-term memory network (LSTM) is used to learn the temporal periodicity and trend of the spatial-temporal data. For each component, input data information is fused by adaptive weights, and the predicted data of each node are obtained after the LSTM network output, so as to predict the production state of the A-class thermal insulation board production line. The proposed Res-LSTM network architecture is shown in Figure 1, and the Res-block applied in the proposed model is shown in Figure 2.

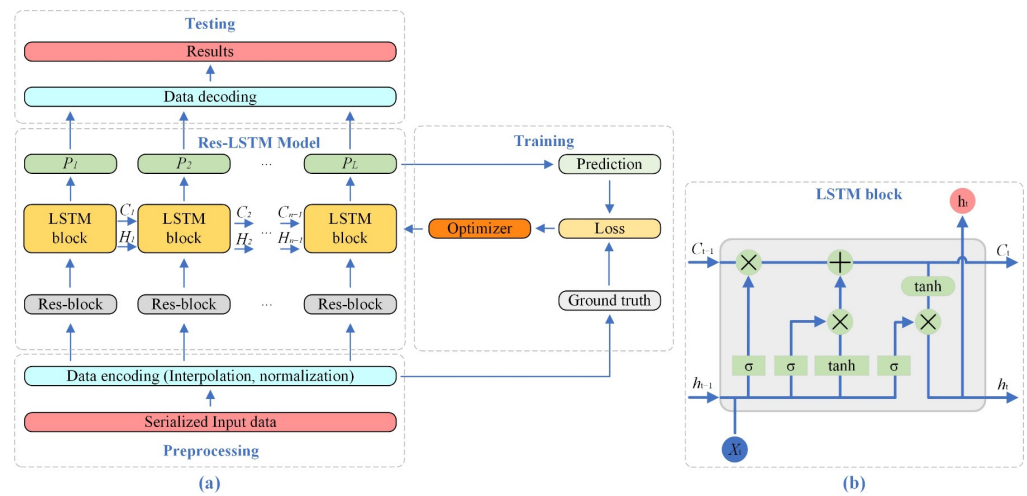


Figure 1. The proposed state prediction network based on Resnet and LSTM. Subplots (a,b) represent the proposed network architecture and LSTM block, respectively.

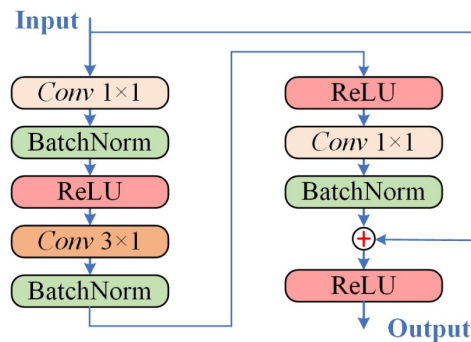


Figure 2. The Res-block applied in the proposed network architecture.

The LSTM block displayed in Figure 1 consists of four gates: forget gate, input gate, update gate and output gate.

(1) Forget gate. The first step is to decide what information to discard from the block, which is achieved by a sigmoid function activation layer. This gate is designed to forget information that leads to false predictions and can be described as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

(2) Input gate. This step decides what information should be kept in the LSTM block. First, the tanh layer creates a vector \tilde{C}_t . Then the sigmoid layer outputs a number between 0 and 1 for each value in \tilde{C}_t to determine which state values to update and by how much. This step can be described as follows:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{3}$$

(3) Update gate. This step updates C_{t-1} to the new state C_t , and this step can be described as follows:

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

(4) Output gate. The last step decides what information should be outputted. The cell state is passed through the tanh network layer, and the result is multiplied by the output of the sigmoid layer. This step can be described as follows:

$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t \times \tanh(C_t) \quad (6)$$

The proposed Res-block consists of three convolutional layers with corresponding batch normalization and activation layers. The beginning and end convolutional layers contain a kernel size of 1×1 , which is used to adjust the channels of input and output. The middle convolutional layer fully extracts hidden features. The architecture of the proposed Res-block is shown in Figure 2.

In this paper, the framework of the state prediction method based on transfer learning for an A-class insulation board production line is mainly divided into three parts: data preprocessing, model training and model validation. The data preprocessing module receives serialized data from various sensors. Because of the different sampling rates and possible missing data, the quadratic interpolation and normalization algorithms are adopted to output the complete time series data which can be used for the proposed network training.

3.2. Transfer Learning

In this paper, the proposed state prediction method is based on multi-sensor data collected in real time. Due to the various production states of the production line, the complicated corresponding relationship between the production line and a variety of sensor data and difficulties in collecting a large amount of real-time sensor data, the model trained with normal steps tends to lack generalization ability in inference procedures. Therefore, a state prediction method for an A-class thermal insulation board based on transfer learning is proposed in this paper; it solves the problem of limited labeled data in the target domain with the help of rich labeled data in other domains. Transfer learning involves two important concepts, domain and task, which are described as follows:

The domain D is defined as a d -dimensional feature space X and a marginal probability distribution $P(x)$, i.e., $D = \{X, P(x)\}, x \in X$. Given a domain D , a task T is defined as consisting of a category space Y and a prediction model $f(x)$, i.e., $T = \{Y, f(x)\}, y \in Y$, according to the statistical prediction model $f(x) = P(y|x)$ interpreted as conditional probability distribution.

Based on the theory of domain adaptation in transfer learning, a state prediction method for an A-class thermal insulation board production line is proposed in this paper based on LSTM. Considering the conditions of the same feature space and same category space and different edge distribution and different condition distribution between domains, homogeneous transfer learning is investigated. That is, the multivariate homogeneous data collected from similar production lines are used to reduce the generalization error of the prediction model of the A-class thermal insulation board production line.

For the prediction error of transfer learning, classical statistical learning theory gives the upper bound guarantee of the generalization error of the learning model under independent and identically distributed conditions. The smaller the training error is, the smaller the generalization error will be, and the more training samples, the closer the training error will be to the generalization error. The theoretical guarantee of the upper bound of generalization error is essential for the success of machine learning based on statistical principles. Considering the upper bound of the generalization error of the transfer learning model, this paper introduces the following assumptions:

$$\varepsilon_T(h) \leq \varepsilon_S(h) + d(D_s, D_T) + C \quad (7)$$

where $\varepsilon_S(h)$ and $\varepsilon_T(h)$ denote the prediction errors of the source domain and the target domain, respectively. $d(D_s, D_T)$ denotes the distance between the source domain and the target domain under a certain metric, and C is a constant term. According to this hypothesis, the optimization objective of the transfer learning model usually consists of two parts: first, the accuracy of the classifier in the source domain should be as high as possible; second, the source domain and the target domain should be as close as possible.

Given training data D of the homogeneous dataset $D_{source} = \{(x_{S_i}, y_{S_i}) | x_{S_i} \in X, y_{S_i} \in Y, i = 1, 2, \dots, n_S\}$ and trained prediction model $f_S(\cdot)$, transfer learning improves the performance of the neural network model $f_T(\cdot)$ of target domain D_{target} by using homogeneous dataset D_{source} and model $f_S(\cdot)$. In this paper, the parameter θ_S of RES-LSTM model trained on the homologous dataset collected for similar production line S is used as the source model. That is, the initialization parameters of RES-LSTM, the prediction model for the production line, are trained on the corresponding dataset. That is, the θ_S is considered as initialization parameters of the RES-LSTM model for the A-class thermal insulation board production line, which are trained on the target dataset. At the same time, the learning rate of training L_r is adjusted, and an adaptation term is added to the loss function to improve the adaptability of the RES-LSTM model in the target domain. The specific structural framework is shown in Figure 3.

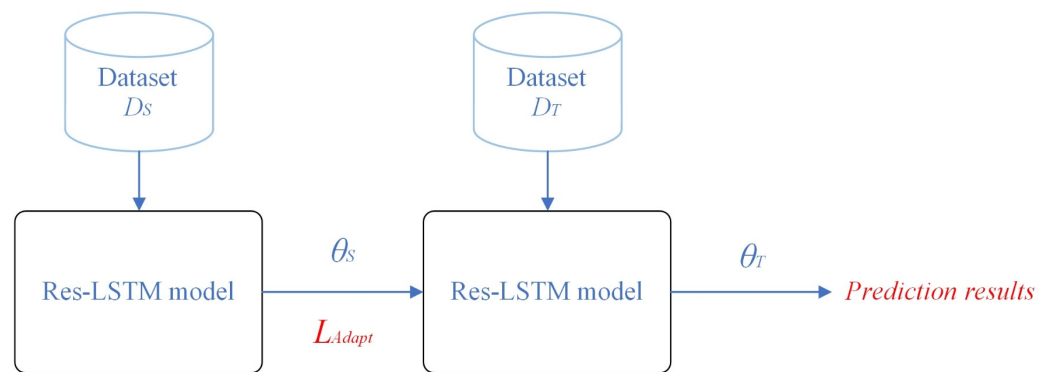


Figure 3. The training architecture based on transfer learning proposed in this paper.

An adaptation term L_{Adapt} has been added to the loss function of the Res-LSTM model on the target domain dataset D_{target} , and the parameter changes of the Res-LSTM model in the target domain can be controlled by using the model information trained on the homogeneous dataset D_{source} . The statistical distance $g(\cdot)$ between the two domains is calculated, and L_{Adapt} is shown as follows:

$$L_{Adapt} = g(H_S, H_T) \tag{8}$$

where $g(\cdot)$ denotes a function measuring the distance between two domains. KL divergence is adopted in this paper, and L_{Adapt} is expressed as follows:

$$L_{Adapt} = g(H_S, H_T) = \frac{1}{n} \sum_i \sum_t (h_S^t)_i \log(h_D^t)_i \tag{9}$$

where i denotes the i -th term of the output vector of the proposed Res-LSTM model.

3.3. Loss Function

A state prediction model based on transfer learning that can effectively predict various states of the production line is proposed in this paper. The loss function of this proposed model consists of two parts: classification loss L_C and adaptation loss L_{Adapt} . The mean square error (MSE) loss function is used for classification loss L_C :

$$L_c = MSE(y_i, \hat{y}_i) = \frac{1}{n} (y_i - \hat{y}_i)^2 \tag{10}$$

where y_i and \hat{y}_i denote prediction results and ground truth, respectively. Specifically, the output results of the proposed model are passed through the softmax function to obtain the probability of each state, and then the mean square error loss is calculated with the real value.

KL divergence is used for the adaptation loss L_{Adapt} , which has been discussed in Section 3.2. In summary, the loss function L of the Res-LSTM model proposed in this paper is

$$\begin{aligned}
 L &= \lambda L_c + \mu L_{Adapt} = \lambda MSE(y_i, \hat{y}_i) + \mu g(H_s, H_T) \\
 &= \lambda \frac{1}{n} (y_i - \hat{y}_i)^2 + \mu \frac{1}{n} \sum_i \sum_t (h_s^t)_i \log (h_D^t)_i
 \end{aligned}
 \tag{11}$$

where λ and μ are constants.

4. Experiments

4.1. Dataset Preparation

In this paper, a state prediction method for an A-class thermal insulation board production line is proposed based on transfer learning. By installing various sensors on two similar production lines, a large number of homogeneous datasets were obtained. The Res-LSTM model proposed in this paper is trained based on the domain adaptation theory of transfer learning. In this study, a variety of sensors were deployed on the two production lines to collect the data of voltage, current, vessel pressure, motor speed and vibration information of the equipment to form the time series datasets. Due to the different sampling rates of various sensors and the possible data loss, it was necessary to preprocess the time series data collected in real time. The algorithms of data interpolation and completion and standardization were applied, and the trainable datasets were obtained. The time series data of the preprocessed target dataset D_t are shown in Figure 4.

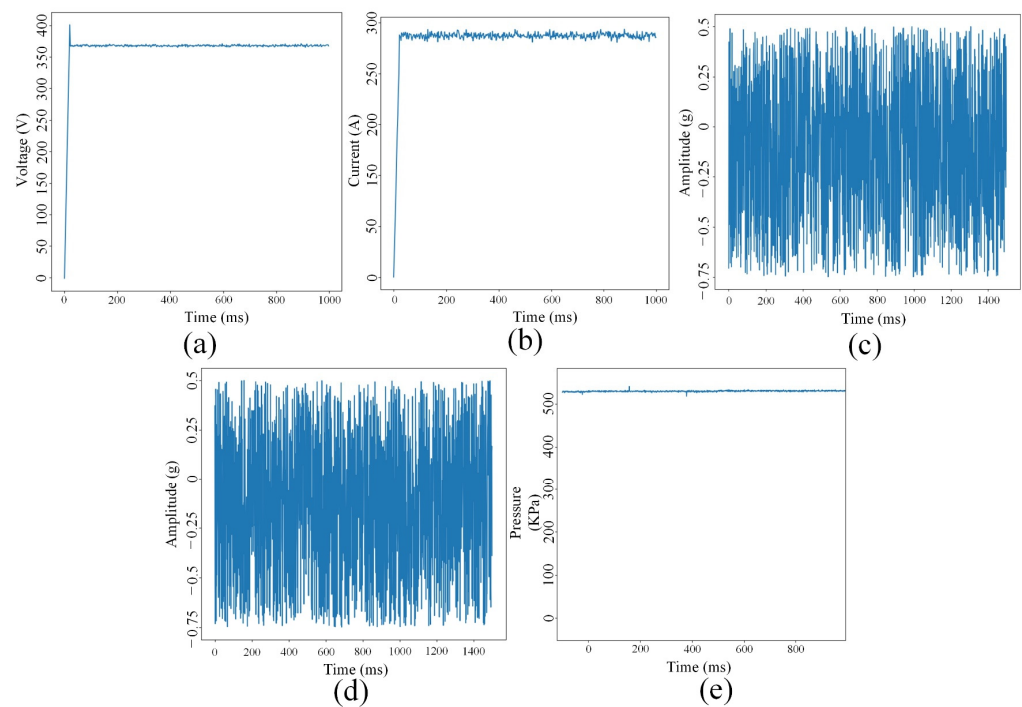


Figure 4. Data of A-class thermal insulation board production line collected in real time. The graphs (a–e) are voltage, current, vibration-x, vibration-y and pressure, respectively.

The source dataset D_s contains a large amount of labeled time series data, and the pre-trained Res-LSTM model is based on this dataset. The parameters of this model will be used as the initial parameters of the state prediction model. The target dataset D_t is used to train the desired RES-LSTM model. The prediction accuracy of Res-LSTM model is improved by reducing the classification loss L_c . By reducing the L_{Adapt} loss, the generalization upper bound of the target model based on transfer learning is reduced as much as possible.

4.2. Model Training

The proposed Res-LSTM state prediction model based on transfer learning can achieve better performance with fewer iterations compared with other common neural network models, which is due to pre-initialized network parameters based on domain adaptation. The training process of the proposed Res-LSTM model based on domain adaptation theory is relatively smooth, and the performance quickly reaches the best level. The proposed Res-LSTM model is implemented by Pytorch, a deep learning framework. Table 1 shows the main configuration list and model parameters of the server used in training.

Table 1. The main configuration and model parameters of the server used in training.

Model	Dell Precision T7820
CPU	Intel Core X-Series
GPU	NVIDIA RTX A4000 (16 G)
RAM	64G ECC
SSD	512 G
HDD	2 T
Power consumption	950 W

Adam optimizer was selected with a momentum of $\beta_1 = 0.92$ and $\beta_2 = 0.995$ in the training process. Batch size and learning rate were set to 96 and 0.00015, respectively. The decay learning rate was introduced, and the decay ratio was 0.97 per five epochs; hence, the final learning rate was $1.5 \times 10^{-4} \times 0.97^{40} \approx 4.44 \times 10^{-5}$. Figure 4 shows the comparison between the training process based on transfer learning and the ordinary training process. In the training process, the changes in mean square error (MSE) loss of the training and testing sets are shown in Figure 5. A rapid decreasing trend was shown in both training and testing processes at the beginning. However, the curve tended to be smooth after epoch 160. To select the model with the best performance, the model parameters after epoch 190 of training were selected as the final CNN-LSTM model parameters. It is obvious that the convergence speed of the proposed Res-LSTM model is faster and the generalization performance is better.

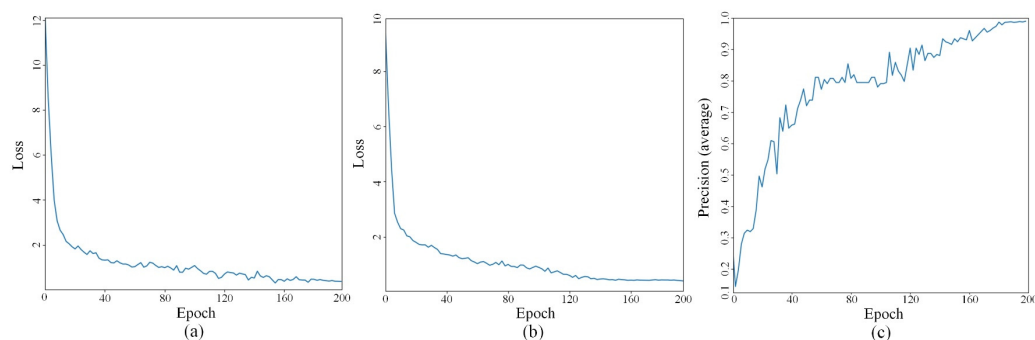


Figure 5. The comparison of transfer learning-based and normal training procedure. Training loss (a), evaluation loss (b) and average precision (c).

4.3. Comparison of Experimental Results

In order to verify the performance of the state prediction method for an A-class thermal insulation panel production line proposed in this paper, the well-trained Res-LSTM model is compared with other algorithms: autoregressive integrated moving average model (ARIMA), Holt–Winters, singular spectrum analysis (SSA), multiple linear regression (MLR) and support vector regression (SVR). The comparisons of above mainstream algorithms are listed in Table 2.

Table 2. The advantages and disadvantages of contrastive algorithms.

Algorithm	Advantages	Disadvantages
ARIMA	Flexible in application, not bound by data types and has strong applicability.	Poor prediction ability for unstable data after difference
Holt–Winters	Different weights to the model for each period, and reasonable prediction	Not suitable for data with enormous changes
SSA	Signal extraction and filtering capabilities	Slow speed for large time series data
MLR	Simple implementation, excellent performance for linear prediction	Interaction and nonlinear effects are ignored. Not suitable for nonlinear prediction
SVR	Processing nonlinear data by Kernel trick. L2 regularization and soft interval lead to strong generalization ability	High computation complexity, slow prediction speed for big data analytics
Proposed	Excellent performance for time series data, fast prediction speed by GPU, high prediction accuracy	Complex model, costs time to train model

ARIMA is a classical theory and method for time series analysis, and its model can be expressed as $ARIMA(p, d, q)$, where p , d and q denote the autoregressive terms, the number of differences and moving average terms, respectively. Holt–Winters is a time series analysis method that can deal with both trend and periodicity components. The idea is to recurse the current data using the different characteristic components of the historical data. SSA is a non-parametric method combining the time domain and frequency domain and can be used to deal with nonlinear, non-stationary and noisy time series. The core idea of SSA is to extract the active components in the series for modeling and prediction. MLR and SVR are widely used machine learning algorithms for time series prediction. Since the original authors know their proposed algorithms best, the optimal parameters used in the above contrast algorithms are set to the values recommended in the original articles in order to obtain the best performance. The comparison between the proposed Res-LSTM model and other models is shown in Table 3.

Table 3. The comparison of the proposed Res-LSTM and other models.

Model	Parameters	Training Set (MSE)	Testing Set (MSE)		
			5 s	15 s	60 s
Proposed	$L = 2, S_{state} = 5$ $seed = 1, total_step = 400$	0.893	0.625	0.752	1.013
Holt–Winters	$\alpha = 0.037, \beta = 0.068, \gamma = 0.198$	2.971	1.594	2.839	3.153
SSA	$L_{ssa} = 80, G_{ssa} = list(1 : 60)$	0.815	1.041	0.938	1.173
MLR	$L_{mlr} = 36$	2.374	1.946	2.216	2.413
SVR	$L_{svr} = 36, C = 5$ $\epsilon = 0.243, \sigma_{svr} = 0.031$	1.748	1.371	1.709	1.814

4.4. Ablation Study

In order to verify that the modules introduced in this paper are helpful for the performance improvement of Res-LSTM model, relevant control experiments were carried out. Cross-entropy, L1-norm and mean absolute error (MAE) were introduced as contrast loss functions to conduct a comparative analysis. The above loss functions were applied to the proposed Res-LSTM model to replace MSE loss, and other hyperparameters remained unchanged. The results of training loss are shown in Figure 6.

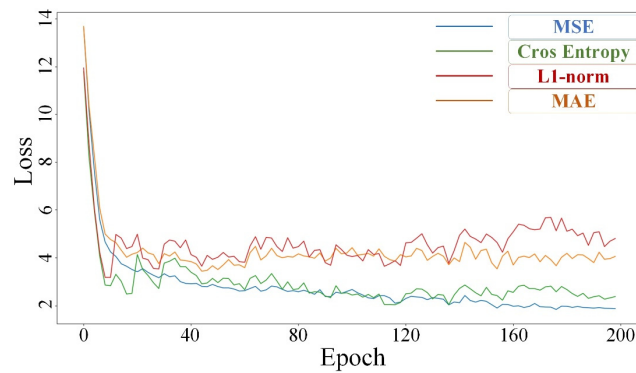


Figure 6. The training results of the proposed Res-LSTM model with different loss functions.

As can be seen in Figure 6, the model with MSE loss obtained fast and stable convergence. At the early training stage, the loss curves of four loss functions showed a rapid decline. However, loss functions of cross-entropy and MAE were unstable for the following training processes. In addition, L1-norm is not appropriate for the proposed model since there is no convergence in the last stage of the training process. The comparison results showed that MSE is a desirable loss function for the proposed model.

5. Application

The A-class insulation board production line consists of mixing equipment, a caterpillar laminating machine and cutting equipment. The proposed state prediction method for an A-class production line based on transfer learning can collect and monitor real-time data from various sensors. The pre-trained Res-LSTM model was used to fuse the real-time data of various sensors and predict the state of the A-class thermal insulation board production line. This proposed method has been applied to industrial production. The pictures of the production line equipment and sensor are shown in Figure 7, and the configuration of the industrial PC for the Res-LSTM model is listed in Table 4.



Figure 7. Pictures of production line equipment and sensors.

Table 4. The IPC configuration for Res-LSTM model inference.

No.	Device Name	Type
1	CPU	I7-8550U
2	RAM	16G DDR4
3	Hard drive	512G SSD
4	GPU	RTX 2070

The A-class insulation board production process consists of several steps, as displayed in Figure 8. Multiple sensors were deployed on the production line to obtain real-time data, which were fed into the well-trained Res-LSTM model.

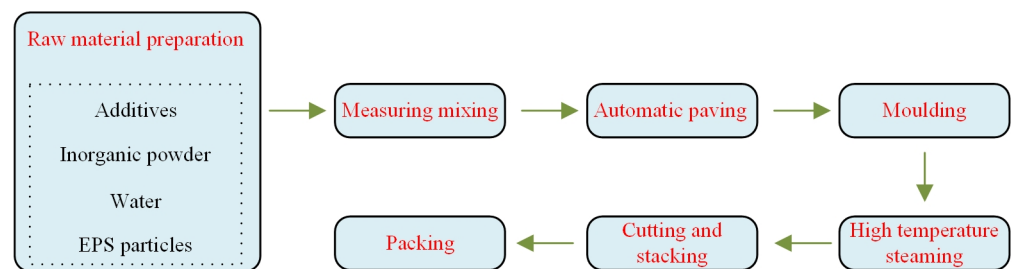


Figure 8. The A-class insulation board technological process.

In order to verify the effectiveness of the proposed prediction model, industrial experiments longer than 15 days were carried out. The prediction results after essential production steps, such as voltage, current and pressure, were recorded and compared with the actual value. The prediction and actual curves are shown in Figure 9.

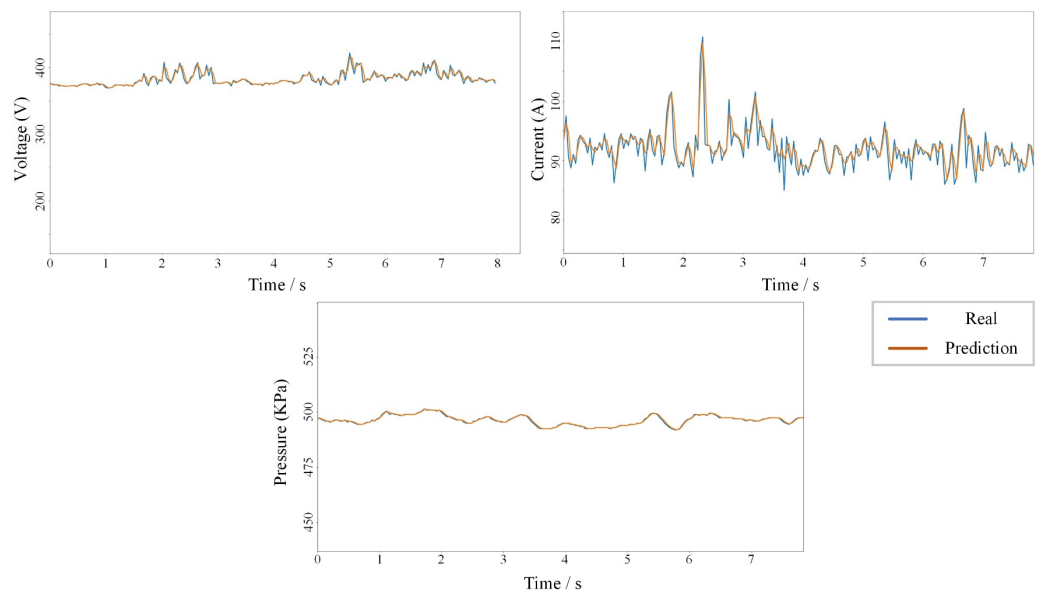


Figure 9. The comparison between captured data and prediction for device voltage and current in one node.

The R^2 -score was introduced to evaluate the prediction of the proposed Res-LSTM model and can be expressed as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_t)^2}{\sum_{i=1}^n (y_i - \hat{y}_p)^2} \tag{12}$$

where y_i denotes the real sample, and \hat{y}_t and \hat{y}_p denote means of real samples and prediction results, respectively. The R^2 -score can reflect the performance of linear regression models, and it outputs a value in the range of (0, 1). The model has a better prediction if the R^2 -score is closer to 1. The R^2 -scores of voltages, currents and pressures in Figure 9 are 0.89, 0.93 and 0.98, respectively. The average of predictions is 0.93, which means the prediction results of the proposed model fit very well with real samples.

6. Conclusions

In this paper, a state prediction method for an A-class insulation panel production line based on transfer learning is proposed. The Res-block is introduced in the proposed model to improve the hidden feature extraction. The transfer learning theory and improved loss function are applied to make the training process of the proposed model fast and stable. Real-time data from an A-class insulation panel production line were collected by placing multiple sensors at key nodes. The well-trained model based on transfer learning was validated by multiple sets of experiments. The R^2 -score is introduced to evaluate the prediction results for industrial experiments. The experimental results show that the accuracy of the proposed state prediction method reached 98.9%, and the average R^2 -score reached 0.93. Therefore, the proposed Res-LSTM model can accurately predict the running state of the production line and fully meet the needs of industrial production.

The method proposed in this paper still can be improved in future work. Considering the stage of multi-sensor data fusion, the weights of data from each sensor cannot be adjusted adaptively, which may reduce the accuracy of the proposed method for the state prediction of the production line. If the model is deployed to another production line, the proposed model needs to be trained in advance and cannot update its weights during inference. This means the model needs to be retrained in the case of migration, which limits the adaptive ability of the proposed model. In future work, an improved data preprocessing algorithm can be researched to enhance the recognition ability of the proposed Res-LSTM model. In addition, self-learning strategy should be studied to improve the migration ability, prediction accuracy and adaptive ability of the proposed Res-LSTM model.

Author Contributions: Conceptualization, Y.W. and X.L. (Xiaowen Liu); methodology, Y.W.; software, Y.W.; validation, H.W., X.G. and X.L. (Xinhua Liu); formal analysis, H.W.; investigation, X.G.; resources, X.L. (Xinhua Liu); data curation, X.G.; writing—original draft preparation, Y.W.; writing—review and editing, H.W.; visualization, Y.W.; supervision, X.L. (Xiaowen Liu); project administration, X.L. (Xiaowen Liu); funding acquisition, H.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Postgraduate Research & Practice Innovation Program of Jiangsu Province, grant number SJCX22_1139.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding the publication of this article.

References

1. Cintura, E.; Nunes, L.; Esteves, B.; Faria, P. Agro-industrial wastes as building insulation materials: A review and challenges for Euro-Mediterranean countries. *Ind. Crops Prod.* **2021**, *171*, 113833. [[CrossRef](#)]
2. Giannotas, G.; Kamperidou, V.; Barboutis, I. Tree bark utilization in insulating bio-aggregates: A review. *Biofuels Bioprod. Biorefin.-Biofpr* **2021**, *15*, 1989–1999. [[CrossRef](#)]
3. Pan, S.C.; Rickard, T.C. Transfer of Test-Enhanced Learning: Meta-Analytic Review and Synthesis. *Psychol. Bull.* **2018**, *144*, 710–756. [[CrossRef](#)]
4. Simoes, R.S.; Maltarollo, V.G.; Oliveira, P.R.; Honorio, K.M. Transfer and Multi-task Learning in QSAR Modeling: Advances and Challenges. *Front. Pharmacol.* **2018**, *9*, 74. [[CrossRef](#)] [[PubMed](#)]
5. Webb, P.; Whitlow, J.W., Jr.; Venter, D. From Exploratory Talk to Abstract Reasoning: A Case for Far Transfer? *Educ. Psychol. Rev.* **2017**, *29*, 565–581. [[CrossRef](#)]
6. Bierly, P.E., III; Daly, P.S. Sources of external organisational learning in small manufacturing firms. *Int. J. Technol. Manag.* **2007**, *38*, 45–68.

7. Song, X.; Gao, S.; Liu, X.; Chen, C. An outdoor fire recognition algorithm for small unbalanced samples. *Alex. Eng. J.* **2021**, *60*, 2801–2809. [[CrossRef](#)]
8. Xue, J.; Chan, P.P.K.; Hu, X. Experimental study on stacked autoencoder on insufficient training samples. In Proceedings of the IEEE International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), Ningbo, China, 9–12 July 2017; pp. 223–229.
9. Zhao, W. Research on the Deep Learning of the Small Sample Data based on Transfer Learning. In Proceedings of the International Conference on Green Energy and Sustainable Development (GESD), Chongqing, China, 27–28 May 2017.
10. Sun, S.-L.; Zhang, L.-Y.; Zhu, J.; Zhu, D.-X. Design of the Control System for the Biomass Briquette Fuel Production Line. In Proceedings of the International Conference of Green Buildings and Environmental Management (GBEM), Qingdao, China, 23–25 August 2018.
11. Izonin, I.; Tkachenko, R.; Kryvinska, N.; Tkachenko, P.; Greguš ml, M. Multiple Linear Regression Based on Coefficients Identification Using Non-iterative SGTm Neural-like Structure. In Proceedings of the Advances in Computational Intelligence, Gran Canaria, Spain, 12–14 June 2019; pp. 467–479.
12. Cui, J.; Zhang, X.; Zhu, J. Design and Realization of Remote Monitoring System for Automatic Forced Fitting Production Line. In Proceedings of the 2nd International Conference on Manufacturing Science and Engineering, Guilin, China, 9–11 April 2011; pp. 2354–2359.
13. Sun, L.; Han, W.; Yan, W.; Sun, L.; Yan, L. Research on Centralized Monitoring System of LiMnO₂ Button Cells Automatic Assembly Line. In Proceedings of the International Industrial Informatics and Computer Engineering Conference (IIICEC), Xi'an, China, 10–11 January 2015; pp. 224–227.
14. Chen, W. Intelligent manufacturing production line data monitoring system for industrial internet of things. *Comput. Commun.* **2020**, *151*, 31–41. [[CrossRef](#)]
15. Tkachenko, R.; Izonin, I. Model and Principles for the Implementation of Neural-Like Structures Based on Geometric Data Transformations. In *Advances in Computer Science for Engineering and Education, Proceedings of the International Conference on Computer Science, Engineering and Education Applications, Hohhot, China, 22–24 October 2018*; Springer: Cham, Switzerland, 2018; pp. 578–587. [[CrossRef](#)]
16. Tkachenko, R. An Integral Software Solution of the SGTm Neural-Like Structures Implementation for Solving Different Data Mining Tasks. In *Lecture Notes in Computational Intelligence and Decision Making*; Springer: Cham, Switzerland, 2022; pp. 696–713.
17. Vijayaraghavan, A.; Dornfeld, D. Automated energy monitoring of machine tools. *CIRP Ann.-Manuf. Technol.* **2010**, *59*, 21–24. [[CrossRef](#)]
18. Duro, J.A.; Padget, J.A.; Bowen, C.R.; Kim, H.A.; Nassehi, A. Multi-sensor data fusion framework for CNC machining monitoring. *Mech. Syst. Signal Process.* **2016**, *66–67*, 505–520. [[CrossRef](#)]
19. Zhao, R.; Yan, R.Q.; Chen, Z.H.; Mao, K.Z.; Wang, P.; Gao, R.X. Deep learning and its applications to machine health monitoring. *Mech. Syst. Signal Process.* **2019**, *115*, 213–237. [[CrossRef](#)]
20. Greff, K.; Srivastava, R.K.; Koutnik, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A Search Space Odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 2222–2232. [[CrossRef](#)]
21. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. The Performance of LSTM and BiLSTM in Forecasting Time Series. In Proceedings of the IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 3285–3292.
22. Shahid, F.; Zameer, A.; Muneeb, M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos Solitons Fractals* **2020**, *140*, 110212. [[CrossRef](#)] [[PubMed](#)]
23. Ordonez, F.J.; Roggen, D. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors* **2016**, *16*, 115. [[CrossRef](#)] [[PubMed](#)]
24. Kong, W.C.; Dong, Z.Y.; Jia, Y.W.; Hill, D.J.; Xu, Y.; Zhang, Y. Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. *IEEE Trans. Smart Grid* **2019**, *10*, 841–851. [[CrossRef](#)]
25. Sun, J.L.; Shi, W.J.; Yang, Z.T.; Yang, J.; Gui, G. Behavioral Modeling and Linearization of Wideband RF Power Amplifiers Using BiLSTM Networks for 5G Wireless Systems. *IEEE Trans. Veh. Technol.* **2019**, *68*, 10348–10356. [[CrossRef](#)]
26. Hua, Y.S.; Mou, L.C.; Zhu, X.X. Recurrently exploring class-wise attention in a hybrid convolutional and bidirectional LSTM network for multi-label aerial image classification. *ISPRS J. Photogramm. Remote Sens.* **2019**, *149*, 188–199. [[CrossRef](#)]
27. Che, Z.P.; Purushotham, S.; Cho, K.; Sontag, D.; Liu, Y. Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Sci. Rep.* **2018**, *8*, 6085. [[CrossRef](#)] [[PubMed](#)]
28. Chen, J.L.; Jing, H.J.; Chang, Y.H.; Liu, Q. Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process. *Reliab. Eng. Syst. Saf.* **2019**, *185*, 372–382. [[CrossRef](#)]
29. Kowald, A.; Barrantes, I.; Möller, S.; Palmer, D.; Murua Escobar, H.; Schwerk, A.; Fuellen, G. Transfer learning of clinical outcomes from preclinical molecular data, principles and perspectives. *Brief. Bioinform.* **2022**, *23*, 330–346. [[CrossRef](#)]
30. Wong, L.J.; Michaels, A.J. Transfer Learning for Radio Frequency Machine Learning: A Taxonomy and Survey. *Sensors* **2022**, *22*, 1416. [[CrossRef](#)] [[PubMed](#)]
31. Lee, C.; Sohn, K.-T. The Relationship Between Learning Flow, Subjective Learning Performance, Self-efficacy to Transfer, and Transfer Intention Perceived by Participants of Synchronous E-learning Program. *Korean Assoc. Learn. -Cent. Curric. Instr.* **2022**, *22*, 133–145. [[CrossRef](#)]

32. Long, M.; Cao, Y.; Wang, J.; Jordan, M.I. Learning Transferable Features with Deep Adaptation Networks. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 97–105.
33. Tzeng, E.; Hoffman, J.; Zhang, N.; Saenko, K.; Darrell, T. Deep Domain Confusion: Maximizing for Domain Invariance. *arXiv* **2014**, arXiv:1412.3474.
34. Ganin, Y.; Lempitsky, V. Unsupervised Domain Adaptation by Backpropagation. *arXiv* **2014**, arXiv:1409.7495.
35. Saito, K.; Ushiku, Y.; Harada, T. Asymmetric Tri-training for Unsupervised Domain Adaptation. In Proceedings of the International Conference on Machine Learning 2017, Sydney, Australia, 6–11 August 2017.
36. Manders, J.; Laarhoven, T.V.; Marchiori, E. Simple Domain Adaptation with Class Prediction Uncertainty Alignment. *arXiv* **2018**, arXiv:1804.04448.
37. Pei, Z.; Cao, Z.; Long, M.; Wang, J. Multi-Adversarial Domain Adaptation. In Proceedings of the 32nd AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018.