machines

MDPI

Article

Bearing Health State Detection Based on Informer and CNN + Swin Transformer

Chunyang Liu [1,2], Weiwei Zou [1], Zhilei Hu [1], Hongyu Li [1,*], Xin Sui [1,3], Xiqiang Ma [1,2], Fang Yang [1,4] and Nan Guo [1,4]

[1] School of Mechatronics Engineering, Henan University of Science and Technology, Luoyang 471003, China; chunyangliu@haust.edu.cn (C.L.); zww02172022@163.com (W.Z.); 15036578316@163.com (Z.H.); sxwd@163.com (X.S.); maxiqiang@haust.edu.cn (X.M.); yangfanghkd@163.com (F.Y.); guonan1860@163.com (N.G.)
[2] Longmen Laboratory, Luoyang 471000, China
[3] Key Laboratory of Mechanical Design and Transmission System of Henan Province, Luoyang 471000, China
[4] Collaborative Innovation Center of Machinery Equipment Advanced Manufacturing of Henan Province, Luoyang 471000, China
* Correspondence: lihy_026@163.com; Tel.: +86-15137100517

Abstract: In response to the challenge of timely fault identification in the spindle bearings of machine tools operating in complex environments, this study proposes a method based on a combination of infrared imaging with an Informer and a CNN + Swin Transformer. The aim is to achieve real-time monitoring of bearing faults, precise fault localization, and classification of fault severity. To accomplish this, an angular contact ball bearing was chosen as the research subject. Initially, an infrared image dataset was constructed, encompassing various fault positions and degrees, by simulating different forms of bearing faults. Subsequently, an Informer-based bearing temperature prediction model was established to select faulty bearing data. Lastly, the faulty data were input into the CNN + Swin Transformer model for bearing fault recognition and classification. The results demonstrate that the Informer model accurately identifies abnormal temperature rises during bearing operation, effectively screening out faulty bearings. Under steady-state conditions, the model achieves a classification accuracy of 97.8%. Furthermore, after employing the Informer screening process, the proposed model exhibits a recognition precision of 98.9%, surpassing other models such as CNN, SVM, and Swin Transformer, which are mentioned in this paper.

Keywords: infrared image; fault recognition; CNN; swin transformer

check for updates

Citation: Liu, C.; Zou, W.; Hu, Z.; Li, H.; Sui, X.; Ma, X.; Yang, F.; Guo, N. Bearing Health State Detection Based on Informer and CNN + Swin Transformer. *Machines* **2024**, *12*, 456. https://doi.org/10.3390/machines12070456

Academic Editor: Davide Astolfi

Received: 3 June 2024
Revised: 28 June 2024
Accepted: 3 July 2024
Published: 4 July 2024

Copyright: © 2024 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

The precision spindle bearings of machine tools play a crucial role in supporting the operation of the spindle and transmitting power. The stability and reliability of these bearings directly impact the machining quality and production efficiency of the machine tool. Operating under high-speed and high-temperature conditions, it becomes challenging to detect and evaluate the health status of bearings. Any malfunction can lead to severe consequences such as equipment downtime. Therefore, monitoring and assessing the health condition of precision spindle bearings are beneficial to ensure the normal and stable operation of the system.

Currently, there are two main methods for diagnosing and evaluating the health status of bearings: signal analysis and deep learning [1]. Signal analysis methods commonly employ time-domain analysis, frequency-domain analysis, and time–frequency-domain analysis to extract signal features and assess their state in conjunction with theoretical models and empirical formulas. Nikula [2] proposed an automatic method to detect bearing faults under low-speed conditions, filtering signals in a specific frequency range and segmenting them into short time windows. Window length is selected based on the

defect frequency of the bearing, and statistical signal time-domain features are calculated. The results demonstrate that this method has good applicability even for short signal lengths. Aasi [3] obtained acoustic signals of defective bearings for investigating the feasibility of using them for characterization for bearing condition monitoring. Results indicate that impulse, kurtosis, and crest factors are appropriate features for diagnosis purposes. Fu [4] proposed a mechanical fault diagnosis method using vibration signals. Different fault types are identified through the calculation of time-domain parameters and adaptive fuzzy mean clustering. The results show that the method is highly sensitive for fault identification, including minor faults. Zhao [5] proposed a novel frequency matching demodulation transform (FMDT) technique for extracting weak fault features in bearings under variable speeds. This method accurately estimates fault characteristic frequencies and identifies fault types without requiring the measured rotational frequency, demonstrating superior performance compared to traditional time–frequency analysis-based methods. Gougam [6] introduced an augmented technique for diagnosing rolling bearing faults, combining two time-domain features (TDFs) with singular value decomposition (SVD) and fuzzy logic systems (FLS), which can improve the accuracy and sensitivity of bearing fault assessment under different working conditions. Zhao [7] developed a high-concentration time–frequency analysis technique called frequency-chirprate synchrosqueezing-based scaling chirplet transform (FCSSCT) for effectively analyzing nonstationary and closely spaced fault frequencies in wind turbine gearboxes and bearings, demonstrating improved energy concentration and fault characterization compared to existing methods. Lu [8] proposed an effective fault diagnosis algorithm for rolling bearings based on clustering and sparse representation. The algorithm first clusters samples using their frequency spectrums, and then trains an adaptive redundant dictionary to reduce noise in test samples. Identification of specific sample categories is achieved by selecting the maximum cosine similarity value between the reconstructed sample and cluster centers. Jawad [9] analyzed the combination of time and frequency domains, studied fault characteristic frequencies under different healthy states, and compared them to theoretical values, proving the effectiveness of this method in fault detection. Frequency-domain analysis is primarily suitable for linear time-invariant systems and may have limited capability in handling nonlinear signals, which can hinder accurate diagnosis of nonlinear fault characteristics in bearing systems.

Based on deep learning, fault diagnosis methods employ neural networks to extract data features, significantly enhancing the accuracy and adaptability of fault recognition algorithms. Liang [10] developed a novel deep learning framework, utilizing hierarchical models and transfer learning, along with sensor/data fusion to enhance fault diagnosis performance for rolling bearings in complex machinery systems with unknown fault locations. This methodology was validated through systematic case studies using publicly available experimental rolling bearing datasets, achieving an accuracy of 94.59%. Sun [11] proposed an optimized model based on convolutional neural networks (CNNs), converting vibration signals into symmetric images in polar coordinates and feeding them into a CNN for automatic fault diagnosis, thereby greatly reducing errors. Chaleshtori [12] introduced a novel approach combining Weighted Principal Component Analysis (WPCA) and a Gaussian Mixture Model (GMM) for bearing fault diagnosis, which was validated through case studies with datasets from the University of Ottawa and Case Western Reserve University, achieving average accuracy rates of 93% and 80%, respectively. Unal [13] explored fault diagnosis in rotary complex machines, focusing on the impact of faulty bearings on neighboring components. They proposed methods to extract features from vibration signals using envelope analysis, Hilbert Transform, and Fast Fourier Transform, and validated an artificial neural network (ANN) based on a fault estimation algorithm through experimental tests, achieving improved classification results with an optimized ANN model using a genetic algorithm. Tang [14] proposed a minimum unscented Kalman filter-aided deep belief network for bearing fault diagnosis, transforming multi-sensor vibration signals into 2-D feature maps and achieving over 98% accuracy on datasets,

which demonstrates high precision and generalization capabilities. Zhang [15] introduced a method that transforms original signals into two-dimensional images and utilizes convolutional neural networks for feature extraction and fault diagnosis. The impact of different sample sizes and load conditions on the diagnostic capabilities of this method were also analyzed. Su [16] developed a knowledge-informed deep learning approach for robust fault diagnosis of rolling bearings, integrating prior knowledge-based features with data-driven machine learning using a Knowledge-Informed Deep Network (KIDN). They also introduced a novel adaptive network design strategy based on a Constrained Gaussian Process (CGP) to enhance generalizability and efficiency in network architecture selection, which was validated through rigorous experimental case studies. Zhang [17] presented a rotating mechanical fault recognition method based on recurrent neural networks (RNN), where one-dimensional time-series vibration signals are converted into two-dimensional images. Gate Recurrent Units (GRU) are employed to process time-series data and learn representative features, followed by the utilization of Multilayer Perceptrons (MLP) for fault recognition. An [18] proposed an intelligent fault diagnosis framework consisting of data segmentation, Long Short-Term Memory (LSTM) networks, and output networks, achieving improved accuracy for bearing fault diagnosis through a simpler structure. However, despite the effective mitigation of issues like vanishing and exploding gradients by LSTM models [19], these types of recurrent neural networks still exhibit inherent limitations in terms of training time when dealing with long sequences.

In recent years, Transformer architecture has continuously garnered favor among researchers. Yang [20] proposed a Transformer neural network-based method for diagnosing bearing faults, leveraging a pure attention mechanism to linearly encode and positionally encode raw vibration signals for feature extraction and fault identification. Validation using public datasets demonstrated that this method achieved a notably high average diagnostic accuracy even without any data preprocessing. Alexakos [21] applied a short-time Fourier transform to vibration data to obtain time–frequency spectrograms. Subsequently, an enhanced Transformer model was employed for image classification through supervised learning, achieving an accuracy of 98.3% in testing. Tang [22] introduced the Vision Transformer into the field of bearing fault diagnosis, addressing the issue of traditional convolutional neural networks failing to capture temporal information in rolling bearing fault diagnosis. This was accomplished by designing multi-scale convolutional fusion layers to obtain multi-scale features and introducing an improved visual Transformer structure to learn long-term temporal correlations, thereby significantly enhancing diagnostic accuracy and noise resistance. Liu [23] presented a visual detection model called the Swin Transformer, which possesses the capability to adapt to different scales and effectively address differences between the image and text domains. It has demonstrated outstanding performance in various visual tasks, such as image classification, object detection, and semantic segmentation, showcasing the potential of Transformer-based models as the backbone network for visual detection.

Regarding the aforementioned study, this paper proposes a novel method for diagnosing the health status of bearings by combining infrared temperature measurement with image detection and recognition. A bearing test platform was established using an infrared camera for temperature measurement. The emissivity of the infrared thermal imager was adjusted and optimized to achieve accurate temperature measurement based on the temperature field image. A temperature rise prediction model based on an Informer structure and a fault type identification model based on a CNN + Swin Transformer were developed, which combined temperature prediction with fault type and severity diagnosis. This approach enables a comprehensive evaluation of the health status of bearings.

2. Test Setup and Data Acquisition

As shown in Figure 1, the experimental setup comprises an infrared thermal imaging camera, an axial loading device, and temperature data acquisition and analysis software on a PC. The experimental bearing is positioned with the end face upwards and is subjected

to axial loading. The axial force generated by the loading mechanism first acts on the bearing housing cover and is then transmitted to the test bearing. The bearing used for the experiment is an H7006C angular contact ball bearing, with the following main structural parameters: an inner ring diameter (d) of 30 mm; an outer ring diameter (D) of 55 mm; a width (B) of 13 mm; a rolling body diameter (Dw) of 5.556 mm; a rolling body count (Z) of 16; and a contact angle (α) of 15°.

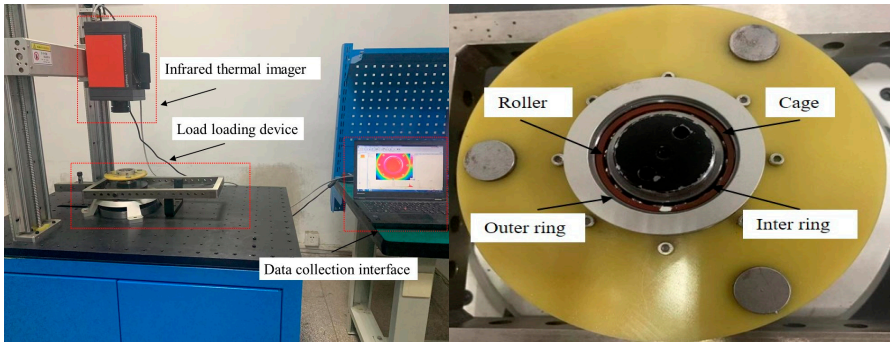


Figure 1. Infrared image acquisition equipment for angular contact ball bearings.

The experimental equipment is capable of monitoring the real-time rotational speed of the inner ring and cage of the bearing, as well as the axial load and surface temperature field. The infrared thermal imager used in this study is the ImagelR8355 model, which has a temperature measurement range of −10~+175 °C; an infrared image sampling frequency of 0~110 Hz; and a measurement accuracy of ±1 °C.

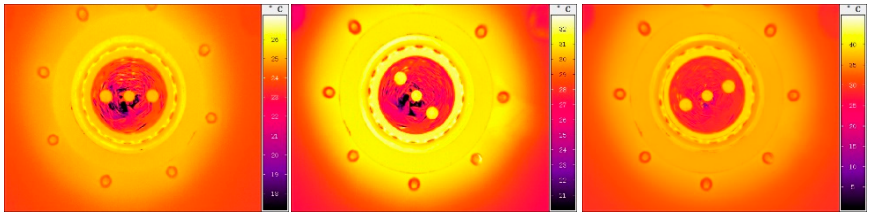
The experimental bearing failure types include slight inner race wear (SIRF), moderate inner race wear (MIRF), severe inner race wear (LIRF), slight outer race wear (SORF), moderate outer race wear (MORF), and severe outer race wear (LORF). The bearing part of the failure is shown in Figure 2.



(**a**) (**b**) (**c**)

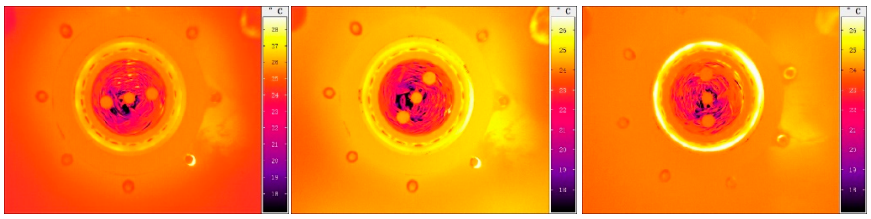
Figure 2. Bearing outer ring defects. (**a**) Slight outer race wear; (**b**) moderate outer race wear; (**c**) severe outer race wear.

The data collection process encompassed 15 different operating conditions, capturing both the warming-up and steady phases, resulting in a total of 4200 images to construct the faulty bearing infrared image dataset. Some of the collected results are depicted in Figure 3.



(**a**) (**b**) (**c**)

Figure 3. *Cont.*



(**d**) (**e**) (**f**)

Figure 3. Infrared images of faulty bearing. (**a**) Light wear on outer ring; (**b**) moderate wear on outer ring; (**c**) heavy wear on outer ring; (**d**) light wear on inner ring; (**e**) moderate wear on inner ring; (**f**) heavy wear on inner ring.

3. Models

3.1. Bearing Health Condition Diagnosis Method Combining Temperature Rise Prediction and Image Classification

This paper presents a bearing health diagnosis method that combines temperature rise prediction and image classification. The specific procedure is illustrated in Figure 4. Firstly, an infrared thermal imaging camera is utilized to monitor the bearing and collect infrared temperature field images, from which temperature values at measurement points are extracted. Next, the Informer model is employed along with historical temperature data of the bearing at the current time to predict the temperature data for the next time step. The predicted values are then compared with the actual measurement results of the next time step. If there is a significant discrepancy between the two, it indicates the possibility of a bearing fault. At this point, the infrared image data are input into a CNN for image feature extraction, followed by employing the Swin Transformer model for fault recognition, determining whether a fault has occurred, and identifying the type and severity of the fault.

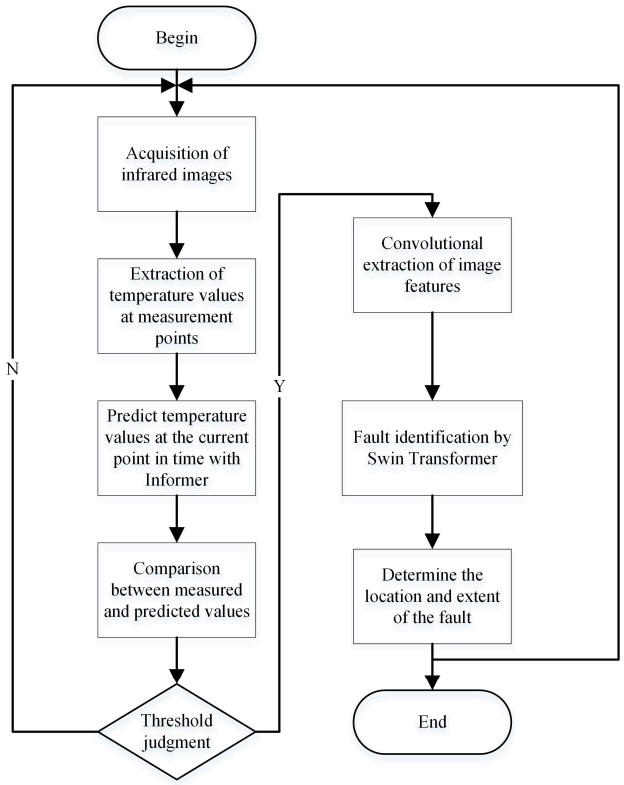


Figure 4. Bearing condition inspection flow.

3.2. Informer

The Informer [24] model is a supervised learning model based on attention mechanisms, consisting of two main components: an encoder and a decoder. The encoder is responsible for extracting features from the original input sequence and modeling its long-term dependencies through multi-head self-attention mechanisms. On the other hand, the decoder predicts future sequences by leveraging the information obtained from the encoder and the history of previous predictions. This architecture makes the Informer highly adaptable to time-series data and capable of effectively handling the problem of long-term dependencies in time-series analysis. Additionally, the overall structure of the Informer is illustrated in Figure 5.

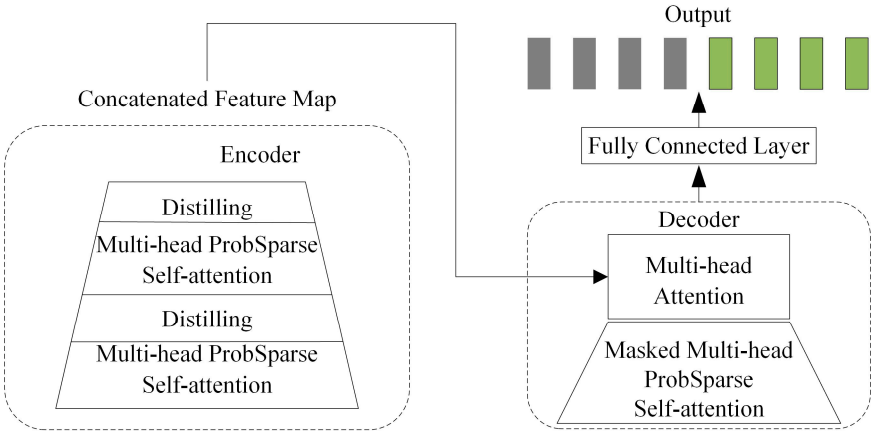


Figure 5. Informer model structure.

The Informer filters out important query matrices by distinguishing the similarity between the probability distribution and the uniform distribution of each query matrix Q through the Kullback–Leibler dispersion formula, which is as follows:

$$KL(q||p) = \ln \sum_{l=1}^{L_K} e^{\frac{q_i k_l^T}{\sqrt{d}}} - \frac{1}{L_K} \sum_{j=1}^{L_K} \frac{q_i k_j^T}{\sqrt{d}} - \ln L_K \tag{1}$$

Since the length of the input sequence is determined in Equation (1), $\ln L_K$ is a constant value that has no effect on the computation process, and so in calculating the sparsity score of the i-th query (Q), the Equation can be expressed as follows:

$$M(q_i, K) = \ln \sum_{l=1}^{L_K} e^{\frac{q_i k_l^T}{\sqrt{d}}} - \frac{1}{L_K} \sum_{j=1}^{L_K} \frac{q_i k_j^T}{\sqrt{d}} \tag{2}$$

where p denotes the self-attentive probability distribution; q denotes the uniform probability distribution; LK denotes the length of the sequence; and d denotes the dimension of the input data sequence after mapping.

The probabilistic self-attention mechanism has been proved to obey a sparse distribution and has the characteristic of long-tailedness, i.e., the query vector (Q) is divided into active and inert query vectors, and in the process of computation, we only care about the dot-product value of the active query, and use the average value of the value vector instead of the dot-product value of the inert vector, and in the process of calculating the dot-product value, only a small portion of dot-products have an effect on the attention mechanism, which reduces the computational task. Based on the above calculation results,

the probabilistic sparse self-attention mechanism is obtained by utilizing Equation (3) as follows:

$$A(Q,K,V) = Soft\ max(\frac{\overline{Q}K^T}{\sqrt{D_K}})V \tag{3}$$

where $\overline{Q}$ is the matrix obtained by the probabilistic sparse of Q; and *softmax* is the normalized activation function.

The encoder employs a "distillation" operation, prioritizing the processing of prominent high-level features and generating concentrated self-attention feature maps in the lower layers, thus producing self-attention feature maps in the upper layers. This helps reduce the input length. At time t, the "distillation" operation from the j-th layer to the $j+1$-th layer can be described as follows:

$$X_{j+1}^t = MaxPool\Big(ELU\Big(Conv1d\Big[X_j^t\Big]_{AB}\Big)\Big) \tag{4}$$

where $[*]_{AB}$ basic operations contain attention blocks and sparse attention mechanisms; $Conv$1d denotes the one-dimensional convolution operation; ELU is the activation function; and $MaxPool$ is the maximum pooling operation.

The purpose of the decoder design is to generate long sequence predictions through a single forward pass. The model adopts a traditional decoder structure, including two identical multi-head attention layers, to address the issue of the high time complexity involved in generating predictions for long sequence data. The input vector representation of the decoder is as follows:

$$X_{de}^t = Concat(X_{token}^t, X_0^t) \in R^{(L_{token}+L_y)d_{model}} \tag{5}$$

where X_{token}^t is the start token; and X_0^t is a placeholder for the target sequence, which is set to 0.

3.3. Diagnostic Modeling with a CNN + Swin Transformer

In order to enhance the accuracy of fault type and severity identification, this paper adopts a two-stage design approach. As shown in Figure 6, initially, a convolutional neural network (CNN) is employed for feature extraction from the input data. The image feature extraction stage involves two convolutional layers and two pooling layers, gradually capturing both local and global features of the image. The first convolutional layer utilizes 32 3 × 3 convolutional kernels, while the second convolutional layer uses 64 3 × 3 convolutional kernels. Pooling layers are utilized to compress the feature map dimensions while retaining essential feature information. The fully connected layer integrates these features. The feature maps obtained by the convolutional neural network are then divided into non-overlapping patches with a height and width of P = 4 and a feature dimension of 48 through a Patch partition layer. This allows the Swin Transformer Block to directly process images with a two-dimensional structure.

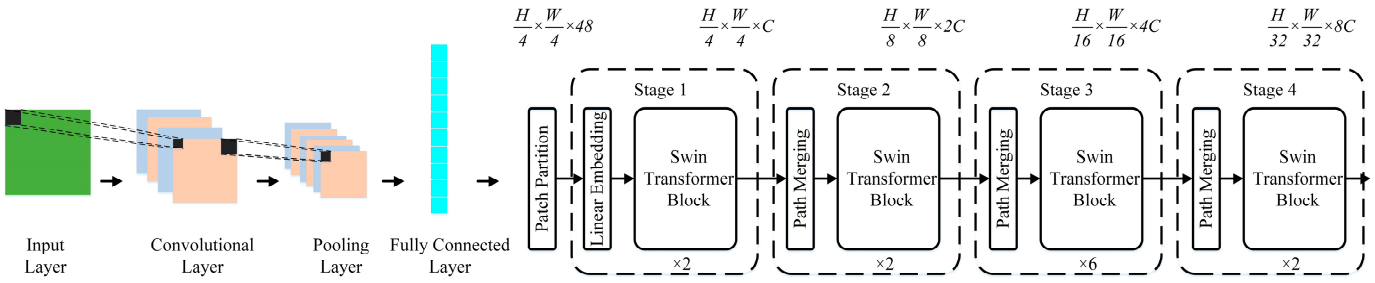


Figure 6. Fault diagnosis recognition model structure of a CNN + Swin Transformer.

The Swin Transformer consists of four stages, each containing similar repeating units. The Patch Partition module splits the input image of size H × W × 3 into non-overlapping

patches of equal sizes. These patches are then transformed into a sequence of vectors. The Linear Embedding operation applies a linear transformation to the input vector sequence, enhancing the network's nonlinear representation capability and aiding in better understanding the semantic and structural information within the image. The transformed vector sequence is then passed through the Swin Transformer Block for feature extraction.

As illustrated in Figure 7, the Swin Transformer Block is composed of Window Multi-head Self-Attention (*W-MSA*) and Shifted-Window Multi-head Self-Attention (*SW-MSA*), which alternate. The *W-MSA* introduces window operation to uniformly divide image patches into non-overlapping windows, each containing adjacent M × M patches. This allows for independent self-attention computation within each window. However, due to a lack of inter-window connections, the *W-MSA* cannot model global features. To address this issue, the Swin Transformer further proposes the *SW-MSA* mechanism. In the *SW-MSA*, the regular windows in the *W-MSA* module are replaced, starting from ([M/2], [M/2]) pixels, which introduces connections between adjacent non-overlapping windows from the previous module to enhance the model's global modeling capability.

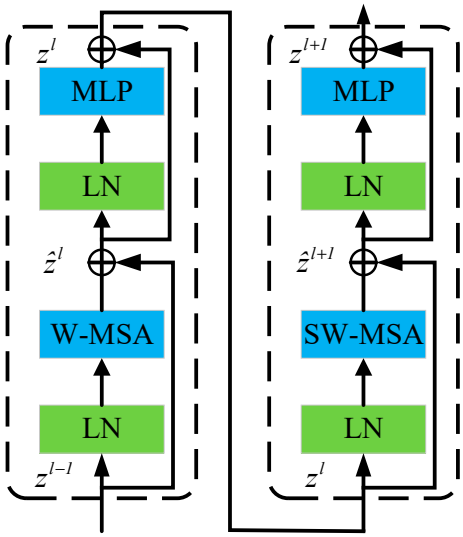


Figure 7. Swin Transformer Block.

The computation for two consecutive Swin Transformer Blocks is as follows:

$$\hat{z}^{l} = W - MSA(LN(z^{l-1})) + z^{l+1} \tag{6}$$

$$\hat{z}^{l} = MLP(LN(\hat{z}^{l})) + \hat{z}^{l} \tag{7}$$

$$\hat{z}^{l+1} = SW - MSA(LN(z^{l})) + z^{l} \tag{8}$$

$$z^{l+1} = MLP(LN(\hat{z}^{l+1})) + \hat{z}^{l+1} \tag{9}$$

where $\hat{z}^{l}$ a and $\hat{z}^{l+1}$ denote the output features of the (S)W-MSA layer and the *MLP* layer, respectively, of module l. The classification results are output through global pooling and the fully connected layer after the execution of the last Swin Transformer Block is completed. *W-MSA* denotes the Window-based Multiple Attention, and *SW-MSA* denotes the Moving Window Multiple Attention.

4. Testing and Analysis

4.1. Informer Model Prediction Results

To validate the performance of the temperature rise prediction model, experiments were conducted using healthy bearings under various operating conditions. Multiple sets of experimental data were collected, and the trained Informer model was used for prediction, with the mean absolute error serving as the evaluation metric. The experimental

results for the measured and predicted data during the bearing heating and steady-state stages are shown below.

By examining Figure 8 on the left, it becomes evident that the predictive performance of the Informer model is remarkably excellent under various operating conditions. The predicted curve gradually aligns with the actual curve. During the initial phase of bearing operation, the temperature progressively rises and the rate of increase gradually decelerates. After 30 min, the bearing enters a stable state of operation, where the temperature stabilizes and converges to different values depending on the specific operating condition. Analyzing the error results depicted in Figure 8 on the right, it can be observed that, across different operating conditions, the mean absolute error (MAE) between the predicted values generated by the Informer model and the actual values of the bearing temperature during both the heating and stable phases is less than 0.1 °C. Thus, the predictive outcomes effectively reflect the genuine temperature variations experienced by a healthy bearing throughout its operational lifespan.

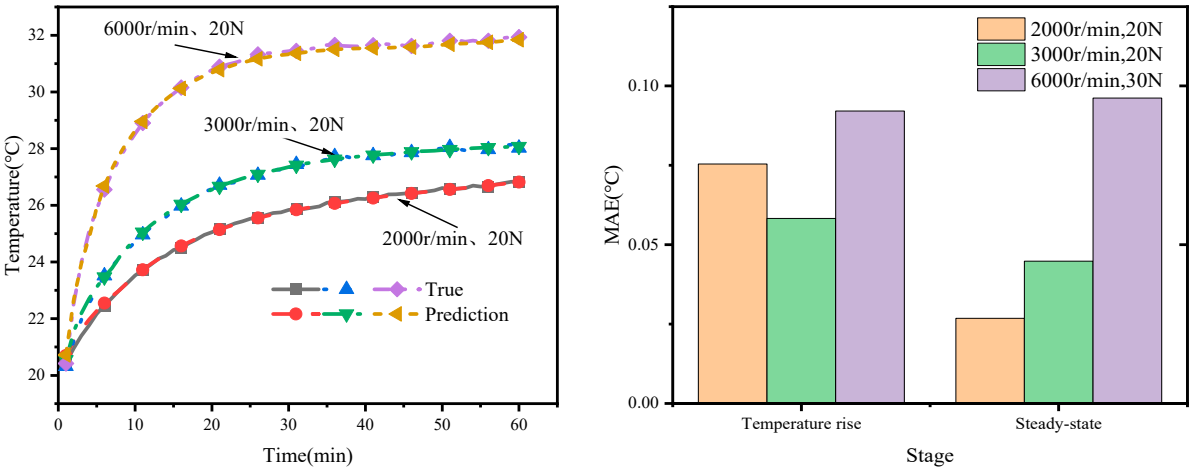


Figure 8. Healthy bearing temperature rise prediction results and errors.

Testing experiments were conducted using a malfunctioning bearing, and the collected temperature rise data were compared with the predicted temperature rise results generated by the Informer model. As shown in Figure 9, notable differences were observed between the predicted outcomes and the actual values during the temperature rise process across multiple operating conditions. By the 10th minute, the error between the measured value and the predicted value had already approached 2 °C. Moreover, as time progressed, the maximum error of the predicted results reached up to 3 °C. It is evident that an increase in frictional power consumption between the various components of the bearing due to wear-induced faults generates a higher amount of heat, causing the temperature value to exceed that of a healthy state. Predicting the temperature rise process of malfunctioning bearings through the Informer model can help determine if the bearing has failed.

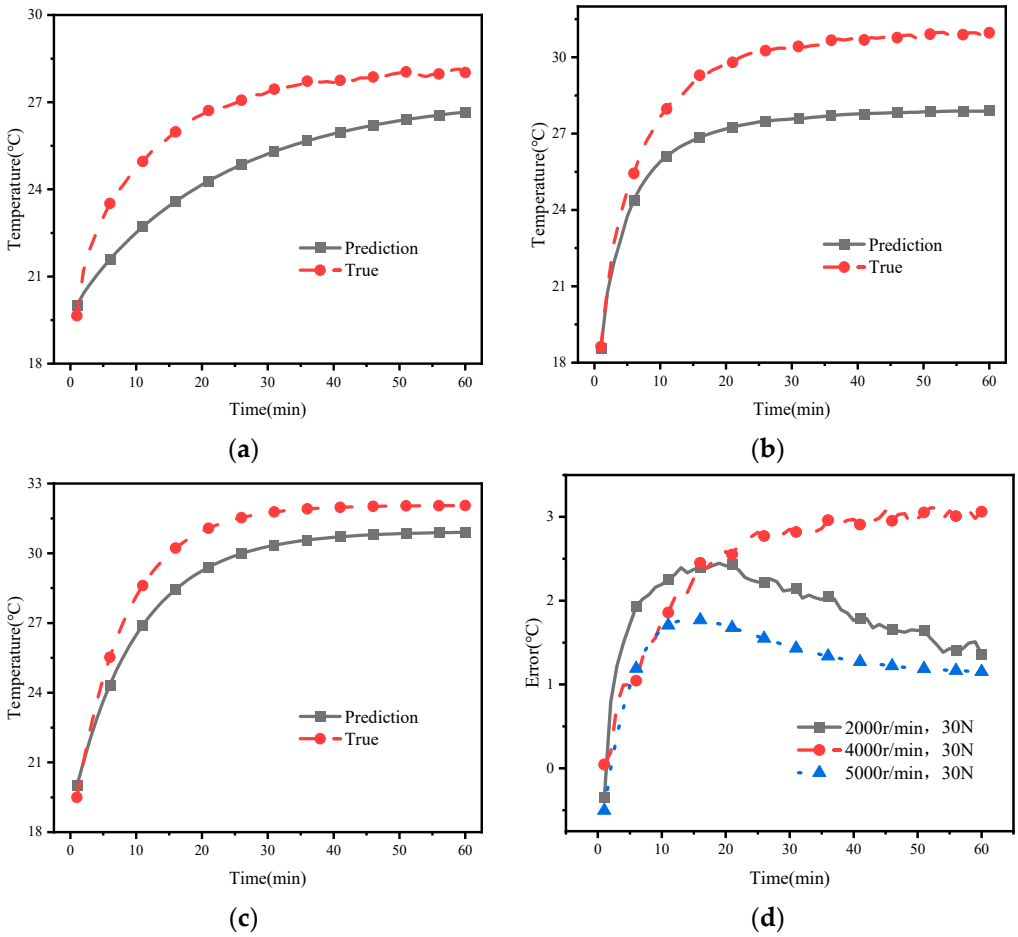


Figure 9. Comparison results of measured and predicted values of faulty bearings. (**a**) 2000 r/min, 30 N; (**b**) 4000 r/min, 30 N; (**c**) 5000 r/min, 30 N; (**d**) prediction error for different operating conditions.

4.2. CNN + Swin Transformer Model Fault Recognition Experiments

In order to ascertain the nature and severity of the occurring faults, the implementation of a CNN + Swin Transformer model was employed to discern various malfunctioning bearings. The dataset was partitioned into training, validation, and test sets in a ratio of 6:2:2. Tables 1 and 2 showcase a comparative analysis of the predictive efficacy pertaining to the identification of fault types in bearings, utilizing both the temperature rise data and the infrared temperature data during the stable phase.

Table 1. Results of two-phase projections.

Temperature Rise								**Steady-State**						
	HB	**SIRF**	**MIRF**	**LIRF**	**SORF**	**MORF**	**LORF**	**HB**	**SIRF**	**MIRF**	**LIRF**	**SORF**	**MORF**	**LORF**
HB	52	3	1	0	2	2	0	57	1	1	0	1	0	0
SIRF	3	53	2	2	0	2	1	1	59	0	0	0	0	0
MIRF	2	0	56	0	1	1	0	0	0	60	0	0	0	0
LIRF	0	4	0	54	2	0	6	0	1	0	59	0	0	0
SORF	2	0	0	1	55	2	0	1	0	0	0	58	0	1
MORF	0	0	3	0	0	57	0	0	0	1	0	0	59	0
LORF	0	0	0	0	1	0	59	0	0	0	0	0	1	59

Table 2. Prediction errors for both phases.

	Temperature Rise				Steady-State			
	Accuracy	**Precision**	**Recall**	**F1 Score**	**Accuracy**	**Precision**	**Recall**	**F1 Score**
HB	0.9650	0.8814	0.8667	0.8740	0.9881	0.9661	0.95	0.9580
SIRF	0.9604	0.8833	0.8413	0.8618	0.9929	0.9672	0.9833	0.9752
MIRF	0.9767	0.9032	0.9333	0.9180	0.9952	0.9677	1	0.9836
LIRF	0.9650	0.9474	0.8182	0.8781	0.9976	0.9833	0.9833	0.9916
SORF	0.9744	0.9016	0.9167	0.9091	0.9929	0.9831	0.9667	0.9748
MORF	0.9767	0.8906	0.95	0.9193	0.9952	0.9833	0.9833	0.9833
LORF	0.9814	0.8939	0.9833	0.9365	0.9952	0.9833	0.9833	0.9833

The table above provides a detailed overview of the predictive results using data from the warming-up stage and the steady-state stage. It can be observed from the results that the predictive accuracy of the warming-up stage data is 91.9%, while the predictive accuracy of the steady-state stage data reaches 97.8%. This unequivocally confirms that the data from the steady-state stage are more reliable and accurate. A comparison of the predictive results with models based on convolutional neural networks, Support Vector Machines, and the Swin Transformer model is illustrated in Figure 10.

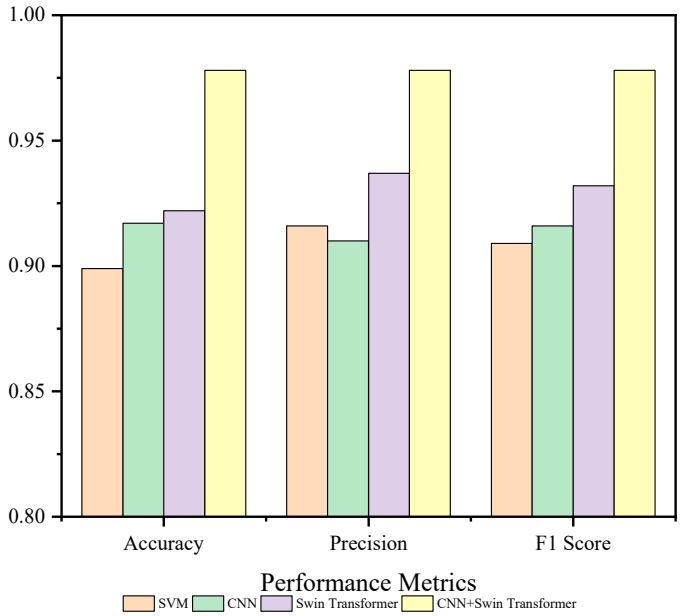


Figure 10. Comparison of classification results of different models.

4.3. Temperature Rise Prediction Combined with Image Recognition for Health Diagnosis

Table 1 provides a detailed overview of the predictive classification results for the fault-bearing temperature rise stage and the steady-state stage. It can be observed that healthy bearings are more prone to misclassification into other categories. This is particularly evident during the temperature rise stage, where it accounts for 23% of misclassifications, and during the steady-state stage, where it rises to 50%.

Therefore, a novel monitoring approach is proposed, which involves using the Informer model for bearing monitoring and integrating anomaly temperature rises for classification. By monitoring temperature changes, we can promptly detect anomalies in the bearings and conduct further classification and processing upon detection. The results of training and predicting classification using anomalous bearing data are presented in Table 3.

Table 3. Faulty bearing steady-state stage data detection results.

Bearing Condition						
	SIRF	**MIRF**	**LIRF**	**SORF**	**MORF**	**LORF**
SIRF	59	0	1	0	0	0
MIRF	1	59	0	0	0	0
LIRF	0	0	60	0	0	0
SORF	0	0	0	60	0	0
MORF	0	1	1	0	58	0
LORF	0	0	0	0	0	60

From Table 3, it can be observed that after filtering out faulty bearings through temperature prediction, the accuracy of fault type classification identification reaches 98.9%, which represents a 1% improvement compared to direct classification identification using infrared temperature field images.

To further evaluate this method, a comparison is made with models based on convolutional neural networks, Support Vector Machines, and the Swin Transformer model, as shown in Figure 11. It can be seen that when combined with temperature rise predictions on the same dataset, the CNN + Swin Transformer model outperforms the CNN, SVM, and Swin Transformer models in terms of accuracy, precision, and F1 score.

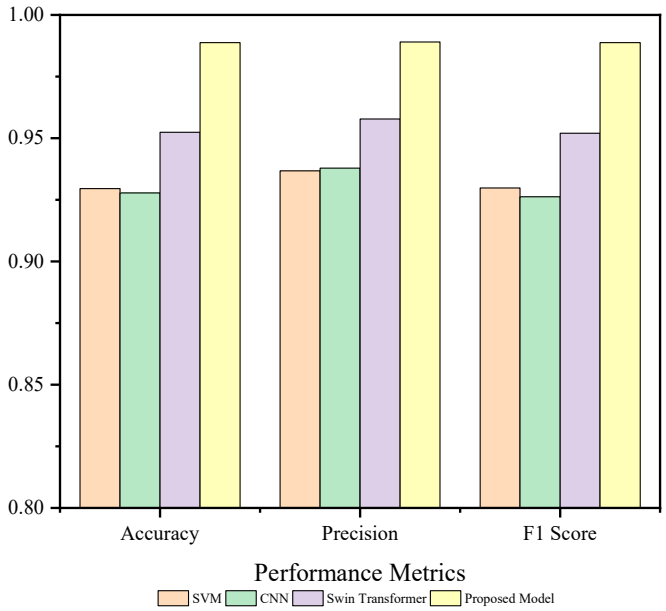


Figure 11. Comparison of classification results of different models.

5. Conclusions

The primary objective of this study is to address the challenge of timely fault detection in spindle bearings of machine tools operating in complex environments. By combining the Informer model with the CNN + Swin Transformer method, we have successfully achieved temperature rise prediction and classification of fault types and severity levels in bearings. The H7006C bearing was selected as the subject of investigation, and a dedicated test bench was constructed to capture infrared thermographic images under various operating conditions. This enabled the creation of a comprehensive dataset for bearing faults. Leveraging the Informer and CNN + Swin Transformer models, we developed a fault recognition model that incorporates temperature rise prediction. The experimental results demonstrate the following improvements:

1. After training the Informer model, it successfully predicts temperature variations in spindle bearings under different operating conditions. The model achieves lower pre-

diction errors during the stable state of healthy bearings compared to the temperature rise phase. When a bearing fault occurs, there is a noticeable difference between the predicted and measured temperatures, especially during the stable operating state. This accurately reflects changes in the health status of bearings.

2. The fault recognition model based on a CNN + Swin Transformer utilizes extracted feature maps and attention mechanisms for training and testing on the dataset. It achieves a recognition accuracy of 98.9% for identifying bearing fault states. Compared to individual models such as CNN, SVM, and Swin Transformer, the proposed method demonstrates superior recognition performance.
3. By combining Informer temperature rise prediction with CNN + Swin Transformer fault diagnosis and recognition, a dual diagnostic approach utilizing temperature differences and image feature differences is achieved. The recognition accuracy for bearing fault states reaches 98.9%. Compared to CNN, SVM, Swin Transformer, and other models, this combined approach accurately reflects the health status of bearings. It provides a solution for assessing the health status of spindle bearings in complex operating environments.

Author Contributions: Conceptualization, H.L. and C.L.; methodology, C.L.; software, H.L. and Z.H.; validation, W.Z. and F.Y.; formal analysis, X.S.; writing—original draft preparation, H.L.; writing—review and editing, W.Z., C.L. and N.G.; supervision, X.M. and X.S.; project administration, N.G. and X.M.; funding acquisition, F.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R & D Program of China (No. 2021YFB2011000), Major Science and Technology Projects of Longmen Laboratory (No. 231100220500).

Data Availability Statement: Data are contained within the article.

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

References

1. Peng, B.; Bi, Y.; Xue, B.; Zhang, M.; Wan, S. A survey on fault diagnosis of rolling bearings. *Algorithms* **2022**, *15*, 347. [CrossRef]
2. Nikula, R.P.; Karioja, K.; Pylvänäinen, M.; Leiviskä, K. Automation of low-speed bearing fault diagnosis based on autocorrelation of time domain features. *Mech. Syst. Signal Process.* **2020**, *138*, 106572. [CrossRef]
3. Aasi, A.; Tabatabaei, R.; Aasi, E.; Jafari, S.M. Experimental investigation on time-domain features in the diagnosis of rolling element bearings by acoustic emission. *J. Vib. Control* **2022**, *28*, 2585–2595. [CrossRef]
4. Fu, S.; Liu, K.; Xu, Y.; Liu, Y. Rolling bearing diagnosing method based on time domain analysis and adaptive fuzzy-means clustering. *Shock. Vib.* **2016**, *2016*, 9412787.
5. Zhao, D.; Cui, L.; Liu, D. Bearing weak fault feature extraction under time-varying speed conditions based on frequency matching demodulation transform. *IEEE/ASME Trans. Mechatron.* **2022**, *28*, 1627–1637. [CrossRef]
6. Gougam, F.; Rahmoune, C.; Benazzouz, D.; Afia, A.; Zair, M. Bearing faults classification under various operation modes using time domain features singular value decomposition fuzzy logic system. *Adv. Mech. Eng.* **2020**, *12*, 1687814020967874.
7. Zhao, D.; Wang, H.; Cui, L. Frequency-chirprate synchrosqueezing-based scaling chirplet transform for wind turbine nonstationary fault feature time–frequency representation. *Mech. Syst. Signal Process.* **2024**, *209*, 111112. [CrossRef]
8. Lu, Y.; Wang, Z.; Zhu, D.; Gao, Q.; Sun, D. Bearing fault diagnosis based on clustering and sparse representation in frequency domain. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–14. [CrossRef]
9. Jawad, S.M.; Jaber, A.A. Bearings health monitoring based on frequency-domain vibration signals analysis. *Eng. Technol. J.* **2022**, *41*, 86–95. [CrossRef]
10. Liang, M.; Zhou, K. A hierarchical deep learning framework for combined rolling bearing fault localization and identification with data fusion. *J. Vib. Control* **2023**, *29*, 3165–3174. [CrossRef]
11. Sun, Y.; Li, S. Bearing fault diagnosis based on optimal convolution neural network. *Measurement* **2022**, *190*, 110702. [CrossRef]
12. Chaleshtori, A.E.; Aghaie, A. A novel bearing fault diagnosis approach using the Gaussian mixture model and the weighted principal component analysis. *Reliab. Eng. Syst. Saf.* **2024**, *242*, 109720. [CrossRef]
13. Unal, M.; Onat, M.; Demetgul, M.; Kucuk, H. Fault diagnosis of rolling bearings using a genetic algorithm optimized neural network. *Measurement* **2014**, *58*, 187–196. [CrossRef]
14. Tang, H.; Tang, Y.; Su, Y.; Feng, W.; Wang, B.; Chen, P.; Zuo, D. Feature extraction of multi-sensors for early bearing fault diagnosis using deep learning based on minimum unscented kalman filter. *Eng. Appl. Artif. Intell.* **2024**, *127*, 107138. [CrossRef]
15. Zhang, J.; Sun, Y.; Guo, L.; Gao, H.; Hong, X.; Song, H. A new bearing fault diagnosis method based on modified convolutional neural networks. *Chin. J. Aeronaut.* **2020**, *33*, 439–447. [CrossRef]

16. Su, Y.; Shi, L.; Zhou, K.; Bai, G.; Wang, Z. Knowledge-informed deep networks for robust fault diagnosis of rolling bearings. *Reliab. Eng. Syst. Saf.* **2024**, *244*, 109863. [CrossRef]
17. Zhang, Y.; Zhou, T.; Huang, X.; Cao, L.; Zhou, Q. Fault diagnosis of rotating machinery based on recurrent neural networks. *Measurement* **2021**, *171*, 108774. [CrossRef]
18. An, Z.; Li, S.; Wang, J.; Jiang, X. A novel bearing intelligent fault diagnosis framework under time-varying working conditions using recurrent neural network. *ISA Trans.* **2020**, *100*, 155–170. [CrossRef]
19. Yu, Y.; Si, X.; Hu, C.; Zhang, J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput.* **2019**, *31*, 1235–1270. [CrossRef]
20. Yang, Z.; Cen, J.; Liu, X.; Xiong, J.; Chen, H. Research on bearing fault diagnosis method based on transformer neural network. *Meas. Sci. Technol.* **2022**, *33*, 085111. [CrossRef]
21. Alexakos, C.T.; Karnavas, Y.L.; Drakaki, M.; Tziafettas, I.A. A combined short time fourier transform and image classification transformer model for rolling element bearings fault diagnosis in electric motors. *Mach. Learn. Knowl. Extr.* **2021**, *3*, 228–242. [CrossRef]
22. Tang, X.; Xu, Z.; Wang, Z. A novel fault diagnosis method of rolling bearing based on integrated vision transformer model. *Sensors* **2022**, *22*, 3878. [CrossRef] [PubMed]
23. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual, 11–17 October 2021; pp. 10012–10022.
24. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 11106–11115.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.