


## Article

# Wind Turbine Predictive Fault Diagnostics Based on a Novel Long Short-Term Memory Model

Shuo Zhang , Emma Robinson and Malabika Basu \*

School of Electrical and Electronic Engineering, Technological University Dublin, D07 H6K8 Dublin, Ireland; d18128381@mytudublin.ie (S.Z.); emma.robinson@tudublin.ie (E.R.)

\* Correspondence: malabika.basu@tudublin.ie; Tel.: +353-851418201

**Abstract:** The operation and maintenance (O&M) issues of offshore wind turbines (WTs) are more challenging because of the harsh operational environment and hard accessibility. As sudden component failures within WTs bring about durable downtimes and significant revenue losses, condition monitoring and predictive fault diagnostic approaches must be developed to detect faults before they occur, thus preventing durable downtimes and costly unplanned maintenance. Based primarily on supervisory control and data acquisition (SCADA) data, thirty-three weighty features from operational data are extracted, and eight specific faults are categorised for fault predictions from status information. By providing a model-agnostic vector representation for time, Time2Vec (T2V), into Long Short-Term Memory (LSTM), this paper develops a novel deep-learning neural network model, T2V-LSTM, conducting multi-level fault predictions. The classification steps allow fault diagnosis from 10 to 210 min prior to faults. The results show that T2V-LSTM can successfully predict over 84.97% of faults and outperform LSTM and other counterparts in both overall and individual fault predictions due to its topmost recall scores in most multistep-ahead cases performed. Thus, the proposed T2V-LSTM can correctly diagnose more faults and upgrade the predictive performances based on vanilla LSTM in terms of accuracy, recall scores, and F-scores.

**Keywords:** operation and maintenance (O&M); wind turbines (WTs); predictive fault diagnostic; supervisory control and data acquisition (SCADA); Time2Vec (T2V); Long Short-Term Memory (LSTM); T2V-LSTM



**Citation:** Zhang, S.; Robinson, E.; Basu, M. Wind Turbine Predictive Fault Diagnostics Based on a Novel Long Short-Term Memory Model. *Algorithms* **2023**, *16*, 546. <https://doi.org/10.3390/a16120546>

Academic Editors: Ying Bi, Mengjie Zhang, Bing Xue and Bo Peng

Received: 21 October 2023  
Revised: 23 November 2023  
Accepted: 25 November 2023  
Published: 28 November 2023



**Copyright:** © 2023 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 global wind energy installations expanded by about 14% annually from 2001 to 2020 [1]. The total wind power capacity increased from 650.8 GW in 2019 to 742.7 GW in 2020, with a spectacular growth of 53% (over 90 GW) since 2019 [2,3]. Due to plentiful wind resources and abundant construction sites in offshore areas, more wind farms are installed with increased seabed depths and remote distances to shore [4]. Using the same commercial wind turbine (WT), offshore power production is at least 1.34 times more than the onshore site with the highest wind energy potential due to stronger and more uniform wind resources in offshore areas [5]. However, offshore WT installation costs are about 2.64 times those of their onshore counterpart [6]. And harsher weather conditions are challenging for the operation and maintenance (O&M) tasks of offshore WTs. Moreover, O&M costs account for a large fraction of total lifecycle costs, with 10–15% and 25–30% for onshore and offshore wind farms, respectively [6,7]. Unexpectedly, sudden faults from high-risk WT components contribute significantly to the increase in O&M costs related to downtimes and discounted revenues [8]. To reduce O&M costs and enhance system reliability, condition monitoring (CM), fault diagnosis, and prognosis are of prior importance through the detection of certain faults before they reach catastrophic fault severity levels. Hence, O&M costs can be decreased along with maintenance interval optimisation [9].

Condition Monitoring System (CMS) can facilitate system failure prevention and WT availability improvement through early-stage fault detections. To diagnose fault-free conditions of WT components, such as the gearbox and drivetrain, CMS has been implemented via vibration analysis [10], oil analysis [11], electrical signature analysis [12], and acoustic emission analysis [13]. CMS-based monitoring is capable of both fault diagnosis and prognosis with a high-frequency resolution, but this approach is more expensive compared to supervisory control and data acquisition (SCADA) [14]. Thus, SCADA systems become more favourable for WT operators to apply the CM technique due to cheaper costs; however, these have a low-frequency resolution [15]. SCADA data are normally collected under a 10 min sampling rate. A number of data-driven studies on SCADA-based monitoring have been utilised for performance monitoring of WT operational conditions in recent years without retrofitting additional sensors.

Stetco et al. [16] investigated the machine learning (ML) applications for CM in WTs, including CM for diagnosis and CM for prognosis. Diagnosis focuses on real-time fault identifications, whilst prognosis is to predict the faults before their occurrences [17].

Classification is a supervised ML approach, applicable for fault detection, diagnosis, and prognosis, to train a classifier by predicting its categorised outputs based on input variables, thus differentiating between healthy and faulty operations. Lu et al. [18] proposed an online fault diagnosis for WT planetary gearbox faults employing a self-powered wireless sensor for signal acquisition. Leahy et al. [19] applied support vector machine (SVM) models to detect, diagnose, and predict faults in a 3 MW direct-drive turbine. However, the classification results on feeding and air-cooling faults had deficient performances due to the problematic classification of the SVM hyperplane. Naik and Koley [20] adopted the  $k$ -nearest neighbour ( $k$ -NN) classifier-based protection to detect and classify multiple types of faults in AC/HVDC transmission systems by varying fault resistance and inception angles with a classification accuracy of 100%. Marti-Puig et al. [21] investigated several automatic feature selection approaches based on the  $k$ -NN classifier for fault prognostics with the use of 36 sensor variables on gearbox and transmission systems. Artificial Neural Network (ANN) was trained by Ibrahim et al. [22] for WT mechanical faults with a median accuracy between 93.5% and 98% in fault detection. For various classification tasks, SVM [18,19,23],  $k$ -NN [20,21], ANN [22,24,25], and RF [25,26] are commonly used with SCADA, simulation, or experimental data. Most importantly, accurate fault diagnosis is the prerequisite for developing any prediction model.

ANN has been widely applied to the ML approach for supervised classification learning [27]. The typical ANN architecture, multi-layer perception (MLP), is a feed-forward multi-layered neural network consisting of an input layer, several hidden layers, and an output layer. The ANN prediction results are determined by data size, data pre-processing, selected neural network structures under their optimum activation functions, etc. [16]. Due to its robustness towards poor-quality data with noise and system disturbances, a well-trained ANN model can still make wise predictions, which cannot be achieved by other ML classifiers [6]. With the escalation of quantitative data sizes and complexity, ANN is a model with ideal predictive results but a slow convergence speed.

Compared to ANN, the Recurrent Neural Network (RNN) is a more promising neural network model for time- and sequence-based tasks because its recurrent structure captures the temporal dependency among inputs with sequential characteristics to predict the next scenario [28]. RNN is a class of deep-learning neural networks designed for variable-length sequence inputs by remembering important events and allowing the previous values as inputs to predict future outputs with recurrent connections in hidden layers. RNN overcomes the over- and under-fitting issues and reduces the convergence time compared to ANN. However, vanishing gradients, caused by error information flowing backward, are large barriers to the success of vanilla RNNs because of the resultant oscillating weights or loss of long-term dependencies [29]. To address vanishing gradients, Long-Short-Term Memory (LSTM), proposed by Hochreiter and Schmidhuber [29], is a remarkable RNN model to control the information flow with additional interacting layers. Based on SCADA

data, Chen et al. [30] verified the outperformance of LSTM over ANN and autoencoder (AE) for anomaly detection. The integrated LSTM-AE model further improved the detection accuracy due to the raw input processed by AE and the time feature managed by LSTM. Based on both single-sensor and multi-sensor signals, LSTM outperformed RNN and ANN on the classification of 11 faults on the wind wheel, bearing, bearing support, and rotor [31]. For a case study of fault classification on inner, outer, and ball faults from rolling bearings [32], LSTM demonstrated higher accuracy than ANN and SVM, and stacked LSTM further enhanced the prediction accuracy. The advantages of LSTM have been validated according to multiple time-series fault diagnosis tasks [30–32].

The time-series events can occur either synchronously or asynchronously. However, most of the RNN or LSTM models fail to make use of time as a feature by considering all inputs to be synchronous. Kazemi et al. [33] proposed a model-agnostic vector representation for time, known as Time2Vec (T2V), to be integrated with the LSTM model to refurbish the architecture with the consumption of time features. The key contributions in this paper can be generalised as follows:

- A feature selection method, Recursive Feature Elimination (RFE) [34], is conducted along with an RF classifier for WT fault prediction. The weights of each feature are computed under the RF classifier, and the RFE application reserves the optimal number of features in order of their significance levels to maintain a balance between prediction accuracy and computational costs.
- By integrating Time2Vec into LSTM, this approach, T2V-LSTM, has been validated to outperform LSTM with a stationary Time2Vec activation function based on several synchronous datasets [33]. In this paper, the data points related to downtimes are removed to reserve only fault-free and fault data provided by the SCADA system for the purpose of fault and no-fault predictions. Thus, a non-stationary Time2Vec activation function is demanded to deal with the yielded asynchronous data.
- A novel deep-learning neural network model, T2V-LSTM, with an optimal non-stationary activation function, is modelled to improve the model performance of LSTM, successfully detecting over 84.97% of faults in advance. The comparative studies between T2V-LSTM, LSTM, and other ML classifiers are investigated for overall and individual fault predictions based on performance metrics, including accuracy, recall scores, precision scores, and F-scores [16].

The paper is organised as follows. Section 2 provides the SCADA operational and status data, and the modelling process is introduced with data pre-processing, feature engineering, and fault prognosis. The methodology studies of T2V-LSTM and the processes of model optimisation are presented in Section 3. Section 4 investigates the comparative predictive results from T2V-LSTM, LSTM, and other classifiers, and Section 5 presents a discussion of this investigation. The key results and contributions are summarised in Section 6.

## 2. SCADA Data

The available data were collected from a 7 MW demonstration offshore WT, owned by the Offshore Renewable Energy (ORE) Catapult [35]. This WT is a three-bladed upwind turbine mounted on a jacket support structure with a total height of 196 m, from blade tip to sea level, located at Levenmouth, Fife, Scotland, UK. The regarded cut-in, rated, and cut-off wind speeds were 3.5 m/s, 10.9 m/s, and 25 m/s, respectively [36]. More detailed information about this WT can be seen in Figure 1. For this turbine, the collected data had two separate groups: operational SCADA data and status data. The investigated datasets of both groups cover a 17-month period from May 2018 to September 2019.

### 2.1. Operational Data and Status Data

The collected SCADA operational data include alarm data, control information, electrical signals, pressure data, temperature data, turbine data, miscellaneous signals, and other signals.

Properties	Values
Wind class	IEC Class 1A
Rotor diameter	171.2 m
Capacity	7 MW
Hub height	110.6 m
Blade length	83.5 m
Total height	196 m blade tip to sea level
Generator	PMG (3.3 kV)
Converter	Full power conversion
Drive train	Medium speed (400 rpm)
Rated frequency	50 Hz
Rotor speed	5.9 ~ 10.6 rpm
Wind speed	3.5 ~ 25 m/s
Rated wind speed	10.9 m/s
Design life	25 years
Certification	DNV

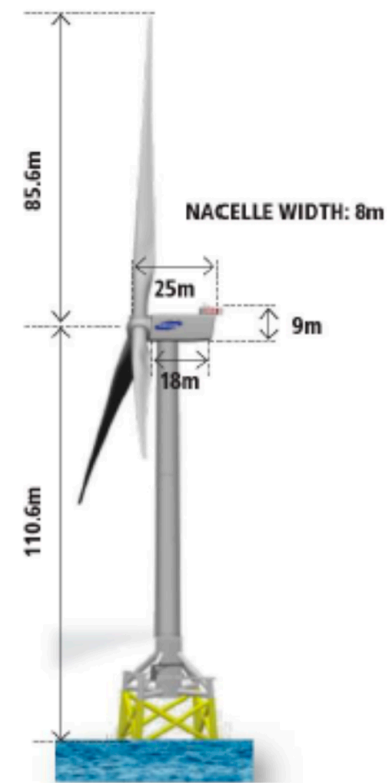


Figure 1. Levenmouth offshore wind turbine [35].

The SCADA system operates at a 10 min sampling rate by monitoring instantaneous parameters, such as wind speed, pitch angle, rotor speed, yaw error, electrical power, currents, voltages, temperatures, and pressures. Taking Table 1 as an example, the minimum, maximum, mean, standard deviation, and ending values of wind speed are collected with the corresponding timestamps. The original dataset includes more than 2000 features and approximately 70,000 data points with regard to the 17 months to be studied.

Table 1. Ten-minute SCADA operational data.

StartTime	WindSpeed _mps_Min	WindSpeed _mps_Max	WindSpeed _mps_Mean	WindSpeed _mps_Stdev	WindSpeed _mps_EndVal
21/05/2018 22:00:00	3.577394	10.11077	6.8690084	1.3447459	5.802108
21/05/2018 22:10:00	3.062414	10.03982	6.7177955	1.108204	6.073331
21/05/2018 22:20:00	4.69204	9.636992	7.1981784	1.0209401	8.475427

The information about requested shutdowns, faults, or warning events is provided by status data. As seen in Table 2, fault and warning events are tracked with respect to their corresponding event codes, on-times, and off-times. There are miscellaneous operating states under the abnormal or faulty conditions of the WT. According to Kusiak and Li [37], the status of fault data is assigned as follows:

$$\text{If } T_{SCADA}(t) < T_{fault} < T_{SCADA}(t+1), \text{ then} \\ \text{Event.Code}(t) = \text{Event.Code}(t+1) = \text{Event.Code}(T_{fault}) \quad (1)$$

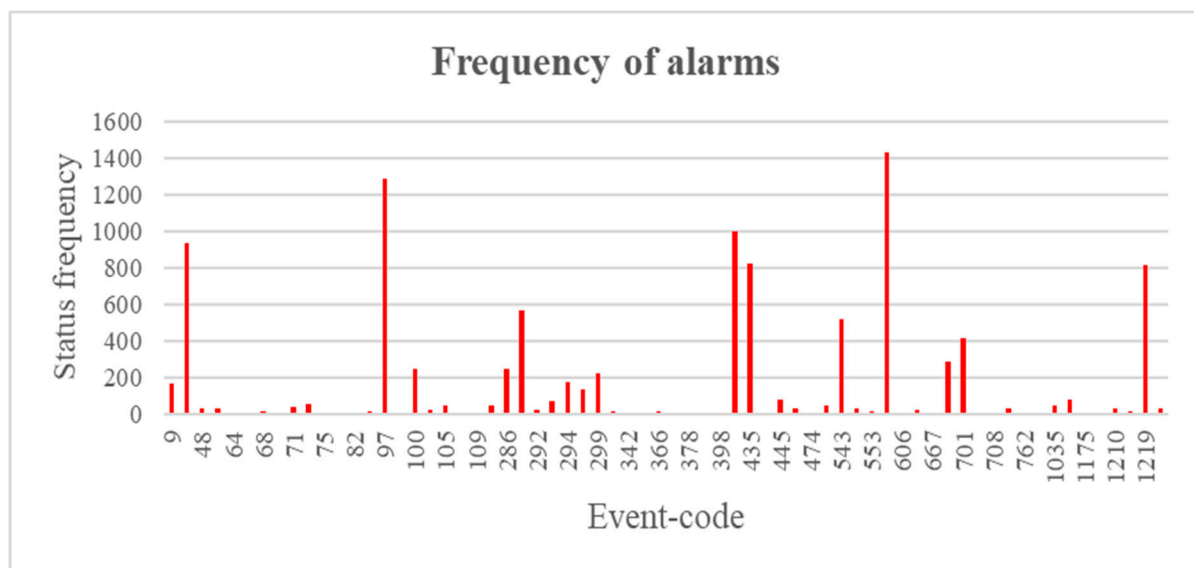
where  $T_{SCADA}(t+1)$  denotes the one-step behind (or 10 min behind) SCADA data since both timestamps,  $t$ , and  $t+1$ , have 10 min intervals.

**Table 2.** SCADA status data.

TimeOn	TimeOff	Event Code	Event Description
21/05/2018 19:38:33	21/05/2018 19:38:39	286	(Demoted) Yaw Hydraulic Pressure Diff Too Large
21/05/2018 20:26:02	21/05/2018 20:26:10	543	Gearbox Cooling Line Pressure Too Low
21/05/2018 20:26:12	21/05/2018 20:29:20	543	Gearbox Cooling Line Pressure Too Low

The 10 min period is applied to capture any fault occurrences. For example, in Table 2, the operational period from “21/05/2018 20:20:00” to “21/05/2018 20:30:00” should be labelled as “Gearbox Cooling Line Pressure Too Low” with its event code of 543 due to its fault occurrence within the 10 min time band.

As seen in Figure 2, the frequency of occurrences of each status varies. Any event code above zero indicates an abnormality. The majority are fault event codes only occurring a few times, but the faults under the event codes 399, 435, 570, and 1219 have occurred more than 800 times within this period. Some event codes, such as 12 and 97, denoting SCADA shutdown request and yaw error, respectively, are not associated with a defined fault status despite their appearances being 932 and 1290, respectively. Aside from these two examples, the majority of event codes are merely warnings, irrelevant to faults, so many event codes are of minor interest in this paper. Additionally, the event codes relating to downtime due to maintenance actions, noise curtailments, and requested owner stops in Figure 2 are to be excluded.

**Figure 2.** Frequency of alarms.

By excluding the data related to downtimes and faults with very limited frequency, only a small number of faults can be reserved according to their relatively frequent occurrences, as seen in Table 3. “HPU 2 Pump Active For Too Long” relates to a fault that occurred in the hydraulic pump unit (HPU), while “PcsOff” and “PcsTrip” relate to shut-off faults and circuit trips within the power conditioning system (PCS), respectively. The deep-learning model must train the classifiers for the specific fault instances defined in Table 3. Hence, the reserved SCADA data can be classified into nine categories: (1) fault-free; (2) HPU 2 pump active fault; (3) Blade 3 slow response; (4) pitch system fatal fault; (5) gearbox cooling pressure fault; (6) (Demoted) gearbox pressure 2 fault; (7) PcsOff fault; (8) PcsTrip fault; (9) sub-pitch fatal fault. The quantity of fault-free cases is much larger than that of any individual fault case, leading to an imbalance in the investigated dataset.



**Table 3.** Fault distributions.

Event Code	Frequency	Description
0	42,598	Fault-free
290	565	HPU 2 pump active for too long
399	998	Blade 3 too slow to respond
435	826	Pitch system fatal error
543	522	Gearbox cooling line pressure too low
570	1436	(Demoted) gearbox filter manifold pressure 2 shutdown
700	701	PcsOff * <sup>1</sup>
701	417	PcsTrip * <sup>2</sup>
1219	816	Sub-pitch priv fatal error has occurred more than 3 times

\*<sup>1</sup> PcsOff represents the shut-off faults of power conditioning system. \*<sup>2</sup> PcsTrip represents the circuit trips within power conditioning system.

## 2.2. Feature Engineering

The major occurrences of the faults are on the HPU, blade, pitch system, gearbox, PCS, and sub-pitch system (see Table 3), and a huge number of original features are Count-False/CountTrue states, apparently irrelevant to those faults. This leads to the principal selection of 60 relevant features, which are only a small subsection of the original 2000 features. Among the 60 features, the deviations of pitch angles, as well as the deviations of sub-pitch positions from blades 1 and 2, 2 and 3, and 3 and 1 are considered because of possible blade angle asymmetry or blade angle implausibility, studied by Kusiak and Verma [38].

Feature engineering aims to reduce dimensionality by eliminating features with lower significance and improving the computational efficiency of deep-learning neural networks. RFE [34] has been commonly applied to fit the model by recursively removing irrelevant or redundant features.

Firstly, an estimator for accurate online fault diagnosis is required to cooperate with RFE for dimensionality reduction. Apart from detecting the abnormality, fault diagnosis can determine the specific fault types with an advanced multi-level fault classification. The accuracy is used to evaluate the performance of classifiers by:

$$Accuracy = \frac{1}{k} \sum_{i=1}^k \frac{TP_i + TN_i}{TP_i + FP_i + FN_i + TN_i} \quad (2)$$

where  $k$  is the total number of classes,  $TP_i$  donates true positive in class  $i$ , when both prediction and actuality are faulty, whilst  $TN_i$  donates true negative in class  $i$ , when both prediction and actuality are fault-free.  $FP_i$  signifies false positive in class  $i$ , when the actual fault-free condition is wrongly predicted to be faulty, whilst  $FN_i$  signifies false negative in class  $i$ , when the actual faulty condition is wrongly predicted to be fault-free.

ML approaches, such as Decision Tree (DT),  $k$ -NN, SVM, RF, ANN, and Gradient Boost (GB), are compared for fault diagnosis. As seen in Figure 3, RF is evidently the finest model among all in terms of its best accuracy (0.98607386). Hence, the RF classifier is chosen to conduct RFE using a 10-fold cross-validation for the test set. Based on the RFE process in Figure 4, the best accuracy (0.9888) is observed by selecting 33 optimal features out of 60 for the fault diagnosis task under RF classification.

As seen in Figure 5, the weighted importance of each feature is depicted under the RF classification. According to the optimal solution given in Figure 4, the 33 top-ranking features in Figure 5 are reserved for predictive fault diagnosis models. The features, such as “AverageMeasuredPtchAngle1\_Max”, “GBoxFilterPres2\_Mean”, etc., have advanced significance levels. However, the features with lower significance levels than “ManualPtch-StateCounter\_EndVal” are excluded in predictive fault diagnosis studies for the purpose of dimensionality reduction.

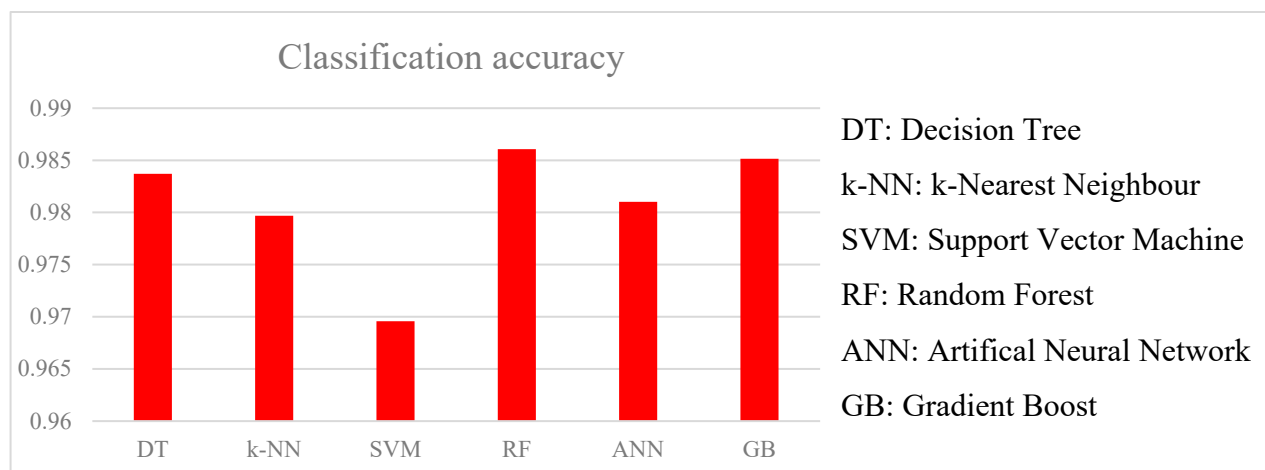


Figure 3. Comparison of classification accuracy.

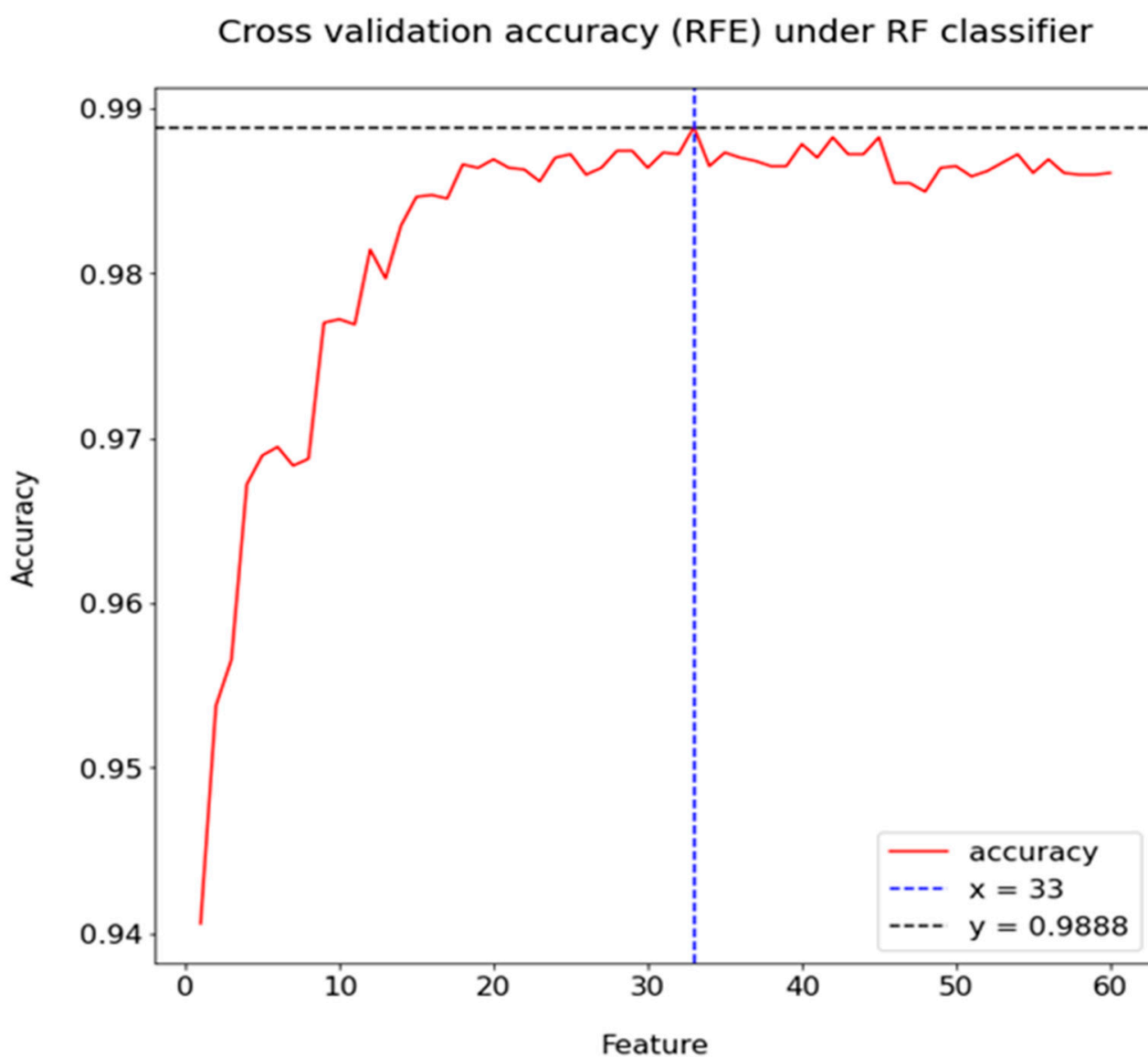


Figure 4. Cross-validation scores plotted against the number of features.

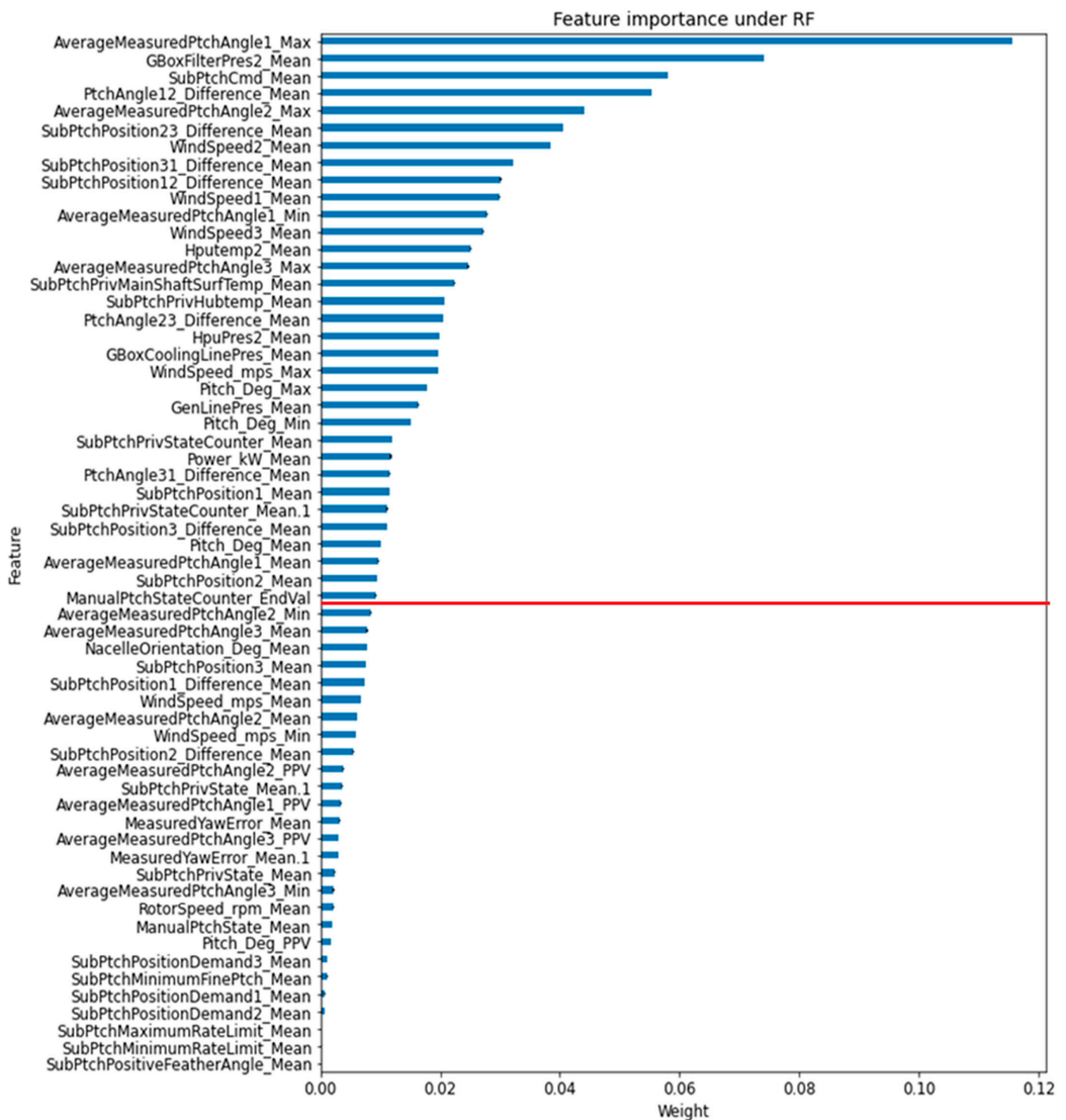


Figure 5. Feature importance under RF classifier (the 33 features above the red line are reserved).

### 2.3. Predictive Fault Diagnosis

Fault diagnosis aims at accurately identifying the fault types within a WT in a real-time application. However, it is insufficient to prevent damage caused by some severe failures only through online fault diagnosis. Then, fault prognosis is recommended by providing the predictive fault diagnosis prior to the fault occurrence, which decreases the maintenance fees and extends machinery life.

The observed dataset for online fault diagnosis is expressed by  $\{X_t, Y_t\}$ , where  $X_t$  and  $Y_t$  are the given input data and the resultant diagnosed fault class. For example,



as the SCADA data are collected at 10 min intervals, the 10 min, 20 min, and 30 min ahead fault predictions can be achieved with the modified datasets,  $\{X_{t-1}, Y_t\}$ ,  $\{X_{t-2}, Y_t\}$ , and  $\{X_{t-3}, Y_t\}$ , respectively, based on the original dataset,  $\{X_t, Y_t\}$ . Thus, the predictive performances under  $n$ -step in advance will be determined using the modified dataset,  $\{X_{t-n}, Y_t\}$ .

### 3. Methods

#### 3.1. LSTM

As vanilla RNNs, with only input gates and output gates, suffer from vanishing or exploding gradients caused by error back-flow problems, the main challenge for vanilla RNNs is to handle the long-term dependencies.

To secure the long-term dependencies, LSTM additionally inserted forget gates for the update and control of cell states, regulating the information flow [29]. LSTM can handle the imbalanced data and efficiently captures a sequence of time-lagged observations as inputs for time-series classification to predict specific faults at any given time ahead. The original LSTM model can be precisely stated as follows:

$$f_j = \sigma(W_f x_j + U_f h_{j-1} + b_f) \quad (3)$$

$$i_j = \sigma(W_i x_j + U_i h_{j-1} + b_i) \quad (4)$$

$$O_j = \sigma(W_o x_j + U_o h_{j-1} + b_o) \quad (5)$$

$$\bar{C}_j = \sigma_c(W_c x_j + U_c h_{j-1} + b_c) \quad (6)$$

$$C_j = f_j \odot C_{j-1} + i_j \odot \bar{C}_j \quad (7)$$

$$h_j = \sigma_h(C_j) \odot O_j \quad (8)$$

Herein,  $x_j$  is the neuron input at the timestamp  $j$ ;  $h_{j-1}$  is the cell state at the previous timestamp;  $j-1$ ,  $f_j$ ,  $i_j$ , and  $O_j$  stand for the forget, input, and output gates, respectively, all determined across the sigmoid nonlinearity,  $\sigma$ , with the given weights  $W_f$ ,  $W_i$ ,  $W_o$ ,  $W_c$ ,  $U_f$ ,  $U_i$ ,  $U_o$ ,  $U_c$ , and the assigned biases,  $b_f$ ,  $b_i$ ,  $b_o$ ,  $b_c$ . The memory cell,  $\bar{C}_j$ , from Equation (6) is estimated through an activation function,  $\sigma_c$ , which is a hyperbolic tangent layer,  $Tanh$ , by default. Then, the current cell state  $C_j$  in Equation (7) is updated regarding the previous cell state,  $C_{j-1}$ , and the estimated cell state,  $\bar{C}_j$ , with the element-wise product operator,  $\odot$ . Finally, the output vector,  $h_j$ , in Equation (8), also known as a hidden layer, is obtained from the element-wise product of the output gate,  $O_j$ , and the cell state,  $C_j$ , across an activation function,  $\sigma_h$ , which is also  $Tanh$  by default.

Based on the dependencies in LSTM, the forget gate,  $f_j$ , controls the fraction of  $C_{j-1}$ , to store in  $C_j$ , filtering  $h_{j-1}$  and  $x_j$  through the sigmoid gate,  $\sigma$ . The input gate,  $i_j$ , controls the fraction of the estimated memory cell,  $\bar{C}_j$ , provided to  $C_j$  through the sigmoid nonlinearity,  $\sigma$ ; the output gate,  $O_j$ , controls the fraction of  $C_j$  flowing into the output vector,  $h_j$ , through  $\sigma_h$ . Therefore, the LSTM architecture can be drawn in Figure 6.

#### 3.2. Time-LSTM-1

Regarding Equation (7),  $C_{j-1}$  covers the information at the previous timestamp, reflecting the long-term interest, and  $x_j$  is the last consumed item, hardly reflecting on current recommendations. Then, Time-LSTM, proposed by Zhu et al. [39], equips LSTM with

time gates to store time intervals in  $C_j, C_{j+1}, \dots$ , controlling the fraction of  $x_j$  on current recommendations. Time-LSTM-1 [39] only adds one time gate,  $T_j$ , to LSTM by:

$$T_j = \sigma(W_t x_j + \sigma(U_t \Delta t_j) + b_t) \quad (9)$$

where  $\Delta t_j$  is defined as the time interval for the  $j$ th event by  $\Delta t_j = t_{j+1} - t_j$ , implemented across a sigmoid function,  $\sigma$ , and  $T_j$  is also determined through  $\sigma$  with the assigned weights,  $W_t$  and  $U_t$ , and the given bias,  $b_t$ .  $\Delta t_j$  can also be recognised as the duration between the current and the last event. Based on the basic LSTM architecture from Equations (3)–(8), Equations (7) and (5) can be revised to:

$$C_j = f_j \odot C_{j-1} + i_j \odot T_j \odot \bar{C}_j \quad (10)$$

$$O_j = \sigma(W_o x_j + U_o h_{j-1} + V_o \Delta t_j + b_o) \quad (11)$$

where  $V_o$  is the added weight to calculate  $O_j$  in Time-LSTM-1. Then, both the input gate,  $i_j$ , and the time gate,  $T_j$ , control the fraction of the estimated memory cell,  $\bar{C}_j$ , provided to the current memory cell,  $C_j$ , in Equation (10). As  $T_j$ , containing the information of interval,  $\Delta t_j$ , is provided to  $C_j$ , and then transferred to  $C_{j+1}, C_{j+2}, \dots$ , the time gate,  $T_j$ , benefits the long-term interests,  $C_j, C_{j+1}, \dots$ , of the LSTM model by storing  $\Delta t_j$ .

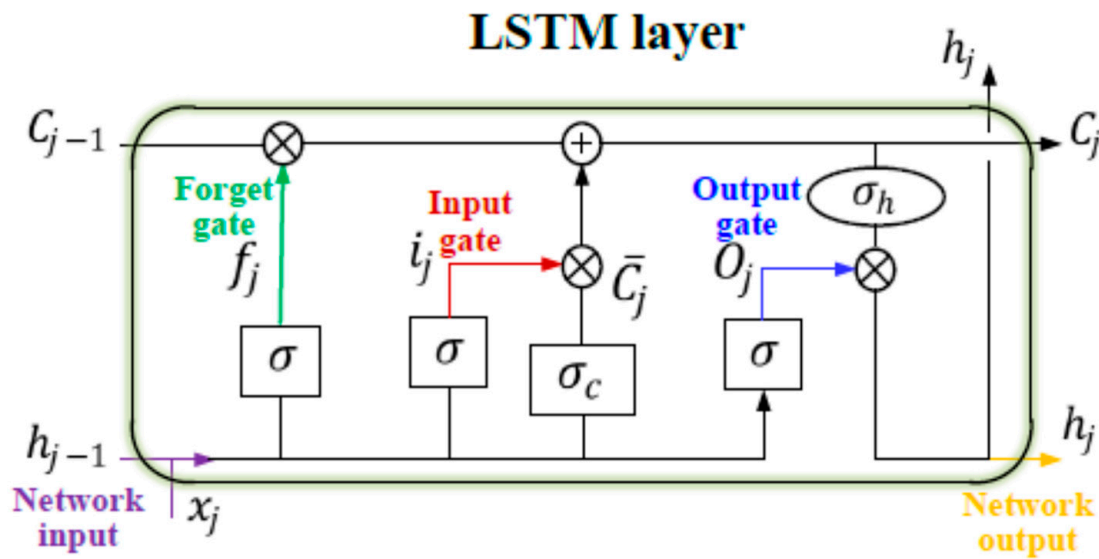


Figure 6. LSTM architecture.

### 3.3. Time-LSTM-3 and T2V-LSTM Models

For Time-LSTM-1, the single time gate  $T_j$  is mainly controlled by  $\Delta t_j$ . Zhu et al. [39] developed two alternative Time-LSTM models, Time-LSTM-2 and Time-LSTM-3, both containing double time gates,  $T1_j$  and  $T2_j$ .  $T1_j$  controls the influence of the last consumed item,  $x_j$ , on current recommendations, while  $T2_j$  stores  $\Delta t_j$  for later recommendations. Based on  $T_j$  in Equation (9),  $T1_j$  and  $T2_j$  can be expressed by:

$$T1_j = \sigma(W_{t1} x_j + \sigma(U_{t1} \Delta t_j) + b_{t1}) \quad (12)$$

$$T2_j = \sigma(W_{t2} x_j + \sigma(U_{t2} \Delta t_j) + b_{t2}) \quad (13)$$

where  $W_{t1}$ ,  $W_{t2}$ ,  $U_{t1}$ , and  $U_{t2}$  are given weights and  $b_{t1}$  and  $b_{t2}$  are given biases. Among three LSTM models with time gates, Time-LSTM-3 is validated with the best predictions by

coupling input and forget gates, inspired by the LSTM variant from Greff et al. [40] and the cell state,  $C_j$ , in Equation (7) under Time-LSTM-3 can be modified by:

$$C_j = (1 - i_j) \odot C_{j-1} + i_j \odot \bar{C}_j \quad (14)$$

Hence, Time-LSTM-3 has a shorter processing time than Time-LSTM-2 due to its simpler architecture and fewer parameters to calculate. By removing the forget gate, Equation (14) can be modified by:

$$\tilde{C}_j = (1 - i_j \odot T1_j) \odot C_{j-1} + i_j \odot T1_j \odot \bar{C}_j \quad (15)$$

$$C_j = (1 - i_j) \odot C_{j-1} + i_j \odot T2_j \odot \bar{C}_j \quad (16)$$

where  $\tilde{C}_j$  is a new cell state to store the result. The output gate,  $O_j$ , in Equation (5) and the output vector,  $h_j$ , in Equation (8) can be replaced by:

$$O_j = \sigma(W_o x_j + U_o h_{j-1} + V_o \Delta t_j + b_o) \quad (17)$$

$$h_j = \sigma_h(\tilde{C}_j) \odot O_j \quad (18)$$

Here, both  $i_j$  and  $T1_j$  are filters for  $\bar{C}_j$ , while  $T2_j$  stores  $\Delta t_j$ , transferred to  $C_j, C_{j+1}, C_{j+2}, \dots$ , for modelling the long-term interests for later recommendations.  $\tilde{C}_j$  is implemented through an activation function,  $\sigma_h$ , influencing the current recommendations.

A model-agnostic vector representation for time, known as Time2Vec, is used to rebuild the architectures of Time-LSTM with the consumption of time features under either stationary or non-stationary activation functions. For this reason, Time2Vec replaces the time interval,  $\Delta t_j$ , by a model-agnostic vector,  $T2V(\Delta t_j)$ , as follows:

$$T2V(\Delta t_j)[i] = \begin{cases} \omega_i \cdot \Delta t_j + \varphi_i, & \text{if } i = 0 \\ \mathcal{F}(\omega_i \cdot \Delta t_j + \varphi_i), & \text{if } 1 \leq i \leq k \end{cases} \quad (19)$$

where  $T2V(\Delta t_j)[i]$  is the  $i$ th element of  $T2V(\Delta t_j)$ ,  $\mathcal{F}$  can be any stationary or non-stationary activation functions, such as *Sine* and *Tanh*, and  $\omega_i$  and  $\varphi_i$  are learnable parameters. Then, in a T2V-LSTM model, all time vectors,  $\Delta t_j$ , should be replaced by Time2Vec elements,  $T2V(\Delta t_j)$ , so the time gates,  $T1_j$  and  $T2_j$ , in Equations (12) and (13), respectively, and the output gate,  $O_j$ , in Equation (17) are modified as follows:

$$T1_j = \sigma(W_{t1} x_j + \sigma(U_{t1} \cdot T2V(\Delta t_j)) + b_{t1}) \quad (20)$$

$$T2_j = \sigma(W_{t2} x_j + \sigma(U_{t2} \cdot T2V(\Delta t_j)) + b_{t2}) \quad (21)$$

$$O_j = \sigma(W_o x_j + U_o h_{j-1} + V_o \cdot T2V(\Delta t_j) + b_o) \quad (22)$$

For T2V-LSTM, the output vector,  $h_j$ , is still controlled according to Equation (18). Therefore, based on Equations (15)–(22), the architecture of Time-LSTM-3 or T2V-LSTM can be plotted in Figure 7. Time2Vec, determined by its selected activation function, has three major advantages: being capable of learning both periodic and non-periodic activation functions, having invariance to time rescaling, and being simple to combine a representation for time with multiple neural networks [33].

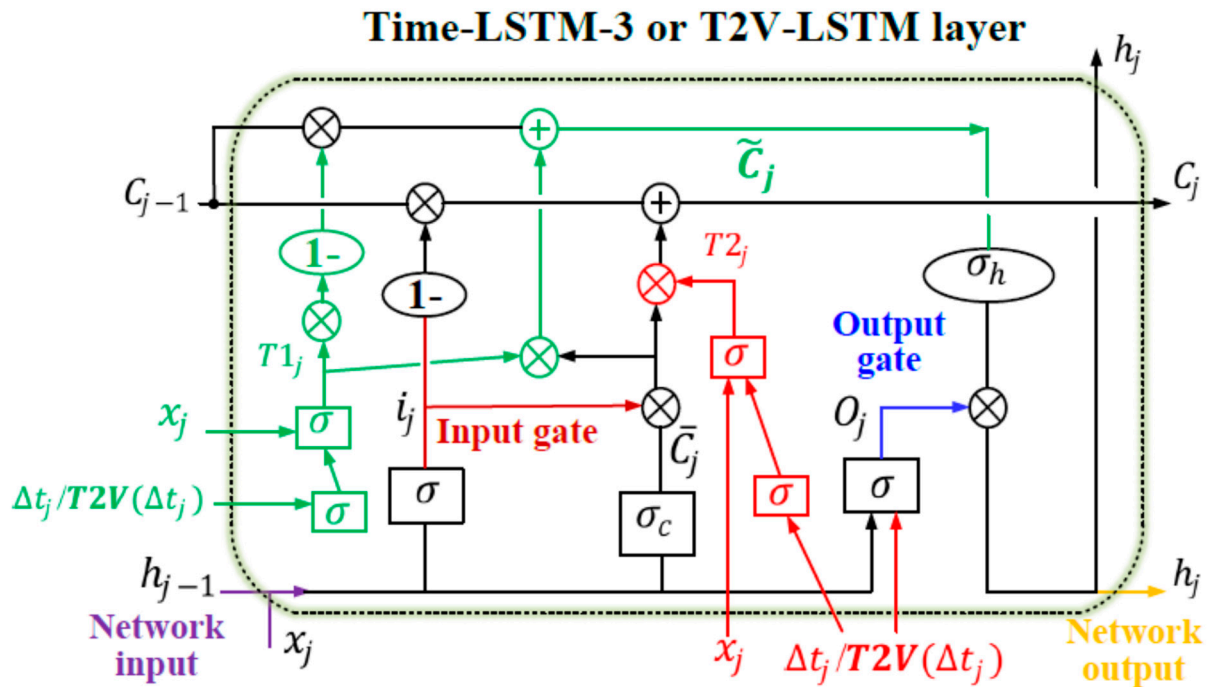


Figure 7. Time-LSTM-3 or T2V-LSTM architecture.

### 3.4. Validations

To evaluate the performances of neural networks and other ML classifiers, it is important to select the appropriate evaluation metrics. The accuracy in Equation (2) is commonly applied, but the overall accuracy of classification results on datasets with a significant imbalance is inappropriate for determining the predictive performance due to far more quantitative fault-free samples than faulty samples. The evaluation of overall fault predictions (FPs) is reflected by the macro precision (MAP) in Equation (23), and FNs are captured by the macro recall (MAR) in Equation (24). Moreover, the performance metrics, micro precision (MIP) and micro recall (MIR), are applied for fault diagnosis on individual faults, as seen in Equations (25) and (26), respectively.

$$MAP = \frac{1}{k} \sum_{i=1}^k \frac{TP_i}{TP_i + FP_i} \quad (23)$$

$$MAR = \frac{1}{k} \sum_{i=1}^k \frac{TP_i}{TP_i + FN_i} \quad (24)$$

$$MIP = \sum_{i=1}^l \frac{TP_i}{TP_i + FP_i} \quad (25)$$

$$MIR = \sum_{i=1}^l \frac{TP_i}{TP_i + FN_i} \quad (26)$$

$$F - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (27)$$

where  $k$  is the total number of classes, and  $l$  is the specific fault class. The F-score in Equation (27) is applied as the harmonic mean of precision and recall scores for both overall and individual fault diagnosis methods.

### 3.5. Model Optimisation

The objective of any neural network is to minimize the cost functions for the most accurate prediction performance by optimising the weights and biases with appropriate activation functions [41]. The Time2Vec activation function,  $\mathcal{F}$ , in Equation (19), as well as the activations functions of the LSTM layer,  $\sigma_c$  and  $\sigma_h$ , in Figure 7, are pivotal to the design of LSTM or T2V-LSTM classifiers by affecting their predictive performance.

GridSearchCV [42] is a hyper-parameter optimisation method based on a given neural network model to optimise the individual model for each combination of hyper parameters, such as the number of epochs, batch sizes, and activation functions. The optimization of hyper parameters intends to maximise prediction accuracy by minimising the cost functions and training times of neural networks. The tunes of hyper parameters are achieved by GridSearchCV, which adopts the  $k$ -fold cross-validation (CV) to train and test the neural network by grid-searching the combination of hyper parameters to generate the highest average score across  $k$  repeated times. The hyper parameters tuned for T2V-LSTM can be seen in Table 4.

**Table 4.** Hyper-parameter optimisation through GridSearchCV.

Hyper Parameter	Grid	Optimisation
Batch size	10, 20, 25, 40, 50, 60, 80, 100	25
Number of Epochs	10, 20, 25, 40, 50, 60, 80, 100	100
Activation function ( $\mathcal{F}$ )	<i>Elu, Relu, Sigmoid, Sine (only Time2Vec), Softmax, Softplus, Softsign, Tanh</i>	<i>Tanh</i>
Activation function ( $\sigma_c$ )	<i>Elu, Relu, Sigmoid, Softmax, Softplus, Softsign, Tanh</i>	<i>Tanh</i>
Activation function ( $\sigma_h$ )	<i>Elu, Relu, Sigmoid, Softmax, Softplus, Softsign, Tanh</i>	<i>Softmax</i>

The optimum activation functions for both Time2Vec,  $\mathcal{F}$ , and the hidden layer,  $\sigma_c$ , are given by *Tanh*, seen in Equation (28), while the optimal activation function for the final classification output layer,  $\sigma_h$ , is yielded by *Softmax* in Equation (29).

$$\text{Tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (28)$$

$$\text{Softmax}(x)_i = \frac{e^{x_i}}{\sum_{j=1}^N e^{x_j}} \quad (29)$$

The softmax activation function [43] is a combination of sigmoid functions applied for multivariate classification tasks by normalising the outputs with probabilities of each class ranging from 0 to 1, so the target class is expressed by the highest probability.

## 4. Results

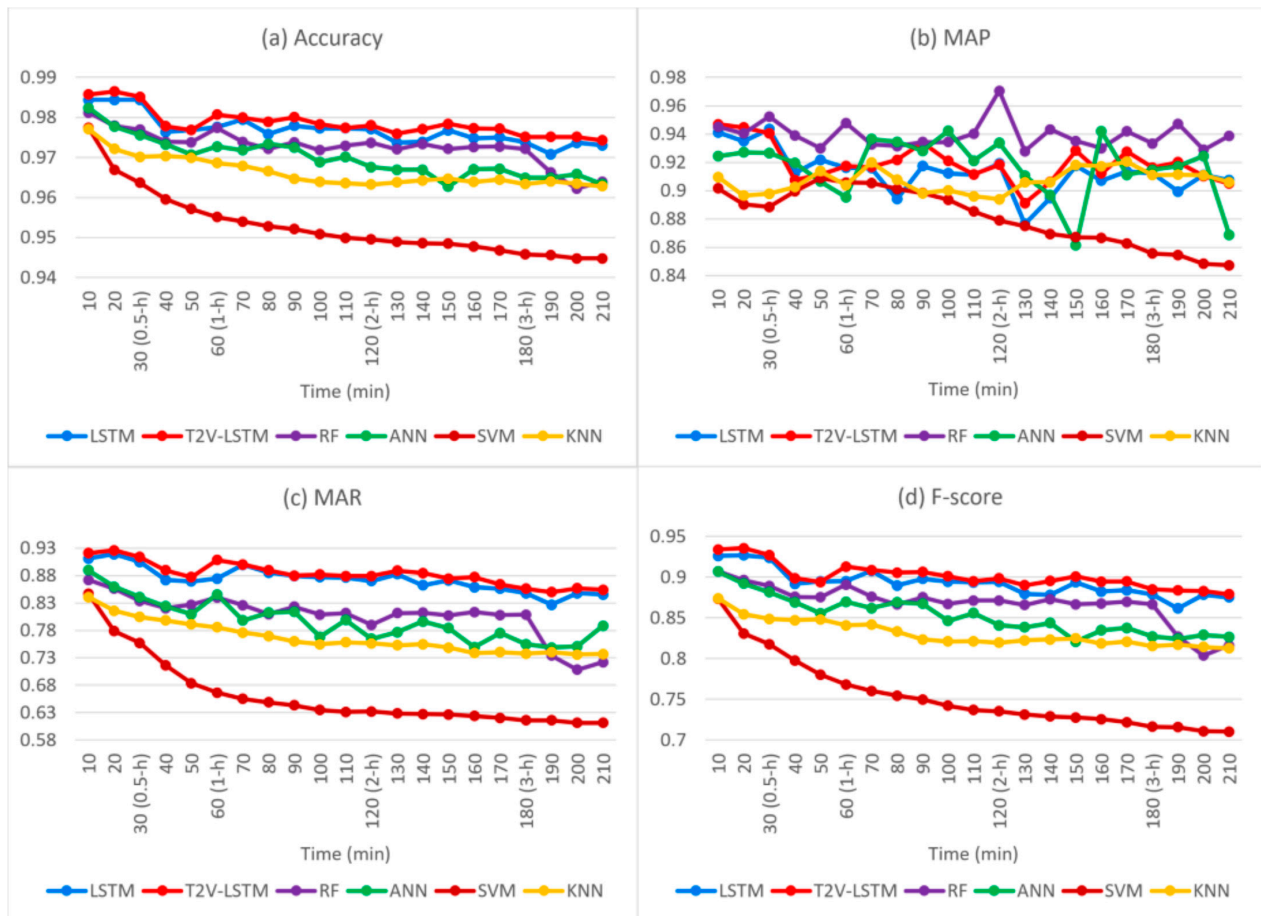
### 4.1. Overall Performance Metrics

In this subsection, the fault prediction models are extracted from timestamps  $t - 1$  (10 min) to  $t - 21$  (210 min). The detailed predictive performances under six classifiers are summarised in Figure 8 in terms of accuracy, MAP, MAR, and F-score with respect to Equations (2), (23), (24), and (27), respectively.

As seen in Table 5, all six classifiers have an upper accuracy of over 94% due to their correct predictions on fault-free cases from the imbalanced dataset. However, the resultant MARs and F-scores have poorer ranges of (0.61156, 0.92622) and (0.71038, 0.93537), respectively, and SVM especially has the poorest MIR range (0.61156, 0.84711). As MAPs (over 0.84) have better ranges than MARs (over 0.61), the resultant F-scores are promoted by precision scores, thereby representing more FNs than FPs. Among all, T2V-LSTM has correctly predicted more fault statuses than other classifiers due to its optimum MAR range.



Moreover, T2V-LSTM also has the finest ranges of accuracy and F-score despite its poorer MAP range in comparison to RF.



**Figure 8.** (a) Accuracy, (b) MAP, (c) MAR, and (d) F-score for fault predictions under six classifiers from  $t - 1$  (10 min) to  $t - 21$  (210 min).

**Table 5.** Validation scores for overall fault prediction.

Scores		Classifier (Overall Fault)						
		LSTM	T2V-LSTM	RF	ANN	SVM	KNN	ALL
Accuracy	MIN	0.97079	<b>0.9742954</b>	0.96222	0.96273	0.94477	0.96284	0.94477
	MAX	0.98441	<b>0.9864767</b>	0.97791	0.9777	0.96697	0.97213	0.98648
MAP	MIN	0.87643	0.8912656	<b>0.92785</b>	0.86133	0.84729	0.89391	0.84729
	MAX	0.94347	0.9469835	<b>0.97052</b>	0.9422	0.90898	0.92044	0.97052
MAR	MIN	0.82667	<b>0.8497778</b>	0.70844	0.74844	0.61156	0.736	0.61156
	MAX	0.91911	<b>0.9262222</b>	0.872	0.88978	0.84711	0.84089	0.92622
F-score	MIN	0.86151	<b>0.8788294</b>	0.80383	0.82085	0.71038	0.81275	0.71038
	MAX	0.92694	<b>0.935368</b>	0.90707	0.9067	0.87351	0.8739	0.93537
Execution time (s)	MIN	299.0876	<b>309.44707</b>	281.5733	232.2455	174.6638	148.6676	148.6676
	MAX	337.6192	<b>346.63702</b>	322.1635	281.5004	326.7903	186.6617	346.63702

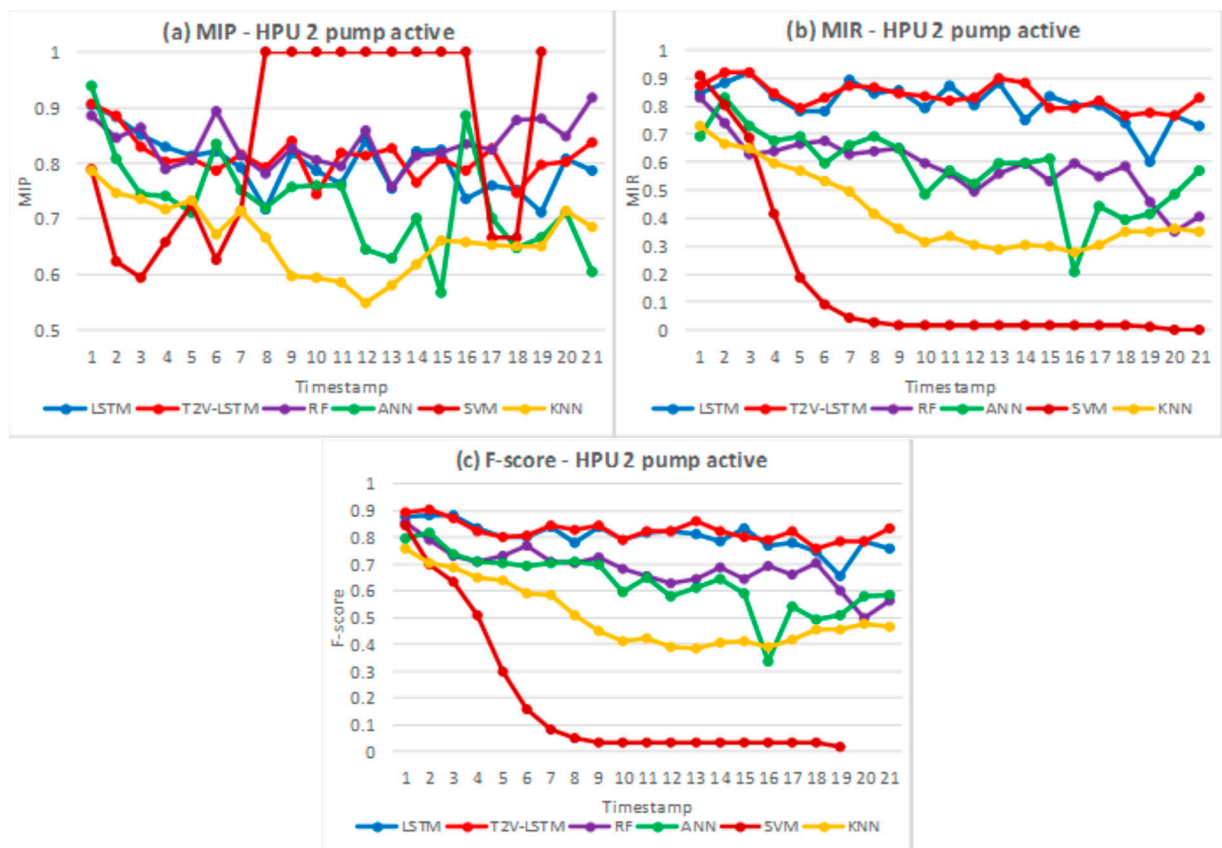
As seen in Figure 8, the time index under 10 min per timestamp in the x-axis denotes the test cases at timestamp  $t - n$ . All classification approaches demonstrate their best accuracy, MAR, and F-score initially at  $t - 1$ , but their predictive results progressively attenuate over time.

As seen in Figure 8b, RF has the best MAPs over time, implying its distinction of diagnosing fault-free conditions precisely with fewer FPs. By comparison, T2V-LSTM has successfully predicted more faults than other classifiers due to its relatively highest MARs in all test cases (see Figure 8c). Recall scores are always prior to precision scores in fault classification models because of the recall and precision scores relating to undetected failures and false fault alarms, respectively [19]. As a result, T2V-LSTM reflects its superiority over all other classifiers across overall fault predictions in terms of accuracy, MAR, and F-score (see Figure 8a,c,d)).

Apart from the best overall prediction scores, the proposed method, T2V-LSTM, requires the longest execution time, as seen in Table 5. However, the maximum execution time of T2V-LSTM (346.63702 s) is still below the minimum 10 min ahead prediction window. Thus, all six classification models can be implemented before any prediction window in all cases under the 10 min SCADA resolution.

#### 4.2. Performance Metrics upon Individual Faults

Herein, individual faults, depicted in Table 3, are examined across timestamps  $t - 1$  (10 min) to  $t - 21$  (210 min) by performance metrics, MIP, MIR, and F-score, with respect to Equations (25)–(27), respectively. The time-domain fault prediction scores for six classification approaches across individual faults are summarised in Figures 9–12. (Demoted) gearbox pressure 2 faults, Blade 3's too-slow response, gearbox cooling pressure faults, and sub-pitch fatal faults witness successful predictions due to the minimum MIR exceeding 86.27% under the proposed T2V-LSTM indicator, studied in Appendix A. Thus, in this paper, the studies on those four individual faults are of less interest.



**Figure 9.** (a) MIP, (b) MIR, and (c) F-score for predictions on HPU 2 pump active faults from  $t - 1$  to  $t - 21$ .

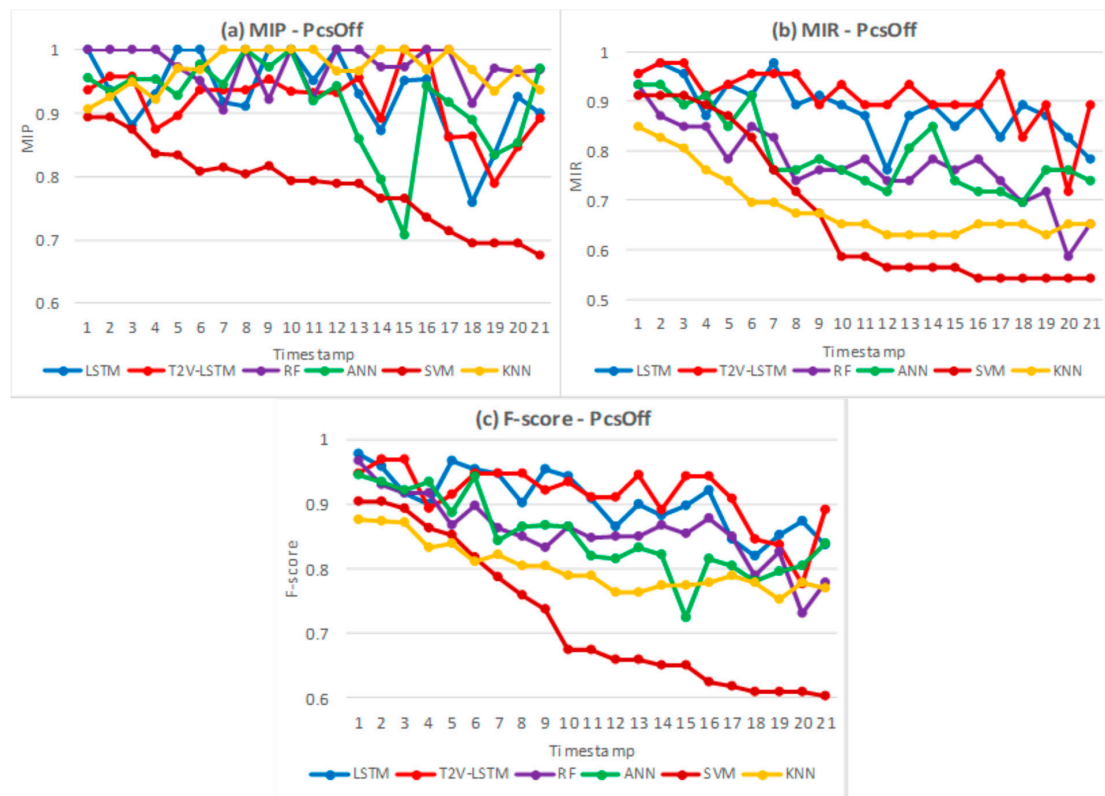


Figure 10. (a) MIP, (b) MIR, and (c) F-score for predictions on PcsOff faults from  $t - 1$  to  $t - 21$ .

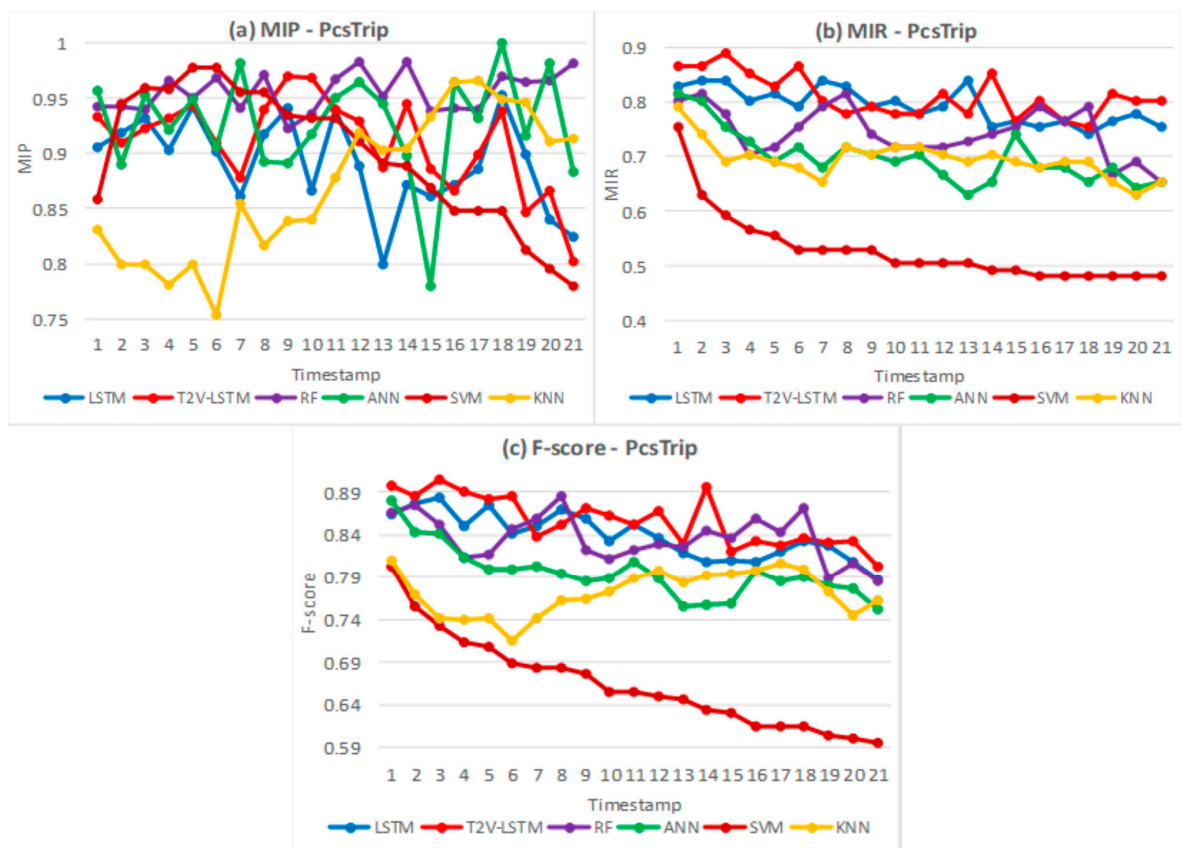


Figure 11. (a) MIP, (b) MIR, and (c) F-score for predictions on PcsTrip faults from  $t - 1$  to  $t - 21$ .

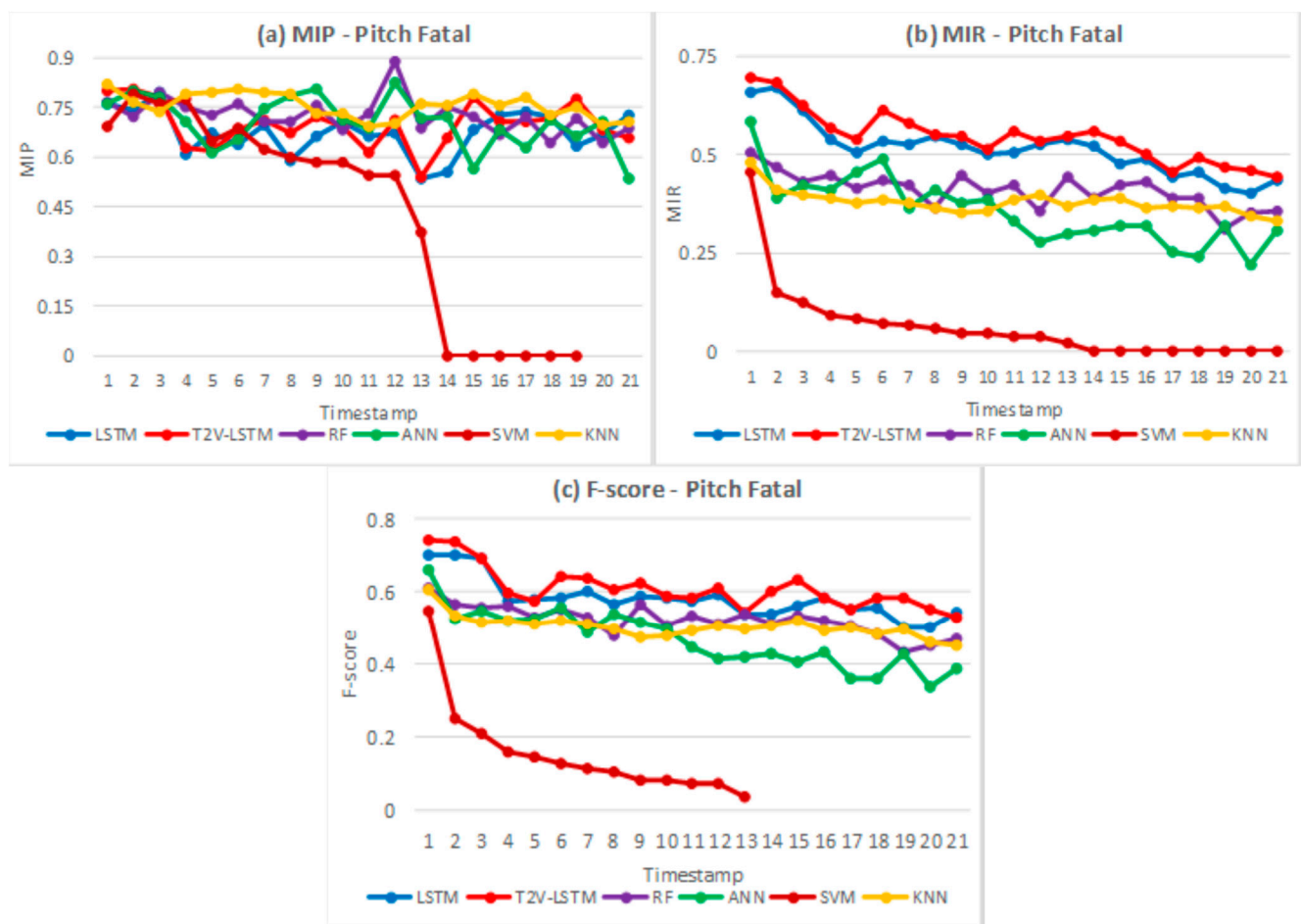


Figure 12. (a) MIP, (b) MIR, and (c) F-score for predictions on pitch fatal faults from  $t - 1$  to  $t - 21$ .

#### 4.2.1. HPU 2 Pump Active

As seen in Table 6, it is most appealing that the maximum MIP under SVM has reached full score, whilst its minimum MIR is the poorest by approaching zero. Thus, SVM is inapplicable for fault diagnosis on HPU 2 pump active. Regarding predictions on fault-free cases, both LSTM models have inferior MIP ranges compared to RF. However, T2V-LSTM is superior to all other classifiers in its best MIR range (0.7657, 0.9189), which consequently leads to its best fault prediction with the fewest FNs and the optimum range of F-score under T2V-LSTM (0.7555, 0.9026).

Table 6. Validation scores for HPU 2 pump active.

Scores		Classifier (HPU 2 Pump Active)						ALL
		LSTM	T2V-LSTM	RF	ANN	SVM	KNN	
MIP	MIN	0.71277	0.744	<b>0.7561</b>	0.56667	0.59375	0.54839	0.54839
	MAX	0.90385	0.9065421	0.91837	0.93902	<b>1</b>	0.78641	1
MIR	MIN	0.6036	<b>0.7657658</b>	0.35135	0.20721	0	0.27928	0
	MAX	<b>0.91892</b>	<b>0.9189189</b>	0.82883	0.82883	0.90991	0.72973	0.91892
F-score	MIN	0.65366	<b>0.7555556</b>	0.49682	0.33577	0.01786	0.38554	0.01786
	MAX	0.88312	<b>0.9026549</b>	0.85581	0.81778	0.84519	0.75701	0.90265

Figure 9 exhibits the time-domain prediction results in advance of HPU 2 pump active faults. As seen in Figure 9b,c, SVM has the steepest downtrend in its MIR and F-score, while T2V-LSTM goes beyond other classifiers in most test cases. Although T2V-LSTM is



the best predictor for HPU2 pump active faults by correctly predicting 76.57~91.89% of faults, HPU 2 pump active beholds smaller MIR ranges, compared to the predictive results on the four specific faults in Table A1.

#### 4.2.2. PcsOff

As seen in Table 7, SVM still has the lowest minimum and maximum values in MIPs, MIRs, and F-scores, thereby the worst prediction results. Except for SVM, all other classifiers have reached full MIP scores, and KNN and RF outscore their counterparts in MIP ranges, with all MIPs surpassing 0.9. Both LSTM models outscore all other classifiers in MIR ranges, but LSTM has a higher minimum MIR and F-score than T2V-LSTM due to a substandard MIR (0.7173913) and F-score (0.7764706) under T2V-LSTM at  $t - 2$ , seen in Figure 10b.

**Table 7.** Validation scores for PcsOff.

Scores		Classifier (PcsOff)						
		LSTM	T2V-LSTM	RF	ANN	SVM	KNN	ALL
MIP	MIN	0.75926	0.7884615	0.90476	0.70833	0.67568	<b>0.90698</b>	0.67568
	MAX	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	0.89362	<b>1</b>	1
MIR	MIN	<b>0.76087</b>	0.7173913	0.58696	0.69565	0.54348	0.63043	0.54348
	MAX	<b>0.97826</b>	<b>0.9782609</b>	0.93478	0.93478	0.91304	0.84783	0.97826
F-score	MIN	<b>0.82</b>	0.7764706	0.72973	0.7234	0.60241	0.75325	0.60241
	MAX	<b>0.97778</b>	0.9677419	0.96629	0.94505	0.90323	0.8764	0.97778

The time-domain predictive results ahead of PcsOff faults are illustrated in Figure 10. The degraded performances under both LSTM models can be recognisably obtained after  $t - 17$ . As seen in Figure 10a, RF and ANN have more optimal predictions on fault-free cases by their MIPs exceeding 0.9 in all cases.

As seen in Figure 10b, both LSTM models have the highest MIRs before  $t - 17$ , despite the poor MIR under LSTM (0.76087) at  $t - 12$ , and T2V-LSTM outclasses all other models due to its MIRs surpassing 0.89.

However, the curtailments in MIRs under both LSTM models are visualised after  $t - 18$ . Thus, both LSTM models can roughly predict over 80% of fault cases.

Consequently, both LSTM models have more balanced F-scores over other classifiers (see Figure 10c). Although LSTM obtains better ranges in MIR and F-score than T2V-LSTM (see Table 7), T2V-LSTM has predicted fault cases more correctly with respect to its greater MIRs in most test cases from Figure 10b.

#### 4.2.3. PcsTrip

As seen in Table 8, the MIP range (0.75342, 1) is expressively upper than the MIR range (0.48148, 0.88889). Hence, the predictions on PcsTrip faults witness relatively lower recall scores than precision scores, resulting in more FNs than FPs. Thus, F-scores are increased by relatively better MIPs. However, SVM still has the worst prediction results on fault cases due to its poorest MIR range. Both LSTM models have satisfying MIPs by surpassing 0.8, but they are outclassed by RF, which has the most correct predictions on fault-free cases due to its highest minimum MIP. Both LSTM models have better MIR ranges and, thereby, more accurate fault predictions. Moreover, T2V-LSTM yields more correct predictions on fault cases with fewer FNs and generates the resultant optimum range on F-scores.

Regarding the time-domain prediction results in Figure 11, the MIPs under both LSTM models underperform RF, ANN, and KNN, whilst T2V-LSTM has superiority on MIRs in most test cases. Therefore, T2V-LSTM is the best fault predictor for PcsTrip faults with its best MIRs, leading to the fewest FNs and most balanced F-scores in Figure 11c.



**Table 8.** Validation scores for PcsTrip.

Scores		Classifier (PcsTrip)						
		LSTM	T2V-LSTM	RF	ANN	SVM	KNN	ALL
MIP	MIN	0.8	0.8024691	<b>0.92308</b>	0.77922	0.78	0.75342	0.75342
	MAX	0.95238	0.969697	0.98361	<b>1</b>	0.97826	0.96552	1
MIR	MIN	0.74074	<b>0.7530864</b>	0.65432	0.62963	0.48148	0.62963	0.48148
	MAX	0.83951	<b>0.8888889</b>	0.81481	0.81481	0.75309	0.79012	0.88889
F-score	MIN	0.7871	<b>0.8024691</b>	0.78519	0.75177	0.59542	0.71429	0.59542
	MAX	0.88312	<b>0.9056604</b>	0.88591	0.88	0.80263	0.81013	0.90566

#### 4.2.4. Pitch Fatal Faults

In addition to the predictions on PcsTrip faults, the predictive results on pitch fatal faults are yielded with an even lower recall range (0, 0.69481) than the corresponding precision range (0, 0.8871), seen in Table 9. By excluding the poorest predictor, SVM, the MIPs and MIRs under other classifiers go beyond 0.53 and 0.22, respectively. Hence, for pitch fatal faults, MIPs are much greater than their corresponding MIRs, resulting in more FNs than FPs, so F-scores are downgraded by relatively poorer MIRs.

**Table 9.** Validation scores for pitch fatal faults.

Scores		Classifier (Pitch Fatal Faults)						
		LSTM	T2V-LSTM	RF	ANN	SVM	KNN	ALL
MIP	MIN	0.53548	0.5419355	0.64286	0.53409	0	<b>0.69412</b>	0
	MAX	0.79661	0.8076923	<b>0.8871</b>	0.82692	0.7931	0.82222	0.8871
MIR	MIN	0.4026	<b>0.4415584</b>	0.31169	0.22078	0	0.33117	0
	MAX	0.66883	<b>0.6948052</b>	0.50649	0.58442	0.45455	0.48052	0.69481
F-score	MIN	0.50196	<b>0.5291829</b>	0.43439	0.33663	0.03704	0.45133	0.03704
	MAX	0.70383	<b>0.7430556</b>	0.60938	0.66176	0.54902	0.60656	0.74306

The time-domain predictions prior to pitch fatal faults are seen in Figure 12. T2V-LSTM is the best fault predictor by having superior MIRs, and LSTM is just second to T2V-LSTM, whilst the other classifiers yield the MIRs below 0.5 in most cases from Figure 12b. It is noteworthy that MIRs under both LSTM models attenuate over time, obtaining lower than 0.5 after  $t - 17$ . By comparison, MIPs mainly surpass 0.5, except for the test cases under SVM in Figure 12a.

Accordingly, as seen in Figure 12c, all F-scores are decreased by their lower MIRs, but both LSTM models outperform other classifiers due to their observably better recall scores in Figure 12b, and T2V-LSTM is still the best fault predictor with the best MIRs and F-scores for pitch fatal faults. However, compared with other faults, pitch fatal faults witness much lower MIRs, and the diagnosed fault cases attenuate over a longer prediction time. Particularly, since  $t - 17$ , the predictions on pitch fatal faults are yielded with minor reliability because of MIRs going below 50%; thus, half of the faults cannot be correctly diagnosed due to yielding more FNs than TPs.

## 5. Discussion

By conducting RFE to remove irrelevant and redundant information from full operational SCADA data, 33 top-ranked features in Figure 5 are reserved for fault predictions.

LSTM has been preferable for prognostics on imbalanced data owing to its ability to store the time-lagged information and exploit the time dependency. LSTM has better recall scores (both MARs and MIRs) and overall predictions than traditional ML classifiers with respect to the results in Sections 4.1 and 4.2. However, there is also a dependence across time among data, and the time feature of inputs can be either synchronous or

asynchronous [33], but vanilla LSTM always fails to recognise time itself as a feature by assuming all inputs to be synchronous. As the data points related to downtimes or maintenance actions are removed, the modelling dataset is asynchronous. Hence, the Time2Vec is adopted to remodel the LSTM architectures into T2V-LSTM (see Figure 7) by way of Time2Vec consuming the time feature under non-stationary activation functions.

The Time2Vec activation function and the other hidden layer of T2V-LSTM in Figure 7 are chosen by *Tanh*, which maps the inputs into a range  $(-1, 1)$ . Like *Sigmoid*, the derivative of *Tanh* is expressed by itself, but the mapping range of *Tanh* is broader than that of *Sigmoid*  $(0, 1)$ . The classification output layer is selected by *Softmax* because its calculated probabilities determine the target classes with given inputs, chiefly implemented for multi-level classifications. The proposed T2V-LSTM model (with a Time2Vec function of *Tanh*) has been certified to upgrade the prediction accuracy of LSTM and outperform all other classifiers on both overall and individual fault predictions.

### 5.1. Overall Performance Metrics

Based on SCADA data with a 10 min sampling rate, T2V-LSTM provides the best adaptability in terms of accuracy, MARs, and F-scores across all timestamps, despite its smaller MAPs compared to RF and ANN (see Figure 8). Hence, the fewest unnecessary maintenance actions can be led by RF, while T2V-LSTM identifies the highest quantity of fault cases, followed by LSTM.

Integrated with Time2Vec, T2V-LSTM outscores vanilla LSTM with regard to accuracy, MAPs, MARs, and F-scores at almost all timestamps in Figure 8. Before  $t - 3$ , T2V-LSTM has its distinguished predictions in terms of accuracy (over 98.5%), MARs (over 91%), and F-scores (over 92.5%). T2V-LSTM marginally attenuates its accuracy, MARs, and F-scores over time by correctly predicting over 87.5% of faults before  $t - 16$ . However, since  $t - 17$ , T2V-LSTM has an unexpected decline in its MARs, and it can only capture 84.97% of faults at  $t - 19$ , while by comparison, MAPs under T2V-LSTM exceed 89% in all cases. Hence, overall fault predictions are validated with fewer FPs than FNs.

### 5.2. Individual Performance Metrics

Individual faults, studied in Section 4.2, witness the most advanced predictions from T2V-LSTM due to its best MIRs and F-scores across most test cases over time. T2V-LSTM has mostly better prediction scores compared to vanilla LSTM, and its resultant F-scores are well adjusted due to its more balanced MIPs and MIRs across Figures 9–12. Regarding the fault studies in Appendix A, T2V-LSTM catches over 86.27% of fault cases and over 88.28% of fault-free cases.

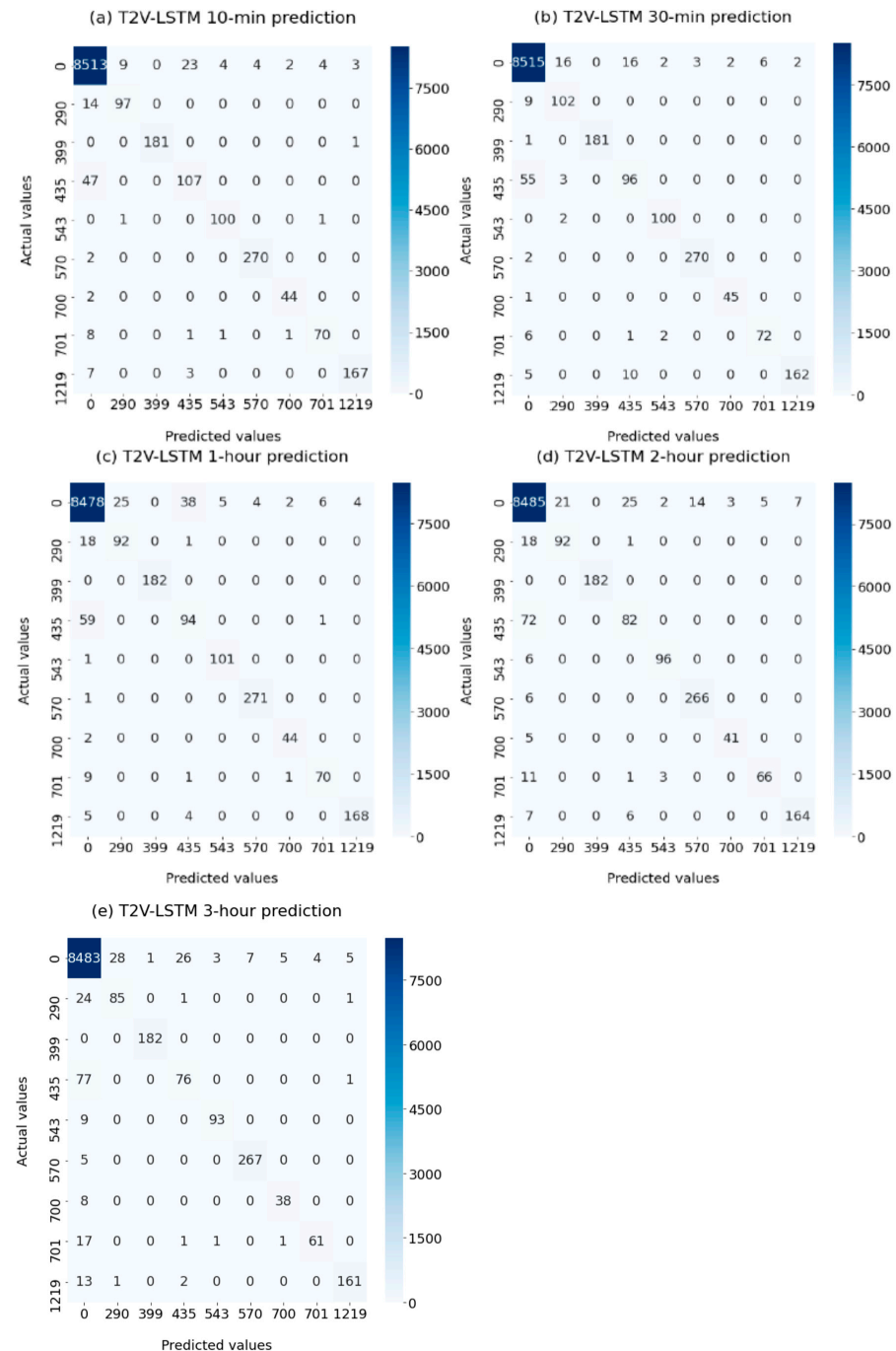
HPU 2 pump active faults exhibit a satisfactory percentage of caught faults via MIRs, mostly over 80%, as seen in Figure 9b. PcsOff faults have both excellent MIPs and MIRs over 89% before  $t - 16$ , but the predicted MIRs have significant relegations by scoring 0.826087 at  $t - 18$  and 0.717391 at  $t - 20$ , as seen in Figure 10b.

Under T2V-LSTM, all above-mentioned faults have balanced precision and recall scores, but PcsTrip and pitch fatal faults see curtailed predictions over time and greater MIPs than MIRs, resulting in both their F-scores downgraded by poorer MIRs. Fewer PcsTrip faults are correctly predicted over time regarding its maximum MIR (88.88%) and minimum MIR (75.30%) at timestamps  $t - 3$  and  $t - 18$ , respectively, as seen in Figure 11b. It is noticeable that the MIRs on pitch fatal faults are even poorer, deteriorating initially from 69.48% to 44.15% over time, as seen in Figure 12b. The forecasts over 40 min ahead on pitch fatal faults show poor results with the least MIRs (below 60%) among all individual faults, and over half of the faults cannot be correctly diagnosed after  $t - 17$ .

T2V-LSTM under a non-stationary *Tanh* function shows its peak effectiveness for both overall and individual fault predictions, according to its overall best accuracy, recall scores (both MARs and MIRs), and F-scores. However, the significant mitigations in accuracy, MARs, and F-scores from Figure 8 are mainly reflected by the attenuated MIRs and F-scores from pitch fatal faults since  $t - 4$  (40 min) in Figure 12b,c.

### 5.3. Confusion Matrix

T2V-LSTM is the best-performing classifier for both overall fault predictions and specific fault predictions. Then, an additive classification step is to visualise the predictions of 10 min, 30 min, 1 h, 2 h, and 3 h in advance via the confusion matrices under T2V-LSTM in Figure 13. The fault-free and fault cases in Figure 13 are represented by their corresponding event codes in Table 3.



**Figure 13.** Confusion matrices under T2V-LSTM at five timestamps: (a)  $t - 1$  (10 min); (b)  $t - 3$  (30 min); (c)  $t - 6$  (1 h); (d)  $t - 12$  (2 h); (e)  $t - 18$  (3 h).

Except for more FNs than FPs from PcsTrip (event code 701) and pitch fatal faults (event code 435), the balances between recall and precision scores are established with regard to their unbiased FPs and FNs from confusion matrices in Figure 13.

The most frequent (Demoted) gearbox pressure 2 faults (event code 570) witness successful fault predictions with few FNs but obtain the 14 FPs at  $t - 12$  from Figure 13d. More accurate predictions can be witnessed on the Blade 3 slow response (event code 399) by yielding, at most, 1 FN or FP. Gearbox cooling pressure faults (event code 543) have great fault-free predictions due to minor FPs, but the relevant misdiagnosed fault cases increasing with time, 6 and 9 FNs at timestamps  $t - 12$  and  $t - 18$ , as seen in Figure 13d,e, respectively.

Sub-pitch fatal faults (event code 1219) have TPs ranging from 161 to 168, with a maximum of 7 FPs, resulting in MIPs over 95%. Since the fewest TPs are obtained at  $t - 18$  in Figure 13e with 16 misdiagnosed fault cases, the minimum MIR reaches 90.96%.

HPU 2 pump active faults (event code 290) obtain the best prediction at  $t - 3$  with the maximum 102 TPs, 9 FNs, and a total of 21 FPs, so the resultant MIP and MIR reach 82.92% and 91.89%, respectively. However, the worst prediction at  $t - 18$  is yielded with a minimum of 85 TPs, 26 FNs, and 29 FPs in total, leading to the poorest MIP (74.56%) and MIR (76.57%). Hence, the predictions on HPU 2 pump active faults witness less success over a longer prediction horizon.

In addition to the HPU 2 pump active faults, the prediction scores on the least frequent PcsOff faults (event code 700) gradually worsen over time. The best prediction on PcsOff is at  $t - 3$ , when only 1 FN and 2 FPs are obtained to confirm its notable MIP (95.74%) and MIR (97.82%). However, the worst prediction at  $t - 18$  generates 38 TPs with a total of 6 FPs and 8 FNs, thereby yielding the resultant MIP (86.36%) and MIR (82.60%).

Regarding PcsTrip faults (event code 701), MIPs are always satisfactory concerning the maximum 7 FPs at  $t - 6$ , whilst MIRs decline over time. The best prediction on PcsTrip faults at  $t - 3$  is provided with 72 TPs, 6 FPs, and 9 FNs, leading to the resultant MIP (92.30%) and MIR (88.88%). By comparison, the worst case at  $t - 18$  yields relatively poorer results with 61 TPs, 4 FPs, and 20 FNs, leading to an agreeable MIP (93.84%) but an undervalued MIR (75.30%). Hence, the predictions on PcsTrip have excellent precision scores, but the resulting F-scores are brought down by gradually declined MIRs.

In addition to PcsTrip faults, the subsequent F-scores of pitch fatal faults (event code 435) are declined by poorer recall scores. Among all faults, the pitch fatal faults witness the most misdiagnosed fault cases, yielding the maximum FNs throughout the time. Initially, at  $t - 1$ , the MIR (69.48%) is acceptable due to 107 TPs out of 154 total fault cases, whilst the MIP (79.85%) is much greater owing to a total of 27 FPs. With a longer prediction horizon, more fault cases are wrongly predicted, accompanied by reduced TPs and increased FNs, which are shown in Figure 12. It is considerable that the prediction at  $t - 18$  yields an MIR of merely 49.35%, along with its relevant MIP scoring 72.38%. Hence, pitch fatal faults have observed extremely lower MIRs in comparison to other faults, and their recall scores are exceptionally exceeded by the relevant precision scores.

## 6. Conclusions

By integrating the vanilla LSTM model with a model-agnostic vector representation for time, Time2Vec, a novel neural network model, T2V-LSTM, is developed to predict multivariate faults with a 7 MW offshore wind turbine based on SCADA data. This approach has shown its efficacy on both overall and specific fault predictions by outperforming LSTM and other ML classifiers in most test cases. It has been proven that all classification models can be implemented prior to the next prediction window in all cases under the 10 min SCADA resolution. Using a feature selection method, RFE, to assess the importance of features for dimension reduction, 33 optimal features are extracted to improve the prediction accuracy and computing efficiency of neural networks. Regarding the T2V-LSTM prediction results, the following conclusions can be noted:

- As there are eight specific faults and massive data imbalances studied in this research, T2V-LSTM can successfully predict all faults 160 min before their occurrence with an overall recall score (MAR) of over 87.5%. T2V-LSTM outperforms LSTM and other classifiers in terms of accuracy, recall scores (both MARs and MIRs), and F-scores in all test cases, but with a longer lagged time, the MAR abruptly falls to roughly 85%.

- T2V-LSTM has satisfactory predictions on (Demoted) gearbox pressure 2 faults, Blade 3 slow response, gearbox cooling faults, and sub-pitch fatal faults, due to its minimum MIP over 88.28% and minimum MIR over 86.27%, shown in Appendix A. Approximately 80% of the HPU 2 pump active faults are correctly predicted along with the relevant MIPs scoring roughly 80%. PcsOff faults exhibit excellent prediction results 160 min before the occurrence, with both recall and precision scores over 89%, but the significantly curtailed MIPs and MIRs take place over a longer prediction horizon. The F-scores on those faults are balanced due to their unbiased and promising precision and recall decisions.
- However, the balance between MIPs and MIRs is demolished under PcsTrip and pitch fatal faults due to their F-scores being brought down by poorer recall scores. PcsTrip and pitch fatal faults behold upper MIP ranges than MIR ranges and degraded predictions over time. PcsTrip faults are successfully predicted 30 min in advance due to their optimal MIR (88.88%), but the minimum MIR (75.30%) is obtained 3 h before occurrence. By comparison, MIRs on pitch fatal faults have an even more critical downtrend, reducing from 69.48% to 44.15% over time. Particularly, over half of pitch fatal faults are misdiagnosed >170 min before occurrence. The curtailments in MIRs on pitch fatal faults over 40 min ahead predominately contribute to the significant degradations of overall accuracy, MARs, and F-scores. Hence, the poorest predictions on pitch fatal faults bear a considerable burden for overall prediction accuracy.
- The confusion matrices visually study the balance between recall and precision scores by predicting the faults 10, 30, 60, 120, and 180 min in advance. Apart from PcsTrip and pitch fatal faults having more biases in FNs over FPs, the other faults can acquire the balanced F-scores due to their FNs roughly equalising FPs. For those faults with balanced F-scores, the resultant MIPs and MIRs mostly surpass 80%, except for the MIP (74.56%) and MIR (76.57%) from the 3 h ahead prediction on HPU 2 pump active faults. PcsTrip faults mainly have excellent MIPs over 90%, but the degradations on their MIRs are expected over time. Hence, the prediction curtailments provided by HPU 2 pump active faults and PcsTrip faults over a longer prediction horizon also contribute to the degradation of overall performance metrics.
- As T2V-LSTM fails to predict over 40% of pitch fatal faults 40 min prior to occurrence, future studies should critically focus on building a performance curve of pitch angle to improve the predictions on pitch fatal faults.

**Author Contributions:** Conceptualisation, S.Z.; methodology, S.Z.; software, S.Z.; validation, S.Z.; formal analysis, S.Z.; investigation, S.Z.; resources, S.Z.; data curation, E.R.; writing—original draft preparation, S.Z.; writing—review and editing, S.Z., E.R. and M.B.; visualisation, S.Z., E.R. and M.B.; supervision, E.R. and M.B.; project administration, M.B.; funding acquisition, M.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Sustainable Energy Authority of Ireland, Grant ref: SEAI 18/RDD/213.

**Data Availability Statement:** Restrictions apply to the availability of these data. Data were obtained from ORE Catapult and are available at <https://ore.catapult.org.uk/what-we-do/offshore-renewable-energy-research/platform-for-operational-data-pod/> (accessed on 23 November 2023) with the permission of ORE Catapult.

**Acknowledgments:** We are grateful to ORE Catapult for giving us the permission to use the SCADA datasets.

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

## Appendix A

As seen in Table A1, the predictive results on (Demoted) gearbox pressure 2 faults, Blade 3's too-slow response, gearbox cooling pressure faults, and sub-pitch fatal faults are validated across timestamps  $t - 1$  (10 min) to  $t - 21$  (210 min) by the minimums and

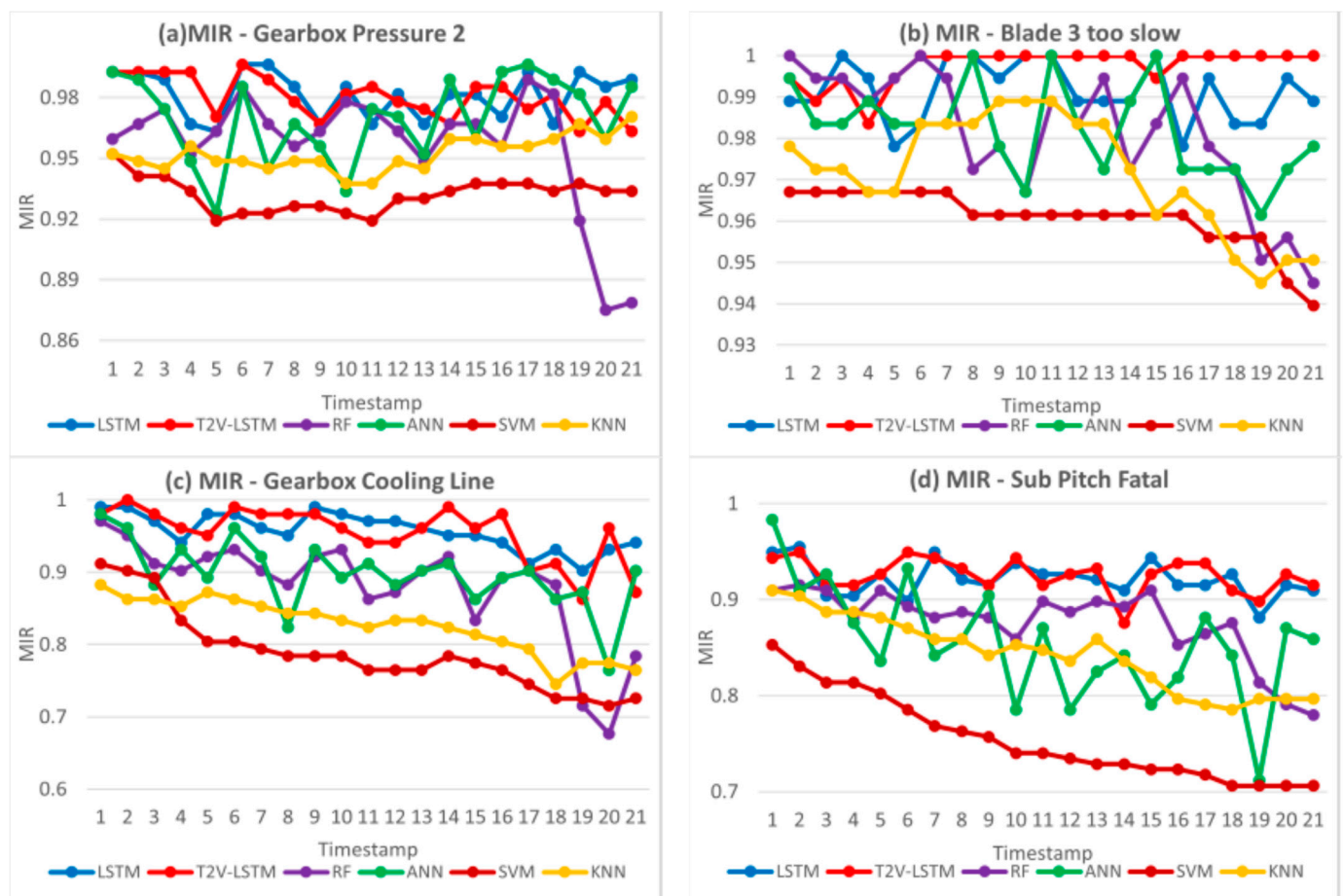


maximums of their performance metrics. Both LSTM models manage to predict over 86.27% of fault cases, outperforming their ML counterparts.

Although LSTM has the finest minimum MIP, MIR, and F-score in Table A1, T2V-LSTM still outperforms LSTM in fault diagnosis because of greater MIRs in most test cases, as seen in the time-domain MIR results from Figure A1.

**Table A1.** Validation scores for (Demoted) gearbox pressure 2 faults, Blade 3's too-slow response, gearbox cooling pressure faults, and sub-pitch fatal faults.

Scores		Classifier (Four Other Individual Faults)						
		LSTM	T2V-LSTM	RF	ANN	SVM	KNN	ALL
MIP	MIN	0.90654	0.8828829	<b>0.90789</b>	0.84444	0.8038	0.85106	0.8038
	MAX	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	0.99435	0.99425	<b>1</b>
MIR	MIN	<b>0.88136</b>	0.8627451	0.67647	0.71186	0.70621	0.7451	0.67647
	MAX	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	0.96703	0.98901	<b>1</b>
F-score	MIN	<b>0.90698</b>	0.9035533	0.77528	0.82353	0.75758	0.82162	0.75758
	MAX	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	0.9805	0.97814	<b>1</b>



**Figure A1.** MIP, MIR, and F-score from  $t - 1$  to  $t - 21$ : (a) (Demoted) gearbox pressure 2 faults; (b) Blade 3's too slow response; (c) gearbox cooling pressure faults; (d) sub-pitch fatal faults.

## References

1. Singh, U.; Rizwan, M.; Malik, H. Wind Energy Scenario, Success and Initiatives Towards. *Energies* **2022**, *15*, 2291. [\[CrossRef\]](#)
2. Sommer, B.; Pinson, P.; Messner, J.W.; Obst, D. Online Distributed Learning in Wind Power Forecasting. *Int. J. Forecast.* **2021**, *37*, 205–223. [\[CrossRef\]](#)
3. Maldonado-Correa, J.; Solano, J.C.; Rojas-Moncayo, M. Wind Power Forecasting: A Systematic Literature Review. *Wind Eng.* **2021**, *45*, 413–426. [\[CrossRef\]](#)
4. Kaldellis, J.K.; Apostolou, D.; Kapsali, M.; Kondili, E. Environmental and Social Footprint of Offshore Wind Energy Comparison with Onshore Counterpart. *Renew. Energy* **2016**, *92*, 543–556. [\[CrossRef\]](#)
5. Al-Hinai, A.; Charabi, Y.; Kaboli, S.H.A. Offshore Wind Energy Resource Assessment across the Territory of Oman: A Spatial-Temporal Data Analysis. *Sustainability* **2021**, *13*, 2862. [\[CrossRef\]](#)
6. Gao, Z.; Liu, X. An Overview on Fault Diagnosis, Prognosis and Resilient Control for Wind Turbine Systems. *Processes* **2021**, *9*, 300. [\[CrossRef\]](#)
7. Zhang, S.; Robinson, E.; Basu, M. Hybrid Gaussian Process Regression and Fuzzy Inference System Based Approach for Condition Monitoring at the Rotor Side of a Doubly Fed Induction Generator. *Renew. Energy* **2022**, *198*, 936–946. [\[CrossRef\]](#)
8. Leite, G.d.N.P.; Araújo, A.M.; Rosas, P.A.C. Prognostic Techniques Applied to Maintenance of Wind Turbines: A Concise and Specific Review. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1917–1925. [\[CrossRef\]](#)
9. Sun, Y.; Kang, J.; Sun, L.; Jin, P.; Bai, X. Condition-Based Maintenance for the Offshore Wind Turbine Based on Long Short-Term Memory Network. *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.* **2022**, *236*, 542–553. [\[CrossRef\]](#)
10. Koukoura, S.; Peeters, C.; Helsen, J.; Carroll, J. Investigating Parallel Multi-Step Vibration Processing Pipelines for Planetary Stage Fault Detection in Wind Turbine Drivetrains Investigating Parallel Multi-Step Vibration Processing Pipelines for Planetary Stage Fault Detection in Wind Turbine Drivetrains. *J. Phys. Conf. Ser.* **2020**, *1618*, 022054. [\[CrossRef\]](#)
11. Raposo, H.; Farinha, J.T.; Fonseca, I.; Galar, D. Predicting Condition Based on Oil Analysis—A Case Study. *Tribol. Int.* **2019**, *135*, 65–74. [\[CrossRef\]](#)
12. Zappalá, D.; Sarma, N.; Djurović, S.; Crabtree, C.J.; Mohammad, A.; Tavner, P.J. Electrical & Mechanical Diagnostic Indicators of Wind Turbine Induction Generator Rotor Faults. *Renew. Energy* **2019**, *131*, 14–24. [\[CrossRef\]](#)
13. Yang, W.; Peng, Z.; Wei, K.; Tian, W. Structural Health Monitoring of Composite Wind Turbine Blades: Challenges, Issues and Potential Solutions. *IET Renew. Power Gener.* **2017**, *11*, 411–416. [\[CrossRef\]](#)
14. Yang, W.; Tavner, P.J.; Crabtree, C.J.; Feng, Y.; Qiu, Y. Wind Turbine Condition Monitoring: Technical and Commercial Challenges. *Wind Energy* **2014**, *17*, 673–693. [\[CrossRef\]](#)
15. Tautz-Weinert, J.; Watson, S.J. Using SCADA Data for Wind Turbine Condition Monitoring—A Review. *IET Renew. Power Gener.* **2017**, *11*, 382–394. [\[CrossRef\]](#)
16. Stetco, A.; Dinmohammadi, F.; Zhao, X.; Robu, V.; Flynn, D.; Barnes, M.; Keane, J.; Nenadic, G. Machine Learning Methods for Wind Turbine Condition Monitoring: A Review. *Renew. Energy* **2019**, *133*, 620–635. [\[CrossRef\]](#)
17. Sahnoun, M.; Baudry, D.; Mustafee, N.; Louis, A.; Smart, P.A.; Godsiff, P.; Mazari, B. Modelling and Simulation of Operation and Maintenance Strategy for Offshore Wind Farms Based on Multi-Agent System. *J. Intell. Manuf.* **2019**, *30*, 2981–2997. [\[CrossRef\]](#)
18. Lu, L.; He, Y.; Wang, T.; Shi, T.; Li, B. Self-Powered Wireless Sensor for Fault Diagnosis of Wind Turbine Planetary Gearbox. *IEEE Access* **2019**, *7*, 87382–87395. [\[CrossRef\]](#)
19. Leahy, K.; Hu, R.L.; Konstantakopoulos, I.C.; Spanos, C.J.; Agogino, A.M.; Sullivan, D.T.J.O. Diagnosing and Predicting Wind Turbine Faults from SCADA Data Using Support Vector Machines. *Int. J. Progn. Health Manag.* **2018**, *9*, 1–11.
20. Naik, S.; Koley, E. Fault Detection and Classification Scheme Using KNN for AC/HVDC Transmission Lines. In Proceedings of the 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 17–19 July 2019; pp. 1131–1135. [\[CrossRef\]](#)
21. Marti-Puig, P.; Blanco-M, A.; Cárdenas, J.J.; Cusidó, J.; Solé-Casals, J. Feature selection algorithms for wind turbine failure prediction. *Energies* **2019**, *12*, 453. [\[CrossRef\]](#)
22. Ibrahim, R.K.; Tautz-Weinert, J.; Watson, S.J. *Neural Networks for Wind Turbine Fault Detection via Current Signature Analysis*; Wind Europe: Hamburg, Germany, 2016.
23. Kandukuri, S.T.; Senanyaka, J.S.L.; Huynh, V.K.; Robbersmyr, K.G. A Two-Stage Fault Detection and Classification Scheme for Electrical Pitch Drives in Offshore Wind Farms Using Support Vector Machine. *IEEE Trans. Ind. Appl.* **2019**, *55*, 5109–5118. [\[CrossRef\]](#)
24. Malik, H.; Mishra, S. Artificial Neural Network and Empirical Mode Decomposition Based Imbalance Fault Diagnosis of Wind Turbine Using TurbSim, FAST and Simulink. *IET Renew. Power Gener.* **2017**, *11*, 889–902. [\[CrossRef\]](#)
25. Yang, X.; Zhang, Y.; Lv, W.; Wang, D. Image Recognition of Wind Turbine Blade Damage Based on a Deep Learning Model with Transfer Learning and an Ensemble Learning Classifier. *Renew. Energy* **2021**, *163*, 386–397. [\[CrossRef\]](#)
26. Zhang, D.; Qian, L.; Mao, B.; Huang, C.; Huang, B.; Si, Y. A Data-Driven Design for Fault Detection of Wind Turbines Using Random Forests and XGboost. *IEEE Access* **2018**, *6*, 21020–21031. [\[CrossRef\]](#)
27. Helbing, G.; Ritter, M. Deep Learning for Fault Detection in Wind Turbines. *Renew. Sustain. Energy Rev.* **2018**, *98*, 189–198. [\[CrossRef\]](#)
28. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv* **2014**, arXiv:1412.3555.

29. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
30. Chen, H.; Liu, H.; Chu, X.; Liu, Q.; Xue, D. Anomaly Detection and Critical SCADA Parameters Identification for Wind Turbines Based on LSTM-AE Neural Network. *Renew. Energy* **2021**, *172*, 829–840. [[CrossRef](#)]
31. Lei, J.; Liu, C.; Jiang, D. Fault Diagnosis of Wind Turbine Based on Long Short-Term Memory Networks. *Renew. Energy* **2019**, *133*, 422–432. [[CrossRef](#)]
32. Yu, L.; Qu, J.; Gao, F.; Tian, Y. A Novel Hierarchical Algorithm for Bearing Fault Diagnosis Based on Stacked LSTM. *Shock Vib.* **2019**, *2019*, 2756284. [[CrossRef](#)]
33. Kazemi, S.M.; Goel, R.; Eghbali, S.; Ramanan, J.; Sahota, J.; Thakur, S.; Wu, S.; Smyth, C.; Poupart, P.; Brubaker, M. Time2Vec: Learning a Vector Representation of Time. *arXiv* **2019**, arXiv:1907.05321.
34. Granitto, P.M.; Furlanello, C.; Biasioli, F.; Gasperi, F. Recursive Feature Elimination with Random Forest for PTR-MS Analysis of Agroindustrial Products. *Chemom. Intell. Lab. Syst.* **2006**, *83*, 83–90. [[CrossRef](#)]
35. ORE Catapult. Levenmouth 7MW Demonstration Offshore Wind Turbine Specification Sheet. 2016, Volume 44, pp. 7–8. Available online: <https://ore.catapult.org.uk/app/uploads/2018/01/Levenmouth-7MW-demonstration-offshore-wind-turbine.pdf> (accessed on 10 September 2023).
36. Serret, J.; Rodriguez, C.; Tezdogan, T.; Stratford, T.; Thies, P. Code Comparison of a NREL-Fast Model of the Levenmouth Wind Turbine with the GH Bladed Commissioning Results. In Proceedings of the ASME 2018 37th International Conference on Ocean, Offshore and Arctic Engineering, Madrid, Spain, 17–22 June 2018; Volume 10. [[CrossRef](#)]
37. Kusiak, A.; Li, W. The Prediction and Diagnosis of Wind Turbine Faults. *Renew. Energy* **2011**, *36*, 16–23. [[CrossRef](#)]
38. Kusiak, A.; Verma, A. A Data-Driven Approach for Monitoring Blade Pitch Faults in Wind Turbines. *IEEE Trans. Sustain. Energy* **2011**, *2*, 87–96. [[CrossRef](#)]
39. Zhu, Y.; Li, H.; Liao, Y.; Wang, B.; Guan, Z.; Liu, H.; Cai, D.; Science, C. What to Do Next: Modeling User Behaviors by Time-LSTM. *IJCAI* **2017**, *17*, 3602–3608. [[CrossRef](#)]
40. Greff, K.; Srivastava, R.K.; Koutn, J.; Steunebrink, B.R. LSTM: A Search Space Odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *28*, 2222–2232. [[CrossRef](#)]
41. Suzuki, K. *Artificial Neural Networks—Methodological Advances and Biomedical Applications*; Intech: Rijeka, Croatia, 2011; Available online: <https://www.intechopen.com/books/citations/artificial-neural-networks-methodological-advances-and-biomedical-applications> (accessed on 10 September 2023).
42. Brownlee, J. How to Grid Search Hyperparameters for Deep Learning Models in Python with Keras. 2016. Available online: <https://machinelearningmastery.com/grid-search-hyperparameters-deeplearning-models-python-keras> (accessed on 10 September 2023).
43. Sharma, S. Activation functions in neural networks. *Towards Data Sci.* **2020**, *4*, 310–316. [[CrossRef](#)]

**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.