Received: April 13, 2022. Revised: May 19 2022.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

RNN-Autoencoder Approach for Anomaly Detection in Power Plant Predictive Maintenance Systems

Budi Santoso[1,2] Wiwik Anggraeni[3] Henry Pariaman[2] Mauridhi Hery Purnomo[4]*

[1]*Interdisciplinary School of Management and Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*
[2]*PT Pembangkitan Jawa Bali, Indonesia*
[3]*Department of Information Systems, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*
[4]*Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*
* Corresponding author's Email: hery@ee.its.ac.id, budisantoso.206032@mhs.its.ac.id

Abstract: Induced Draft Fans (IDF) and Primary Air Fans (PAF) are critical equipment in steam power plants. Anomaly detection based on machine learning models is an approach that is currently being developed for optimization and increasing the effectiveness of predictive maintenance (PDM) as well as increasing the reliability of thermal power plants. The research aims to develop a data-driven model for diagnostic and prognostic equipment, produce accurate predictions using many sensor data taken in real-time from the SCADA system, and design a PDM management framework using an anomaly detection system. This research proposes to use a combined recurrent neural network (RNN)-autoencoder approach as a "normal" behavior model (NBM) with the Mahalanobis Distance (MD) statistical method for the detection of anomalies in power plant equipment. Based on time-series input sensor data, the RNN – autoencoder is utilized to predict the behavior of the equipment in health circumstances. In contrast, the MD is used to determine the distance between the actual parameters of the equipment and its "normal" behavior prediction to determine the anomaly condition. This study examined the performance of the long short-term memory (LSTM) and gated recurrent unit (GRU) models in modeling normal behavior with hyperparameter optimization. The LSTM with the best hyperparameters had a validation loss of 5,690 x 10-4 and a validation accuracy of 93.36 percent, whereas the GRU with the best hyperparameters had a validation loss of 4.484 x 10-4 and a validation accuracy of 93.47 percent. GRU can outperform the LSTM model. The proposed framework can detect anomaly conditions in various cases of IDF and PAF equipment disturbances, both in early warning of equipment failure and during downtime conditions.

Keywords: Anomaly detection, LSTM, GRU, Autoencoder, Normal behavior model, Mahalanobis distance, Predictive maintenance.

1. Introduction

One of the main targets for power plant performance is the reliable and efficient operation of the plant. Power plant reliability and efficiency are closely related to corporate performance. Power generation companies with large business scales typically handle many power plants, and the issue is that the number of technical professionals required to manage these power plants is expanding. One possible option is to centralize experts to serve all power plant units, backed by the power plant's digitization strategy. [1].

Big data and artificial intelligence technologies are developing rapidly and have been widely researched to apply to the broad sector [2-4]. Previous studies have developed many methods to utilize big data in the electricity sector to increase the reliability and effectiveness of maintenance [4-8]. Big data analysis applications were developed in the electricity sectors, among others, to improve demand response, high-accuracy forecasting, failures/ downtimes detection, high-accuracy operation planning model, equipment monitoring, and lifetime

extension with failure detection as well as feature extraction and advanced visualization [2, 9].

Power plants, especially coal-fired ones, have a large amount of data generated periodically and in real-time from many sensors connected to the supervisory control and data acquisition (SCADA) system. Many previous studies have been carried out focusing on diagnostics of engine failures to minimize the occurrence of engine downtime. This diagnostic capability makes it possible to optimize machine maintenance and shift the maintenance strategy from time-based to condition-based methods.

Induced Draft Fans (IDF) and Primary air fans (PAF) equipment play a critical role in coal power plants. Induced Draft Fans are used in steam power plants to maintain a balanced draft system in the furnace by sucking in flue gas from the combustion process to maintain a steady furnace pressure. While PAF supplies the primary air used to transport pulverized coal material to the boiler furnace, it works on the principle of a centrifugal fan with variable speed regulation. Historically, this equipment has experienced several failures (breakdowns), which resulted in downtime losses. Therefore, this equipment is used as the object of a case study in this research.

Maintenance activities are carried out to avoid failure events and improve equipment performance. The presence of anomaly detection allows practical maintenance tasks planning on equipment that can reduce high losses due to unplanned downtime and has great significance in the condition based maintenance (CBM) [10-12]. Online equipment anomaly detection will indicate equipment abnormality and initiate component inspection activities in the field [10].

The use of real-time data from SCADA system for condition monitoring has been extensively researched in the past [10]. In the application of equipment failure prediction, the most chosen method is the data-driven approach, using historical data [12-14]. Data-driven methods are preferred over mathematical or physical models of equipment due to the dynamic and non-linear nature of the system [13, 15, 16].

One of the essential jobs in diagnostic and prognostic systems to improve CBM capabilities is equipment anomaly detection. This abnormality cannot be detected using the fixed threshold method because it cannot detect early detection.

This research uses an unsupervised learning method with a "normal" behavior model (NBM) approach. The model will predict the behavior pattern of equipment under health conditions using certain input features. With this method, the model will predict the continuous variable of the equipment and compare it with the actual measurement value of the variable in real-time.

This study contributes to the suggested framework for using unsupervised learning algorithms to detect abnormalities in power production equipment subjected to condition-based and predictive maintenance. The unsupervised learning approach is proposed as a solution to the difficulties of providing a training dataset for a supervised learning model that incorporates a sample of extremely rare anomalous events in power plant equipment, as well as a different pattern for each failure mode. The proposed approach aims to overcome the weakness of anomaly detection methods proposed in previous studies, whose performance is still inadequate when dealing with imbalanced datasets and large datasets with a large number of model features for equipment, without reducing the size of the dataset, which can reduce the sensitivity of anomaly detection in real-time. The methodology provided in this study is distinct from earlier work which employs recurrent neural network (RNN) variations, specifically long short-term memory (LSTM) and gated recurrent unit (GRU), as autoencoder models to represent the normal behavior of equipment using actual data from power generation. The RNN-autoencoder approach is combined with the Mahalanobis Distance approach to determine the equipment's health index. Normalized individual reconstruction error signals to determine the root cause of an anomaly state are also novel in previous research, allowing for more precise planning of maintenance tasks in a power plant based on equipment damage detection.

The remainder of this paper is organized in the following manner. Section 2 reviews some relevant work, including supervised and unsupervised learning approaches, autoencoders, LSTMs, GRUs, and the overall residual prediction value, as well as their numerous applications for anomaly detection. Then, in Section 3, we offer our proposed approach. The following Section 4 details the experimental setup for detecting equipment anomalies in a power plant. Section 5 discusses the findings and analyses. Finally, Section 6 discusses the conclusion and future works.

2. Related works

2.1 Supervised learning approach for failure detection

Machine learning based on supervised learning is one of the previously investigated ways for detecting

equipment abnormalities. In modeling the problem of anomalies classification, supervised learning approaches such as Principal Component Analysis [17], neural network [7], K-Nearest Neighbors, Logistic Regression, Multilayer Perceptron [18], and support vector machines [19] have been studied previously.

Supervised deep learning method has been used to overcome problems in the multiple kernel detection method based on one-class SVM for equipment condition classification [20], which requires dimension reduction for large datasets, lacks sensitivity to detect short-term anomalies, and is unable to detect latent features.

Supervised deep learning architectures such as LSTM [19, 20], and GRU [20] successfully handle multivariate time-series sequential data without requiring dimension reduction, could detect latent features, and may be used for real-time anomaly detection [20].

2.2 Unsupervised learning method with normal behavior model approach

The supervised learning method requires the training data with an almost balanced proportion of normal and abnormal labels to train the model. The quantity and quality of training data are also considerations in the supervised learning method. [15]. However, this is a problem for applying in power plant equipment because most equipment is healthy throughout its entire operating cycle.

Many studies have developed unsupervised learning methods to overcome the problems of unbalancing categorical training datasets [8, 11, 21, 22]. Anomaly detection was previously performed using the pattern recognition method, which detects deviations in specific datasets using the "normal" behavior model (NBM). NBM is designed to investigate the equipment operating variable's healthy behavior patterns under certain conditions and at particular times. This method involves rebuilding the 'typical' behavior of time-series parameter data and then detecting abnormalities using reconstruction errors [10-12, 15]. The data used to form the NBM of this equipment is obtained from the historical operation of the equipment with special treatment, including eliminating data when equipment downtime occurs, eliminating data when a failure occurs according to operating records, and eliminating abnormal data based on statistical characteristics [11].

2.3 Normal behavior model with autoencoder (AE)

One model that is widely used to build NBM for anomaly detection applications is the autoencoder (AE) model, which is an unsupervised learning method in which the output of the model is a reconstruction of the input [11, 14, 15] with latent-space representation techniques [23-25]. AE architecture is widely used because it can overcome the difficulty of determining features manually, which requires domain knowledge about data and processes related to it [11, 26] and can automatically extract representative features from raw data so that it significantly improves failure detection accuracy [22, 27]. AE can also improve the model's generalizability with data reduction [15, 28].

The AE approach has been applied in previous studies for the anomaly detection in power plants, including the reheater metal temperature [14], primary air fan generator [21], Pulverizer [15], the vibration of the CWP motor bearing [13], and the motor temperature of 10kV [16].

AE-based methods have been investigated to be combined with neural networks [21] and deep learning [22] to determine the non-linear relationship between features. Traditional AE neural network approaches, on the other hand, ignore the dynamic aspect of time series data [11]. Higher prediction accuracy can be achieved by paying attention to the dynamic nature of time series data. The method used to model dynamic properties and long-distance dependencies between sequence data is RNN that allows mapping input to output sequences of different lengths [14].

However, the limitation of the conventional RNN approach is the issue of vanishing gradient and exploding gradient that will limit the network's ability to explore these dependencies or relationships [29-32]. This study proposes a method for overcoming the drawback of standard RNN models, especially the presence of vanishing and exploding gradients during model training.

2.4 Autoencoder in combination with long short-term memory (LSTM) and gated recurrent unit (GRU)

The long short-term memory (LSTM) and gated recurrent unit (GRU) methods have previously been successfully developed to overcome the weaknesses of exploding gradients in RNN and overfitting in time-series applications with deep learning models [13-16, 33].

Prior research combined the Autoencoder approach for feature extraction, generalization of the model by reducing data, and the LSTM model [8], [15] to address the issue of gradient reduction in extensive periods and long sequences. The capability of the LSTM-AE model to handle multi-output, time series, and long-term dependence problems is another advantage [11].

A combination model of LSTM and autoencoder has been employed in several studies to determine the condition of power plant equipment. [13, 15, 16].

Several studies have shown that GRU and LSTM show very similar performance in a wide variety of applications, but neither of them outperforms the other in all job applications [14]. Therefore, the performance of these two types of models, LSTM and GRU, in the application of equipment anomaly detection at power plant equipment should be evaluated in this study.

2.5 Anomaly detection based on overall residual prediction value

Anomalous conditions are detected by comparing the expected behavior with the actual behavior of the equipment. When abnormal conditions are found in the equipment, the predictions generated by the model will be much different from the measurement data, resulting in gaps or prediction errors. Whether the data is included in the abnormality will be decided after computing the gap value. [10-12, 15, 21].

The use of a simple root means squared (RMS) value or the fixed threshold method of the prediction error value has not been able to early detect equipment abnormalities in anomaly detection applications in previous studies [10-12, 22]. In many cases of equipment failure, the monitored parameter values may not deviate significantly from the normal operating range, even when anomalous conditions occur [10].

One of the methods proposed in previous studies to determine abnormal conditions is statistical process control with the control chart method. These parameters are abnormal on the control chart method when a prediction error is outside the upper or lower limit based on the AE signal output's average value and standard deviation. [7, 15, 34].

Many studies use the Mahalanobis Distance (MD) method, which measures real-time "overall residual" (OR) to detect anomalies or outliers in a given data set [10, 11, 22]. Even with a small error value between the estimated NBM parameters and the real measurement value, this MD statistical approach has been proven to detect anomalies. [10, 11].

The kernel density estimate (KDE), which defines the MD distribution of healthy data, is used to determine the threshold value for anomalous conditions. [22, 35]. After an anomaly is detected, as failure early warning, some studies also analyzed the cause of the abnormality. Some studies use the total residual value of the model to detect abnormal conditions and use the individual variable model's residual value to diagnose the cause of the abnormality [11]. The approach proposed in this study aims to solve the limitation of the fix-threshold method proposed in earlier studies, which is less sensitive to detect abnormal conditions.

3. Proposed approach

3.1 Framework

The proposed method in this study aims to detect anomalies in equipment through an unsupervised learning method that employs the Normal Behavior Model (NBM) to predict normal equipment behavior. Fig. 1 and 2 respectively illustrate the offline and online stages of this study's equipment condition monitoring architecture.

3.1.1. Offline phase

The offline model development stages are carried out regarding reference condition equipment data as the training and validation datasets. A Normal Behavior Model of the equipment and a threshold value is generated from this offline phase will be used in the online condition monitoring phase.

The Normal Behavior Model of the equipment is a data-driven model that shows the relationship between sensor parameters under normal operating conditions. The model development process consists of dataset formation (training, validation, and testing), data pre-processing, machine learning algorithm selection, model training, model validation, and testing.

As shown in the Fig. 1, the NBM model development process begins with taking data from historians during the period of equipment normal conditions (normal state), which then the data is used for the autoencoder model training process. The offline phase's training will produce in a model architecture and a set of weights exported through the pipeline and used to forecast normal behavior in the offline and online phases.

The trained model is then utilized in this offline phase to forecast the healthy behavior of the training dataset, resulting in a distribution of residual values. The threshold value for the online condition

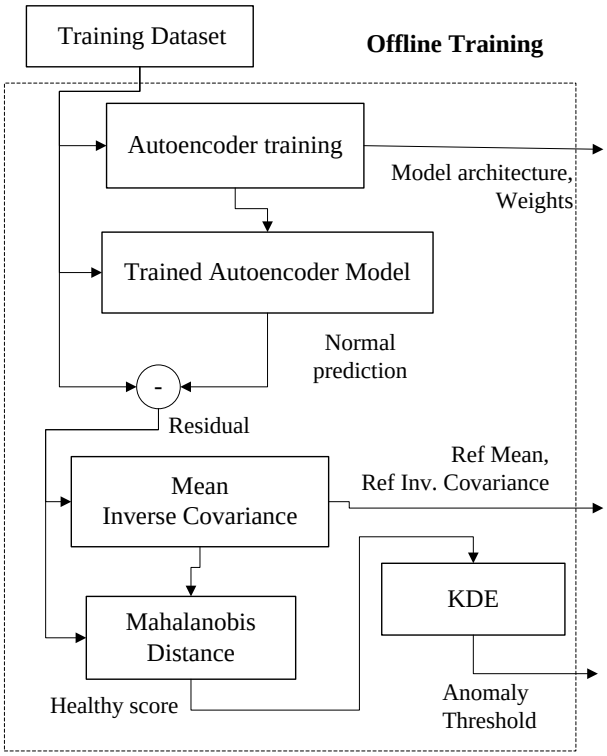

Figure. 1 Framework of the proposed approach in offline stages

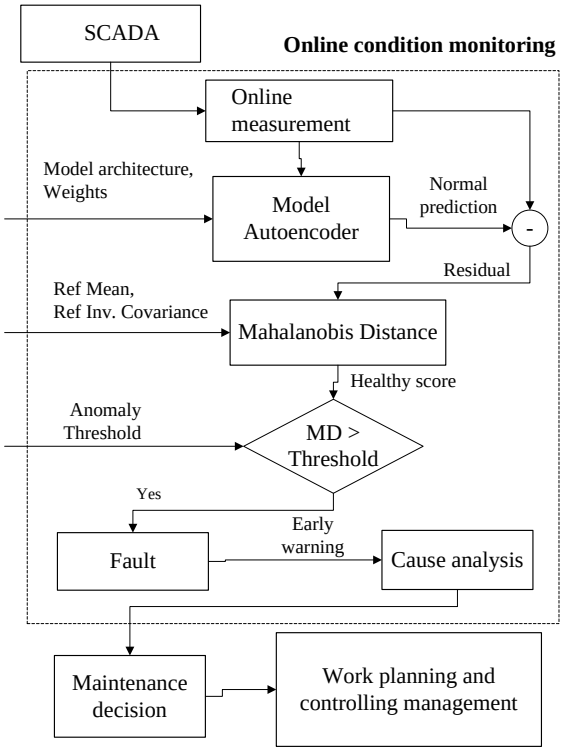

Figure. 2 Framework of the proposed approach in online condition monitoring stages

monitoring phase is established using the Mahalanobis Distance distribution from the reference dataset. Based on the 99th percentile of the cumulative distribution function (CDF) of the MD value of the training data, the threshold value is set at a specific value.

3.1.2. Online phase

The model that has been trained in the offline phase is then used in the anomaly detection system to predict the expected behavior of the equipment online based on the input of equipment parameters measured in real-time every specific period as shown in Fig. 2.

The prediction results of normal behavior generated by the NBM are then compared with the actual parameter measurements of the equipment so that prediction errors or residual values are obtained. The residual value is the discrepancy between the actual parameter measurement and the expected model outcomes.

Anomaly detection in this study was carried out by calculating the equipment health score using the Mahalanobis Distance method to express the multivariate distance between the residuals of a sample and the residual distribution of the reference data.

The calculation of the Mahalanobis Distance value requires the mean and inverse covariance values obtained from the reference dataset. Based on the threshold value set in the offline phase, whether the equipment is in a normal condition (healthy) or an anomaly can be determined. If the Mahalanobis Distance value exceeds the threshold value, then the equipment condition will be detected as an anomaly condition.

The occurrence of an anomaly event is an early warning of equipment failure. The anomaly detection results are then validated by an engineer or an expert and are used to decide the maintenance tasks that need to be carried out for the equipment. The maintenance tasks planning based on this anomaly prediction is proposed to be integrated with the existing Work Planning and Controlling system to determine scheduling, resource planning, material planning, and control the execution of maintenance work.

3.1.3. Autoencoder model

The model architecture proposed in this study is an autoencoder. The encoder section reduces the dimensionality of the data by converting the input into latent-space representation, which is the code with reduced dimensions. The decoder section uses the latent-space representation to reconstruct the input.

Since the autoencoder model has to reconstruct the compressed data, the model only learns to retain relevant information and ignore the noise. This way,

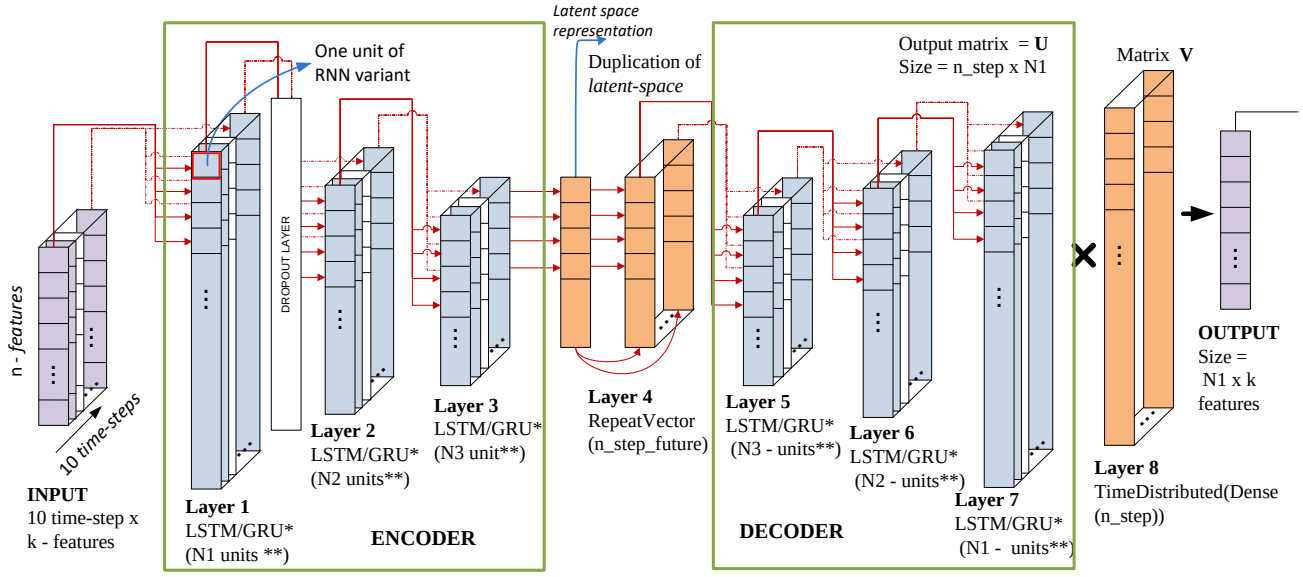

Figure. 3 The architecture design of RNN-Autoencoder (*One of LSTM/GRU is selected, **The number of units N1, N2, N3 varies to get the best model by hyperparameter optimization)

the autoencoder model will focus only on the essential features and eliminate meaningless information and noise. The autoencoder model must create a latent-space representation representing the original data features. In other words, the model must learn the data features and simplify their representations to make them easier to analyze.

3.1.4. Model architectural design

The proposed model was built with Keras [36] backend library from Tensorflow [37] version 2.5.0, with Python version 3.8.8 on Jupyter Notebook software.

The NBM model is designed with an autoencoder architecture, consisting of an encoder and a decoder section with details as shown in Fig. 3.

The encoder obtains its input from a dataset arranged in a multivariate time series in 3D shape (samples number x n time-steps x k features), as shown in Fig. 3. The three-layer of RNN sequence (with LSTM or GRU variants) is used in the encoder section, where the unit number, one of the hyperparameters (HP), is changed to get the best performance. The activation function in layers 1, 2, and 3 are obtained during the HP optimization search process to determine the parameters with the highest performance. The "return_sequence" parameter in layer-1 and layer-2 is enabled, which means each cell provides outputs passed to every cell in the next layer. While on layer 3, the “return_sequence" parameter is disabled, which means that only the last cell gives the output.

The decoder section employed 3-layers of RNN arranged sequentially. The "number of units" in layers 5-7 mirrors the "number of units" in the encoder section. Like the encoder section, the activation function for layers 5, 6, and 7 are also determined during the HP optimization search phase. The arrangement of the three layers gets input from the latent space, and then the compressed data must be parsed by the RNN array into the original form of the data. The output of the RNN in the last layer is forwarded to the *TimeDistributed* layer from Dense, which functions as a layer wrapper to form the decoder output into a 3D time-series format.

3.2 Anomaly detection

The Mahalanobis Distance was employed in this study to determine a healthy score for monitoring equipment conditions in CBM. Mahalanobis Distance is calculated based on the distance of the reconstruction error value of a point to be detected whether an anomaly occurs, with the distribution of the reconstruction error on standard equipment. In this case, the terms of reference normal equipment conditions are the training datasets.

Thus, if the sample of equipment data has the Mahalanobis Distance value outside the normal equipment distribution, it will be recognized as an outlier condition or anomaly.

In the offline phase of the research framework, the trained NBM model is used to predict normal behavior with the input of the training dataset as a reference. From the vector of the training dataset, $\boldsymbol{Y}_{ref}$, and the vector of model prediction, $\widehat{\boldsymbol{Y}}_{ref}$, represented in Eq. (1), the multivariate reconstruction error of the training dataset, $\mathbf{E}_{\text{ref}}$, can be computed.

$$\boldsymbol{E}_{ref} = \widehat{\boldsymbol{Y}}_{ref} - \boldsymbol{Y}_{ref} \quad (1)$$

The Mahalanobis Distance of Reconstruction Error of the Training Dataset, which is the reference healthy score value, can be stated in the Eq. (2).

$$MD_{ref} = \sqrt{\left(\boldsymbol{E}_{ref} - \boldsymbol{\mu}_{ref}\right)^{T} . {C_{ref}}^{-1} . \left(\boldsymbol{E}_{ref} - \boldsymbol{\mu}_{ref}\right)} \quad (2)$$

where $\boldsymbol{\mu}_{ref}$ is mean of the error signal reconstruction from training dataset and C_{ref} is covariance matrix of the training dataset reconstruction error.

The covariance of the variables X and Y is expressed by Eq. (3).

$$COV(X,Y) = \frac{\sum_{i=1}^{n}(X_i-\bar{x})(Y_i-\bar{y})}{n-1} \quad (3)$$

The mean and inverse covariance of the residual distribution of the training data will be utilized as reference data for anomaly detection in the online phase, where $\bar{x}$ and $\bar{y}$ are the averages of X and Y, respectively.

The Mahalanobis Distance is calculated using the residual of each sample, the means of error reconstruction of the reference dataset, and the inverse covariance of error reconstruction of the reference dataset. This parameter determines the equipment's health score.

Anomaly conditions are detected if the Mahalanobis Distance value exceeds the threshold. In this study, the kernel density estimation (KDE) method was used to simulate the Mahalanobis Distance training data's probability density function (PDF). The PDF of the Mahalanobis distance, $\hat{p}(x)$, can be expressed in Eq. (4).

$$\hat{p}(x) = \frac{1}{N\sigma}\sum_{i=1}^{N} K\left(\frac{x-x_i}{\sigma}\right) \quad (4)$$

x_i is the i-th sample of the Mahalanobis Distance value, N is the number of samples, and σ is the bandwidth parameter. The Kernel function K(.) can be expressed in Eq. (5).

$$K\left(\frac{x-x_i}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(-\frac{(x-x_i)^2}{2\sigma^2}\right)} \quad (5)$$

The anomaly condition threshold value, xd, in this study was determined based on the cumulative distribution function (CDF) of the Mahalanobis distance value by integrating the PDF function. The confidence level value, α is determined as the threshold value for the anomaly distribution, wherein in this study, the value 0.99 was chosen. The CDF value can be expressed in Eq. (6).

$$\alpha = P(x < x_d) = \int_0^{x_d} \hat{f}(x)dx \quad (6)$$

In the context of online equipment condition monitoring, the Mahalanobis Distance technique as the equipment's healthy score can merely detect the presence of an equipment anomaly without elucidating which feature is the most significant contributor to the anomalous condition. One purpose of this research was to determine the parameters that contributed to the source of the anomaly using a reconstructed error signal from each normalized individual variable. This normalization is necessary since the maximum scale of each variable varies. Eq. (7) illustrates how to calculate scaled RE, a normalized error signal.

$$scaled\ RE_i = \frac{(x_i-\hat{x}_i)}{std\ (x_i)} \quad (7)$$

Where x_i is the actual value of the i-th input variable, $\hat{x}_i$ is the reconstructed value of feature x_i, std (x_i) is calculated based on the standard deviation of the equipment's reference dataset when in healthy condition, namely training data, for each variable.

4. Experimental setup

4.1 Dataset

The data used for this study was obtained from the Tanjung Awar-Awar Steam Power Plant in Tuban Regency, East Java, Indonesia, which has 2x350 MW of capacity. These data include:

- Historical data from sensors installed on IDF and PAF equipment taken from the DCS control system,
- Historical data on equipment abnormalities from operation and maintenance records. After collecting all data, we picked equipment data in a healthy state with sufficient variety in operating conditions to obtain the relationship between parameters.

The dataset used for modeling is taken from a known period that the equipment is in a healthy state. At the first step, the equipment sensor's historical data was taken from the DCS control system historian database for one year (January 2021 – December 2021). The raw data has been taken is then resampled the average value every 1 minute to overcome the problem of noisy and spiked values on bad sensors. Then the healthy state and anomaly state period of equipment can be mapped from this data. Training

and validation data are selected from the equipment's healthy state period.

In this research, the variables extracted from historical equipment will be used in their entirety as a model feature, except for data recognized as anomalous and redundant sensors, to see how good the autoencoder model is at extracting essential features from massive data.

This research proposes a multivariate autoencoder time series model with 21 model features in 10 timesteps that would be used to predict the value of one future timestep of 21 target features which are the same as the input feature. This means that the data from the input sequences $\{\mathbf{x}_{(t-9)}, \mathbf{x}_{(t-8)}, ..., \mathbf{x}_{(t)}\}$ is utilized to forecast the value of the future sequence $\mathbf{x}_{(t+1)}$.

4.2 Data preprocessing

Several variables in the raw data are identical sensors because the sensor is redundant, so from this redundant sensor pair, only one sensor will be selected as a model feature to reduce the size of the model.

The variables used as model features were selected after raw data exploration. The sensor variables selected as the model features are those whose historical data do not indicate a flat value and are not in "bad quality" status due to sensor damage. Flat value data can be detected from the trending and minimum standard deviation values.

After removing various factors, the NBM model's features are obtained, notably as many as twenty variables for the IDF equipment model as shown at Table 1 and twenty-one variables for the PAF equipment model, shown at Table 2.

Correlation analysis was used to determine the correlation coefficient between the model feature variables using the Pearson correlation method. According to the correlation matrix between the variables presented in Fig. 4, there is a high correlation between the model feature variables for the most part, except for a few variables with just a correlation coefficient less than 0.3.

After obtaining raw data from DCS historians, it is pre-processed to fit the machine learning model that will be constructed. The initial data processing stage includes handling missing values (NaN), separating the dataset into training, validation, testing subsets, and scaling the dataset.

Data normalization is done to efficiently train the model because the data values are in a small range. The Scikit-Learn's MinMaxScaler is employed in this study to normalize the data, which means that the original data range was converted to a new range between 0 and 1 [38]. The training, validation, and testing datasets are all scaled using the same settings to ensure that the original value has not deviated from the scaling process mistake. Eq. (8) represents the output of data transformation using the MinMaxScaler, y_i, from the variable x_i.

$$y_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{8}$$

Following that, the shape dataset is converted to a three-dimensional format (batch size, time steps, and feature count) corresponding to the shape of the data required by the LSTM/GRU model. In the case of PA FAN 1A equipment, the training dataset had the shape (152812, 10, 21), whereas the validation dataset had the shape (21591, 10, 21). Meanwhile, a training dataset containing shapes (180945, 10, 20) and a shape dataset validation (67681, 10, 20) were employed for the ID Fan 2B equipment.

4.3 Model training

The model was trained using the healthy equipment dataset, then partitioned into training, validation, and testing datasets. The model is trained using training data. The training in this research was carried out using the Google Collabs platform.

Validation occurs after model training, aiming to determine the optimized model with the best performance. The validation data in this study is derived from a different part of the training dataset. The model testing procedure is used to determine how well the previously trained algorithm performs while predicting new data that has never been seen before. The purpose of employing dataset testing is to evaluate the trained model's generalizability [39].

In this study, the training dataset comprised 72.8 percent, while the validation dataset comprised 27.2 percent, with the validation set being distinct from the training data. The ratio of the training and validation datasets is chosen to approximate 70/30, which produces the best performance in some circumstances [40]. The optimizer "adam" and the loss function "mse" are used to train the model. The activation function used at each layer of the encoder and decoder is selected using the results of a HP search method that selects the optimal activation function from multiple options, including "ReLu", "Tanh", "Sigmoid", and "Softmax". This study chose the "sigmoid" function as the recurrent activation function.

Table 1. Variable as model features for ID fan 2B equipment

CT2E046304	Motor Current	RT2E030604	Mid Bearing Temp 1	RT2E046601	Inlet Temp
ZT2B021704	Fixed Blade Position	RT2E030602	Rlr Bearing Temp 1	RT2E046602	Outlet Temp
VT2E031306	Horizontal Vibration	RT2E032502	Motor Stator Ph-U Temp 1	DISP2ACTP	Generator Power
VT2E033305	Vertical Vibration	RT2E032503	Motor Stator Ph-V Temp 1	PT2B021101	Furnace Press
PT2B020507	Inlet Press	RT2E030504	Motor Stator Ph-W Temp 2	PT2E032604	Air Pre-Heater inlet Flue Gas Press
PT2B021202	Outlet Press	RT2E030601	Thrust Bearing Temp 1	PT2E033604	Air Pre-Heater outlet Flue Gas Press
RT2E030305	Motor Drive End Bearing Temp	RT2E032306	Motor N-Drive Ed Bearing Temp		

Table 2. Variable as model features for PA fan 1A equipment

ID	Description	ID	Description	ID	Description
DISP1ACTP	Generator Power	PT1B020201	Inlet Press.	RT1D012304	Motor Non-Drive End Bearing Temp.
RT1D013304	Hydraulic Coupler Bearing Oil Out Temp.	RT1D026401	Inlet Temp.	RT1D010604	Motor Stator Phase-U Temp. 1
RT1D011303	Hydraulic Coupler Oil Inlet Temp.	RT1D026402	Out Temp.	RT1D010605	Motor Stator Phase-V Temp. 1
PT1D011305	Hydraulic Coupler Oil Out Press.	PT1B020202	Outlet Press.	RT1D012705	Motor Stator Phase-W Temp. 2
RT1D013303	Hydraulic Coupler Oil Out Temp.	CT1D013306	Motor Current	RT1B023601	Primary Cool Air Temp. 1
ZT1B020307	Hydraulic Coupler Scope Tube Position	RT1D012305	Back 2# Temp.	VT1D011302	Horizontal Vibration
AI1B020606	Hydraulic Coupler Speed	RT1D010303	Motor DE Bearing Temp.	VT1D013302	Vertical Vibration

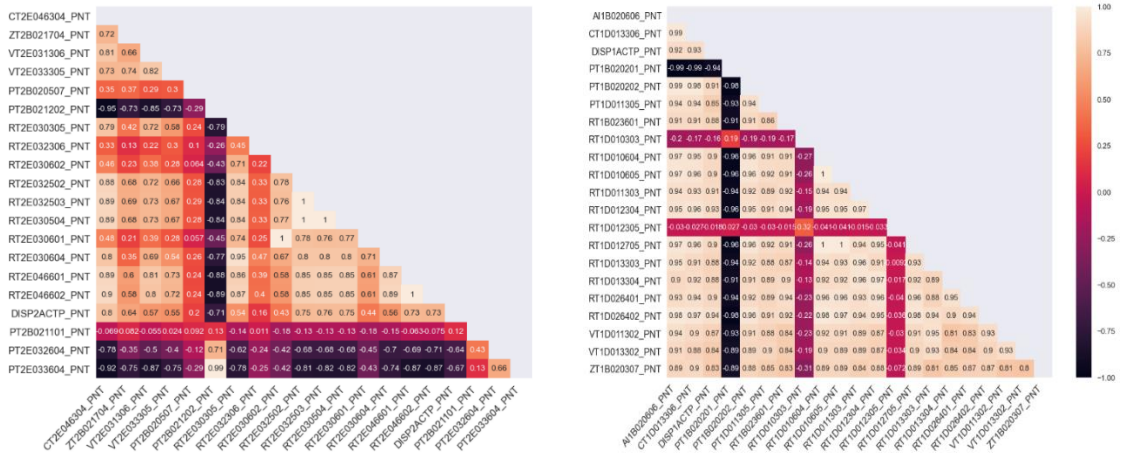

Figure. 4 Graph of heatmap correlation matrix between model feature variables for ID Fan 2B equipment on the left side and PA Fan 1A on the right side

5. Result and analysis

5.1 Hyperparameter search and trained model Performance

In this study, an experimental training model was established using a variety of hyperparameters (HP) combinations on the autoencoder model using RNN variations, namely LSTM and GRU. Then the performance of each model was analyzed to determine the best model.

The training process used several HP searches stages in this study. This research employs an

Table 3. Hyperparameter search space stage 1

Hyperparameter	Min value	Max value	Step
No. of unit N1	96	224	32
No. of unit N2	32	64	8
No. of unit N3	8	12	2
Best performance HP			
	LSTM	GRU	
Result of N1, N2, N3	160, 40, 10	160, 40, 10	
Val. accuracy	87.22%	91.37%	

Table 4. Hyperparameter search space stage 2

Hyperparameter	Choice	
Activation function	Relu, Tanh, Sigmoid	
Learning rate	0.1, 0.01, 0.001, 0.0001	
Best performance HP		
	LSTM	GRU
Activation function	Tanh	Tanh
Learning rate	0.001	0.001
Val. accuracy	92.4%	94.04%

automatic HP search method to accelerate the HP search process, which takes a long time when performed manually with many parameter variations. This study tuned the autoencoder model using the Keras Tuner library [41] and the Hyperband tuner [42] method. Keras Tuner optimizes HP selection automatically through a series of loop iterations that include HP combination selection, model training, and model evaluation. The HP optimization process begins with defining a set of HP combinations to be examined.

Hyperband determines the maximum number of tuning trials possible given the training resources available [42]. Hyperband is based on random configuration sampling from HP and iteratively develops one of the most promising HPs by avoiding inefficient time-consuming configuration optimization attempts or reaching convergence early. Hyperband will train the model with fewer epochs and then gradually train it with more epochs on some of the best models.

This research divides the HP search process into several stages to avoid a lengthy process if all HP combinations are tuned concurrently. The HP search process begins by determining the ideal parameter for the number of units in each layer (N1, N2, N3) on the encoder and decoder sections, as seen in Fig. 3. N1 specifies the number of RNN units in the encoder's first layer and the decoder's third layer. N2 denotes the encoder and decoder sections' second layer. In contrast, N3 represents the third layer of the encoder section and the first layer of the decoder section. The number of RNN units is designed symmetrically between the encoder and decoder sections. Table 3 summarizes the search space for stage 1 of HP optimization. The first stage employs a fixed learning rate of 10^{-2}, and a dropout layer with a rate of 0.1 is introduced after the first layer in the decoder section. "Tanh" is the activation function that is being employed. This process is carried out for each LSTM and GRU model architecture separately.

HP tuning stage 1 coincidentally produces the same optimal number of unit combination outputs between the GRU and LSTM models. The number of N1, N2, and N3 for the two models is 160, 40, and 10. The GRU model produces the best validation accuracy score of 91.37%, while the LSTM model produces the best validation accuracy score of 87.22%.

The second stage of the HP search process is to determine the optimal configuration of the learning rate and activation function. The search space for this step of HP search is shown in Table 4, which uses the most optimal number of unit parameters found during the first stage of HP tuning. At this level, the activation function's search space summary is "Relu", "Tanh", and "Sigmoid", while the learning rates are 0.1, 0.01, 0.001, and 0.0001. The most optimal combination of HP was obtained by this process, which used a learning rate of 0.001, and the activation function "Tanh", turned out to be the same for both LSTM and GRU models. The validation accuracy of the LSTM model is 92.4 percent, while the GRU model is 94.04 percent.

The third stage is a manual training method that utilizes the best HPs generated throughout the first two stages. Manual training is conducted with various batch sizes, including 128, 256, 384, 512, and 640. Training is carried out with a maximum of 1000 epochs by activating the checkpoint model to save the model that has improved the validation loss results at each epoch and activating the early stop to avoid the training model becoming overfit.

Table 5 summarizes the observations made during the training and validation processes of the performance model for each experiment, including the best performance epoch, training loss, validation loss, accuracy, validation accuracy, and training duration at the best performance epoch. Fig. 5 compares the MAE values obtained during the training and validation process for each model, which shows the training progress for every epoch. These findings indicate that the GRU's best model was obtained with a training batch size of 384 and a validation loss of 4.484x10-4 at the 65th epoch. The training loss was 2.187x10-4, and the validation accuracy was 93.47 percent at this epoch. The LSTM model performed optimally at the 92nd epoch, with a validation loss of 5,690x10-4, a training loss of

Table 5. The outcomes of model training by varying the model type and batch size

Architecture	GRU	GRU	GRU	GRU	GRU	LSTM	LSTM	LSTM	LSTM	LSTM
Batch size	128	256	384	528	640	128	256	384	528	640
Best performance epochs	60	51	65	88	62	99	92	61	133	219
Performance at best epochs										
Loss (x10^{-4})	2,028	2,178	2,187	2,178	2,299	1,926	2,062	2,632	2,227	2,028
Val Loss (x10^{-4})	4,602	4,633	4,484	4,708	4,635	5,954	5,690	7,195	7,492	5,881
Accuracy (%)	96,89	96,40	96,54	96,60	96,10	96,87	96,94	95,89	96,20	96,60
Validation Accuracy (%)	94,65	93,69	93,47	94,10	94,63	91,86	93,36	91,29	91,36	92,66
Training time (s)	4425,2	2582,3	2471,5	3156,2	1773,4	8893,3	5217,6	2984,8	6023,3	8755,4

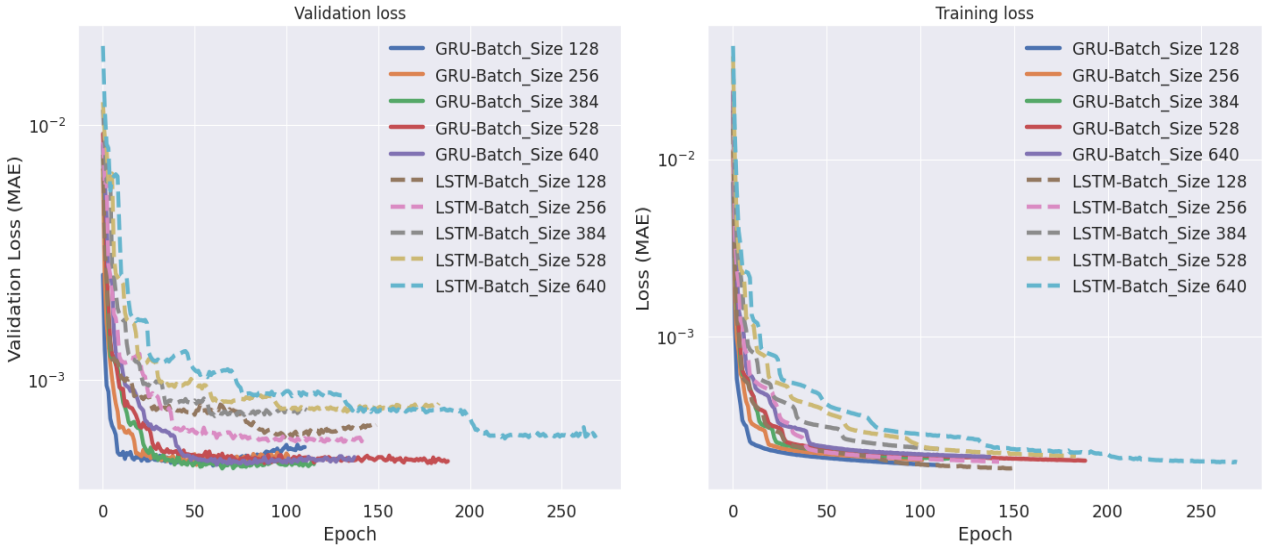

Figure. 5 Trends in the progress of multiple models' performance throughout training, training loss on the left side and validation loss on the right side

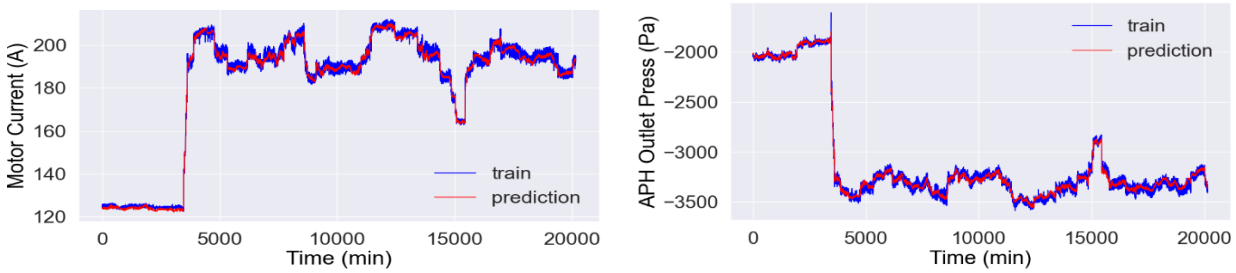

Figure. 6 The model's prediction results using the training dataset for Motor Current (CT2E046304) and Air Preheater flue gas outlet pressure (PT2E033604), respectively, on the left and right sides

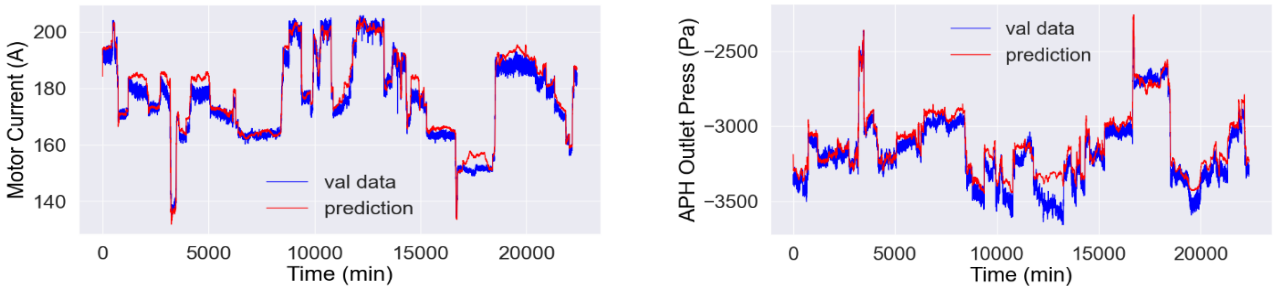

Figure. 7 The model's prediction results using the validation dataset for Motor Current (CT2E046304) and Air Preheater flue gas outlet pressure (PT2E033604), respectively, on the left and right sides

2,062x10-4, and a validation accuracy of 93.36 percent. The time required for training to reach the optimal performance epoch is considerably different for the two LSTM and GRU models, namely 5217

seconds and 2471 seconds, respectively. The time disparity between these two models is because their total parameters are significantly different, 220,240 for GRU and 290,900 for LSTM. The GRU model performs better with fewer epochs than LSTM. The same approach was utilized in this study to construct the model for PA Fan 1A equipment.

To illustrate how the model predicts the training dataset, Fig. 6 and 7 show several examples of variables whose prediction results are presented alongside their actual values to illustrate how the model predicts the training dataset. As shown in Fig. 6 and 7, the model does a good job of predicting the sequence of these properties.

5.2 Online anomaly detection

The best normal behavior model of the HP tuning and training process, as given in Table 5, will be used to forecast the equipment's condition in this study. The trained model is utilized to predict the training dataset, and then the MD health score is calculated as described in Eq. (2) to determine the threshold value. The threshold value of MD for the anomalous condition is determined numerically using the CDF with PDF integration of the MD distribution with the training dataset, as indicated in Eqs. (4) to (6). The confidence level is set to 99.9%. The MD threshold value is set to 25.0 in the case of ID Fan 2B equipment.

A simulation was conducted in this study utilizing a test dataset to evaluate the proposed anomaly detection approach's performance. The test dataset for this simulation is drawn from a period with several abnormal conditions, as determined by operational data, namely from July 1, 2021, to December 30, 2021. Then, see how anomaly detection performs in each case of equipment abnormality.

Based on the predictions of model $\hat{Y}$ and the actual measured parameter value Y with a test dataset to simulate actual data, the residual error value E is obtained as stated in Eq. (9).

$$\boldsymbol{E} = \hat{\boldsymbol{Y}} - \boldsymbol{Y} \tag{9}$$

The MD_k Health index value can be determined using Eq. (10), where μ_{ref} and C_{ref} are the mean and covariance parameters of the residual signal of the reference data extracted from the training data, respectively.

$$MD_k = \sqrt{\left(\boldsymbol{E}_k - \boldsymbol{\mu}_{ref}\right)^T . {C_{ref}}^{-1} . \left(\boldsymbol{E}_k - \boldsymbol{\mu}_{ref}\right)} \tag{10}$$

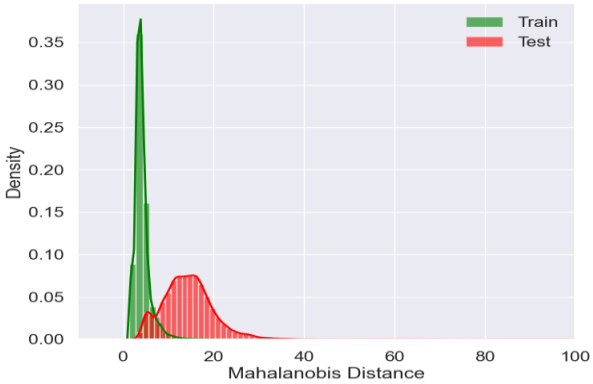

Figure. 8 Mahalanobis distance distribution comparison between training datasets under normal conditions and test datasets containing several anomalous conditions

The difference in the distribution of the MD parameter plots between the two datasets, namely training and validation as a reference for equipment in good health and a test dataset with multiple anomalous circumstances, is illustrated in Fig. 8. Fig. 8 demonstrates differences in the center and standard deviation of the MD distribution between healthy and anomalous equipment, allowing normal and anomalous conditions to be recognized in the test dataset by choosing a specified threshold value.

Numerous disturbance events occurred on the IDF 2B and PAF 1A equipment, resulting in downtime or equipment failure. The equipment events are utilized as simulations to evaluate the effectiveness of the proposed anomaly detection approach in this study.

5.2.1. Case study 1: ID fan 2B breakdown due to stall phenomena

The ID Fan 2B equipment exhibited abnormal behavior on several occasions due to the stall phenomena, namely a decrease of suction force from the fan. As a result of this incident, the boiler derated and could not run at full load. The operator will notice a significant decrease in the IDF motor's current and increased furnace pressure, indicating that the fan suction power has been drastically lowered.

On October 28, 2021, at 19.30, one of the breakdown occurrences on IDF 2B equipment induced by the stall phenomena happened. The proposed approach successfully recognized an anomaly during equipment failure, as indicated by a significant rise in the Mahalanobis Distance value used as a healthy score parameter for the equipment, as illustrated in Fig. 9, where a red dot indicates the observed anomaly event.

As shown in Fig. 9, before the equipment breakdown occurs, the proposed approach has

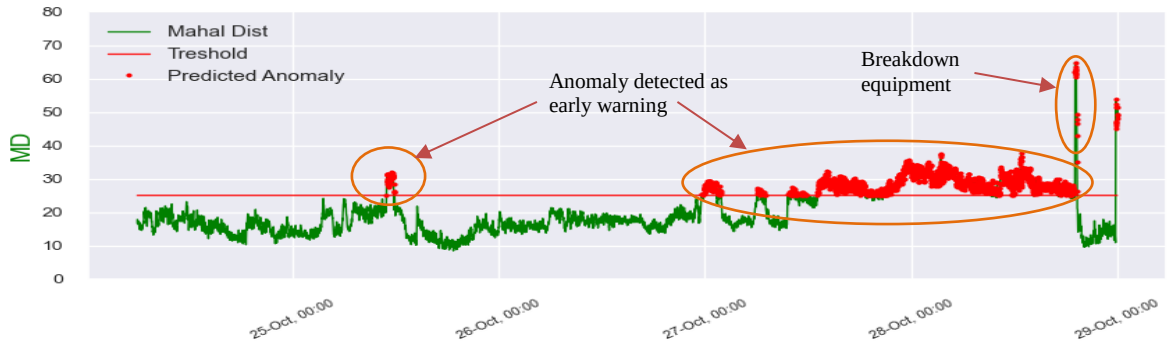

Figure. 9 Mahalanobis distance trend in IDF 2B stall phenomena occurred on Oct 28 at 19:30

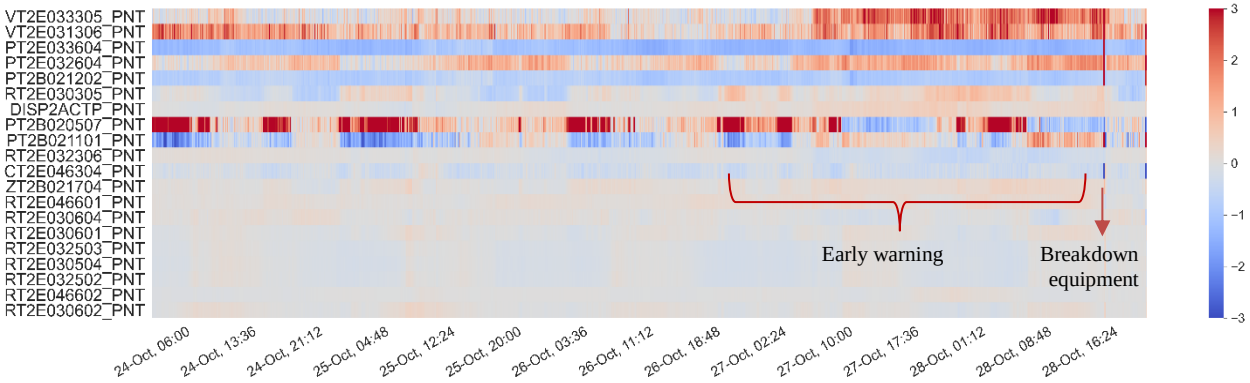

Figure. 10 Heatmap diagram of normalized reconstruction error in IDF 2B stall phenomena occurred on Oct 28 at 19:30.

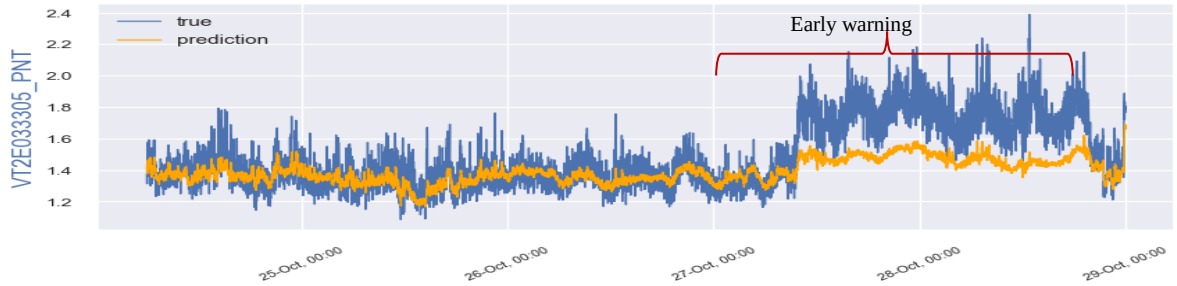

Figure. 11 The trend of variable's tag name VT2E033305 in IDF 2B stall phenomena occurred on Oct 28 at 19:30.

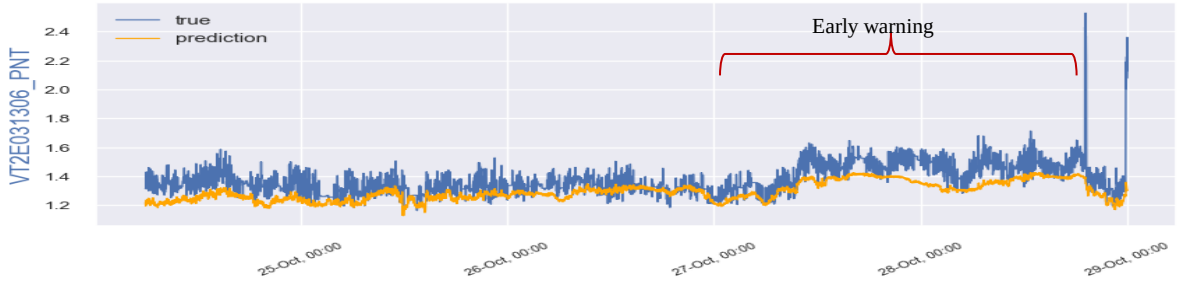

Figure. 12 The trend of variable's tag name VT2E031306 in IDF 2B stall phenomena occurred on Oct 28 at 19:30

succeeded in detecting the anomaly condition as an early warning three days before the breakdown event with an increasing frequency of anomaly conditions and an increasing healthy score amplitude. The increase in the healthy score value indicates a condition in which the equipment parameters have failed or have moved out of the normal distribution of the health equipment parameters.

This healthy score represents the combination of all features' distances to the distribution of equipment parameter values in a healthy state. It is unclear which parameters contribute to the abnormal situation while considering the healthy score. This study employs the method of calculating the reconstructed error signal separately for each feature variable to ascertain which parameters are responsible for the anomaly. The

reconstructed error signal is then normalized to the training dataset's standard deviation, allowing for the scaling of different ranges of each variable within the same range. Visualization of the error reconstruction signal from each variable as a "heatmap" graphic, as illustrated in Fig. 10, will aid in determining the root causes of the anomaly, as the normalized error signals for each variable are arranged from most significant to almost negligible contribution to anomalous condition.

This study computes the difference in reconstruction error values between the early warning anomaly period and the baseline period, which is a healthy period. After this computation, the reconstruction error difference value is sorted to show the most significant parameter contribution to the anomaly. Fig. 9 illustrates that the top variable has the most significant effect on the anomaly condition. The "cool" color scheme, which ranges from gray to blue, shows that the actual value of an equipment parameter is less than the healthy prediction. The "warm" hue ranges from gray to red, suggesting that the variable's actual value exceeds the projected usual behavior. Vertical vibration VT2E033305, horizontal vibration VT2E031306, and air pre-heater outlet flue gas pressure PT2E033604 contribute to the anomaly condition shown in Fig. 10 to 12 provide the trending comparison between the actual and expected healthy values of the variables generating the anomaly to illustrate how different their values and patterns are.

Several other stall events at IDF 2B between October and November 2021, including October 28, at 23:57, November 02, November 19, and November 20 followed a pattern identical to the stall event on October 28, 2021, 19:30 described above. Anomaly conditions can be recognized using a health score parameter that exceeds its threshold value at the moment of breakdown and multiple anomaly conditions that act as an early warning system for several periods prior to the breakdown. The presence of an anomaly that is discovered as an early warning sign before this equipment failure provides a chance to plan PDM work and execute an action plan that the operator can execute in a short-term, controllable period.

It is possible to deduce that there was an early warning of the ID Fan and APH outlet pressure vertical and horizontal vibration parameters prior to the breakdown event. This analysis examines the anomalous pattern in the IDF 2B stall phenomena on October 28, 2021, at 19:30, as depicted in Fig. 9 to 12. These vibration parameters increased beyond the predicted value for their typical behavior, whereas the APH outlet pressure value decreased below the expected value. There was also an increase in Furnace pressure parameters prior to the breakdown. The interpretation of these parameters suggests that the ID Fan 2B is operating at a higher load than its typical, increasing vibration. This failure occurs due to increased resistance or pressure drop on the IDF suction line, as indicated by the APH outlet pressure value parameter.

Figure. 13 Finding of plugging in air preheater element during overhaul inspection

Numerous defects were discovered during the overhaul inspection from December 2021 to January 2022, including a substantial blockage of the APH element due to fly ash deposition. The IDF 2B blade and the IDF 2B lock static blade were eroded. As illustrated in Fig. 13, the accumulation of fly ash in the APH causes clogs in the ducting inlet line.

The IDF must work extra, increase vibration, and change other parameters that eventually break down due to loss of suction power (stall). The findings of this overhaul validate the prediction of anomalies detected by the approach proposed in this study.

5.2.2. Case study 2: PA fan 1A breakdown due to misalignment

On May 13, 2021, and May 30, 2021, PA Fan 1A encountered an alignment issue. The malfunction resulted in downtime due to the need to repair the PA Fan 1A shutdown. The effectiveness of the anomaly detection has been evaluated using the same method as mentioned in Section 5.2.1 for IDF 2B equipment.

This misalignment problem is indicated by an increase in the vibration parameters of the axial motor and the axial of variable speed coupling to values of 6.7 mm/s and 5.74 mm/s, respectively. Most vibration problems as parameters that contribute to the misaligned state are only monitored via portable vibration measurements and not in DCS. Therefore, they are not included in the NBM feature. Although the root cause of the primary problem is not included in the model features, anomalous conditions were

recognized 105 minutes before the equipment downtime on May 13, 2021, due to changes in operating patterns prior to the power plant unit's shutdown. As illustrated in Fig. 14, the Mahalanobis Distance gradually increases until it approaches the threshold value, indicating the presence of a significant distance with the distribution of healthy equipment. In comparison, the Megawatt Generator is the parameter that makes the most significant contribution to the anomaly's cause.

5.2.3. Case study 3: PA fan 1A breakdown due to hydraulic pressure low-low trip

On May 30, 2021, at 07.42, PA Fan 1A also experienced a hydraulic pressure Low-Low Trip issue. As an early warning, the anomaly condition is detected 12 hours before equipment downtime, as evidenced by an increase in the MD value above the threshold, as illustrated in Fig. 15. To determine the

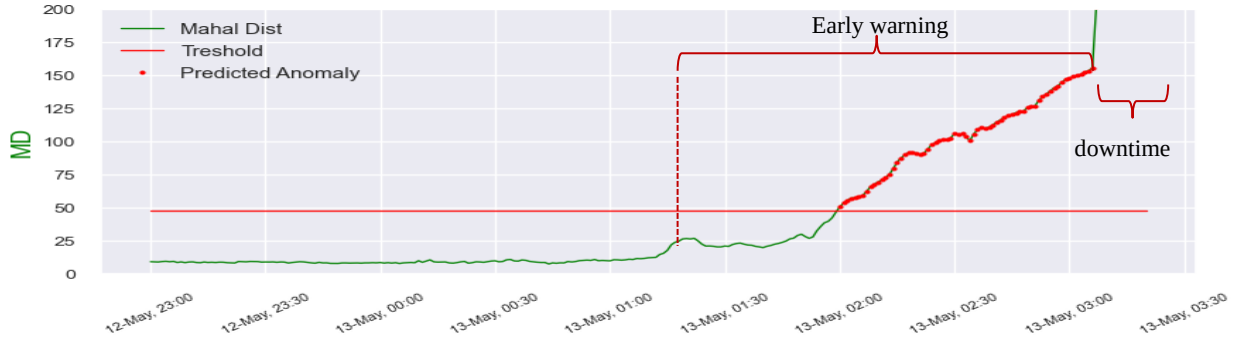

Figure. 14 The Mahalanobis distance trending of PA fan 1A during a misalignment disturbance on May 13, 03:06

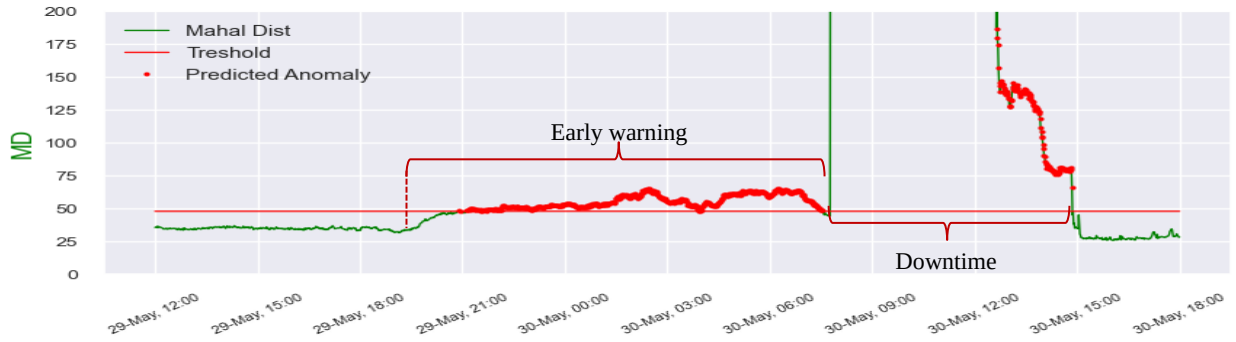

Figure. 15 The Mahalanobis distance trending of PA fan 1A during hydraulic pressure low-low on May 30, 07:42

Table 6. Performance comparison of the proposed anomaly detector model with other models

Model	Dataset	TN	FP	FN	TP	Accuracy	Precision	Recall	F1 Score	Weighted F1 score β=2
RNN-AE	Test	265148	5307	4333	3182	0,965	0,375	0,423	0,398	0,407
	Train	180938	7	0	0	1,000	0,000			
K-Mean	Test	133663	1565	3471	287	0,964	0,155	0,076	0,102	0,090
	Train	94150	41078	3757	0	0,677	0,000	0,000		
OCSVM	Test	177754	92701	1649	5866	0,661	0,060	0,781	0,111	0,165
	Train	180582	363	0	0	0,998	0,000			
RF	Test	101177	34051	308	3450	0,753	0,092	0,918	0,167	0,244
	Train	135227	0	0	3757	1,000	1,000	1,000	1,000	1,000
SVM	Test	85751	49477	349	3409	0,642	0,064	0,907	0,120	0,181
	Train	135227	0	0	3757	1,000	1,000	1,000	1,000	1,000
KNN	Test	86650	48578	366	3392	0,648	0,065	0,903	0,122	0,182
	Train	135227	0	0	3757	1,000	1,000	1,000	1,000	1,000
LR	Test	76238	58990	339	3419	0,573	0,055	0,910	0,103	0,157
	Train	135227	0	0	3757	1,000	1,000	1,000	1,000	1,000
MLP	Test	66937	68291	337	3421	0,506	0,048	0,910	0,091	0,139
	Train	135227	0	0	3757	1,000	1,000	1,000	1,000	1,000

root cause of the anomaly, the same approach as described in Section 5.2.1 is used. The degree of the change in the magnitude of the individual normalized reconstruction error signals from the baseline condition to the anomalous condition ranked from biggest to smallest value. The variable with the most significant increase in normalized error reconstruction value corresponds to the signal with the most significant contribution. The horizontal vibration of VT1D011302 and the temperature of the motor DE RT1D010303 are determined to be the root cause signals that contribute the most to the anomaly condition using this method. It can be interpreted that there has been a decrease in lubricating pressure, increasing the value of horizontal vibration, and an increase in the DE bearing temperature, which was detected 12 hours before the breakdown occurred.

5.3 Performance comparison of proposed method

To verify the effectiveness of the proposed method approach, this study validated the proposed model by comparing its performance to other models. Each model's ability to detect anomalous conditions as an early warning before or during downtime is compared. Other models used for comparison include the K-Means unsupervised learning method, the one-class support vector machine (OCSVM), and the supervised learning classifier methods such as random forest (RF), support vector machine (SVM), k-nearest neighbors (KNN), logistic regression (LR), and multi-layer perceptron (MLP).

The comparison model is given the same number of features and dataset size as the proposed model to evaluate the effectiveness of each model in handling a big imbalanced dataset with numerous features. The type of supervised learning classifier model was evaluated using a dataset that was stratifying and non-randomly split into a train and test dataset with a percentage of 50 percent:50 percent. The non-random data sharing option is implemented so that there is no data leakage between the training dataset and the testing dataset.

Table 6 displays the comparison of each model's performance outcomes. The higher accuracy, precision, recall, and F1-score parameter values of the anomaly detection model show its superior performance. The F1-score of 0.398 for the RNN-AE model presented in this work is greater than that of other comparable models utilizing the test dataset, as shown in Table 6. Due to the imbalance of the dataset, the F1 score is a more important metric than accuracy and recall. By assuming the cost difference between false positives (FP) and false negatives (FN), we additionally construct the parameter weighted F1-score with a value of $\beta = 2$, indicating that a false negative condition produces a loss of twice as many false positives. The RNN AE approach has the highest weighted F1-score value of all comparable models, 0.407.

Table 6 demonstrates that the unsupervised K-Mean model is insufficient to handle a large number of model characteristics and huge data dimensions. The same thing happened to the OCSVM model and the supervised learning models revealed overfitting, so that it functioned nearly perfectly only on the training dataset, however, the performance in classifying the testing dataset was insufficient, as indicated by the high FP value.

6. Conclusion and future works

This study proposes a framework for using an unsupervised learning approach with the RNN-autoencoder method in a predictive maintenance (PDM) system applied to power plants. The normal behavior model is used to detect anomalous conditions in this study. It is combined with variants of the RNN model, particularly LSTM and GRU, to construct an autoencoder network arrangement. The condition monitoring phase compares the measurement signal for the actual equipment parameters to the predicted model, which is interpreted as the equipment's normal behavior, calculates the reconstruction error, and then evaluates the distance between the two using the health score parameter. The health score parameter is the multivariate distance between online data and the distribution of healthy equipment, calculated using the Mahalanobis Distance method in this study. A threshold value based on the 99th percentile of the Mahalanobis distance distribution of healthy equipment can be used to determine the equipment's anomalous condition.

During the model development phase, the performance model employing GRU and the LSTM with HP optimization are compared to determine the most optimal parameters. According to the experimental results on a real dataset of power plant equipment, the LSTM model with the best HPs achieved a validation loss of 5,690 x 10-4 and a validation accuracy of 93.36 percent, whereas the GRU model with the best HPs achieved a validation accuracy of 4.484 x 10-4 and a validation accuracy of 93.47%. GRU can achieve better performance than the LSTM model, although only slightly different in performance.

This proposed framework was demonstrated in a case study of IDF and PAF equipment disturbances at a power plant. It can detect anomalous conditions

such as the IDF stall phenomenon, changes in the operating pattern of the PA Fan prior to unit shutdown, and hydraulic pressure Low-Low problems on the PA Fan. Anomaly situations observed include those that occur during equipment downtime and precursor to equipment downtime or failure. A rise in the healthy score value until the threshold value indicates the early warning pattern. However, the pattern and duration of the early warning are case-specific. The detection of an anomaly initiates condition-based maintenance tasks or provides operational recommendations for conducting operator controllable maneuvers.

Additional research can be conducted in the future to refine the framework provided in this study for renewable energy plants.

Conflicts of Interest (Mandatory)

The authors declare no conflict of interest.

Author Contributions (Mandatory)

Conceptualization, Budi Santoso, and Mauridhi Hery Purnomo; methodology, Budi Santoso; software, Budi Santoso; validation, Wiwik Anggraeni, and Henry Pariaman; formal analysis, Budi Santoso, Wiwik Anggraeni, Mauridhi Hery Purnomo; data curation, Henry Pariaman; writing—original draft preparation, Budi Santoso; writing—review and editing, Budi Santoso, Wiwik Anggraeni, Henry Pariaman, Mauridhi Hery Purnomo; visualization, Budi Santoso; supervision, Mauridhi Hery Purnomo;

All authors read and approved the final manuscript.

Acknowledgments

We would like to express our gratitude to PT Pembangkitan Jawa Bali for giving scholarships and funding for master's education and research and for providing access to the data used in this study.

References

[1] P. PJB, “Intelligence Center of Optimization Reliability and Efficiency”, 2020. https://www.ptpjb.com/icore/ (accessed Jun. 03, 2020).

[2] F. V. Scheidt, H. Medinová, N. Ludwig, B. Richter, P. Staudt, and C. Weinhardt, “Data analytics in the electricity sector - A quantitative and qualitative literature review”, *Energy AI*, Vol. 1, p. 100009, 2020, doi: 10.1016/j.egyai.2020.100009.

[3] M. Song and L. Jia, “Big data mining method of thermal power based on spark and optimization guidance”, In: *Proc. of 2018 IEEE 7th Data Driven Control Learn. Syst. Conf.*, pp. 514-520, 2018, doi: 10.1109/DDCLS.2018.8516098.

[4] P. Zhang, Y. Xu, F. Luo, and Z. Y. Dong, “Power Big Data: New Assets of Electric Power Utilities”, *J. Energy Eng.*, Vol. 145, No. 3, p. 04019009, 2019, doi: 10.1061/(asce)ey.1943-7897.0000604.

[5] B. Liu, Z. Fu, P. Wang, L. Liu, M. Gao, and J. Liu, “Big-data-mining-based improved K-means algorithm for energy use analysis of coal-fired power plant units: A case study”, *Entropy*, Vol. 20, No. 9, 2018, doi: 10.3390/e20090702.

[6] M. Moleda and D. Mrozek, “Big data in power generation”, *Springer International Publishing*, Vol. 1018, 2019.

[7] D. Hu, G. Chen, T. Yang, C. Zhang, Z. Wang, Q. Chen, and B. Li, “An Artificial Neural Network Model for Monitoring Real-Time Parameters and Detecting Early Warnings in Induced Draft Fan”, *Volume 3: Manufacturing Equipment and Systems*, Vol. 3, No. May 2019, 2018, doi: 10.1115/MSEC2018-6370.

[8] T. Ergen and S. S. Kozat, “Unsupervised anomaly detection with LSTM neural networks”, *IEEE Trans. Neural Networks Learn. Syst.*, Vol. 31, No. 8, pp. 3127-3141, 2020, doi: 10.1109/TNNLS.2019.2935975.

[9] H. A. Hejazi and H. M. Rad, “Power systems big data analytics: An assessment of paradigm shift barriers and prospects”, *Energy Reports*, Vol. 4, pp. 91-100, 2018, doi: 10.1016/j.egyr.2017.11.002.

[10] P. Bangalore and L. B. Tjernberg, “An Artificial Neural Network Approach for Early Fault Detection of Gearbox Bearings”, *IEEE Trans. Smart Grid*, Vol. 6, No. 2, pp. 980-987, 2015, doi: 10.1109/TSG.2014.2386305.

[11] D. Hu, C. Zhang, T. Yang, and G. Chen, “Anomaly Detection of Power Plant Equipment Using Long Short-Term Memory Based Autoencoder Neural Network”, *Sensors (Basel)*, Vol. 20, No. 21, p. 6164, 2020, doi: 10.3390/s20216164.

[12] M. C. Garcia, M. A. Sanz-Bobi, and J. del Pico, “SIMAP: Intelligent System for Predictive Maintenance Application to the health condition monitoring of a windturbine gearbox”, *Computers in Industry*, Vol. 57, No. 6, pp. 552–568, 2006, doi: 10.1016/j.compind.2006.02.011.

[13] M. K. Wisyaldin, G. M. Luciana, and H. Pariaman, “Using LSTM network to predict circulating water pump bearing condition on coal fired power plant”, In: *Proc. of 2nd Int.*

Conf. Technol. Policy Electr. Power Energy, Vol. 3, pp. 54-59, 2020, doi: 10.1109/ICT-PEP50916.2020.9249905.

[14] R. Laubscher, "Time-series forecasting of coal-fired power plant reheater metal temperatures using encoder-decoder recurrent neural networks", *Energy*, Vol. 189, p. 116187, 2019, doi: 10.1016/j.energy.2019.116187.

[15] H. Pariaman, G. M. Luciana, M. K. Wisyaldin, and M. Hisjam, "Anomaly Detection Using LSTM-Autoencoder to Predict Coal Pulverizer Condition on Coal-Fired Power Plant", *Evergreen*, Vol. 8, No. 1, pp. 89-97, 2021, doi: 10.5109/4372264.

[16] M. K. Wisyaldin, G. M. Luciana, and H. Pariaman, "Pendekatan Long Short-Term Memory untuk Memprediksi Kondisi Motor 10 kV pada PLTU Batubara", *Jurnal Kajian Ilmu dan Teknologi*, Vol. 9, No. 2, pp. 311–318, 2020, [Online]. Available: https://stt-pln.e-journal.id/kilat/issue/view/76.

[17] A. Ajami and M. Daneshvar, "Data driven approach for fault detection and diagnosis of turbine in thermal power plant using Independent Component Analysis (ICA)", *Int. J. Electr. Power Energy Syst.*, Vol. 43, No. 1, pp. 728-735, 2012, doi: 10.1016/j.ijepes.2012.06.022.

[18] J. Mulongo, M. Atemkeng, T. A. Narh, G. M. Nguegnang, and M. A. Garuti, "Anomaly Detection in Power Generation Plants Using Machine Learning and Neural Networks Anomaly Detection in Power Generation Plants Using Machine Learning and Neural Networks", *Appl. Artif. Intell.*, Vol. 00, no. 00, pp. 1-16, 2019, doi: 10.1080/08839514.2019.1691839.

[19] M. Yuan, Y. Wu, and L. Lin, "Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network", In: *Proc. of 2016 IEEE International Conference on Aircraft Utility Systems*, pp. 135-140, 2016, doi: 10.1109/AUS.2016.7748035.

[20] A. Nanduri and L. Sherry, "Anomaly detection in aircraft data using Recurrent Neural Networks (RNN)", In: *Proc. of ICNS 2016: Securing an Integrated CNS System to Meet Future Challenges*, pp. 1-8, 2016, doi: 10.1109/ICNSURV.2016.7486356.

[21] K. Lu, S. Gao, W. Sun, Z. Jiang, X. Meng, Y. Zhai, Y. Han, and M. Sun, "Auto-encoder based fault early warning model for primary fan of power plant", In: *Proc. of IOP Conference Series: Earth and Environmental Science*, Vol. 358, No. 4, 2019, doi: 10.1088/1755-1315/358/4/042060.

[22] Y. Li, F. Hong, L. Tian, J. Liu, and J. Chen, "Early warning of critical blockage in coal mills based on stacked denoising autoencoders", *IEEE Access*, Vol. 8, pp. 176101-176111, 2020, doi: 10.1109/ACCESS.2020.3026918.

[23] C. M. A. Roelofs, M. A. Lutz, S. Faulstich, and S. Vogt, "Autoencoder-based anomaly root cause analysis for wind turbines", *Energy AI*, Vol. 4, p. 100065, 2021, doi: 10.1016/j.egyai.2021.100065.

[24] C. Mortensen, T. Nolte, and C. Mortensen, "Fault Detection in Mobile Robotics Using Autoencoder and Mahalanobis Distance", *Malardalen University, Vasteras, Sweden*, 2021.

[25] K. Zope, K. Singh, S. H. Nistala, A. Basak, P. Rathore, and V. Runkana, "Anomaly Detection and Diagnosis In Manufacturing Systems", *Annu. Conf. PHM Soc.*, Vol. 11, No. 1, pp. 1-10, 2019, doi: 10.36001/phmconf.2019.v11i1.815.

[26] M. Roy, S. K. Bose, B. Kar, P. K. Gopalakrishnan, and A. Basu, "A Stacked Autoencoder Neural Network Based Automated Feature Extraction Method for Anomaly Detection in Online Condition Monitoring", *IEEE Symposium Series on Computational Intelligence*, pp. 1501-1507, 2018.

[27] A. Gensler, J. Henze, B. Sick, and N. Raabe, "Deep Learning for solar power forecasting - An approach using AutoEncoder and LSTM Neural Networks", In: *Proc. of 2016 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 2858-2865, 2017, doi: 10.1109/SMC.2016.7844673.

[28] J. Zhang, X. Jiang, X. Chen, X. Li, D. Guo, and L. Cui, "Wind power generation prediction based on LSTM", In: *Proc. of ACM International Conference Proceeding Series*, pp. 85-89, 2019, doi: 10.1145/3325730.3325735.

[29] J. D. Kelleher, B. Mac Namee, and A. D'Arcy, *Fundamentals of Machine Learning for Predictive Data Analytics, second edition: Algorithms, Worked Examples, and Case Studies*, Second ed., The MIT Press, 2020.

[30] G. Yang, Y. Wang, and X. Li, "Prediction of the NO emissions from thermal power plant using long-short term memory neural network", *Energy*, Vol. 192, No. x, p. 116597, 2020, doi: 10.1016/j.energy.2019.116597.

[31] R. Zhao, R. Yan, J. Wang, and K. Mao, "Learning to monitor machine health with conVolutional Bi-directional LSTM networks", *Sensors (Switzerland)*, Vol. 17, No. 2, p. 273, 2017, doi: 10.3390/s17020273.

[32] A. Singh, "Anomaly Detection for Temporal Data using Long Short-Term Memory (LSTM)",

KTH Royal Institute of Technology, 2017.
[33] D. P. Guillen, N. Anderson, C. Krome, R. Boza, L. M. Griffel, J. Zouabe, and A. Y. Rashdan "A RELAP5-3D/LSTM model for the analysis of drywell cooling fan failure", *Prog. Nucl. Energy*, Vol. 130, No. October 2019, p. 103540, 2020, doi: 10.1016/j.pnucene.2020.103540.
[34] P. B. Dao, W. J. Staszewski, T. Barszcz, and T. Uhl, "Condition monitoring and fault detection in wind turbines based on cointegration analysis of SCADA data", *Renew. Energy*, Vol. 116, pp. 107-122, 2018, doi: 10.1016/j.renene.2017.06.089.
[35] X. Liang, F. Duan, I. Bennett, and D. Mba, "A sparse autoencoder-based unsupervised scheme for pump fault detection and isolation", *Appl. Sci.*, Vol. 10, No. 19, 2020, doi: 10.3390/app10196789.
[36] F. Chollet and others, "Keras.", *GitHub*, 2015, [Online]. Available: https://github.com/fchollet/keras.
[37] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, and J. Dean, "Tensorflow: A system for large-scale machine learning", In: *Proc. of 12th Symposium on Operating Systems Design and Implementation*, pp. 265-283, 2016.
[38] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, and O. Grisel, "Scikit-learn: Machine Learning in Python", *J. Mach. Learn. Res.*, Vol. 12, pp. 2825-2830, 2011.
[39] E. Alpaydin, "Introduction to machine learning.", *MIT Press*, 2020.
[40] Q. H. Nguyen, H. B. Ly, L. S. Ho, N. A. Ansari, H. V. Le, V. Q. Tran, I. Prakash, and B. I. Pham, "Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil", *Math. Probl. Eng.*, Vol. 2021, p. 4832864, 2021, doi: 10.1155/2021/4832864.
[41] T. O'Malley, E. Bursztein, J. Long, F. Chollet, H. Jin and L. Invernizzi, *Keras Tuner*, 2019. [Online]. Available: https://github.com/keras-team/keras-tuner.
[42] L. Li, K. Jamieson, G. Desalvo, A. Rostamizadeh, and A. Talwalkar, "Hyperband: A novel bandit-based approach to hyperparameter optimization", *J. Mach. Learn. Res.*, Vol. 18, pp. 1-52, 2018.