

Article

Not peer-reviewed version

---

# Extruder Machine Gear Fault Detection Using Autoencoder LSTM via Sensor Fusion Approach

---

Joon-Hyuk Lee , [Chibuzo Nwabufo Okwuosa](#) , [Jang-Wook Hur](#) \*

Posted Date: 9 October 2023

doi: [10.20944/preprints202310.0467.v1](https://doi.org/10.20944/preprints202310.0467.v1)

Keywords: Anomaly detection; autoencoder; long short-term memory; deep learning; discrete wavelet transform, feature extraction, outlier detection.)



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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.

Article

# Extruder Machine Gear Fault Detection Using Autoencoder LSTM via Sensor Fusion Approach

Joon-Hyuk Lee, Chibuzo Nwabufo Okwuosa and Jang-Wook Hur \*

Department of Mechanical Engineering (Department of Aeronautics, Mechanical and Electronic Convergence Engineering), Kumoh National Institute of Technology, 61 Daehak-ro, Gumi-si 39177, Gyeongsang-buk-do, Korea; okwuosachibuzo333@gmail.com

\* Correspondence: hhjw88@kumoh.ac.kr

**Abstract:** In industrial settings, gears play a crucial role by assisting various machinery functions such as speed control, torque manipulation, and altering motion direction. The malfunction or failure of these gear components can have serious repercussions, resulting in production halts and financial losses. As a result, there is an increasing requirement to monitor the state of these components in order to avoid such issues from occurring. To address this need, research efforts have focused on early defect detection in gears in order to reduce the impact of possible failures. This study focused on analyzing vibration and thermal datasets from two extruder machine gearboxes using an autoencoder Long Short-Term Memory (LSTM) model. The major goal is to implement an outlier detection approach to detect and classify defects. The results of this study highlighted the extraordinary performance of the Autoencoder LSTM model, which achieved an impressive accuracy rate of 94.42% in recognizing malfunctioning gearboxes within the extruder machine system. Furthermore, the study used a thorough global metrics evaluation methodology to further test the model's dependability and efficacy, consequently substantiating the proposed approach's validity.

**Keywords:** anomaly detection; autoencoder; long short-term memory; deep learning; discrete wavelet transform; feature extraction; outlier detection

---

## 1. Introduction

The recent industrial revolution has had a significant impact on the production and manufacturing sectors, with more advancements on the way. Many industry revolutions have come and gone in the world of research and academia with the aim of increasing and modifying the relationship between man and machine, which is accomplished through the intelligent integration of the Internet of Things (IoT) and cyber-physical systems [1–3]. The current phase of the industrial revolution, industry 4.0, has made technology available in a way that allows for seamless and smart decision-making, which has generally improved efficiency and revenue in these sectors. One of the most important aspects of Industry 4.0 is its application in the prognostics and health management (PHM) of equipment, which entails detecting anomalies as well as predicting when these systems may fail, thereby providing a healthy and conducive environment for production. Most of the time, this is accomplished through the use of various types of sensors to collect rich data from these machines in collaboration with a machine learning algorithm that sequentially trains to understand the nature of the data and then aids in the detection and early prediction of any future fault occurrences using advanced technology [4,5]. Most of the means associated with useful data collection from machines for adequate health monitoring include, but not limited to, vibration sensors, acoustic emission sensors, thermal sensors, current sensors, and so on; the selection of the appropriate sensor to be used on equipment is solely dependent on the nature of the machine, the type of environment, and the researcher's expertise. The next industrial revolution, industry 5.0, is intended and projected to be more sophisticated and advanced in the sense that machines not only assist humans but also collaborate with humans in solving technical problems [6].



The plastic/fabric extruder machine is a critical piece of equipment in the plastic manufacturing industry. Its working principle entails collecting plastic raw materials through its hopper, which is then directed to the extruder screw, whose pressure and rotation are provided by the gear and an electric motor combined with the machine's heater pad to melt the plastic raw materials as well as transport them to the chamber where the molten plastic would be used to achieve whatever purpose it has been assigned in terms of shape. Nonetheless, the machine's effectiveness is dependent on the flawless performance of its components. Among these, the extruder screw stands out as a critical component of the plastic extruder machine. As a result, ensuring the continuing functionality of this specific component becomes critical. The reason for this monitoring is that any fault in the extruder screw could have serious consequences for the manufacturing process. The induction motor provides the required rotary power and torque, while the gear as shown in Figure 1 reduces the rotational speeds of the extruder screw, resulting in more torque for crushing and transporting raw materials. There has been a lot of research on fault diagnostics and prediction in induction motors [8–10], but the focus of this research is on the gearbox, whose failure would be catastrophic in the sense that the operational movement and rotation of the extruder screw, which has been designed to provide the required amount of torque and pressure for the manufacturing process, would be greatly affected. In our study, the gear component of the plastic extruder gearbox is made up of helical gears, which are known for their high contact ratio and thus provide high torque [7].

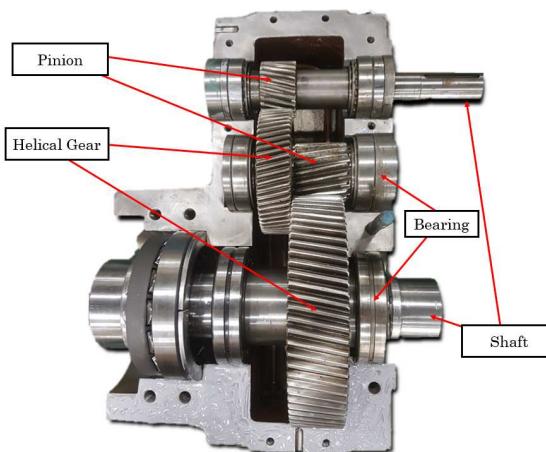


Figure 1. Top view of the extrude gearbox.

While vibrations are frequently anticipated to come from the running gears, vibration sensors are among the most popular sensors for condition-based gearbox monitoring. Unfortunately, unlike spur gears, helical gears are known to produce less noisy and non-stationary vibration signals due to their contact ratio difference; however, noise and non-stationary signal generation are common with gears, which are regarded as challenges for most signal processing methods [11,12]. Nonetheless, scientists have devised solutions to this problem, such as fusing vibration data with other sensor data like sound and thermal data, or de-noising and decomposing vibration data signals to extract the important spectral properties of the signal [13,14]. Most known classes of helical gear failures often cause unusual friction between the meshing and/or mating gear components, which generates heat, making thermal data useful for fault analysis in helical gearbox fault detection and isolation (FDI) [15].

Generally, time-frequency signal transformations are preferred especially in helical gearbox vibration signals because they present the signal in such a way that useful information can be easily extracted or detected in a signal, which aids in fault diagnosis. Nonetheless, the auto-encoder LSTM is well recognized for its capacity to identify faults or function in multivariate time-domain signals. As a result, the LSTM-auto-encoder is an excellent tool for processing and training a fusion of thermal and vibration datasets for proper FDI [16,17]. Machine learning (ML) algorithms have long been used by scientists owing to their efficiency and adaptability to small data sets, as well as their high diagnostic

and prognostic accuracy, low computational cost, and ease of implementation. However, due to some of its well-known issues, including its propensity for over-fitting, poor performance with complex datasets, and high parameter dependence, Artificial Neural Network (ANN)-based algorithms, such as Feed-forward neural networks (FNNNs), Long Short-Term Memories (LSTMs), Deep neural networks (DNNs), Deep belief networks (DBNs), Recurrent neural networks (RNNs), Convolutional neural networks (CNNs), etc., have presented the ideal sophisticated diagnostic and prognostic tool, despite their unique challenges such as high computational cost and interpretability issues; however, their uniqueness and robustness in PHM is no match for traditional ML algorithms.

## 2. Literature Review and Related Works

As previously stated in the preceding section, gearboxes play a critical role in industrial domains, particularly in contexts involving torque transfer, speed reduction, and motion dynamics modification, among other activities. As a result, the consequences of gearbox failure resonate far and wide. Any failure, regardless of the underlying reason, has the potential to produce unneeded downtime. The resulting operational stop reduces productivity by hindering industrial processes. Furthermore, the resulting output shortage directly leads to a revenue loss. The interdependence of gearbox functionality and industrial processes emphasizes the importance of preventive maintenance and constant monitoring to avoid potential failures. Mitigating the danger of gearbox failure by such methods not only ensures operational continuity but also protects against the tangible economic ramifications of stopped output and financial losses. As a result, grasping the basic properties of faults is a fundamental prerequisite in the pursuit of effective fault mitigation within mechanical systems. This includes determining their frequency, patterns of occurrence, and severity. This early understanding serves as the foundation for developing meticulous tactics for correcting these flaws as effectively as possible. A significant insight comes when considering the gearbox in a plastic extruder machine, where the prevalent failures are directly linked with gear-related concerns. While helical gears are more resistant to failure than spur gears, failure is an unavoidable possibility.

Numerous failures have been extensively researched and documented in the academic world. These include broken teeth, fissures, the occurrence of pitting corrosion, uniform wear, axis alignment inconsistencies, fatigue-induced difficulties, instances of impact induced fractures, and the likes. Surprisingly, amid this spectrum of failures, those linked to fatigue phenomena have emerged as the most common [18–21]. Notably, the prevalence of fatigue-related failures broadens its impact, serving as a critical precursor for additional severe and catastrophic defects within the system [21]. Tooth bending fatigue and surface contact fatigue are the two main types of fatigue failure, which are typically linked to issues with gear assembly, misalignment, unintentional stress concentration, and unsuitable material choice or heat treatment [21,22]. Gear tooth wear is a similarly common form of failure to fatigue in terms of prominence. This failure mechanism involves the loss of gear material and frequently results from many triggers that include mechanical, electrical, and chemical effects [22]. Fundamentally speaking, abrasive and adhesive wear are distinguish modes of tooth wear failures. Adhesive wear is characterized by material transfer between teeth, which leads to propensities for ripping and welding, as opposed to abrasive wear, which includes material removal as a result of inter-tooth contact [22,23]. Scuffing is a key failure mode that is frequently ignored in gear analysis. This occurrence results from sliding motions interacting with lubricated contacts, which generate high temperatures. These elevated temperatures can consequently cause the surface film that coats the gears to deteriorate, leading to deformations and eventually the melting of the relatively softer gear components [22–25].

The accuracy of relying exclusively on vibration signals for precise defect identification may be compromised by the elevated levels of noise and temperature that frequently accompany malfunctioning gear conditions. As a result, many researchers have implemented techniques to improve diagnostic precision. These methods often entail either applying de-noising techniques to separate important signals from the noise-contaminated vibration data originating from gear

components or combining vibration signals with other sensor outputs to create comprehensive diagnostic models. The incorporation of vibration and acoustic sensor data helped the development of a thorough diagnostic model, as demonstrated by the researchers in this specific study [26]. Their method involved the independent extraction of statistical features from each sensor. Relevant attributes were identified using a cutting-edge feature selection method. In the end, a comprehensive diagnostic model specifically designed to solve chipped gear defects was developed by synergistically combining the chosen features from both sensors. In a different study [27], the author skillfully combined current and vibration sensors operating over a range of frequencies to create a condition-based monitoring framework for spotting gear wear issues. The study's conclusions emphasized not only the attainment of desired results but also a calculated approach for reducing the computing demands generally connected with data fusion. This was accomplished by carefully assessing the dataset to only include the most pertinent qualities, and then strategically incorporating statistical and heuristic feature engineering techniques.

Additionally, Zhang Y. and Baxter K. proposed a cross-domain fault diagnostic framework by synergistically combining vibration and torque information from a gearbox in a different exploration [28]. Their ground-breaking approach addressed a common issue that arises when utilizing different statistics from diverse sensors. To counteract this, they used a fusion strategy in which the various sensor datasets were combined into a single 1-D sample array. Then, as a crucial element of their cross-domain fault diagnostic approach, a CNN-based classifier was used. This innovative method made it possible to integrate several sensor outputs, improving the system's capacity for diagnostics. Several researchers have made considerable advances in refining sensor fusion approaches, as demonstrated by the approach used in this study [29]. To build a diagnostic model, the author used a trio of sensors—a vibration accelerometer, a microphone, and sound emission sensors—across a variety of operational circumstances. Their process entailed extracting wavelet features from each sensor's data stream, followed by identifying relevant features. This technique resulted in a powerful model that validated their intended aim. A similar three-sensor fusion technique was discovered in another study involving the prediction of the remaining usable life (RUL) of a hydraulic gear pump in the presence of variable pollution levels [30]. The researchers used a Kalman Filter-based linear model to smoothly fuse fault features from three distinct sensor data streams—vibration, flow rate, and pressure signals—in this case. These fused properties were then used as input for a Bidirectional Long Short-Term Memory (BI-LSTM) network, resulting in the creation of a strong RUL architecture.

Due to the inherent characteristics and different origins of sensor data commonly used in sensor fusion, these datasets often contain intrusive background noise, lack stable patterns across time, and depart from a normal Gaussian distribution. As described in this specific research study [31,32], these variables collectively restrict the extraction of important information from the data. Consequently, it is necessary to use supplementary signal processing techniques to present these datasets in a way that allows for effective information extraction. In the context of our investigation, the vibration datasets acquired from machinery necessitate undergoing a de-noising process. This procedure is critical for extracting relevant information from vibration data. The effectiveness of this process is dependent on the robustness of the signal processing algorithms used and the expertise of the analyst. Numerous methods for denoising and decomposing signals have been introduced, including discrete wavelet transform (DWT), Bayesian filter-based methods, and empirical mode decomposition, the latter of which is based on the Hilbert Huang transform (HHT) [33,34]. Among these techniques, discrete wavelet transform and Bayesian-filter-based algorithms are well-known for their effectiveness and robustness. However, when it comes to performance, empirical evidence has shown that discrete wavelet transform is a better option for both signal de-noising and decomposition tasks [34]. This insight acted as a catalyst for its preferential use in our ongoing inquiry.

One of the primary goals of this research is to properly combine vibration sensor data with thermal sensor data to build a reliable Prognostics and Health Management (PHM) scheme. To achieve this integration, appropriate fusion techniques must be used to develop strong health indicators

(HIs) for an efficient diagnostic model [35–37]. It is critical to note that, while the requirement for a fusion algorithm is undeniable, the technique to be used is significantly dependent on the specific challenge at hand. Local Linear Embedding (LLE), for example, can be sensitive to the choice of nearest neighbors, whereas Principal Component Analysis (PCA) may encounter difficulties when dealing with datasets having a normal distribution. Independent Component Analysis (ICA), on the other hand, is dependable when dealing with non-Gaussian input distributions, particularly when these inputs display statistical independence, as demonstrated in previous studies [39,40]. In a related study [39], the authors conducted a thorough comparison of Independent Component Analysis (ICA) and Autoencoder (AE) approaches. The goal of this study was to synchronize data collection from numerous IJTAG-compatible Embedded Instruments (EIs) and build a machine learning-based system-level model for forecasting the end of life (EOL) in safety-critical systems that use multiple on-chip embedded instruments. According to the findings of the study, the ICA and EI fusion strategy excelled in capturing latent variables for model training, hence improving the EOL prognostic power. In addition, J. Weidong introduced the FastICA compound neural network, an original ICA-based network that makes use of feature extraction from multi-channel vibration measurements [41]. This method shows how ICA has the potential to be used as a strong feature extraction tool for challenging sensor data fusion problems.

As a result, the techniques outlined across the spectrum of reviewed research highlight a common theme: the inherent limits of relying simply on vibration signals for diagnosing gear-related difficulties. This collaborative knowledge acts as a catalyst, propelling us to incorporate a unique methodology into our model. Our method combines vibration and thermal sensor data from a plastic extruder machine's gearbox. While it has been recognized that malfunctioning gearboxes frequently generate heat due to irregular gear meshing, little study has been conducted to harness thermal signals for comprehensive defect investigation which most recorded studies often focus on thermal imagining rather than thermal data signals. This undertaking is a unique step, resulting from the inspiration obtained from the combination of earlier study findings. Therefore, with all these findings in view, the contributions of this sensor fusion plastic extruder gearbox outlier detection fault-based model are highlighted as follows:

- A Discrete Wavelet Decomposition for enhanced Vibration Signal Analysis in Plastic Extruder Gearbox Fault Diagnosis: By Incorporating a Discrete wavelet decomposition strategy, we aim to extract invaluable insights from the vibration signals entrenched in noise. This technique seeks to bolster the efficacy of diagnosing faults within the plastic extruder gearboxes.
- An effective statistical time-frequency domain feature extraction and correlation Filter-Based Selection technique: We presented an effective method for extracting features in the time-frequency domain. We further introduced a correlation filter-based feature selection process, which is commonly used in feature engineering, to highlight salient and vital information, hence enhancing the model's overall performance.
- A multi-Sensor Fusion Using the FastICA Technique: Our strategy includes a multi-sensor fusion paradigm aided by the fast Independent Component Analysis (fastICA) technique. The proposed technique harmoniously blends selected information from multiple separate sensor datasets. This fusion not only condenses data to a single-dimensional array but also preserves the unique characteristics of each source.
- An LSTM-Autoencoder Outlier Detection Using a Fused Multi-Sensor Dataset. We achieved an outlier detection by leveraging an LSTM-Autoencoder, which is enabled by a fusion of multi-sensor data techniques. This comprehensive methodology results in a strong framework ready for defect detection in the context of a plastic extruder gearbox.
- A Framework Validation and Global Evaluation Metrics Proposed. We provide a set of global evaluation indicators to validate our suggested approach. These evaluations highlight the framework's efficiency and efficacy, demonstrating its ability to manage the complexities of defect detection within plastic extruder gears.

The rest of the paper is structured thus: Section 3 and section 4 covers the theoretical background and the outlier detection model of the paper. Section 5 breaks down the data collection and methodological concept of our study. Conclusively, section 6 and section 7 summarizes experimental results and conclusion of the study respectively.

### 3. Theoretical Background

This section explains the essential principles underlying the key elements that constitute the foundation of our research. These parts include DWT for de-noising and signal decomposition, FastICA for feature dimension reduction, and the proposed Autoencoder-LSTM fault detector.

#### 3.1. DWT De-Noising/Decomposition

The wavelet transform is a signal analysis mathematical tool. Through a succession of wavelets, it decomposes signals into multiple frequency components at different scales, capturing both time and frequency information. This enables localized signal feature analysis, which is important for tasks like de-noising, compression, and feature extraction. These series are produced by orthogonal functions and indicate a square-integrable function, whether real or complex-valued [31,32]. Just like DFT and STFT which are often used in situations where the fast Fourier transform falls short in performance, the wavelet transform as highlighted earlier is a time-frequency signal process tool that is a unique and efficient tool that can present a signal in an orthogonal or non-orthogonal format using basic a function known as wavelet [32,38]. Generally, the essential difference is in the decomposition approach: the Fourier transform divides a signal into its sinusoidal components, whereas the wavelet transform employs localized functions (wavelets) that exist in both real and Fourier space. Because of this localization in both domains, the wavelet transform can provide more intuitive and interpretable information about a signal. Wavelet transform, as opposed to Fourier transform which focuses on the frequency of a signal in most cases, incorporates both time and frequency characteristics, allowing for a more dense study of signals with localized features.

As a mathematical tool, the general equation of a wavelet transform is presented thus:

$$WT(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi(t)^* \left( \frac{t-b}{a} \right) dt \quad (1)$$

where  $a$  and  $\sqrt{a}$  represent the scale parameter and the normalization factor for energy conservation, which regulates the dilation of the wavelet function of the transform,  $b$  represents the translation parameter across the time axis. The mother wavelet is represented by  $\psi(t)$  while the  $\psi(t)^*$  represents the complex conjugate of the presented mother wavelet.

In academia, the two most prevalent wavelet transforms are discrete wavelet transform and continuous wavelet transform. Their main distinction is the function used in their computation. For example, in the creation of a DWT, an orthogonal wavelet is frequently used, whereas CWT adapts a non-orthogonal wavelet. Because of the nature of the signal retrieved from the extruder gearbox, which is embedded with noise, we concentrated our research on DWT. DWT is well-known for its usage in signal de-noising and decomposition into distinct levels. The Discrete Wavelet Transform (DWT) transforms a signal into approximation and detail coefficients at different scales. The approximation coefficients indicate the low-frequency content of the signal, whereas the detail coefficients represent high-frequency features. This iterative technique gives a multiscale examination of the signal, allowing for efficient representation, compression, and signal processing. The coefficients can be used for signal reconstruction and additional analysis.

The general equation for obtaining the wavelet transform is shown in Equation [].

$$W(k, m) = \sum_{n=-\infty}^{\infty} x[n] \cdot \frac{1}{\sqrt{2^k}} \psi \left( \frac{n-m}{2^k} \right) \quad (2)$$

where  $k$  represents the scale or level of decomposition,  $m$  represents the translation or position in each decomposed level.  $X[n]$  is the discrete-time signal being transformed.  $\psi[n]$  represent discrete wavelet function.

The performance of a wavelet is solely based on the wavelet function (mother wavelet). Therefore, it is important to note that the wavelet function's specific form differs depending on the wavelet family (e.g., Haar, Daubechies, Morlet, etc.). The aforementioned formulas represent the wavelet transform's conceptual structure, while the actual computation includes evaluating the integral or sum over the proper ranges.

### 3.2. FastICA for Dimension Reduction

Primarily, Independent Component Analysis (ICA) was created to solve the problem of blind source separation in image and audio processing. Its major goal was to extract from observed signals a set of statistically independent components. FastICA was created in response to the potential of ICA for dimensionality reduction, specifically for feature fusion [39]. In many circumstances, the mutual information among numerous aspects is buried by high-order statistical characteristics, and FastICA is successful at minimizing high-order correlations while maintaining mutual independence among these features. FastICA is thus a useful tool for reducing dimensionality by merging characteristics while keeping their independence [39,40,42].

FastICA is a signal decomposition algorithm that divides observed signals into statistically independent components. It assumes the signals are a mixture of unknown sources and attempts to estimate the original sources by maximizing their independence. The procedure begins by centering the signals and then whitening them to remove correlations and equalize variances. To quantify the divergence from Gaussianity in the altered signals, a measure of non-Gaussianity, such as negentropy, is used [38,42]. FastICA maximizes this metric iteratively by updating the weights of linear combinations of the observed signals. After obtaining the independent components, dimensionality reduction can be accomplished by picking a selection of components that capture the most relevant information or contribute the most to the original signals. The dimensionality of the data is efficiently decreased by removing less relevant components. The reconstructed signals can then be obtained by projecting the independent components back. For our study, we choose FastICA because in fault detection scenarios the more discriminant the data the better it is for the training model to easily adapt and classify and/or detect the presence of abnormality in a set of data.

### 3.3. Correlation Coefficients

Correlation coefficients are statistical measurements that assess the degree and direction of a relationship between two variables. The Pearson correlation coefficient, Spearman rank-order correlation coefficient, and Kendall rank correlation coefficient are three regularly used correlation measurements. The Pearson correlation coefficient evaluates the linear relationship between variables. It is calculated by dividing the covariance of the variables by the product of their standard deviations. Pearson correlation coefficients vary from -1 to 1. A value of -1 indicates a strong negative linear association, 0 shows no linear relationship, and 1 suggests a strong positive linear relationship. It is commonly symbolized by the symbol (rho). The Spearman rank-order correlation coefficient, on the other hand, is a non-parametric statistic that assesses the strength of a monotonic relationship between variables. It is based on the data ranks rather than the actual data values. Its range, like the Pearson correlation coefficient, is from -1 to 1, with -1 indicating a strong negative monotonic association, 0 suggesting no monotonic link, and 1 indicating a strong positive monotonic relationship. The Kendall rank correlation coefficient is another non-parametric statistic that assesses the strength of the monotonic association between variables. It takes into account the number of concordant and discordant pairs in the data.

Thus, in our study and the majority of studies involving linear variables, the Pearson correlation coefficient is frequently selected above alternative correlation coefficients. The other two types,

however, operate more effectively than the Pearson correlation in situations involving non-linear variables. The correlation coefficient has generally been used successfully in academia for feature reduction, selection, diagnostics, prognosis, and other tasks. The Pearson coefficient was used in this study to extract meaningful and discriminant features, which is essential for effective problem diagnosis and fault detection [8,9,38,43].

$$\rho_p = \frac{n (\sum xy) - (\sum x) (\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2] [n \sum y^2 - (\sum y)^2]}} \quad (3)$$

$$r_s = 1 - \frac{6 \sum d_i^2}{n (n^2 - 1)} \quad (4)$$

$$t_b = \frac{P - Q}{\sqrt{(P + Q + X_0) (P + Q + Y_0)}} \quad (5)$$

### 3.4. Autoencoder

Autoencoders are a form of artificial neural network that is used to learn input data representations. They are made up of three basic parts: an encoder, a bottleneck layer, and a decoder. In the bottleneck layer, the encoder maps the input data to a compressed representation. The bottleneck layer acts as a bottleneck for information flow, lowering the input's dimensionality. The latent space representation is the learned representation in the bottleneck layer. The decoder attempts to recover the original input data using the latent space representation. The autoencoder's purpose is to reduce the reconstruction error, which is the difference between the input data and the reconstructed output.

By defining the problem as a supervised learning task, autoencoders can be trained with the aid of unlabeled data. The goal is to produce an output that closely resembles the original input. This is accomplished by reducing the reconstruction error, for instance  $(x, \bar{x})$ , where  $x$  is the initial input sequence and  $\bar{x}$  is the resultant reconstruction sequence. The autoencoder learns to extract relevant features from input data and build a compressed representation in the latent space by iteratively modifying the network's parameters. As a result, autoencoders can be used for tasks like dimensionality reduction, data de-noising, and anomaly detection [44–47].

### 3.5. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks were created to get around regular RNNs' limitations when processing lengthy sequences. To capture and hold long-term dependencies in sequential data, they contain memory cells and gating mechanisms. A memory cell used by LSTMs serves as a conveyor belt for information as it moves through the sequence. Long-term memories are stored in the cell state, and what should be discarded is decided by the forget gate. The output gate regulates the output depending on the cell state, whereas the input gate controls fresh information that is added to the cell state. Because they have the capacity to learn and spread pertinent information over lengthy sequences, LSTMs excel at jobs involving sequential data.

The mathematical expression for the LSTM architectural structure is defined with the following equation:

$$i_t = \sigma (W_i * [h_{t-1}, x_t]) + b_i \quad (6)$$

$$f_t = \sigma (W_f * [h_{t-1}, x_t]) + b_f \quad (7)$$

$$O_t = \sigma (W_o * [h_{t-1}, x_t]) + b_o \quad (8)$$

$$C_t = \tanh (W_c * [h_{t-1}, x_t]) + b_c \quad (9)$$

$$C_t = f_t * c_{t-1} + i_t * \tilde{c}_t \quad (10)$$

$$\sigma_t = \sigma (W_o * [h_{t-1}, x_t]) + b_o \quad (11)$$

$$h_t = o_t * \tanh (c_t) \quad (12)$$

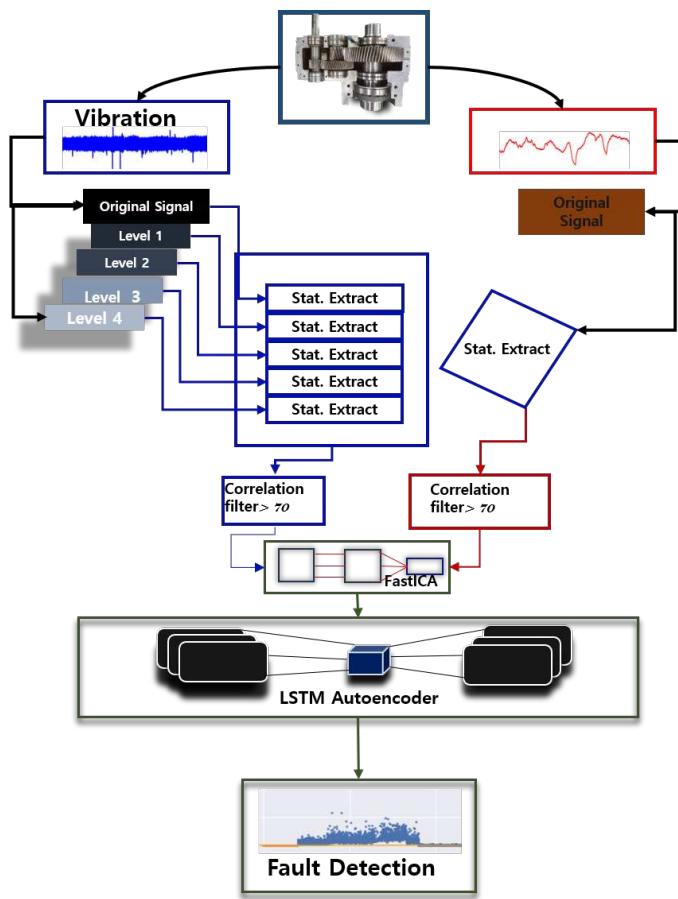
where  $i, f, O$  represents the input, forget, and output gates.  $x_t$  describes the current input to the LSTM architectural structure,  $C_t, c_{t-1}, h_t, h_{t-1}$  represents the cell state, previous cell state, the hidden cell state, and the previous hidden cell state respectively.  $\sigma, W, b$  represents the the sigmoid function, weight and bias of each gate [48–51].

For a more insightful explanation of the structure of the LSTM; LSTMs employ gates that permit selective information memory and forgetting, allowing them to update the cell state based on the current input and past state. The input gate applies an activation function to the input and previous hidden state (such as sigmoid, ReLU, or softmax), yielding values between 0 and 1. These values are then multiplied element by element-wise with the input, with their importance scaled accordingly. The forget gate generates values between 0 and 1 by applying a sigmoid function to the input and prior concealed state. These values are then multiplied element by element with the prior cell state, with the previous values scaled according to their importance. Values between 0 and 1 are produced by the output gate after applying a sigmoid function to the input and prior concealed state. The output of applying a hyperbolic tangent function to the current cell state is then multiplied element-wise by these values to produce the LSTM's final output. A vector of values that is updated at each time step makes up the cell state of LSTMs. Utilizing the current input, the prior cell state, and the prior concealed state, the cell state is updated. Following that, the hidden state, which is utilized to make predictions, is updated using the revised cell state [52–54].

#### 4. The Proposed Outlier Detection Model

In our study, we used an Autoencoder LSTM deep learning approach to create an anomaly detection model. Anomaly detection entails recognizing patterns that differ clearly from the usual pattern in a given dataset. Anormality detection seeks to distinguish uncommon datasets, known as anomaly datasets, from normal datasets. Many strategies have been developed in academia to detect anomalies [9,43,47,57,58], such as Statistical methods, machine learning algorithms and data visualization approaches, supervised, semi-supervised, and unsupervised learning approaches, outlier detection, clustering technique, and so on are some of the commonly used techniques employed for anomaly detection, where presented models learn the normal patterns or structures from the data without explicitly labeled anomalies. Once trained, the models can detect outlier from learned usual behavior and highlight them as potential abnormalities.

Figure 2 displays the anomaly detection model employed in our study; we employed an outlier detection methodology with the aid of an Autoencoder LSTM deep learning approach. The model basically comprises of five (5) major steps which are summarized below.



**Figure 2.** Proposed LSTM Autoencoder Anomaly Detection Model.

- Data acquisition: Both vibration and thermal data were collected in order to construct an appropriate model for monitoring extruder gear performance. The incorporation of several data sources is prompted by the fact that vibration signals obtained from gearboxes are prone to noise contamination, making it difficult to extract valuable insights on their own. A more comprehensive and useful picture can be built by adding additional data, such as temperature measurements. Vibration data is critical for detecting anomalies or inconsistencies in the operation of the gear. However, because of the existence of noise, it is frequently impossible to distinguish important patterns or trends purely from vibration signals. This is when the extra thermal data comes into play. By combining vibration and thermal data, it is possible to identify hidden links and correlations between the performance of the gear and the accompanying temperature fluctuations. The use of both vibration and thermal data seeks to improve the accuracy and usability of the model built to monitor the extruder gear. This method allows for a more comprehensive study, allowing for the detection of potential problems such as high friction, overheating, or abnormal operating circumstances. Finally, by combining multiple data sources, a more robust and efficient model may be constructed, providing useful insights for optimizing extruder gear performance, maintenance, and dependability.
- Signal processing and feature extraction: The second key aspect of the model is revolves around signal processing, with the aim of extracting valuable information from gearbox vibration data while minimizing the inherent noise. The Discrete Wavelet Transform (DWT) was used as a method for deconstructing, filtering, and preprocessing the vibration signals to achieve this. The DWT extracted time-frequency statistical information from both the original signal and each vibration signal decomposition level. A full analysis of the vibration data was performed by performing decomposition at various levels, collecting variances across different scales and frequencies. Thermal data, on the other hand, as a time-varying signal, did not go through

decomposition. Instead, from the raw temperature data, time statistical features were extracted. The goal of this method was to capture the temporal patterns and trends revealed by temperature readings. The study aims to improve the quality and usability of the information gained by applying the DWT to vibration signal processing and extracting time statistical features from temperature data. This methodology allowed us to identify key trends, correlations, and anomalies in the vibration and temperature data, allowing us to gain a more comprehensive understanding of the extruder gear's behavior and performance.

- Feature selection: To obtain an effective diagnosis in the setting of anomaly detection, discriminant traits are required. A correlation filter technique was used to guarantee that the features extracted had enough discriminative power. This technique ensures that only features with a correlation percentage of 70% or above are deemed closely connected. By removing characteristics that do not match this correlation threshold, the resulting feature set is tailored to include informative and discriminating features, improving the accuracy and effectiveness of the diagnosis process.
- Signal Fusion: The integration of data from numerous sources while keeping their different characteristics is a critical step in our suggested model's signal data fusion. FastICA was used as the signal-processing method in our study for this reason. FastICA aided us in the merging of data from several sources, allowing us to mix and extract important information while preserving the distinctive qualities of each data source. We accomplished effective signal integration using FastICA, allowing for a thorough analysis that captures the synergistic effects and correlations across the various data sources in our investigation.
- Diagnosis/outlier detection: The entire model's procedures are built with the goal of detecting faults, specifically through outlier detection. The model's structure is deliberately constructed to accomplish this aim. As the AI tool of choice in our investigation, we used an autoencoder LSTM. Details concerning the implementation and operation of the LSTM autoencoder have been discussed earlier in this section. The overarching goal is to use this AI tool to discover issues by finding anomalies in data, allowing for prompt diagnosis and intervention.

#### 4.1. Model Hyper-Parameter Function

In the hidden layers of neural networks, activation functions are used to introduce nonlinearity, which is critical for representing complex input. For instance, linear regression models are insufficient for most data representations because they lack nonlinear activation functions. Sigmoid, *tanh*, and ReLU (Rectified Linear Unit) are examples of common activation functions that are often employed in deep-layer neural networks. In binary classification tasks, the sigmoid function transfers inputs to a range of 0 to 1. However, given big input values, it can saturate, inhibiting learning. The *tanh* function is similar to the sigmoid function, however, it maps inputs to a range of -1 to 1.

On the other hand, ReLU has grown in popularity as a result of its capacity to improve training efficiency and effectiveness. Positive inputs are kept while negative inputs are set to 0. ReLU can experience the "dying ReLU" problem when neurons stuck in the negative area become inactive, despite its simplicity and computational efficiency. Loss functions, also known as cost functions, estimate how much the actual ground truth departs from the outputs that were projected. Various task kinds are catered for by various loss functions. Cross-entropy loss is appropriate for classification jobs while mean squared error loss is frequently utilized for regression activities. When developing deep learning models, the loss function is minimized by changing the model's weights and biases. Iterative optimization is used to improve the model's performance and accuracy. The mathematical equations for some of the regularly employed activation functions *sigmoid*, *relu*, and *softmax* are presented in Equations (13)–(15) respectively.

$$f_x = \frac{1}{1 + e^{-x}} \quad (13)$$

$$f_x = g_x = \max(0, x) \quad (14)$$

$$\text{softmax}(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \quad (15)$$

The success of a model is often determined by the architecture chosen, a decision that is often reliant on the researcher's knowledge and experience. Table 1 details the Architecture Parameters of the model used in our analysis.

**Table 1.** LSTM-Autoencoder Architecture Model Parameter.

Model Architecture	Description
Number of Classes	2
Number of Layers	7
Batch Size	32
Number of Epochs	100
Dropout Rate	0.001
Optimizer	Adam
Activation Function	Relu
Loss Function	MSE
Validation Split	0.2

#### 4.2. Model Global Performance Evaluation Metrics

It is critical to thoroughly examine the diagnostic skills of various deep learning while taking into account variables like model complexity, computational needs, and parameterization in order to accurately estimate their capabilities. This makes the use of defined criteria for assessing performance and discriminating necessary. These parameters include F1-score, accuracy, sensitivity, precision, and false alarm rate. By using these measurements, we can compare and objectively assess the performance of various models, allowing us to make well-informed decisions based on their individual advantages and disadvantages. Some of the known global evaluation metrics employed in studies are presented thus in Equations (16)–(20).

$$\text{Accuracy} = \frac{TP}{TP + FP + TN + FN} \quad (16)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (17)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (18)$$

$$\text{F1-Score} = \frac{2 * \text{Sensitivity} * \text{Precision}}{\text{Precision} + \text{sensitivity}} \quad (19)$$

$$\text{FAR} = \frac{FP}{FP + TN} \quad (20)$$

where  $TP$ ,  $FP$ ,  $TN$ , and  $FN$ , respectively, are the numbers of accurately classified groups, numbers of inaccurately classified groups, numbers of inaccurately labeled samples that belong to a group that was accurately classified, and the number of inaccurately labeled samples belonging to a group that was inaccurately classified.

It is essential to evaluate categorization models in order to judge their effectiveness and dependability. Although metrics like true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) give a general picture of classification accuracy, it is frequently required to assess the performance of each specific class to get a more complete picture. Take the case of a classifier that completes a five-class issue with an overall accuracy of 95%. This apparently great accuracy may really be the consequence of the model's ability to categorize three or four of the five classes accurately while misclassifying the other one or two classes. However, in the case of an outlier detection model as in the case of our model, determining these metrics helps assess the model's performance in accurately identifying outliers while minimizing false positives and false negatives, ensuring effective outlier detection not just for the instance but also when employed in other instances.

This discrepancy underlines the necessity for the confusion matrix, which enables us to assess the diagnostic efficacy of each class in a model. The confusion matrix gives a thorough examination of the predictions made by the classifier/outlier, dividing them into true positives, false positives, true negatives, and false negatives for each class. We may evaluate a model's performance more carefully by understanding where it performs best or worse by examining this matrix.

In conclusion, while global metrics offer an overall evaluation of classification accuracy, assessing class-specific performance using the confusion matrix is essential to spot any inconsistencies or biases and to make defensible choices about the validity of a classification model.

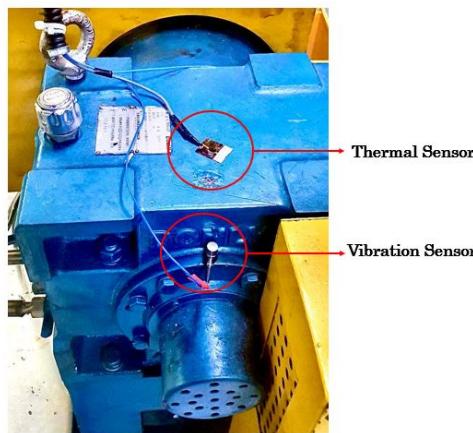
## 5. Data Collection and Methodology

This section discusses the data acquisition process, sensor placement, signal processing, feature extraction, feature selection, and signal fusion. The data employed in the study were acquired from two independent plastic extruder machines (healthy and faulty machines) in SPONTECH.

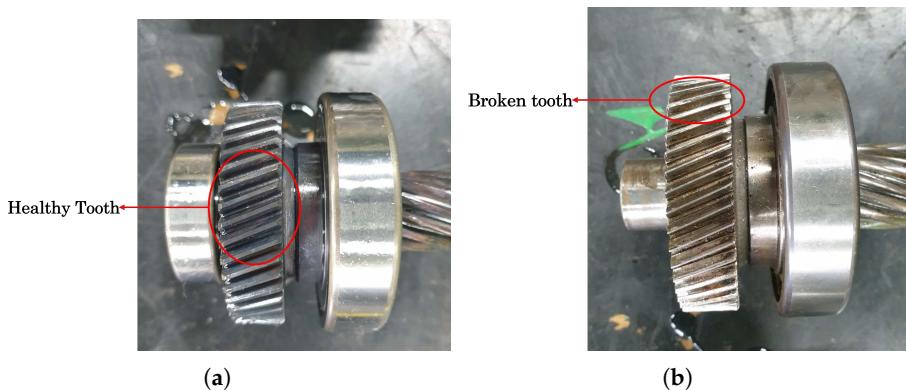
SPONTECH is a subsidiary of Toray Inc; Toray Co. is Japan's premier chemical and textile conglomerate, with an unrivaled No. 1 position in carbon fiber, as well as Japan's leading material giant, producing engineering plastics, IT materials, and chemical fibers in addition to carbon fibers.

Figure 3 depicts the setup of individual sensors on the plastic extruder machine. These sensors are deliberately critically placed to collect crucial information that will be used to generate a dataset with useful data when examined. As previously stated, we used two plastic extruder machines in our research. The first machine as seen in Figure 4a had been running for less than four months, and its data had been designated as the healthy dataset.

The second machine, represented in Figure 4b, had been in service for more than two years and had a chipped gear tooth. This machine (with the chipped gear tooth) was used to create the faulty dataset. We sought to evaluate and analyze the variations between the healthy and faulty situations by incorporating data from these two machines in order to get insights into the performance and potential concerns of the plastic extruder machines.

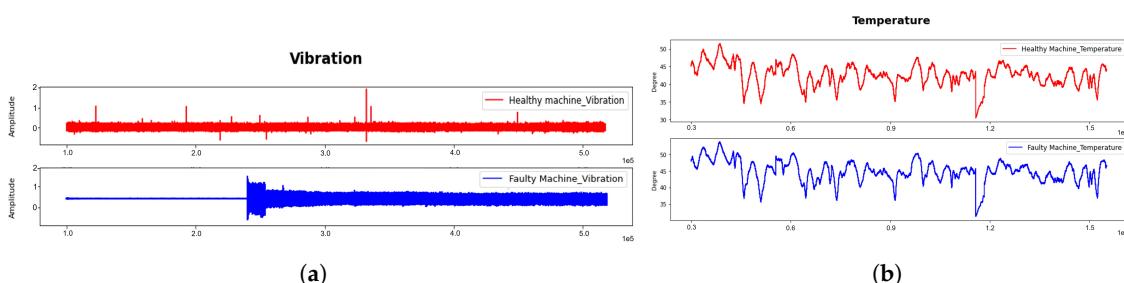


**Figure 3.** Sensor placement for data acquisition.



**Figure 4.** Pictorial View of two extruder gearbox: (a) Healthy plastic extruder gearbox overview (b) Faulty plastic extruder gearbox overview.

Figure 5a and 5b depict a visualization of time-domain data gathered from the vibration and thermal sensors respectively in order to have a better understanding of the data collected from both machines. In our investigation, We used a shear Piezotronics accelerometer model 353B33 with a vibration sensitivity of 99.2mv/g for our research. To acquire thermal data, we also used an RTD PT100 thermal sensor. This visualization provides a comprehensive perspective of the data collected from various sensors, allowing us to assess and comprehend the nature of the measurements gathered from the plastic extruder machines used in our research.



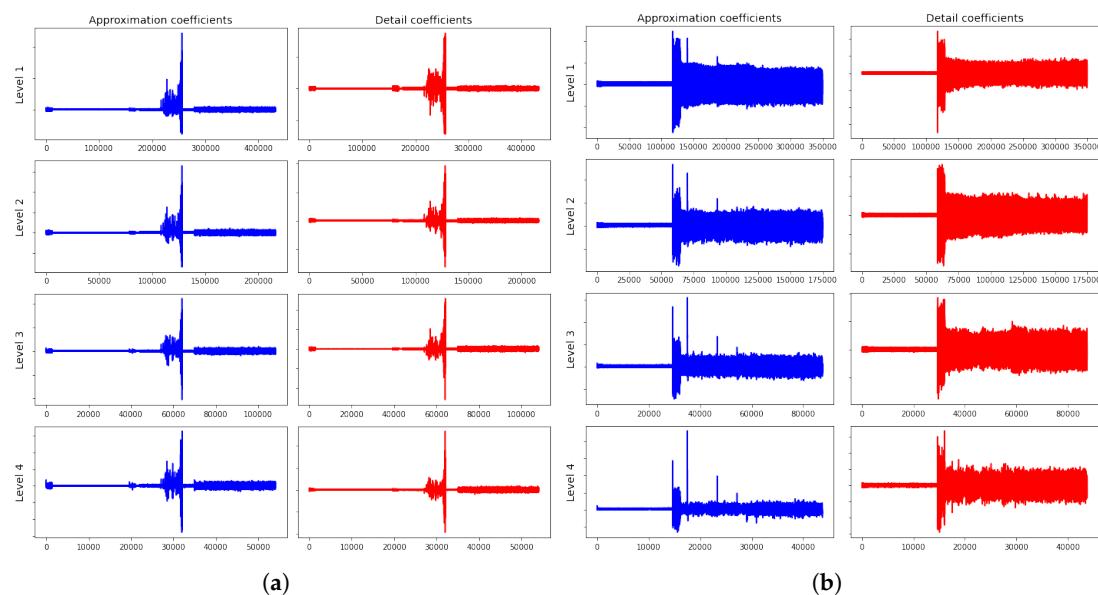
**Figure 5.** Raw motor time-domain signal from both extruder gearboxes: (a) Vibration signals (b) thermal signals.

From the observation in Figure 5a, a little distinction can be seen between the vibration data generated from the healthy and faulty gearbox conditions. The healthy data displayed a uniform

periodic pattern throughout the whole range of the dataset, occasionally modulated at various intervals. On the other hand, the faulty data visualization shows a non-consistent behavior across the whole data range; the early part displayed a non-constituent data display while the remaining part of the dataset displayed to an extent a uniform visualization of the dataset. However, it is important to note that this visual representation alone might not necessarily indicate the needed discriminative information for ensuring the effectiveness of an anomaly detection model.

Additionally, Figure 5b shows a comparable data visualization of temperature signals for both the healthy and faulty gearbox. On the other hand, it is noticeable that the temperature measurements in the healthy dataset are a little lower than those in the faulty dataset. The difference is normal given that a damaged gearbox will probably produce higher temperatures than a healthy one. These temperature changes can help spot abnormalities in gearbox performance and offer useful insights into possible variances between the two circumstances.

The substantial quantity of noise contained in the vibration signals produced by gearboxes must be addressed in order to efficiently extract or enhance vital information. As a result, we used (DWT) to de-noise the signals for both healthy and malfunctioning gearboxes. The visual representation of the decomposition and de-noising of the vibration signal received from the gearboxes is shown in Figure 6. By separating the wanted signal components from the noise using this method, we are able to emphasize and extract the essential data required for additional analysis and diagnostics.



**Figure 6.** Discrete wavelet transform for : (a) Healthy plastic extruder gearbox decomposition (b) Faulty plastic extruder gearbox decomposition.

The Discrete Wavelet Transform (DWT) was applied to both the healthy and faulty gearbox signals, as shown in Figure 6a and 6b, resulting in a four (4) level decomposition. This decomposition efficiently decreases the effects of noise in the signals, revealing the time-frequency domain properties of the processed signals. As discussed earlier in the previous chapter, the DWT decomposition generates the approximate and detailed coefficients that represent the signal's low and high frequencies, respectively. In our investigation, we concentrated on the approximate coefficient because it offers more detailed information on the gearbox signal's significant frequencies and features.

### 5.1. Feature Extraction

DWT are signal processing tools that uniquely transform a signal to its time-frequency domain; thus, time-frequency domain features are frequently used to ensure that useful information is

successfully extracted from these signal-presenting features that are rich and contain all of the useful details of a given signal. In our investigation, we used a multi-sensor approach with only the vibration signal being subjected to a DWT; on the other hand, we retrieved only time-domain statistical features from the temperature, which is time-variant data.

**Table 2.** Statistically extracted features and their mathematical formulas.

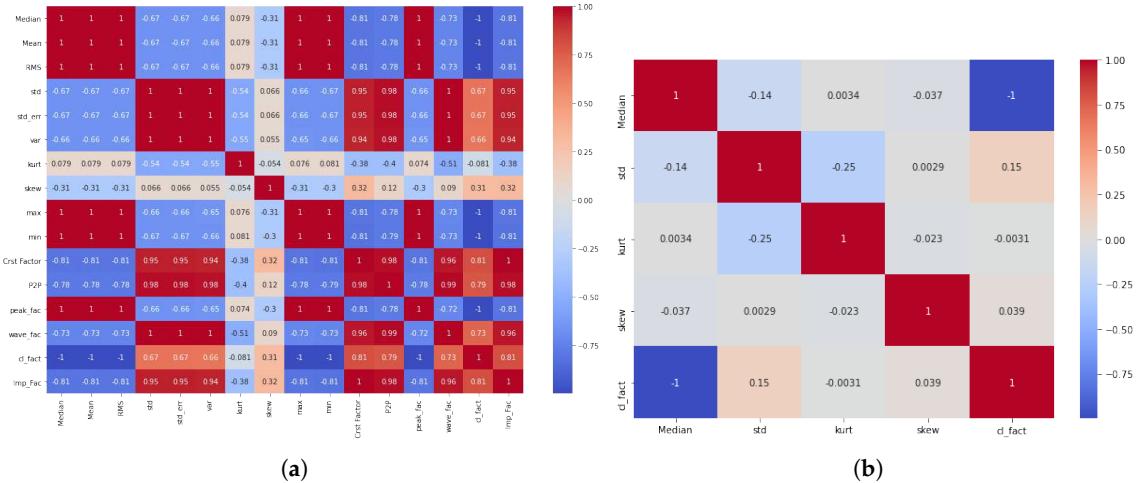
Domain	Feature Name	Definition
	Median	$\left(\frac{n+1}{2}\right)^{th}$ sample
	Mean	$\bar{x} = \frac{1}{n} (\sum_{i=1}^n x_i)$
	Root Mean Square	$X_{rms} = \sqrt{\frac{\sum_{i=1}^n (x_i)^2}{n}}$
	Standard Deviation	$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{n}}$
	Variance	$S^2 = \frac{\sum (x_i - \mu)^2}{n}$
	Standard Error	$SE = \frac{\sigma}{\sqrt{n}}$
	Kurtosis	$X_{kurt} = \frac{1}{N} \sum \left( \frac{(x_i - \mu)^3}{\sigma} \right)$
Time-Domain	Skewness	$X_{skew} = E \left[ \left( \frac{(x_i - \mu)^3}{\sigma} \right) \right]$
	Max	$X_{max} = \max(x_i)$
	Min	$X_{min} = \min(x_i)$
	Crest Factor	$X_{CF} = \frac{x_{max}}{x_{rms}}$
	Peak-to-peak	$X_{p-p} = x_{max} - x_{min}$
	Peak factor	$x_{PF} = \frac{x_{max}}{\sqrt{x_s}}$
	Wave Factor	$x_{WF} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n  x_i ^2}}{\frac{1}{n} \sum_{i=1}^n  x_i }$
	Clearance factor	$x_{CF} = \frac{x_{max}}{\text{mean}  2x }$
	Impulse factor	$X_{IF} = \frac{x_{max}}{\frac{1}{N} \sum_{i=1}^N  x_i }$
	Maximum Frequency	$X_{MF} = x_{max} (\frac{1}{N} \sum_{i=1}^N  x_i ^2)$
Frequency-Domain	Variance Frequency	$X_{VF} = x_{var} (\frac{1}{N} \sum_{i=1}^N  x_i ^2)$
	Spectral Kurtosis	$SK = \frac{2 \sum_{k=0}^{B_L/2-1} ( X(k,n)  - \mu X )^4}{B_L \cdot \sigma_{ X }^4} - 3$
	Spectral Skewness	$SS = \frac{2 \sum_{k=0}^{B_L/2-1} ( X(k,n)  - \mu X )^4}{B_L \cdot \sigma_{ X }^3}$

In this study, we used sixteen (16) statistical features in the time domain to evaluate temperature data. In addition, we all employed sixteen time-frequency domain features in analyzing the DWT decomposition of the vibration signal as well as the original signal; the sixteen statistical time-frequency features included twelve (12) time-domain features and four (4) frequency-domain features. Our goal was to extract useful information from the signals in order to improve the model's efficiency.

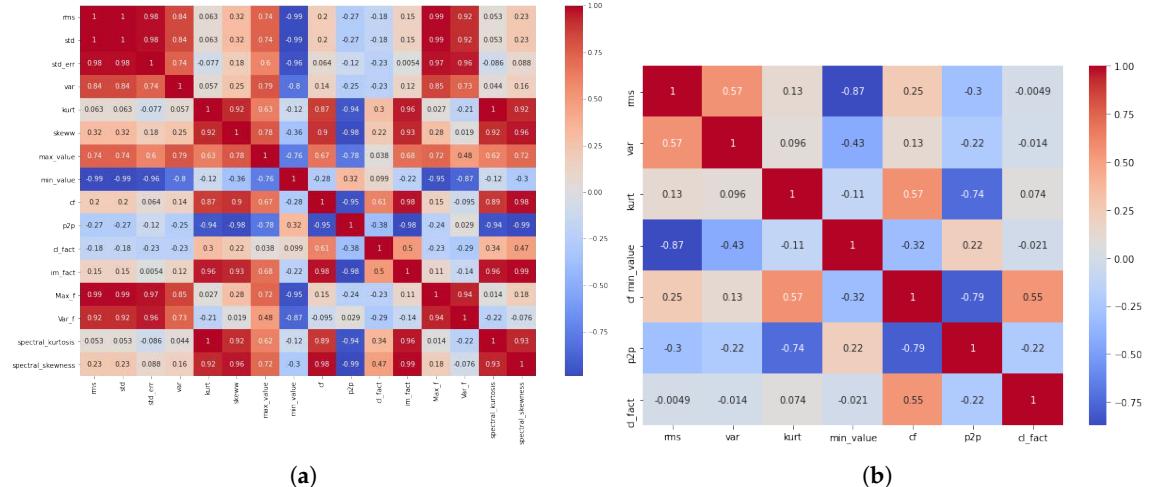
It is vital to highlight that we did not adhere to any precise criteria when selecting statistical features. Instead, we chose based on the popularity of specific characteristics in the area and the authors' experience.

## 5.2. Feature Selection and Sensor Fusion

To evaluate the adequacy of the extracted features for our model, we conducted a discriminant test using a Pearson correlation-filter-based approach. This method involved assessing the correlation between features and dropping the features with a correlation of 70% or more leaving behind features below the 70% similarity threshold. Figures 7 and 8 show the correlation plot and the filter-based correlation plot for both the thermal and vibration datasets employed in our study, respectively.



**Figure 7.** Correlation plot for : (a) All extracted statistical features for healthy gearbox (b) all selected statistical features from healthy gearbox.



**Figure 8.** Correlation plot for : (a) All extracted statistical features for faulty gearbox (b) all selected statistical features from faulty gearbox.

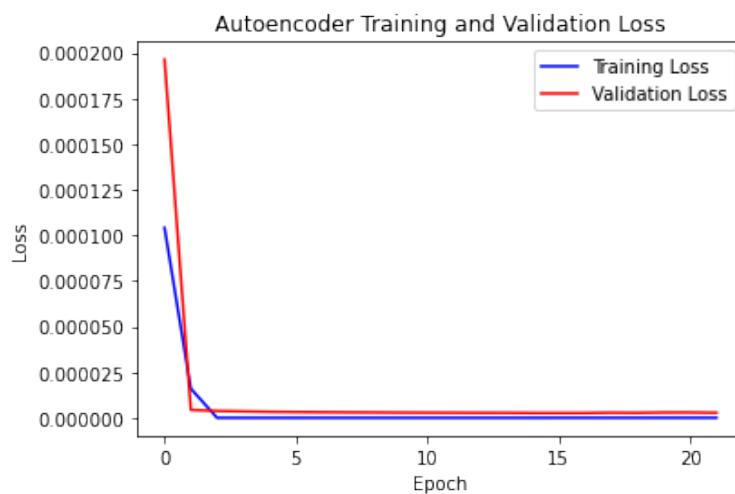
By this means, the correlation filter-based model selected five (5) features from the thermal data (shown in Figure 7b) and seven (7) features from the vibration data (shown in Figure 8b). This feature selection process efficiently reduced the dataset, retaining only the relevant and most discriminant features necessary for optimal model performance.

To integrate the multi-sensor data in our study, FastICA (Fast Independent Component Analysis) was used to combine multi-sensor data in our investigation. This method was used to keep the distinguishing characteristics of each sensor's separate qualities while blending them together. The FastICA ensures that the fused data keeps the distinct properties of each sensor, allowing us to gather and exploit the essential information from all sensors in a cohesive manner.

## 6. Experimental Results and Model Evaluation

An autoencoder is a neural network that learns to reconstruct input data from a compressed representation known as encoding. On the other hand, LSTM, a recurrent neural network, is frequently employed in language processing and captures long-term dependencies. In order to take input data, learn a concise representation, and reconstruct the original input, an LSTM can be incorporated into an autoencoder architecture. The reconstruction loss is used to train the autoencoder. Although the number of features has no direct impact on performance, large amounts of input data could make it harder to learn a decent representation. Performance is influenced by the size, reliability, and selection of the hyperparameters.

LSTM-autoencoders, like other neural networks, can be trained using common approaches such as stochastic gradient descent and back-propagation. The reconstruction loss, which measures the discrepancy between the input and output sequences, is often used to evaluate network performance. The LSTM-autoencoder can learn to decrease this reconstruction loss and provide accurate reconstructions of the input data by refining the model's parameters using gradient-based optimization approaches. To create a model that could efficiently determine anomaly in a plastic extruder gearbox, an autoencoder LSTM architecture was employed in our study; where the LSTM captured the long-term dependency of a given data which in this case a fusion of the machine's vibration and thermal sensor data, while the autoencoder helps in dimension and also for feature learning. The mean squared error (MSE) is a popular loss function used in the training of an LSTM-autoencoder. The MSE quantifies the average squared difference between the expected and true outputs, indicating the goal of accurately reconstructing the input data. The model is trained on labeled training data and its performance is evaluated on a separate validation set during the training phase. The validation loss is computed on the validation set, while the training loss is computed on the training set. It is critical to monitor the trend of these losses, as a considerable difference between them can suggest over-fitting. Over-fitting happens when the model fits the training data too closely, resulting in poor performance on unknown data. In academia lots of methodologies have been presented in mitigating over-fitting in autoencoder LSTM, these techniques include dropout, early stop, regularization, data augmentation, earlier data fusion, and reducing model complexity [59]. Some of these techniques we consciously ensure we implement in our model setup to enable and efficiently model performance that is void of over-fitting. Figure 9 the training and validation loss for the LSTM-autoencoder model employed in our study.

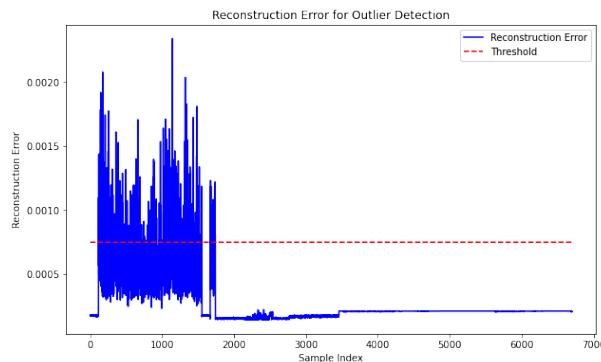


**Figure 9.** Training and validation loss plot for our autoencoder LSTM model.

Figure 9 depicts our model's effective training, adaption, and validation. It clearly shows the discernible difference between validation and training loss. Notably, throughout the early stages of

model training, the difference between the training and validation losses shrank significantly, reaching a point of negligible importance around the 22nd epoch. Because of this convergence, model training was halted at that epoch.

The key objective of our model is to create a framework capable of quickly identifying instances of anomalies within a plastic extruder gearbox. As a result, it is critical to evaluate the model's performance by subjecting it to our faulty collected data. This technique seeks to assess the model's competency and precision in finding faults using an outlier fault detection methodology included in the model's architecture. We also employed other evaluation metrics to authenticate our model such as precision, F1-score, accuracy, and accuracy to achieve this goal. These metrics prove valuable when an individual possesses the actual labels of input data and seeks to group the signal. The LSTM-autoencoder's architecture is composed of seven layers and encompasses 247,937 parameters; the seven layers include an input layer, four encoder LSTM layers, a repeated vector layer, and a time-distributed layer. A comprehensive depiction of the LSTM-autoencoder model is available in Table 1, showcasing the architecture of our model such as a dropout rate set at 0.001, a total of 100 epochs, a batch size of 32, seven layers, and two classes. After successfully training our model, we used an outlier detection technique to predict the presence of an abnormality in our faulty dataset. Figure 10 depicts the reconstruction error and the threshold, which demonstrate the principle of the outlier detection technique used in our model; reconstruction error is the difference or discrepancy between the input data and the output data of a model, which occurs frequently when the input data is fed through an encoding and decoding process. In the case of anomaly detection, as used in our model, reconstruction error is frequently used to highlight the dissimilarity between input and output data, perhaps indicating if the data is anomalous or faulty. In our situation, we used the dataset gathered from the healthy extruder gearbox to train our model, and the dataset collected from the faulty extruder gearbox as the output dataset.



**Figure 10.** Visualization plot for the set threshold with reconstruction error for the proposed model's outlier detection.

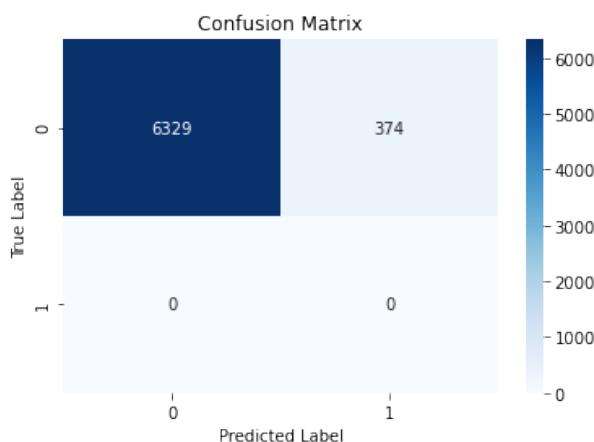
Global metrics are critical in measuring the performance and efficacy of models across entire datasets or any particular problem the model is intended to address in the realms of data analysis, machine learning, and assessment. When looking for a full understanding of how a model performs across numerous classes, categories, or instances, using global evaluation metrics becomes very important. This method allows for a comprehensive assessment of a model's strengths and shortcomings across all groups and categories.

To evaluate the model performance in our study, we used global evaluation metrics such as Accuracy, F1-Score, Recall, and Precision. These metrics provide a comprehensive picture of how well the model performs in various settings. Our findings are summarized in Table 3, which summarizes the conclusions of our inquiry.

**Table 3.** Global Evaluation Metrics values for the Autoencoder-LSTM Model.

Definition	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
AE-LSTM	94.42	100	93.67	91.72

To emphasize the importance of model validation, we present the concept of a confusion matrix. This tool we used to determine whether the presented accuracy corresponds to the predicted labels' class classifications. Figure 11 presents an overview of the confusion matrix technique on our model. The derived confusion matrix shows that the model's predictions are proportionally consistent with the accuracy metric, reinforcing the model's reliability.

**Figure 11.** Confusion matrix for faulty class anomaly outlier detection for the proposed model.

## 7. Discussion and Conclusion

Breakdowns in machinery can have serious consequences, including downtime and financial losses. Our research focused on the gearbox of a plastic extruder machine, which is primarily made up of helical gears. helical gears, despite their lower susceptibility to failure, are not immune to breakdowns, necessitating the implementation of Condition-Based Monitoring (CBM). Our analysis used a multi-sensor approach, including vibration and thermal sensors. Traditional vibration measurement can be hampered due to the tendency of defective gearboxes to generate excessive noise. Our research resulted in a solid framework that included various methodologies such as the Discrete Wavelet Transform (DWT) for vibration signal decomposition, time-frequency statistical feature extraction, correlation filter-based feature selection, the Fast Independent Component Analysis (FASTICA) sensor fusion technique, and an outlier fault detection approach.

One of our study's main goals was to create a model capable of seamlessly merging different sensors while retaining their inherent properties, which aligned with the study's overall goal. We extracted 16 time-domain features from temperature signals and time-frequency features from vibration signals to do this. Following that, we used a careful procedure to choose the five and seven most important features from the thermal and vibration datasets respectively. Using the Fast Independent Component Analysis (FASTICA) approach, these selected features were harmoniously blended into a single-dimensional representation. pleasantly, our innovative implementation of the Autoencoder-LSTM outlier fault detection technique achieved a remarkable prediction performance accuracy of 94.42%, setting an impressive milestone in our research path. In a commitment to ensuring the integrity of our model, we thoroughly scrutinized our outcomes using a variety of global evaluation metrics. This extensive study served to validate and highlight the robustness and dependability of our proposed system. Overall, our study's multidimensional approach not only addressed sensor fusion but also demonstrated the potential of our model for effective problem detection and classification in the context of plastic extruder gearbox systems.

A pivotal hurdle encountered in our study revolved around data acquisition, which had a substantial impact on the data collection procedure. This highlighted the need to use appropriate data extraction strategies to achieve the goals of our study. Future studies could compare the accuracy and performance of our model to alternative models that use different feature extraction and selection approaches. Additionally, the incorporation of more advanced deep learning algorithms could be investigated in order to evaluate whether improved performance can be attained within feasible computing timeframe. Notably, despite our model's 94.42% accuracy, there is still room for significantly better performance results. These efforts would contribute to a thorough understanding of fault detection models, helping the improvement and optimization of our proposed system.

**Author Contributions:** Conceptualization, J.-H.L. and C.N.O.; methodology, C.N.O.; software, J.-H.L. and C.N.O.; formal analysis, C.N.O.; investigation, J.-H.L. and C.N.O.; resources, J.-H.L., C.N.O. and J.-W.H.; data curation, J.-H.L.; writing—original draft, J.-H.L. and C.N.O.; writing—review and editing, C.N.O.; visualization, J.-H.L. and C.N.O.; supervision, J.-W.H.; project administration, J.-W.H.; funding acquisition, J.-W.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Innovative Human Resource Development for Local Intellectualization support program (IITP-2023-2020-0-01612) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation)

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

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

**Sample Availability:** The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to laboratory regulations.

## Abbreviations

The following abbreviations are used in this manuscript:

RUL	Remaining Useful Life
FN	False Negative
TN	True Negative
TP	True Positive
FDI	False Detection and Isolation
IoT	Internet of Things
PHM	Prognostics and Health Management
LSTM	Long Short Term Memory
ANN	Artificial Neural Network
ML	Machine Learning
DL	Deep Learning
FNNN	Feed-forward neural networks
DNN	Deep neural networks
CNN	Convolutional neural networks
DBN	Deep belief networks
DWT	Discrete Wavelet Transform
LLE	Local Linear Embedding
PCA	Principal Component Analysis
ICA	Independent Component Analysis
AE	Autoencoder
AI	Artificial Intelligence
FastICA	Fast Independent Component Analysis
MSE	Mean Square Error
CBM	Condition Based Monitoring

## References

1. Kumar, S.; Tiwari, P.; Zymbler, M. Internet of Things is a revolutionary approach for future technology enhancement: a review. *J Big Data.* **2019**, *6*, 111. <https://doi.org/10.1186/s40537-019-0268-2>.
2. Zhou, I.; Makhdoom, I.; Shariati, N.; Raza, M. A.; Keshavarz, R.; Lipman, J.; Abolhasan, M.; Jamalipour, A. Internet of Things 2.0: Concepts, Applications, and Future Directions. *IEEE Access.* **2021**, *9*, 70961-71012. <https://doi.org/10.1109/ACCESS.2021.3078549>.
3. Do, J.S.; Kareem, A.B.; Hur, J.-W. A. LSTM-Autoencoder for Vibration Anomaly Detection in Vertical Carousel Storage and Retrieval System (VCSRS). *Sensors.* **2023**, *23*, 1009. <https://doi.org/10.3390/s23021009>.
4. Wu, G.; Yan, T.; Yang, G.; Chai, H.; Cao, C. A Review on Rolling Bearing Fault Signal Detection Methods Based on Different Sensors. *Sensors.* **2022**, *22*, 8330. <https://doi.org/10.3390/s22218330>.
5. Singh, V.; Mathur, J.; Bhatia, A. A comprehensive review: Fault detection, diagnostics, prognostics, and fault modeling in HVAC systems. *International Journal of Refrigeration.* **2022**, *144*, 283-295. <https://doi.org/10.1016/j.ijrefrig.2022.08.017>.
6. Ghobakhloo, M.; Iranmanesh, M.; Tseng, M-L.; Grybauskas, A.; Stefanini, A.; Amran, A. Behind the definition of Industry 5.0: a systematic review of technologies, principles, components, and values. *Journal of Industrial and Production Engineering.* **2023**, *40*:6, 432-447. <https://doi.org/10.1080/21681015.2023.2216701>.
7. Wang, H.; Zhou, C.; Hu, B.; Liu, Z. Tooth wear prediction of crowned helical gears in point contact. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology.* **2020**, *6*, 947-963. <https://doi.org/10.1177/1350650119896467>.
8. Okwuosa, C.N.; Akpudo, U.E.; Hur, J.-W. A Cost-Efficient MCSA-Based Fault Diagnostic Framework for SCIM at Low-Load Conditions. *Algorithms.* **2022**, *15*, 212. <https://doi.org/10.3390/a15060212>.
9. Okwuosa, C.N.; Hur, J.-w. A Filter-Based Feature-Engineering-Assisted SVC Fault Classification for SCIM at Minor-Load Conditions. *Energies.* **2022**, *15*, 7597. <https://doi.org/10.3390/en15207597>.
10. Gundewar, S.K.; Kane, P.V. Condition Monitoring and Fault Diagnosis of Induction Motor. *J. Vib. Eng. Technol.* **2021**, *9*, 643-674. <https://doi.org/10.1007/s42417-020-00253-y>.
11. Zhou, J.; Sun, W.; Wang, Z. Vibration and noise characteristics of a gear reducer under different operation conditions. *Journal of Low Frequency Noise, Vibration and Active Control.* **2019**, *2*, 574-591. <https://doi.org/10.1177/1461348419825603>.
12. Amarnath, M.; Krishna, I. R. P. Vibration and noise characteristics of a gear reducer under different operation conditions. *Local fault detection in helical gears via vibration and acoustic signals using EMD based statistical parameter analysis.* **2014**, *58*, 154-164. <https://doi.org/10.1016/j.measurement.2014.08.015>.
13. Karabacak, Y. E.; Özmen, N. G.; Gümüşel, L. Intelligent worm gearbox fault diagnosis under various working conditions using vibration, sound and thermal features. *Applied Acoustics.* **2022**, *58*, 108463. <https://doi.org/10.1016/j.apacoust.2021.108463>.
14. Tang, X.; Xu, Y.; Sun, X.; Liu, Y.; Jia, Y.; Gu, F.; Ball, A. D. Intelligent fault diagnosis of helical gearboxes with compressive sensing based non-contact measurements. *Applied ISA Transactions.* **2023**, *133*, 559-574. <https://doi.org/10.1016/j.isatra.2022.07.020>.
15. Roda-Casanova, V.; Gonzalez-Perez, I. Investigation of the effect of contact pattern design on the mechanical and thermal behaviors of plastic-steel helical gear drives. *Mechanism and Machine Theory/Mechanism and Machine Theory.* **2021**, *164*, 104401. <https://doi.org/10.1016/j.mechmachtheory.2021.104401>.
16. Homayouni, H.; Ghosh, S.; Ray, I.; Gondalia, S.; Duggan, J.; Kahn, M. G. An Autocorrelation-based LSTM-Autoencoder for Anomaly Detection on Time-Series Data. *IEEE International Conference on Big Data (Big Data).* **2020**, 5077. <https://doi.org/10.1109/BigData50022.2020.9378192>.
17. Mallak, A.; Fathi, M. Sensor and Component Fault Detection and Diagnosis for Hydraulic Machinery Integrating LSTM Autoencoder Detector and Diagnostic Classifiers. *Sensors.* **2021**, *433*, 104401. <https://doi.org/10.3390/s21020433>.
18. Kishore, K.; Sharma, A.; Mukhopadhyay, G. Failure Analysis of a Gearbox of a Conveyor Belt. *J Fail. Anal. and Preven.* **2020**, *20*, 1237-1243. <https://doi.org/10.1007/s11668-020-00928-4>
19. Yong, C.; Jinkai, L.; Libin, Z.; Yiqi, L.; Wangyang, B.; Xiaopeng, Y. Dynamic Simulation and Experimental Identification for Fatigue Pitting Helical Gear Fault[J]. *Journal of Mechanical Engineering.* **2021**, *57*, 61-70. <https://doi.org/10.3901/JME.2021.09.061>

20. Nejad, A. R.; Gao, Z.; Moan, T. Fatigue Reliability-based Inspection and Maintenance Planning of Gearbox Components in Wind Turbine Drivetrains. *Energy Procedia*. **2014**, *53*, 248-257. <https://doi.org/10.1016/j.egypro.2014.07.234>

21. Asi, O. Fatigue failure of a helical gear in a gearbox. *Engineering Failure Analysis*. **2006**, *13*, 1116-1125. <https://doi.org/10.1016/j.engfailanal.2005.07.020>

22. Kale, A. S. Bending Fatigue Failure In Gear Tooth. *IJERT*. **2013**, *02*, 2278-0181. <https://doi.org/10.17577/IJERTV2IS60387>

23. Zhang, S.; Zhou, J.; Wang, E.; Zhang, H.; Gu, M.; Pirttikangas, S. State of the art on vibration signal processing towards data-driven gear fault diagnosis. *Institution of Engineering and Technology*. **2022**, *4*, 249-266. <https://doi.org/10.1049/cim2.12064>

24. Zhang, S.; Zhou, J.; Wang, E.; Zhang, H.; Gu, M.; Pirttikangas, S. State of the art on vibration signal processing towards data-driven gear fault diagnosis. *Institution of Engineering and Technology*. **2022**, *4*, 249-266. <https://doi.org/10.1049/cim2.12064>

25. Poletto, J.C.; Fernandes, C.M.C.G.; Barros, L.Y.; Neis, P.D.; Pondicherry, K.; Fauconnier, D.; Seabra, J.H.O.; De Baets, P.; Ferreira. N.F. Identification of gear wear damage using topography analysis. *Wear*. **2023**, *522*, 204837. <https://doi.org/10.1016/j.wear.2023.204837>

26. Vanraj, ; Dhami, S. S.; Pabla, B. S. Gear fault classification using Vibration and Acoustic Sensor Fusion: A Case Study. *Condition Monitoring and Diagnosis (CMD)*. **2018**, *1-6*. <https://doi.org/10.1109/CMD.2018.8535974>.

27. Jaen-Cuellar, A.Y.; Trejo-Hernández, M.; Osornio-Rios, R.A.; Antonino-Daviu, J.A. Gear Wear Detection Based on Statistic Features and Heuristic Scheme by Using Data Fusion of Current and Vibration Signals. *Energies*. **2023**, *16*, 948. <https://doi.org/10.3390/en16020948>.

28. Zhang, Y.; Baxter, K. Deep Transfer Multi-sensor Fusion for Gearbox Diagnostics. *International Research Journal of Modernization in Engineering Technology and Science*. **2020**, *02*, PP 15. e-ISSN: 2582-5208.

29. Kumar, T.P.; Saimurugan, M.; Haran, R. B. H.; Siddharth, S.; Ramachandran, K. I.A multi-sensor information fusion for fault diagnosis of gearbox utilizing discrete wavelet features. *Measurement Science and Technology*. **2019**, *30*, 085101. <https://doi.org/10.1088/1361-6501/ab0737>.

30. Lee, M. -S.; Shifat, T. A.; Hur, J. -W. Kalman Filter Assisted Deep Feature Learning for RUL Prediction of Hydraulic Gear Pump. *in IEEE Sensors Journal*. **2022**, *22*, 11088-11097. <https://doi.org/10.1109/JSEN.2022.3167926>

31. Yuan, W. Study on Noise Elimination of Mechanical Vibration Signal Based on Improved Wavelet. *12th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Phuket, Thailand*. **2020**, 141-143. <https://doi.org/10.1109/ICMTMA50254.2020.00039>

32. Chen, X.; Yang, Y.; Cui, Z.; Shen, J. Wavelet Denoising for the Vibration Signals of Wind Turbines Based on Variational Mode Decomposition and Multiscale Permutation Entropy. *in IEEE Access*. **2020**, *8*, 40347-40356. <https://doi.org/10.1109/ACCESS.2020.2975875>

33. Chen, G.; Zhao, Y. JWavelet-based denoising: A brief review. *in Proc. 4th Int. Conf. Intell. Control Inf. Process. (ICICIP)*. **2013**, 570-574. <https://doi.org/10.1109/ICICIP.2013.6568140>

34. Chatterjee, S.; Thakur, R. S.; Yadav, R. N.; Gupta, L.; Raghuvanshi, D. K. Review of noise removal techniques in ECG signals. *IET Signal Process*. **2020**, *9*, 569-590. <https://doi.org/10.1049/iet-spr.2020.0104>

35. Sun, C.; He, Z.; Cao, H.; Zhang, Z.; Chen, X.; Jian M. Z. A non probabilistic metric derived from condition information for operational reliability assessment of aero-engines. *IEEE Trans. Rel.*. **2015**, *64*, 167-181. <https://doi.org/10.1109/TR.2014.2336032>.

36. Roemer, M. J.; Kacprzynski, G. J.; Schoeller, M. H. Improved diagnostic and prognostic assessments using health management information fusion. *in Proc. IEEE Autotestcon IEEE Syst. Readiness Technol. Conf., Valley Forge*. **2001**, 365-377. <https://doi.org/10.1109/AUTEST.2001.948984>.

37. Duan, L.; Zhao, F.; Wang, J.; Wang, N.; ZhangL, J. An integrated cumulative transformation and feature fusion approach for bearing degradation prognostics. *Shock Vibrat...* **2018**, *64*, 1-15. <https://doi.org/10.1155/2018/9067184>.

38. Akpudo, U. E.; Hur, J. -W. An Automated Sensor Fusion Approach for the RUL Prediction of Electromagnetic Pumps. *in IEEE Sensors Journal*. **2021**, *9*, 38920-38933. <https://doi.org/10.1109/ACCESS.2021.3063676>

39. Bagheriye, L.; Ali, G.; Kerkhoff, H. G. Life-time prognostics of dependable VLSI-SoCs using machine-learning. *in Proc. IEEE 26th Int. Symp. Line Test. Robust Syst. Design (IOLTS)*. **2020**, 1-4. <https://doi.org/10.1109/IOLTS50870.2020.9159753>

40. T.-W. Lee, M. Girolami, and T. J. Sejnowski, Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Comput.* **1999**, *11*, 417–441
41. Weidong, J. Fault diagnosis of gearbox by FastICA and residual mutual information based feature extraction. *in Proc. Int. Conf. Inf. Autom.* **2009**, *13*, 928–932. <https://doi.org/10.1109/ICINFA.2009.5205051>
42. Hyvärinen, A.; Oja, E. Independent component analysis: algorithms and applications. *Neural Networks*. **2000**, *13*, 411-430. [https://doi.org/10.1016/S0893-6080\(00\)00026-5](https://doi.org/10.1016/S0893-6080(00)00026-5)
43. Okwuosa, C.N.; Hur, J.-w. An Intelligent Hybrid Feature Selection Approach for SCIM Inter-Turn Fault Classification at Minor Load Conditions Using Supervised Learning. *IEEE ACCESS*. **2023**, *11*, 89907 - 89920. <https://doi.org/10.1109/ACCESS.2023.3266865>
44. Kang, J.; Kim, C.-S.; Kang, J.W.; Gwak, J. Anomaly Detection of the Brake Operating Unit on Metro Vehicles Using a One-Class LSTM Autoencoder. *Appl. Sci.* **2021**, *11*, 9290. <https://doi.org/10.3390/app11199290>
45. I.T.; Park, J.R.; Jung, K.; Lee, J.S.; Kang, D.-K. Anomaly Detection of Water Level Using Deep Autoencoder. *Sensors*. **2021**, *21*, 6679. <https://doi.org/10.3390/s21196679>
46. Givnan, S.; Chalmers, C.; Fergus, P.; Ortega-Martorell, S.; Whalley, T. Anomaly Detection Using Autoencoder Reconstruction upon Industrial Motors. *Sensors*. **2022**, *22*, 3166. <https://doi.org/10.3390/s22093166>
47. Givnan, Do, J.S.; Kareem, A.B.; Hur, J.-W. LSTM-Autoencoder for Vibration Anomaly Detection in Vertical Carousel Storage and Retrieval System (VCSRS). *Sensors*. **2022**, *23*, 1009. <https://doi.org/10.3390/s23021009>
48. Jin, J.; Chung, Y.; Park, J. Development of a Flowmeter Using Vibration Interaction between Gauge Plate and External Flow Analyzed by LSTM. *Sensors*. **2020**, *20*, 5922. <https://doi.org/10.3390/s20205922>
49. Lee, J.-H.; Hong, J.-K. Comparative Performance Analysis of Vibration Prediction Using RNN Techniques. *Electronics*. **2022**, *11*, 3619. <https://doi.org/10.3390/electronics11213619>
50. Huang, J.; Mo, J.; Zhang, J.; Ma, X. A Fiber Vibration Signal Recognition Method Based on CNN-CBAM-LSTM. *Appl. Sci.* **2022**, *12*, 8478. <https://doi.org/10.3390/app12178478>
51. Gu, K.; Zhang, Y.; Liu, X.; Li, H.; Ren, M. DWT-LSTM-Based Fault Diagnosis of Rolling Bearings with Multi-Sensors. *Electronics*. **2021**, *10*, 1009. <https://doi.org/10.3390/electronics10172076>
52. Kwon, J.E.; Shifat, T.A.; Kareem, A.B.; Hur, J.-W. RUL Prediction of Switched Mode Power Supply Using a Kalman Filter Assisted Deep Neural Network. *Processes*. **2022**, *10*, 55. <https://doi.org/10.3390/pr10010055>
53. Curreri, F.; Patanè, L.; Xibilia, M.G. RNN- and LSTM-Based Soft Sensors Transferability for an Industrial Process. *Sensors*. **2021**, *21*, 823. <https://doi.org/10.3390/s21030823>
54. Wang, X.; Huang, T.; Zhu, K.; Zhao, X. LSTM-Based Broad Learning System for Remaining Useful Life Prediction. *Mathematics*. **2022**, *10*, 2066. <https://doi.org/10.3390/math10122066>
55. Wang, X.; Huang, T.; Zhu, K.; Zhao, X. LSTM-Based Broad Learning System for Remaining Useful Life Prediction. *Mathematics*. **2022**, *10*, 2066. <https://doi.org/10.3390/math10122066>
56. Zaman, S. M. K.; Liang, X. An Effective Induction Motor Fault Diagnosis Approach Using Graph-Based Semi-Supervised Learning. *in IEEE Access*. **2021**, *9*, 7471-7482. <https://doi.org/10.1109/ACCESS.2021.3049193>
57. Zaman, S. M. K.; Liang, X. An Effective Induction Motor Fault Diagnosis Approach Using Graph-Based Semi-Supervised Learning. *in IEEE Access*. **2021**, *9*, 7471-7482. <https://doi.org/10.1109/ACCESS.2021.3049193>
58. Amruthnath, N.; Gupta, T. A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. *5th International Conference on Industrial Engineering and Applications (ICIEA)*. **2018**, 355-361. <https://doi.org/10.1109/IEA.2018.8387124>
59. Jabbar, H. K; Khan, R. Z. Methods to Avoid Over-Fitting and Under-Fitting in Supervised Machine Learning (Comparative Study. *semantic scholar*. [https://doi.org/10.3850/978-981-09-5247-1\\_017](https://doi.org/10.3850/978-981-09-5247-1_017)

**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.