



INDUSTRIAL MACHINE FAILURE PATTERN ANALYSIS FOR AUTOMATED DEFECT DIAGNOSIS IN ADVANCED MANUFACTURING SYSTEMS

Annotation:

This study investigates the application of exploratory data analysis as a tool to facilitate automated defect diagnosis and reliability engineering of advanced industrial automation systems. The research is based on a quantitative exploratory design with the help of the Predictive Maintenance Dataset with Irregularities to explore the machine operating conditions and failure patterns. The focus of the analysis is on the distribution of machine failures, variations in operating parameters, and relationships between the operating parameters of the machines. These results show that the dataset is dominated by normal operating conditions, and the failure events make up a relatively small fraction of the dataset. The most common fault categories are heat dissipation failure and overstrain failure. It is discovered through comparative analysis of the operating conditions that various types of failure have different characteristics in operation, specifically in terms of torque, rotational speed, tool wear, temperature, etc. The air temperature and process temperature are highly positively correlated, and the rotational speed and torque are highly negatively correlated. Failure status is correlated with air temperature and torque, with weak to moderate positive correlations, which could be used as diagnostic criteria. Air temperature and torque are correlated with failure status with weak to moderate positive correlation, indicating that they could be used as diagnostic variables. The results suggest the importance of exploratory analysis to detect patterns of operation and factors that repeat as failure in an industrial setting. The study is an important contribution to reliability engineering by providing evidence-based information that could help maintenance personnel in detecting problems early and making informed decisions, as well as help design future automated defect diagnosis frameworks.

Defect, Diagnosis, Reliability, Engineering, Industrial, Automation, Predictive, Dataset

Keywords:

Information about the authors

1)Umme Habiba Aesha

Department of Mechanical Engineering, Minnesota State University,
Mankato, Mankato,
Minnesota, USA
email: habibaaesha00009@gmail.com

2) Reyan Hridoy Bhuiyan

Department of Mechanical Engineering, Minnesota State University,
Mankato, Mankato,
Minnesota, USA
Email: reyanhridoyj7500@gmail.com

3)Sumaya Mahajabin Khan

Department of Electrical Engineering, Minnesota State University, Mankato,
Mankato,
Minnesota, USA
Email: Sumayamahajabin19@gmail.com

4)Yasir Arafat Ayan

Department of Computer Information Science Minnesota State University,
Mankato, Mankato,
Minnesota, USA
Email: yeasirarafatayan007@gmail.com

Abstract

This study investigates the application of exploratory data analysis as a tool to facilitate automated defect diagnosis and reliability engineering of advanced industrial automation systems. The research is based on a quantitative exploratory design with the help of the Predictive Maintenance Dataset with Irregularities to explore the machine operating conditions and failure patterns. The focus of the analysis is on the distribution of machine failures, variations in operating parameters, and relationships between the operating parameters of the machines. These results show that the dataset is dominated by normal operating conditions, and the failure events make up a relatively small fraction of the dataset. The most common fault categories are heat dissipation failure and overstrain failure. It is discovered through comparative analysis of the operating conditions that various types of failure have different characteristics in operation, specifically in terms of torque, rotational speed, tool wear, temperature, etc. The air temperature and process temperature are highly positively correlated, and the rotational speed and torque are highly negatively correlated. Failure status is correlated with air temperature and torque, with weak to moderate positive correlations, which could be used as diagnostic criteria. Air temperature and torque are correlated with failure status with weak to moderate positive correlation, indicating that they could be used as diagnostic variables. The results suggest the importance of exploratory analysis to detect patterns of operation and factors that repeat as failure in an industrial setting. The study is an important contribution to reliability engineering by providing evidence-based information that could help maintenance personnel in detecting problems early and making informed decisions, as well as help design future automated defect diagnosis frameworks.

Keywords: Defect, Diagnosis, Reliability, Engineering, Industrial, Automation, Predictive, Dataset

I. Introduction

A. Background of Study

The growing trend in modern industrial automation is toward smart machines, sensors, control systems, and data-driven technologies that are interconnected to enable efficient and reliable manufacturing operations. As Industry 4.0 technologies have been adopted, manufacturers are able to gather vast amounts of data about their operations and leverage that information to monitor equipment conditions, identify abnormal equipment operating behaviour, and improve production performance. Researchers shed light on predictive maintenance, and Industry 4.0 explains that data from industrial machines can be used for condition monitoring, fault detection, maintenance planning, and minimizing unplanned machine failures [1]. In automated factories, equipment failures and defects can cause disruptions to manufacturing, maintenance needs, and impact product quality. Thus, the capacity to interpret and understand data related to machines and identify early failure patterns has grown in significance for ensuring reliable industrial operations. Historical and real-time data can also be used to determine abnormal operating conditions and patterns that may lead to failure, which can be accomplished by using automated defect diagnosis to help maintenance personnel. New strategies like data analytics and machine learning will enable the shift from reactive maintenance to proactively managing reliability. As part of this, exploratory analysis of industrial machine data can be useful in uncovering relationships between operating conditions, machine failures, and possible contributing factors. This analysis can be the basis of evidence-based development of automated solutions to aid in the diagnosis, troubleshooting, and reliability enhancement of advanced industrial automation systems.

B. Problem Statement

Machine failure, production defects, unexpected downtime, and recurring operational issues pose constant reliability problems in industrial manufacturing systems, especially when affecting continuous production. A piece of research on the topic of predictive maintenance for industrial applications stressed the need of sensing equipment status and detecting unexpected patterns and failure situations to minimize unexpected failures and maintenance disruptions [2]. Manufacturing may fail due to various operating conditions, which can be a challenge for the maintenance operator to determine the exact condition of equipment failures. The occurrence of defects can also lead to loss of product quality and result in further inspection, repair or replacement needs. An unexpected failure in a machine can disrupt manufacturing schedules and raise the operating cost. The troubleshooting process can be complicated by a number of parameters in the machines changing simultaneously, or by not having the appropriate operational data analyzed systematically. Producing the product repeatedly presents a problem, as the patterns that may lead to failure must be studied to determine the source of the problem. The traditional reactive approach may thus offer only a partial solution to supporting early identification of emerging problems. A data-driven approach is required to analyze the data of industrial operations, to characterize meaningful failure patterns, to detect potential factors related to defects, and to facilitate more systematic diagnosis. Automated analysis may be a way to help make machine data useful for making decisions for reliability.

C. Aim of the Research

The purpose of this research is to create a data-driven method for automatic fault diagnosis based on the analysis of machine operating data, failure pattern identification, and factors influencing industrial reliability.

D. Objectives of the Research

The objectives of the research are mentioned below:

1. To create an automated, data-driven process to detect manufacturing defects and possible causes.
2. To study the pattern of failures and operating conditions of the machines, which may facilitate automated defect diagnosis.
3. To investigate the use of identified failure-related patterns for better reliability on an automated industrial system.
4. To present evidence from exploratory data analysis that can be used to help identify equipment issues and corrective actions more quickly.

C. Research Question

What role can exploratory analysis of industrial machine data play in automated identification of manufacturing defects, possible contributing causes, and increased reliability of automated industrial systems?

F. Significance of the Study

The value of this study is understood from the fact that industrial machine data can be explored to enable automated defect diagnosis and decision-making for reliability. The study reveals patterns that occur during manufacturing defects and equipment failures instead of only troubleshooting reactions. The study uses exploratory analysis to uncover and categorize failure conditions, to investigate relationships between machine operating variables, and to identify potential causes of repeat failures. These patterns can be helpful for diagnostic purposes to determine why specific failure conditions might be present under specific operating conditions. The findings can also help maintenance and engineering staff to systematically detect abnormal conditions and pinpoint parts that need further investigation. The research can help to streamline operations and translate operational data into meaningful visual and analytical evidence, which can aid in quicker identification of potential issues and more intelligent corrective measures. Subsequently, the assessment offers a data-driven basis for enhancing reliability and minimizing the effects of manufacturing failure on system operation in automated manufacturing environments.

2. Literature Review

2.1 Industrial Automation Reliability and Manufacturing Failure Challenges

Industrial automation has moved the notion of reliability engineering away from the evaluation of single components to the study of cyber-physical systems, which include mechanical, electrical, software, communication, and control elements that can interact with each other. Tordeux et al. [3] state that, besides new opportunities arising from the implementation of sensor-based monitoring, artificial intelligence, digital twins, or predictive maintenance, the complexity of systems, the volume of data, and the need for real-time processing also make reliability assessment more difficult in Industry 4.0. This generates a crucial conflict for a manufacturing setting. Failures may be more observable when there is more connectivity, but the data they generate could also be heterogeneous, dynamic, nonlinear, and hard to interpret. Soleimani et al. [4] classify the major challenges in complex engineered system diagnostics as variable operating conditions, non-stationarity, nonlinearity, and uncertainty. This is a very relevant review as it demonstrates that there is no single diagnostic approach that is universally suitable. Data-driven methods can make use of very rich sets of operational records, but might be dependent on representative data and might not offer any causal explanation. Physics-based approaches give a better mechanistic interpretation, but may be challenging when the behaviour of a system or interactions between its components are not well understood. For this reason, reliability issues no longer only involve failure detection, but also the identification of analytical method(s) that will be credible in the face of varying operating conditions. This is particularly crucial in automated factories where production states vary and equipment may exhibit different sensor patterns with different process or load situations.

Further, reliability issues arise from the organisation of maintenance and troubleshooting activities of industrial operations. The effectiveness of reliability-centred maintenance cannot be seen just as a technical analysis exercise as it is affected by the asset management discipline, quality practices and the maturity with which the asset management processes are implemented, as argued by Anick and Kabir [5]. They found that there is a correlation between improved execution maturity and reported reliability results, such as less downtime and fewer repeat failures, which indicate that achieving improved reliability requires making maintenance knowledge more consistent in practice. This view is crucial for automated manufacturing because a technically advanced diagnostic system can be of limited use if issues identified are not related to maintenance planning and follow-up corrective measures. From a more technology-oriented point of view, Shamim and Ruddro [6] review the PLC integration and smart diagnostics in CT tube manufacture. They recognize recurring degradation mechanisms and demonstrate how vibration, acoustic, thermal, and electrical signs and symptoms can offer information for condition assessment. But they also point to interoperability, data governance, and regulatory needs as challenges to implementation. The question here is not so much whether the sensors or diagnostic algorithms are able to detect abnormal conditions, but how they can be combined with existing control, maintenance, and governance processes without adding to the operational complexity.

The manufacturing literature also suggests that reliability problems frequently stem from equipment design, integration, testing, monitoring, and troubleshooting, more often than not, not from individual failures. Chowdhury and Nuruzzaman [7] highlight the integration of modular equipment, dynamic testing, fault detection, monitoring using sensors, and the linking of predictive maintenance with maintenance and enterprise platforms. Their review can prove to be useful in determining the overall infrastructure necessary to support reliable automated equipment, especially when legacy equipment, varied communication protocols, and multiple technology generations exist in the same environment. Meanwhile, the width of the proposed integration framework shows a practical limitation. Technologies do not necessarily lead to more reliable diagnosis since interoperability of data, quality of testing, diagnostic delay, and integration of maintenance workflows can contribute to the usefulness of the information generated. This worry is in line with the trend of reliability problems observed throughout Industry 4.0, which has increased complexity with the added benefit of monitoring. In the current research setting, the important conclusion is that manufacturing failures should be considered as identifiable patterns in machine operating data, not as individual incidents. Distributions, relationships, abnormal operating conditions, and recurring failure patterns can be spotted using exploratory analysis that may suggest potential contributing factors. This kind of evidence can be used to enable automated defect diagnosis, whilst maintaining an appropriate separation between statistical association and confirmed physical root cause. This is important to note because exploratory relationships can help lead to further investigation, but don't be presented as evidence of causation.

2.2 Automated Defect Detection and Failure Diagnosis in Industrial Systems

The shift from traditional “rule-based” defect detection to AI-based techniques, which can be trained to identify complex patterns from machine-generated signals, has become more prevalent. Bhuiyan and Uddin [8] state that measurement of vibration and acoustic signals is important for intelligent fault diagnosis, as defects can cause characteristic changes in measurable signals. Their analysis reveals their use of deep transfer learning to transfer learned representations between similar diagnostic tasks in order to decrease reliance on large task-specific datasets. This is especially applicable to an industrial environment where failure data that is labeled is more difficult to come by, as serious failures are less common than normal operating conditions. However, the actual usefulness of transfer learning is limited by the similarity between the source and target domains. Transfer of representations can be less reliable if machines are not configured exactly the same, if the sensors are not in the same position, if the operating load is different, if there is different noise, or if there are different fault characteristics. The review thus suggests that automated diagnosis cannot be considered as an algorithm selection problem; therefore, more complex challenges arise. The accuracy of the sensor, the representation of the features, the variation between different domains and the availability of representative examples of the faults are all factors that may affect the ability of an automated system to discriminate between true faults and normal operating variation.

Signal-based diagnosis offers a more direct link between the machine's behaviour and automated diagnosis of the fault. Iqbal and Madan [9] show how the CNC machine's bearing faults can be detected using vibration and acoustic signals along with a CNN. To generate the TFRs, they applied a short-time Fourier transform to raw signals before feeding them into a convolutional learning module for classification, and they reported very high classification rates. The study demonstrates the benefit of integrating several signal sources as vibration and acoustic signal measurements can provide complementary indications of bearing degradation. However, the quoted performance should be understood in the context of the dataset used for the experiments and not understood as typical for industrial environments. Controlled experiments typically yield better fault signatures than production environments with the noise of sensors, changing loads, machine interactions, and potentially unobserved operating conditions, all of which can make fault classification more difficult. This restriction points out an important difference between the definition of fault in given experimental conditions and the automatic diagnosis in continuously changing manufacturing systems. However, exploratory analysis of operational variables is still useful as it can help determine if there are distinct changes in machine conditions related to failure categories before more comprehensive diagnostic models are applied.

The other significant problem is how to divide up the available set of fault data. Ileri et. al [10] show that data balancing and feature normalization can significantly influence the classification of faults

automatically, and their empirical results indicate that use of balanced datasets and one-dimensional deep learning yields good results. The study was important since it was rare for the industrial failure data to be evenly distributed between all possible failure categories. Observations made while the machine is running can greatly outnumber the observations made with specific defects, leading to better recognition of common class observations and poorer recognition of rare defect observations by a learning algorithm. They have repeatedly developed balanced datasets that show that classification performance may vary depending on the structure of the training data, and thus high performance does not automatically lead to a robust diagnostic capability. Methodological implications are that the class distribution and the proper processing of the data must be carefully analyzed with automated defect detection prior to the interpretation of the classification results. In exploratory industrial research, the frequency of failure states and the distribution of operating variables can therefore be examined to determine whether what appear to be patterns are real machine behavior or whether they are due to unequal representation of failure categories. This preliminary evidence can be especially valuable when the goal is to determine diagnostic indicators, rather than to determine readiness for deployment of a predictive model.

Another key factor in AI-based diagnosis is the division between detection and feature extraction and classification. Lang et al. [11] define fault diagnosis as feature extraction and fault classification and summarize the traditional machine learning and deep learning methods in motor fault diagnosis. They conclude that signal processing still plays a critical role due to the impact of the extracted information on further classification. This defies the notion that more complex deep-learning models will solve increasingly complex diagnosis challenges. A model can have good classification accuracy with little insight into what physical or operational properties produced the classification. In an industrial reliability application, this interpretability problem becomes important since maintenance staff must have some idea of the abnormal condition to determine a suitable corrective action. Although automated diagnosis is of most assistance across the reviewed approaches, it can be useful when the classification is linked to interpretable machine behaviour, representative operating data, and clearly defined fault categories. This facilitates an exploratory way of working where machine variables are first explored for distributions, relationships, and failure-specific patterns, where only potential diagnostic indicators are identified without assuming these statistical associations are physical causes.

2.3 Data-Driven Failure Pattern Analysis and Root-Cause Identification

In the analysis of failure patterns, one increasingly has to make use of meaningful relationships that can be derived from multidimensional observations of the sensors rather than the sole use of fault rules that have been pre-defined. Rasekh et al. [12] argue that classification using artificial neural networks can convert the intricate response of the sensors to patterns, which can be used for accurate separation of the various conditions. The researchers found that in one type of analysis, a smaller, optimal subset of sensors provided better classification accuracy than did their full sensor array. It's significant for failure analysis because it means that the more data provided by the sensors, the less diagnostic information can be gained. Rather, informative variables can be found to eliminate redundancy without losing the desirable features for condition discrimination. The focus of the study, however, was on classification of essential oils and not on mechanical/industrial failures, which makes it more difficult to directly transfer to situations where faults occur in a gradual manner, and where the signal from the sensors may have any kind of dynamic interactions. The methodological contribution of this study is thus to illustrate the power of pattern recognition and feature selection for the identification of discriminative signatures, and to point out the importance of domain-specific validation when using these techniques for defect diagnosis in industry.

In Tao et al. [13], multi-dimensional sensor information is especially useful if the deterioration manifests over a series of related wear indicators instead of one measurable fault signal. In wind turbine gearboxes, they introduced a novel method that integrates several wear sensors and a multidimensional Transformer network for online monitoring and warning, highlighting the need to account for dependencies among different wear-sensor dimensions and to continuously monitor equipment condition. This approach takes failure analysis beyond purely threshold-based detection by employing learned representations of changing wear conditions to assist with the warning decision.

This is important for root cause analysis as it can provide information about how abnormal conditions are developing and changing over time and across dimensions. However, if a monitoring model can be developed to accurately detect abnormal wear, it does not necessarily explain the physical process causing the wear. This distinction is crucial since it indicates different needs for reliability engineering: predictive accuracy and causal interpretability. A developing failure pattern can be identified in a model without requiring engineering knowledge, and without needing to use other analytical methods to determine if the cause of the failure is lubrication, component degradation, operating conditions, or sensor failure.

Another difficulty is when the abnormal observations are from the sensing system itself, not from the equipment under monitoring. Haldimann et al. [14] tackled this issue using a disentangled recurrent neural network aimed at detecting multiple sensor faults and minimizing the misleading propagation of one sensor fault to other sensor residuals. Utilization of analytical redundancy and probabilistic residual modelling offers a more structured approach for identifying process abnormalities from measurement-related anomalies, and the linear computational complexity reported for increasing numbers of sensors allows for scalability. However, it is not just a classification task: Bali et al. [15] also show that Random Forest prediction can be used in a sensor-based monitoring environment, and that predictive modelling can be linked to operational energy optimisation. Based on these studies, it is concluded that fusing pattern recognition, temporal modelling, sensor fault isolation, and predictive decision support will be useful in the application of data-driven diagnosis. However, they also highlight one of the key analytical needs for automated defect diagnosis: if a pattern is detected as abnormal, then the variables of the detected pattern must be meaningful to be interpreted. This needs to be done through models and exploratory analyses that can link observed failure signatures to possible contributing factors, and not just by accurate prediction of the failure.

2.4 Predictive Maintenance and Reliability Improvement Through Data Analytics

Predictive maintenance has grown from scheduled maintenance to real-time equipment condition assessment based on data, especially in an Industry 4.0 environment. In the field of predictive maintenance, Achouch et al. [16] defined it as a process that combines sensing technologies, data acquisition, communication systems, machine learning, and maintenance decision-making to predict the degradation of the equipment before reaching functional failure. They note the shift from conventional corrective and preventive maintenance to models that leverage past and current data to help predict equipment condition and useful life. It was found that the implementation of predictive maintenance is not only limited to the choice of learning algorithm, but also to the quality of the data, the features from which it is derived, the placement of the sensors, and the availability of failure data. This has a practical limitation as industrial failures may be infrequent, are diverse, and vary depending on operating conditions. Thus, accurate models obtained by using a small set of data might not be as reliable when applied to a different production environment. Predictive maintenance thus relies on the ability of analytics-based predictions to be integrated into the operational context and not considered a standalone maintenance solution.

The importance of data analytics is even greater when equipment condition has to be continually monitored and when abnormal behaviour has to be detected early. As per Yousuf et al. [17] the concept of IoT-based health monitoring of AC induction motor is the integration of sensor data acquisition and fault detection for predictive maintenance of the motor. This integration enables electrical and operational parameters to be monitored continuously, providing opportunities to identify deviations and potentially prevent major equipment issues. The approach is also applicable to reliability improvement since continuous monitoring will diminish the need for fixed inspection periods, and condition information will be available to maintenance personnel near the time of deterioration. Furthermore, the analytical difficulty doesn't end there: it's important to figure out if the motor is running in a normal or abnormal manner. Various faults can create similar sensor patterns, and load, temperature, and operating conditions can also affect measurements. So, to make a good fault detection model, it requires enough context so that it can differentiate between a real degradation and a normal operating variation.



This is particularly relevant for industrial systems that rely on automated processes, where false alarms may trigger unnecessary maintenance work and lead to higher operating expenses.

Another basis for the connection between data analysis and maintenance decisions is condition-based monitoring. According to Ali and Abdelhadi [18], condition-based maintenance is based on the assessment of health indicators of equipment, which is done continuously or periodically and maintenance is to be initiated based on the assessment of these indicators. This method can help to optimize the use of resources; maintenance can be linked to signs of deterioration instead of being carried out in line with elapsed operating time. However, it is crucial that the indicators of a condition are selected correctly and that changes in conditions are understood correctly in order to be effective. A measurement can signify deterioration, but it doesn't necessarily reveal the cause of failure, and a fault that occurs slowly might be hard to tell from normal variability. This has led to a need for analytical methods to uncover meaningful patterns across multiple variables and not just using thresholds. The difficulty is compounded if there are interconnected components within any industrial system, where degradation of one component can impact the behaviour of another and can lead to indirect monitoring of failure. Data analytics must thus be able to assist in recognising conditions and interpreting the evolving behaviour of equipment.

In heavy machinery, the situation is even more complex as there are many different kinds of sensors delivering very diverse data; this data must be processed in a systematic way before a predictive model can provide valuable information. Hassan et al. [19] investigate the processes of data processing in predictive maintenance and highlight the significance of preprocessing, feature extraction, dimensionality reduction, data transformation, and other analytical techniques in pre-processing industrial data for predictive applications. The processes have a knock-on effect on the trustworthiness of the downstream models, as noisy, incomplete, redundant or poorly structured observations can mask downstream degradation patterns. One key methodological issue is that more complex predictive models will not compensate for poor data preparation. In the case of heavy equipment, sensor signals can be even more complicated to interpret due to variations in operating loads and environmental factors, and direct relationships between the operating conditions and the failure of one or more of the variables may be difficult to establish. This means there is a need for a set of analytical workflows which integrate data quality assessment with pattern identification and predictive modelling. In the field of reliability engineering, these workflows can offer a more robust foundation for the detection of new abnormal conditions, identification of associated variables, and help in maintenance decisions while keeping the focus on the uncertainty of real industrial operating data.

2.5 Research Gap

Industrial reliability is well documented in the literature, as is defect detection by automatic vision and signal-based fault diagnosis, as well as data-driven classification. However, not much emphasis is put on the use of exploratory analysis of operational variables, failure modes, and the relationships amongst the machine parameters to systematically diagnose and identify potential failure root causes. This creates a gap between classification of faults and interpretable reliability analysis.

III. Methodology

A. Research Design

This study uses a quantitative exploratory research design using secondary data that is industrial machine condition data from the AI4I PMDI dataset. The research neither develops nor trains a predictive machine learning model. Rather, it employs exploratory data analysis to uncover the distribution of machine failures, variability of operating conditions, relationships between pertinent variables, and patterns of abnormal machine states. The analytical approach is geared toward the overall goal of automated defect diagnosis and reliability engineering of discovering patterns that can be observed and used as useful diagnostic cues. Descriptive and graphical methods are used to determine differences between normal and failure-related observations on both numerical and categorical variables. Special focus is placed on repeated patterns and relationships that can be a sign of potential failure contributing factors. The results of exploratory analysis are used to identify statistical associations and not experimental mechanisms yet, so they may be used to guide subsequent diagnostic investigations, but should not be considered as proof of experimental mechanisms.

B. Data Source and Dataset Description

The study is conducted on the AI4I PMDI, which is a Predictive Maintenance Dataset with Irregularities along with an industrial machine condition observation dataset that includes operational and failure data [16]. The dataset was chosen because its structure makes it possible for exploratory investigation of machine behavior during various operating conditions, and it contains information potentially useful for research on equipment failure and predictive maintenance. The data include numerical information about the machines that are used to describe the machine operating characteristics, plus categorical information describing the condition of the machines and information describing machine failure status [16]. These variables offer the opportunity to explore the individual distributions and relationships between operating variables and failure conditions. Irregularities are significant with regard to the present research as they are likely to appear in real industrial monitoring data, making it difficult to interpret the machine behaviour simply. The data set thus serves as a sound foundation to study failure distributions, to seek out operational variables, to compare machine conditions and to examine relationships between variables, and to determine combinations of conditions that may be useful for automated defect diagnosis and reliability evaluation.

C. Data Preparation and Cleaning

The data were preprocessed in R to improve consistency and understanding before exploratory analysis. The CSV file was first loaded into the R environment, and the structure and data types of the variables were examined to determine which were measurements and which were attributes. There were no missing values or duplicate observations detected in the dataset, and this did not impact on any further visualizations. Observed ranges and distributions of the numerical variables were also examined for unusual or irregular observations that needed to be addressed when interpreting the results. Where necessary, categorical variables were recoded into their appropriate factor form to accurately depict the conditions of the machine and failure categories in the graphical analyses. There are no transformations were applied only to obtain a desired analytical pattern. Preparation instead was geared toward maintaining the information in the original observations and structuring the variables in the proper way for exploratory visualization. This cleaned data was then analyzed for descriptive statistics and graphical exploration of failure patterns in machines.

D. Exploratory Data Analysis and Visualization

The main analytical method employed to analyze the machine condition data and detect patterns for automated defect diagnosis was exploratory data analysis. Graphical representations were created using R to provide complementary analyses of the characteristics of failure observations. The analysis was broken down into four dimensions, which started with the distribution of failure conditions to define the prevalence and structure of failure conditions seen. The second dimension explored the differences in operating conditions to determine if the machine parameters show differences for normal conditions and failure-related observations. The third dimension explored relationships among relevant numerical variables for the purpose of discovering more robust and more fragile relationships that might be worth focusing on in the reliability analysis. Patterns of failure conditions were explored in the fourth dimension to determine if multivariable patterns could be used as additional diagnostic information for failure conditions and whether combinations of machine characteristics can be found that provide this information. This visualization approach enables examination of the intricate relationships without any assumption based on a pre-established prediction model.

E. Analytical and Interpretation Framework

Each exploratory result is related to the research goal, which is to understand the patterns of failure in industry and possible diagnostic indicators. Failure distributions are used to determine the distribution and frequency of abnormal machine conditions, and comparisons of operating conditions are used to determine patterns of abnormal conditions associated with the different machine conditions. Associations between numerical variables are explored and specific parameters are examined to show meaningful statistical associations that may be the basis for further exploration. Multivariable visualizations are subsequently presented to investigate if it is possible to visually discriminate failure-related observations better when presented as combinations of machine conditions than when presented as individual variables.

Interpretation is done by making sense of the patterns that are observed in the dataset and the operational meaning of variables, where appropriate. Notably, no causal connection is assumed throughout the analysis; only statistical association. A variable that has been shown to be associated with failure is thus a possible contributing factor or diagnostic indicator, but no more. This framework aids in interpretation by evidence but does not lead to false causal claims.

F. Limitations

There are some limitations to be taken into account when considering the results of this exploratory research. The analysis is based on secondary data observations and thus depends on the variables, measurement conditions, data quality, and irregularities of the source data. The results could consequently be representative of the dataset, not necessarily of all the conditions that would be encountered in industrial automation systems. Exploratory analysis can detect associations or recurring patterns, but is not able to draw clear causal links between machine parameters and failures. Failure categories and possibly other irregular observations represented in the data might also affect what patterns are apparent in the visual analysis. Also, the dataset might not contain full details of the operation that lead up to every failure event, such as detailed maintenance history, environmental conditions, component age, or exact physical mechanisms. Thus, whenever patterns are detected, they must be taken as analytical evidence to help with further diagnostic investigation, but not as fully quoted generalizations about the failure of industrial machines.

G. Ethical Considerations

The secondary industrial data for this study are available to the public and are meant for use in research and analyses to lessen the concerns about the collection of personally identifiable data from human participants. The dataset and its originator(s) are properly cited and the data is utilized in an academic context. Ethical interpretation is important due to the fact that failure patterns in the industry can affect the maintenance decisions, continuity of operations, and reliability of equipment. The analysis does not, therefore, make statistical associations seem to be physical causes and does not make any physical claims without strong evidence for them. Readers are warned of any analytical limitations, irregular observations and uncertainties in a transparent way, so that they can make proper judgement on the findings. The exploratory procedures are also clearly explained to enable reproducing them and other investigators to appreciate the results obtained. The study is also free from selective interpretation of visual patterns, and it advocates reporting based on evidence instead of individual observation as definitive representations of the reliability of industries.

VI. Conceptual Framework



Figure 1. Conceptual Framework Diagram
(Source: Created by learner)

V. Results

A. Distribution and Characteristics of Industrial Machine Failures

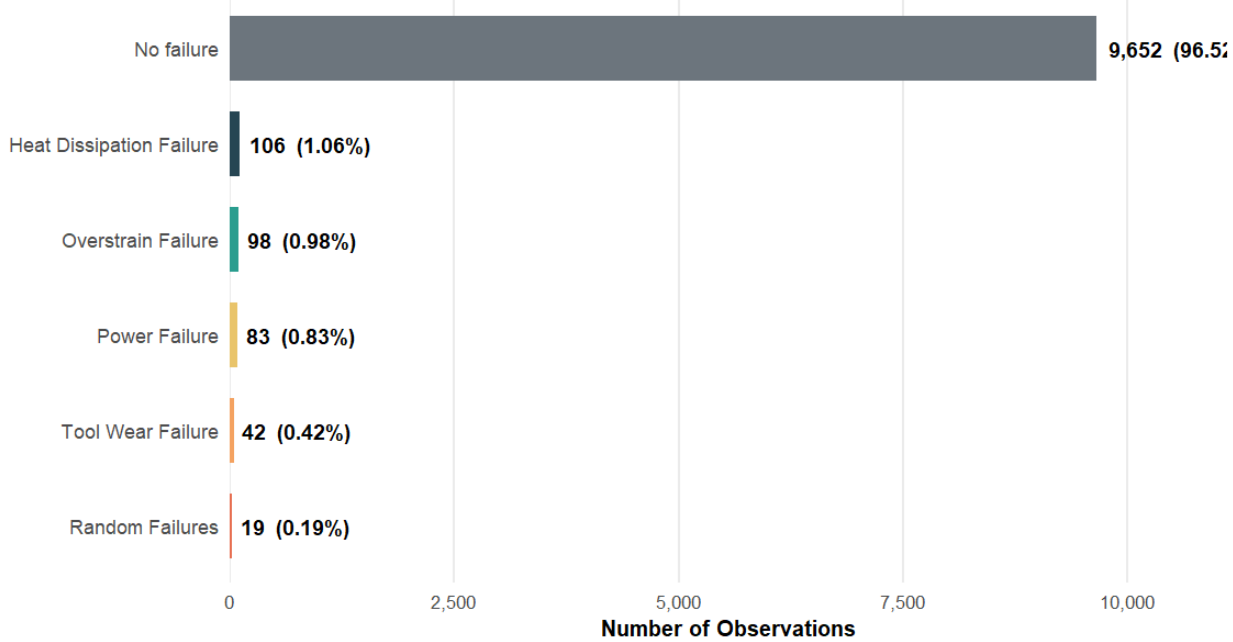


Figure 2. Frequency and Prevalence of Normal and Abnormal Diagnostic Conditions
(Source: Created by learner)

The results of the analysis of the states of failure of the industrial machines were extremely skewed distributions. The dataset consisted mainly of normal operating conditions (9,652 observations, 96.52%), and failure events combined accounted for less than 4% of the total number of observations. Failures due to Heat Dissipation (1.06%) and Overstrain (0.98%) were the most common failures, followed by Power failure (0.83%), Tool wear failure (0.42%), and Random failures (0.19%). These results suggest that even though failures of the machines are not frequent, there are clear diagnostic patterns that need to be monitored in order to have the best chance of catching them before they occur.

B. Operating Conditions Associated with Machine Failures

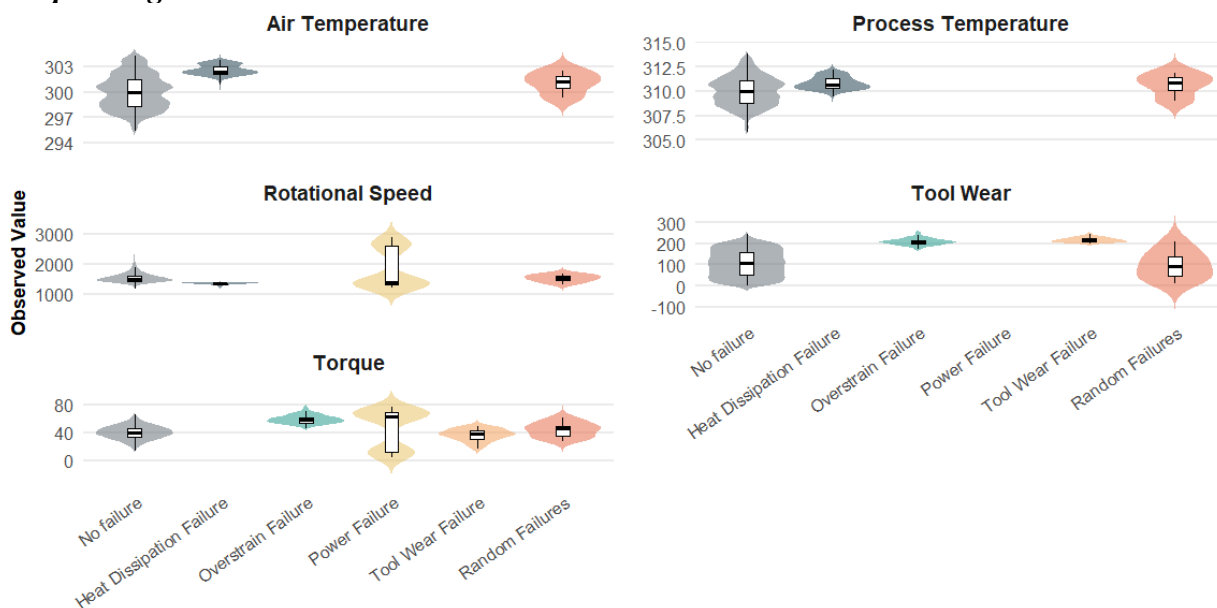


Figure 3. Distribution, Central Tendency and Variation of Key Machine Operating Variables
(Source: Created by learner)

There were significant differences in operating conditions between the categories of machine failure. Overstrain and tool wear failures showed high levels of torque and tool wear, while power failures showed the highest variance in rotational speed and torque.

The distributions of air and process temperatures within failure states were relatively small, with heat dissipation failures occurring at slightly higher temperatures. Several operating variables were found to be more variable for random failures, indicating less predictable behavior for these failures. These patterns show the significance of operational metrics for the identification of failure mechanisms.

C. Relationships Among Machine Variables and Potential Failure-Contributing Factors

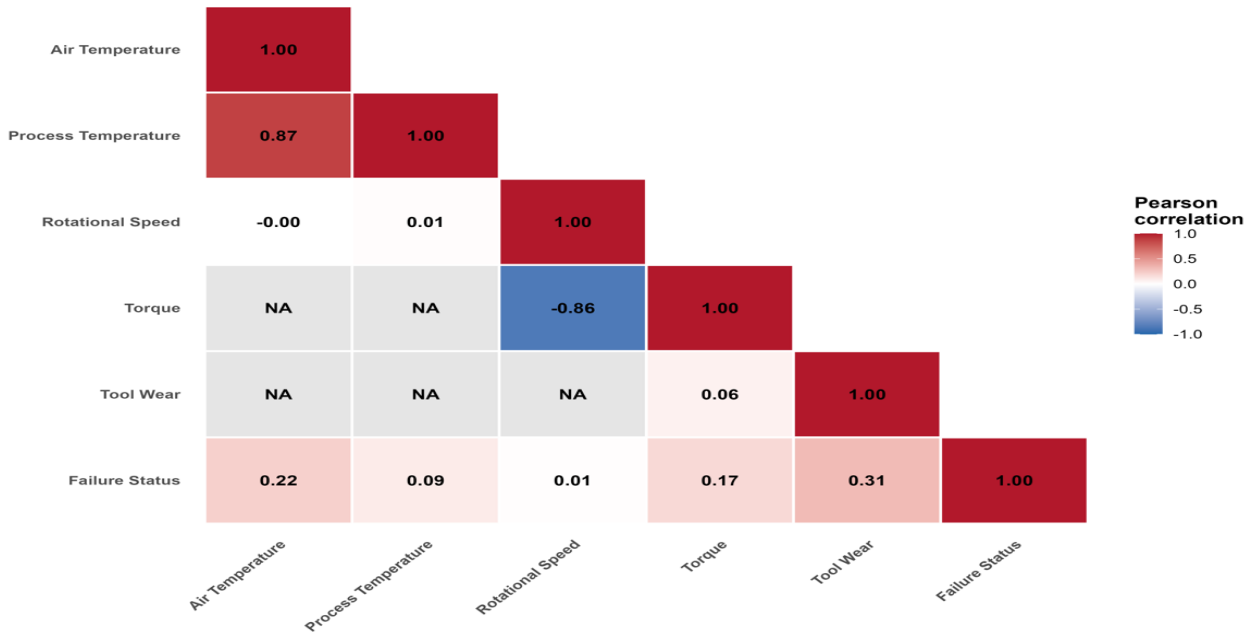


Figure 4. Pairwise Pearson Correlation Based on Available Observations
(Source: Created by learner)

Several significant machine variable versus failure status relationships were found using the correlation analysis. In the present study, air and process temperature showed a strong positive correlation between them ($r = 0.87$), which shows that there is a close relationship between the thermal conditions of the machine operation. There was a significant negative correlation between rotational speed and torque ($r = -0.86$), indicating that, in general, higher speed was correlated with lower torque. Tool wear, air temp, and torque exhibited weak-to-moderate positive associations with failure ($r = 0.31$, 0.22 , and 0.17 , respectively), indicating these as potential factors in the occurrence of a machine failure.

VI. Analysis of Results

The results show that machine failures are much less likely to happen than normal operating conditions, which indicates that the distribution of machine states is very unbalanced. Failure events are relatively rare but there are frequent failure categories, indicating that failure of the machine is not just a random event, but is related to patterns of machine operation. Failure of heat dissipation and overstrain failure appeared to be the most common abnormal conditions, and the rates of tool wear failure and random failure were relatively low. The number of observations in the normal category is large, indicating typical behavior of industrial systems in which the equipment is used for the majority of its lifecycle. However, the separation of the failure groups suggests that it may be possible to focus on specific conditions that are abnormal and that are systemically identified and monitored for diagnostic purposes. The operating conditions of the machines were analyzed, and it was found that there are unique operating characteristics of failure categories, which can be used to help diagnose defects. These differences between the failure states were found to be meaningful when comparing the distribution of temperatures, rotational speeds, torques, and tool wear. Results showed that overstrain failure had higher torque values, and tool wear failure had a higher degree of wear, which was apparent. Power failure was found to have a high degree of torque and rotational speed variation, indicating more unstable operating characteristics. Failure due to heat dissipation was observed with relatively higher thermal conditions than other categories. Random failures exhibited more variability in several machine parameters and thus a more complicated and less predictable operational profile. The observations indicate that certain operating conditions can be considered as diagnostic parameters for machine

failure mechanism classification and for reliability assessment.

The correlation analysis results add more support about the relationships between those machine variables and the relationship between the occurrence of failure and those machine variables. The high positive correlation of air temperature and process temperature is found to be good, which reflects the tight connection of operating temperature for the processes inside the industrial equipment. Likewise, the negative correlation between rotational speed and torque (for varying operational demands) is a mechanical characteristic that is well understood. Failure status was most strongly correlated with tool wear and next with air temperature and torque, which suggests that factors of degradation and operational stress can be of importance in the failure development. While these relationships do not imply causality, they do show statistically significant associations that can be used to help inform automated defect diagnosis and reliability engineering efforts. The findings of the different analyses collectively illustrate that exploratory analysis can reveal useful diagnostic patterns that can be used to help make maintenance decisions and provide more reliable industrial operations.

VII. Future Work

The current exploratory study could be expanded in future studies by creating and testing predictive machine learning models for automated defect diagnosis. Operational patterns identified can be classified with algorithms to assess their ability to predict the failure categories before the event. Other analytical methods like feature selection, clustering, anomaly detection, and ensemble learning can also be explored for enhancing the diagnostic performance. The use of temporal machine information may also help to elucidate the evolution of conditions of degradation and failure over time.

More research should examine the inclusion of other operational data, such as maintenance history, environmental factors, component age, and production factors. This information can give a deeper understanding of the failure behaviors that are observed and can help develop a more complete reliability assessment process. Validation with data collected from other industrial sectors would increase the generalizability of results and increase its applicability to practice. Further research could also examine the possibility of incorporating exploratory analytics and real-time monitoring systems together to facilitate ongoing assessment of equipment and to inform proactive decisions regarding maintenance. The developments could bolster the performance of computer-based defect diagnosing systems and lead to better operational reliability in today's industrial world.

VIII. Conclusion

In this paper, the use of exploratory data analysis for automated defect diagnosis and reliability engineering for advanced industrial automation systems was studied with the AI4I Predictive Maintenance Dataset with Irregularities. The analysis included a study of the failure distributions of the machines and the study of operating conditions and correlations between important machine variables to identify recurring patterns for abnormal machine behaviour. The results showed that machine failures constitute a small percentage of observations, but have distinct operational characteristics that can be used to help with diagnosing problems. Temperature, torque, rotation speed, and tool wear variations between failure groups indicated the significance of operational parameters to machine condition. The correlations between the selected machine parameters and failure status were also found to be meaningful and were identified by correlation analysis. Although the study does not establish causal relationships, it does show the importance of exploratory analysis for identifying diagnostic indicators based on evidence. The research offers a platform for future predictive maintenance programs and helps in developing more reliable, data-driven methods for monitoring and making maintenance decisions for industrial systems.

References

- [1] A. Sahli, R. Evans, and A. Manohar, "Predictive maintenance in Industry 4.0: Current themes," *Procedia CIRP*, vol. 104, pp. 1948–1953, 2021, doi: 10.1016/j.procir.2021.11.329
- [2] B. Taşcı, A. Omar, and S. Ayvaz, "Remaining useful lifetime prediction for predictive maintenance in manufacturing," *Computers & Industrial Engineering*, vol. 184, Art. no. 109566, 2023, doi: 10.1016/j.cie.2023.109566
- [3] A. Tordeux, T. M. Julitz, I. Müller, Z. Zhang, J. Pietruschka, N. Fricke, et al., "System Reliability Engineering in the Age of Industry 4.0: Challenges and Innovations," *arXiv preprint arXiv:2411.08913*, 2024, doi: 10.48550/arXiv.2411.08913.
- [4] M. Soleimani, F. Campean, and D. Neagu, "Diagnostics and Prognostics for Complex Systems: A Review of Methods and Challenges," *Quality and Reliability Engineering International*, vol. 37, no. 8, pp. 3746–3778, 2021, doi: 10.1002/qre.2947.
- [5] K. M. T. A. Anick and T. Kabir, "Reliability-Centered Maintenance of Electrical Power and Control Systems Using Manufacturing-Based Asset Management and Quality Models," *American Journal of Advanced Technology and Engineering Solutions*, vol. 2, no. 3, pp. 29–59, 2022, doi: 10.63125/xq6a0793.
- [6] M. M. R. Shamim and R. A. Ruddro, "Integration of PLC and Smart Diagnostics in Predictive Maintenance of CT Tube Manufacturing Systems," *International Journal of Scientific Interdisciplinary Research*, vol. 1, no. 1, pp. 62–96, 2022, doi: 10.63125/gspb0f75.
- [7] A. Chowdhury and M. Nuruzzaman, "Design, Testing, and Troubleshooting of Industrial Equipment: A Systematic Review of Integration Techniques for U.S. Manufacturing Plants," *Review of Applied Science and Technology*, vol. 2, no. 1, pp. 53–84, 2023, doi: 10.63125/893et038.
- [8] M. R. Bhuiyan and J. Uddin, "Deep transfer learning models for industrial fault diagnosis using vibration and acoustic sensors data: A review," *Vibration*, vol. 6, no. 1, pp. 218–238, 2023, doi: 10.3390/vibration6010014.
- [9] M. Iqbal and A. K. Madan, "CNC machine bearing fault detection based on convolutional neural network using vibration and acoustic signal," *Journal of Vibration Engineering & Technologies*, vol. 10, no. 5, pp. 1613–1621, 2022, doi: 10.1007/s42417-022-00468-1.
- [10] U. Ileri, Y. Altun, and A. Narin, "An efficient approach for automatic fault classification based on data balance and one-dimensional deep learning," *Applied Sciences*, vol. 14, no. 11, Art. no. 4899, 2024, doi: 10.3390/app14114899.
- [11] W. Lang, Y. Hu, C. Gong, X. Zhang, H. Xu, and J. Deng, "Artificial intelligence-based technique for fault detection and diagnosis of EV motors: A review," *IEEE Transactions on Transportation Electrification*, vol. 8, no. 1, pp. 384–406, 2022, doi: 10.1109/TTE.2021.3110318.
- [12] M. Rasekh, H. Karami, A. D. Wilson, and M. Gancarz, "Performance analysis of MAU-9 electronic-nose MOS sensor array components and ANN classification methods for discrimination of herb and fruit essential oils," *Chemosensors*, vol. 9, no. 9, Art. no. 243, 2021, doi: 10.3390/chemosensors9090243.
- [13] H. Tao, W. Feng, G. Yang, R. Du, and Y. Zhong, "A multiple wear sensors online monitoring and warning method in lubricating oil using multidimensional transformer network for wind turbine gearboxes," *IEEE Sensors Journal*, vol. 25, no. 2, pp. 3503–3519, 2025, doi: 10.1109/JSEN.2024.3504495.
- [14] D. Haldimann, M. Guerriero, Y. Maret, N. Bonavita, G. Ciarlo, and M. Sabbadin, "A scalable algorithm for identifying multiple-sensor faults using disentangled RNNs," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 3, pp. 1093–1106, Mar. 2022, doi: 10.1109/TNNLS.2020.3040224.
- [15] A. Bali, P. E. Hladik, H. Gandji, A. Gherbi, and M. Cheriet, "Integrating random forest prediction for energy optimization in solar-powered environmental monitoring," in *Proc. 2024 IEEE 27th Int. Symp. Real-Time Distributed Computing (ISORC)*, Tunis, Tunisia, May 2024, pp. 1–10, doi: 10.1109/ISORC61049.2024.10551350.
- [16] J.-V. Autran, "AI4I-PMDI: Predictive Maintenance Dataset with Irregularities," *Figshare*, dataset, Jun. 13, 2024. doi: 10.6084/m9.figshare.24718428.
[https://figshare.com/articles/dataset/AI4I-PMDI Predictive Maintenance Dataset with Irregularities/24718428?file=43463241](https://figshare.com/articles/dataset/AI4I-PMDI_Predictive_Maintenance_Dataset_with_Irregularities/24718428?file=43463241)