International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 08 | August - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Early Detection and Classification of Bearing Faults in Semiconductor Materials towards Sustainable Manufacturing Process

Prof. M K DHANANJIAYA, SAPTHAMI B K PRADEEP, SHASHIKUMAR, JANARDHAN

STUDENTS OF DEPARTMENT INFORMATION SCIENCE AND ENGINEERING

ACHARYA INSTITUTE OF TECHNOLOGY, SOLADEVANAHALLI. 560107

---***---

Abstract - Machine learning (ML) approaches have gained significant attention in recent years for their ability to identify patterns and anomalies in large and complex datasets. In the context of semiconductor manufacturing, ML techniques are being used for fault detection and classification to improve the overall quality of the manufacturing process. Fault detection is critical in ensuring that semiconductor materials meet the required specifications and standards, which is essential for sustainable manufacturing processes. ML approaches for fault detection in semiconductor materials involve the use of algorithms to analyze data from various sensors and measurements, including temperature, pressure, and vibration sensors. The algorithms can identify patterns in the data and detect any anomalies that may indicate a fault or defect in the manufacturing process. ML algorithms can also be used to classify faults and identify their root causes, which can help improve the manufacturing process and reduce waste. One of the most technologically challenging industrial processes is the production of semiconductors. Long-established machine learning techniques like univariate and multivariate analysis have been used to create predictive models that can identify failures. Predictive modelling has been the subject of large collaborative research programme over the past ten years between universities and fantastic businesses. In this article, we take a closer look at some of these study topics before recommending machine learning techniques to automatically produce an accurate predictive model to forecast equipment failures during the semiconductor industry's wafer production process. The goal of this research project is to create a decision model that will aid in immediately identifying any equipment breakdown in order to maintain high process yields in production.

1. INTRODUCTION

A few years ago, A field of control engineering known as "detection, isolation, and recovery of faults" is concerned with keeping an eye on a system, identifying when a fault has occurred, locating the flaw, and then taking the necessary steps to stop more failures or unfavourable incidents. Fault detection is essential in operations that are expensive and safety-sensitive because early identification stops unanticipated occurrences from happening. Many industrial sectors are interested in using fault detection and diagnostics to enhance various process performances. The warning symptoms, which can be visual, audible, or even olfactory, are indications that the machine or system is usually communicating before the real failure takes place. These signals are recognisable because they are wholly extraneous, do not correspond to the usual operation of the machinery, A group of classification algorithms built on the Bayes' Theorem are known as naive Bayes classifiers. It is a family of algorithms rather than a single algorithm, and they all operate under the same general idea. Every pair of traits being classed, thus, is unrelated to one another The Bayes Theorem determines the likelihood of an event occurring given the likelihood of an earlier event occurring. The following equation is the mathematical formulation of Bayes' theorem. $P(A|B)=P(B|A)P(A)/ P(B)$ Supervised machine learning algorithms like random forest are frequently employed in classification and regression issues. On various samples, it constructs decision trees and uses their average for classification and majority vote for regression.

Machine Learning approaches for fault detection in

Semiconductor Materials The Random Forest Algorithm's ability to scale is one of its most crucial characteristics. The supervised learning algorithms family includes the decision tree algorithm. The decision tree technique, in contrast to other supervised learning methods, is capable of handling both classification and regression issues. By learning simple decision rules inferred from prior data (training data), a decision tree is used to generate a training model that may be used to predict the class or value of the target variable.

2.PROPOSED SYSTEM

This data set is challenging to effectively analyse due to the imbalance in passed and failed samples as well as the volume of measurement data collected from hundreds of sensors. Therefore, creating a strategy based on machine learning techniques to create an accurate fault detection model is our key objective. Our study's feature selection methods range from the simple elimination of features with constant values and features with a large number of missing values (more than 55% of the values are missing) to statistical analyses like principal and square component analysis (PCA) and theoretical data like profit ratio.

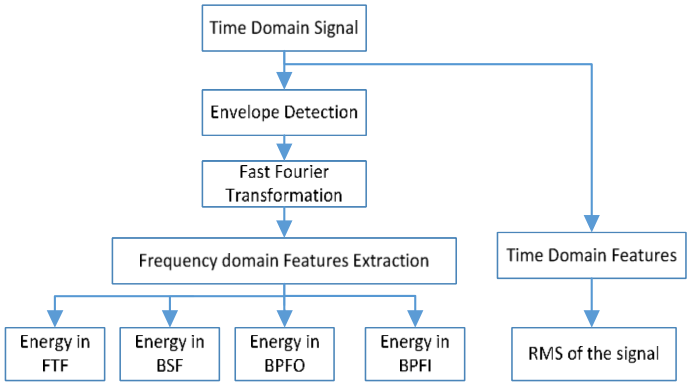
3.PROBLEM STATEMENT

To develop and implement a Predictive Maintenance technique (PdM) that is both cost-effective and efficient in order to minimise process tool downtime, schedule maintenance, and unscheduled downtime while maximising uptime. Advanced maintenance techniques are needed for larger, more complex building facilities in order to maintain the highest operating ratio. Most maintenance businesses that carry out routine inspections and repairs offer remote monitoring services and continuously check on the facilities' health in order to address the aforementioned issue. According to industry estimates, most capital equipment loses at least 8% of its productivity due to unplanned downtime and another 7% due to regular maintenance. The production of sophisticated semiconductors has long been hampered by the need to maximise process tool uptime.

4.Fault diagnosis

A. Signal Processing and feature extraction

The steps of signal processing and the feature extraction process are shown in Fig. 2. First, the time domain signal is collected, and the RMS of the signal is calculated. Then the Hilbert transformation is applied to detect the envelope of the time domain signal. Fast Fourier transformation is used to convert the envelope signal into the frequency domain. Finally, fault specific frequencies located in the frequency spectrum and energy associated with each frequency bands are extracted. Five fault cases were considered: healthy, inner race degradation (IR_D), inner race failure (IR_F), outer race degradation (OR_D), and outer race failure (OR_F). SVM classification algorithm is used to classify the faults.



B. SVM-based Classification algorithm

In the previous section, five features used to predict about bearing faults were selected. Since the status of bearings in each data sample is already known, it is possible to train a SVM classifier using collected data. However, only 80% of the collected data is used to train the classifier, and the remaining

$2 \quad s \quad D_c$

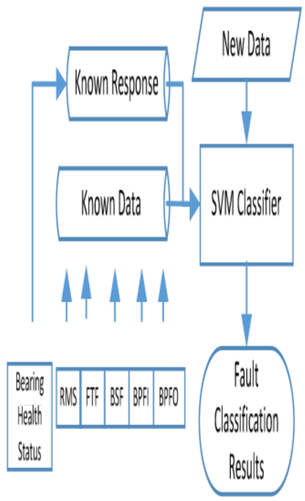
$$REF = \frac{D_c}{D_b} f_s \left(1 - \left(\frac{D_b}{D_c} \cos\theta\right)^2\right) \qquad (5)$$

N_b

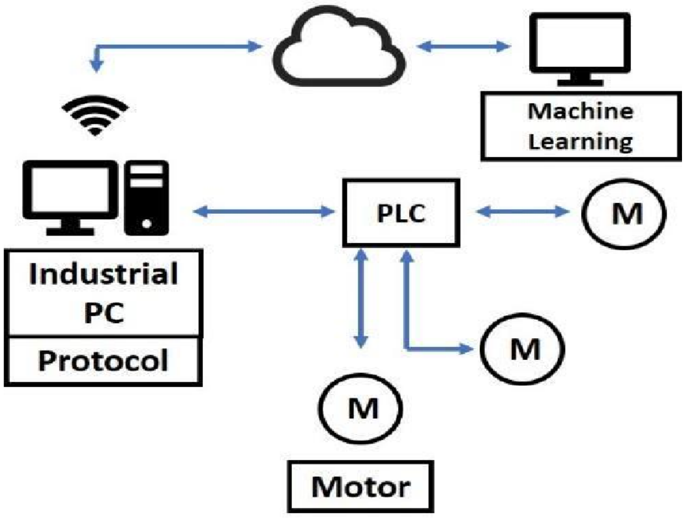
D_c

θ

f_s



5.System set up



A. Exploratory Data Analysis

The sensors send the data when the machine shows a change in state which usually is sampled per second. Hence, the data received consists of many Null values. The first step towards preprocessing [1] includes replacing the Null values with the sliding mean values. Since, the failure points in the data are less as compared to the data points representing good production cycles, we tried outlier detection using clustering and studied the outliers to gain insights.

B. System Setup

The PLC on the machine stores data for various parameters which need to be monitored. The concept of Daisy Chaining is used to reduce the wiring between the machines and PLCs. Using an RS485 port the data is sent to the adaptor which then converts the data into TCP form which is fed to the Industrial Personal Computer (IPC). The IPC is connected to the internet and pushes the data to the cloud using MQTT protocol in the form of data packets. The figure shows a block diagram for the process of data transmission from the motor (control of slitter rolls) to PLC to the IPC which is connected to the internet for pushing the data to the cloud

6.IMPLEMENTATION

The imbalance of pass and fail examples in addition to the large number of measurement data obtained from hundreds of sensors make this dataset a difficult one to accurately analyze. It is thus our main focus to devise a method based on machine learning techniques to build an accurate model for fault detection. Feature selection techniques in our study are ranging from simply removing features with a constant value and features containing numerous missing values (more than 55% of values are missing), to statistical based analysis such as chi-square and principal component analysis (PCA) and information theoretical based such as gain ratio. We also devise a cluster based technique call Average Diff to analyze discrimination power of each feature. On the model building phase, we apply four methods to induce the fault- detection model namely decision tree, naïve Bayes and k-nearest neighbor. The dataset is in a form of matrix; rows represent each observation or instance and columns represent features which are values recorded from each sensor.

1. Data Preparation Phase: This phase involves collecting and preparing the raw data for analysis. It includes tasks such as data acquisition, data integration, data cleaning, and data transformation. The main goal is to ensure that the data is in a suitable format and structure for further analysis.
2. Data Cleansing Phase: This phase focuses on identifying and correcting or removing errors, inconsistencies, and inaccuracies in the data. It involves tasks such as handling missing values, handling outliers, removing duplicate records, and resolving inconsistencies in data formats. The purpose is to improve the quality and reliability of the data.
3. Feature Scaling: Feature scaling is a preprocessing step that aims to standardize or normalize the features in the dataset. It ensures that all features have a similar scale, which can be important for certain machine learning algorithms. Common techniques for feature scaling include standardization (e.g., z-score normalization) and normalization (e.g., min-max scaling).
4. Feature Reduction: Feature reduction refers to the process of reducing the number of features in a dataset while preserving the most important information. It is often done to address the curse of dimensionality, improve computational efficiency, and prevent overfitting. Techniques for feature reduction include principal component analysis (PCA), linear discriminant analysis (LDA), and feature extraction methods like factor analysis.
5. Feature Selection: Feature selection is the process of selecting a subset of relevant features from the original dataset. The goal is to choose the most informative and discriminative features that contribute the most to the predictive performance of a model. Feature selection helps in reducing complexity, improving model interpretability, and avoiding overfitting. It can be done using various methods such as filter methods, wrapper methods, and embedded methods.

7.Disadvantage:

- Implementation Complexity: Implementing predictive maintenance systems can be complex and require significant investments in infrastructure, data collection, analytics tools, and skilled personnel. It may involve integrating various data sources, developing predictive models, and setting up monitoring systems.
- Data Quality and Availability: Predictive maintenance relies heavily on accurate and timely data from sensors, equipment, and other sources. In some cases, obtaining high-quality data can be challenging due to sensor limitations, data variability, or data accessibility issues. Lack of data or poor data quality can hinder the effectiveness of predictive maintenance.
- False Alarms and False Positives: Predictive maintenance systems may generate false alarms or false positives, leading to unnecessary maintenance activities or disruptions to production. Tuning the predictive models and

fine-tuning the thresholds is essential to minimize false alarms and ensure accurate predictions.

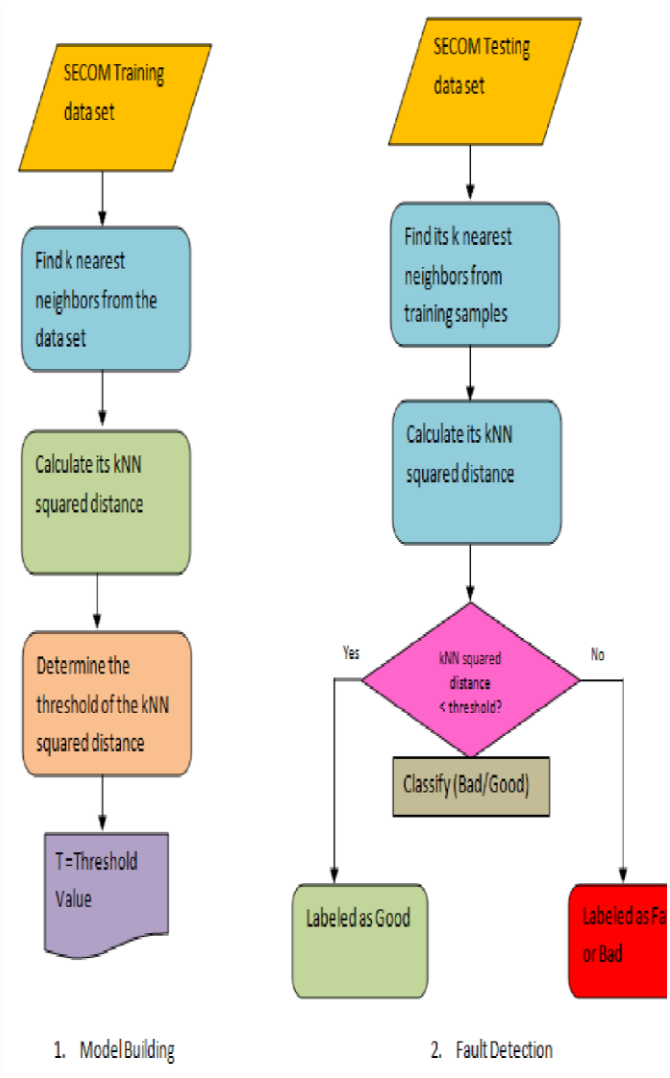
8.Advantage:

- Increased Equipment Availability: By detecting and addressing potential issues before they cause equipment failures, predictive maintenance improves equipment availability. This leads to increased production uptime and reduced production losses.
- Improved Product Quality: Equipment malfunctions or failures can have a negative impact on product quality in semiconductor manufacturing. Predictive maintenance helps in ensuring that equipment is functioning optimally, reducing the risk of defects and improving product quality.

9. Algorithm Used:

- K-Nearest Neighbor: The k-nearest neighbors (KNN) algorithm is a simple, easy-toimplement supervised machine learning algorithm that can be used to solve both classification and regression problems. The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other.
- Decision Tree: A decision tree is a supervised machine Learning algorithm that is flowchalike structure in which each internal node represents a "test" on an attribute (eg whether a coin flip comes up heads or tails), each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes). The paths from root to leaf represent classification rules.
- Naive Bayes Classifiers: In machine learning, naïve Bayes classifiers is a supervised machine learning algorithm that is a family of simple "probabilistic classifiers" based on applying Bayes 'theorem with strong (naïve) independence assumptions between the features. They are among the simplest Bayesian network models.

K nearest Neighbor



Due to the unique characteristics of the semiconductor processes, such as nonlinearity in the data, a predictive model using the k-nearest neighbor rule (PD-kNN) is developed in this paper. The kNN rule is an intuitive concept and its basic idea is given as the following: For a given unlabeled sample x, the kNN rule finds the knearest labeled samples in the training data set and assigns x to the class that appears most frequently within the k-subset (i.e., k-nearest neighbors).

The proposed predictive model using the kNN rule (PDkNN) is based on the fault sample's distance to the nearest neighboring training samples must be greater than a normal samples' distance to the nearest neighboring training samples. The idea is that to determine a threshold (*t*) with certain confidence level.

The proposed method consists of two parts:

1) Predictive Model Building

2) Fault Detection using Classification

1) Predictive model building

The kNN squared distance is defined as the sum of squared distances of sample *i* to its k-nearest neighbors.

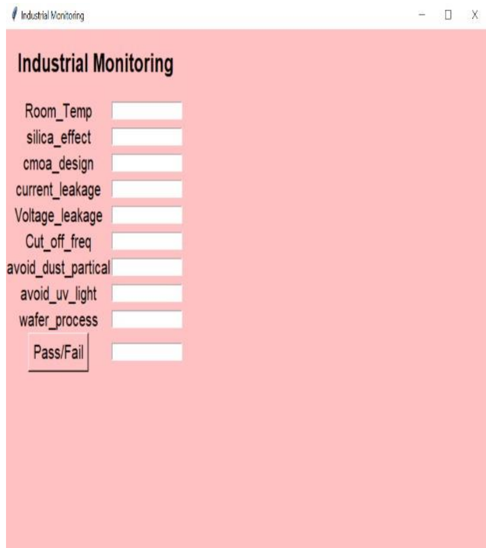
2)Naïve Bayes

A Naïve Bayes classifier is a probabilistic-based classifier that has been initially introduced by Duda and Hart [36] in 1973. SECOM data set is an imbalanced dataset and Navïe Bayes is a frequently proposed to imbalanced dataset problems. Naive Bayes induction algorithm is very popular in classification field. The Naive Bayes Classifier technique is based on the socalled Bayesian theorem and is particularly suited in Semiconductor industry, when the dimensionality of the inputs is high. Despite its simplicity.

3)Decision Tree

Decision tree induction is a powerful technique to discover a tree model for future event prediction. A decision tree is a tree-shaped structure that represents sets of decisions

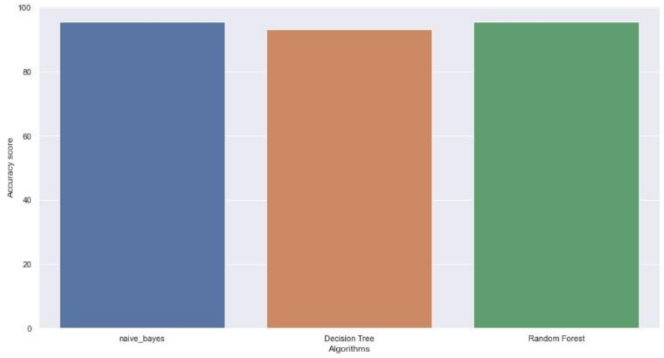
10.SNAPSHOTS



1. **FIG 1**

The use of predictive analysis proves to be a viable design solution for industrial machine prognostics. We have used Autoregressive Integrated Moving Average (ARIMA) [4] for forecasting the machine parameters to map the future states

2. FIG 2



Hence, this analysis helped us to map the failures with characteristic changes in the parameter values.

112.CONCLUSION

machine learning will help overcome major limitations in productivity and maintenance costs associated with it. The supervised models can be used to find insights from the data and the subsequent use of prognostics and forecasting will make sure that the production process runs efficiently with minimal costs incurred for maintenance and reduce product quality degradation.

ML techniques applied in PdM of industrial components.

Recent applications within the timeframe of ten years (i.e., 2010–2020) for several ML algorithms were reviewed and presented. Finally, some discussions have been drawn based on the literature review performed. It is observed that predictive maintenance has enormous market opportunities, and that machine learning is an innovative solution to predictive maintenance implementation. Yet, according to a PwC survey, only 11% of the

The occurrence of such phenomena leads to product quality degradation. Then a thorough analysis was carried out on the outliers as well as the failure points which were marked in the incoming labelled data to find the differences from the good cycles of production. Deeper analysis was done to understand the effects of the parameters on the product quality

13.REFERENCE

[1] P.K. Sethy, N. K. Barpanda, A. K. Rath, and S. K. Behera, "Deep feature based Rice leaf disease identification using support vector machine,"Comput. Electron. Agricult., vol. 175, Aug. 2020, Art. no. 105527

[2] V.Singh and A. K. Misra, "Detection of plant leaf diseases using image segmentation and soft computing techniques," Inf. Process. Agricult.,vol. 4, pp. 41–49, Mar. 2017.

[3] S.H. Lee, C. S. Chan, S. J. Mayo, and P. Remagnino, "How deep learningextracts and learns leaf features or plant classification 1-13, Nov 2017.

[4] S.Geetharamani and P. J. Arun, "Identification of plant leaf diseases usinga nine-layer deep convolutional neural network,' Jun 2019.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. Adv. Neural Inf.

Process. Syst. (NIPS), vol. 25, Dec. 2012, pp. 1097–1105.

[7] Ren, W. Liu, and G. Wu, ‘‘Feature reuse residual networks for insect pest recognition,’’ IEEE Access, vol. 7, pp. 122758–122768, 2019.

[8] V. Hoang Trong, Y. Gwang-hyun, D. Thanh Vu, and K. Jin-young, ‘‘Latefusion of multimodal deep neural networks for weeds classification,’’Comput. Electron. Agricult., vol. 175, Aug. 2020, Art. no. 105506.

[9] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, ‘‘Using deep transfer learning for image-based plant disease identification,’’ Comput.Electron. Agricult., vol. 173, Jun. 2020, Art. no. 105393.

[10] G. Farjon, O. Krikeb, A. B. Hillel, and V. Alchanatis, ‘‘Detection andcounting of flowers on apple trees for better chemical thinning decisions,’’Precis. Agricult., vol. 21, pp. 1–19, Aug. 2019.