The present work was submitted to the Institute of General Mechanics


**Rheinisch Westfälische Technische Hochschule Aachen**
**Faculty of Mechanical Engineering**


**Project Report**


# Project P06 – Group 08

# Intelligent Monitoring of the production system

## Intelligent Monitoring of Engineering Systems


| | |
|---|---|
| Presented by: | Praveen Rajamurugesan (453146), MME-CAME |
| | Gokul Visakakannan (452779), MME-CAME |
| | Somesh Kawale (452255), MME-CAME |
| | Yuvaraj Subramanyam (453330), MME-PS |
| Group: | 08 |

**Declaration**

We declare to the best of our knowledge that we have written this report by ourselves and only with the help of all references listed in the bibliography.

Aachen, July 24, 2024

| | |
|---|---|
| Praveen Raaj Rajamurugesan | |
| Gokul Visakakannan | |
| Somesh Kawale | |
| Yuvaraj Subramanyam | |

Content

# List of Figures

# 1 Introduction

Inefficiencies in the production line due to machine failure or external disturbances can have a huge influence on the systems and can result in delays, malfunctions, and increased operation costs. So, maintaining the efficiency of the production line is quite important in order to achieve maximum output and thus higher profits. Higher efficiency of production lines can be achieved by continuously monitoring the state of the system and implementing necessary corrective measures. This continuous monitoring can be done either by manual examinations, which can be time-consuming and also prone to human mistakes, or by implementing the use of various sensors and then using data obtained from these sensors to detect anomalies through the application of artificial intelligence and machine learning.

We have tried to achieve this in our project, by using a smartphone to collect various sets of data from a model of a production line and use this data to train a computational model which will be capable of detecting anomalies from real-time data obtained from a similar production line. We aim to eventually explore the potential of integrating various sensors and using advanced data analysis techniques to improve the robustness of a production line. By automating the disturbance detection process we mean to achieve higher product quality, reduce production downtime, and to optimize overall production efficiency.[1,2]

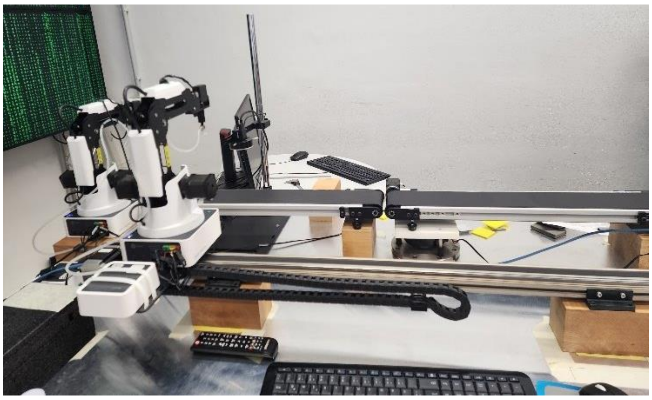

**Figure 1: Experimental Setup**

**Figure 3: Application of project in computer chip manufacturing**

**Figure 2: Experimental Setup**

**Division of responsibilities:**

| Member | Responsibilities |
|---|---|
| Praveen Rajamurugesan | - Algorithm implementation in MATLAB and Python<br>- Research of Time and Frequency Domain analysis<br>- Presentation, Report preparation |
| Gokul Visakakannan | - Study on Testing and Validating the Model<br>- Research of Time and Frequency Domain analysis<br>- Presentation, Report preparation |
| Somesh Kawale | - Study of Confusion Matrix ROC and AUC curves<br>- Research of Frequency domain analysis<br>- Presentation, Report preparation |
| Yuvaraj Subramanyam | - Study on Denoising methods<br>- Research of Time domain analysis<br>- Presentation, Report preparation |

# 2 Methods

## 2.1. Signal (pre-)processing

In our Project, we used a detailed signal pre-processing methodology to ensure the integrity and usability of the input data for our computational models. The signal data, collected from nine teams, which includes five cases per team. These signal data consisted of timestamps, Acceleration x, Acceleration y, Acceleration z and Total acceleration. All these signal data were fed into MATLAB, followed by resampling signal data from different sources to standardized sampling frequency of 200 Hz. Then further analysis of these signal was proceeded. To reduce noise in the time series data, a Gaussian filter with a window size of 10 and a sigma value of 2 was applied. These values for filter were tuned considering data loss and signal distortion. This filtering process significantly cleared the noise, enhancing the data interpretability and signal quality.

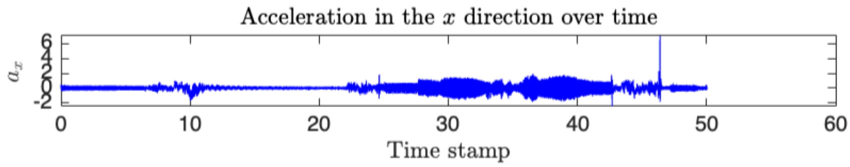


**Figure 3: Acceleration in the x-direction over time for case 8.1**

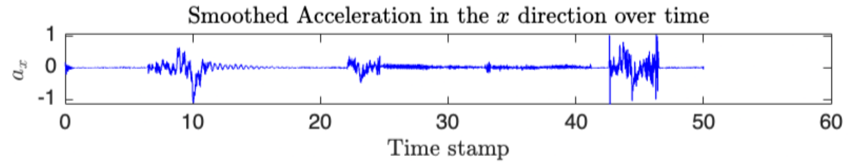


**Figure 4: Smoothed acceleration in the x-direction over time for case 8.1**

After noise reduction, we extracted key features from both time and frequency domains for examination.  performed data slicing and labeling, and used a random drop function to ensure an equal weightage of classes and split into input features and output target vectors. These pre-processing steps were also applied to the test dataset, to prepare it for computational model testing to predict these locations.

## 2.2. Computational Model

In this section, we discuss the computational models used and implemented to analyze and predict disturbances within the system. To locate frequency disturbances, we utilized both the Fourier Transformation Technique (FFT) based model and the Short-Term Fourier Transformation (STFT) based model techniques. These methods were chosen for their ability to effectively convert time-domain signals into the frequency domain, facilitating the identification and analysis of frequency-related disturbances. The FFT-based method yielded better results compared to the STFT-based model, as the resolution of the STFT model depends on the window size parameter, which can pose an issue.

For locating the damper and inclination, we implemented machine learning algorithms such as the Decision Tree Classifier and Random Forest Classifier. Due to its superior performance in our tests, we selected the Random Forest Classifier as the primary model.

# 2.3. Best-performing model

## 2.3.1. Model architecture

This section details the architecture and parameters of our most effective models. To determine frequency distribution, we first calculate the dominant frequency for the Acceleration Z channel in the given datasets. Using the FFT, we derive the single-sided spectrum value, from which the dominant frequency for the Z sensor channel is identified by locating the maximum value of the single-sided spectrum. Subsequently, a band-pass filter centered around the dominant frequency is applied, which identifies the peak in the signal and extracts the corresponding time.

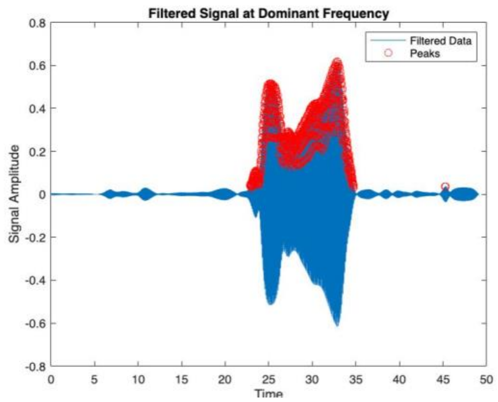


Figure 5: Filtered Signal at Dominant Frequency for Test Case 3

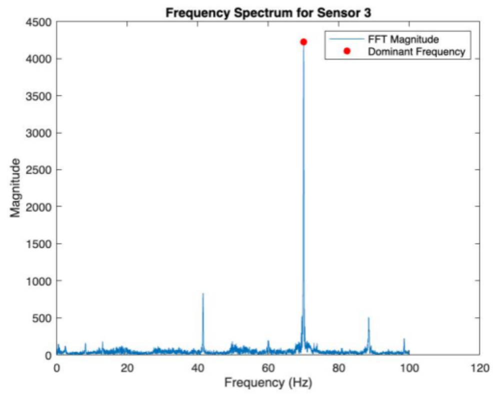


Figure 6: Dominant Frequency for Test Case 3

To determine the location of the Damper and Inclination, we employed a methodology involving machine learning techniques. For data preparation, the time series data was segmented based on the locations of interest into three distinct regions. Specifically, locations 3 and 4 were combined into a single region, designated as Region 3; location 5 was referred to as Region 2; and locations 6 and 7 were merged into Region 1. The Regions and corresponding time stamps were matched manually based on the video recorded on the day of taking measurement and interpreting the time domain graph. Prior to analysis, we ensured that all data columns were converted to their appropriate data types.

Subsequently, the data was split into feature vectors and target vectors. The feature vector included the timestamp, acceleration in the x, y, and z directions, total acceleration, and location. The target vector indicated the presence of the damper and inclination. The dataset was then divided, with 70% allocated for training the model and the remaining 30% reserved for testing purposes.

The best performing model, we utilized was Random Forest Classifier. The model operated with default parameters: n_estimators set to 100 and criterion set to 'gini'. The model predicted the value of the target vectors using binary values, where 0 indicated no disturbances in the region and 1 indicated disturbances. This binary classification allowed for the clear identification of regions with potential disturbances. Finally, the value of the output vectors, including Damper and Inclination, were manually calculated concerning the corresponding time step, ensuring precise localization.

# 3 Results and discussion

## 3.1. Cross-validation

The model successfully predicted the training dataset with an accuracy of 96%. A classification report and confusion matrix were obtained for the testing dataset. The model predicted the location of damper and Inclination with an accuracy of 96% respectively and F1-score of 0.96. The use of the Random Forest method enhanced the precision and balance of the model in identifying true positives and true negatives. The model obtained a mean cross-validation score of 0.9108 which showcases that the model is consistent. Also from Confusion matrix we can see the model has high Recall and High specificity.

Classification report for Damper:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.98 | 0.96 | 0.97 | 11742 |
| 1 | 0.93 | 0.96 | 0.94 | 5808 |
| accuracy | | | 0.96 | 17550 |
| macro avg | 0.95 | 0.96 | 0.96 | 17550 |
| weighted avg | 0.96 | 0.96 | 0.96 | 17550 |

**Figure 7: Classification report for Damper**

Classification report for Damper:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.98 | 0.96 | 0.97 | 11742 |
| 1 | 0.93 | 0.96 | 0.94 | 5808 |
| accuracy | | | 0.96 | 17550 |
| macro avg | 0.95 | 0.96 | 0.96 | 17550 |
| weighted avg | 0.96 | 0.96 | 0.96 | 17550 |

**Figure 8: Classification report for Inclination**

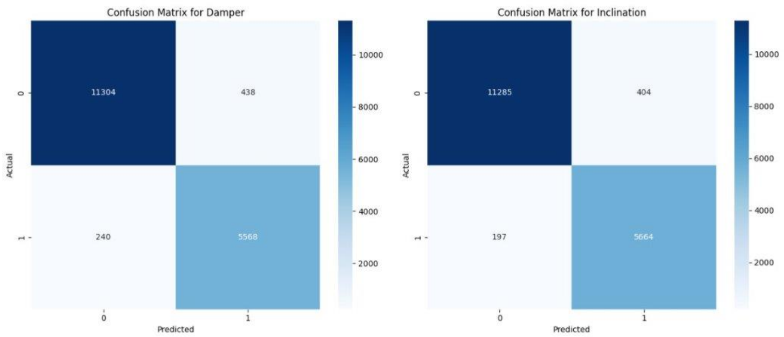


**Figure 9: Confusion Matrix**

| File name | Frequency value | Frequency location | Inclination | Damper |
|---|---|---|---|---|
| Test Case 1 | 70 Hz | Location 5 | Location 3 & 4 | None |
| Test Case 2 | 45 HZ | Location 5 | Location 3 & 4 | Location 3 & 4 |
| Test Case 3 | 70 Hz | Location 5 | None | Location 3 & 4 |
| Test Case 4 | 60Hz | Location 6 and 7 | Location 3 & 4 | Location 3 & 4 |
| Test Case 5 | None | None | None | Location 3 &4 |

**Figure 10: Test Case**

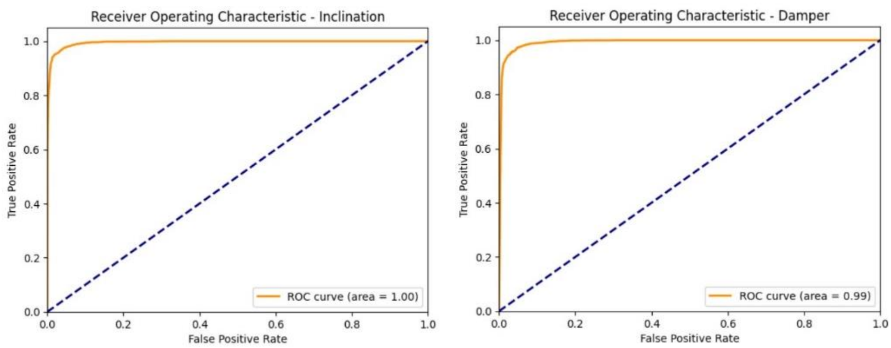


**Figure 11: ROC  Curve for Inclination and damper**

High AUC indicates good classification performance but also higher chance of overfitting, which also explains the poor performance of model to test data which is new and unseen.

## 3.2. Discussion and Feedback

Based on the feedback received from the presentation, the following changes were implemented and documented:

For the frequency disturbance model, an error in the sampling frequency was identified and rectified. Additionally, for the machine learning model used to identify the location of the damper and inclination, Short-Term Fourier Transform (STFT) features were incorporated along with the time-domain features. The parameters for `n_estimators` and `max_depth` were tuned to reduce the complexity of the model. The `n_estimators` parameter was tested with values of 10, 20, 50, and 100, while the `max_depth` parameter was tested with values of 5, 10, and 15. `max_depth` :10 and `n_estimators` :50 gave better classification and also reducing complexity of model.

```
Classification Report for Damper:
              precision    recall  f1-score   support

         0.0       0.85      0.94      0.90     11742
         1.0       0.85      0.67      0.75      5808

    accuracy                           0.85     17550
   macro avg       0.85      0.81      0.82     17550
weighted avg       0.85      0.85      0.85     17550
```

**Figure 12: Classification report for Damper after changes**

```
Classification Report for Inclination:
              precision    recall  f1-score   support

         0.0       0.87      0.92      0.90     11689
         1.0       0.83      0.73      0.77      5861

    accuracy                           0.86     17550
   macro avg       0.85      0.83      0.84     17550
weighted avg       0.86      0.86      0.86     17550
```

**Figure 13: Classification report for Inclination after changes**

After implementing the updates, the frequency disturbance model yielded perfect results for the test cases. Although the accuracy has decreased, the model performed better on the test data compared to the previous version, indicating a reduction in overfitting. The machine learning model demonstrated improved generalization in identifying the damper; however, it continues to face challenges in accurately identifying inclination in previously unseen regions. Also since inclination is possible only in 3 and 4 locations and we assumed them to be of one region, even if the model predicts the inclination over the specified region it proves to be not useful in real life scenario, further investigation and analysis to be done for segregating nearby regions for developing better model which can be implemented in industry.

| File name | Frequency value | Frequency location | Inclination | Damper |
|---|---|---|---|---|
| Test Case 1 | 70 Hz | Location 5 | Location 3 & 4 | None |
| Test Case 2 | 70 HZ | Location 5 | Location 3 & 4 | None |
| Test Case 3 | 70 Hz | Location 5 | None | Location 5 |
| Test Case 4 | 20Hz | Location 6 and 7 | Location 3 & 4 | None |
| Test Case 5 | 40Hz | Location 5 | Location 3 & 4 | Location 5 |

**Figure 14: Test Case after updating**

Significant improvements are still necessary to enhance the prediction of inclination and damper, ensuring the model's applicability in real-world production plants.

# 4 Conclusion and outlook

## 4.1. Conclusion

Our machine learning model has helped us in predicting effects caused by various disturbances like frequency, inclination, and dampening which were introduced in the model of the production line. We have managed to achieve an impressive accuracy of 96% for both damper and inclination classification with the help of the Random Forest Multiclass Classifier with time domain features and combing them with frequency domain features we get accuracy of 86% but the model's performance is better in unseen data for damper case. Further improvement can be achieved by enhancing the model's generalization ability for unseen data and scalability to unforeseen situations. We need to improve this in order to achieve robust performance in real-world scenarios.

## 4.2. Outlook

Although these findings are encouraging, it is critical to recognize the model's limits. Subsequent versions of the model could concentrate on improving features, adjusting hyperparameters, or incorporating additional relevant factors in order to provide a prediction of production line anomalies that is more precise, more adaptable, and scalable to various applications and new environments. We can achieve this by refining our model to better handle diverse and previously unseen data. This project can be further enhanced to implement real-time disturbance detection coupled with automated correction mechanisms. This will minimize the need for human intervention, thereby optimizing resource allocation and improving the efficiency of the entire production system. The soul concept of this project is to prove that time and energy can be efficiently used and allocated with help of computational methods and techniques to avoid catastrophic events.

# References

[1] J. V. Abellan-Nebot und . F. R. Subirón , „A review of machining monitoring systems based on artificial intelligence process models," *The International Journal of Advanced Manufacturing Technology,* Bd. Volume 47, p. pages 237–257, 2010.

[2] C. Wei, „Intelligent manufacturing production line data monitoring system for industrial internet of things," *Computer Communications,* Bd. 151, pp. Pages 31-41, 2020.

[3] T. Morimoto und Y. Hashimoto, „AI approaches to identification and control of total plant production systems," *Control Engineering Practice,* Bd. 8, Nr. 5, pp. 555-567, 2000.