**FAULT CLASSIFICATION IN BEARINGS AND GEARS**

**FINAL REPORT**

***Prepared by***

**SAI NIVEDH V - (CB.SC.U4AIE23062)**

**SYED ANEES ASHRAF - (CB.SC.U4AIE23068)**

**C V SHREYAAS ADITYA - (CB.SC.U4AIE23018)**

**MATHIVANAN S - (CB.SC.U4AIE23042)**

**SEMESTER –1**

**ENDSEM PROJECT REPORT**

***As a part of the subject***

22PHY106 - COMPUTATIONAL PHYSICS



AMRITA SCHOOL OF ENGINEERING

AMRITA VISHWA VIDYAPEETHAM

COIMBATORE – 641 112

**DECEMBER - 2023**

**DECLARATION**

*We hereby declare that our project work entitled* ***“Fault classification in Bearings and Gears”*** *submitted to*

**Mr. Jithin Velayudhan**

Assistant Professor

Centre of Excellence and Networking (CEN)

Amrita School of Engineering

*Place :* Coimbatore

*Date:* 24-12-2023

**ACKNOWLEDGEMENT**

We would like to express our sincere appreciation to our supervisor,

**Mr. Jithin Velayudhan**, for his invaluable guidance, constructive feedback, and unwavering support throughout this project.

We would also like to extend our heartfelt thanks to **Dr. K.P. Soman**, Head of the Department, for his valuable suggestions and encouragement at various stages of the work. We are grateful to all the staff and members of the Department of Artificial Intelligence, Amrita Vishwa Vidyapeetham Coimbatore, for their support and cooperation.

Lastly, we would like to thank our colleagues for their flow of ideas and dear ones who helped us in completing this work. Your contributions have been instrumental in making this project a success.

**CONTENT**

**TITLE PAGE NUMBER**

|  |  |
| --- | --- |
| **ABSTRACT**  **INTRODUCTION**  **PROBLEM DEFINITION**  **MOTIVATION TO DEFINE PROBLEM DEFINITION**  **OBJECTIVE TO SOLVE**  **FLOW CHART**  **PROPOSED METHODOLOGY**  **FUTURE SCOPE**  **CONCLUSION**  **SOURCE CODE**  **REFERENCES** | **5**  **6**  **8**  **9**  **10**  **11**  **12**  **24**  **25**  **26**  **27** |

**ABSTRACT**

This work is dedicated to improving the monitoring of external organs using useful raw data from the IMS Information System. This data is particularly useful as it provides 1-second snapshots of vibration signals recorded at specific times, allowing fine-grained examination of behavior. The research focused on three different types of bearings: solid bearings, bearings with inner ring defects, and bearings with outer rings. This method involves carefully pre-processing the raw vibration data and finally determining the crest factor, kurtosis value, peak value, etc. It is presented by extracting the best features such as; Each of these is carefully selected specifically for its meaning. This extraction process reveals unique patterns in the data, creating important symbols. The subsequent classification of bearings is intricately based on abnormalities discerned within their respective graphical representations. To further refine the analysis, dimensionality reduction becomes imperative. Employing Principal Component Analysis (PCA), the original set of eight features is judiciously condensed to a more manageable two features. This reduction not only streamlines the dataset but also facilitates a robust validation of the proposed classification method across all fault categories.

Expanding the scope of inquiry, unlabelled datasets sourced from Kaggle, featuring various loads and fault classifications, are incorporated into the study of Gears. In a pioneering move, these datasets are meticulously labeled, enhancing the breadth of the analysis. The next stage involves using different learning models to use the recording data to classify the bearing according to the type of fault and load. It is worth noting that our scheme has the best accuracy; findings are confirmed by rigorous validation using confusion matrix analysis.

**Key words :**

* **IMS bearing data** - *NSF I/UCR Centre for Intelligent Maintenance Systems (IMS – www.imscenter.net) with support from Rexnord Corp. in Milwaukee, WI.*
* **Principal Component Analysis**

**INTRODUCTION**

The industrial environment is characterized by mechanical components such as bearings and gears that play an important role in ensuring the smooth operation of machines. The reliability and performance of these components are crucial to the overall performance of a variety of machines, from heavy machinery to precision instruments. However, these materials tend to wear and deteriorate over time; It requires high-performance equipment to provide timely warranty and prevent major damage.

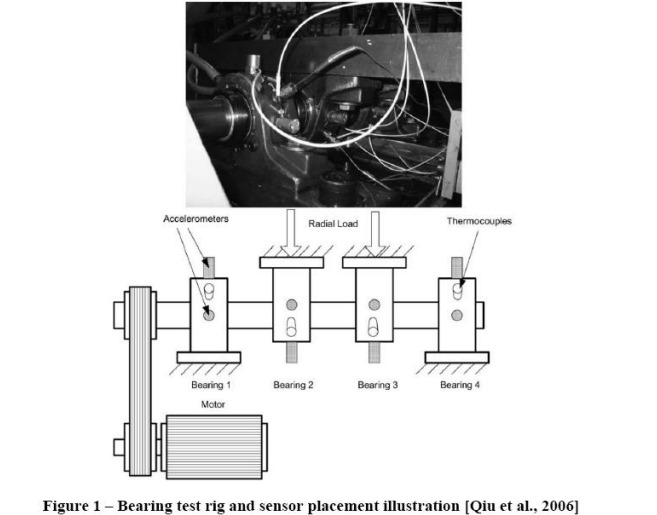
This project delves into the complex field of bearing and gear failure classification, with a special focus on outer ball bearings. The research was supported by a large amount of raw data provided by the IMS Data Center, known for its 1-second information on vibration signals. These photographs recorded at a specific time provide detailed information about the behavior of the organism, especially in different conditions.

The research focused on three different types of bearings: bearings in healthy condition, inner ring bearings and outer ring bearings. The process begins with careful pre-processing of raw vibration data from which the best features are extracted. These features (including best peak, kurtosis value, maximum value, etc.) were carefully selected because they have clear meaning in capturing differences in the data. These features, as important landmarks, form the basis for the subsequent distribution of the bed, as ambiguity is shown in the image representation.

Dimensionality reduction is an important step in qualitative analysis and Principal Component Analysis (PCA) is used for this purpose. The first set of eight features is compressed into two more manageable features, simplifying the dataset and contributing to a robust validation of the proposed distribution method for all error categories.

Extending the query, this project includes anonymous features from the Kaggle dataset with various loading and error distributions. At the beginning of its kind, these documents were carefully labeled, increasing the scope of analysis. The next stage involves using different learning models to use the recording data to classify the bearing according to the type of fault and load. One of the main advantages of this approach is the good results in accuracy, which are confirmed by stringent validation using confusion matrix analysis.

Here we use a convolutional neural network (CNN), which is a feed-forward neural network that self-learns feature engineering through filter (or kernel) optimization>Importantly, the goal of this project is: Organ and equipment separation To contribute to the field of health by breaking The combination of advanced signal processing, feature extraction and machine learning techniques will increase the accuracy and efficiency of fault detection, thereby simplifying monitoring strategies and reducing the risk of operational disruption.



**PROBLEM DEFINITION**

The primary challenge addressed in this project is the accurate and timely identification of faults in bearings and gears within industrial machinery, with a focus on ball bearings and Gears.

The diverse nature of faults, ranging from misalignments and imbalances to wear and fatigue, demands a robust classification system that can differentiate between various fault types. The goal is to develop a solution that not only detects the presence of faults but also provides insights into the specific type and severity, enabling proactive maintenance measures.

The challenge lies in effectively harnessing this data's depth and breadth, considering the variety of fault types and the nuances within each class.

**MOTIVATION TO DEFINE PROBLEM DEFINITION**

The motivation to rigorously define the problem at hand stems from the critical role that bearings and gears play in the seamless operation of industrial machinery. As the backbone of mechanical systems, these components are subjected to a myriad of operational challenges, ranging from misalignments and imbalances to wear and fatigue. The consequences of undetected faults in bearings and gears can be severe, leading to unplanned downtime, costly repairs, and, in extreme cases, catastrophic failures with potential safety implications.

The motivation is further fueled by the evolving landscape of industrial maintenance practices. Traditional methods relying on manual inspections and routine maintenance schedules often fall short in identifying latent or early-stage faults, leaving machinery vulnerable to unexpected breakdowns. The need for a more proactive and data-driven approach to fault detection becomes evident, with the aim to transition from reactive maintenance to predictive maintenance strategies.

Additionally, the advent of advanced sensor technologies and the availability of comprehensive datasets, such as the IMS Bearing Data and Kaggle, present an unprecedented opportunity to delve deeper into the intricacies of bearing behavior. The motivation to define the problem with precision arises from the desire to harness the full potential of these datasets, extracting meaningful insights that can revolutionize the way faults in bearings and gears are identified and addressed.

Moreover, the incorporation of machine learning techniques in fault classification opens up new horizons for efficiency and accuracy.

**OBJECTIVE TO SOLVE**

The aim is to use machine learning algorithms such as tracking classification patterns to train and develop the classification system that can detect and identify the fault of the bearing and extract Energy-based features.

Next, we must resort to a Convolutional Neural Network (CNN), a powered neural network that independently learns the architecture through filter (or kernel) optimization. It has many layers, including layers, outer layers, and full layers. Another reason to do this is that CNNs are very good at classifying signal information.

Convolutional layers apply filters to the input image to extract features. The pooling layer subsamples the image to reduce the computational cost. All layers for final classification.

In our project we will explore and use graphical methods to represent visual anomalies in vibration data, helping to diagnose and explain some faults in bearings and power.

Finally, we will examine the possibility of using crime classification in a real business environment, solving problems regarding the performance of calculations, the integration of the sensor and the inconsistency of the integration with existing observations.

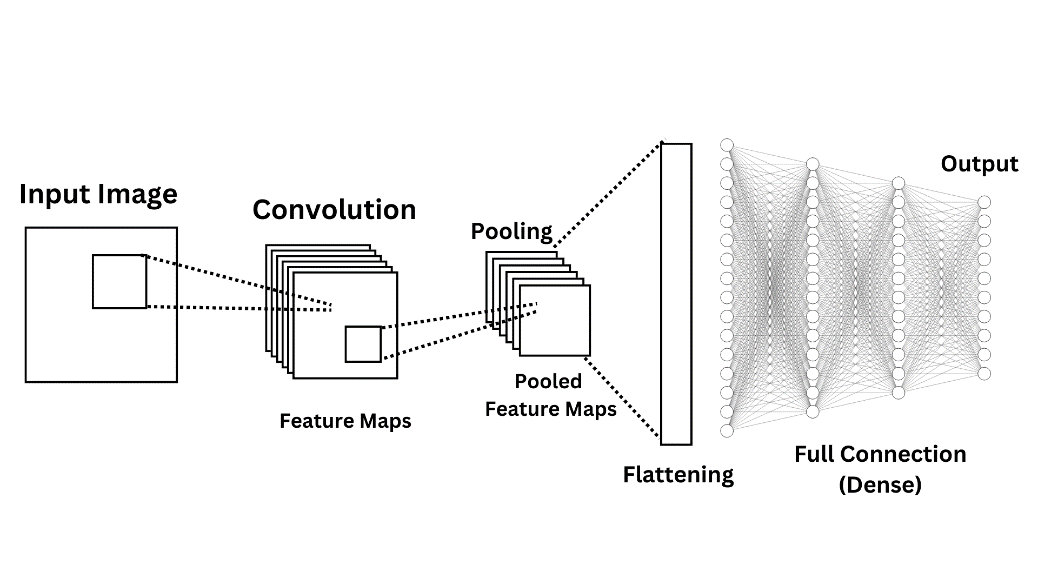


Figure 2 – Architecture of Convolution neural network

**FLOW CHART**

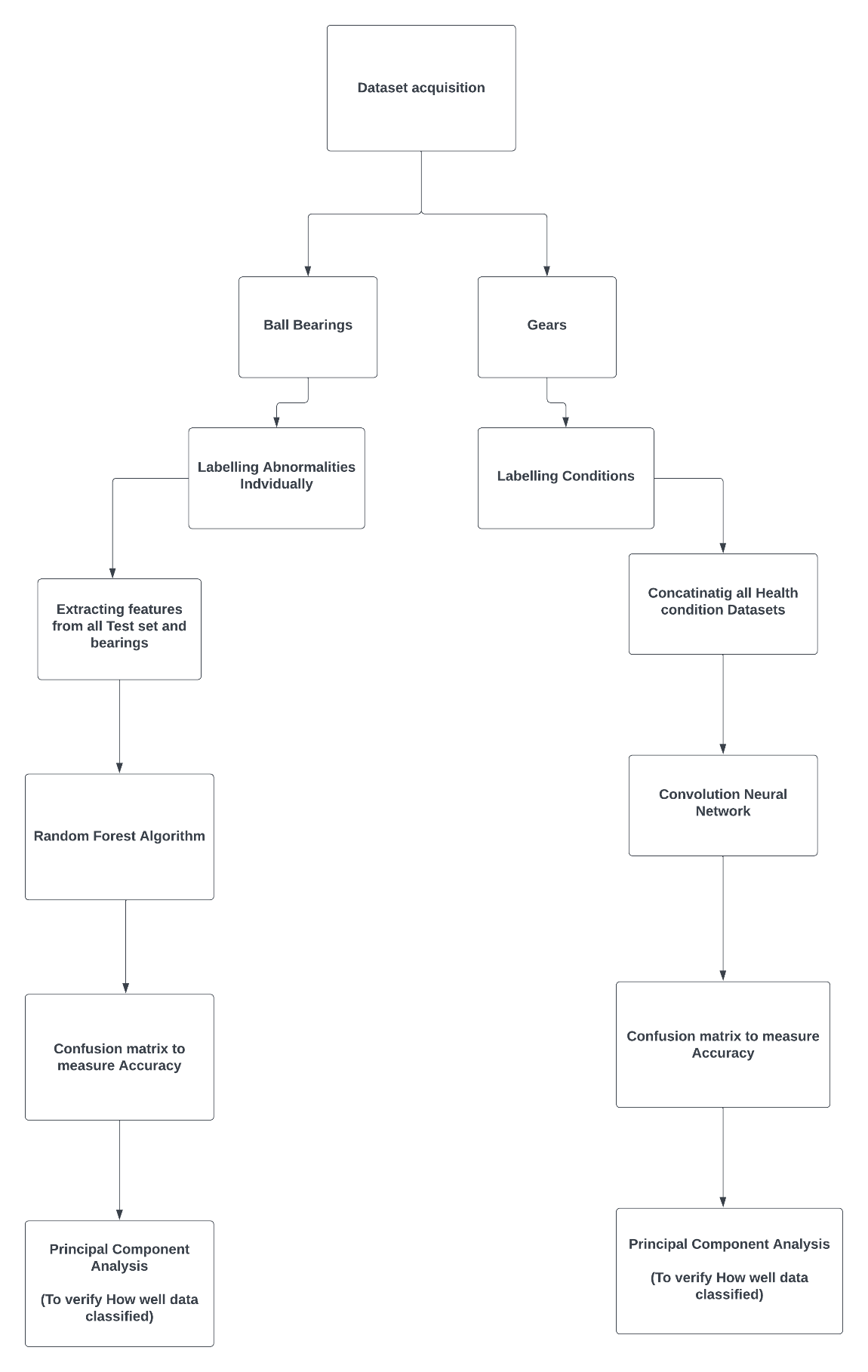


Figure 3 – Flowchart that depicts Brief overview of our Project

**PROPOSED METHODOLOGY**

**Ball Bearing:**

**1.Data Importing and Preprocessing:**

**1.1 Data Importation**

The first step of the plan is to import the Dataset, which contains data from the four accelerometers associated with the ball bearings. This file contains three test files, each containing information about four viruses and specific errors that occurred for each virus during the tests.

**1.2 Feature Extraction**

For each bearing in each test data, features will be extracted to characterize the vibration signal. These functions include maximum, minimum, std (model mean square), Rokewness, Kurtosis, Crest values ​​and standard deviation. The aim is to develop a comprehensive system that stores information about vibration patterns associated with different types of faults.

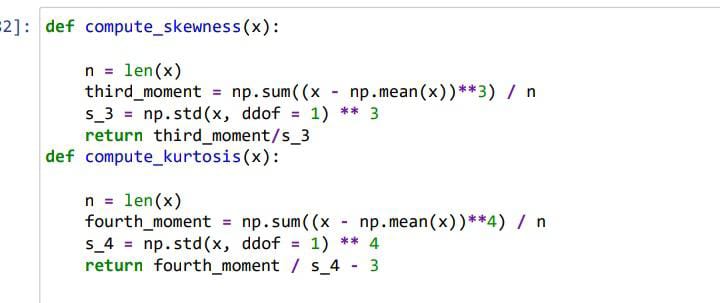


Figure 4 – Function definitions of two of the total features

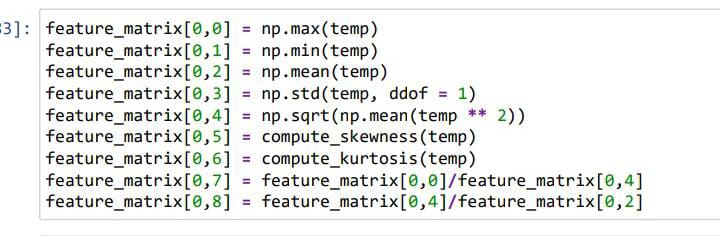


Figure 5 – Defining Features using scipy library in python

**2.Fault Classification:**

**2.1 Default fault labelling**

The data set provides information about the fault that occurred in each bearing during the test. Faults will be classified according to this preset label. We will be doing this for all data sets.



Figure 6 – Max vs time for an Bearing



Figure 7 – Code to classify abnormal curve as Specific defect

**2.2 Dimensionality reduction with PCA**

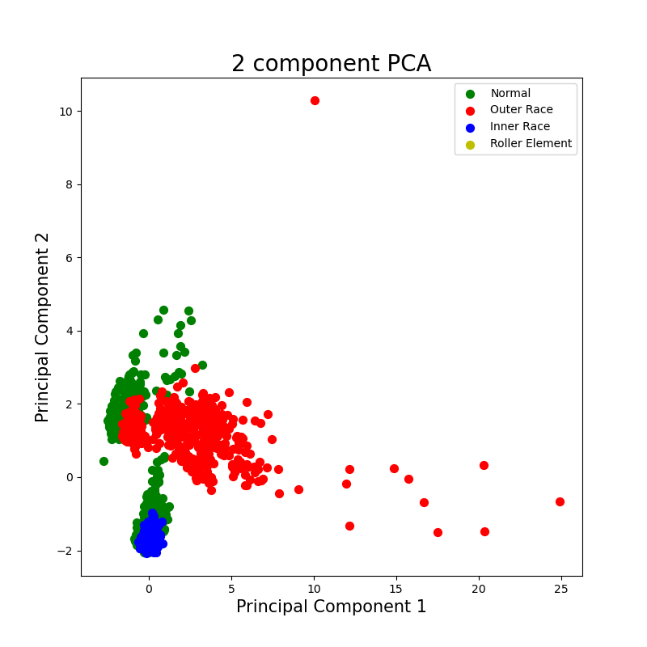
 To streamline the feature set, Principal Component Analysis (PCA) will be applied to reduce the dimensions of the extracted features to two principal components. This step aims to retain the essential information while simplifying the dataset.

Figure 8 – Plotting of classification after performing PCA

**2.3 Model Training**

Reduction technique will be used to train the machine learning model. Attributes related to vibration properties (except incorrect characters) will be evaluated as "x", incorrect characters will be evaluated as "y". The model will be trained to predict errors based on vibration characteristics.

**2.4 Model Validation**

The performance of the training model will be evaluated using random forest distribution. The performance of the model will be evaluated by preparing a confusion matrix that will give an idea about the accuracy and efficiency of fault classification.

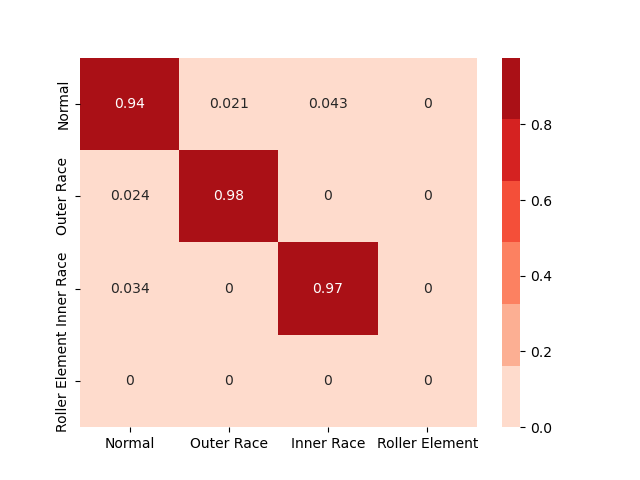


Figure 9 – Confusion matrix (which shows that how well our model has classified )

**3.Visualization and Analysis:**

**3.1 Plotting classified values**

Error results obtained from the training model have been prepared for all test forms and organs. This visualization will help to check how changes are made at the same test time.

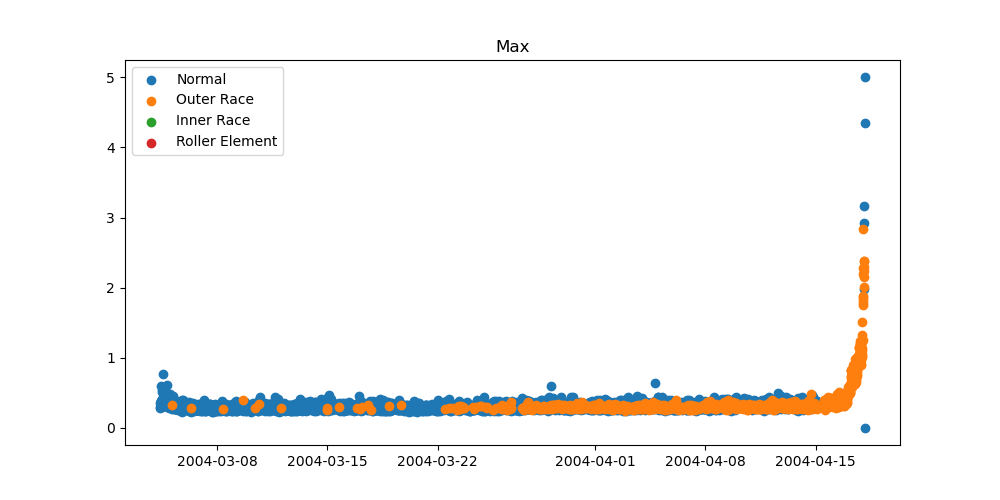


Figure 10 – It’s the visualization of Defect of bearing 3 from test 3 data and it’s clear that at end, Bearing 3’s Outer race Damaged severely

**Gears :**

**1.Data Preprocessing:**

* 1. **Data collection**

The file contains two folders labeled "Dead Tooth" and "Healthy", containing accelerometer data from the gearbox. Each folder has nine subfolders that represent different failure modes, such as Load (labeled from 0% to 90%).

* 1. **Data Labelling**

All data files are marked as "OK" or "Error" based on the relevant data. We are sure that there is another exception in the table called "fault".

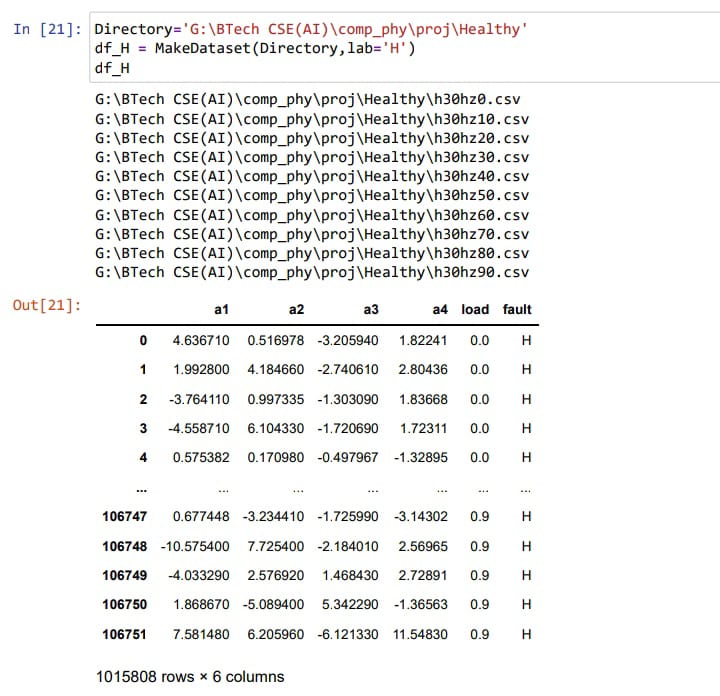


Figure 11 – Python code to Label fault type

**1.3 Data Concatenation:**

Status and fault information are combined into a single configuration, resulting in six features: four electronic accelerometers, electrical load lines, and fault lines.

**1.3 Data Splitting:**

The dataset is separated into two subsets: X, containing all highlights but the blame name, and Y, containing as it were the blame names.

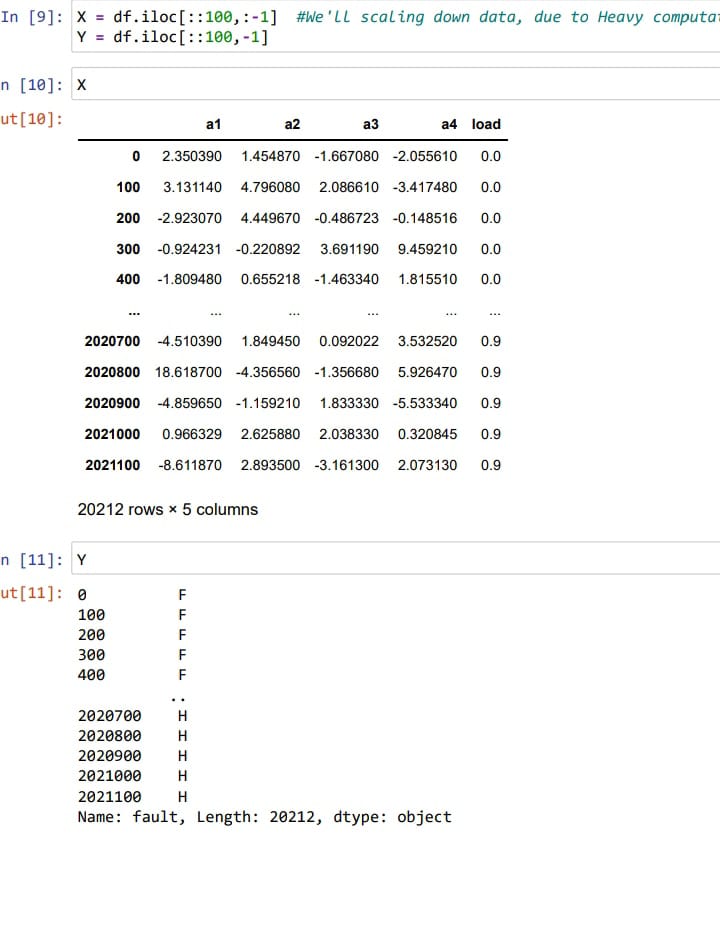


Figure 12 – Splitting Dataset

**2. Dimensionality Reduction:**

**2.1 t-Distributed Stochastic Neighbour Embedding (t-SNE):**

t-SNE is connected to the highlight set X to imagine and assess its dissemination. Be that as it may, due to confinements in precisely recognizing between Solid and Blame information, an elective approach is investigated.

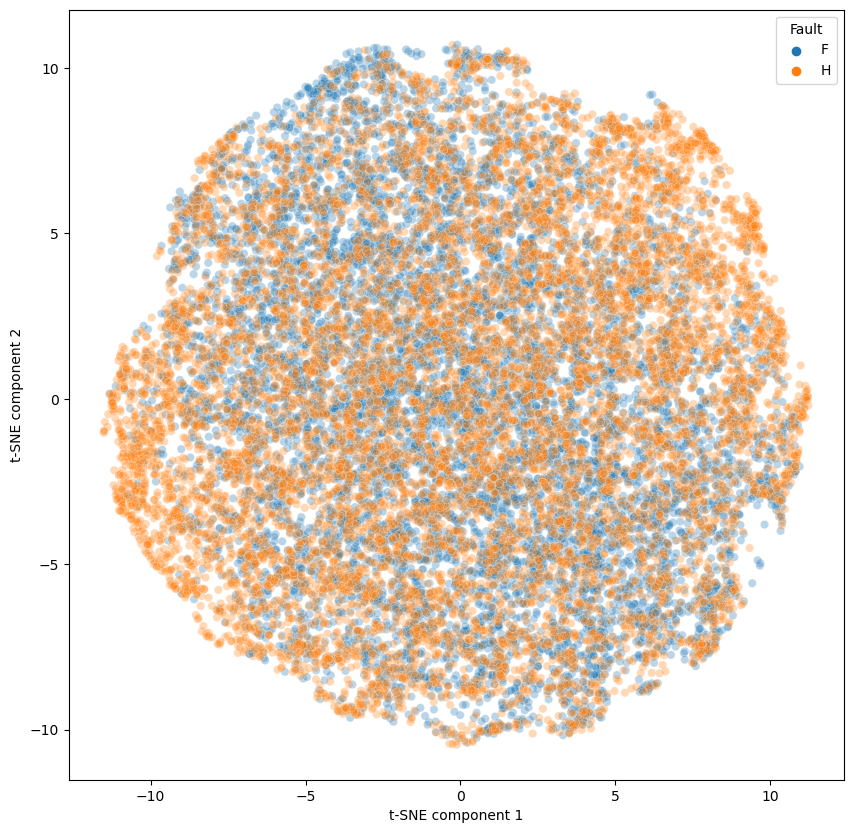


Figure 13 – TSNE plot to illustrate how well data set splitted up

**2.2 Random Classifier:**

A Irregular Classifier is utilized to disperse the information and produce a perplexity network, uncovering troubles in recognizing due to overfitting and fluctuating speeding up information.

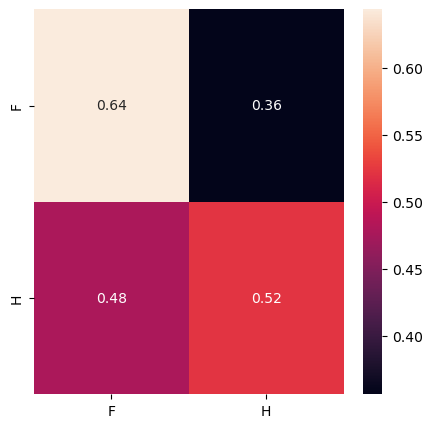


Figure 14 – Confusion matrix for Random Forest Algorithm

**3.Convolution Neural Networks:**

**3.1 Data Windowing:**

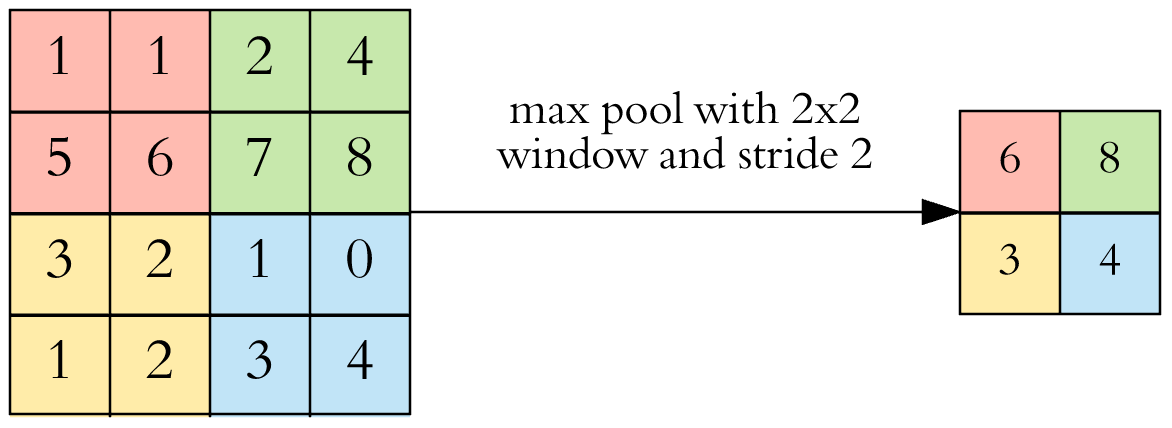
A sliding window approach is actualized with parameters window length as 100 and walk as 200 to extricate arrangements from the dataset for preparing the CNN.

Figure 15 – Stride concept in Pooling (convolution neural networks)

**3.2 Data Encoding:**

Information is prepared to form input highlights (X) and yield names (Y) reasonable for preparing the CNN. The Y names are one-hot encoded utilizing categorical encoding.

Figure 16 – Code to Implement convolutional

**3.3 CNN Architecture:**

A CNN demonstrate is built with layers counting convolutional and pooling layers, taken after by smoothing and thick layers. The demonstrate is compiled utilizing categorical cross entropy as the misfortune work and Adam optimizer.

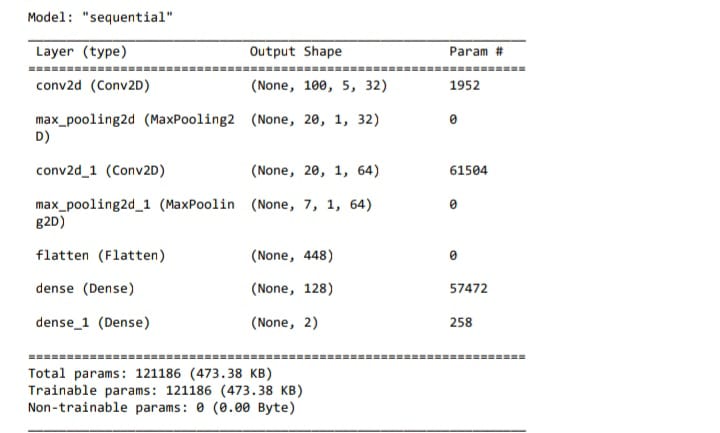


Figure 17 – summary of Layers used in our Models

**4. Model Evaluation:**

**4.1 Confusion Matrix:**

The perplexity lattice is produced to survey the CNN model's capacity to precisely classify Sound and Blame information. In our case This Disarray framework outlines us that Our Show has classified information with 100 % accuracy.

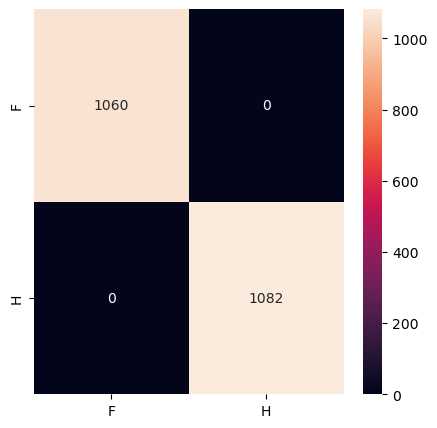


Figure 18 – Confusion matrix of Outputs from CNN

**4.2 t-SNE Plot:**

t-SNE is connected to the CNN-generated highlights to imagine the dissemination and division of the dataset. It’s exceptionally fine to plot with any of the two Thick layers.

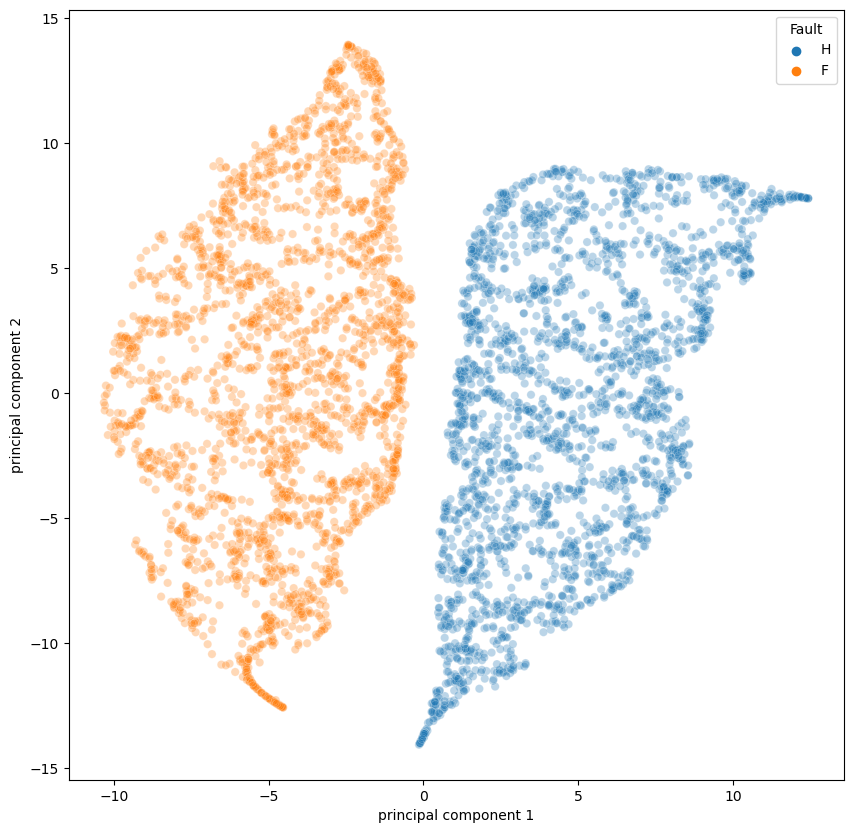
This plot clearly portrays that our demonstrate Classified the total information with Tall precision than Irregular Timberland Calculation.

Figure 19 – TSNE plot to illustrate how well our data has been segregated using CNN

**FUTURE SCOPE**

Consolidating information from extra sensors, such as temperature sensors or vibration sensors, can give a more comprehensive understanding of equip wellbeing. This development may contribute to a more vigorous and exact blame discovery framework. Testing with energetic windowing procedures seem improve the flexibility of the CNN demonstrate to changing blame conditions. Versatile window sizes and strides may be investigated to optimize the extraction of significant highlights from the accelerometer information. Execution of outfit learning strategies, combining the qualities of different models, can possibly make strides in general classification execution. Gathering strategies like sacking or boosting may be investigated to relieve the affect of overfitting. Exploring the application of exchange learning utilizing pre-trained CNN models on related errands or spaces may speed up the preparing handle and make strides generalization. This approach may be especially advantageous when constrained labeled information is accessible. Expanding the venture to real-time checking scenarios seem include conveying the show on edge gadgets for on-site blame discovery. This move toward edge computing can upgrade the responsiveness of the framework and diminish reliance on centralized preparing. Utilizing information increase procedures, such as turn, scaling, or clamor infusion, can falsely grow the dataset, possibly making strides the model's capacity to generalize to concealed information and varieties in working conditions. Improving the interpretability of the model's choices is vital for picking up client believe and encouraging the integration of the framework into mechanical situations. Methods for clarifying CNN expectations, such as LIME (Neighborhood Interpretable Model-agnostic Clarifications), can be investigated. Joining the created blame location framework with existing support frameworks or setting up communication conventions with upkeep faculty might streamline the sending of preventive support activities based on the distinguished deficiencies. Executing ceaseless checking instruments and versatile learning techniques can guarantee that the demonstrate remains up-to-date with advancing blame designs and natural changes over time.

**CONCLUSION**

Within the interest of viably classifying equip case information into Solid and Blame categories, our venture set out on a comprehensive travel through various methodologies.

Initially, the endeavor to imagine the dataset utilizing t-Distributed Stochastic Neighbor Inserting (t-SNE) uncovered challenges in accomplishing a clear partition between Solid and Blame information. The confinements watched incited an investigation into the application of a Irregular Classifier. In any case, the comes about demonstrated issues related to overfitting and the fluctuating nature of speeding up information, inciting the require for a more modern approach.

The CNN's engineering, comprising convolutional and pooling layers, demonstrated instrumental in capturing hierarchical features from the information. The model's capacity to memorize complicated designs in adapt case information was apparent in its execution measurements and the coming about perplexity lattice, which showcased a clear refinement between Sound and Blame classifications. This effective classification was advance confirmed by the t-SNE plot, which outlined the model's adequacy in accomplishing a well-separated dispersion of the dataset. Our venture not as it were tended to the inborn challenges in classifying equip case information but moreover illustrated the predominance of Convolutional Neural Systems in this space. The extend results emphasize the significance of leveraging progressed profound learning methods for perplexing blame conclusion errands. The experiences picked up from this endeavor contribute to the developing body of information in prescient upkeep and condition observing, with potential applications in different mechanical divisions. As innovation proceeds to advance, our work serves as a venturing stone towards more vigorous and exact blame discovery techniques.

**SOURCE CODE**

* [Gear fault classification](https://drive.google.com/file/d/10gJwfeTAG1pfFlVzr4HlTsFqx54YsBWC/view?usp=sharing)
* [Ball Bearing classification](https://drive.google.com/file/d/1c3Ogz62Cj_uhpgjUWnZ3aNSVVewuohbV/view?usp=sharing)
* [Ball Bearing Dataset](https://www.kaggle.com/datasets/vinayak123tyagi/bearing-dataset)
* [Gears Dataset](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbUFEblBic0kzQmdZYV9Ec1dDSHI4VmpkTG5fd3xBQ3Jtc0trR3d6U29sLU5DN1NYZWdueGd4dWE5ZU0tWVVxOHN2aUR4c1lWUXdKR05taEtYZGVvU19STkNyUGNYMWl4T01yZlpieFFFT2E2b0xnYTdiQmFTM01ZU3NJYVFqZHZ0aHFnaURISlpIWkF6eEstQUxpYw&q=https%3A%2F%2Fwww.kaggle.com%2Fbrjapon%2Fgearbox-fault-diagnosis&v=km9HVtDRC5k)

**REFERENCES**

**1.Recent Trends in Engineering Design Select Proceedings of ICAST 2020 -** [**https://link.springer.com/book/10.1007/978-981-16-1079-0**](https://link.springer.com/book/10.1007/978-981-16-1079-0)

**2.An Overview of Vibration Analysis Techniques for the Fault Diagnostics of Rolling Bearings in Machinery -**

[**https://www.hindawi.com/journals/sv/2022/6136231/**](https://www.hindawi.com/journals/sv/2022/6136231/)

**3.A Review on Vibration Signal Analysis Techniques Used for Detection of Rolling Element Bearing Defects -**

[**https://www.researchgate.net/publication/349918774\_A\_Review\_on\_Vibration\_Signal\_Analysis\_Techniques\_Used\_for\_Detection\_of\_Rolling\_Element\_Bearing\_Defects**](https://www.researchgate.net/publication/349918774_A_Review_on_Vibration_Signal_Analysis_Techniques_Used_for_Detection_of_Rolling_Element_Bearing_Defects)

**4.Machinery Fault Diagnosis Using Signal Analysis -**

[**https://www.sciencedirect.com/science/article/pii/S2351978919302914?ref=pdf\_download&fr=RR-2&rr=82fc06fc8bb9937f**](https://www.sciencedirect.com/science/article/pii/S2351978919302914?ref=pdf_download&fr=RR-2&rr=82fc06fc8bb9937f)

**5. A Comparison of Model-Based and Machine Learning Techniques for Fault Diagnosis**

[**A Comparison of Model-Based and Machine Learning Techniques for Fault Diagnosis | IEEE Conference Publication | IEEE Xplore**](https://ieeexplore.ieee.org/abstract/document/10021712)