Structural Health Monitoring using Random Forest Classifier

Abisheik Visagan M

August 11, 2025

Abstract

This report explores the use of machine learning for structural health monitoring, specifically in detecting anomalies in sensor data (e.g., vibration, stress, temperature, humidity, and traffic load). A Random Forest classifier was trained on synthetic data generated to mimic real-world conditions, and its performance was evaluated using various metrics such as accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix.

1 Introduction

Structural health monitoring (SHM) systems use sensor data to monitor the integrity of structures such as buildings, bridges, and dams. Anomalies in sensor readings may indicate potential structural issues or failure. Predicting these anomalies is crucial for preventive maintenance and ensuring safety. This study utilizes a Random Forest classifier, a powerful ensemble learning technique, to predict anomalies in structural health based on sensor data.

2 Data Description

The data used for this analysis was synthetically generated, including features like vibration, stress, temperature, humidity, and traffic load. Anomalies were introduced to simulate real-world structural issues, such as high vibration or stress, abnormal temperature variations, and unusual traffic load.

- **Vibration**: Measured in units representing mechanical vibrations.
- **Stress**: Measured in units representing stress levels on the structure.
- **Temperature**: Measured in degrees Celsius.
- **Humidity**: Measured in percentage.
- **Traffic Load**: Measured in the number of vehicles or load on the structure.

Anomalies were randomly injected into the data to simulate structural issues.

3 Methodology

3.1 Preprocessing

The data was preprocessed by scaling the features using the `StandardScaler` to standardize the range of the features. This step ensures that no feature dominates due to differences in scale.

3.2 Modeling

A Random Forest classifier was used for the classification task. Hyperparameter optimization was performed using `RandomizedSearchCV`, which tested several configurations of the following parameters:

- `n_estimators`: Number of trees in the forest.
- `max_depth`: Maximum depth of each tree.
- `min_samples_split`: Minimum number of samples required to split a node.
- `min_samples_leaf`: Minimum number of samples required to be at a leaf node.
- `bootstrap`: Whether bootstrap samples are used when building trees.

The model was trained using 80% of the data, and the remaining 20% was used for testing.

3.3 Evaluation Metrics

The model's performance was evaluated using the following metrics:

- **Confusion Matrix**: Shows the number of true positives, true negatives, false positives, and false negatives.
- **Classification Report**: Includes precision, recall, and F1-score for each class.
- **ROC-AUC Score**: Measures the ability of the model to distinguish between the two classes.
- **Precision-Recall Curve**: Evaluates the trade-off between precision and recall for different thresholds.

4 Results

4.1 Confusion Matrix

The confusion matrix for the model is as follows:

$$\begin{bmatrix} 935 & 5 \\ 2 & 58 \end{bmatrix}$$

Here, the true positives (TP) are 58, the true negatives (TN) are 935, the false positives (FP) are 5, and the false negatives (FN) are 2.

4.2 Classification Report

The classification report provides detailed metrics for precision, recall, and F1-score for each class. The results are as follows:

Class	**Precision**	**Recall**	**F1-Score**
0.0 (Normal)	1.00	0.99	1.00
1.0 (Anomalous)	0.92	0.97	0.94
Accuracy	0.99		
Macro Avg	0.96	0.98	0.97
Weighted Avg	0.99	0.99	0.99

4.3 ROC-AUC Score

The ROC-AUC score for the model is 1.00, indicating that the model perfectly distinguishes between normal and anomalous instances.

4.4 Precision-Recall Curve

The Precision-Recall curve was plotted to evaluate the performance of the model on imbalanced classes. The average precision score was computed as 0.94, which is quite good for an imbalanced dataset.

4.5 Feature Importance

The feature importance plot highlights which features contributed the most to predicting anomalies. The following plot shows the relative importance of each feature:

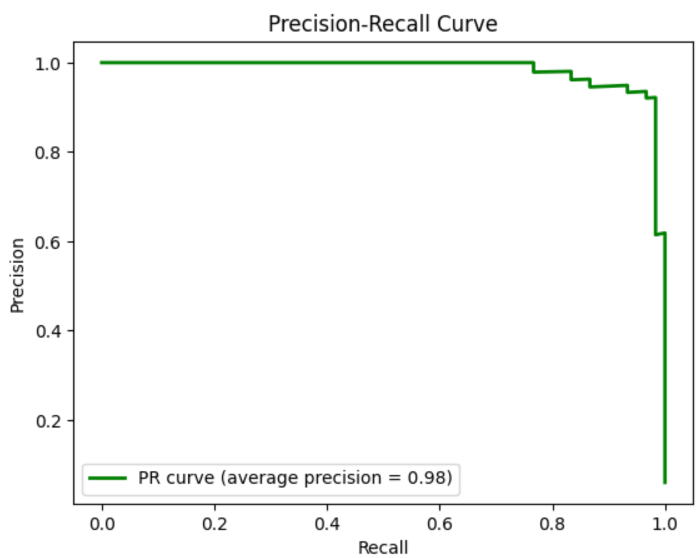

Figure 1: Precision recall curve

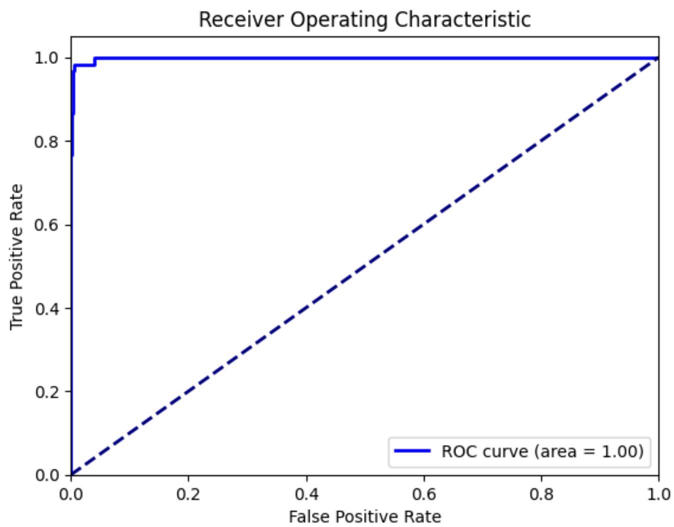

Figure 2: Receiver operating charecteristic

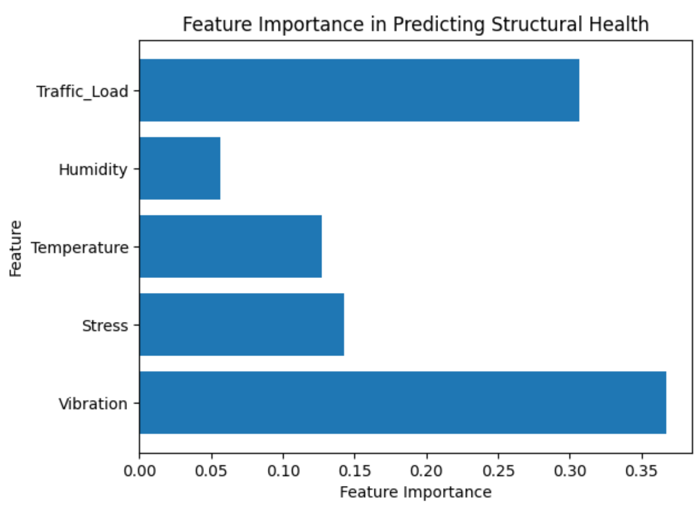

Figure 3: Feature Importance for Structural Health Prediction

5 Discussion

The model performed excellently, with high precision, recall, and F1-score for both the normal and anomalous classes. The confusion matrix and ROC-AUC score indicate that the model has a strong ability to differentiate between normal and anomalous states. Additionally, the precision-recall curve confirms that the model is robust, even in cases where anomalies are less frequent.

6 Conclusion

This report demonstrates the effectiveness of the Random Forest classifier for predicting structural health from sensor data. The model achieved high accuracy and AUC, making it a suitable choice for anomaly detection in SHM systems. Future work could involve exploring other algorithms such as XGBoost or testing the model on real-world data to further validate its performance.