The University of Azad Jammu & Kashmir Muzaffarabad

DEPARTMENT OF SOFTWARE ENGINEERING

A project report on

Automatic machine fault detection from Acoustic Data using Deep Learning

By

Hanzala Khalid	**2021-SE-05**
Ahtsham Ul Haq	**2021-SE-25**
Abdullah Shahid	**2021-SE-32**

Submitted to: Engr. Dr. Ahmad Khawaja

Abstract

This report presents an automated fault detection framework for electrical machines using audio signal analysis named as “Automatic machine fault detection from Acoustic Data using Deep Learning”. The approach extracts time–frequency features by converting audio signals into spectrogram images and additional feature representations (such as MFCC, STFT magnitude, and phase). These features are pre-processed including segmentation, augmentation, normalization, and resizing into square images and then used to train a convolutional neural network (CNN) for fault classification. Experimental evaluation shows a test accuracy of approximately 80.76% and performance is visualized using training/validation graphs and a confusion matrix.

Keywords**:** Fault detection, spectrogram, audio processing, convolutional neural network (CNN), MFCC, STFT, data augmentation.

1. Introduction

Fault detection in electrical machines is essential to ensure uninterrupted operation, reduce maintenance costs, and avoid catastrophic failures. Traditionally, methods such as vibration and thermal monitoring have been used for diagnostic purposes; however, these techniques can require invasive installations and high costs. Recently, audio-based methods have emerged as a non-invasive alternative, offering a cost-effective solution that leverages the rich information contained in sound signals.

Audio signals captured during machine operation contain characteristic patterns that change when a fault occurs. By transforming these signals into time–frequency representations—most commonly spectrograms one can visualize the distribution of energy across different frequency bands over time. Spectrograms reveal both transient and steady-state characteristics of the machine’s operation, which are useful for differentiating between fault types such as Arcing, Corona, Looseness, and Tracking.

In this project, we propose a methodology that converts audio recordings into both raw and processed (square) spectrogram images. Additional features, including Mel Frequency Cepstral Coefficients (MFCCs), STFT magnitude (in decibels), and phase information, are extracted from segmented audio clips. Data augmentation techniques (time stretching, pitch shifting, and noise addition) are applied to improve model robustness. Finally, these features are used to train a convolutional neural network (CNN) that automatically classifies the type of fault.

2. Data Acquisition and Preprocessing

2.1 Audio Data Collection

Audio samples are stored in a Google Drive folder at:
/content/drive/MyDrive/samples
Each audio file is labeled with one of four fault types: Arcing, Corona, Looseness, or Tracking. The file names include fault-type keywords, which are later used for labeling and organizing the data into class-specific subfolders.

2.2 Spectrogram Generation

For each audio file, the Short-Time Fourier Transform (STFT) is computed. The amplitude spectrum is then converted to decibels to form a spectrogram. These raw spectrogram images are saved in the directory:
/content/drive/MyDrive/spectrogram_images
with separate subfolders for each fault type.

2.3 Feature Extraction and Data Augmentation

The audio signals are segmented into overlapping chunks (e.g., 0.3–0.5 seconds with 20% overlap) to capture short-term features. For each segment, the following features are extracted:

- **MFCCs:** Typically, 13 coefficients capturing the spectral envelope.
- **STFT Magnitude:** Converted to a decibel scale.
- **Phase Information:** Captures the phase shift in the STFT.

Data augmentation techniques are then applied to each audio file, including:

- **Time Stretching:** Slightly speeding up or slowing down the audio.
- **Pitch Shifting:** Shifting the frequency content up or down.
- **Noise Addition:** Introducing slight random noise.

These operations increase the diversity of the dataset, allowing the model to learn more robust features.

2.4 Image Resizing to Square Format

Each feature matrix (initially of rectangular shape) is normalized and resized to a square dimension (e.g., 34×34 or 64×64 pixels) using interpolation. This step ensures that all images fed into the CNN have consistent dimensions. The resulting square images are stored in:
/content/drive/MyDrive/square_feature_images

3. Convolutional Neural Network Architecture

A convolutional neural network (CNN) is employed to classify the processed square images into the four fault categories. The architecture consists of:

- **Convolutional Layers:**
  Three layers with 32, 64, and 128 filters extract hierarchical features from the input images.
- **Pooling Layers:**
  Max pooling layers reduce the spatial dimensions and help in feature generalization.
- **Dropout Layers:**
  Dropout (with rates of 25–50%) is used after convolutional and dense layers to mitigate overfitting.
- **Fully Connected Layers:**
  The flattened features are passed through a dense layer (with 128 neurons) before reaching the output layer.
- **Output Layer:**
  A softmax layer provides probability distributions over the four fault classes.

The network is compiled with the Adam optimizer and uses categorical cross-entropy as the loss function. Early stopping based on validation loss is applied to prevent overfitting.

4. Experimental Results

4.1 Training and Validation Performance

The CNN is trained over multiple epochs (typically 20–50) with a batch size (e.g., 16 or 32) using the preprocessed square images. The dataset is split into training, validation, and testing sets (70% training, 15% validation, 15% testing).

Figure 4 shows the training and validation accuracy curves over epochs, and *Figure 5* displays the corresponding loss curves. These graphs provide insights into the convergence behavior of the model and indicate that the model achieves a test accuracy 80.76%.

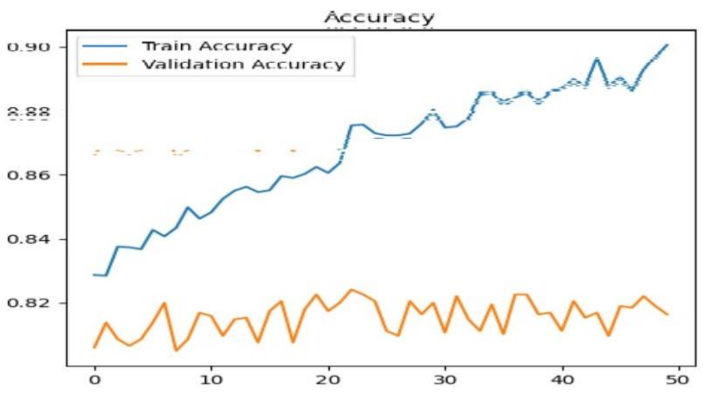


Figure 1 Training Validation Accuracy Curve

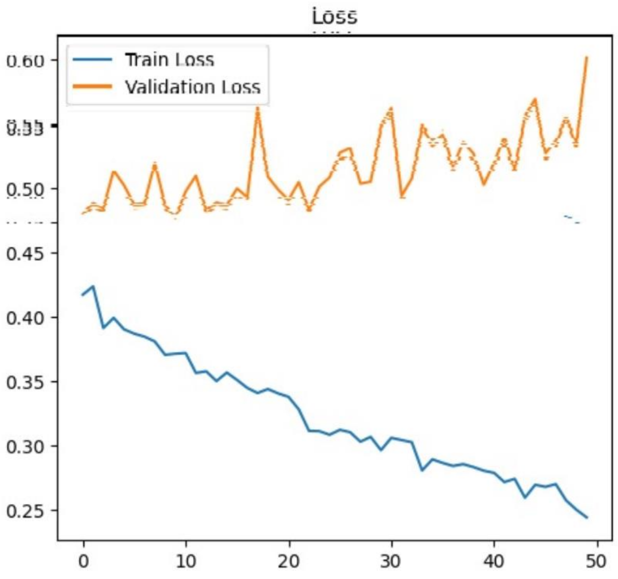


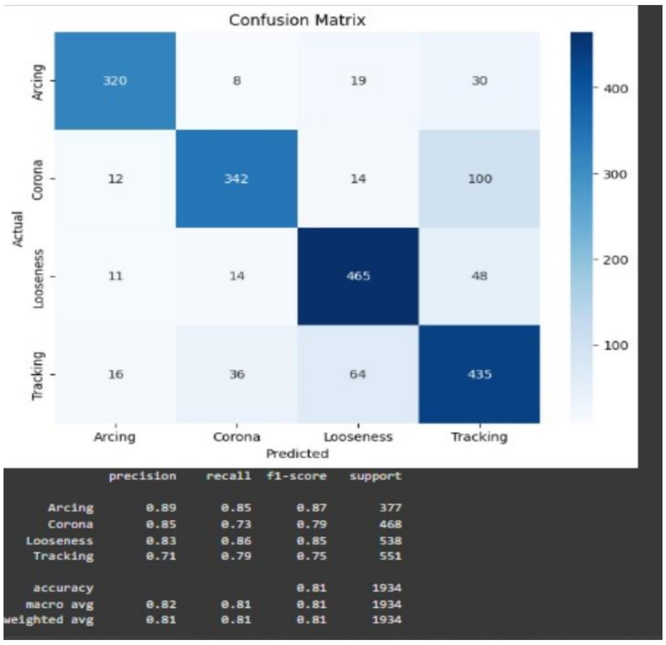
Figure 2 Training validation loss cuve

4.2 Confusion Matrix and Classification Report

A confusion matrix is computed on the test set to evaluate the classification performance across the four fault types. The classification report summarizes precision, recall, and F1-score for each class:

- **Arcing:** High precision and recall.
- **Corona:** Slightly lower recall.
- **Looseness:** Balanced performance.
- **Tracking:** Moderate performance, with room for improvement.

Figure 3 present the confusion matrix, and additional tables (or text) should summarize the classification report details.



```
              precision    recall  f1-score   support

      Arcing       0.89      0.85      0.87       377
      Corona       0.85      0.73      0.79       468
   Looseness       0.83      0.86      0.85       538
    Tracking       0.71      0.79      0.75       551

    accuracy                           0.81      1934
   macro avg       0.82      0.81      0.81      1934
weighted avg       0.81      0.81      0.81      1934
```

Figure 3 confusion matrix

5. Discussion

The experimental results confirm that the proposed method combining spectrogram analysis with CNN-based classification is effective for fault detection in electrical machines. Key observations include:

- **Data Preprocessing:**
  The segmentation, augmentation, and normalization steps significantly enhance the robustness of the features extracted from the audio signals.
- **CNN Performance:**
  The chosen CNN architecture is able to learn discriminative features from the square images, achieving high classification accuracy. However, some classes (e.g., Tracking) exhibit lower recall, suggesting the need for further model tuning.

Future work may explore deeper network architectures, alternative feature extraction methods, and additional data augmentation strategies to further improve performance.

6. Conclusion

This report details an effective approach for fault detection in electrical machines through audio signal processing and deep learning. By converting audio signals into both raw and square spectrogram images and employing a CNN for classification, the system achieves a competitive test accuracy of approximately 81%. The methodology is thoroughly documented, from data acquisition and preprocessing to model training and evaluation. The results, as visualized in the provided figures and performance metrics, demonstrate the viability of this approach for real world fault diagnosis applications. Future work will focus on further optimization of the network architecture and additional feature engineering.