Name: Aya AbdElmoniem Mohamed

Id: 20221462478

Report

1. Introduction

- The primary goal is to develop a model that accurately classifies sounds into two categories: cat and dog.

- Key findings include the dataset structure, data pre-processing steps, modeling techniques used, and evaluation metrics for the model's performance.

Section 1: Dataset Overview

- The dataset consists of 115 audio files, categorized into training and test sets for both cats and dogs.

Test Set:

- 49 entries of cat sounds.
- 49 entries of dog sounds.

Training Set:

- 115 entries of cat sounds.
- 64 entries of dog sounds.

- The data is structured to ensure a balanced representation but varies in the number of samples per category.

Section 2: Data Pre-Processing

- Data loading and cleaning processes are crucial for accurate model training.

- The data is loaded from a CSV file that lists audio file names along with their corresponding labels.

- Missing values in the dataset are handled by dropping entries that do not have corresponding audio files.

- Labels are assigned as follows:

- `0` for cat sounds.

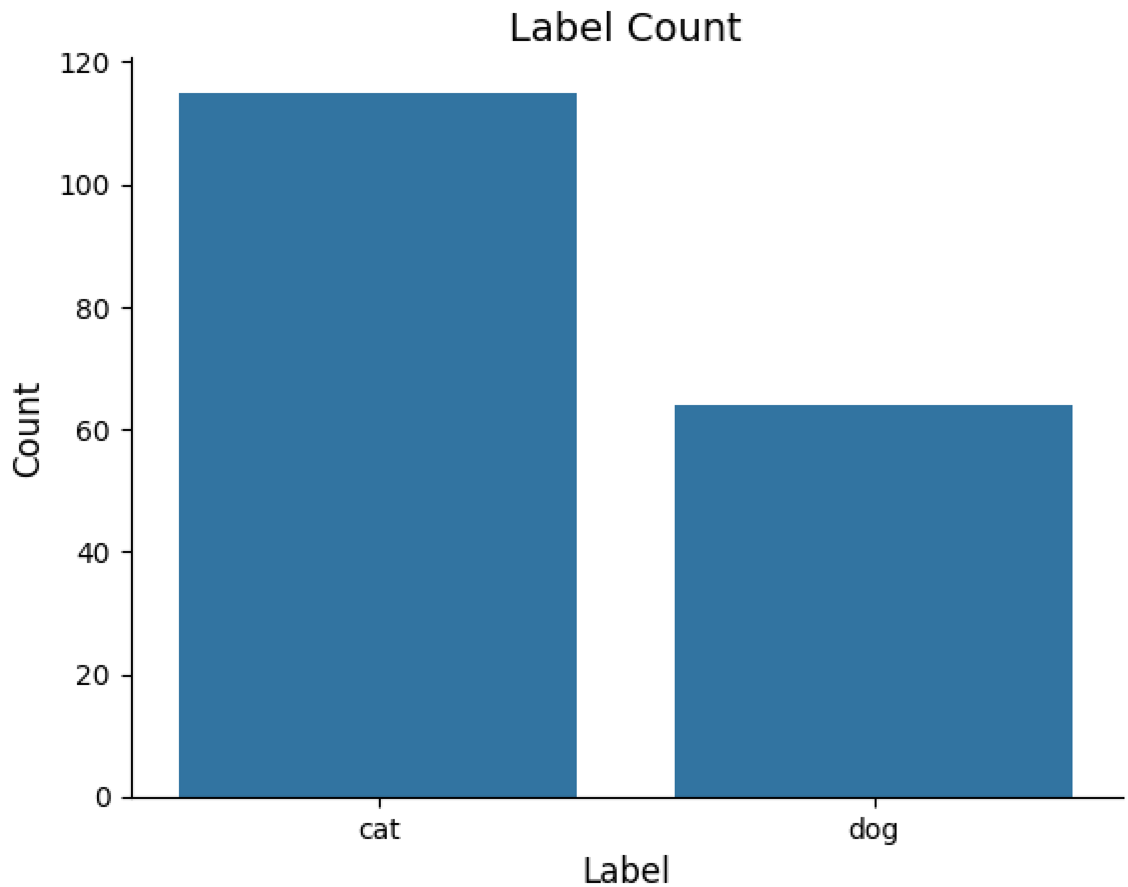
- `1` for dog sounds.

- The training and test datasets are shuffled to ensure randomness in model training.

Section 3: Data Visualization

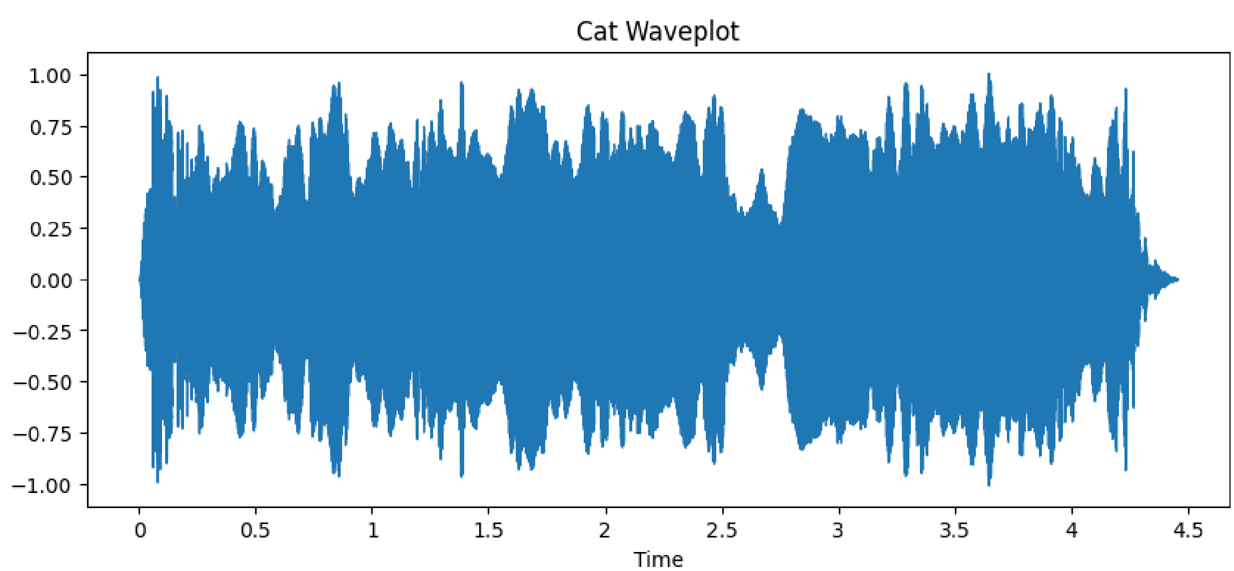
- Data visualization techniques are employed to understand the distribution of the dataset.

- A count plot illustrates the number of entries for each label (cat vs. dog).

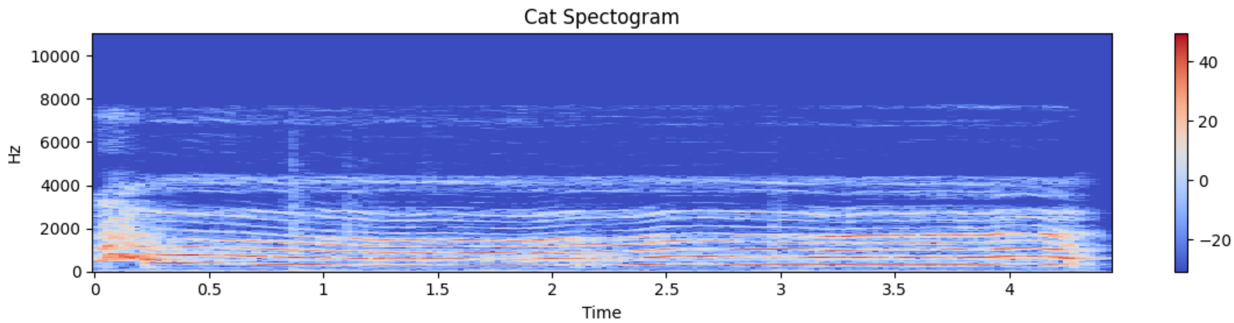


- Waveplots and spectrograms are generated for audio samples to visualize sound characteristics.

1.

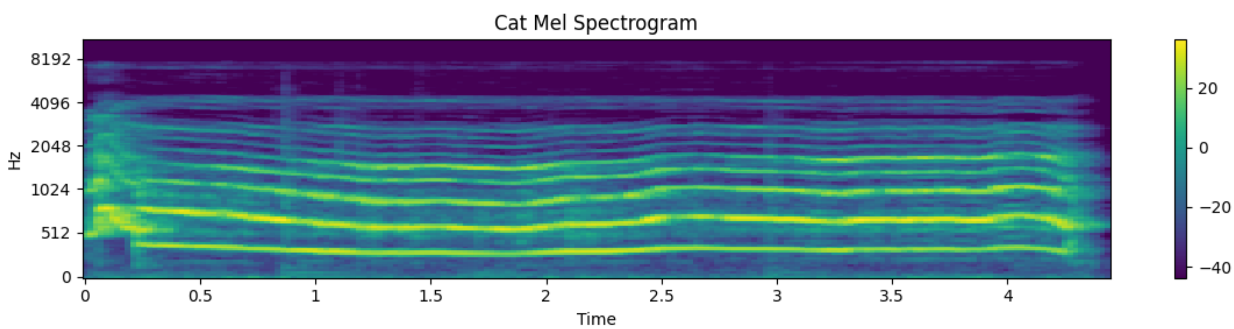


2.



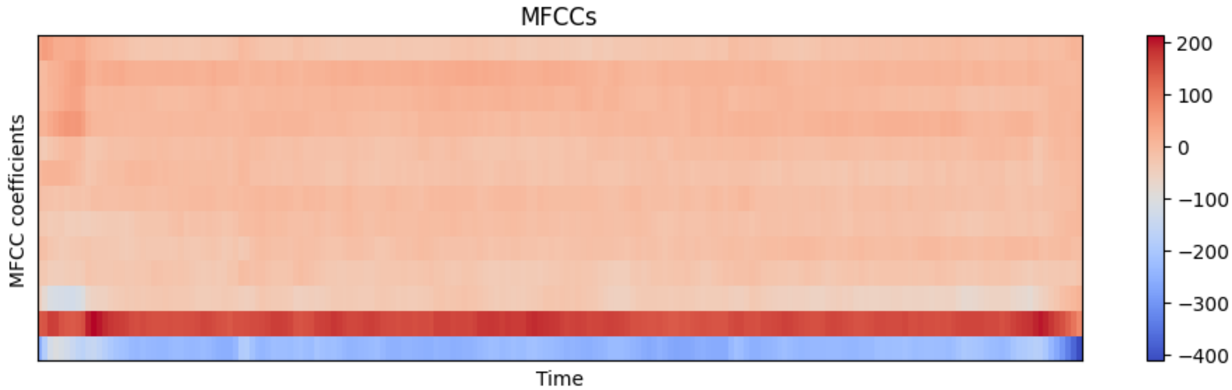
The spectrogram presented illustrates the audio characteristics of cat sounds, with the x-axis representing time in seconds and the y-axis indicating frequency in Hertz (Hz) up to 10,000 Hz. The color gradient reflects the amplitude of the sound, where lighter shades (orange and yellow) indicate higher intensity. This visual representation allows for the analysis of how the sound evolves over time and highlights specific frequency bands that may correspond to different vocalizations, such as meows or purrs. Such spectrograms are essential for audio classification tasks, as they help machine learning models identify distinguishing features of various sounds, aiding in applications like animal behavior studies and sound recognition systems.

3.



the x-axis representing time in seconds and the y-axis showing frequency in Hertz (Hz), reaching up to 8192 Hz. The color gradient indicates amplitude, with brighter colors representing higher intensity sounds. This visualization allows for the analysis of the frequency content of the cat's vocalizations over time, highlighting specific patterns and characteristics of the sounds.

4.



The x-axis represents time, while the y-axis shows the MFCC coefficients. The color gradient indicates the amplitude of these coefficients, with varying shades representing different intensity levels.

MFCCs are crucial for capturing the timbral characteristics of audio signals and are often utilized in machine learning models for tasks such as sound classification and speech analysis. This representation helps in analyzing how the sound features evolve over time, making it easier to distinguish between different audio patterns

Section 4: Feature Extraction

- Audio features are extracted using libraries like `librosa`, which provides tools for audio analysis.

- Features include Mel-frequency cepstral coefficients (MFCCs), spectrograms, and waveforms.

- Extracted features are crucial for training the classification model, as they represent the audio signals in a numerical format.

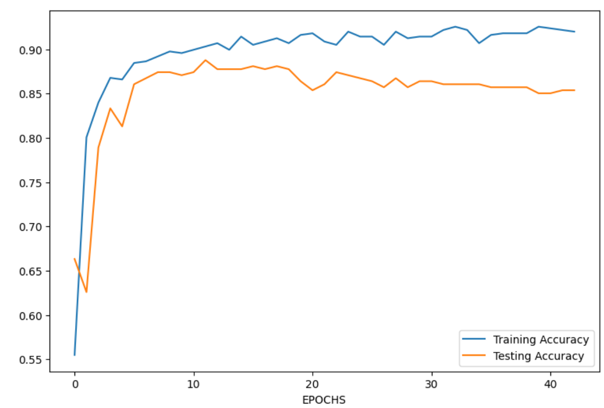
Section 5: Modeling

- Various modeling techniques are applied to classify the audio data.

- A machine learning model a Convolutional Neural Network (CNN), is trained on the extracted features.

- The model's performance is evaluated using metrics like accuracy, precision, recall, and confusion matrices.

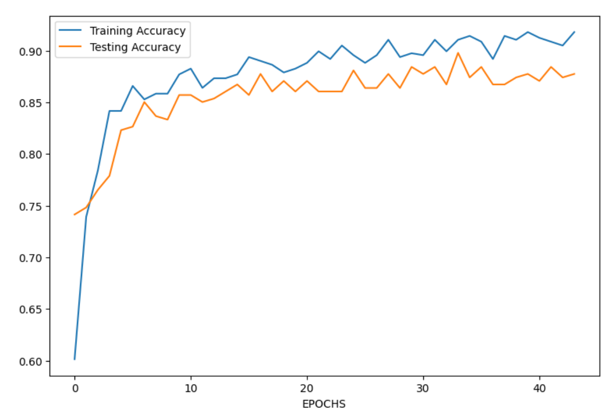
First Model plot :



With loss: 0.3409

Accuracy of the model on the test data : 87.41496801376343 %

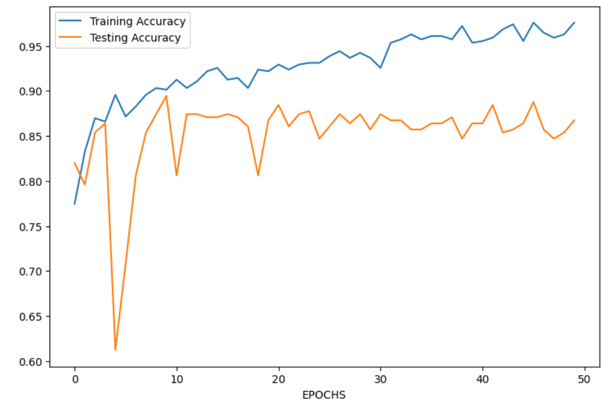
Model 2:



With loss: 0.3270

Accuracy of the model on the test data : 89.79591727256775 %

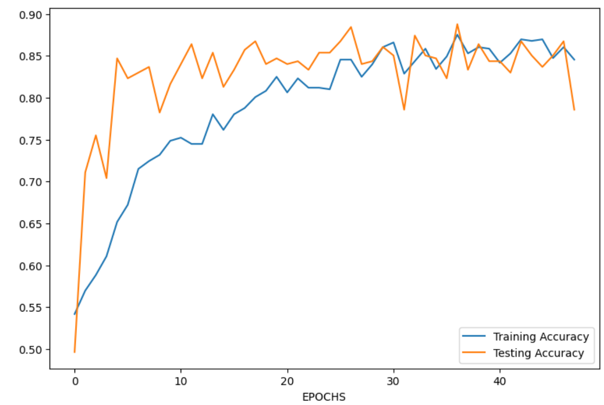
Model 3:



With loss: 0.4720

Accuracy of the model on the test data : 86.73469424247742 %

Model 4:



With loss: 0.3297

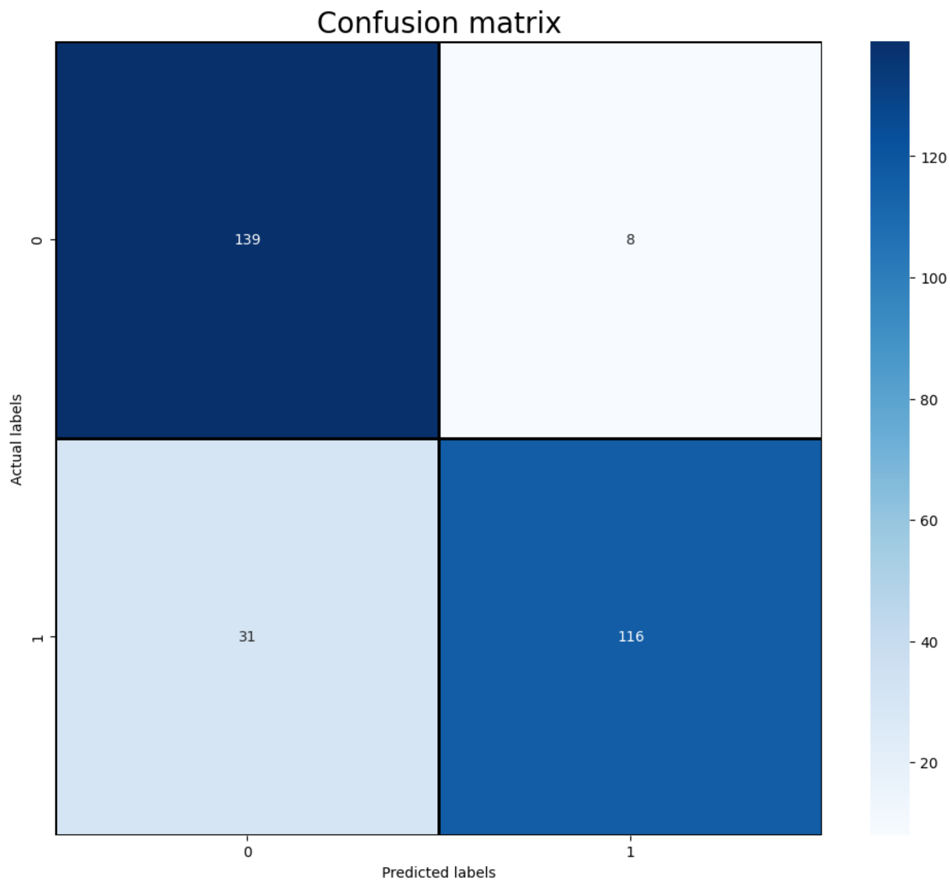
Accuracy of the model on the test data: 84.01360511779785 %

Section 6: Evaluation

- The model is evaluated on the test set to determine its effectiveness in classifying cat and dog sounds.

- Metrics such as accuracy and confusion matrices provide insights into the model's strengths and weaknesses.

- The results indicate areas for improvement and potential adjustments to the model or data processing techniques.



the calculation of key metrics such as accuracy, precision, and recall, providing insights into the model's performance. The color gradient enhances the visualization, with darker shades indicating higher counts. Overall, the matrix suggests that the model performs well, particularly in classifying instances of class 1, though some misclassifications are evident