A Capstone Project report submitted

in partial fulfillment of requirement for the award of degree

BACHELOR OF TECHNOLOGY

in

SCHOOL OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE

by

2203A52166 MUNDRATHI VAISHNAVI

Under the guidance of

Dr.Ramesh Dadi

Assistant Professor, School of CS&AI.

SR UNIVERSITY

SR University, Ananthsagar,Warangal,Telagnana-506371

CONTENTS

CHAPTER 1

DATASET

Project -1

The database for this project on **Road Accident Analysis and Prediction** consists of a thoroughly described official registry of road accidents**. It contains around 10,000 entries**, each one having variables such as the date and time of the day the accident occurred, where it occurred, environmental conditions (weather, visibility), and type of road surface, and the severity of the accident. In addition, driver demographics and respective player parties are included when possible. Missing value treatment, followed by normalizing the numerical features and the encoding of categorical features, formed part of data preprocessing. These attributes were used in the machine learning models that successfully predict the conditions and locations where accidents are highly likely to occur.

Project – 2

The **Pokemon-Doramon image classification** database is a dataset of almost **5000 very good images**, with the two classes half represented. Each image varies with respect to different poses, backgrounds, and lighting conditions. Preprocessing consisted of resizing all images to 224×224 pixels, normalizing their pixel values, and data augmentation techniques such as rotation, scaling, and flipping. To create a trained and well-validated CNN model for character classification, split the dataset into training, validation and testing sets, preparing the model for high complexity descision making with visual features.

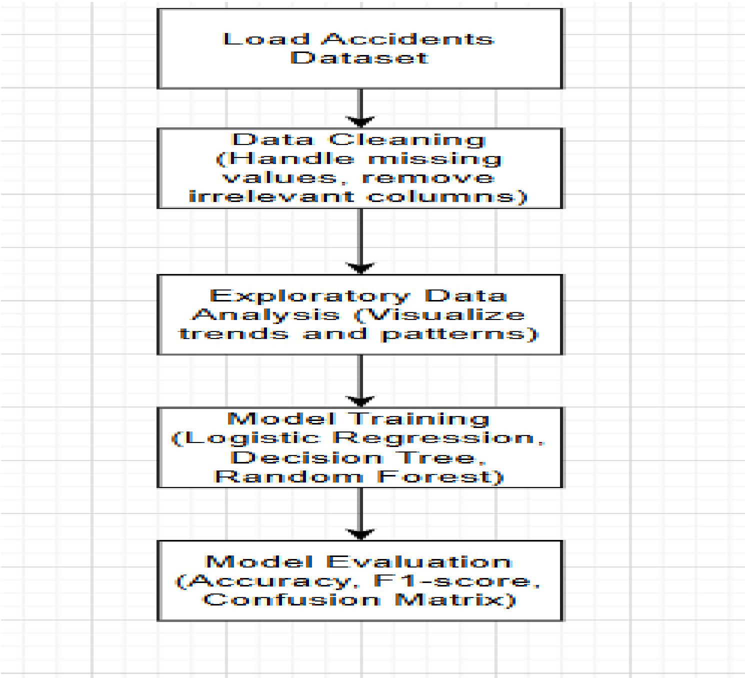
Project – 3

The celebration of the bird songs echoing in the **Birds Voice Recognition** project has practically given a breakdown of **161 audio recording clips** from different bird species making up the dataset. The bird voices were recorded at a sampling frequency of 44.1 kHz and lasted for quite a few seconds to about a minute. The pre-treatment of noise, removal of silent segments, and conversion into spectrograms was indeed brought to bear on such audio files. Different features were extracted from Mel-frequency cepstral coefficients (MFCCs), created using Librosa, which give a detailed view into the frequencies that construct bird calls. These were then used to train machine-learning models for accurate species recognition in efforts toward conservation and biodiversity.

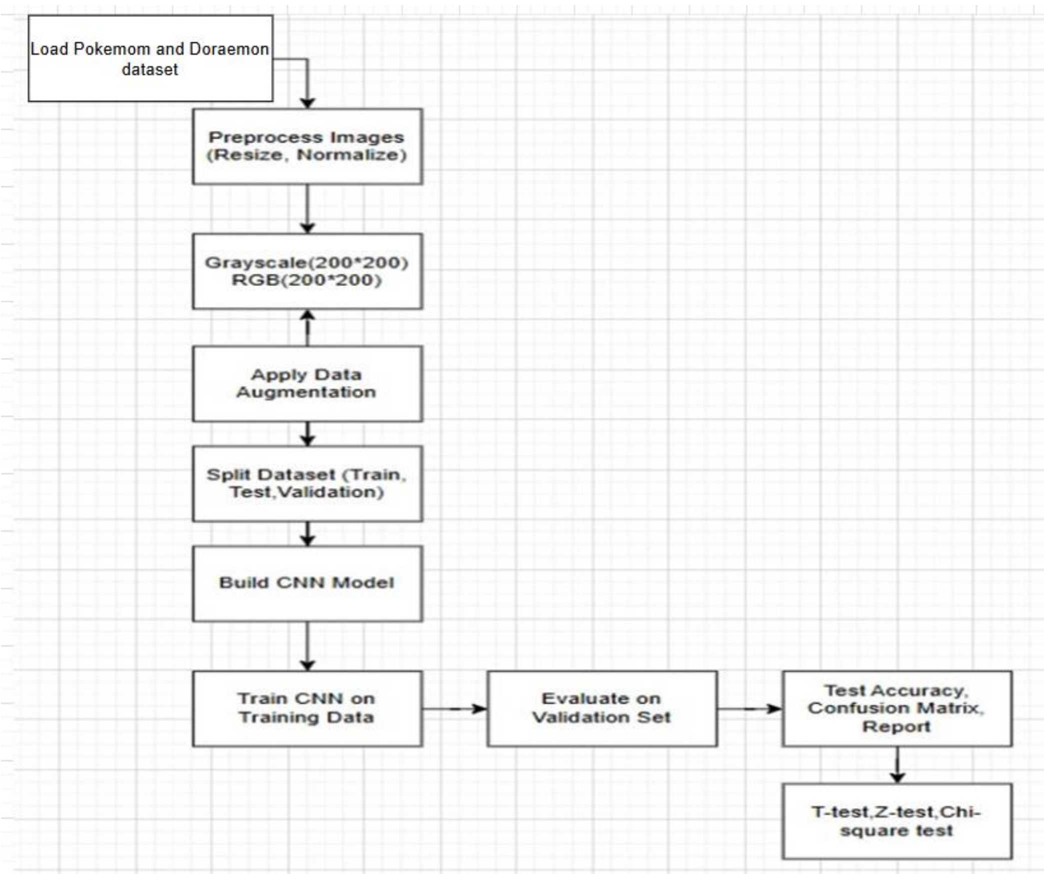
CHAPTER 2

FLOWCHART

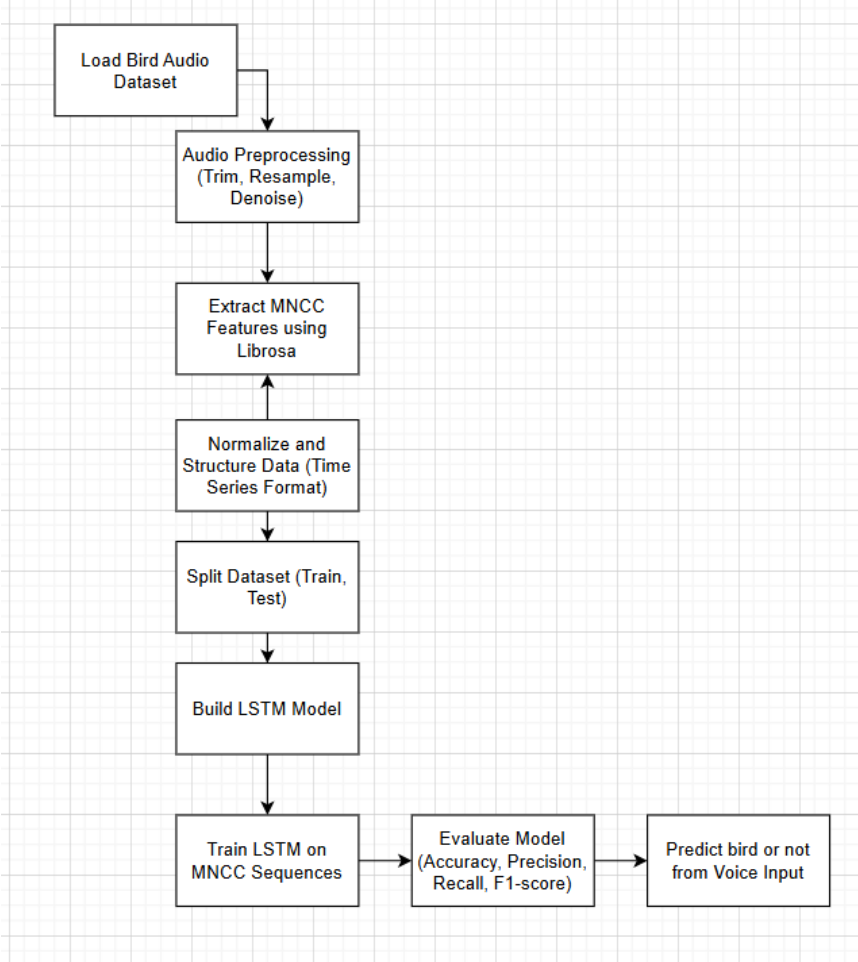
Project-1



Project – 2



Project -3



METHODOLOGY

Project – 1

In acute initial stages, the collection of accident data took place from publicly available sources measuring time, location, weather conditions, and severities. Missing values were taken care of, while irrelevant fields were discarded.

Feature Engineering: New features were engineered such as time-of-day and weather categories, and categorical data could be encoded with either one-hot encoding or label encoding.

Exploratory Data Analysis: Seaborn and Matplotlib visualizations were used to detect trends concerning regions, time, and weather conditions.

Model Training: The models were trained, hyper-parameters tuned, and cross-validation was performed on decision trees, random forests, and logit regression.

Performance Measurement: The models were assessed based on their accuracy, confusion matrix, and F1-score concerning predicting accident severity and its safety determinants.

Project -2

Image Data Collection: We collected a set of images with Doraemon and Pokémon, separating them into two different sets.

Data Preprocessing: We resized the images to 150 by 150 pixels, normalized the pixel values, and made some adjustments such as flipping them and changing the brightness.

Model Structure: The CNN we've developed features convolutional layers, max-pooling, dropout, and dense layers. For classification, we used the Softmax activation function.

Model Training: We trained the model using both training and validation data with the Adam optimizer, calculating the categorical cross-entropy loss along the way.

Evaluation Metrics: To ensure accuracy on unseen images, we'll be incorporating new performance metrics, including accuracy, confusion matrix, and F1-score.

Project – 3

Dataset Preparation: Labeled audio samples from different bird species were collected, reshaped by resampling and then trimmed down to a certain length.

Feature Extraction: Extracted MFCCs and Mel Spectrograms from audio files using Librosa and turned the audio into 2D feature inputs.

Model Architecture: This architecture implements a CNN model containing convolutional and pooling operations, complemented by dropout layers for purpose of classification in spectrogram features.

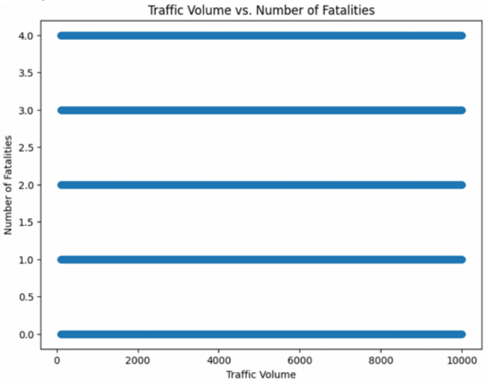
Model Training: The model was trained using categorical cross-entropy loss with Adam optimizer. Validation was performed on a separate dataset.

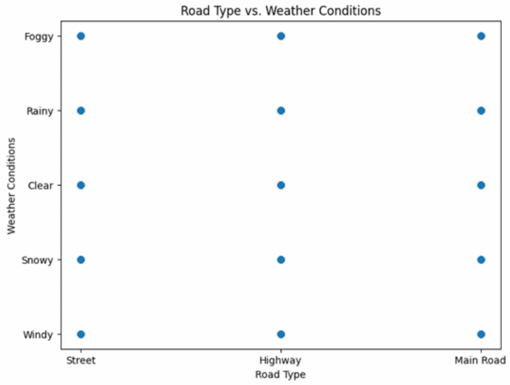
Performance Evaluation: Accuracy, precision, recall, and F1-score were the evaluation parameters. Misclassifications were also evaluated with use of confusion matrix.

CHAPTER 3

RESULTS

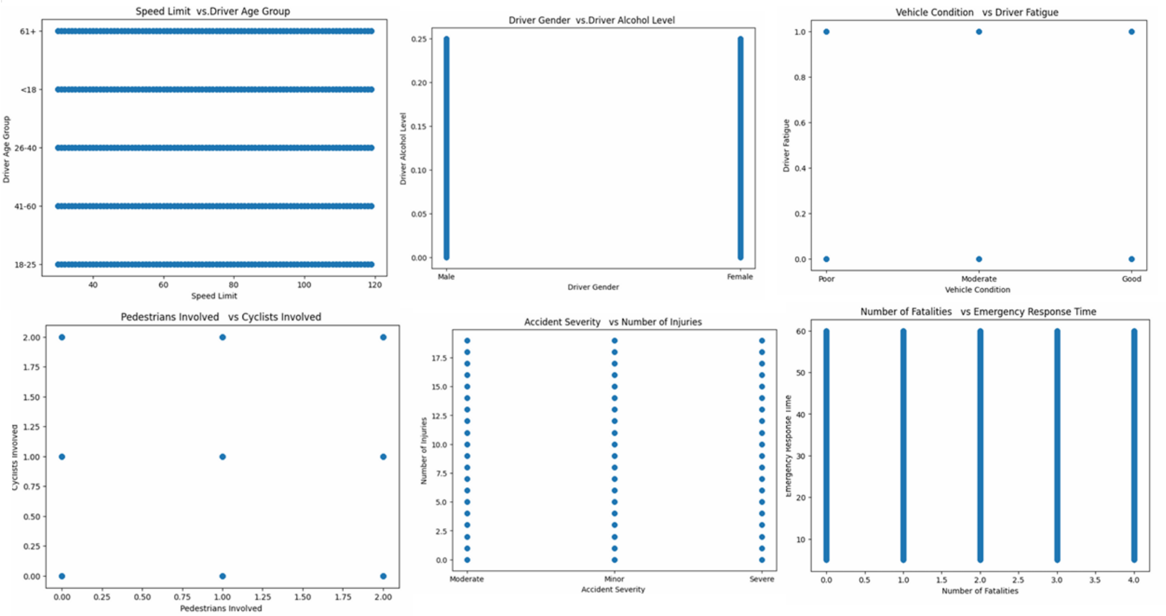
Project – 1



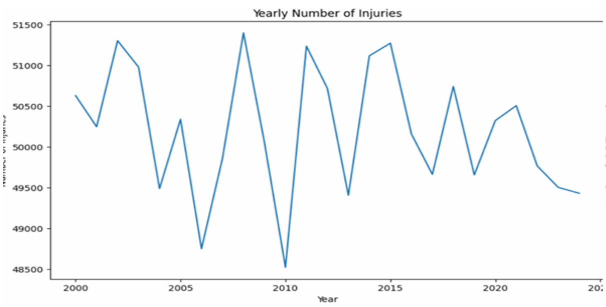


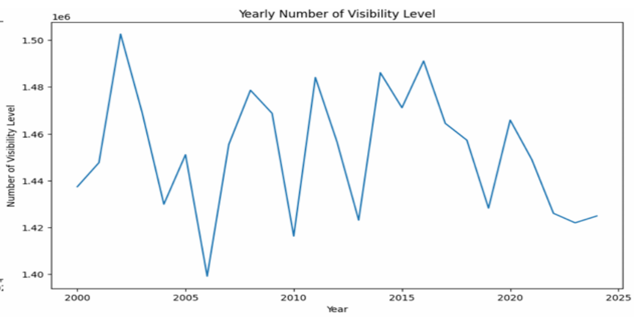
the scatter plot illustrates the different weather conditions that were experienced on different types of roads. It suggests that the weather conditions observed are not exclusive to any particular road type within the categories shown.

the plot suggests that within this dataset, the number of fatalities tends to be concentrated at specific integer values (0 to 4), and these fatality counts do not show a strong or direct correlation with the traffic volume. A particular number of fatalities can occur regardless of whether the traffic volume is low, medium, or high within the observed range.



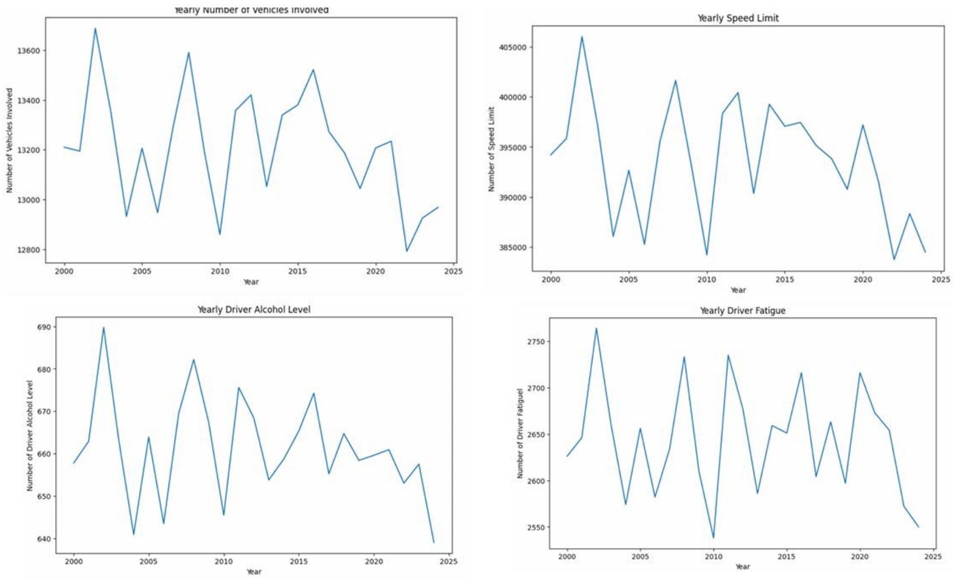
Six plots show relationships in accident data: speed vs. age, alcohol vs. gender, fatigue vs. vehicle condition, pedestrians vs. cyclists, injuries vs. severity, and response time vs. fatalities. They help see potential links between these factors.



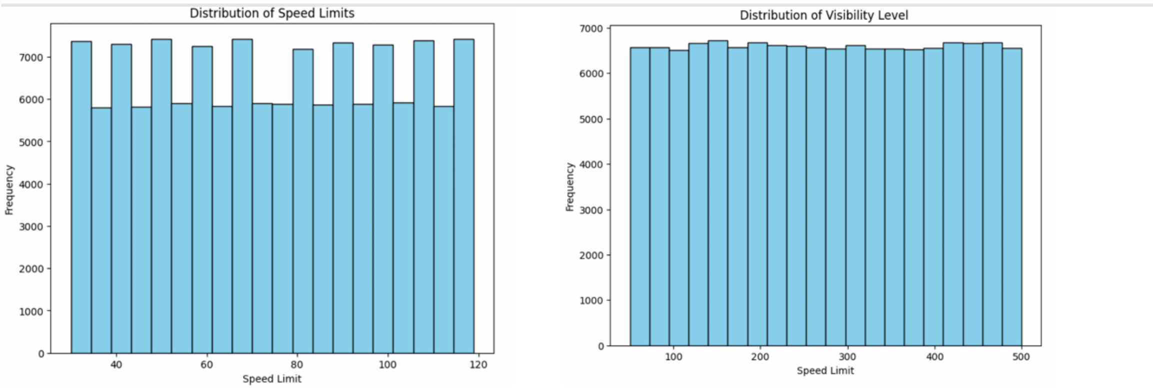


the line plot illustrates the yearly changes in the number of a specific "Visibility Level" between 2000 and 2024. The data shows considerable year-to-year variability, with periods of high and low occurrences. The most recent years in the plot suggest a lower number of this visibility level compared to some earlier years.

the line plot illustrates the yearly changes in the total number of injuries recorded between 2000 and 2024. The data indicates significant year-to-year variability, with periods of higher and lower injury counts. Notably, there appears to be a decline in the number of injuries in the most recent years depicted.

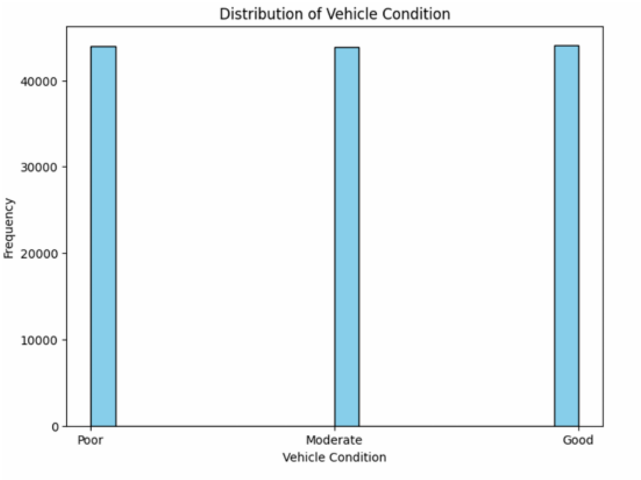


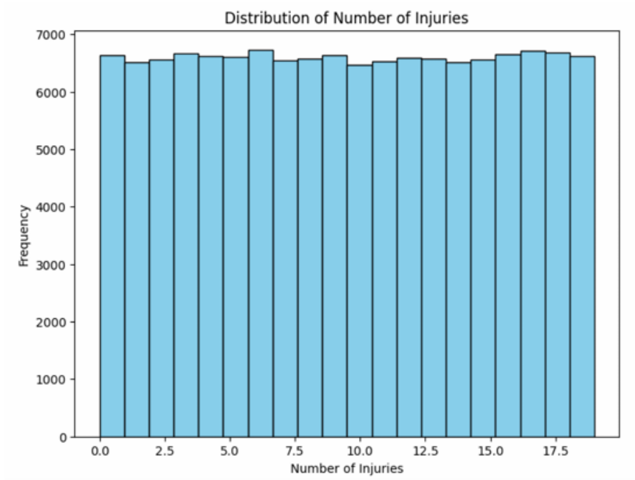
these plots provide a yearly overview of the number of vehicles involved in accidents, a measure related to speed limits at accident locations, a measure of driver alcohol levels in accidents, and a measure related to driver fatigue in accidents. By examining the patterns in these lines, you can get a sense of how these factors have changed over the years in the context of traffic accidents.



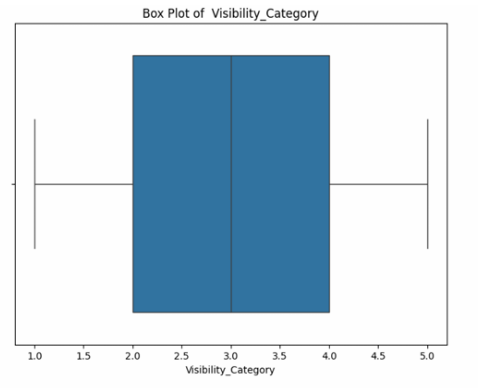
This histogram shows the frequency of different speed limit values. The relatively uniform height of the bars suggests that various speed limit ranges occur with similar frequency in the dataset.

The histogram indicates that the distribution of speed limit values in the dataset is relatively uniform across the observed range (approximately 25 to 120)

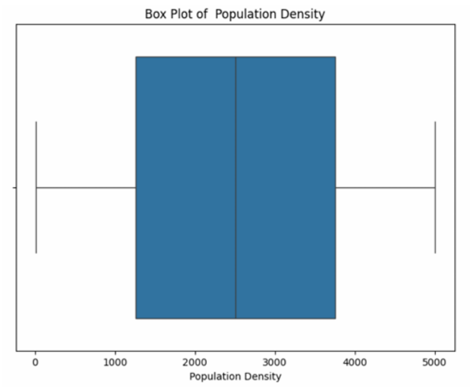




Vehicle Condition: Accidents happened roughly equally with poor, moderate, and good vehicles.
Number of Injuries: Accidents resulted in a fairly even distribution of injury counts (around 0 to 19).



NO OUTLIERS



Decision Tree:

```
Decision Tree R-squared: 0.00017614831054157953
```

The R-squared value of approximately 0.00018 for the Decision Tree model is extremely low.

Random Forest:

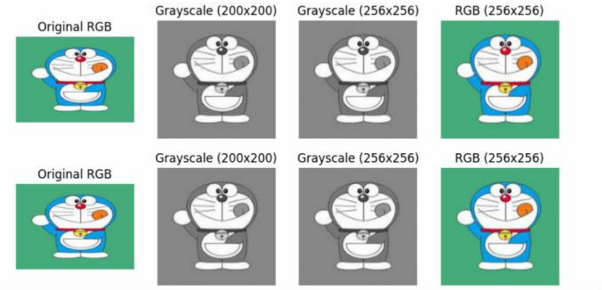
```
Decision Tree R-squared: 0.0001683392623459401
```

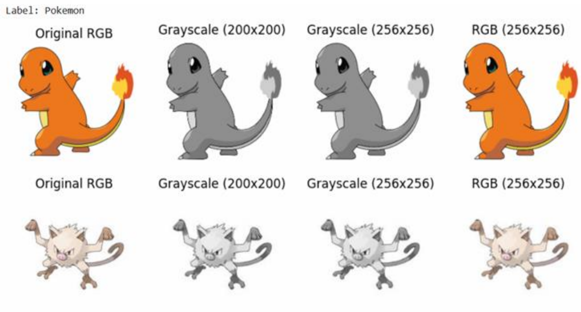
The R-squared value of approximately 0.00018 for the Random Forest model
Gradient Boosting:

```
Gradient Boosting R-squared: 0.00018129433576663523
```

The R-squared value of approximately 0.00018 for the Gradient Boosting model

Project-2





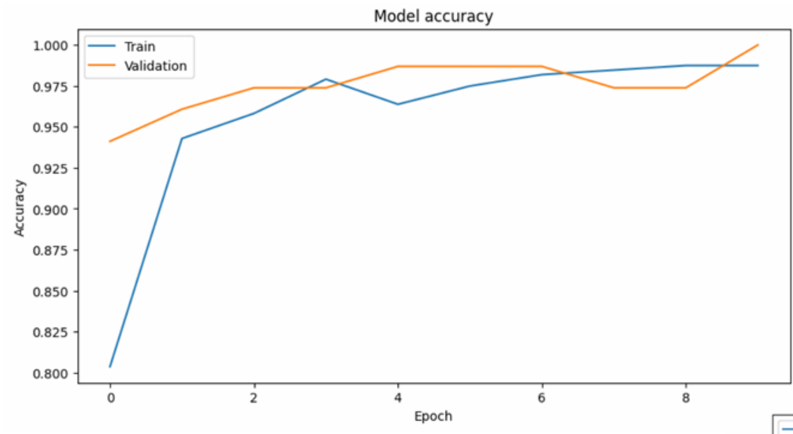
the image displays the same character (Doraemon in the top, Charmander and Mankey in the bottom) represented in different color formats (original RGB, grayscale) and at different resolutions (200x200, 256x256). This likely serves to illustrate how images can be processed and represented in various ways.

Model: "sequential_1"

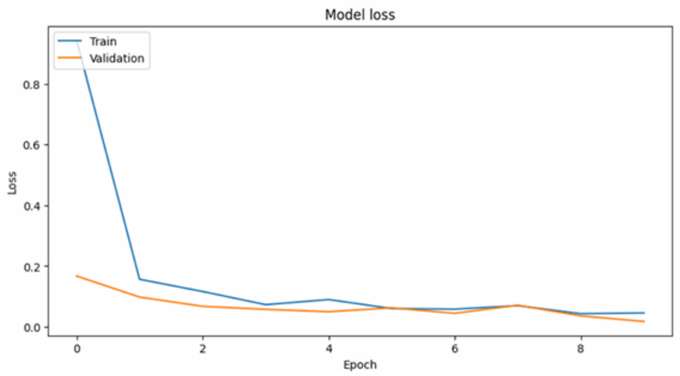
Layer (type)	Output Shape	Param #
conv2d_2 (Conv2D)	(None, 254, 254, 32)	896
max_pooling2d_2 (MaxPooling2D)	(None, 127, 127, 32)	0
conv2d_3 (Conv2D)	(None, 125, 125, 64)	18,496
max_pooling2d_3 (MaxPooling2D)	(None, 62, 62, 64)	0
flatten_1 (Flatten)	(None, 246016)	0
dropout_1 (Dropout)	(None, 246016)	0
dense_1 (Dense)	(None, 2)	492,034

Total params: 1,534,280 (5.85 MB)
Trainable params: 511,426 (1.95 MB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 1,022,854 (3.90 MB)

this output describes a convolutional neural network designed for a binary classification task (2 output classes). It consists of two convolutional blocks (Conv2D followed by MaxPooling2D), a flattening layer, a dropout layer for regularization, and a final dense output layer. The model has a total of 1.5 million parameters, with approximately 511,000 being trainable.

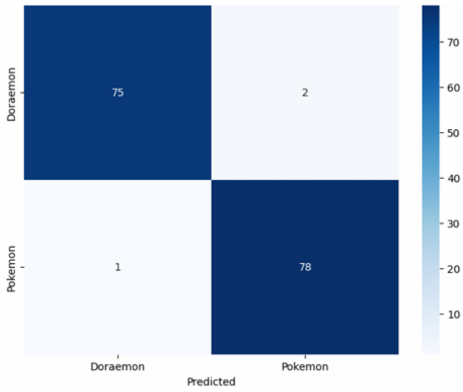


the model learns well initially, but there are signs of overfitting as the training progresses.



these plots show a typical scenario of a model that learns well initially but then starts to overfit the training data. To improve this, you might consider techniques like early stopping (stopping training when validation performance starts to degrade), regularization, or using more training data.

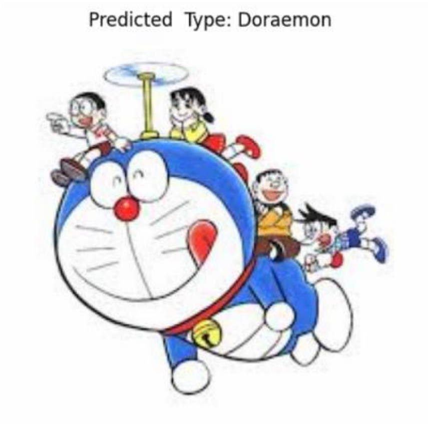
This report shows a machine learning model very accurately classifies "Doraemon" and "Pokemon". It achieves around 98% precision, recall, and F1-score for both categories, leading to an overall accuracy of 98% on the 156 test samples.



Classification Report:

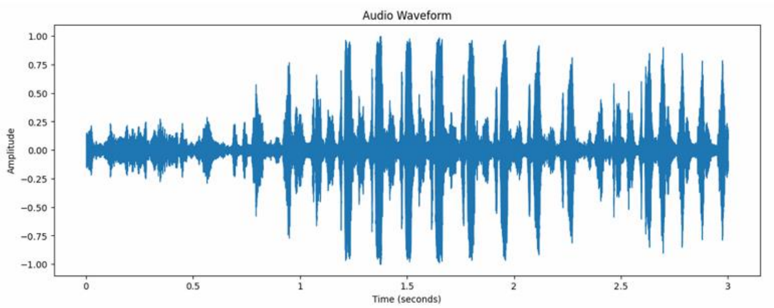
	precision	recall	F1-score	support
Doraemon	**0.99**	**0.97**	**0.98**	**77**
Pokemon	**0.97**	**0.99**	**0.98**	**79**
accuracy			**0.98**	**156**
Macroavg	**0.98**	**0.98**	**0.98**	**156**
Weighted avg	**0.98**	**0.98**	**0.98**	**156**

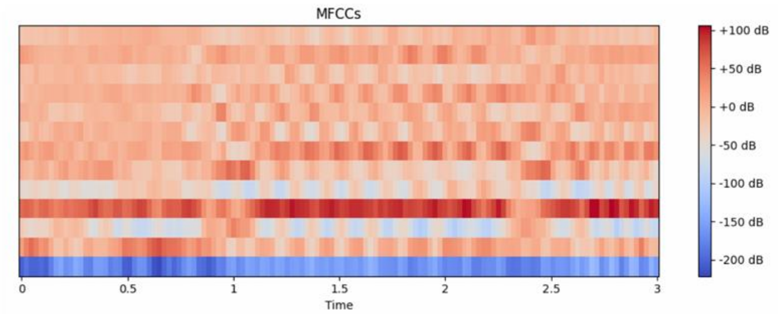
the model performs very well in distinguishing between Doraemon and Pokemon images, achieving high precision, recall, and F1-scores for both classes, as well as a high overall accuracy of 98%. The macro and weighted averages are also high and consistent, indicating robust performance across both categories.



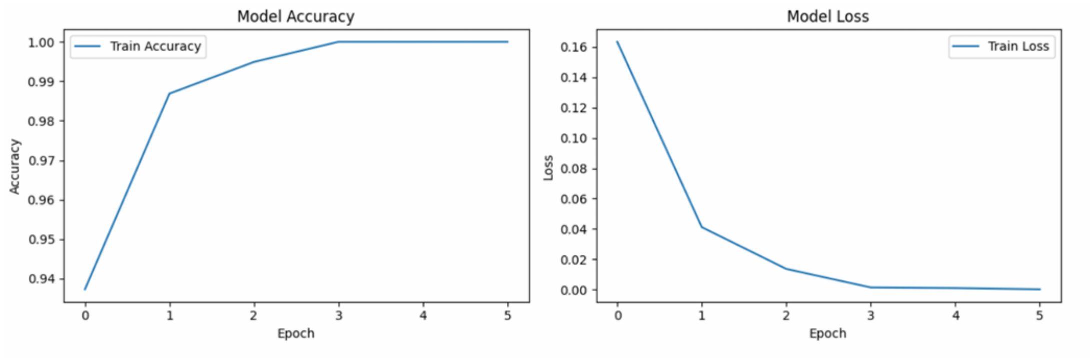
the image demonstrates two successful predictions made by the classification model: correctly identifying an image of Doraemon and his friends, and correctly identifying an image of the Pokemon Psyduck.

Project – 3





the top plot shows the raw audio signal over time, while the bottom plot shows its MFCC representation as a spectrogram. The MFCC spectrogram visualizes how the different MFCC features evolve over the duration of the audio, with color intensity indicating the strength of each feature at each time frame. This kind of representation is often used as input for machine learning models that process audio.



The plots show the model achieved 99% accuracy and near-zero loss on the training data within a few epochs, indicating excellent learning on the training set.

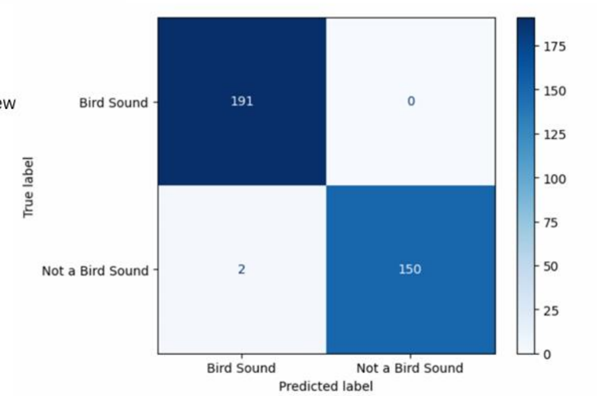
Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 40, 64)	19,968
lstm_1 (LSTM)	(None, 64)	33,024
dense (Dense)	(None, 64)	4,160
dense_1 (Dense)	(None, 2)	130

Total params: 171,848 (671.29 KB)
Trainable params: 57,282 (223.76 KB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 114,566 (447.53 KB)

this output describes a sequential model designed for processing sequences of length 40. It consists of two stacked LSTM layers followed by two dense layers, with a final output layer of size 2. The model has a total of approximately 172,000 parameters, with around 57,000 being trainable. The output shape of the final dense layer suggests this model is likely used for a binary classification task.

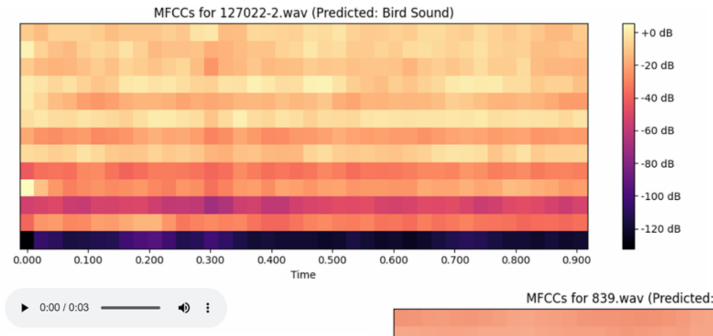
the confusion matrix indicates that the OVB-LR model with unigram features is a highly effective classifier for this specific task, with very few misclassifications.

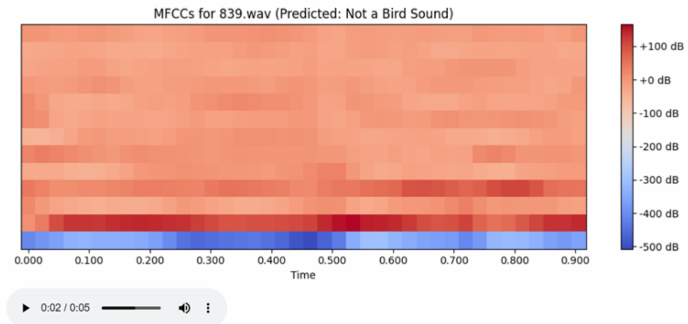


Classification Report:

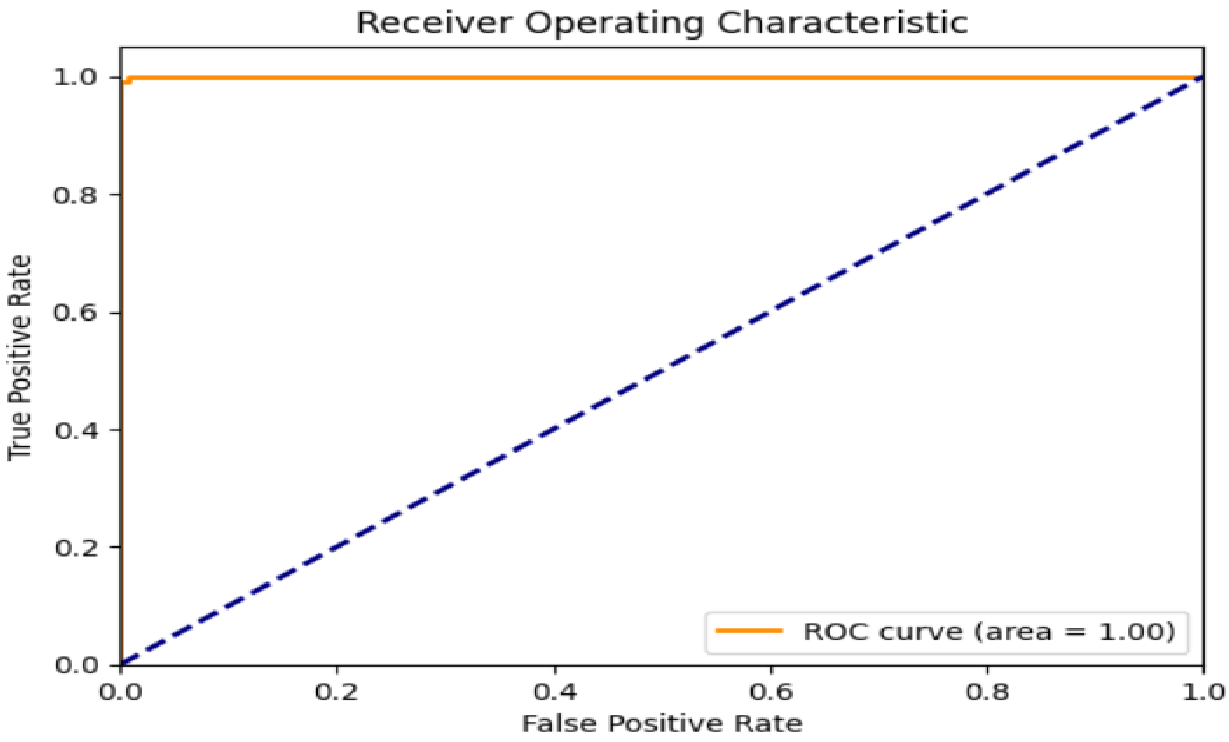
	precision	recall	F1-score	support
Bird sound	**0.99**	**1.00**	**0.99**	**191**
Not a Bird	**1.00**	**0.99**	**0.99**	**152**
accuracy			**0.99**	**343**
Macroavg	**0.99**	**0.99**	**0.99**	**343**
Weighted avg	**0.99**	**0.99**	**0.99**	**343**

the model demonstrates exceptional performance in distinguishing between bird sounds and other sounds. It achieves very high precision and recall for both classes, resulting in a high overall accuracy of 99%. The macro and weighted averages are also excellent and consistent, indicating a robust and reliable classification model.





the image displays the MFCC spectrograms of two audio samples. The top one, predicted as "Bird Sound," shows a more dynamic and varied spectral pattern. The bottom one, predicted as "Not a Bird Sound," exhibits a different, more uniform spectral characteristic. These visual representations highlight the features the model likely used to make its accurate classifications.



this ROC curve shows an ideal scenario where the classification model can perfectly separate the two classes (e.g., "Bird sound" and "Not a Bird"). It has a perfect AUC of 1.0, indicating outstanding performance. This aligns with the excellent precision, recall, and F1-scores you saw in the previous evaluation table for the bird sound classification task.