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

2203A52169 **PININTTI SUNIL**

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

Kaggle supplied the data that was utilized in the Netflix Show Recommender System. The dataset contains a wealth of rich and detailed information on 8,800+ **movies and TV shows on Netflix**. Each entry contains relevant fields such **as title, director, cast, genres, country of origin, year, rating, duration, and short description.** After preprocessing, these attributes were combined together to create a content-based recommendation system. The textual information was vectorized using CountVectorizer, and the similarity of each title was calculated using cosine similarity. The missing or null values were sufficiently handled in the preprocessing phase for data integrity and model accuracy.

Project – 2

The Rice Image Classification task consists of a dataset complete with around 7,500 beautiful, high-resolution images of five different types of rice: **Arborio, Basmati, Ipsala, Jasmine, and Karacadag**. The images were captured in a controlled environment for a standard background for clarity and consistency. There are between 1,200 and 1,500 images for each type of rice. After preprocessing, all images were resized to 150×150 pixels, normalized, and label-encoded to suit the CNN model architecture. To make it more generalizable, we utilized different data augmentation strategies, i.e., rotations, zooms, and horizontal flips. The well-formatted image dataset is ideal for the CNN model to learn the spatial patterns and properly distinguish between the types of rice.

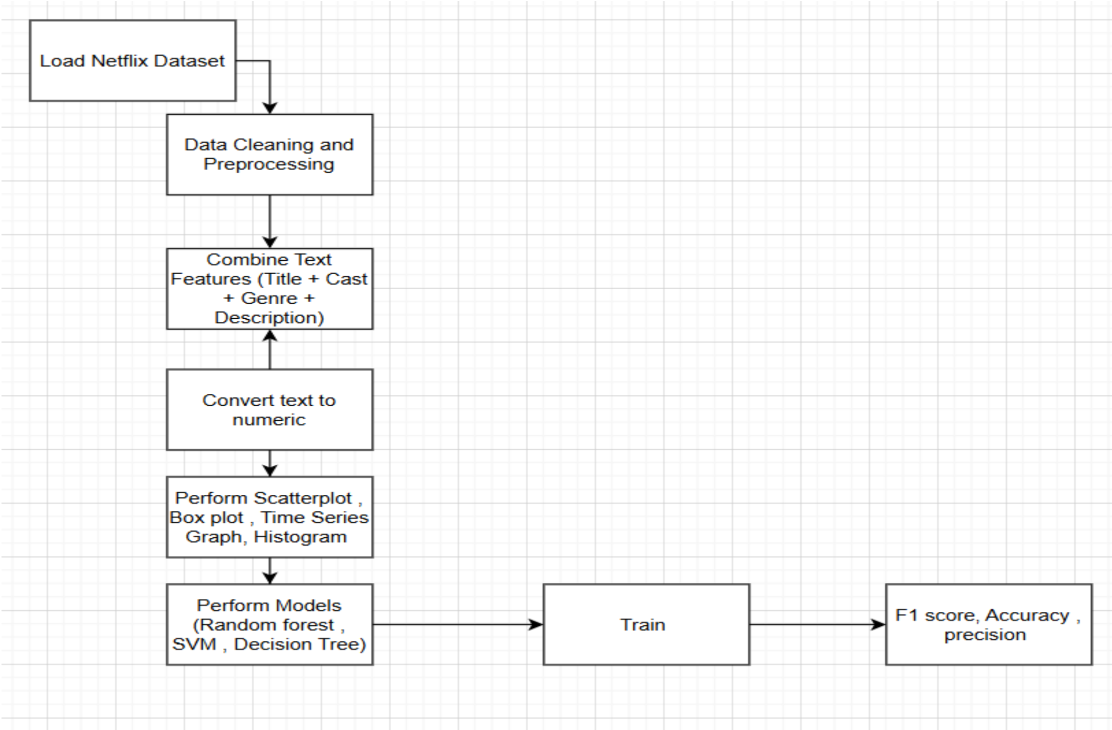
Project – 3

The Gender Voice Recognition project obtained a voice dataset consisting of about **3,168 audio samples**, which were classified as either male or female. Once the audio samples were obtained, the data was preprocessed to extract the audio features by using **Modified Normalized Cepstral Coefficients (MNCC).** These MNCC parameters were used as the input to the LSTM model. The voice data was recorded at a sampling rate of **44.1 kHz,** with each clip lasting between 1 to 3 seconds. The cleaning process involved trimming silences, eliminating background noise, extracting time-series features with librosa, and normalizing the data. The MNCC features were specifically structured into sequential arrays to give the LSTM model insights into temporal voice patterns, allowing it to predict gender effectively.

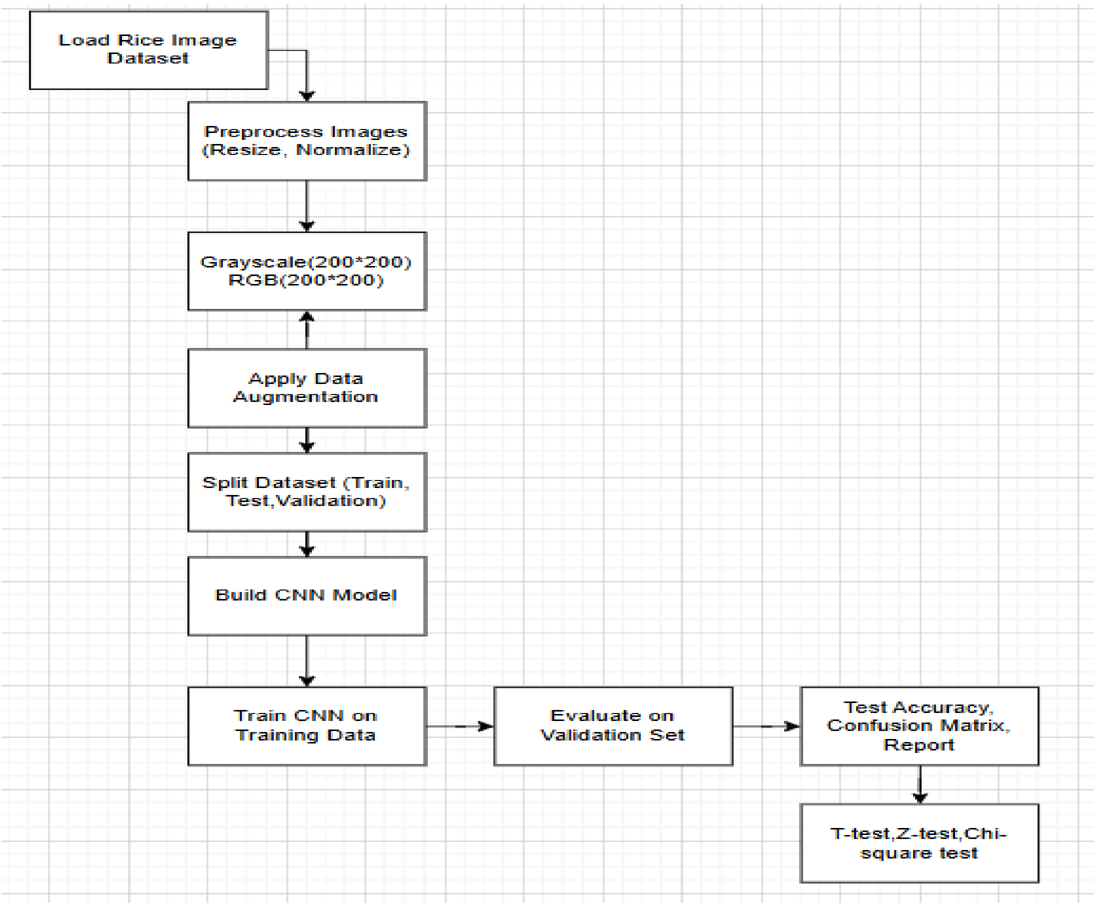
CHAPTER 2

FLOWCHART

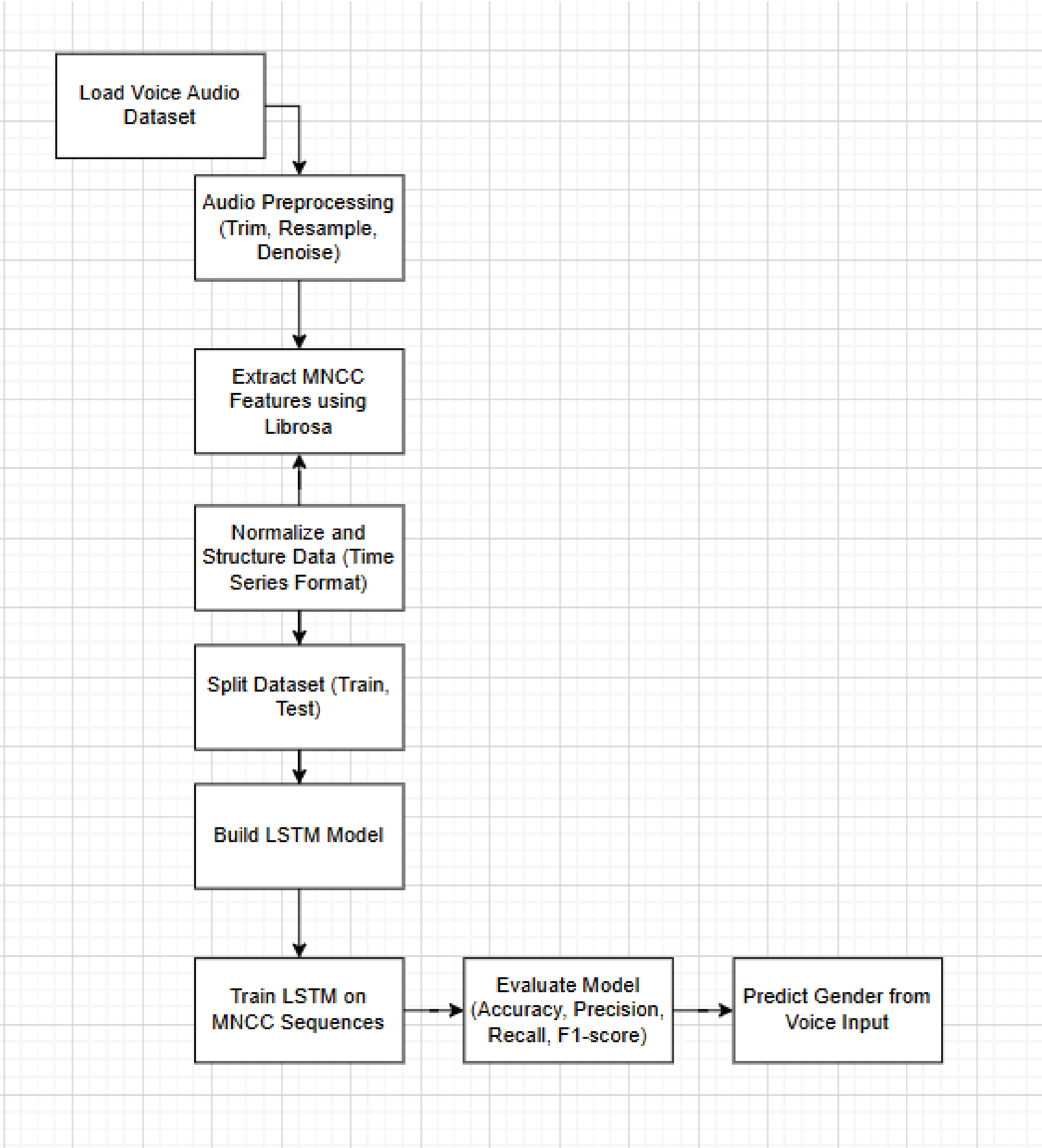
Project-1



Project – 2



Project -3



METHODOLOGY

Project – 1

Dataset Preparation: The Netflix dataset includes fields like title, director, cast, genre, and description. Initially, null values were handled, and irrelevant columns were removed.

Data Preprocessing: All text-based features were combined to form a single composite feature that encapsulates the essence of each show. Text cleaning involved lowercasing, removing stopwords, and punctuation.

Vectorization: We used CountVectorizer to convert the text data into a numerical matrix, where each row represents a show and each column represents a word frequency.

Similarity Matrix: Using cosine similarity, we calculated the similarity scores between all show combinations.

Recommendation Engine: A function was implemented to fetch the top 10 most similar shows based on the cosine similarity scores when a title is provided by the user.

Project -2

Dataset Acquisition: The rice image dataset is loaded with thousands of labeled images containing different varieties of rice grains such as Basmati, Arborio, Karacadag, and others.

Preprocessing: We ensured the resizing of images uniformly to a size of 150x150 pixels, normalized pixel values between 0 and 1, and used some augmentation techniques in order to increase the model's capacity to generalize.

Model Architecture: We set up a sequential CNN architecture featuring multiple convolutional layers paired with max-pooling layers to reduce spatial dimensions. To combat overfitting, we also added dropout layers.

Training: The model was trained using the training data with categorical cross-entropy loss and then validated against the validation data.

Evaluation Metrics: Finally, we evaluated the model's performance on the completed architecture using metrics like accuracy, confusion matrix, and F1-score.

Project – 3

Dataset Preparation: The dataset consists of male and female voice samples with different pitches, tones, and lengths.

Preprocessing: All audio samples were resampled, denoised, and trimmed to maintain uniformity.

Feature Extraction: With the help of Librosa, MNCC features were extracted, preserving frequency-based patterns from voice signals.

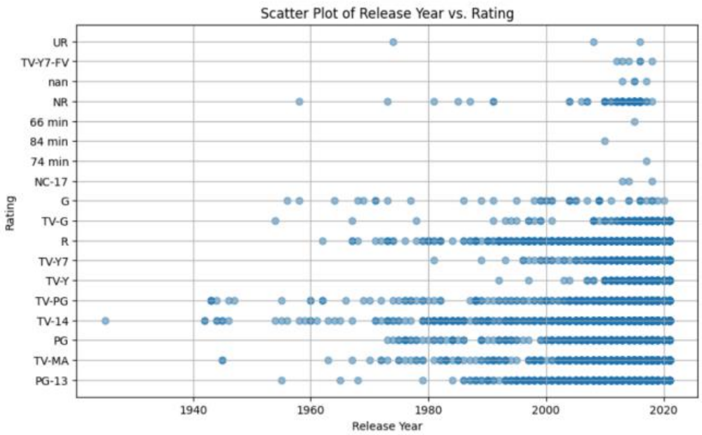
Model Architecture: To uncover the temporal patterns in the MNCC sequences, we developed a deep LSTM model. This model features a robust output layer with a sigmoid activation function for binary predictions, along with dropout layers to help avoid overfitting.

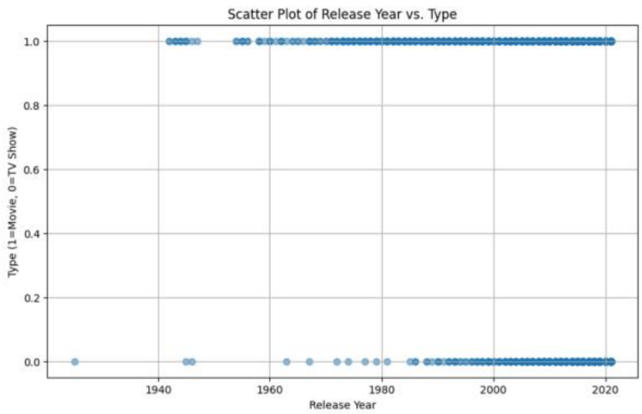
Model Training: We utilized real-time audio samples to test the model after training it using binary cross-entropy loss.

Performance Evaluation: We assessed the model's ability to predict gender by using various metrics, including accuracy, recall, precision, and F1-score.

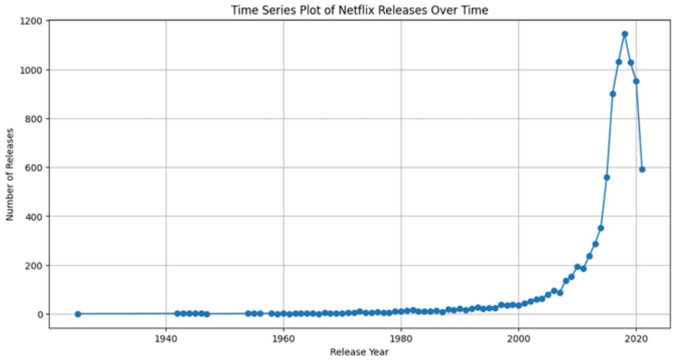
CHAPTER 3
RESULTS

Project – 1

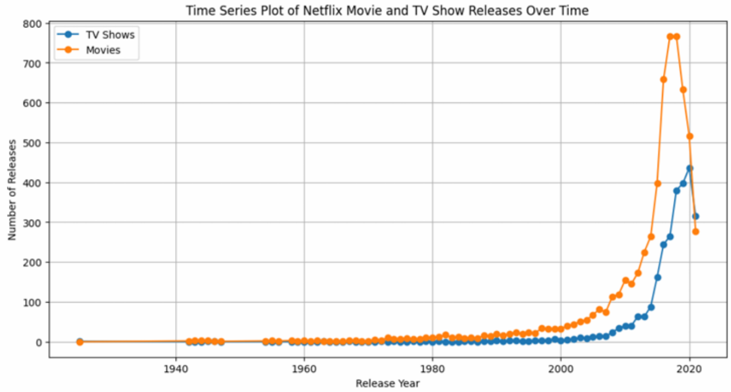




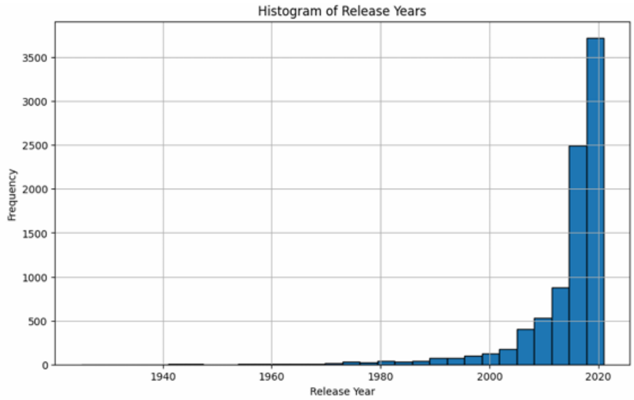
The scatter plot shows the distribution of Netflix content by release year and type, where 1 represents Movies and 0 represents TV Shows. Most of the content is movies, especially from the 2000s onwards. TV shows are fewer and began increasing mainly after 2010, indicating Netflix's recent focus on expanding its TV show library.

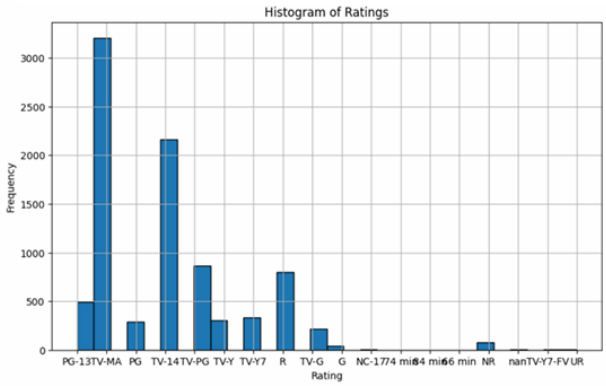


This graph shows the number of Netflix releases over time. The number of releases started rising significantly around 2010, peaked around 2020, and then declined slightly. It visually captures Netflix's content boom and recent slowdown



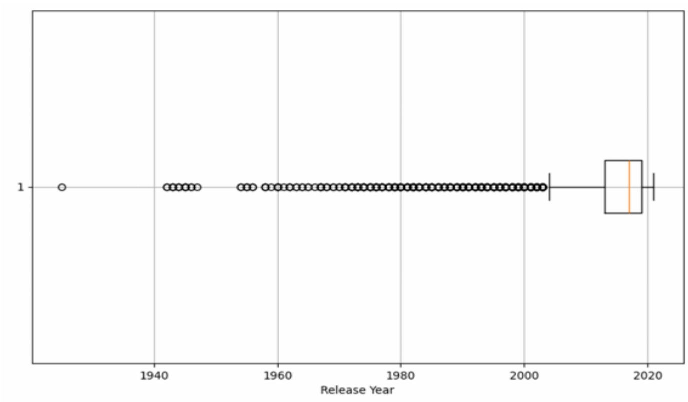
Netflix movie and TV show releases rose sharply after 2010, peaked around 2019, and then declined. Movies consistently outnumbered TV shows.

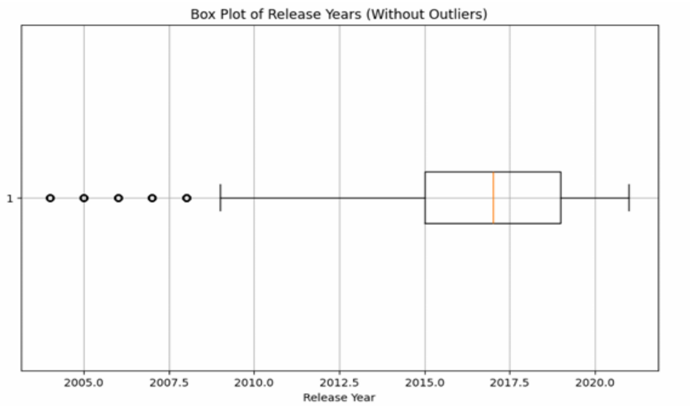




This graph shows how Netflix releases increased sharply from 2000, peaked around 2020, and then slightly declined. It reflects Netflix's growing content production and recent slowdown.

This graph shows that most content on Netflix is rated TV-MA, meaning it's for mature audiences, followed by TV-14 and TV-PG. Netflix has a strong focus on adult-oriented content.





This graph shows that most releases are from 2000 to 2020, with only a few outliers from earlier years. It highlights a surge in modern content.

This graph shows that most Netflix releases are between 2015 and 2019, with a median around 2017.5. A few older releases appear as outliers.

SVM:

	precision	recall	F1-score	support
0	**0.00**	**0.00**	**0.00**	**548**
1	**0.69**	**1.00**	**0.82**	**1214**
accuracy			**0.69**	**1762**
Macroavg	**0.34**	**0.50**	**0.41**	**1762**
Weighted avg	**0.47**	**0.69**	**0.56**	**1762**

The model excels at classifying class 1 (high precision and perfect recall, F1-score 0.82) but completely fails for class 0 (all metrics are 0.00). The overall accuracy (0.69) is driven by the better performance on the more frequent class 1.

Random Forest:

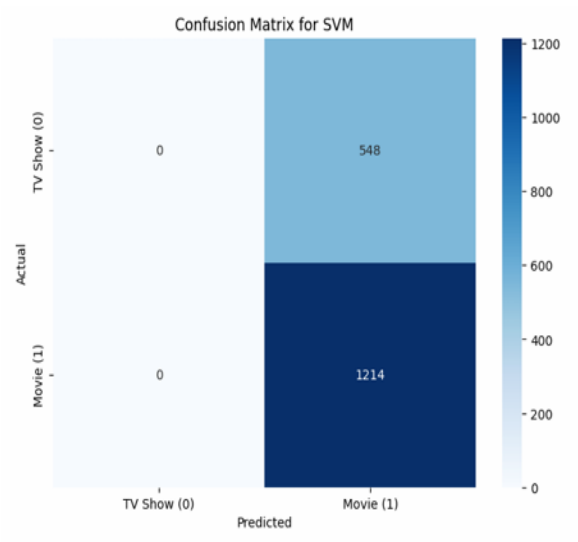
	precision	recall	F1-score	support
0	**0.56**	**0.14**	**0.22**	**548**
1	**0.71**	**0.95**	**0.81**	**1214**
accuracy			**0.70**	**1762**
Macroavg	**0.63**	**0.54**	**0.52**	**1762**
Weighted avg	**0.66**	**0.70**	**0.63**	**1762**

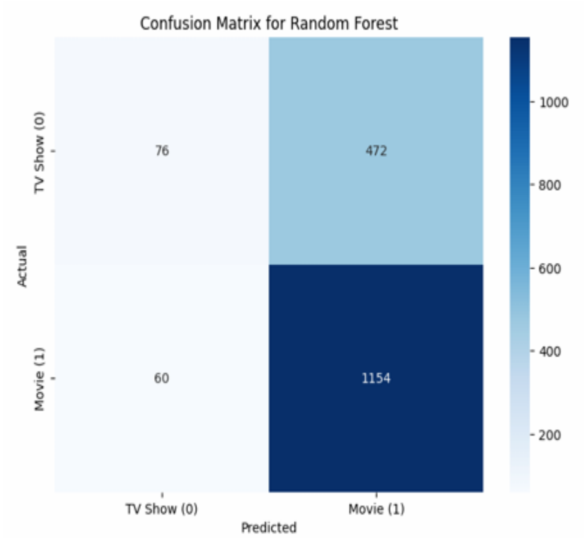
The model is good at classifying class 1 (high recall, good precision), but struggles with class 0 (low recall, decent precision). Overall accuracy is decent (0.70) but skewed by the more frequent class 1. There's a clear performance difference between the classes.

Decision Tree:

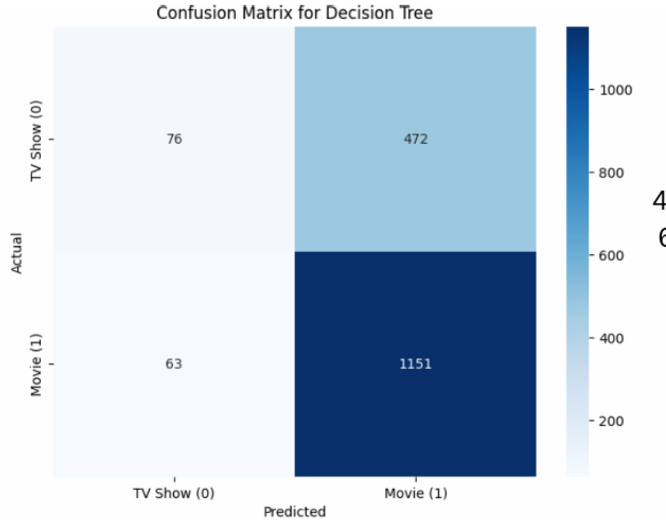
	precision	recall	F1-score	support
0	**0.55**	**0.14**	**0.22**	**548**
1	**0.71**	**0.95**	**0.81**	**1214**
accuracy			**0.70**	**1762**
Macroavg	**0.63**	**0.54**	**0.52**	**1762**
Weighted avg	**0.66**	**0.70**	**0.63**	**1762**

Model strongly favors classifying class 1 (high recall, good precision) but poorly identifies class 0 (very low recall, decent precision). Overall accuracy is moderate (0.70), skewed by the more frequent class 1.

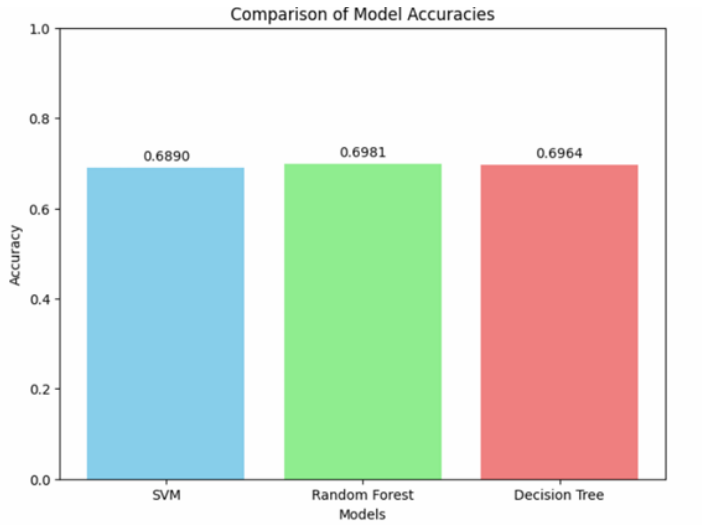




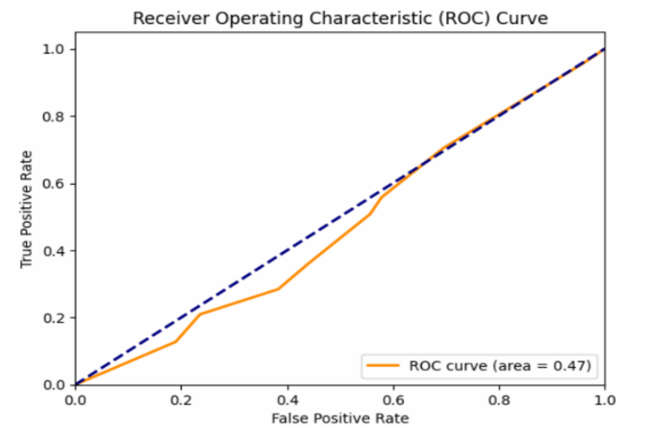
The confusion matrix shows that the SVM classifier predicts Movies (1) well but incorrectly classifies all TV Shows (0) as Movies. There are 1214 correct predictions for Movies but 548 misclassified TV Shows—suggesting a strong bias towards predicting content as Movies.



Confusion Matrix for Decision Tree:
76: Correctly classified TV shows.
472: TV shows misclassified as movies.
63: Movies misclassified as TV shows.
1151: Correctly classified movies.

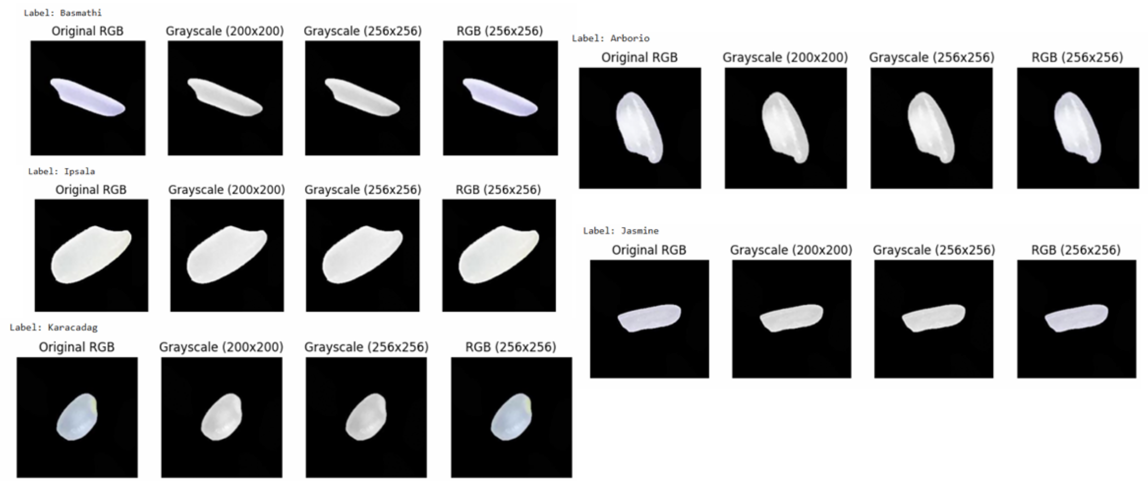


The bar chart compares the accuracy of three models: SVM (0.6890), Random Forest (0.6981), and Decision Tree (0.6964). Random Forest has the highest accuracy, followed by Decision Tree, with SVM being the least accurate.



The ROC curve shows a classification model's performance. The x-axis is the False Positive Rate, and the y-axis is the True Positive Rate. The orange curve represents the model's performance across different thresholds. The dashed blue line is a random classifier. The AUC (Area Under the Curve) is 0.47, which is below 0.5, indicating the model performs worse than random guessing.

Project – 2



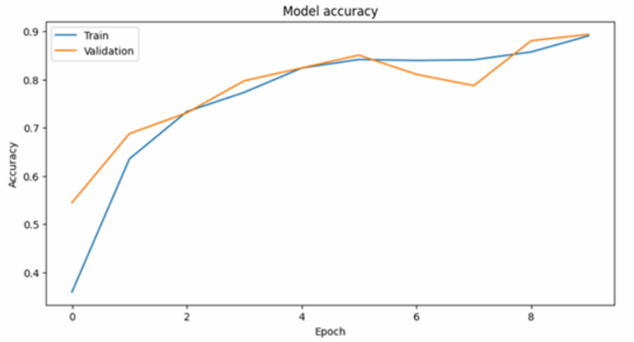
The image provides a visual comparison of five different rice varieties, showing each variety in its original RGB color and in grayscale at two different resolutions, as well as the original RGB resized. This standardized presentation is useful for various image processing and analysis tasks.

Model: "sequential"

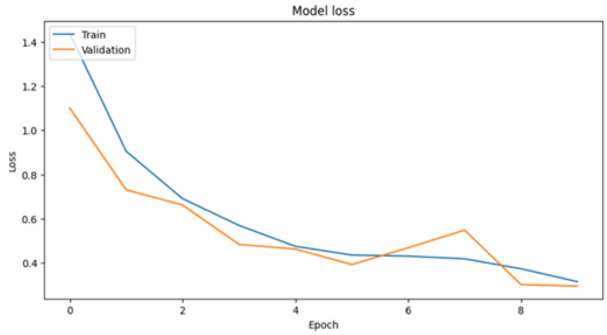
Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 254, 254, 32)	896
max_pooling2d (MaxPooling2D)	(None, 127, 127, 32)	0
conv2d_1 (Conv2D)	(None, 125, 125, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 62, 62, 64)	0
flatten (Flatten)	(None, 246016)	0
dropout (Dropout)	(None, 246016)	0
dense (Dense)	(None, 5)	1,230,085

Total params: 3,748,433 (14.30 MB)
Trainable params: 1,249,477 (4.77 MB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 2,498,956 (9.53 MB)

This describes a neural network model with convolutional (conv2d), pooling (max_pooling2d), flattening (flatten), dropout, and dense layers. It takes an input of an unspecified batch size and processes it through these layers. The final dense layer outputs 5 values, likely for a 5-class classification. The model has a total of 3,748,433 parameters, with 1,249,477 being trainable.



The graph shows training accuracy (blue) increasing, while validation accuracy (orange) plateaus and slightly drops after epoch 7, indicating potential overfitting.

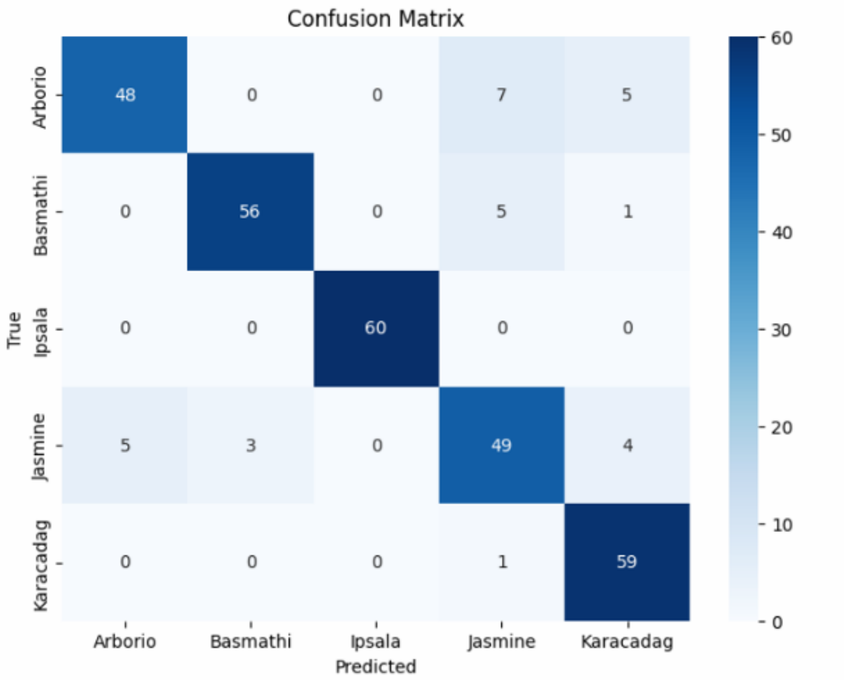


The graph shows training and validation loss over epochs. Training loss (blue) decreases. Validation loss (orange) decreases initially, then increases after epoch 6, indicating potential overfitting.

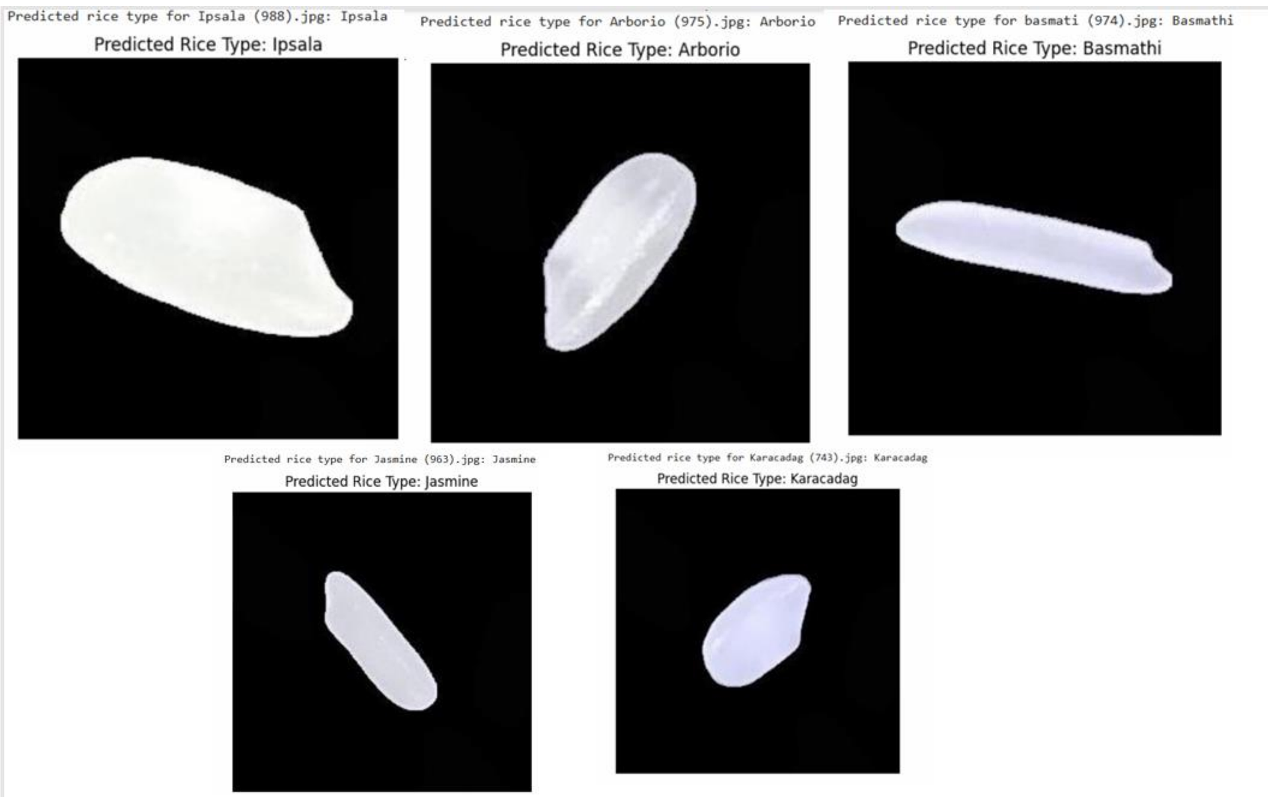
Classification Report:

	precision	recall	F1-score	support
Arborio	0.91	0.80	0.85	60
Basmathi	0.95	0.90	0.93	62
Ipsala	1.00	1.00	1.00	60
Jasmine	0.79	0.80	0.89	62
Karacadag	0.86	0.98	0.91	60
Accuracy			0.90	303
macro avg	0.90	0.90	0.90	303
weighted avg	0.90	0.90	0.90	303

The model performs very well in classifying different rice types, achieving high precision, recall, and F1-scores for each variety (Arborio, Basmathi, Ipsala, Jasmine, Karacadag). Ipsala is perfectly classified. The overall accuracy is 90%, and both macro and weighted averages are also around 90%, indicating strong and balanced performance across all rice types.



This matrix shows how often the model correctly and incorrectly classified different rice types. High numbers on the diagonal (48, 56, 60, 49, 59) mean many correct guesses. Off-diagonal numbers show misclassifications (e.g., 7 Arborio wrongly called Jasmine). The model generally did a good job distinguishing the rice types.



The image shows a rice classification model correctly identifying five different rice types (Ipsala, Arborio, Basmathi, Jasmine, Karacadag) with high confidence scores for each prediction. This visual output confirms the model's strong performance in distinguishing between these rice varieties.

```
T-statistic: 1.9999999999999933
P-value: 0.11611652351681642
Fail to reject the null hypothesis: There's no significant difference in average accuracy from 0.8.
```

```
Z-statistic: 1.9999999999999933
P-value: 0.0455002638963591
Reject the null hypothesis: The average accuracy is significantly different from 0.8.
```

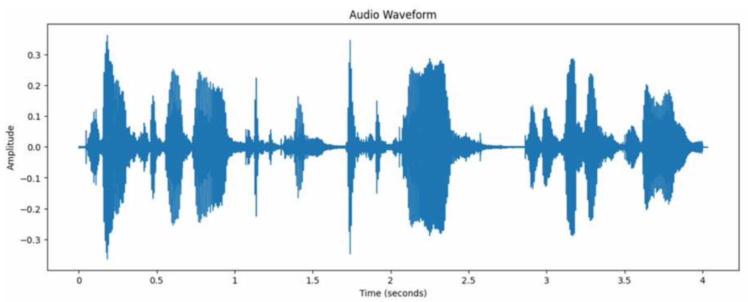
```
Chi-square statistic: 0.5998763141620285
P-value: 0.8964606878074289
Fail to reject the null hypothesis: No significant difference between observed and expected frequencies.
```

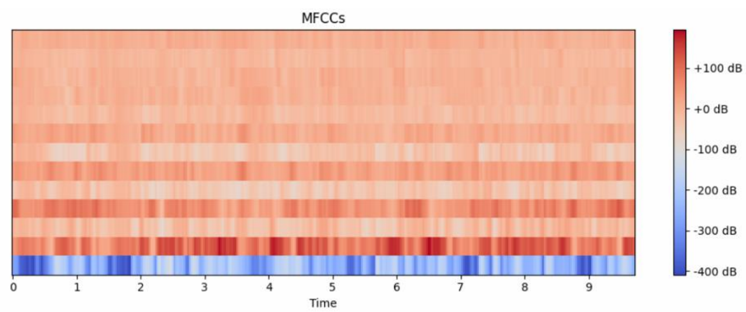
```
Z-statistic: 4.022870397928839
P-value: 5.749313666325864e-05
Reject the null hypothesis: There is a statistically significant difference between the groups.
```

The image shows results from four statistical tests.

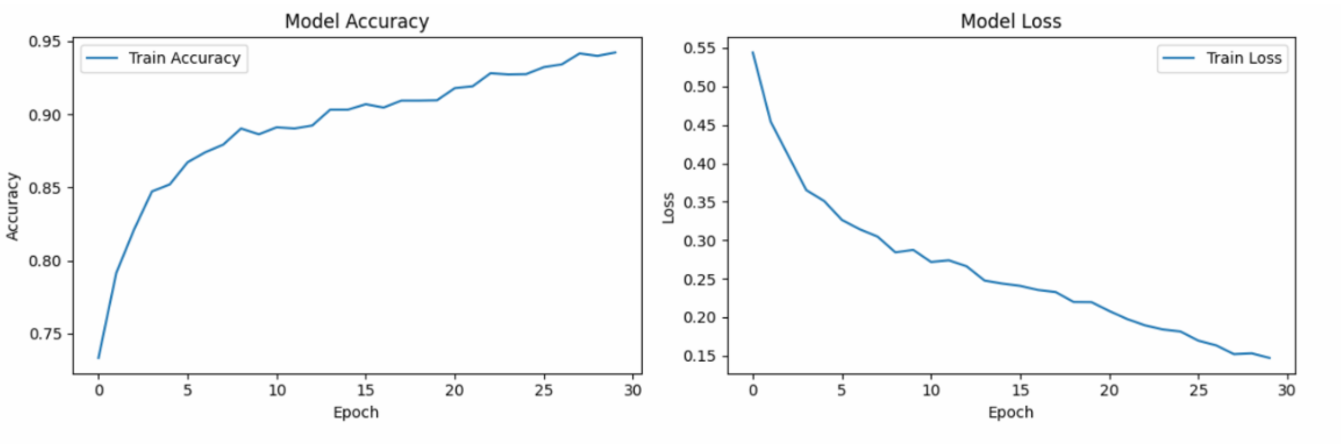
1. **T-test:** No significant difference in average accuracy from 0.8 (p=0.116).
2. **Z-test:** Average accuracy is significantly different from 0.8 (p=0.0455).
3. **Chi-square:** No significant difference between observed and expected frequencies (p=0.896).
4. **Z-test:** Statistically significant difference between the groups (p=5.75e-05).

Project-3





The top plot shows the raw audio signal's amplitude over time. The bottom plot shows the MFCCs, which are a time-frequency representation of the audio that is more perceptually relevant for human hearing and commonly used in audio analysis. Different patterns in the MFCC heatmap correspond to different sounds or phonetic elements present in the audio waveform.



The left graph shows training accuracy increasing over 30 epochs, reaching about 94%. The right graph shows training loss decreasing over the same period, reaching about 0.15. This indicates the model is learning well from the training data.

Model: "sequential_1"

Layer (type)	Output Shape	Param #
lstm_2 (LSTM)	(None, 40, 64)	19,968
lstm_3 (LSTM)	(None, 64)	33,024
dense_2 (Dense)	(None, 64)	4,160
dense_3 (Dense)	(None, 2)	130

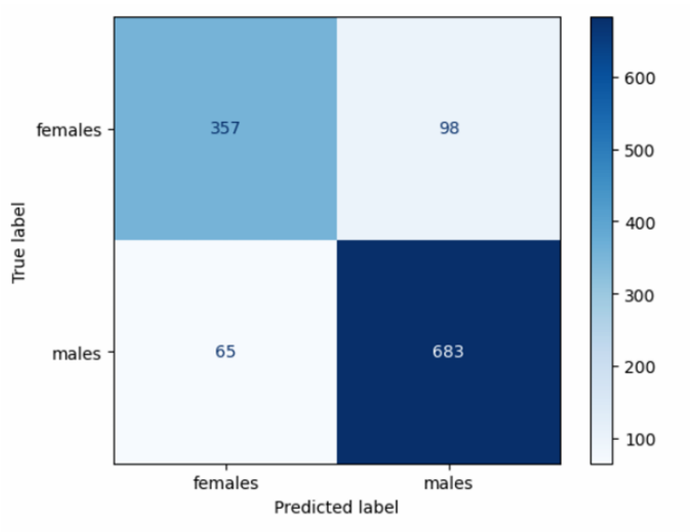
Total params: 171,848 (671.29 KB)
Trainable params: 57,282 (223.76 KB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 114,566 (447.53 KB)

This model is a sequential neural network designed for a task with sequential input (due to the LSTM layers) and likely a 2-class output (due to the final dense layer with output shape (None, 2)). It consists of two LSTM layers followed by two dense layers. The model has a total of 171,848 parameters, of which 57,282 are trainable. The optimizer being used has an additional 114,566 parameters it manages during training.

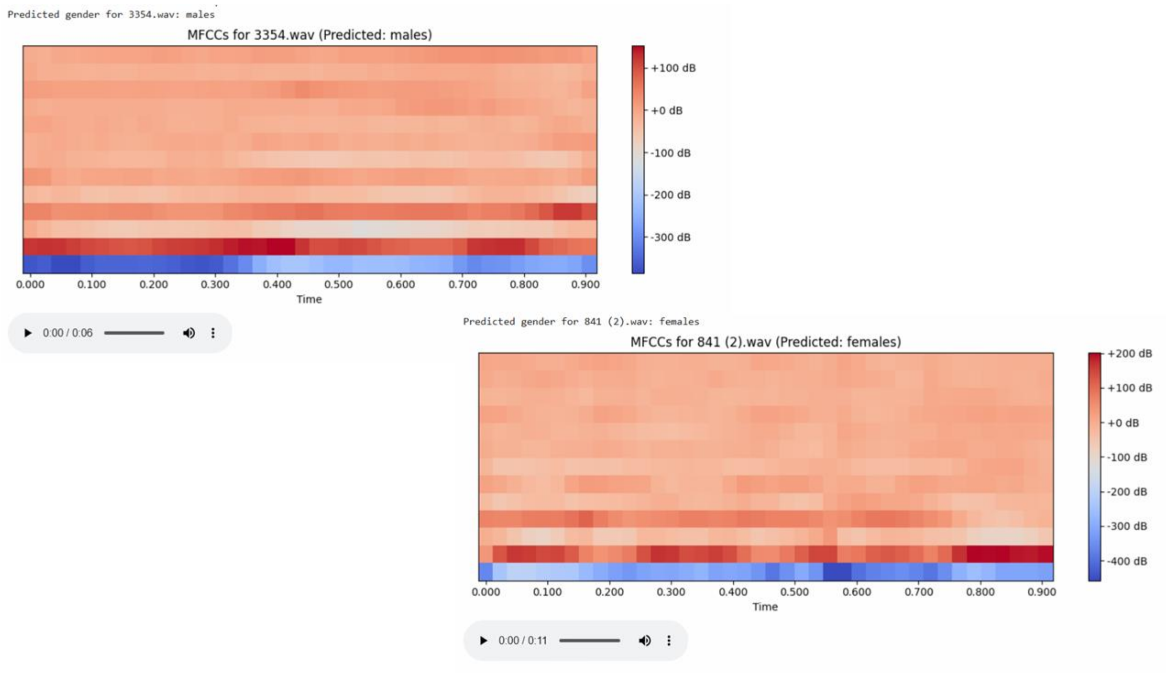
Classifiaction Report:

	precision	recall	F1-score	support
Females	**0.85**	**0.78**	**0.81**	**455**
Males	**0.87**	**0.91**	**0.89**	**748**
accuracy			**0.86**	**1203**
Macroavg	**0.86**	**0.85**	**0.85**	**1203**
Weighted avg	**0.86**	**0.86**	**0.86**	**1203**

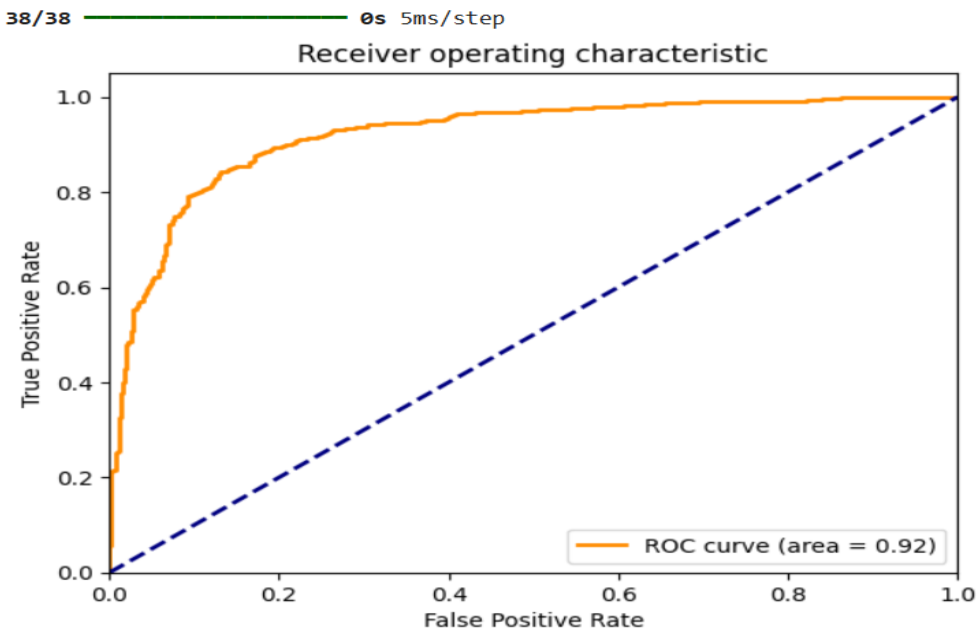
The gender classification model performs well, with around 85-87% precision and 78-91% recall for Females and Males respectively. The overall accuracy is 86%, and the F1-scores are also high (0.81 for Females, 0.89 for Males), indicating a good balance between precision and recall for both genders. The model shows slightly better performance for Males, likely due to the larger number of male samples.



The confusion matrix shows the model correctly predicted 357 females and 683 males. It misclassified 98 females as males and 65 males as females.



The image visualizes the MFCCs of two audio samples, one predicted as "males" and the other as "females" by a gender classification model. The MFCC spectrograms provide a visual representation of the frequency characteristics of the audio over a short time segment. The different patterns in the spectrograms are the features that the model likely learned to associate with male and female voices. The presence of audio players suggests the possibility of listening to the original audio signals. The titles clearly state the filenames and the model's gender predictions for each audio sample.



The ROC curve shows the performance of a binary classifier. The orange curve is significantly above the blue random chance line, and the area under the curve (AUC) is 0.92. This high AUC indicates that the classifier has a very good ability to discriminate between the positive and negative classes.