Model Performance Report

Training and Classification Times

Model: resnet | Training Time: 36.56s | Classification Time: 0.55s

Model: lstm | Training Time: 75.05s | Classification Time: 0.69s

Model: transformer | Training Time: 140.43s | Classification Time: 0.91s

AUC Scores

Model: resnet | AUC Score: 0.73

Model: lstm | AUC Score: 0.60

Model: transformer | AUC Score: 0.53

Model: resnet

Class: Class 0

precision: 0.62

recall: 0.52

f1-score: 0.57

support: 25.00

Class: Class 1

precision: 0.60

recall: 0.71

f1-score: 0.65

support: 38.00

Class: Class 2

precision: 0.54

recall: 0.40

f1-score: 0.46

support: 35.00

Class: Class 3

precision: 0.59

recall: 0.74

f1-score: 0.65

support: 23.00

accuracy: 0.59

Class: macro avg

precision: 0.59

recall: 0.59

f1-score: 0.58

support: 121.00

Class: weighted avg

precision: 0.58

recall: 0.59

f1-score: 0.58

support: 121.00

Model: lstm

Class: Class 0

precision: 0.50

recall: 0.16

f1-score: 0.24

support: 25.00

Class: Class 1

precision: 0.53

recall: 0.71

f1-score: 0.61

support: 38.00

Class: Class 2

precision: 0.36

recall: 0.60

f1-score: 0.45

support: 35.00

Class: Class 3

precision: 0.75

recall: 0.13

f1-score: 0.22

support: 23.00

accuracy: 0.45

Class: macro avg

precision: 0.54

recall: 0.40

f1-score: 0.38

support: 121.00

Class: weighted avg

precision: 0.52

recall: 0.45

f1-score: 0.41

support: 121.00

Model: transformer

Class: Class 0

precision: 1.00

recall: 0.04

f1-score: 0.08

support: 25.00

Class: Class 1

precision: 0.57

recall: 0.21

f1-score: 0.31

support: 38.00

Class: Class 2

precision: 0.30

recall: 0.91

f1-score: 0.45

support: 35.00

Class: Class 3

precision: 0.00

recall: 0.00

f1-score: 0.00

support: 23.00

accuracy: 0.34

Class: macro avg

precision: 0.47

recall: 0.29

f1-score: 0.21

support: 121.00

Class: weighted avg

precision: 0.47

recall: 0.34

f1-score: 0.24

support: 121.00

Figure 1

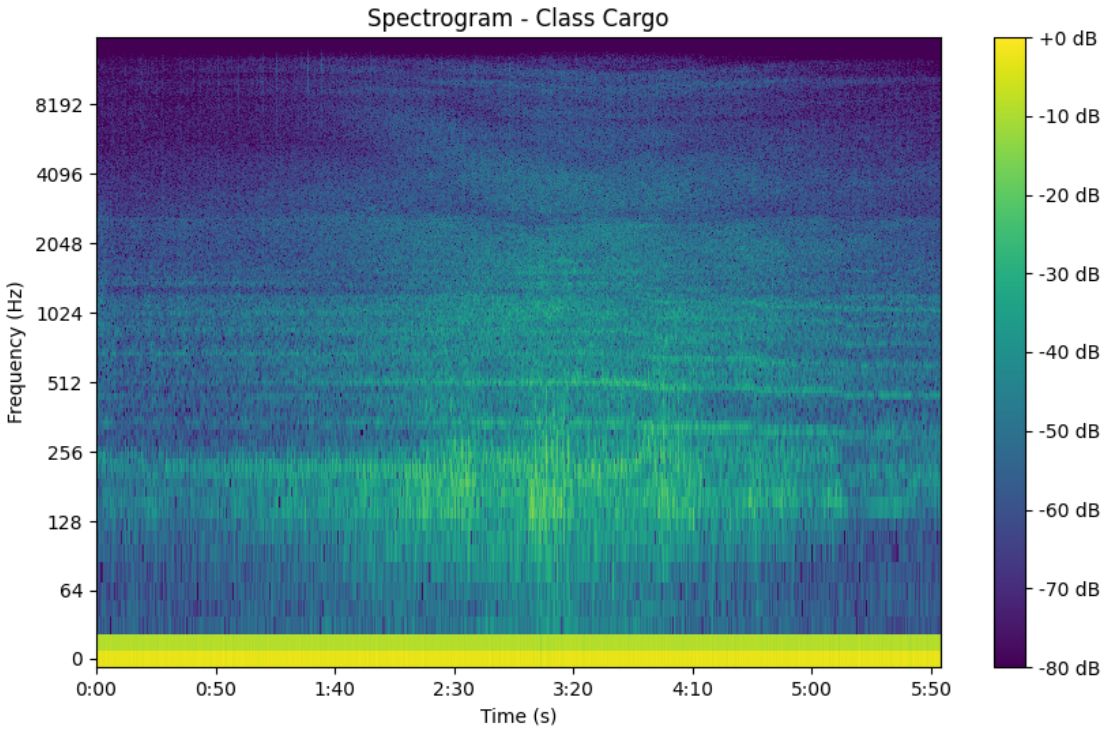

Spectrogram - Class Cargo
Frequency (Hz)
8192
4096
2048
1024
512
256
128
64
0
0:00
0:50
1:40
2:30
3:20
4:10
5:00
5:50
Time (s)
+0 dB
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

Figure 2

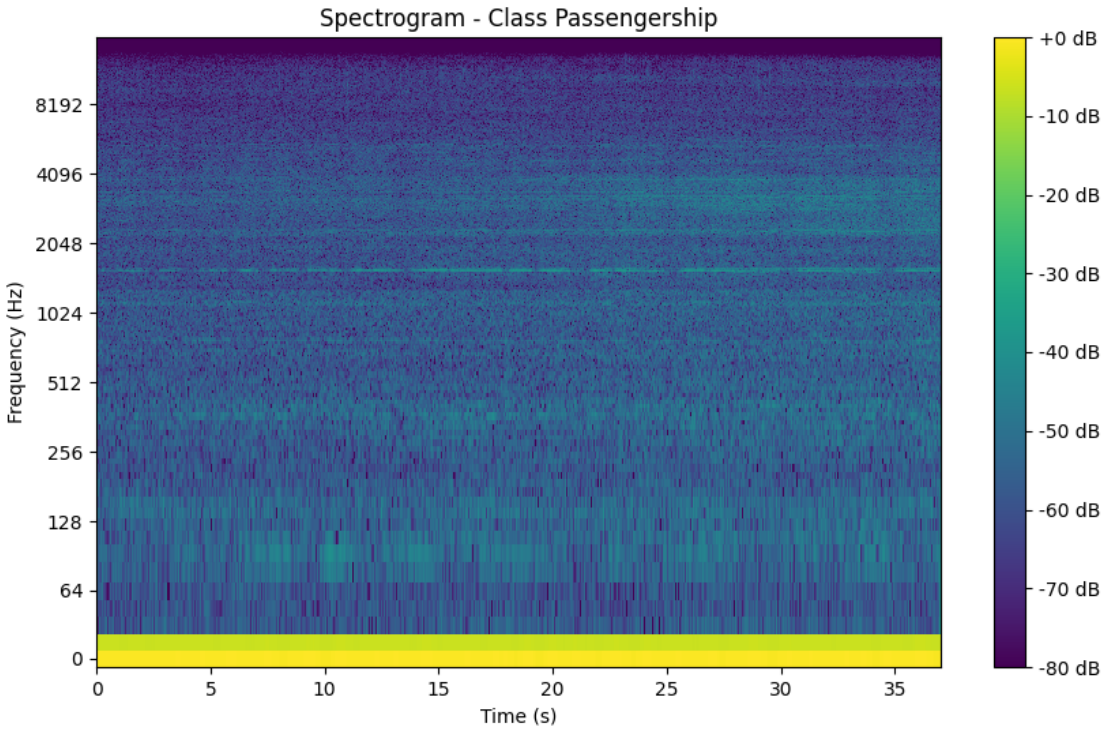

Spectrogram - Class Passengership
Frequency (Hz)
8192
4096
2048
1024
512
256
128
64
0
0
5
10
15
20
25
30
35
Time (s)
+0 dB
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

Figure 3

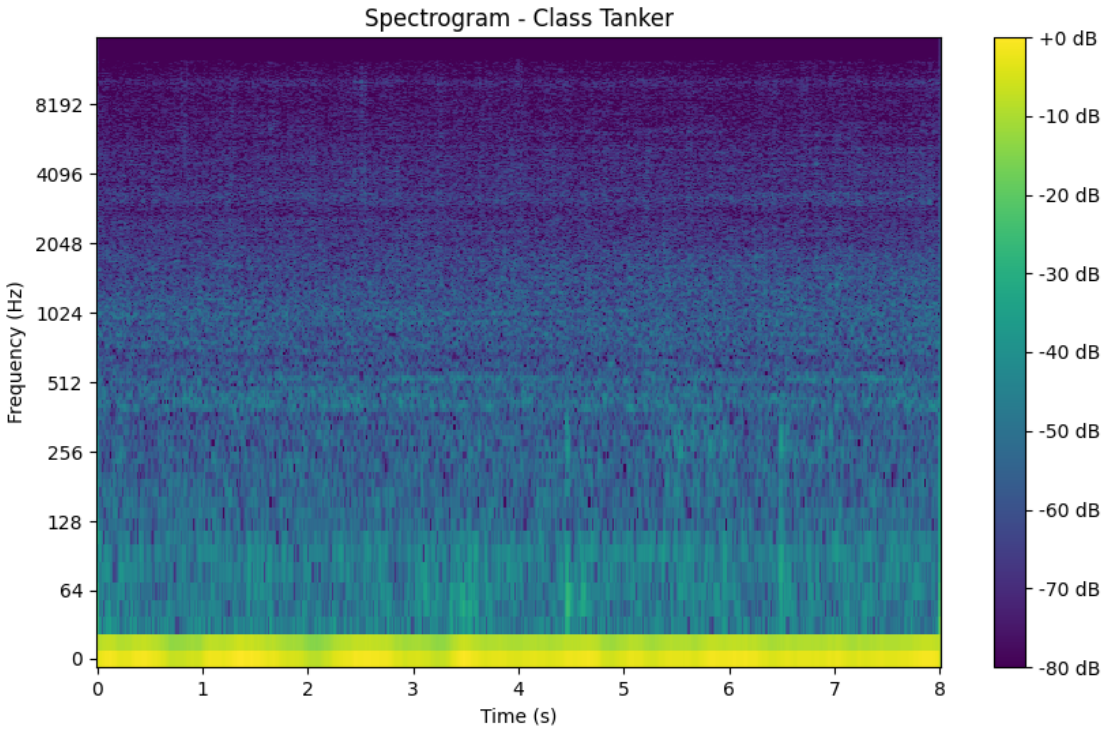

Spectrogram - Class Tanker
Frequency (Hz)
8192
4096
2048
1024
512
256
128
64
0
0
1
2
3
4
5
6
7
8
Time (s)
+0 dB
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

Figure 4

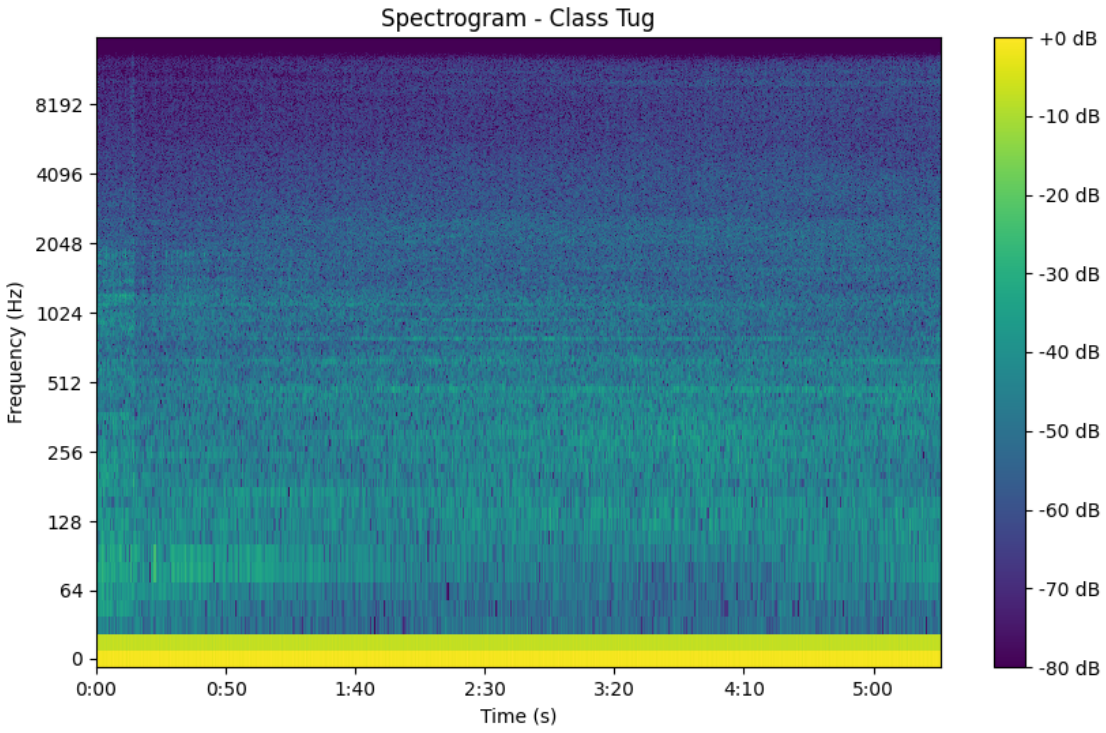

Spectrogram - Class Tug
Frequency (Hz)
8192
4096
2048
1024
512
256
128
64
0
0:00
0:50
1:40
2:30
3:20
4:10
5:00
Time (s)
+0 dB
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

Figure 5

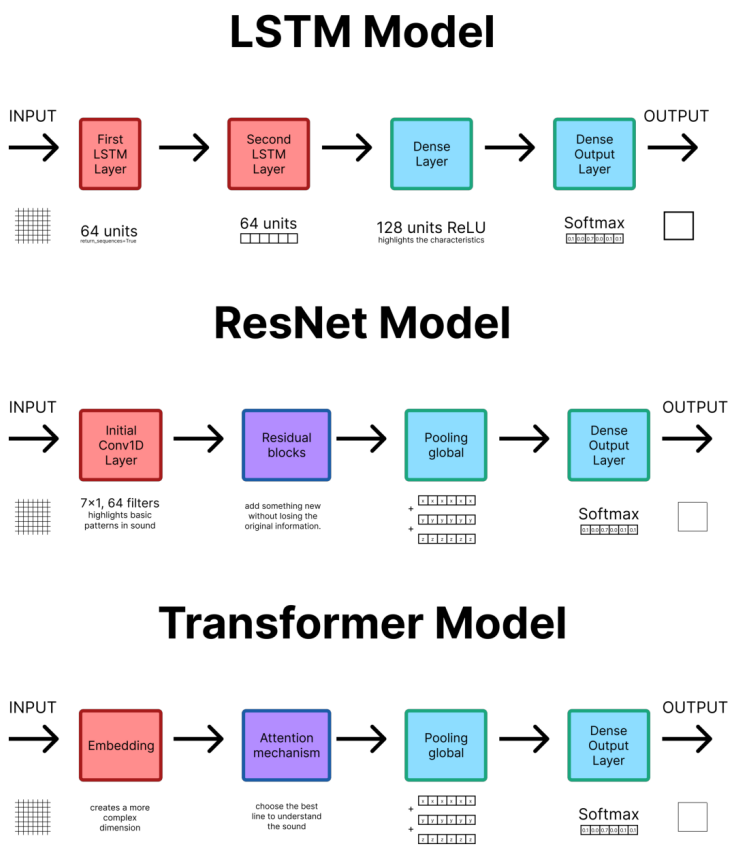

LSTM Model
INPUT
First LSTM Layer
Second LSTM Layer
Dense Layer
Dense Output Layer
OUTPUT
64 units
64 units
128 units ReLU
highlights the characteristics
Softmax
ResNet Model
INPUT
Initial Conv1D Layer
Residual blocks
Pooling global
Dense Output Layer
OUTPUT
7×1, 64 filters
highlights basic patterns in sound
add something new without losing the original information.
Softmax
Transformer Model
INPUT
Embedding
Attention mechanism
Pooling global
Dense Output Layer
OUTPUT
creates a more complex dimension
choose the best line to understand the sound
Softmax

Figure 6

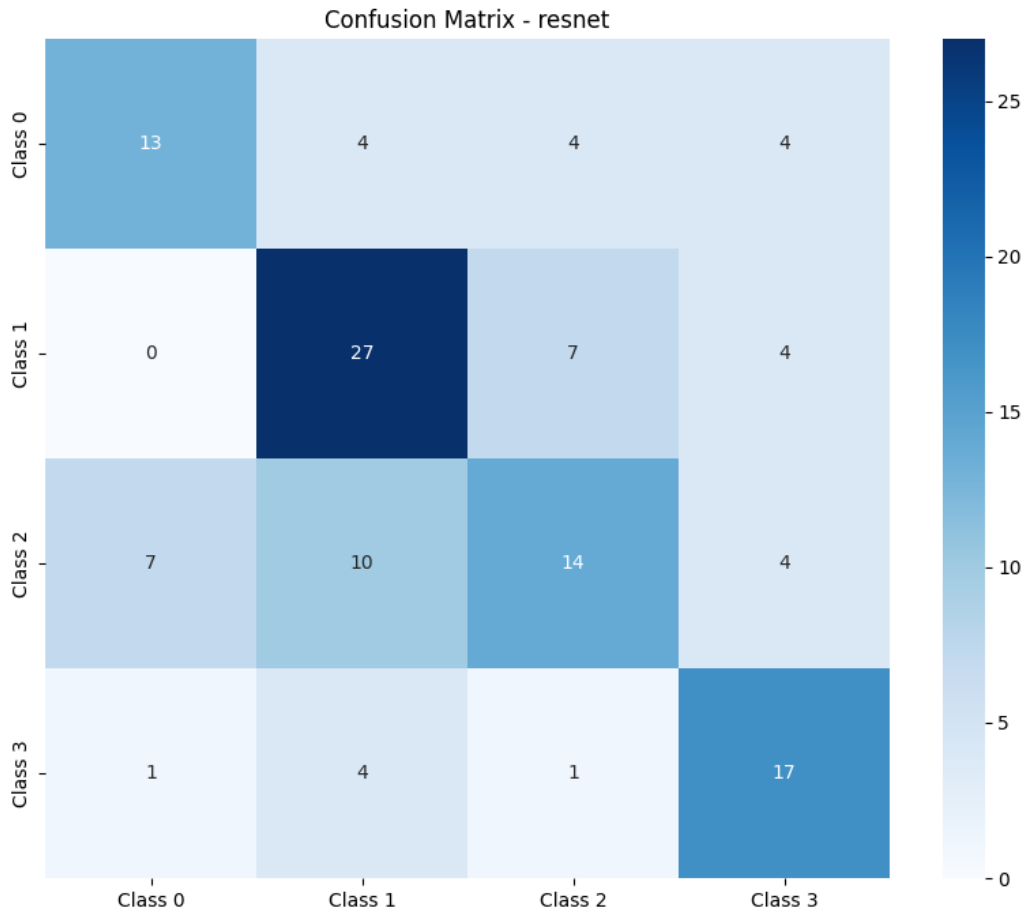

Confusion Matrix - resnet
Class 0
Class 1
Class 2
Class 3
13
4
4
4
0
27
7
4
7
10
14
4
1
4
1
17
Class 0
Class 1
Class 2
Class 3
25
20
15
10
5
0

Figure 7

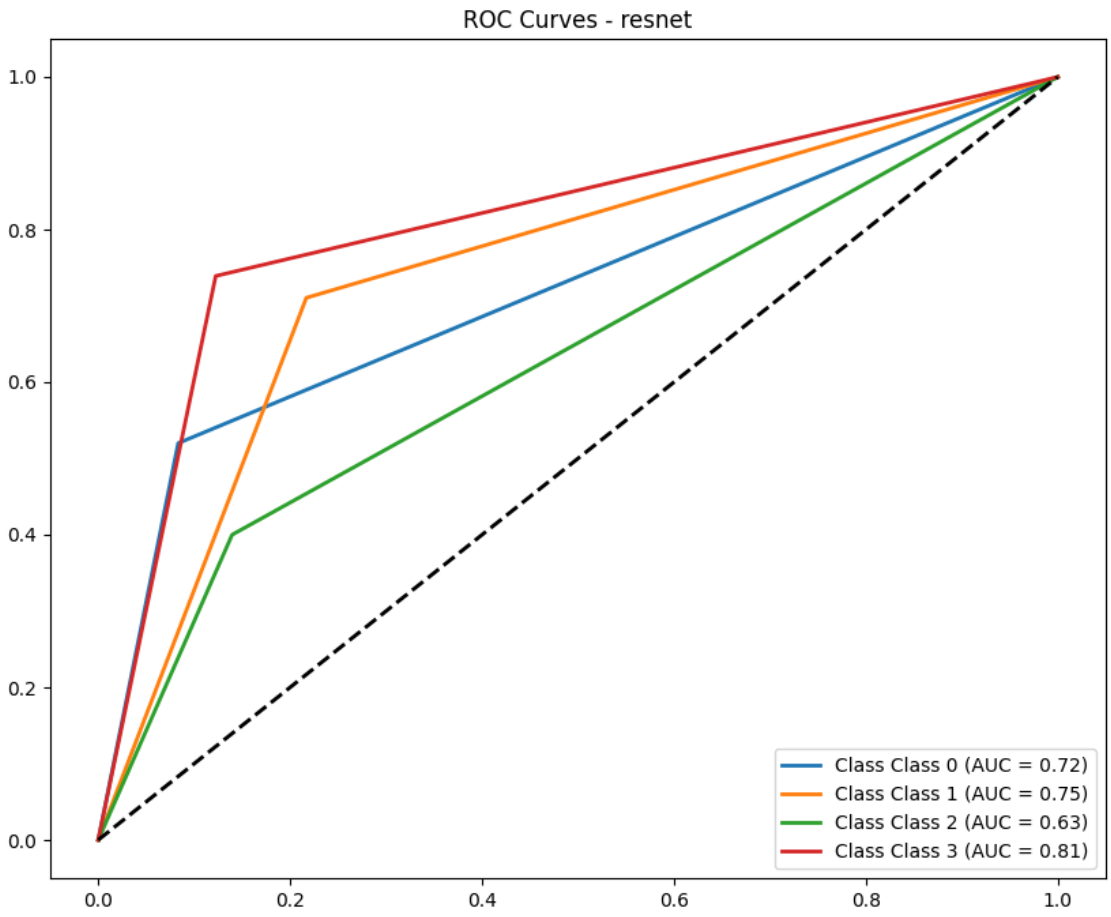

ROC Curves - resnet
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Class Class 0 (AUC = 0.72)
Class Class 1 (AUC = 0.75)
Class Class 2 (AUC = 0.63)
Class Class 3 (AUC = 0.81)

Figure 8

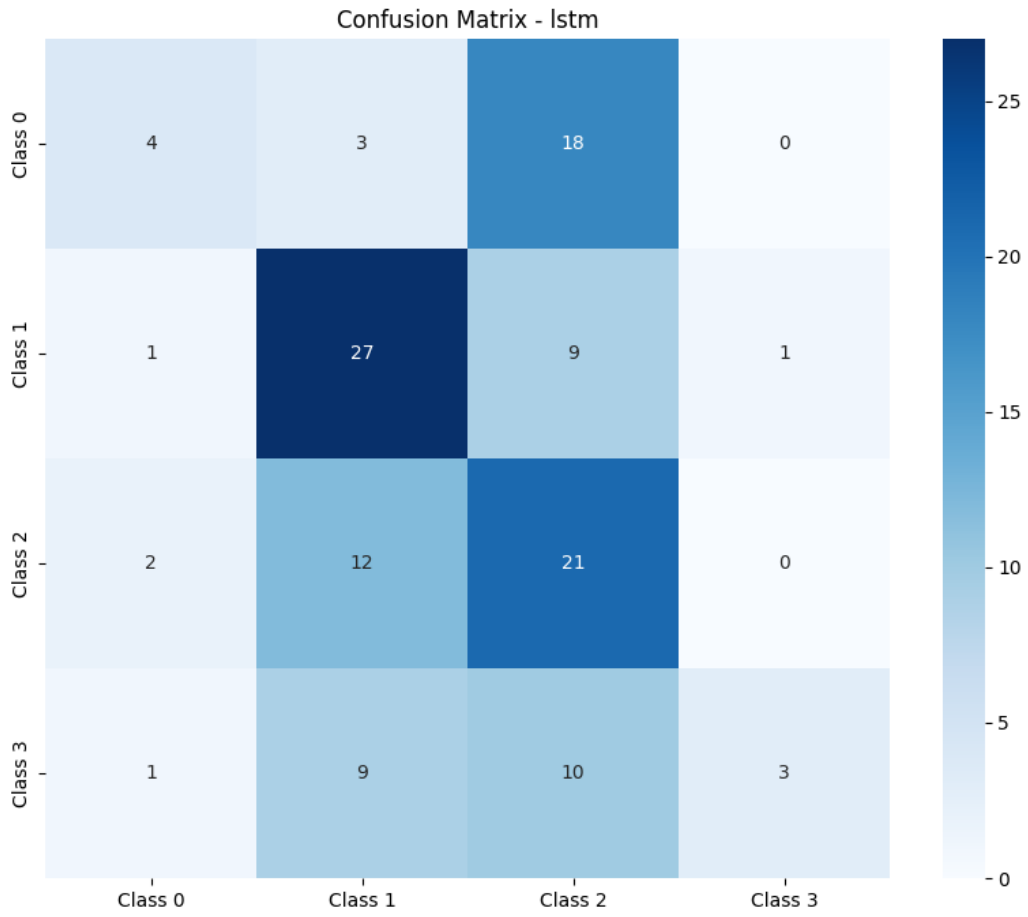

Confusion Matrix - lstm
Class 0
Class 1
Class 2
Class 3
4
3
18
0
1
27
9
1
2
12
21
0
1
9
10
3
Class 0
Class 1
Class 2
Class 3
25
20
15
10
5
0

Figure 9

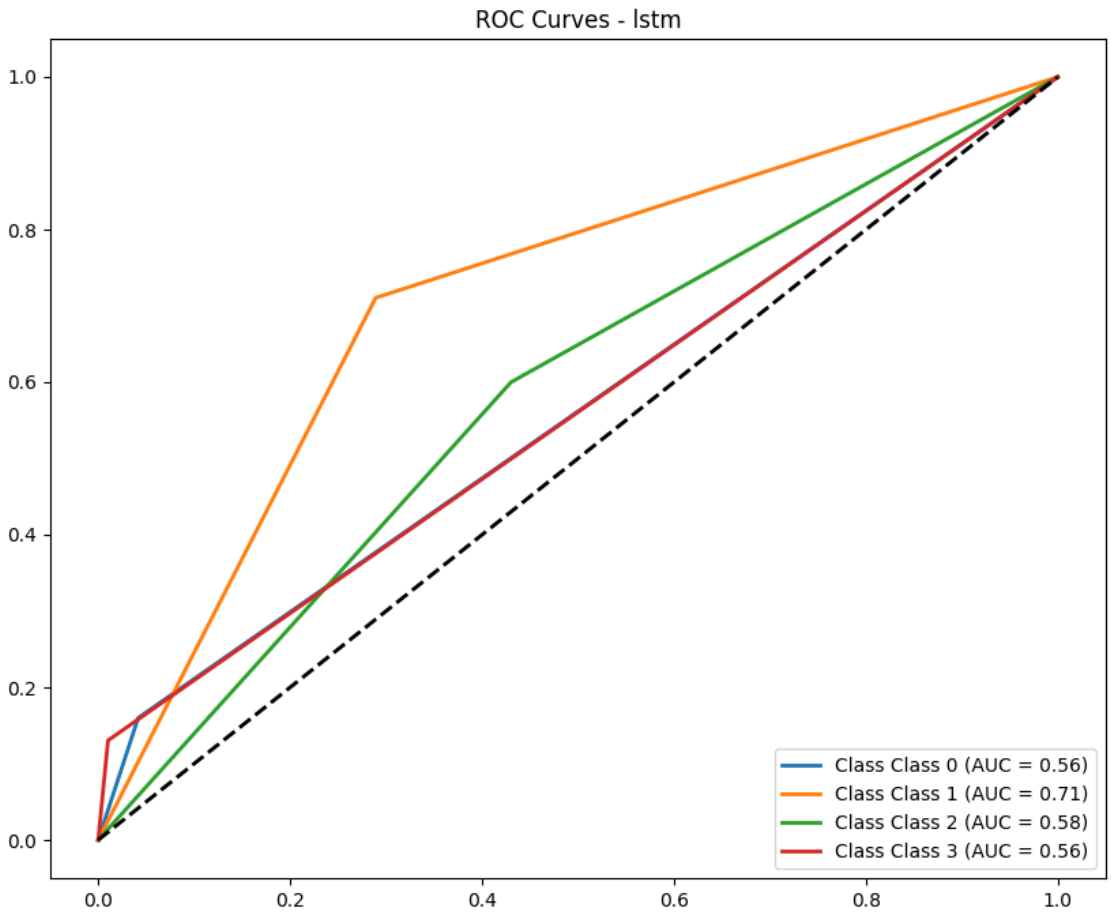

ROC Curves - lstm
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Class Class 0 (AUC = 0.56)
Class Class 1 (AUC = 0.71)
Class Class 2 (AUC = 0.58)
Class Class 3 (AUC = 0.56)

Figure 10

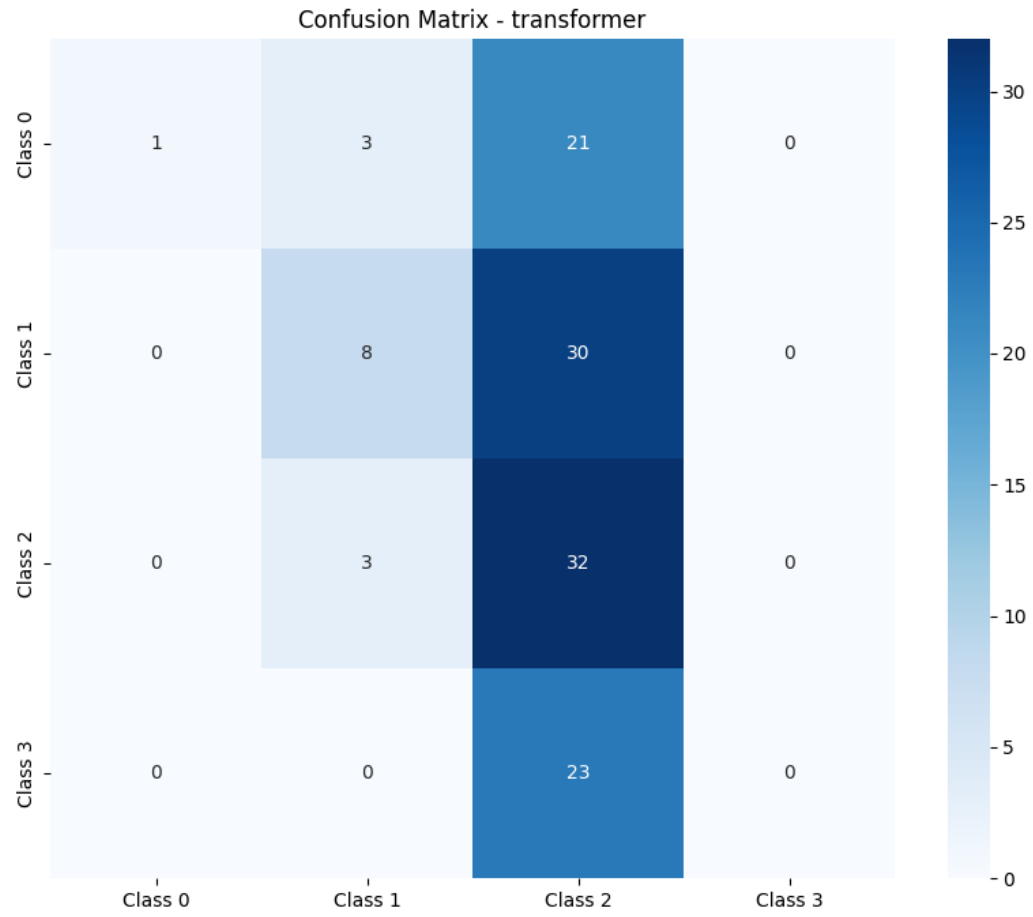

Confusion Matrix - transformer
Class 0
Class 1
Class 2
Class 3
1
3
21
0
0
8
30
0
0
3
32
0
0
0
23
0
Class 0
Class 1
Class 2
Class 3
30
25
20
15
10
5
0

Figure 11

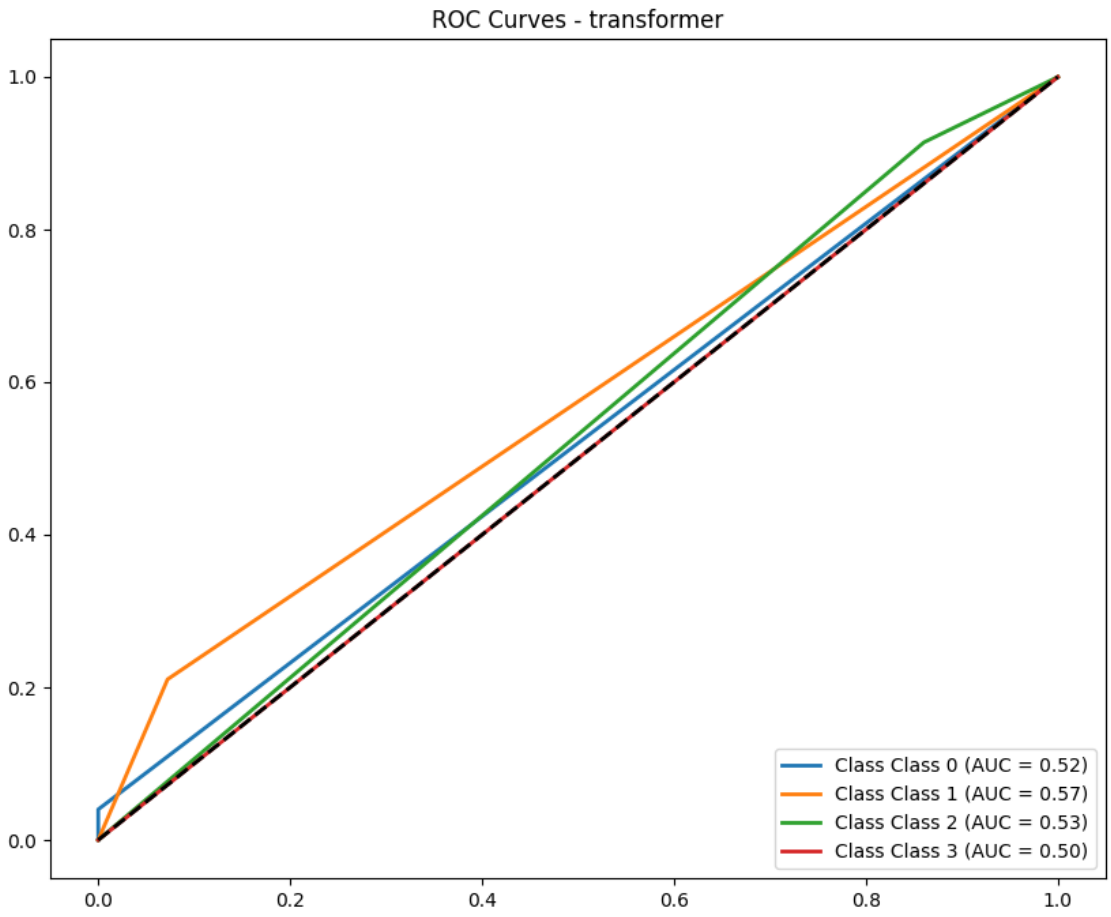

ROC Curves - transformer
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Class Class 0 (AUC = 0.52)
Class Class 1 (AUC = 0.57)
Class Class 2 (AUC = 0.53)
Class Class 3 (AUC = 0.50)

Figure 12

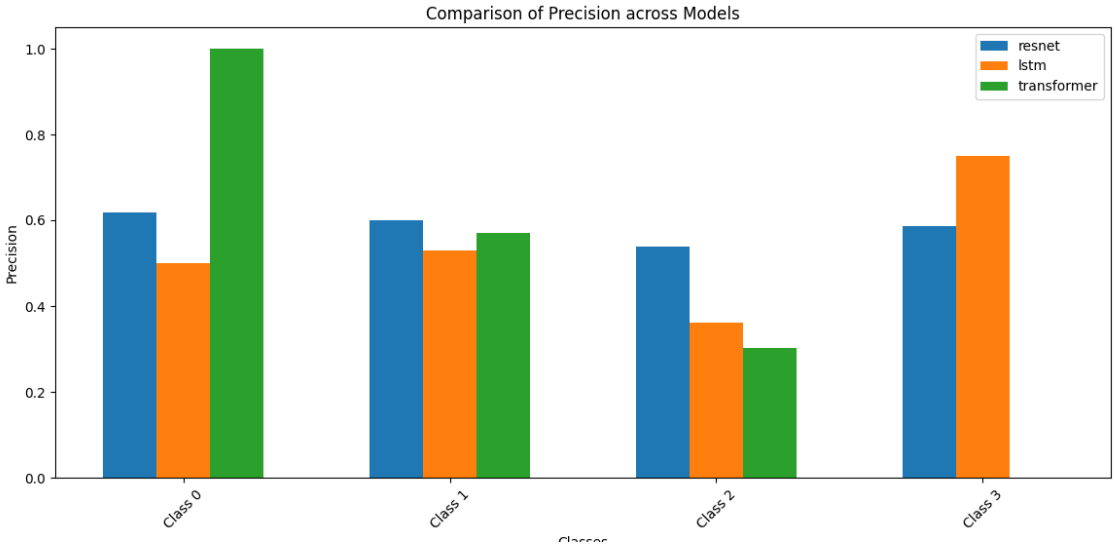

Comparison of Precision across Models
resnet
lstm
transformer
Precision
1.0
0.8
0.6
0.4
0.2
0.0
Class 0
Class 1
Class 2
Class 3

Figure 13

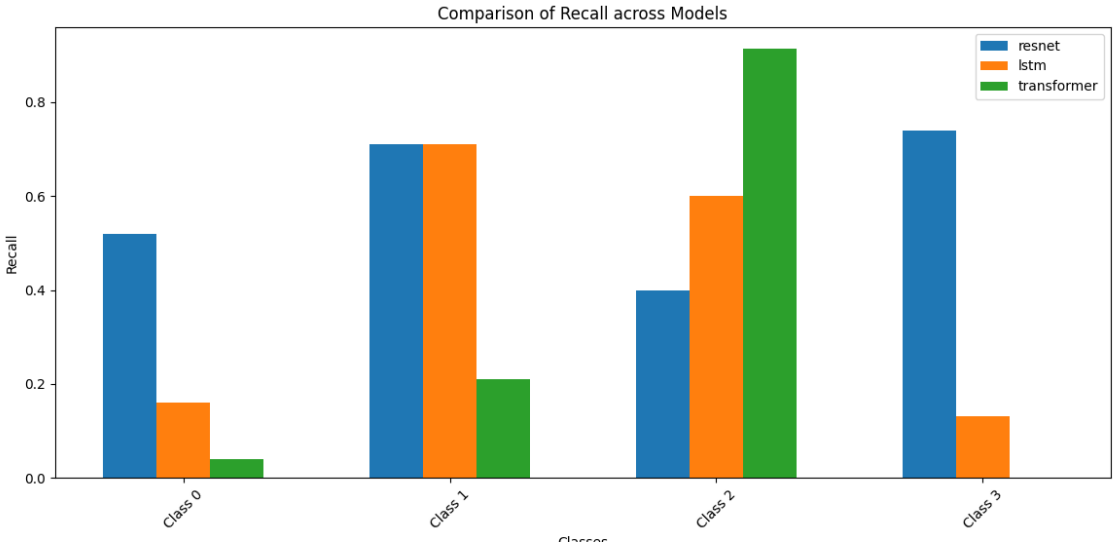

Comparison of Recall across Models
resnet
lstm
transformer
Recall
0.0
0.2
0.4
0.6
0.8
Class 0
Class 1
Class 2
Class 3

Figure 14

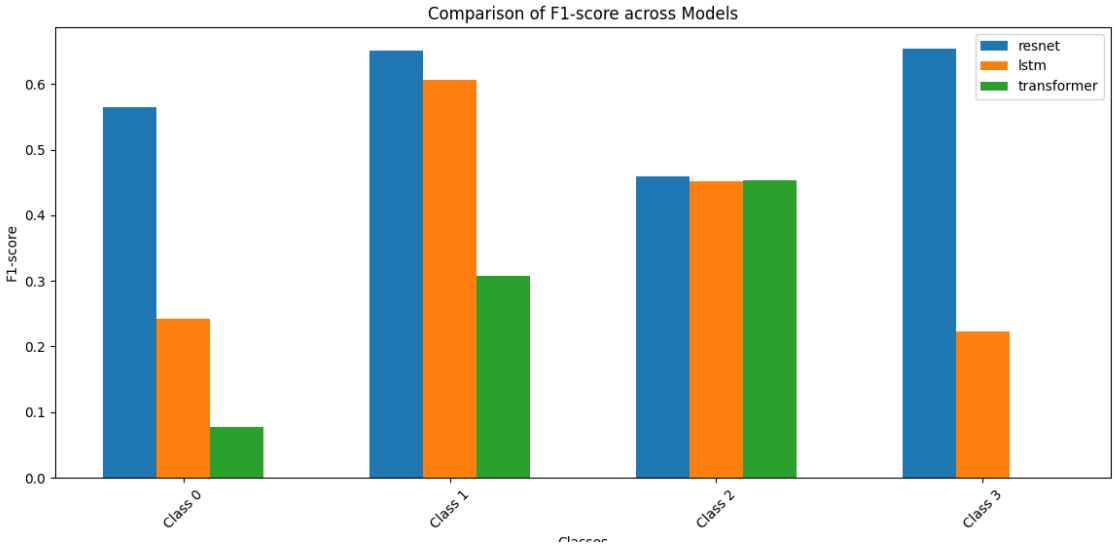

Comparison of F1-score across Models
resnet
lstm
transformer
F1-score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Class 0
Class 1
Class 2
Class 3

Figure 15

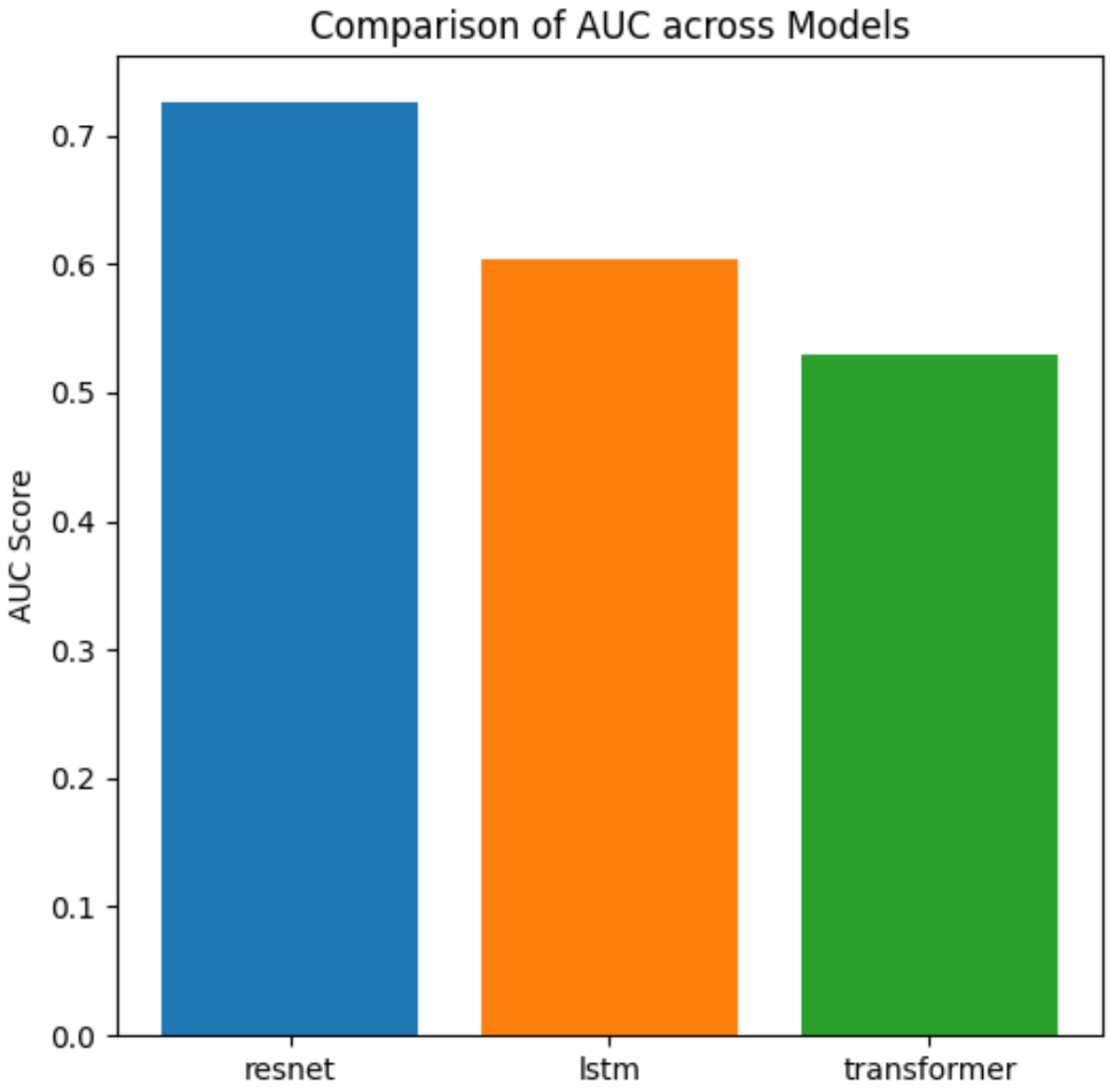

Comparison of AUC across Models
AUC Score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
resnet
lstm
transformer

Figure 16

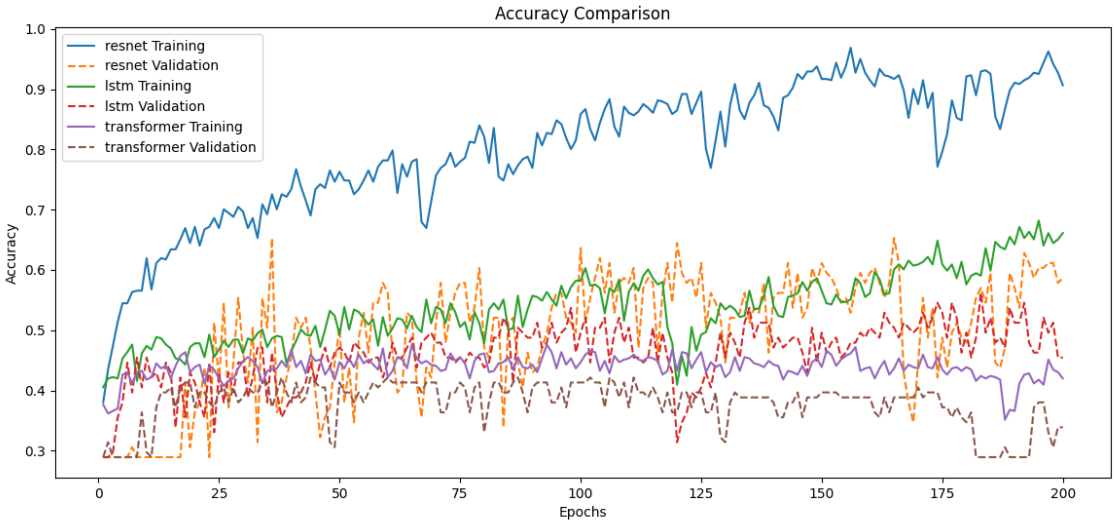

Accuracy Comparison
resnet Training
resnet Validation
lstm Training
lstm Validation
transformer Training
transformer Validation
Accuracy
Epochs
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
25
50
75
100
125
150
175
200

Figure 17

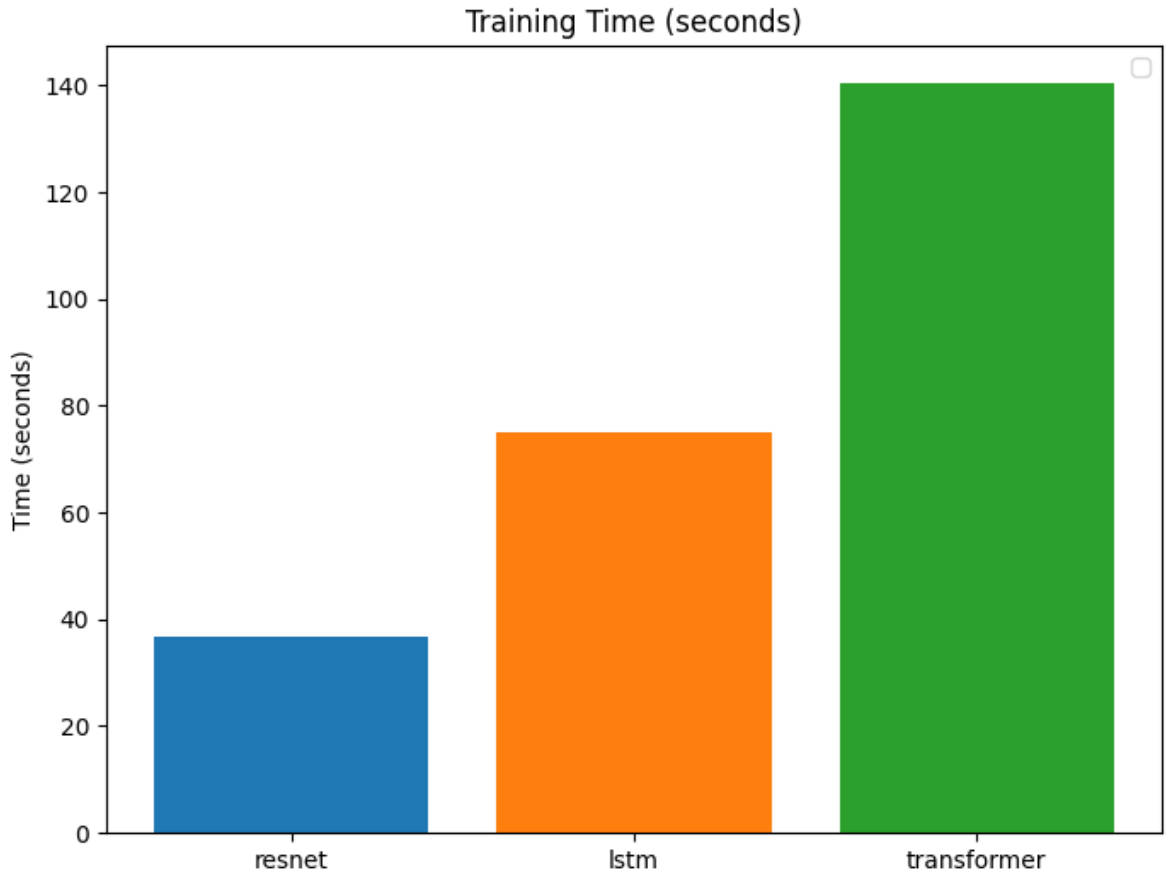
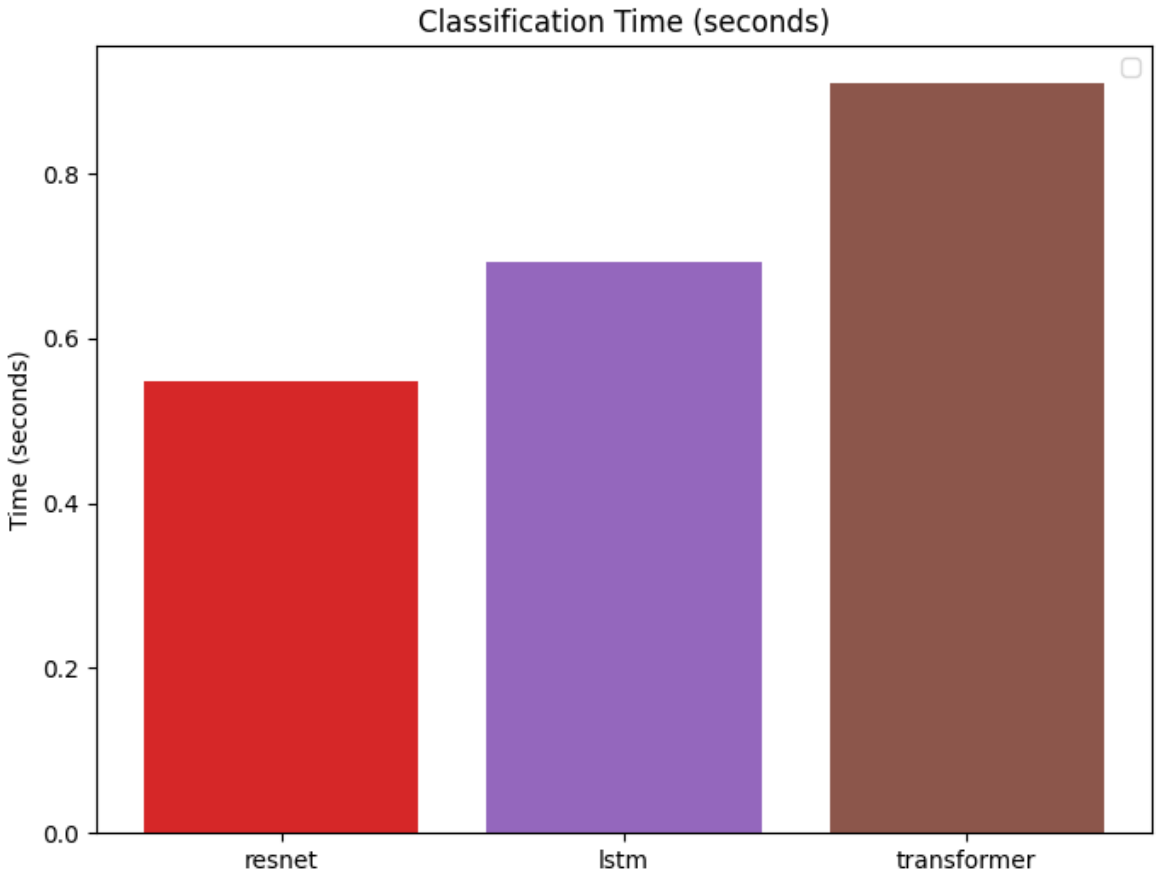

Training Time (seconds)
Time (seconds)
0
20
40
60
80
100
120
140
resnet
lstm
transformer

Figure 18


Classification Time (seconds)
Time (seconds)
0.0
0.2
0.4
0.6
0.8
resnet
lstm
transformer

Configuration Settings

data_path: C:\Users\gumar\OneDrive\Área de Trabalho\Pesquisa UBO\Revisão Literaria\UrbanSound8K\UrbanSound8K\audio

metadata_path: C:\Users\gumar\OneDrive\Área de Trabalho\Pesquisa UBO\Revisão Literaria\UrbanSound8K\UrbanSound8K\m

epochs: 200

batch_size: 32

test_size: 0.2

random_state: 42

num_classes: 4

learning_rate: 0.001

dense_units: 128

dropout_rate: 0.3

lstm_units: 64

lstm_dropout: 0.2

resnet_filters: 32

resnet_kernel_size: 3

resnet_blocks: 3

transformer_heads: 4

transformer_ff_dim: 128

transformer_layers: 2

transformer_epsilon: 1e-06

preprocessing_methods:

  MFCC: True

  GFCC: False

  STFT: False

  LOFAR: False

  DEMON: False

selected_models: ['resnet', 'lstm', 'transformer']