Model Performance Report

Training and Classification Times

Model: resnet | Training Time: 257.62s | Classification Time: 1.35s

Model: lstm | Training Time: 26.30s | Classification Time: 0.52s

Model: transformer | Training Time: 42.48s | Classification Time: 0.41s

AUC Scores

Model: resnet | AUC Score: 0.68

Model: lstm | AUC Score: 0.58

Model: transformer | AUC Score: 0.56

Model: resnet

Class: Class 0

precision: 0.85

recall: 0.44

f1-score: 0.58

support: 25.00

Class: Class 1

precision: 0.44

recall: 0.74

f1-score: 0.55

support: 38.00

Class: Class 2

precision: 0.56

recall: 0.54

f1-score: 0.55

support: 35.00

Class: Class 3

precision: 0.80

recall: 0.35

f1-score: 0.48

support: 23.00

accuracy: 0.55

Class: macro avg

precision: 0.66

recall: 0.52

f1-score: 0.54

support: 121.00

Class: weighted avg

precision: 0.63

recall: 0.55

f1-score: 0.54

support: 121.00

Model: lstm

Class: Class 0

precision: 0.50

recall: 0.04

f1-score: 0.07

support: 25.00

Class: Class 1

precision: 0.47

recall: 0.74

f1-score: 0.57

support: 38.00

Class: Class 2

precision: 0.40

recall: 0.66

f1-score: 0.49

support: 35.00

Class: Class 3

precision: 1.00

recall: 0.04

f1-score: 0.08

support: 23.00

accuracy: 0.44

Class: macro avg

precision: 0.59

recall: 0.37

f1-score: 0.31

support: 121.00

Class: weighted avg

precision: 0.55

recall: 0.44

f1-score: 0.35

support: 121.00

Model: transformer

Class: Class 0

precision: 0.00

recall: 0.00

f1-score: 0.00

support: 25.00

Class: Class 1

precision: 0.49

recall: 0.55

f1-score: 0.52

support: 38.00

Class: Class 2

precision: 0.36

recall: 0.80

f1-score: 0.50

support: 35.00

Class: Class 3

precision: 0.00

recall: 0.00

f1-score: 0.00

support: 23.00

accuracy: 0.40

Class: macro avg

precision: 0.21

recall: 0.34

f1-score: 0.25

support: 121.00

Class: weighted avg

precision: 0.26

recall: 0.40

f1-score: 0.31

support: 121.00

Figure 1

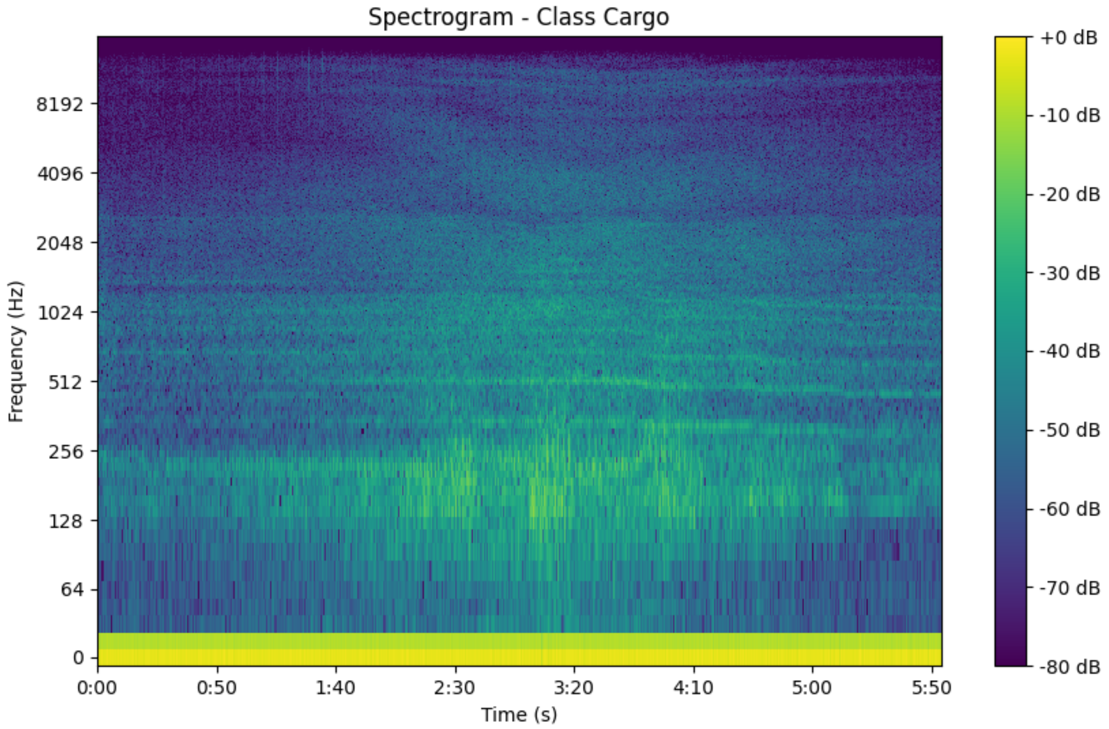

Spectrogram - Class Cargo
Frequency (Hz)
8192
4096
2048
1024
512
256
128
64
0
0:00
0:50
1:40
2:30
3:20
4:10
5:00
5:50
Time (s)
+0 dB
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

Figure 2

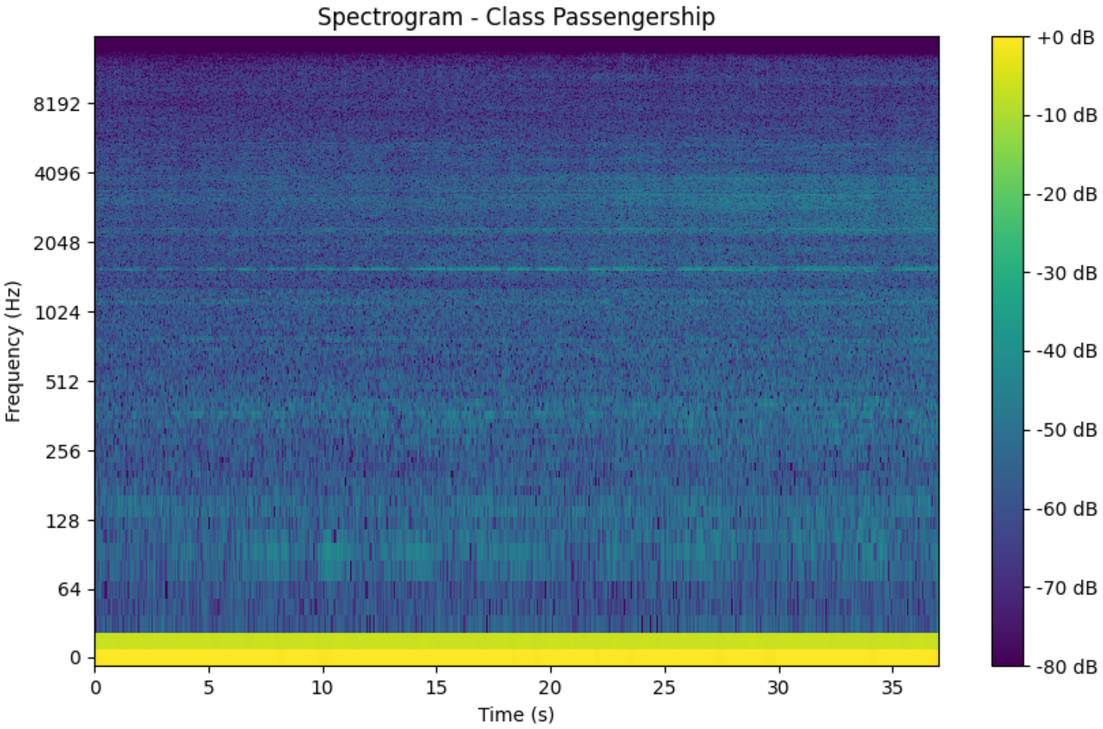

Spectrogram - Class Passengership
Frequency (Hz)
8192
4096
2048
1024
512
256
128
64
0
0
5
10
15
20
25
30
35
Time (s)
+0 dB
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

Figure 3

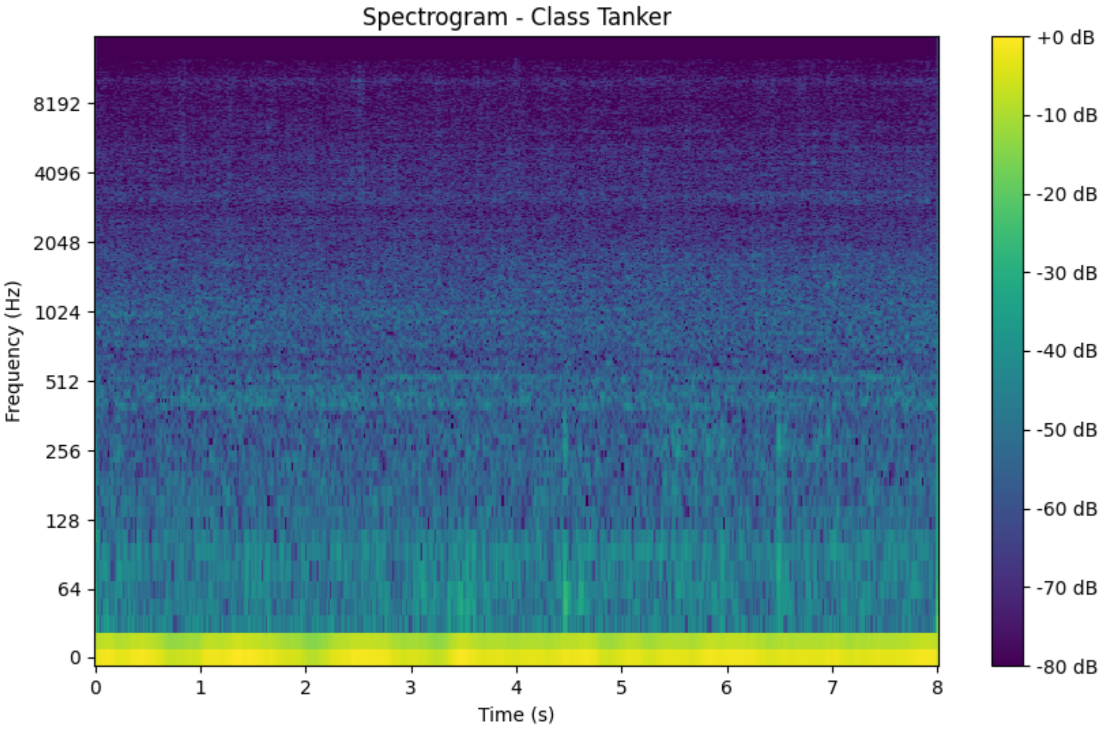

Spectrogram - Class Tanker
Frequency (Hz)
8192
4096
2048
1024
512
256
128
64
0
0
1
2
3
4
5
6
7
8
Time (s)
+0 dB
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

Figure 4

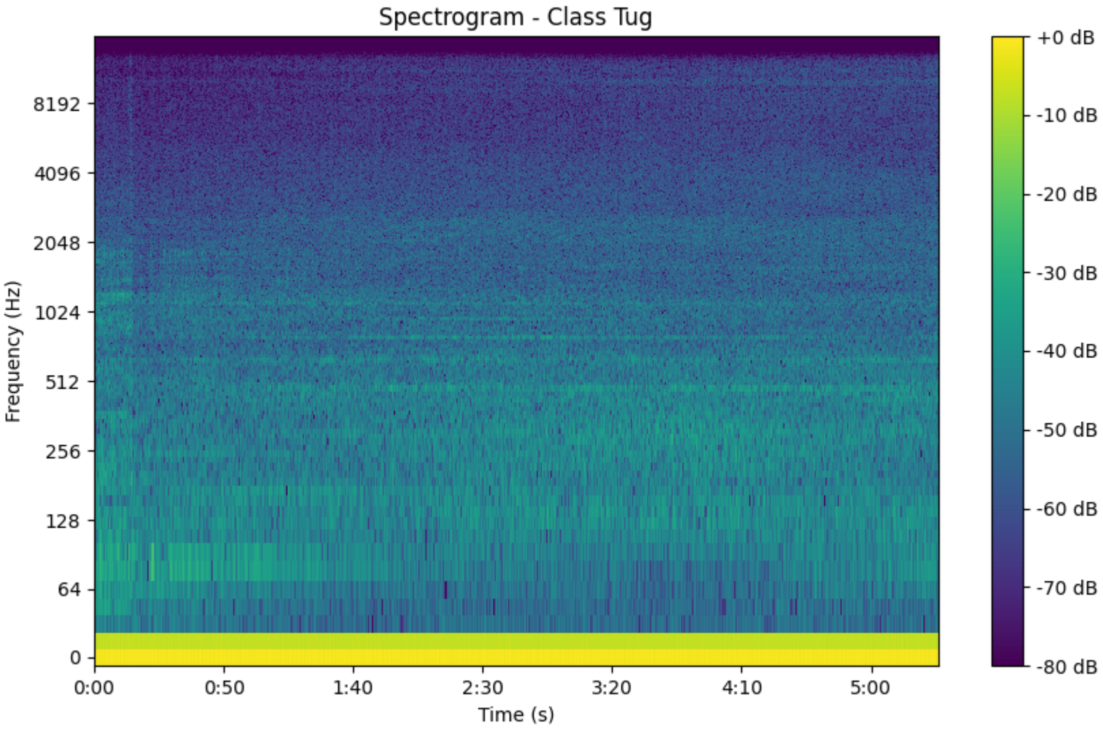

Spectrogram - Class Tug
Frequency (Hz)
8192
4096
2048
1024
512
256
128
64
0
0:00
0:50
1:40
2:30
3:20
4:10
5:00
Time (s)
+0 dB
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

Figure 5

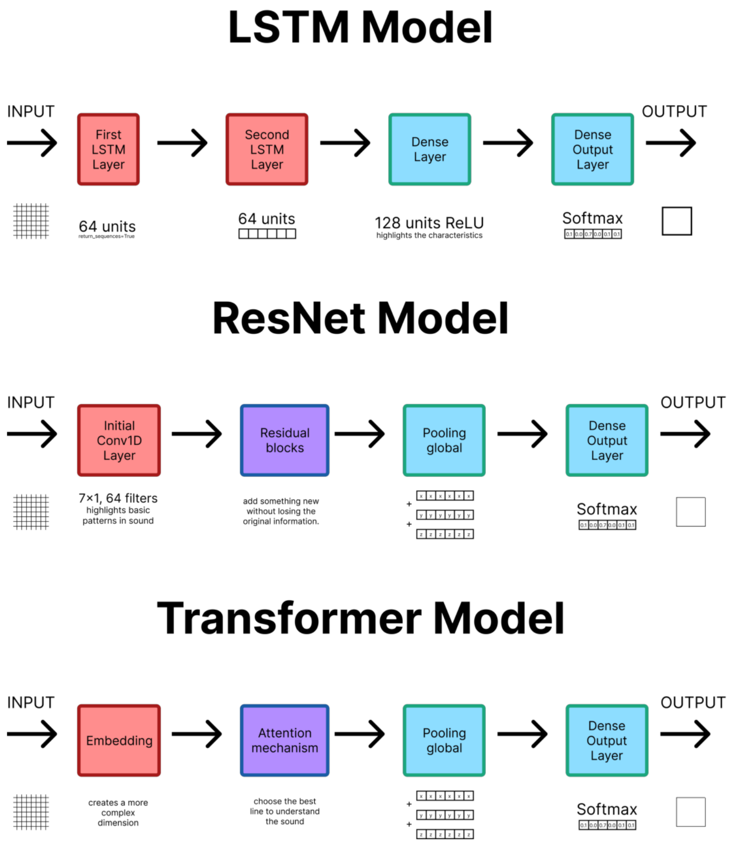
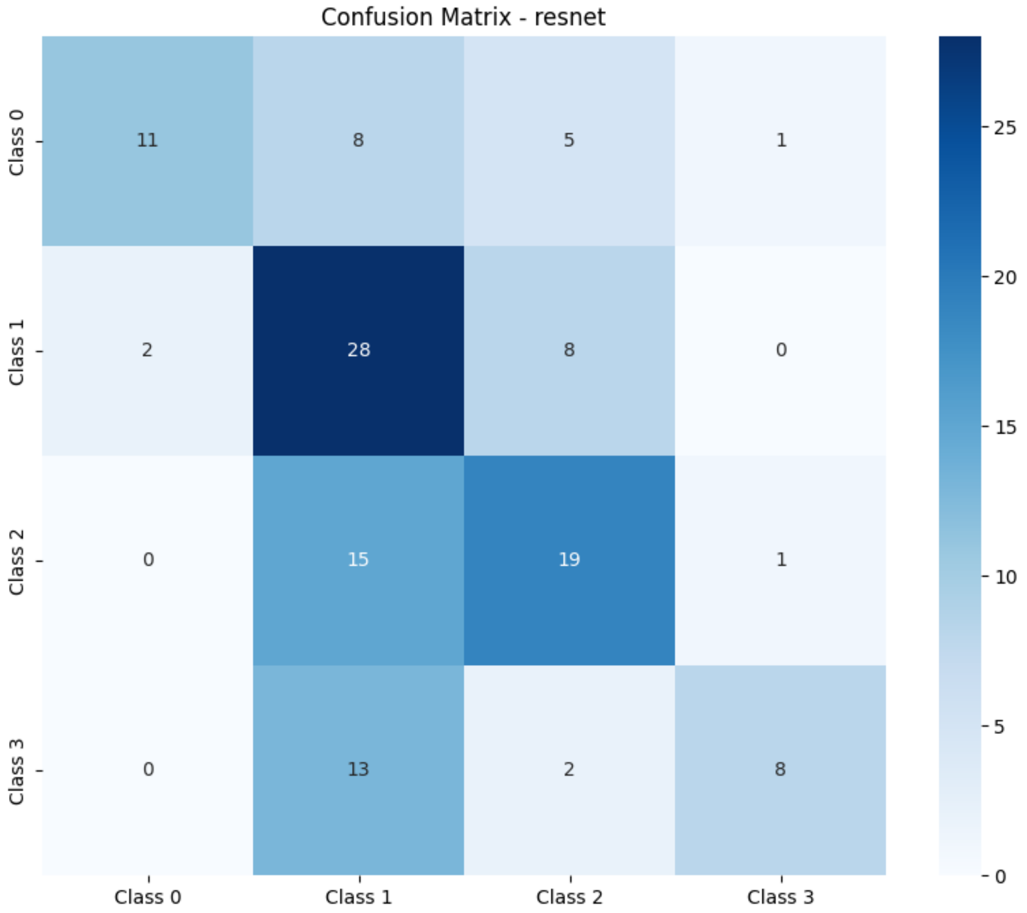

LSTM Model
INPUT
First LSTM Layer
Second LSTM Layer
Dense Layer
Dense Output Layer
OUTPUT
64 units
return_sequences=True
64 units
128 units ReLU
highlights the characteristics
Softmax
0.1 0.0 0.7 0.0 0.1 0.1
ResNet Model
INPUT
Initial Conv1D Layer
Residual blocks
Pooling global
Dense Output Layer
OUTPUT
7×1, 64 filters
highlights basic patterns in sound
add something new without losing the original information.
x x x x x x
+
y y y y y y
+
z z z z z z
Softmax
0.1 0.0 0.7 0.0 0.1 0.1
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
x x x x x x
+
y y y y y y
+
z z z z z z
Softmax
0.1 0.0 0.7 0.0 0.1 0.1

Figure 6



	Class 0	Class 1	Class 2	Class 3
Class 0	11	8	5	1
Class 1	2	28	8	0
Class 2	0	15	19	1
Class 3	0	13	2	8

Figure 7

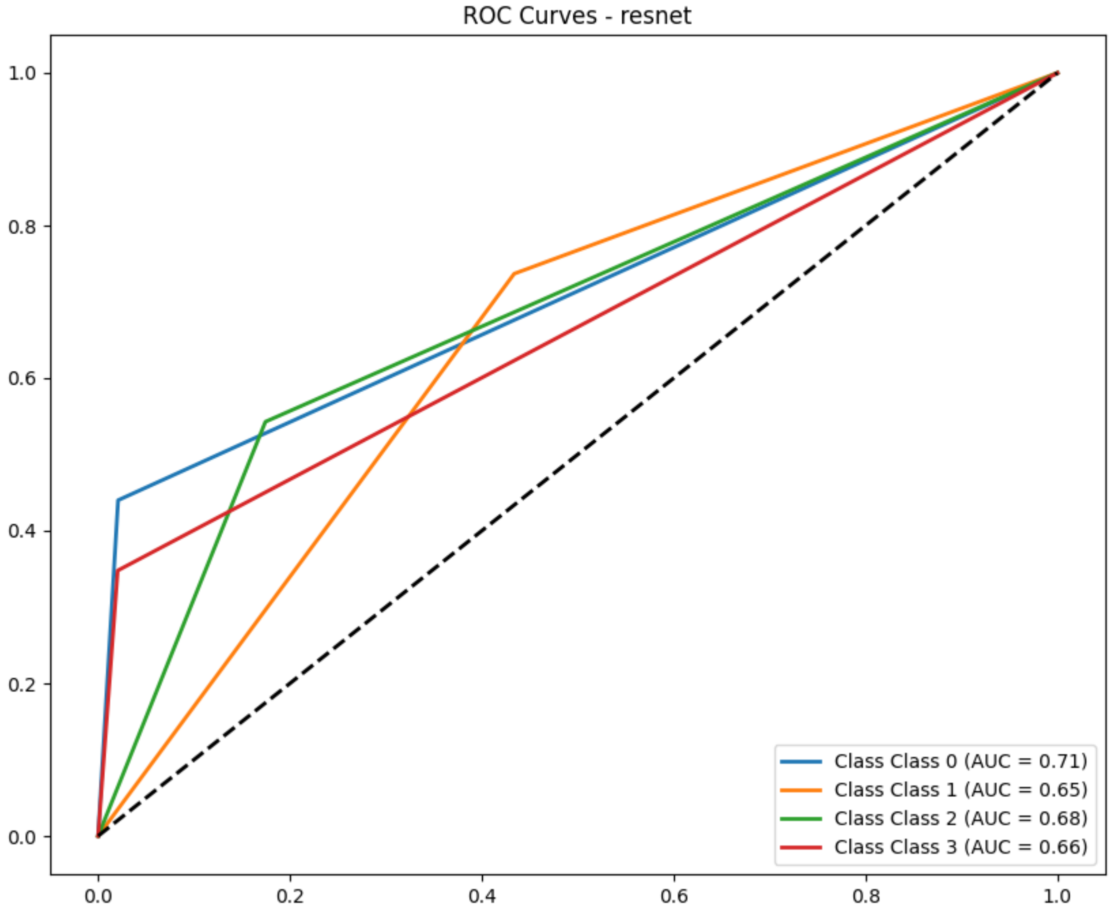

ROC Curves - resnet
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Class Class 0 (AUC = 0.71)
Class Class 1 (AUC = 0.65)
Class Class 2 (AUC = 0.68)
Class Class 3 (AUC = 0.66)

Figure 8

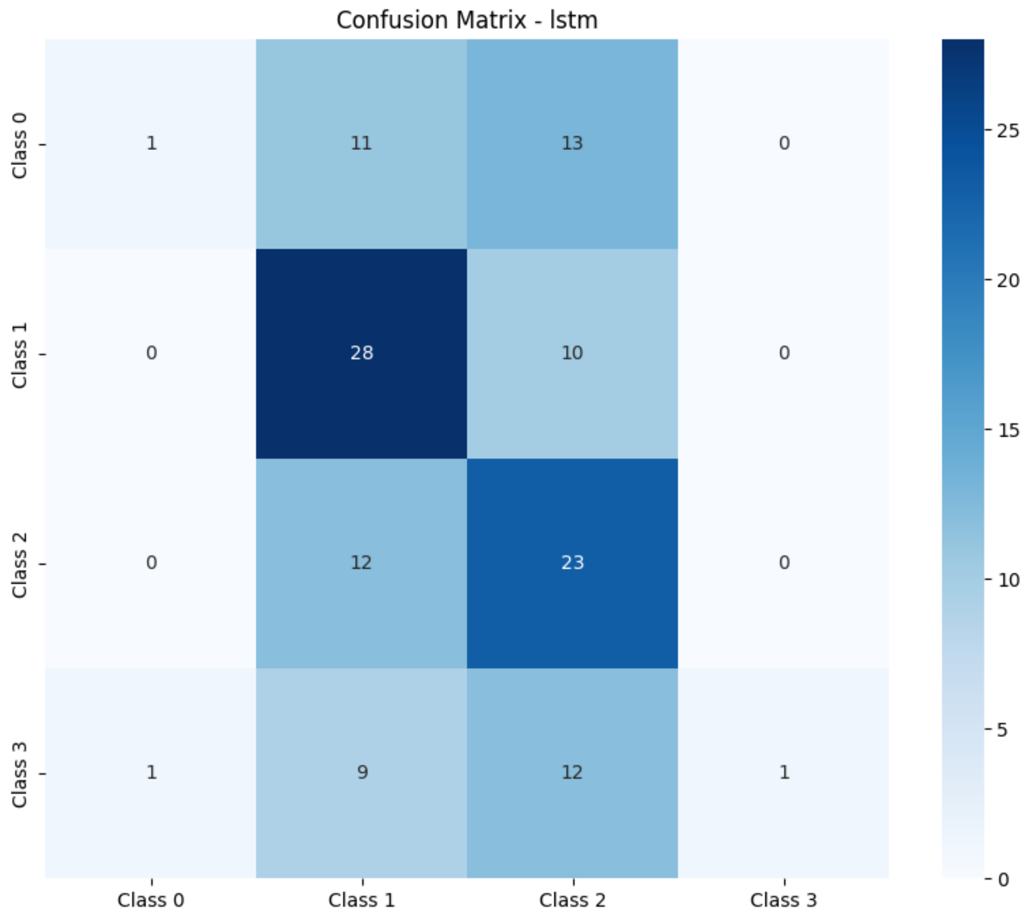

Confusion Matrix - lstm
Class 0
Class 1
Class 2
Class 3
1
11
13
0
0
28
10
0
0
12
23
0
1
9
12
1
Class 0
Class 1
Class 2
Class 3
25
20
15
10
5
0

Figure 9

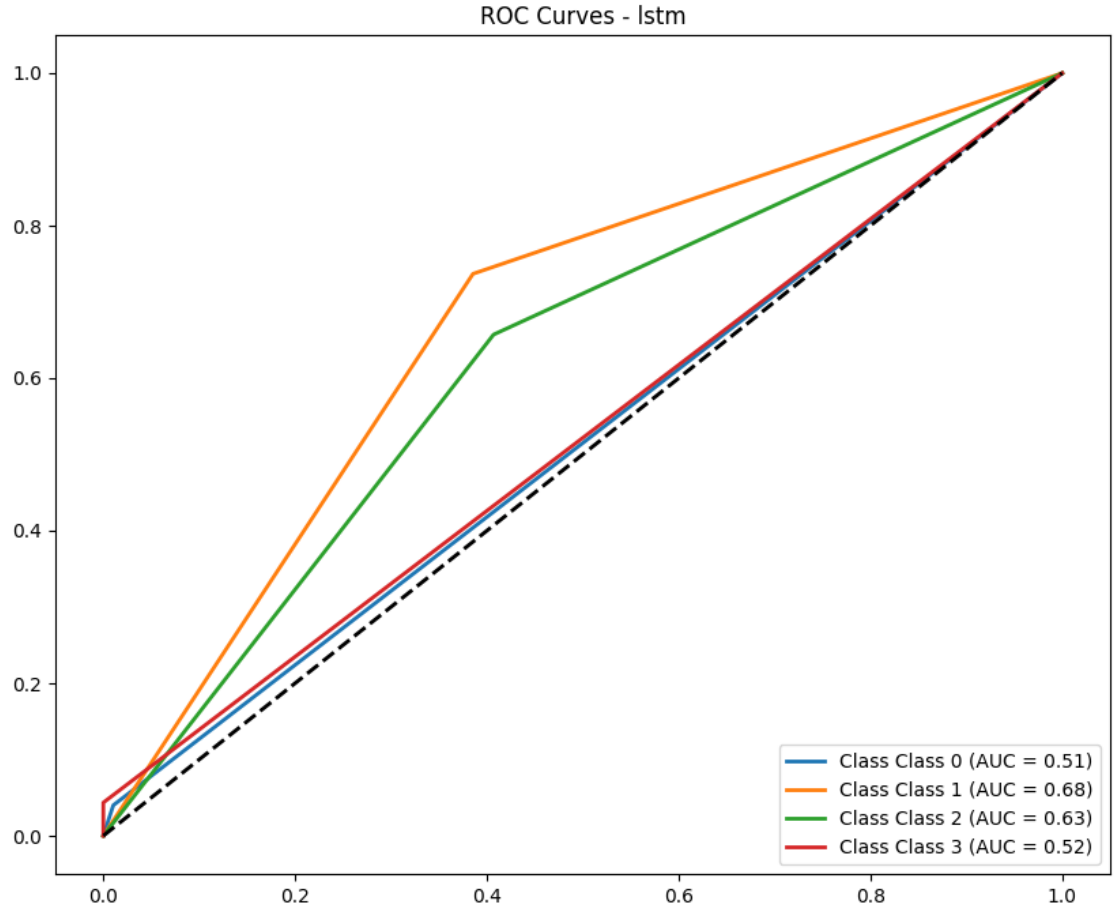

ROC Curves - lstm
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Class Class 0 (AUC = 0.51)
Class Class 1 (AUC = 0.68)
Class Class 2 (AUC = 0.63)
Class Class 3 (AUC = 0.52)

Figure 10

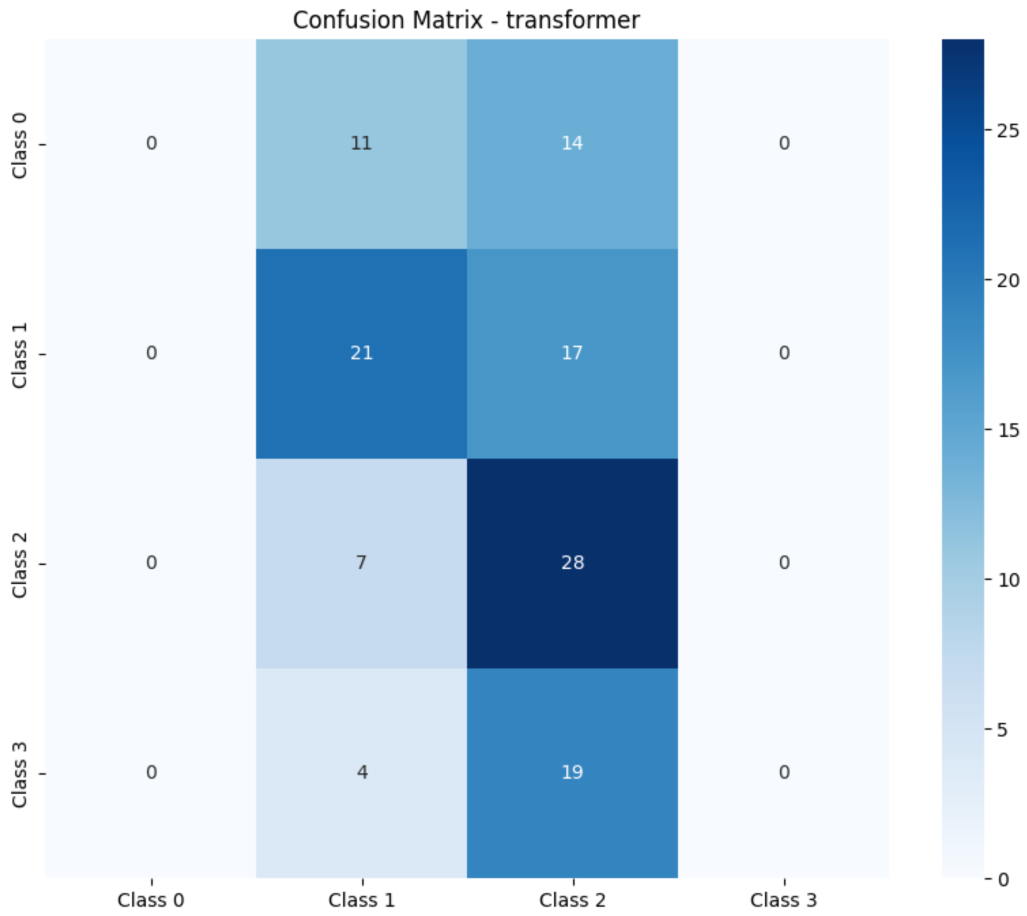

Confusion Matrix - transformer
Class 0
Class 1
Class 2
Class 3
0
11
14
0
0
21
17
0
0
7
28
0
0
4
19
0
Class 0
Class 1
Class 2
Class 3
25
20
15
10
5
0

Figure 11

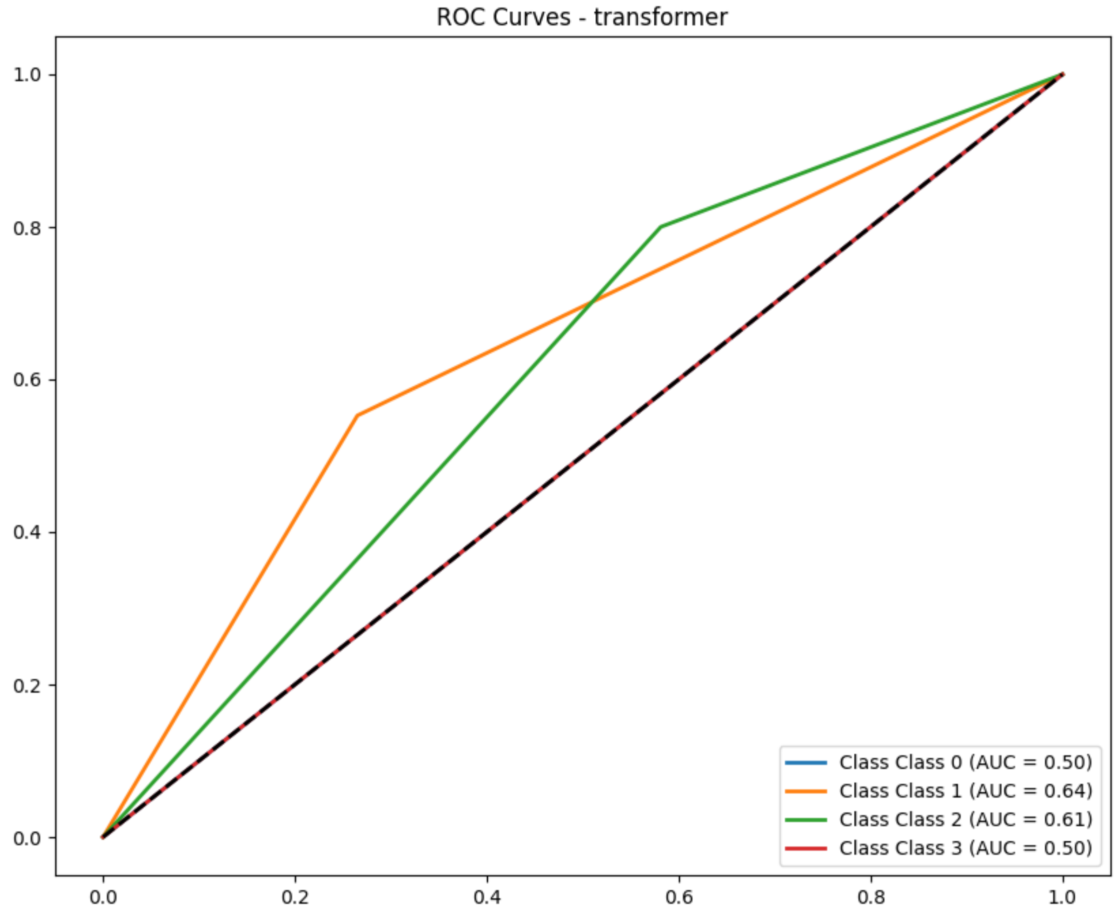

ROC Curves - transformer
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Class Class 0 (AUC = 0.50)
Class Class 1 (AUC = 0.64)
Class Class 2 (AUC = 0.61)
Class Class 3 (AUC = 0.50)

Figure 12

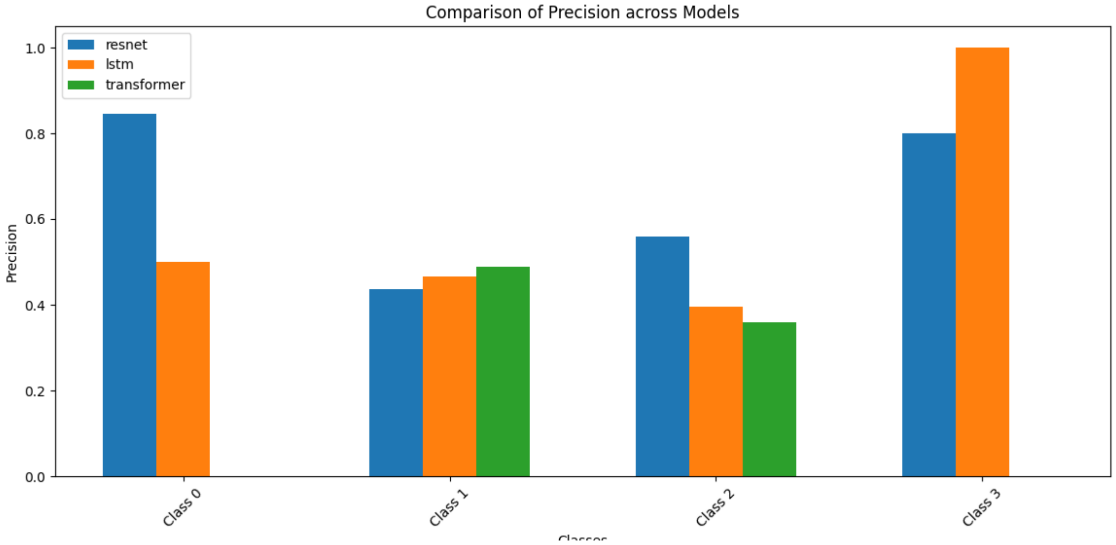

Comparison of Precision across Models
resnet
lstm
transformer
Precision
1.0
0.8
0.6
0.4
0.2
0.0
Class 0
Class 1
Class 2
Class 3
Classes

Figure 13

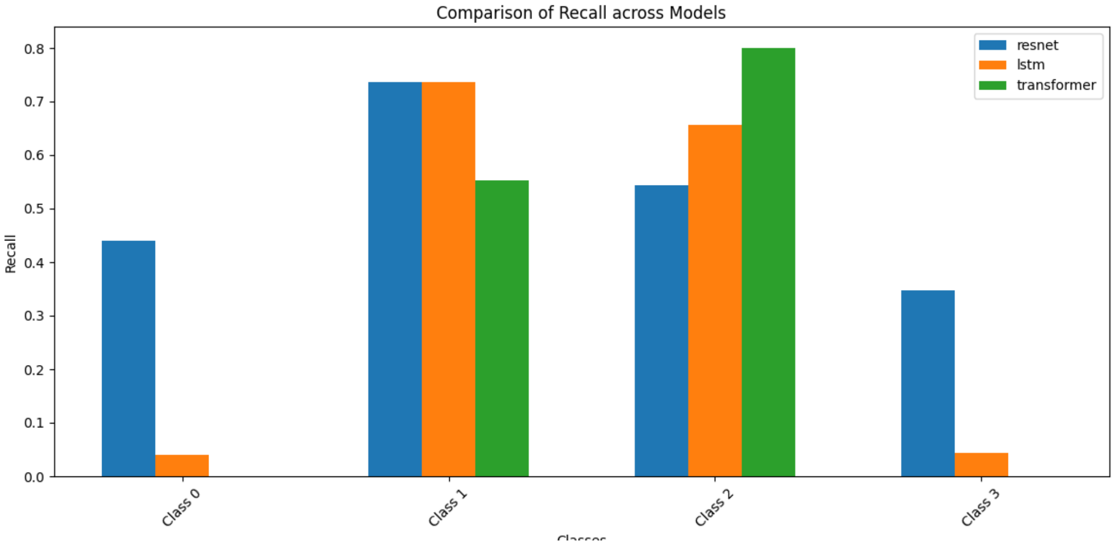

Comparison of Recall across Models
Recall
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
resnet
lstm
transformer
Class 0
Class 1
Class 2
Class 3
Classes

Figure 14

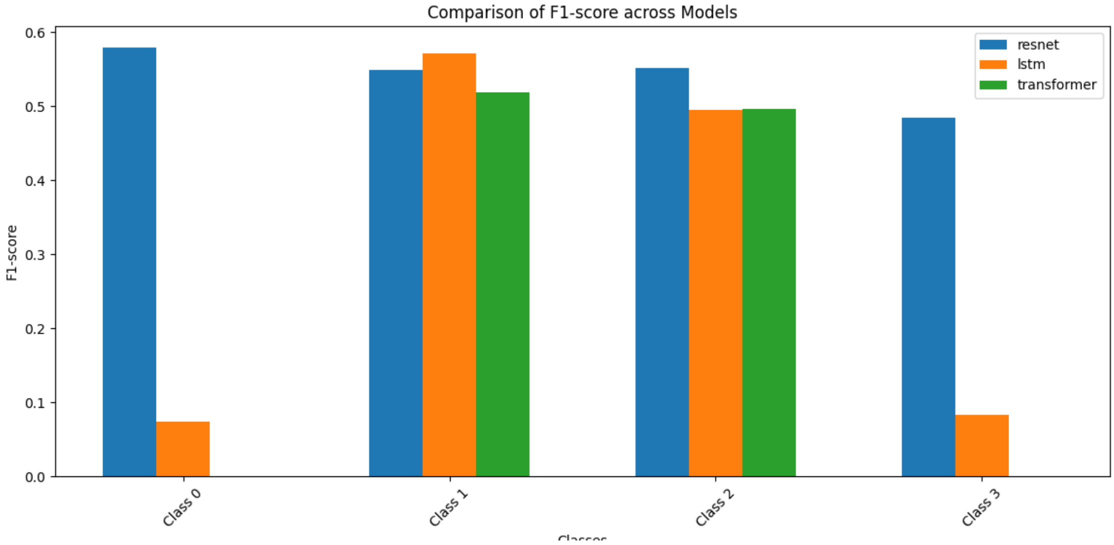

Comparison of F1-score across Models
F1-score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
resnet
lstm
transformer
Class 0
Class 1
Class 2
Class 3
Classes

Figure 15

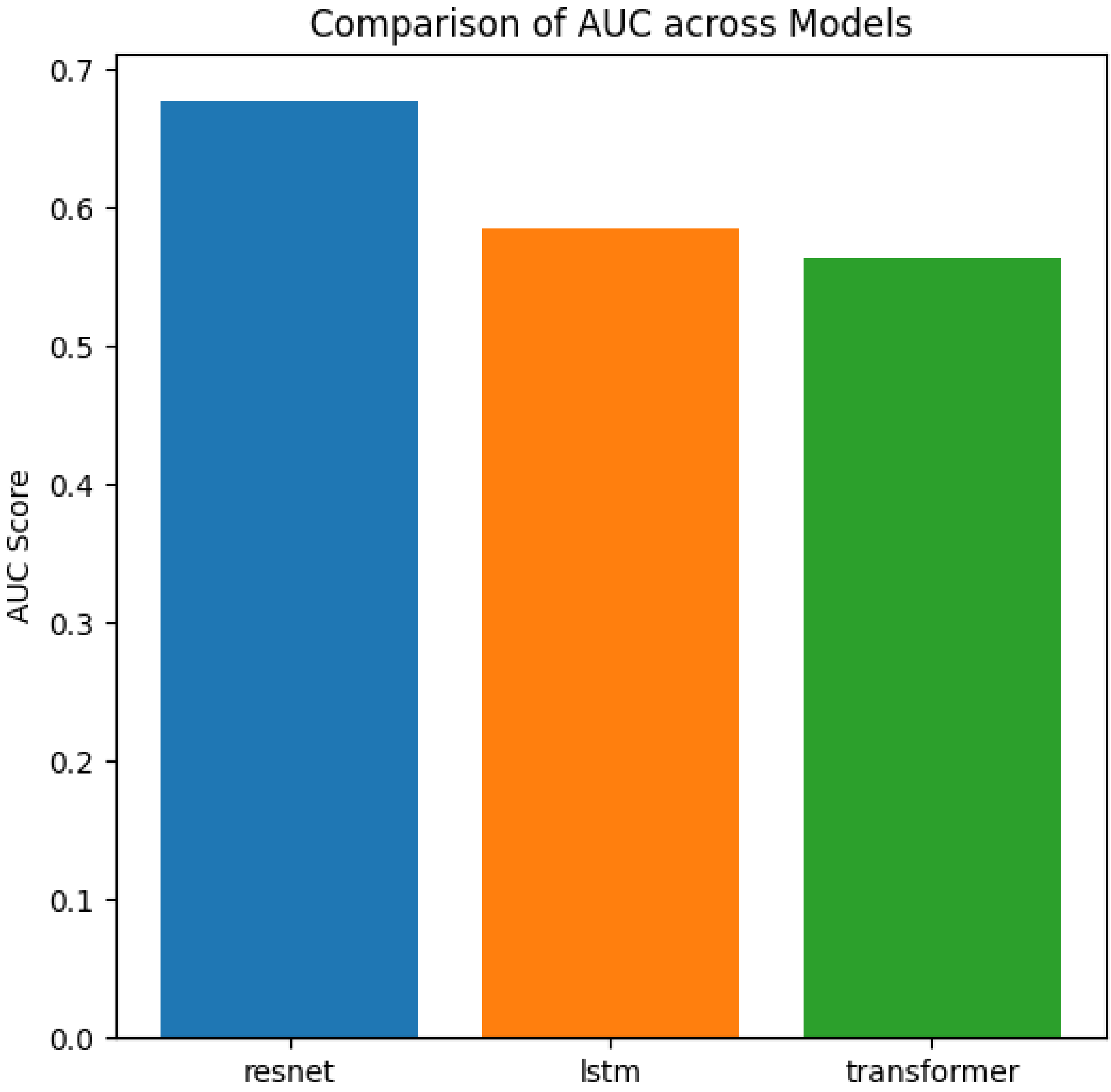

Comparison of AUC across Models
AUC Score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
resnet
lstm
transformer

Figure 16

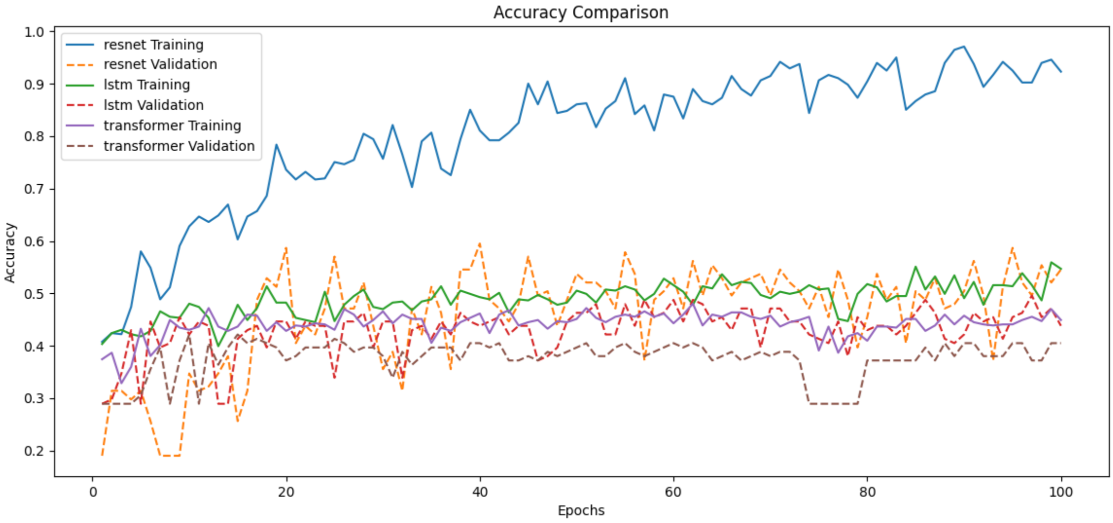

Accuracy Comparison
resnet Training
resnet Validation
lstm Training
lstm Validation
transformer Training
transformer Validation
Accuracy
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
20
40
60
80
100
Epochs

Figure 17

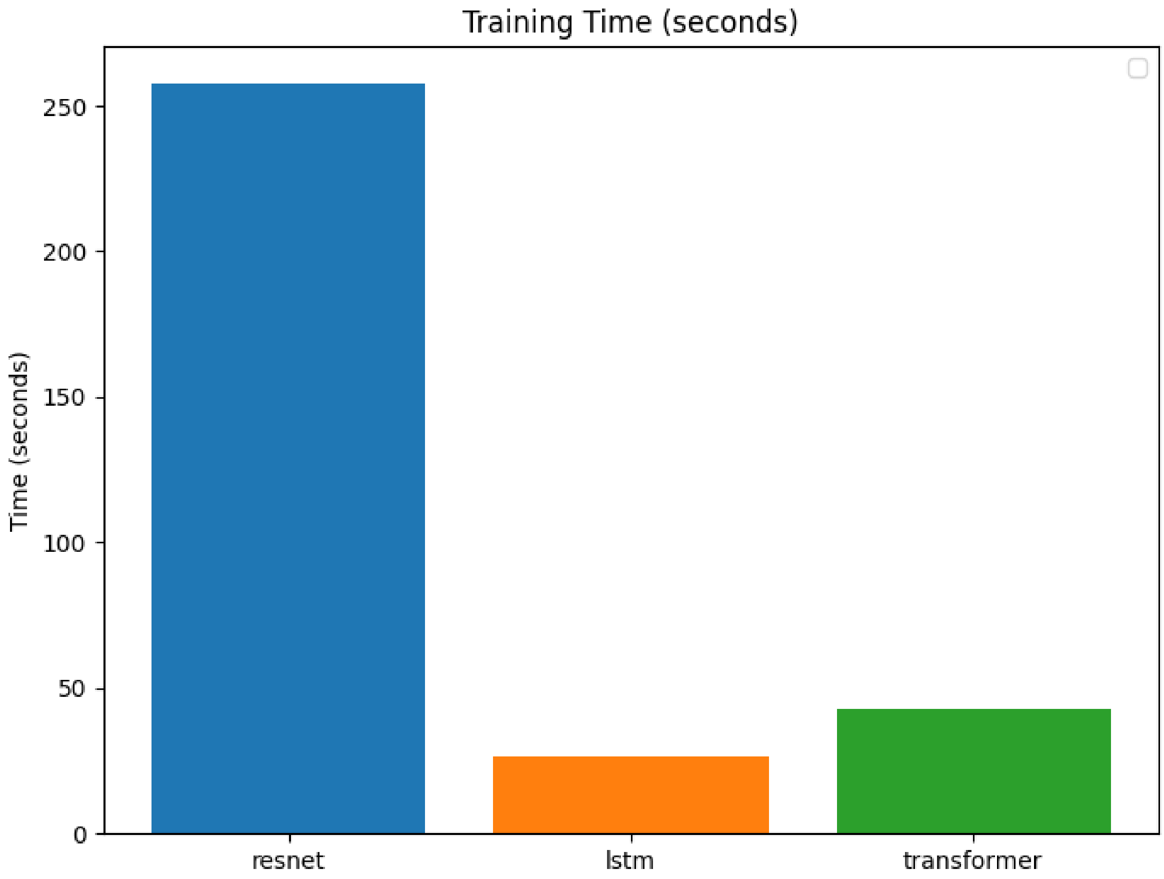

Training Time (seconds)
Time (seconds)
0
50
100
150
200
250
resnet
lstm
transformer

Figure 18

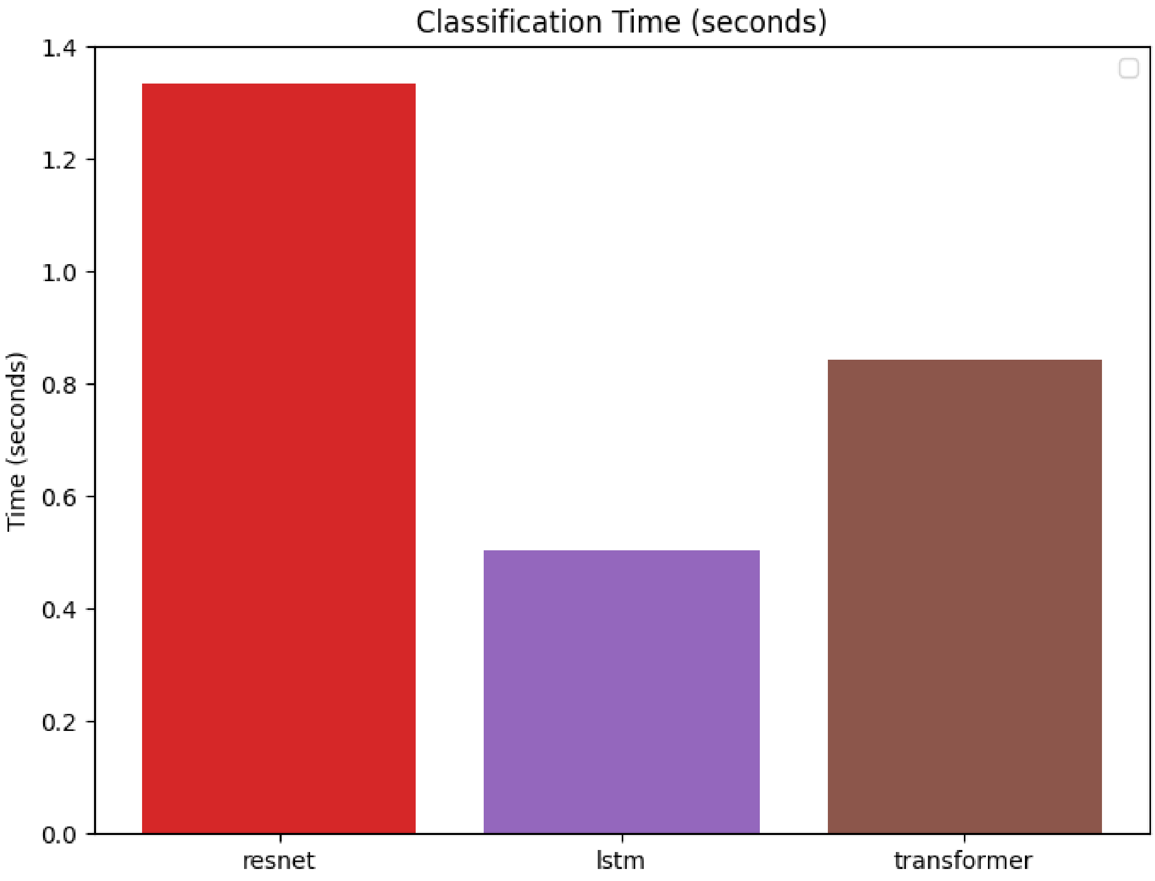
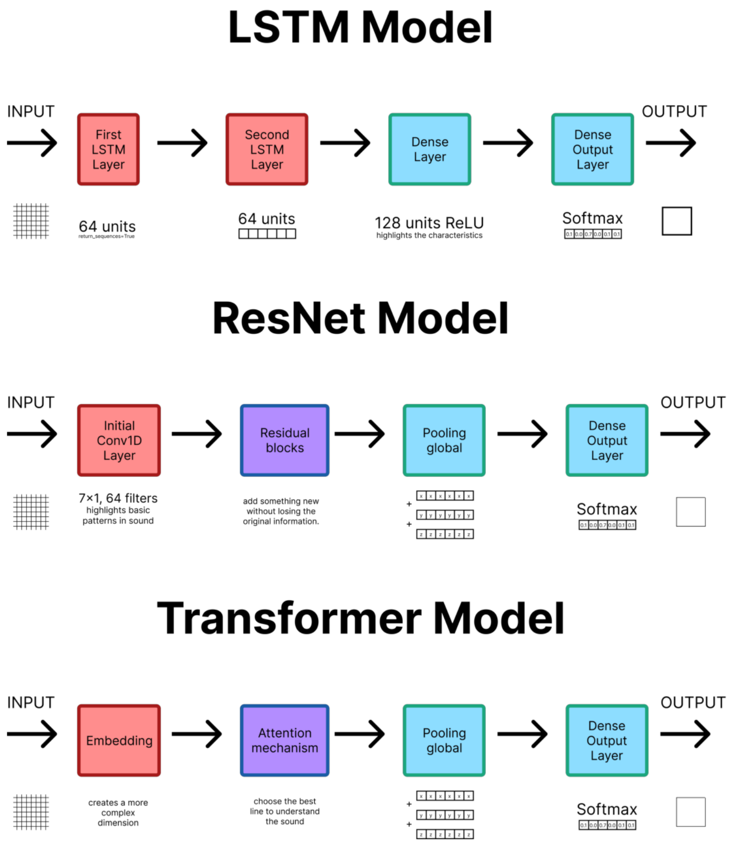

Classification Time (seconds)
Time (seconds)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
resnet
lstm
transformer

Figure 19



LSTM Model

INPUT
First LSTM Layer
Second LSTM Layer
Dense Layer
Dense Output Layer
OUTPUT
64 units
return_sequences=True
64 units
128 units ReLU
highlights the characteristics
Softmax
0.1 0.0 0.7 0.0 0.1 0.1

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.
x x x x x x
+
y y y y y y
+
z z z z z z
Softmax
0.1 0.0 0.7 0.0 0.1 0.1

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
x x x x x x
+
y y y y y y
+
z z z z z z
Softmax
0.1 0.0 0.7 0.0 0.1 0.1

Figure 20

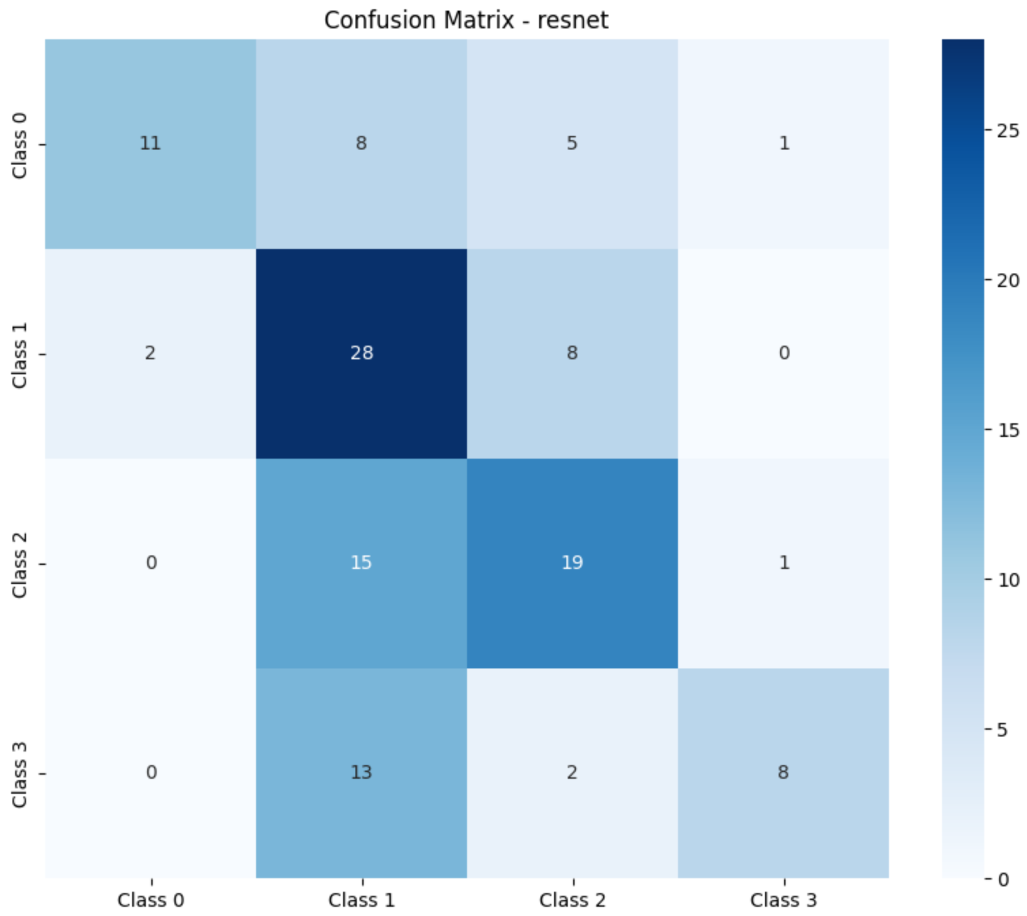

Confusion Matrix - resnet
Class 0
Class 1
Class 2
Class 3
11
8
5
1
2
28
8
0
0
15
19
1
0
13
2
8
Class 0
Class 1
Class 2
Class 3
25
20
15
10
5
0

Figure 21

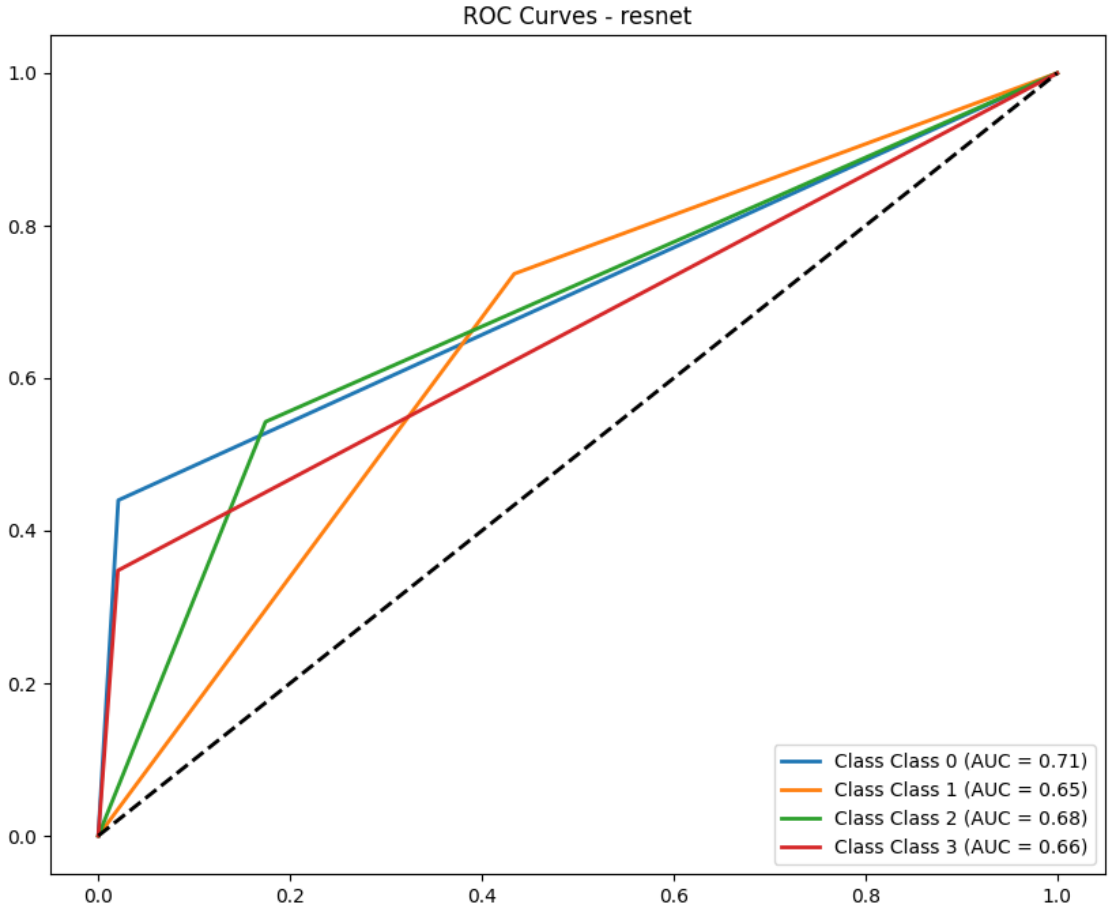

ROC Curves - resnet
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Class Class 0 (AUC = 0.71)
Class Class 1 (AUC = 0.65)
Class Class 2 (AUC = 0.68)
Class Class 3 (AUC = 0.66)

Figure 22

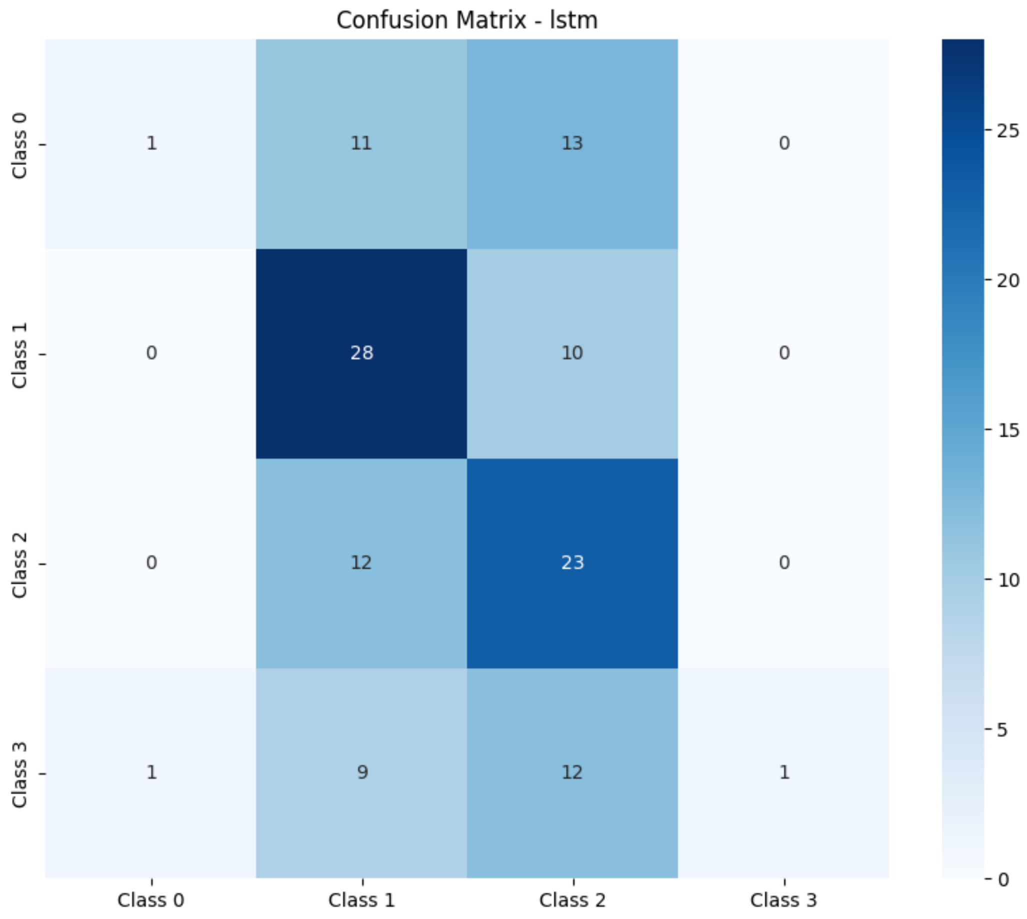

Confusion Matrix - lstm
Class 0
Class 1
Class 2
Class 3
1
11
13
0
0
28
10
0
0
12
23
0
1
9
12
1
Class 0
Class 1
Class 2
Class 3
25
20
15
10
5
0

Figure 23

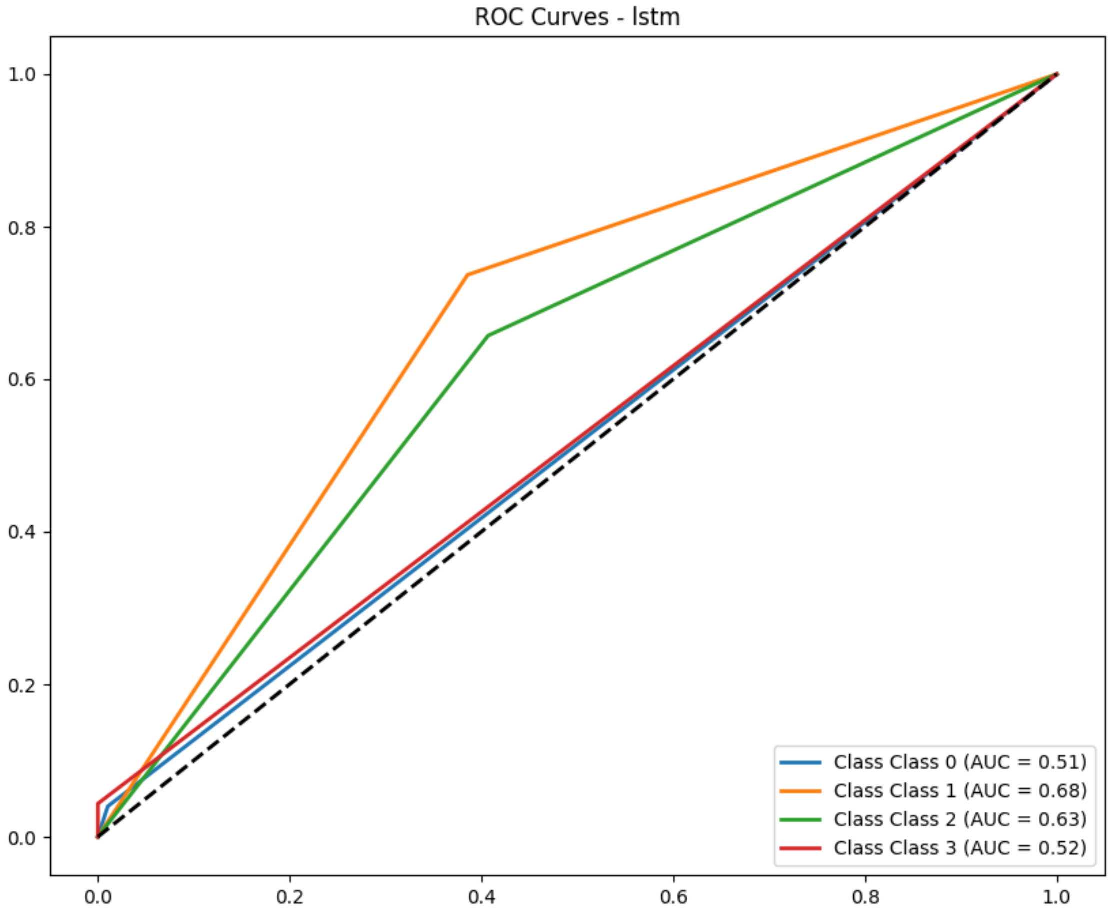

ROC Curves - lstm
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Class Class 0 (AUC = 0.51)
Class Class 1 (AUC = 0.68)
Class Class 2 (AUC = 0.63)
Class Class 3 (AUC = 0.52)

Figure 24

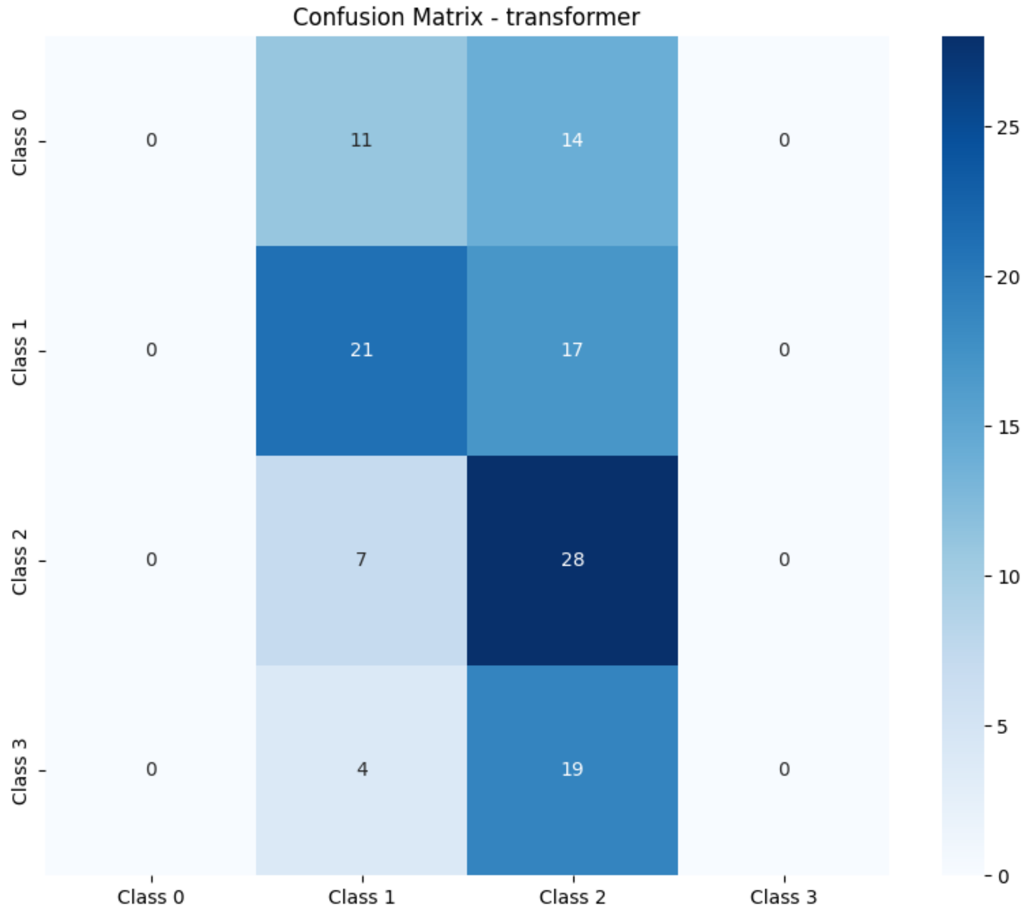

Confusion Matrix - transformer
Class 0
Class 1
Class 2
Class 3
0
11
14
0
0
21
17
0
0
7
28
0
0
4
19
0
Class 0
Class 1
Class 2
Class 3
25
20
15
10
5
0

Figure 25

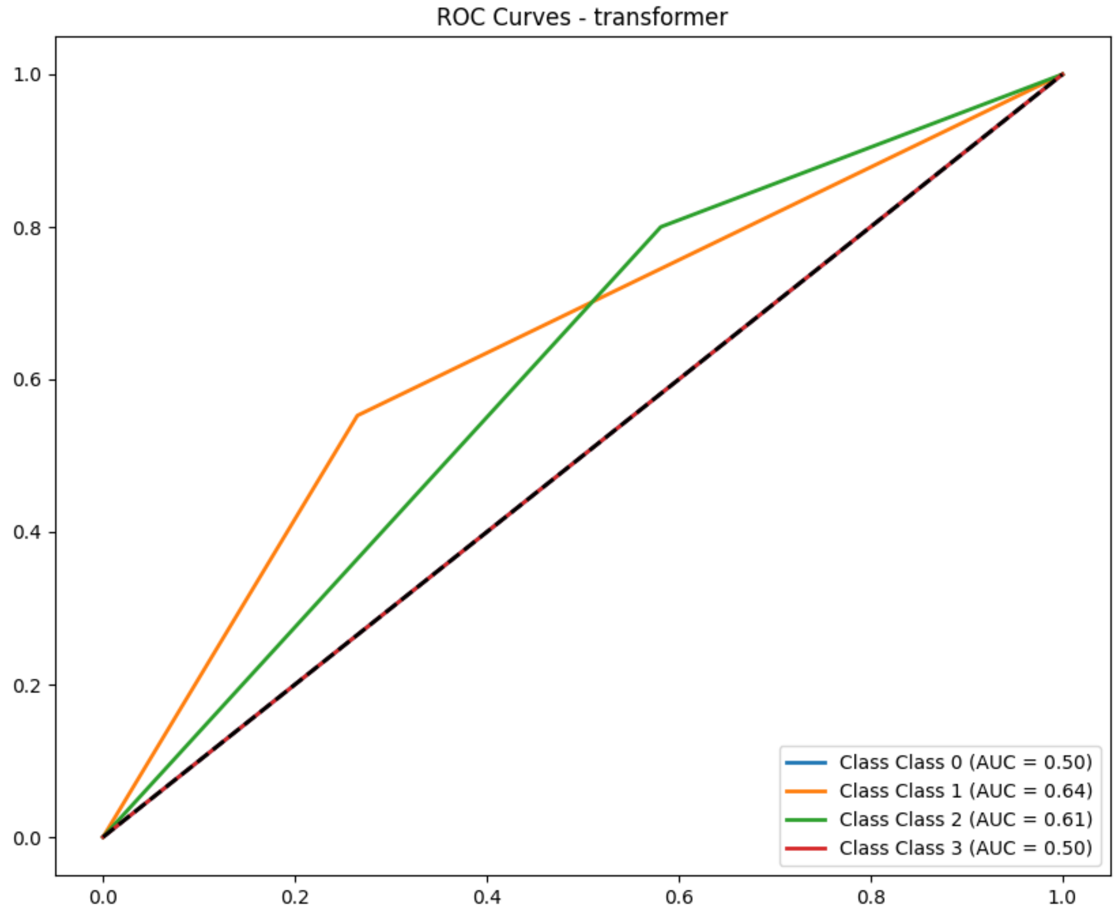

ROC Curves - transformer
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Class Class 0 (AUC = 0.50)
Class Class 1 (AUC = 0.64)
Class Class 2 (AUC = 0.61)
Class Class 3 (AUC = 0.50)

Figure 26

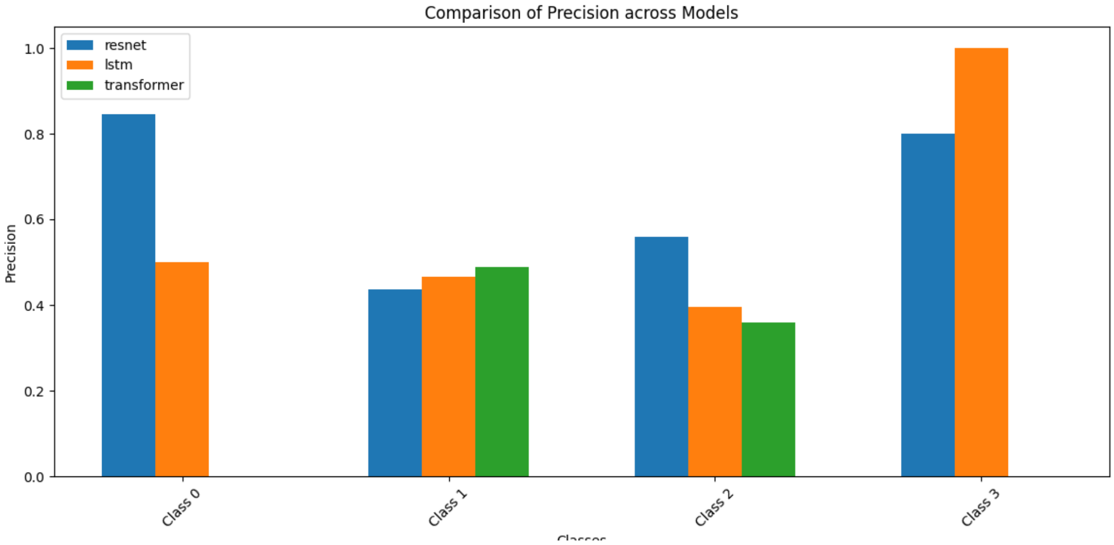

Comparison of Precision across Models
resnet
lstm
transformer
Precision
1.0
0.8
0.6
0.4
0.2
0.0
Class 0
Class 1
Class 2
Class 3
Classes

Figure 27

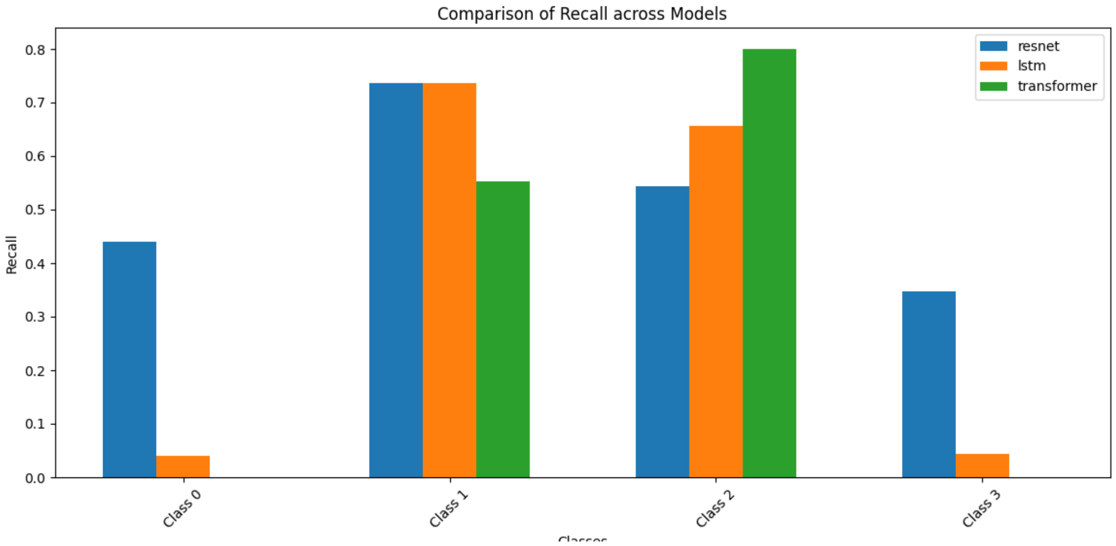

Comparison of Recall across Models
Recall
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
resnet
lstm
transformer
Class 0
Class 1
Class 2
Class 3
Classes

Figure 28

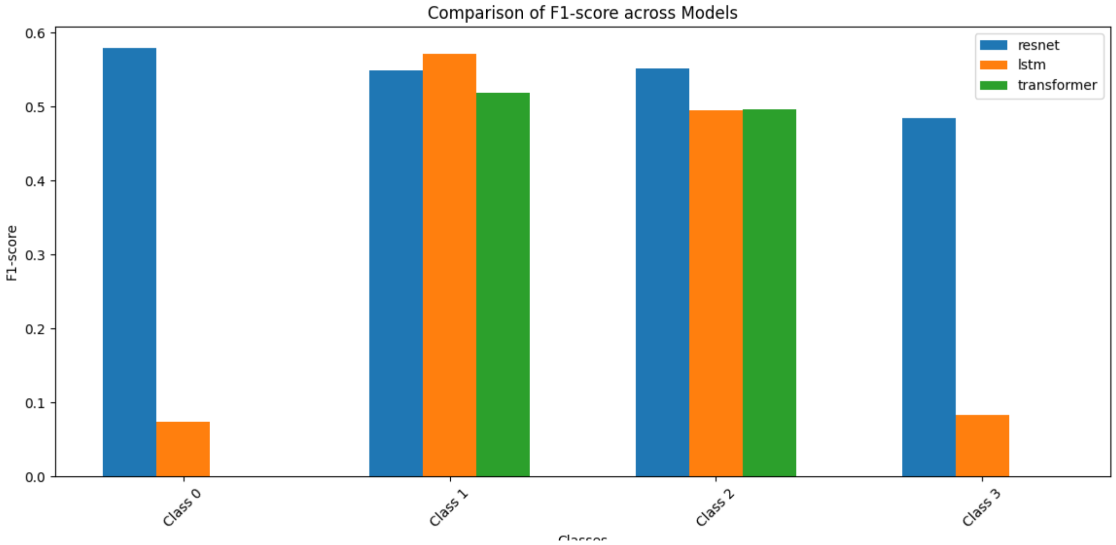

Comparison of F1-score across Models
F1-score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Class 0
Class 1
Class 2
Class 3
resnet
lstm
transformer

Figure 29

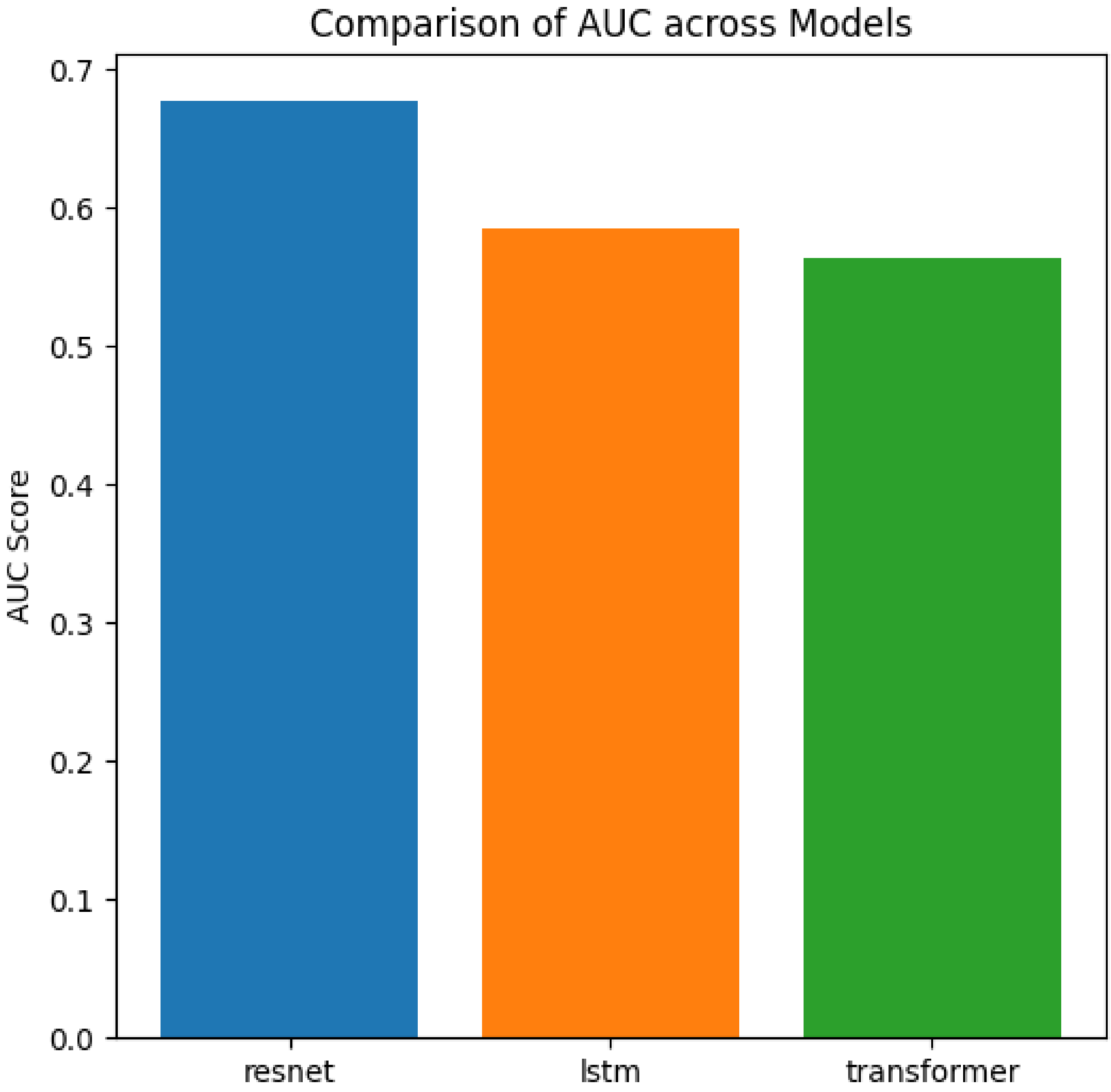

Comparison of AUC across Models
AUC Score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
resnet
lstm
transformer

Figure 30

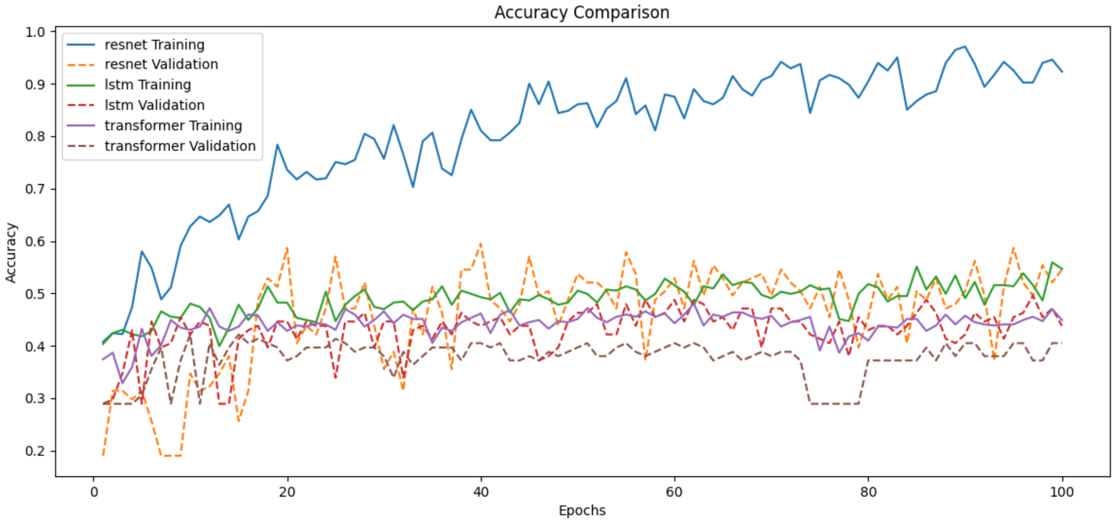

Accuracy Comparison
resnet Training
resnet Validation
lstm Training
lstm Validation
transformer Training
transformer Validation
Accuracy
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
20
40
60
80
100
Epochs

Figure 31

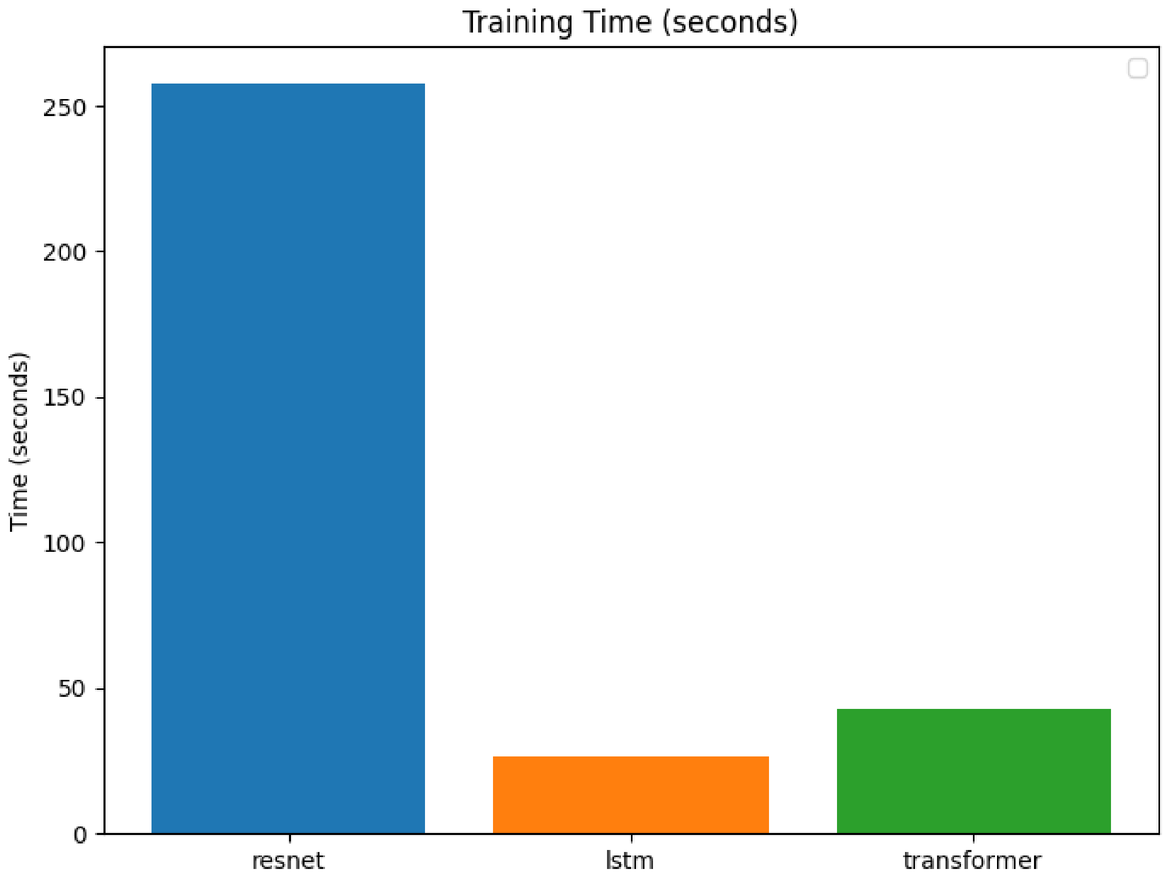
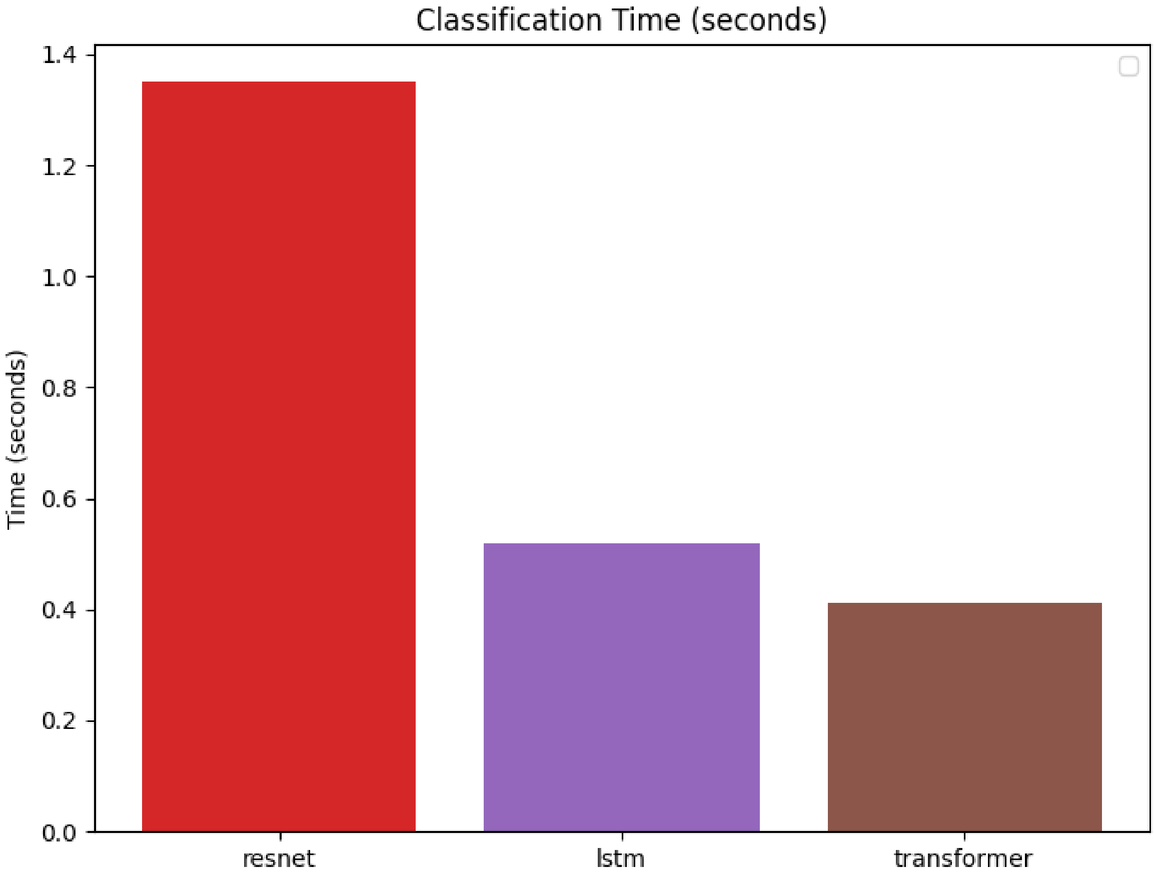

Training Time (seconds)
Time (seconds)
0
50
100
150
200
250
resnet
lstm
transformer

Figure 32


Classification Time (seconds)
Time (seconds)
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
resnet
lstm
transformer

Configuration Settings

```
data_path: D:\deepship\original
metadata_path: C:\Users\gumar\OneDrive\Área de Trabalho\Pesquisa UBO\Revisão Literaria\UrbanSound8K\UrbanSound8K\m
epochs: 100
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
  Delta: False
  LOFAR: False
  DEMON: False
selected_models: ['resnet', 'lstm', 'transformer']
```