Music Emotion Recognition

Group 12

E/17/040 : Chandrasena M.M.D.
E/17/356 : Upekha H.P.S.
E/17/407 : Wijesooriya H.D.

www.menti.com

Introduction

- **What is Music Emotion Recognition**
  - **Recognize the emotion that music expresses.**
    - Extract and analyze music features, and map the relations between music features and the emotion space.

- **Importance of Music Emotion Recognition**
  - MER has an enormous significance in real world applications such as,
    - Music recommendation
    - Music therapy
    - Music data management

Research Problem and Objectives

Problem:

➔ Accuracy of MER (Music Emotion Recognition) systems are much lower.

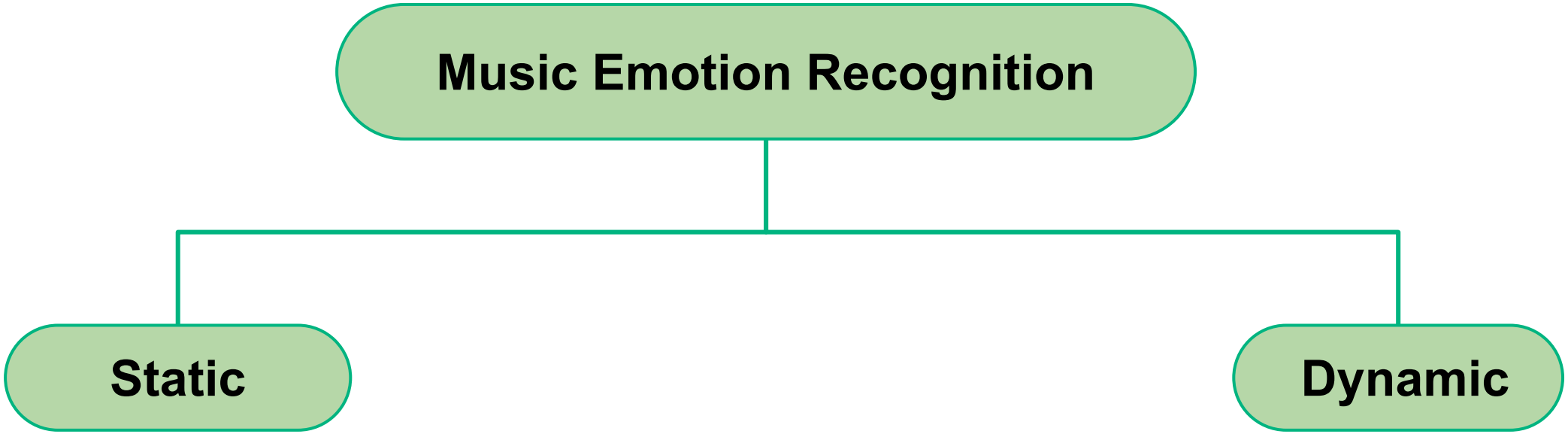
Objectives:

➔ Identifying the drawbacks of existing systems

➔ Increasing the accuracy of those existing systems

➔ Implementing a dynamic MER system using DNN concepts with high accuracy

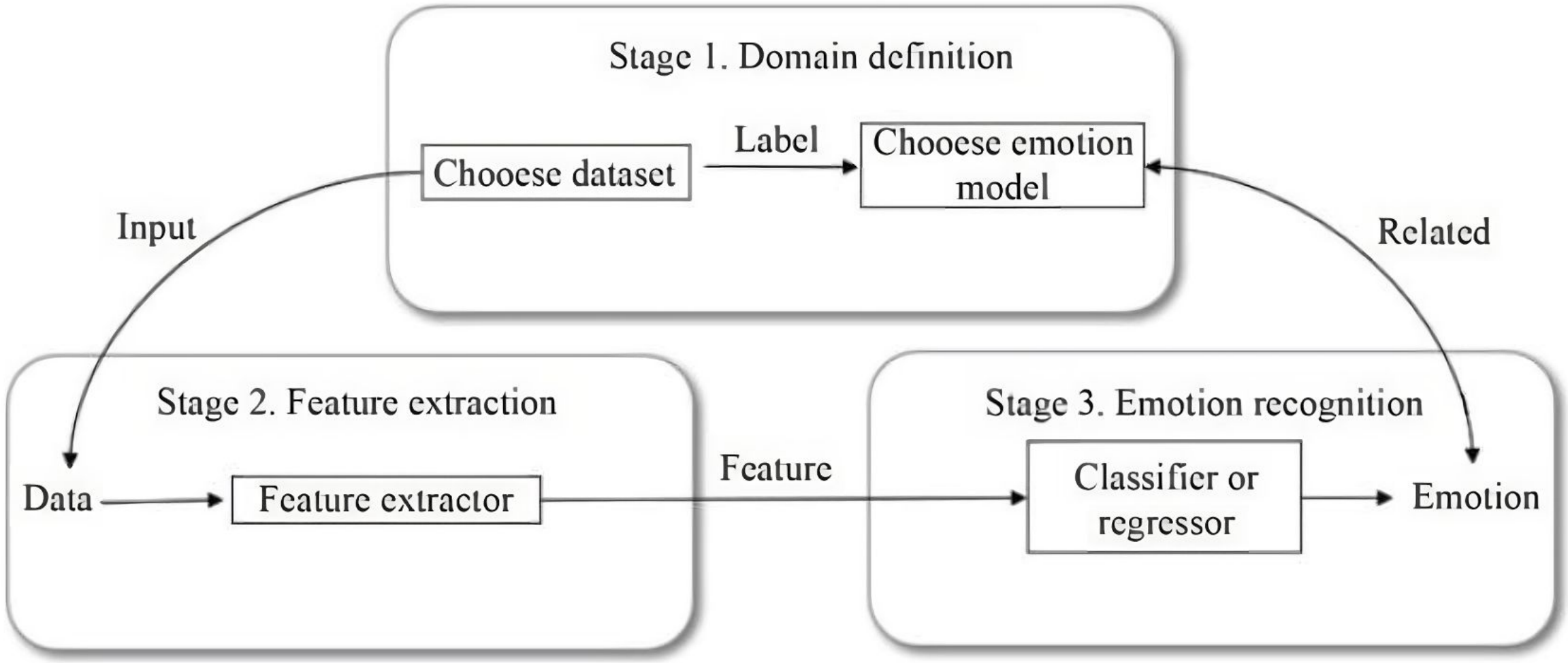
Music Emotion Recognition Methods



- Song level MER
- Assign overall emotion label

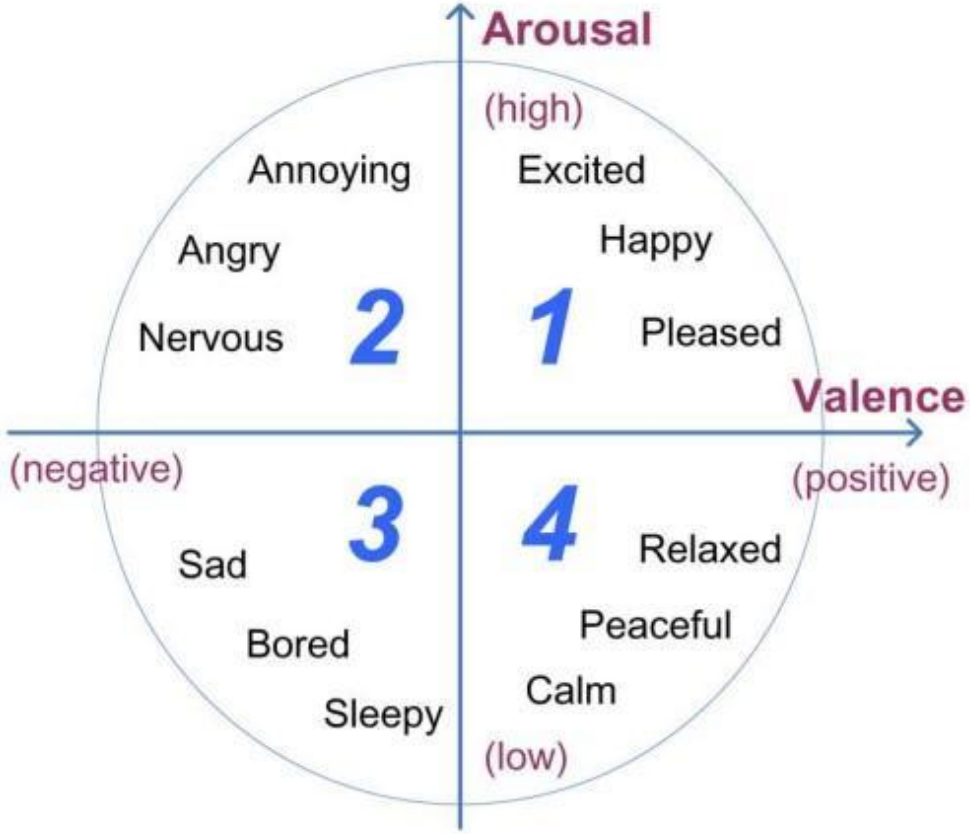
- Music Emotion Variation Detection (MEVD)
- Consider emotion in music as a changing process

Music Emotion Recognition Framework



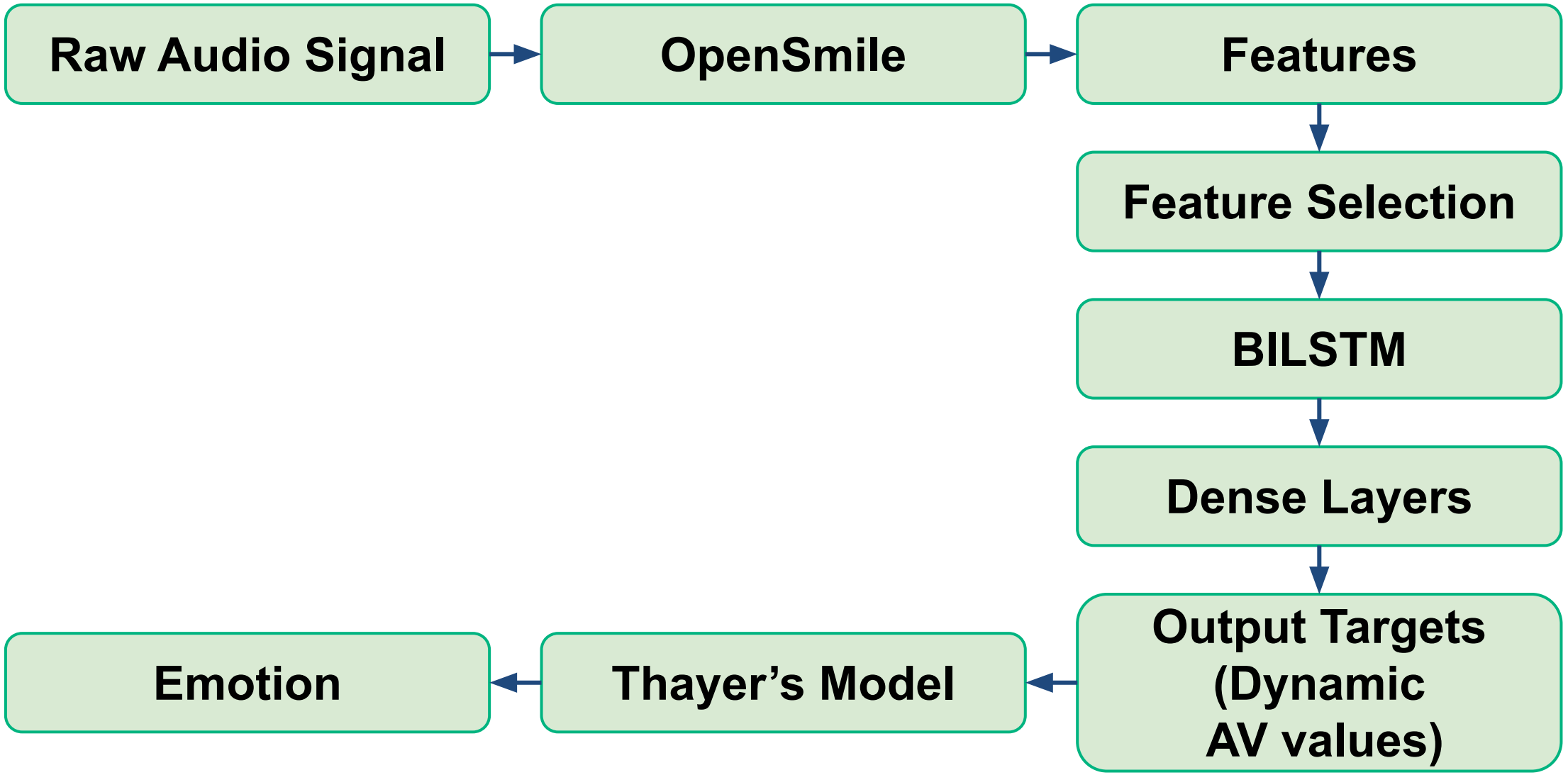
Emotion Models

Model name	Application Domain	Emotion conceptualization	Number of classes /dimensions
Hevner affective ring	Music	Categorical	67
Russell's model	General	Dimensional	2
Thayer's model	General	Dimensional	2



Thayer's arousal-valence plane

Proposed Methodology



Experiments and Findings

Dataset

- **MediaEval Dataset for Emotional Analysis in Music - DEAM dataset**

- **Obtained from Kaggel**
  https://www.kaggle.com/datasets/imsparsh/deam-mediaeval-dataset-emotional-analysis-in-music

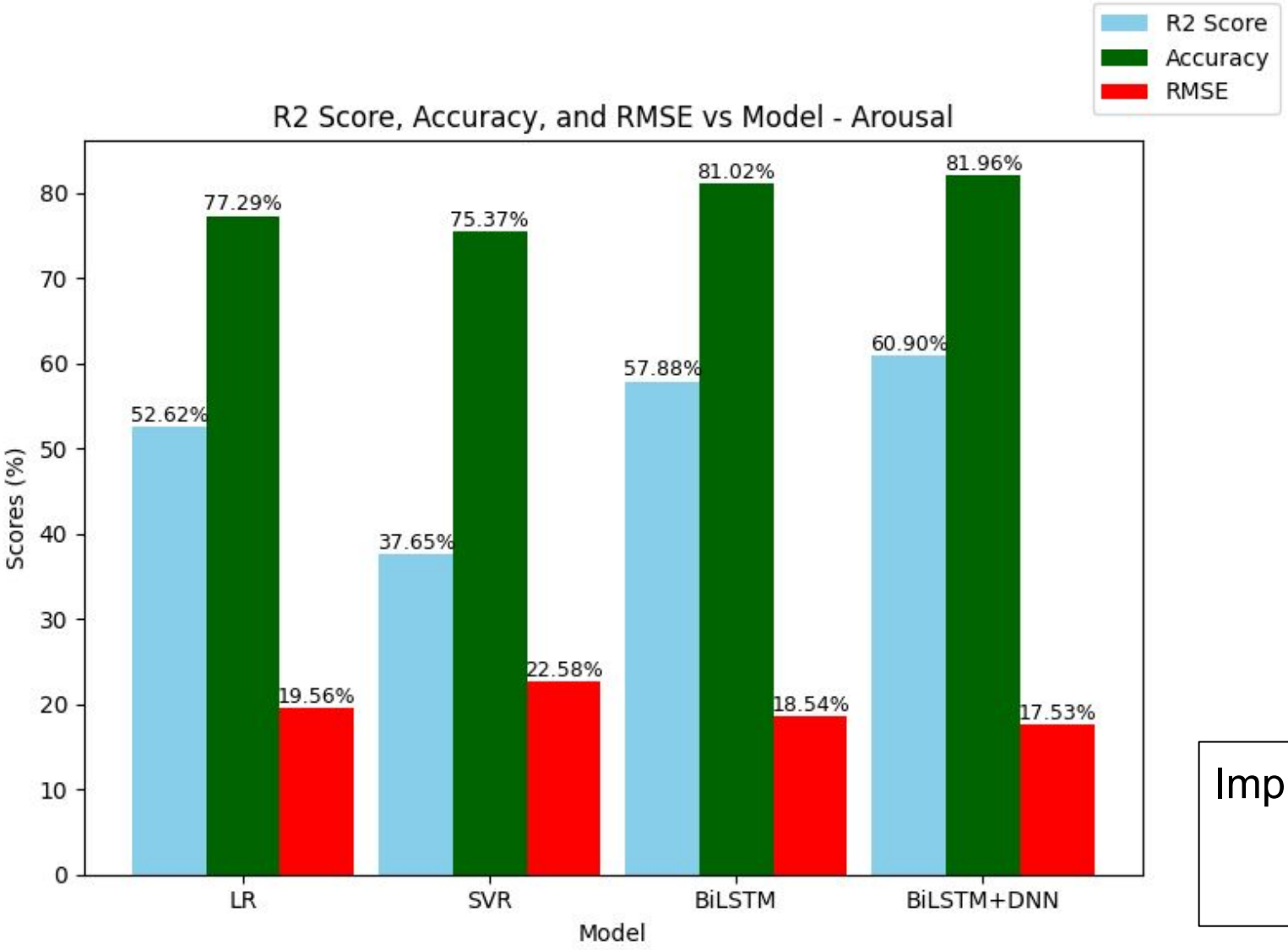
- **1802 songs - annotated with both arousal and valence values for every 0.5s**
- **Feature variables = 260**
- **Target variables = 2 (Arousal and Valence)**
- **Number of data points = 106132**

frameTime	F0final_sma_stddev	F0final_sma_amean	voicingFinalUnclipped_sma_stddev	voicingFinalUnclipped_sma_amean	jitterLocal_sma_stddev	jitterLocal_sma_amean	jitterDDP_sma_stddev	jitterL
15.0	11.31167	71.10648	0.030939	0.776627	0.080092	0.071759	0.060902	
15.5	10.70966	75.66485	0.021004	0.761171	0.101999	0.135746	0.072888	
16.0	17.01124	84.17995	0.021248	0.760293	0.141607	0.195160	0.124378	
16.5	34.97898	101.08820	0.021063	0.754002	0.201426	0.216901	0.129330	
17.0	38.56500	110.98110	0.026652	0.756074	0.188570	0.154176	0.115892	
17.5	37.20660	83.88412	0.030379	0.747746	0.088179	0.087981	0.095062	
18.0	107.86010	112.24430	0.034547	0.739723	0.112436	0.096967	0.097900	
18.5	106.71090	119.76980	0.032857	0.745989	0.170641	0.145237	0.144709	
19.0	97.60259	97.29289	0.028055	0.738480	0.172240	0.155860	0.157357	
19.5	116.56300	120.72790	0.036618	0.750792	0.120717	0.077591	0.112953	

Baseline, Reference and Proposed Model

Model	Parameters
Linear Regression (baseline model)	Folds = 5
SVR	Folds = 5
BiLSTM (reference model)	Folds = 5, Learning Rate=0.001, Epochs=25, Batch Size=32, Optimizer=Adam, Activation Function=relu
BiLSTM + DNN (proposed model)	Folds = 5, Learning Rate=0.001, Epochs=25, Batch Size=32, Optimizer=Adam, Activation Function=relu, Dense Layer Units=512

Model Comparison - Arousal

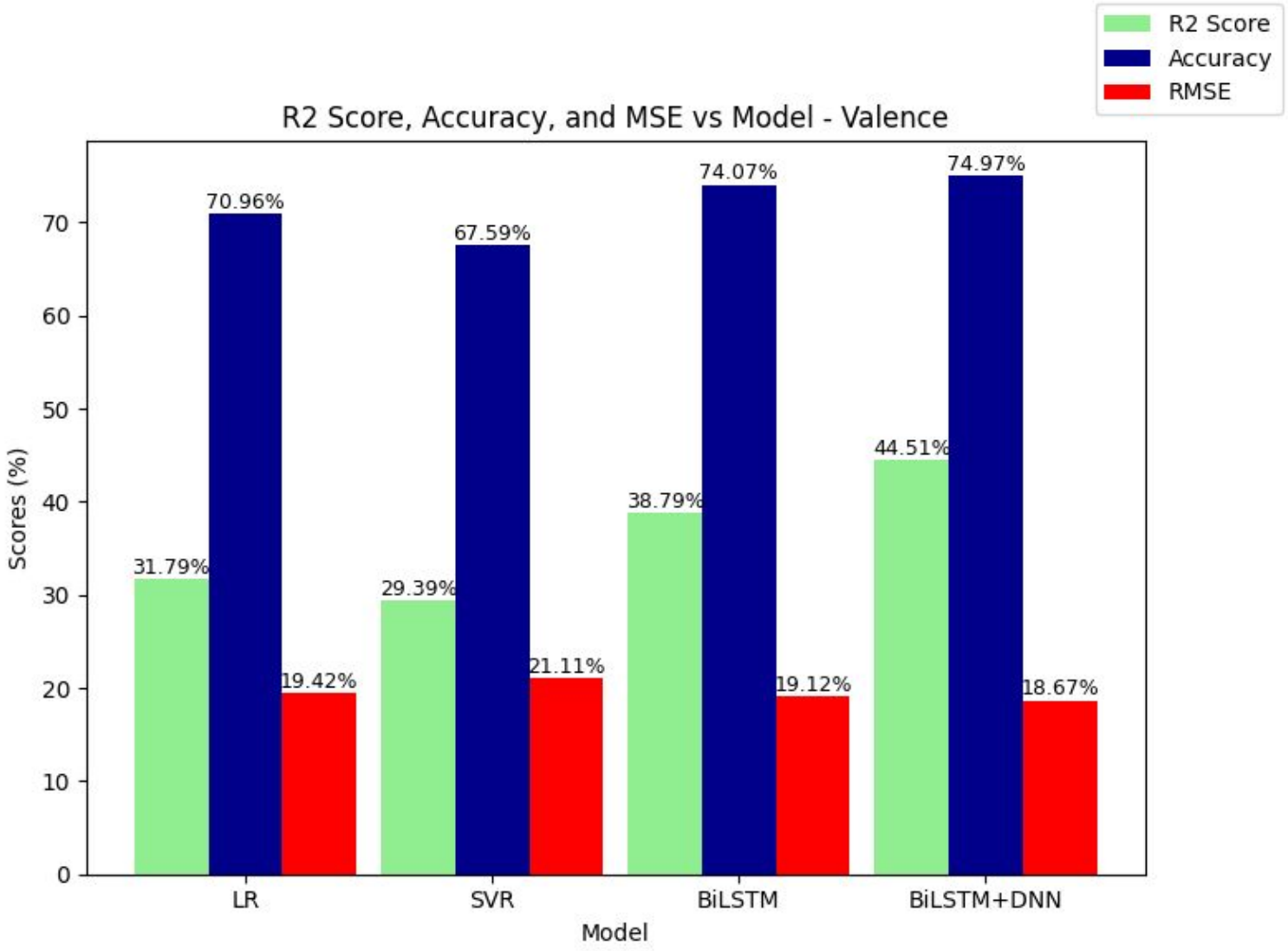


Compared to Baseline Model

Score	Improvement (%)
R2 Score	15.08%
Accuracy	6.04%
RMSE	10.38%

$$\text{Improvement} = \frac{(\text{our model} - \text{baseline})}{\text{baseline}} \times 100$$

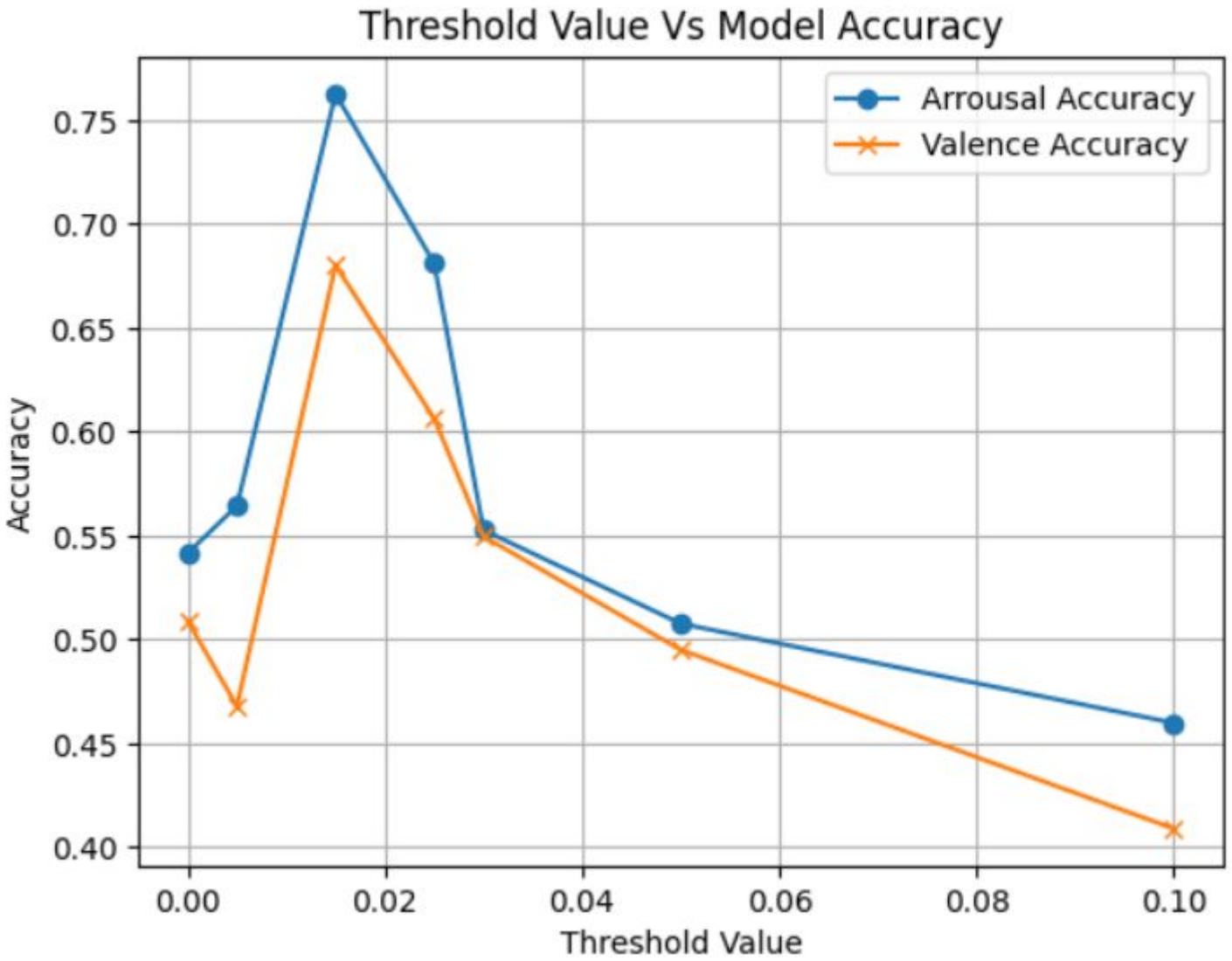
Model Comparison - Valence



Compared To Baseline Model

Score	Improvement (%)
R2 Score	40.01%
Accuracy	5.65%
RMSE	3.86%

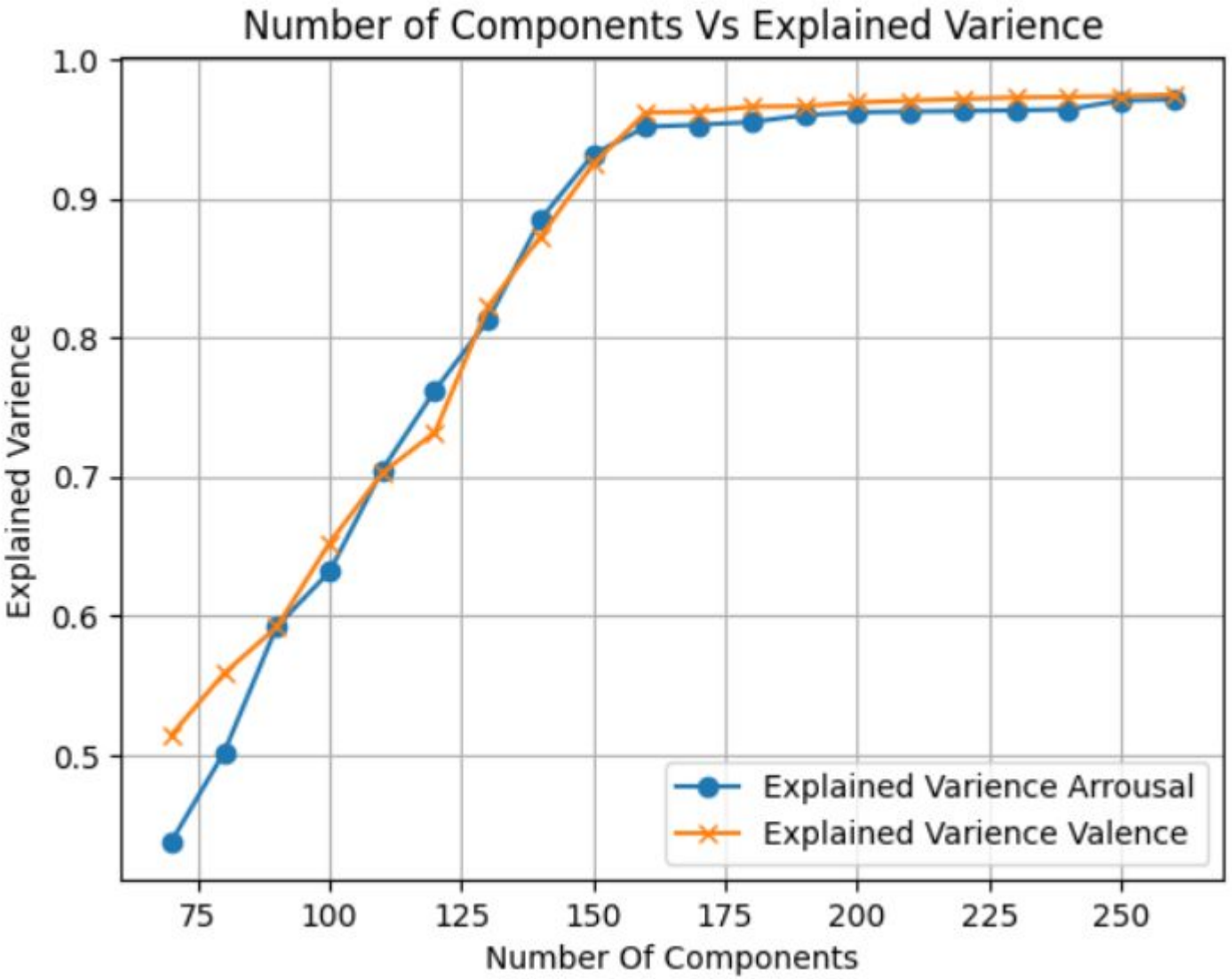
Feature Selection - CFS Method (Correlation Based Feature Selection)



Selected Threshold value : 0.015

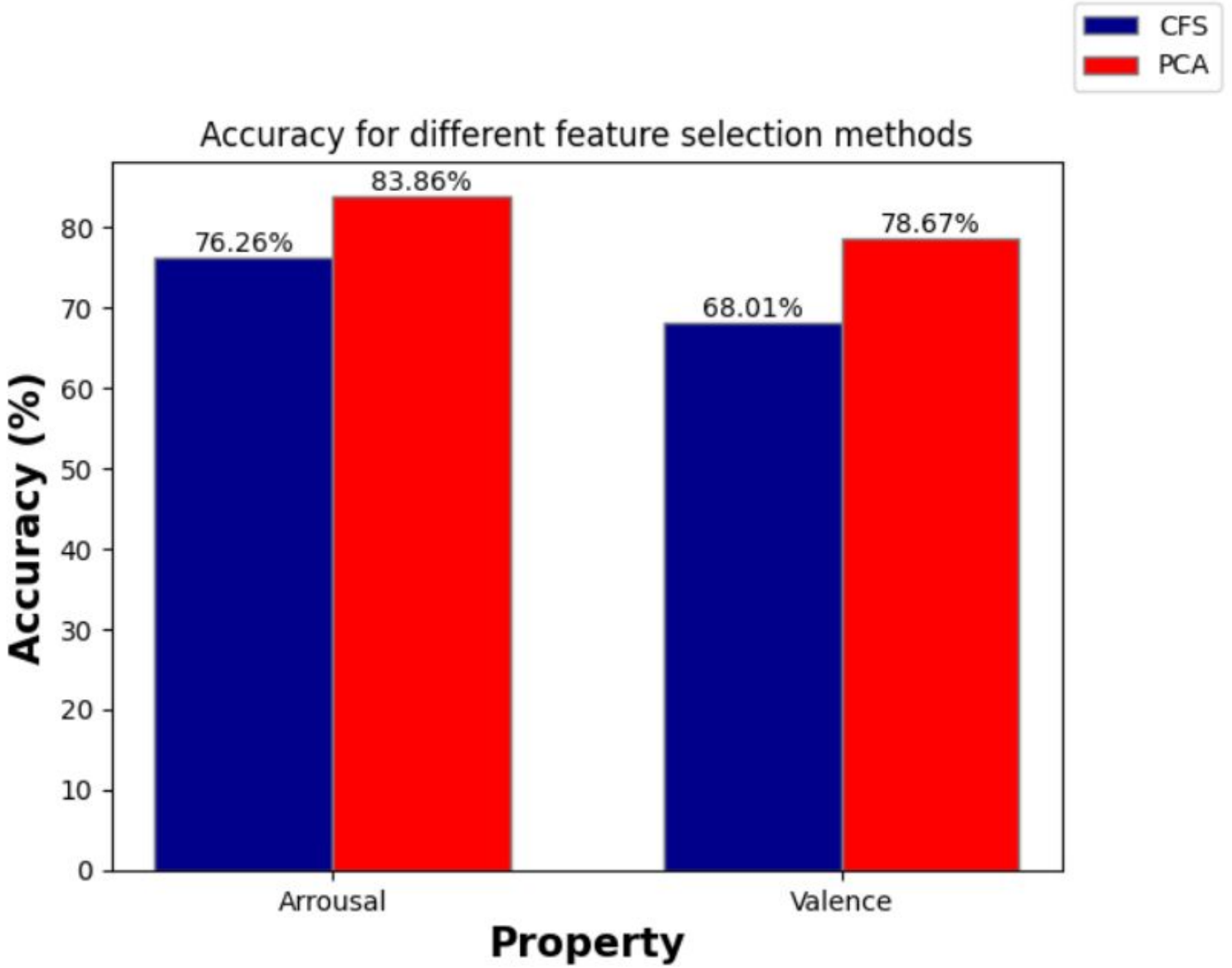
Number of features in the Feature set :150

Feature Selection - PCA Method (Principal Component Analysis)

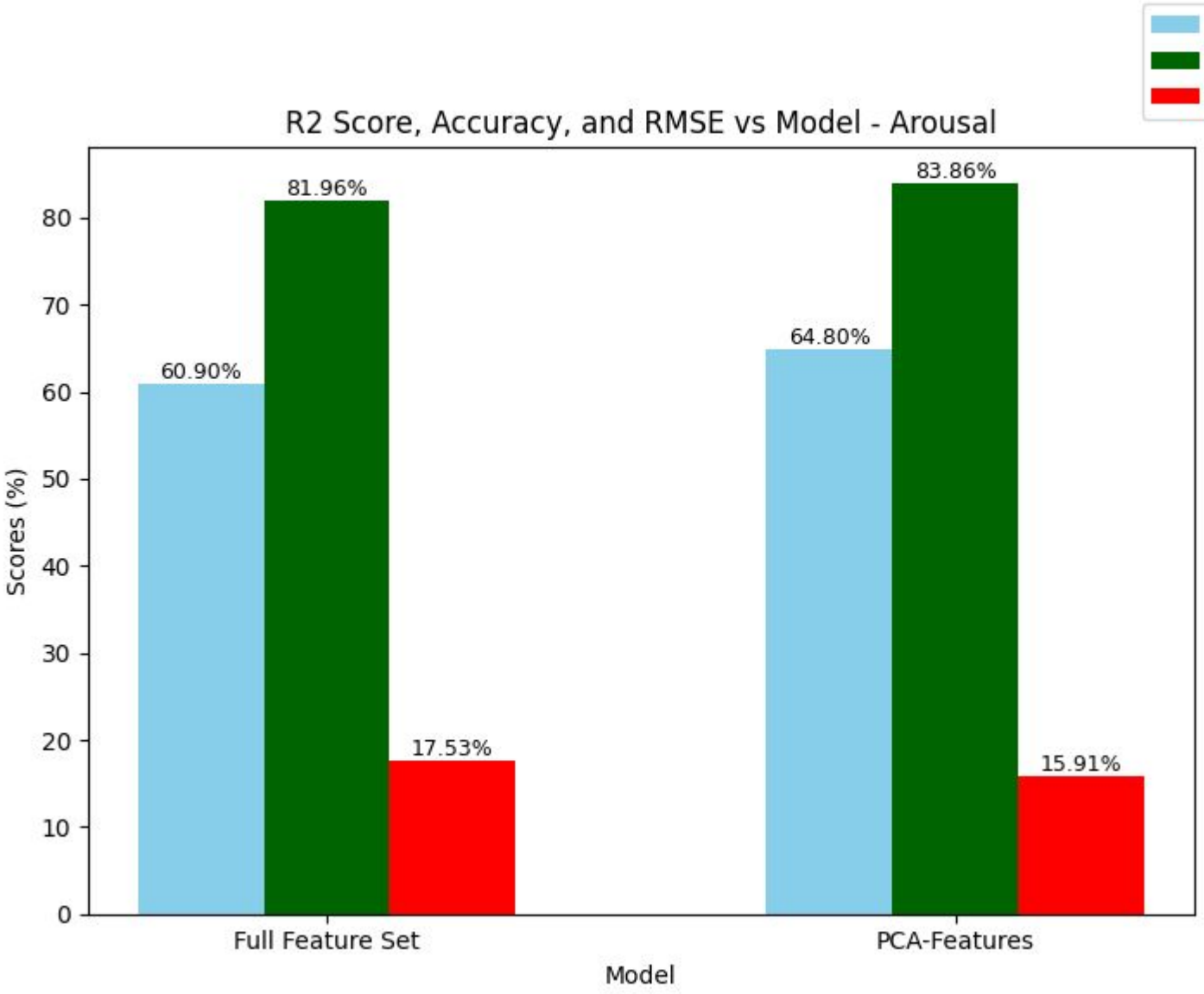


Optimum number of Components : 159

Comparison of Accuracies of Feature Selection

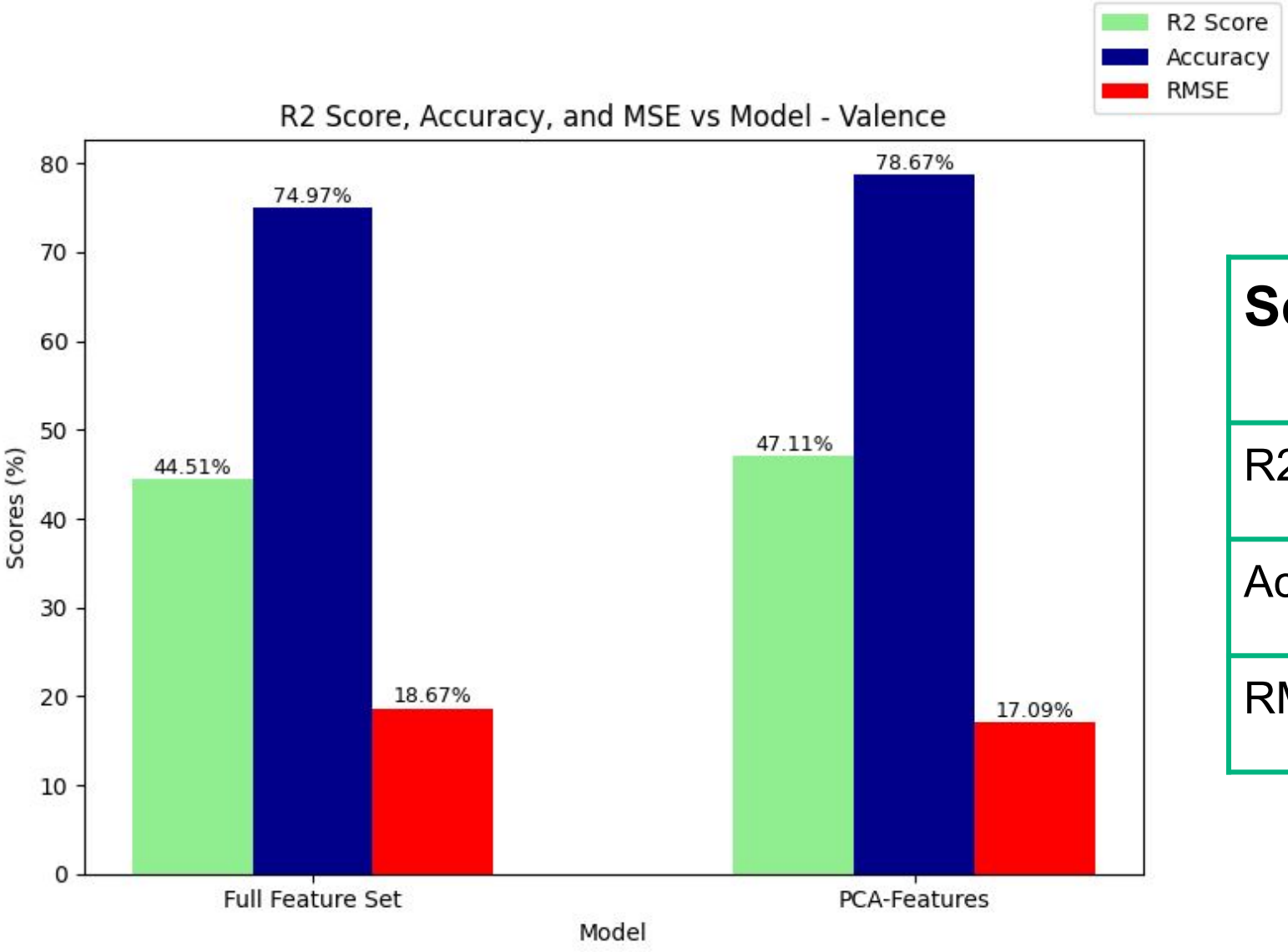


Model Accuracies for Full Feature Set and PCA Features - Arousal



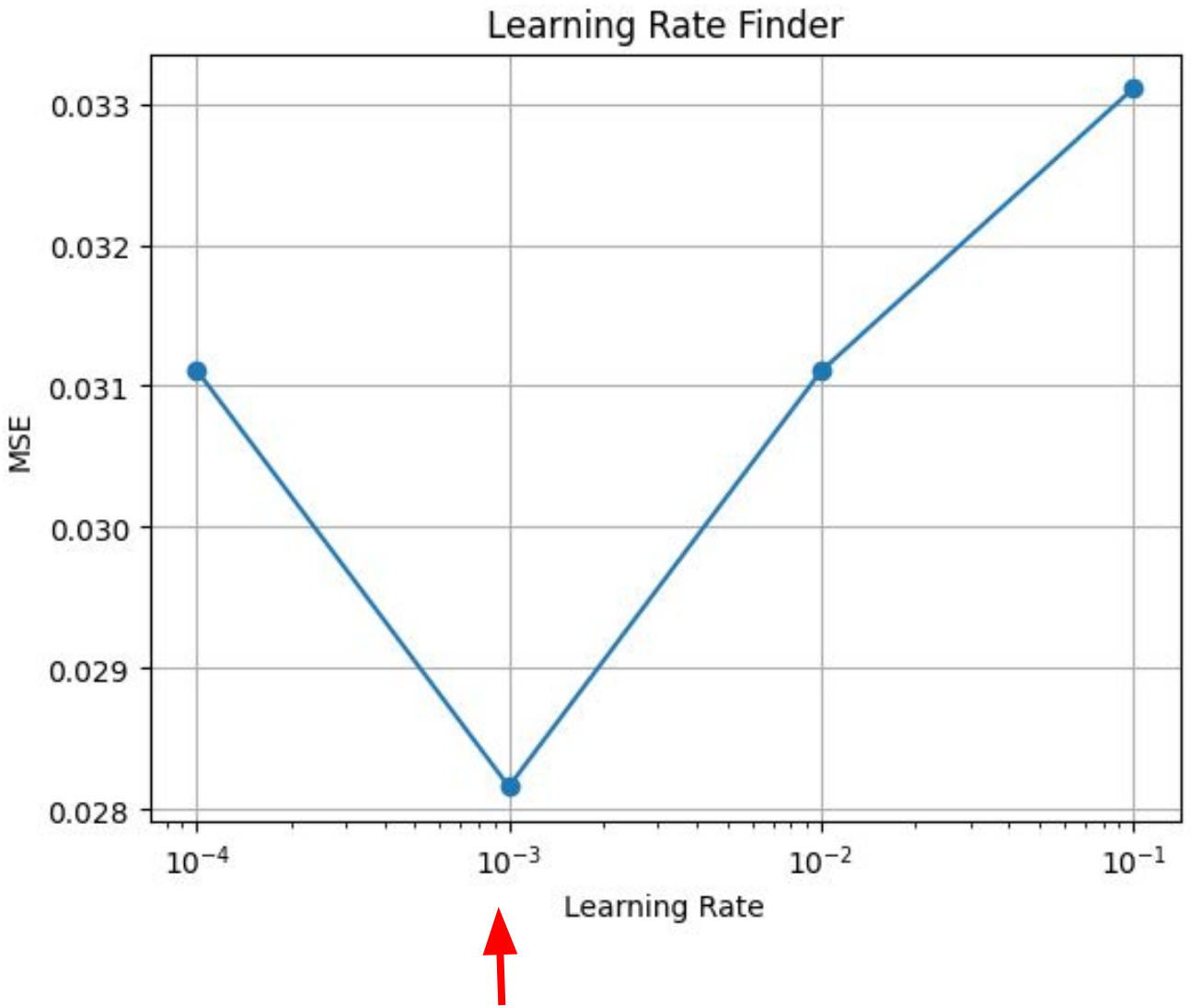
Score	Improvement (%)
R2 Score	6.40%
Accuracy	2.37%
RMSE	9.24%

Model Accuracies for Full Feature Set and PCA Features - Valence

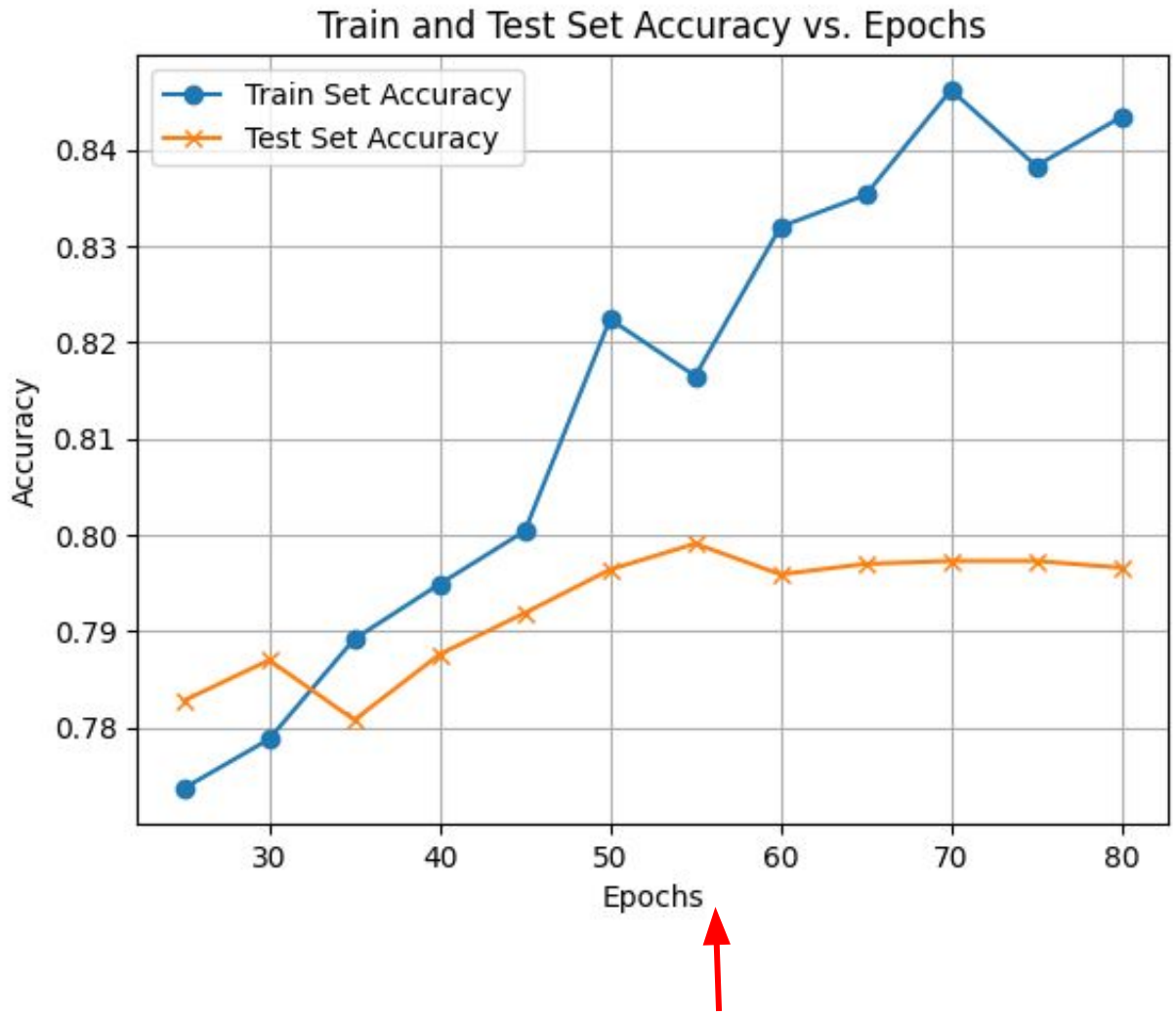


Score	Improvement (%)
R2 Score	5.84%
Accuracy	4.94%
RMSE	8.46%

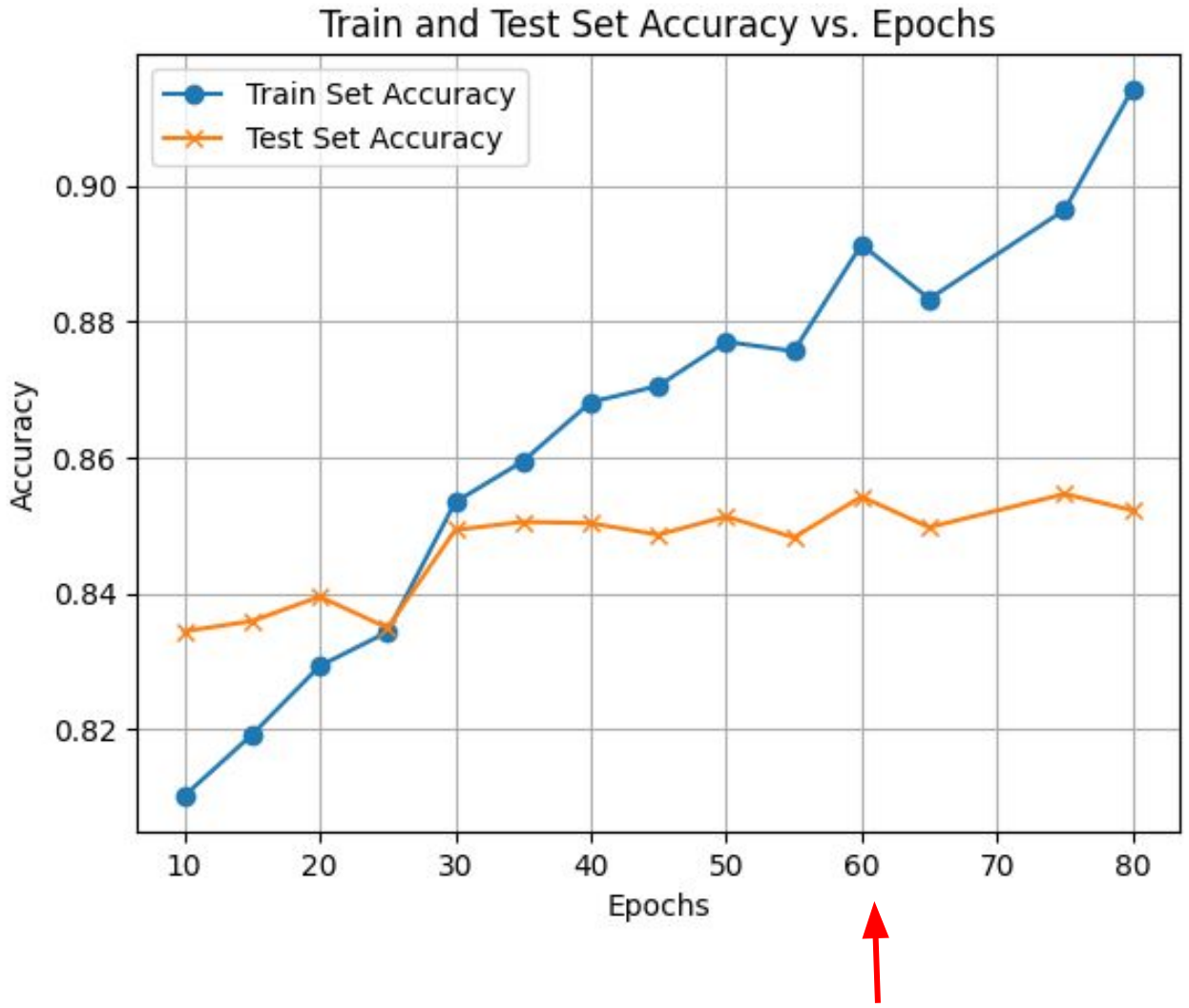
Parameter Tuning - Optimum Learning Rate



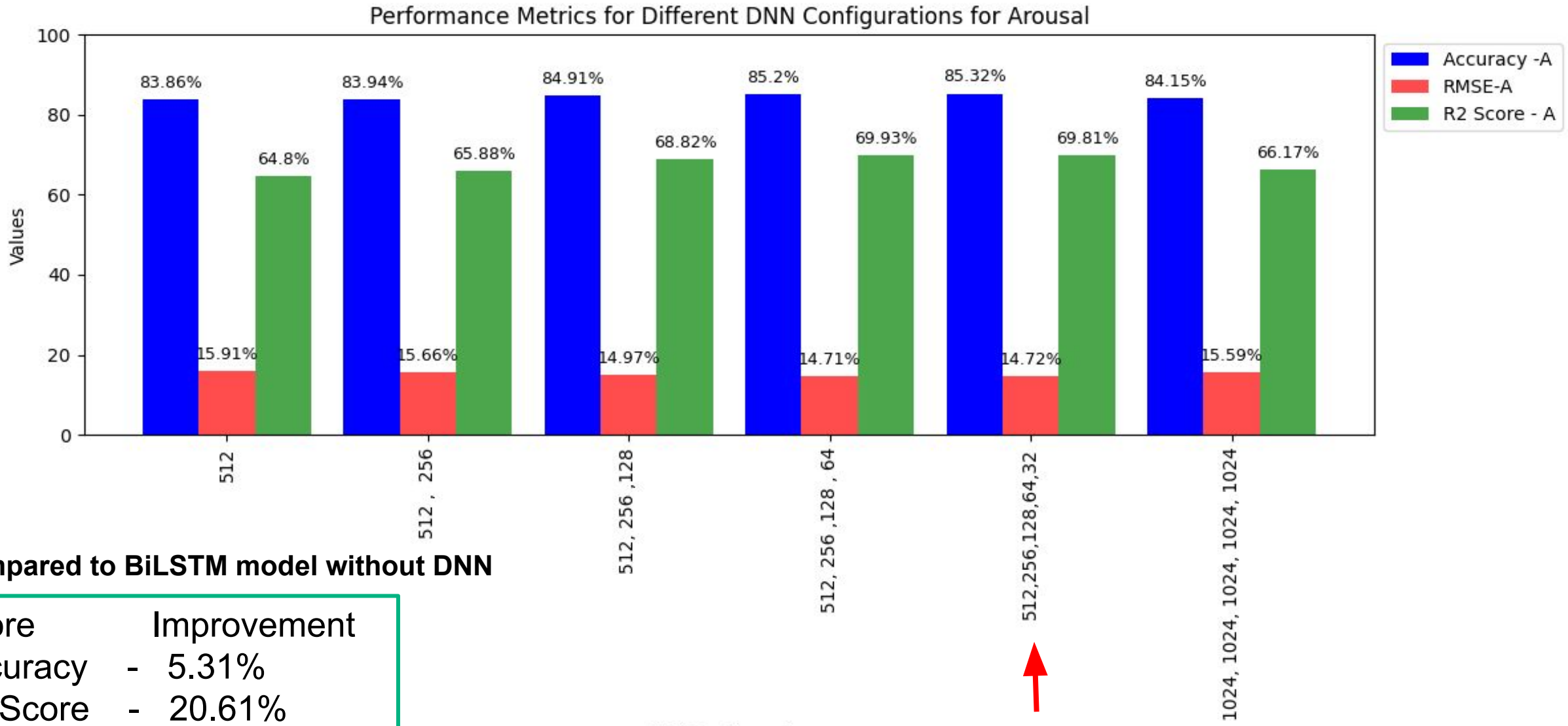
Parameter Tuning - Optimum Number of Epochs (Valence)



Parameter Tuning - Optimum Number of Epochs (Arousal)



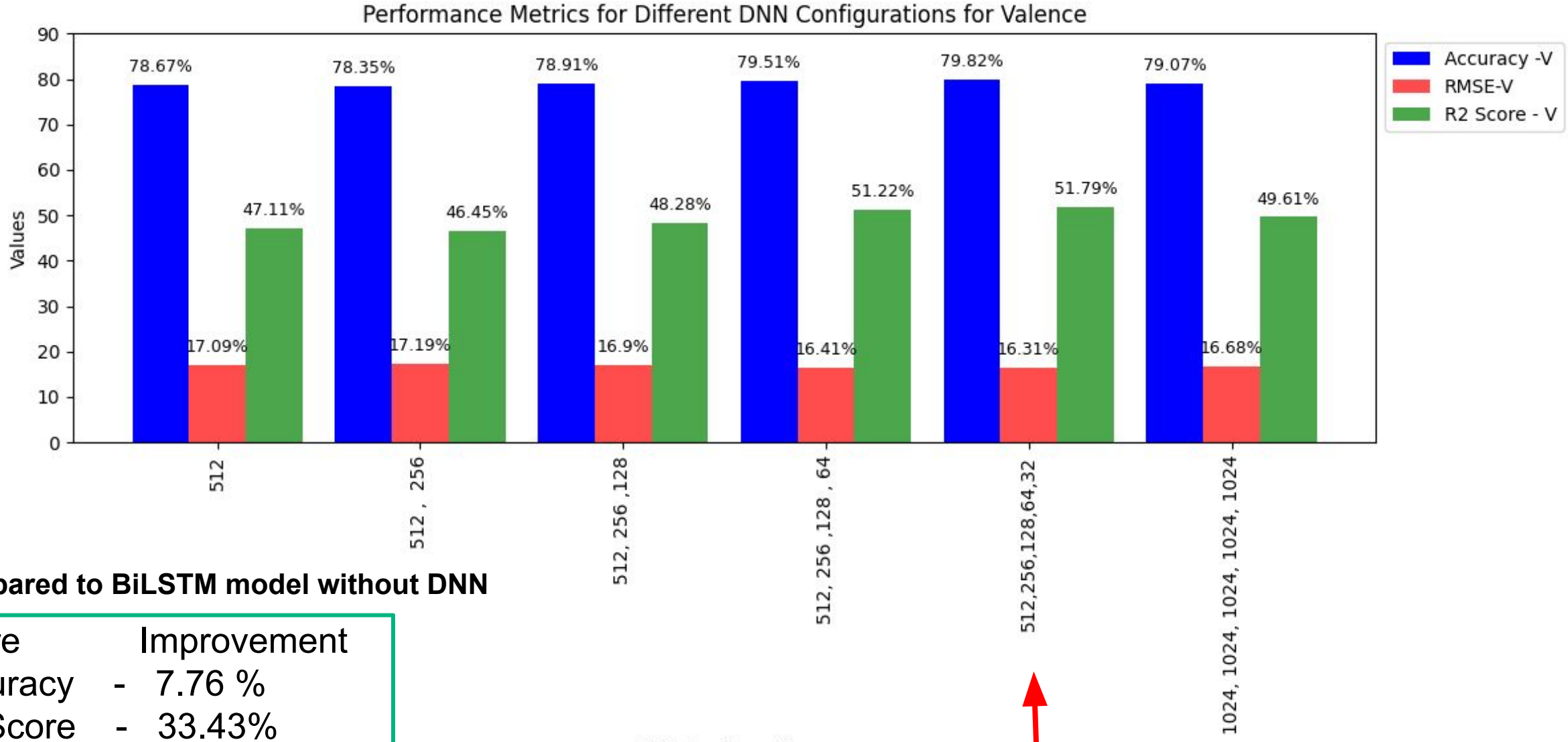
Different Dense Layer Combinations - Arousal



Compared to BiLSTM model without DNN

Score		Improvement
Accuracy	-	5.31%
R2 Score	-	20.61%
RMSE	-	20.6%

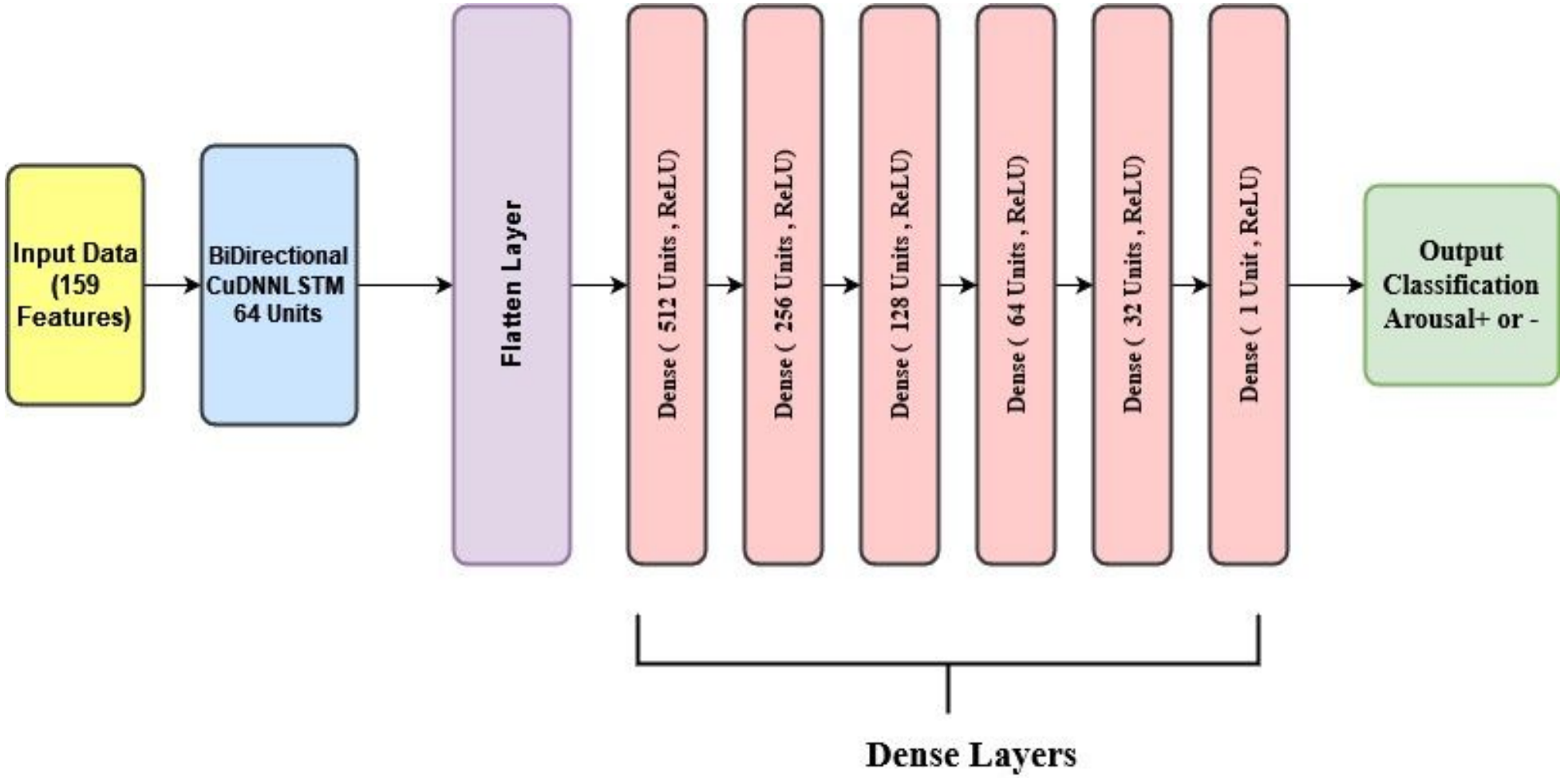
Different Dense Layer Combinations - Valence



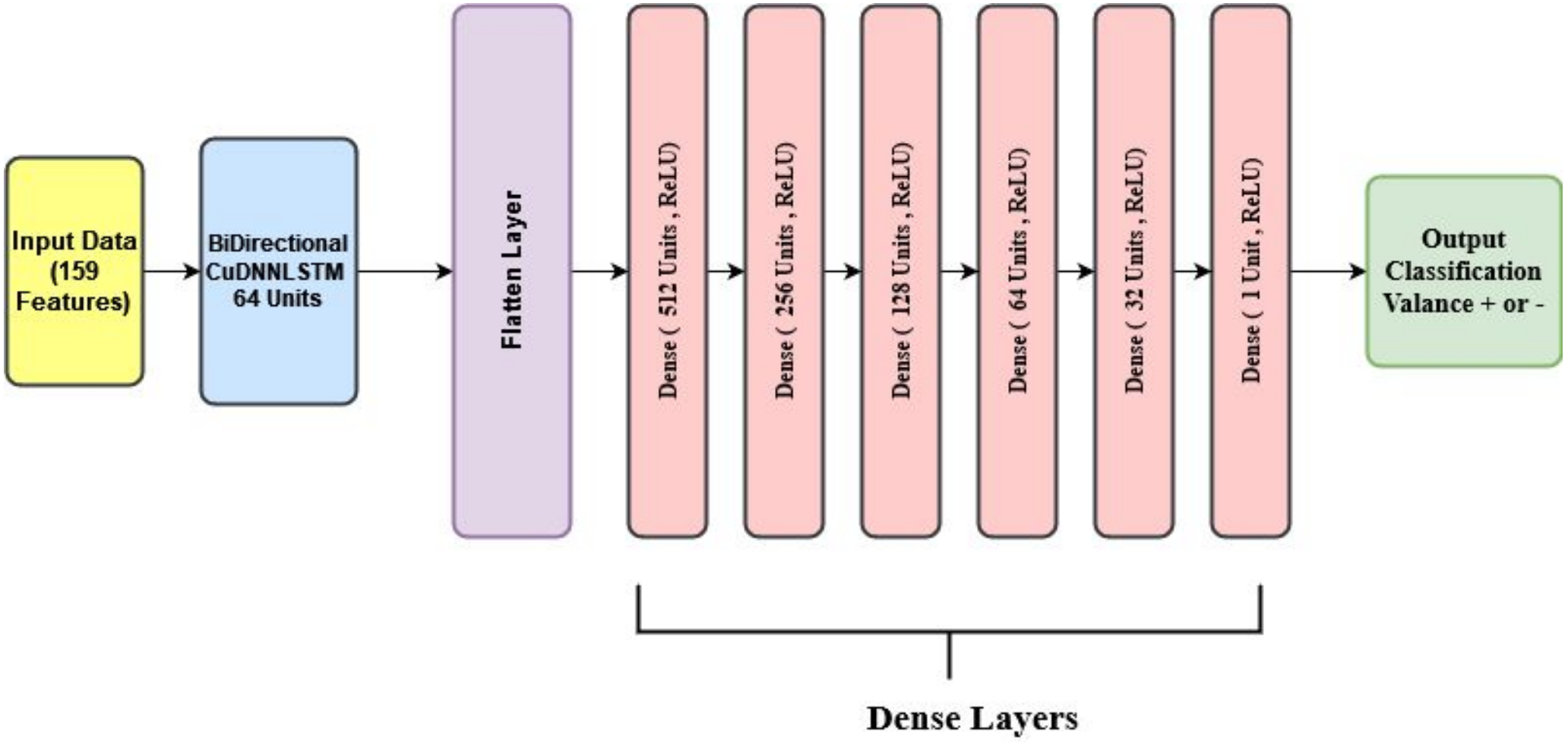
Compared to BiLSTM model without DNN

Score		Improvement
Accuracy	-	7.76 %
R2 Score	-	33.43%
RMSE	-	14.69 %

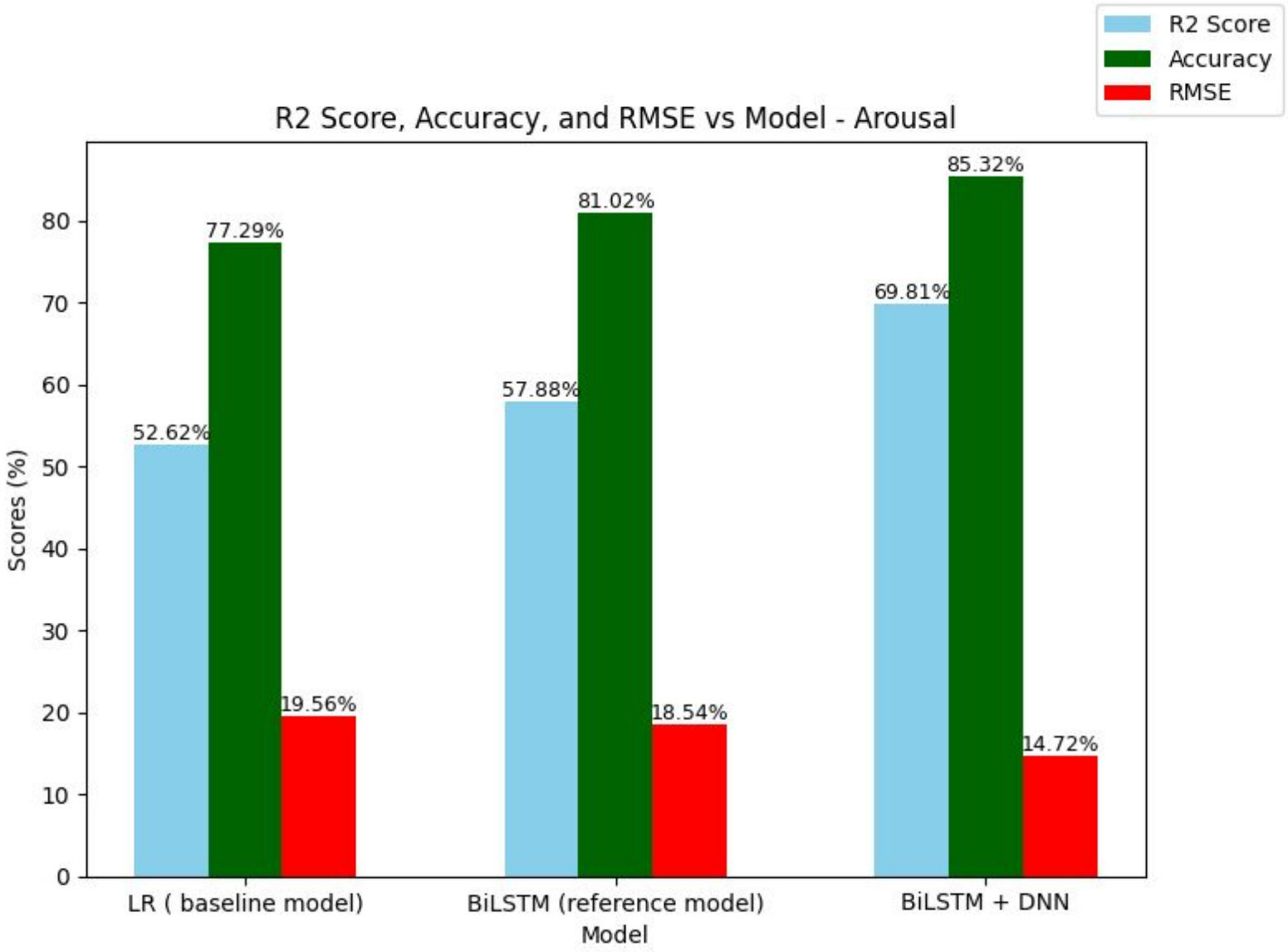
Arousal Model Architecture



Valance Model Architecture



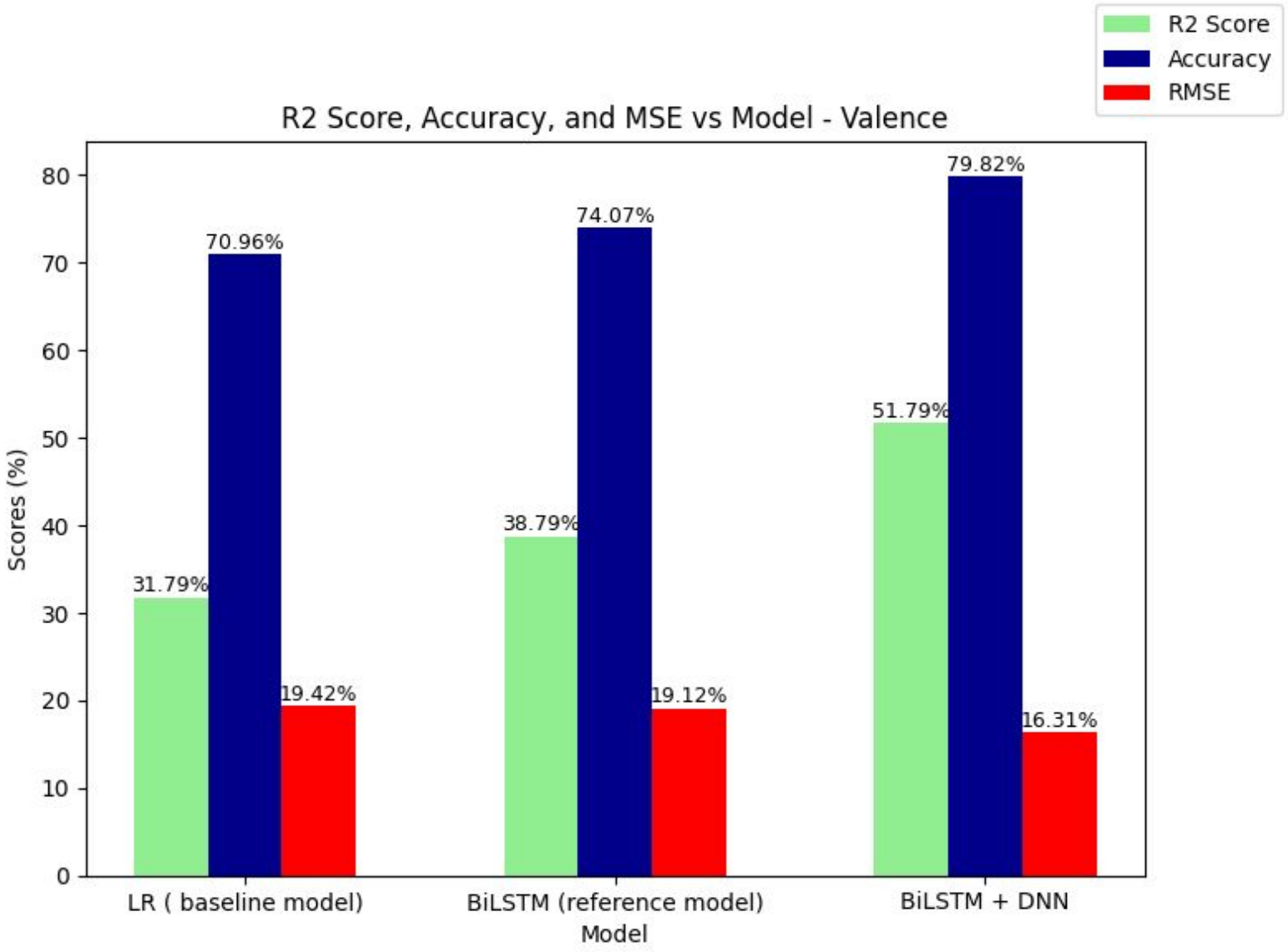
Reference Model , Baseline Model Vs Our Model (Arousal)



Compared to Baselin

Score	Im (%
R2 Score	
Accuracy	
RMSE	

Reference Model , Baseline Model Vs Our Model (Valance)



Compared to Baseli

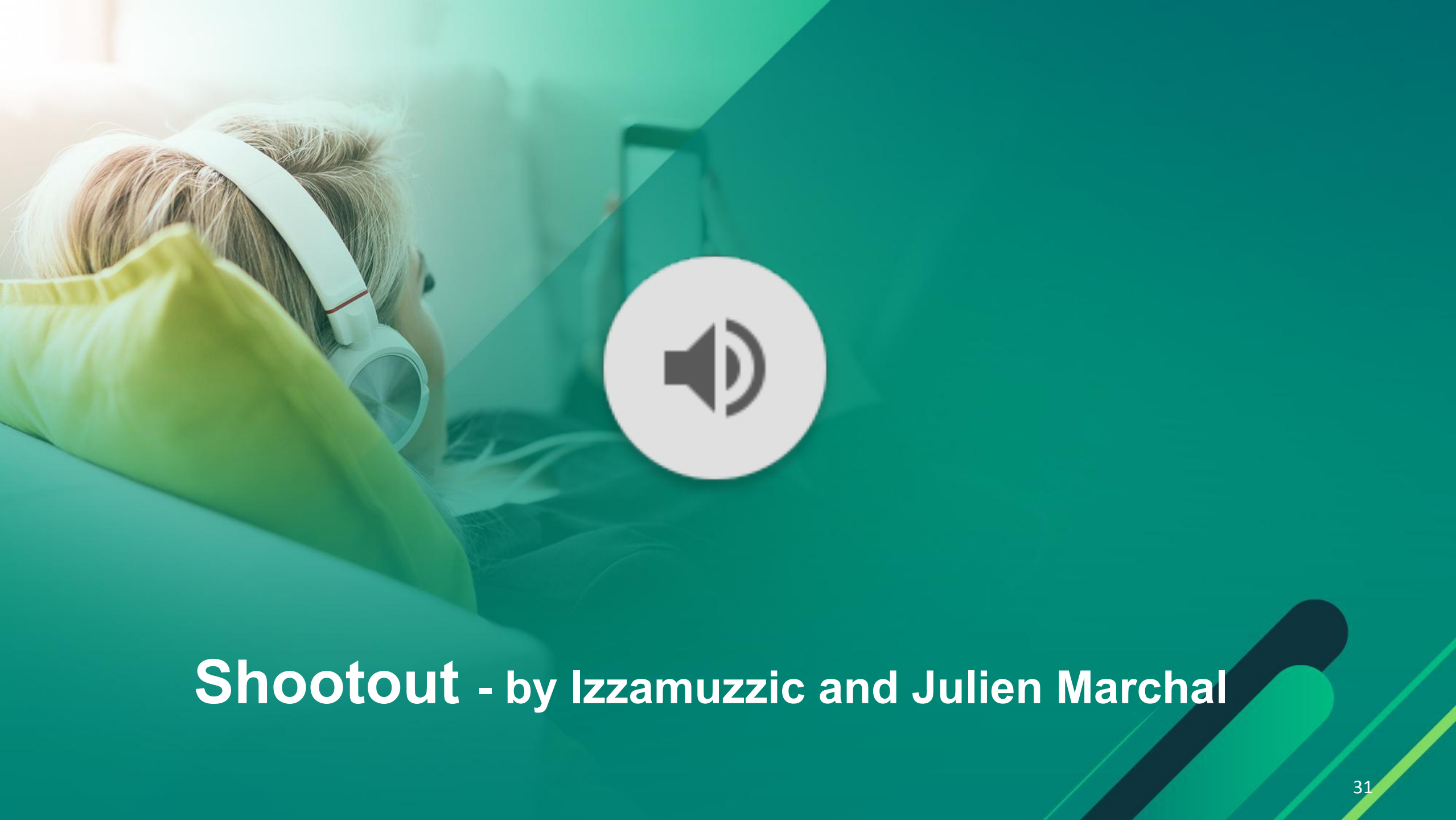
Score	Im (%
R2 Score	
Accuracy	
RMSE	

Findings

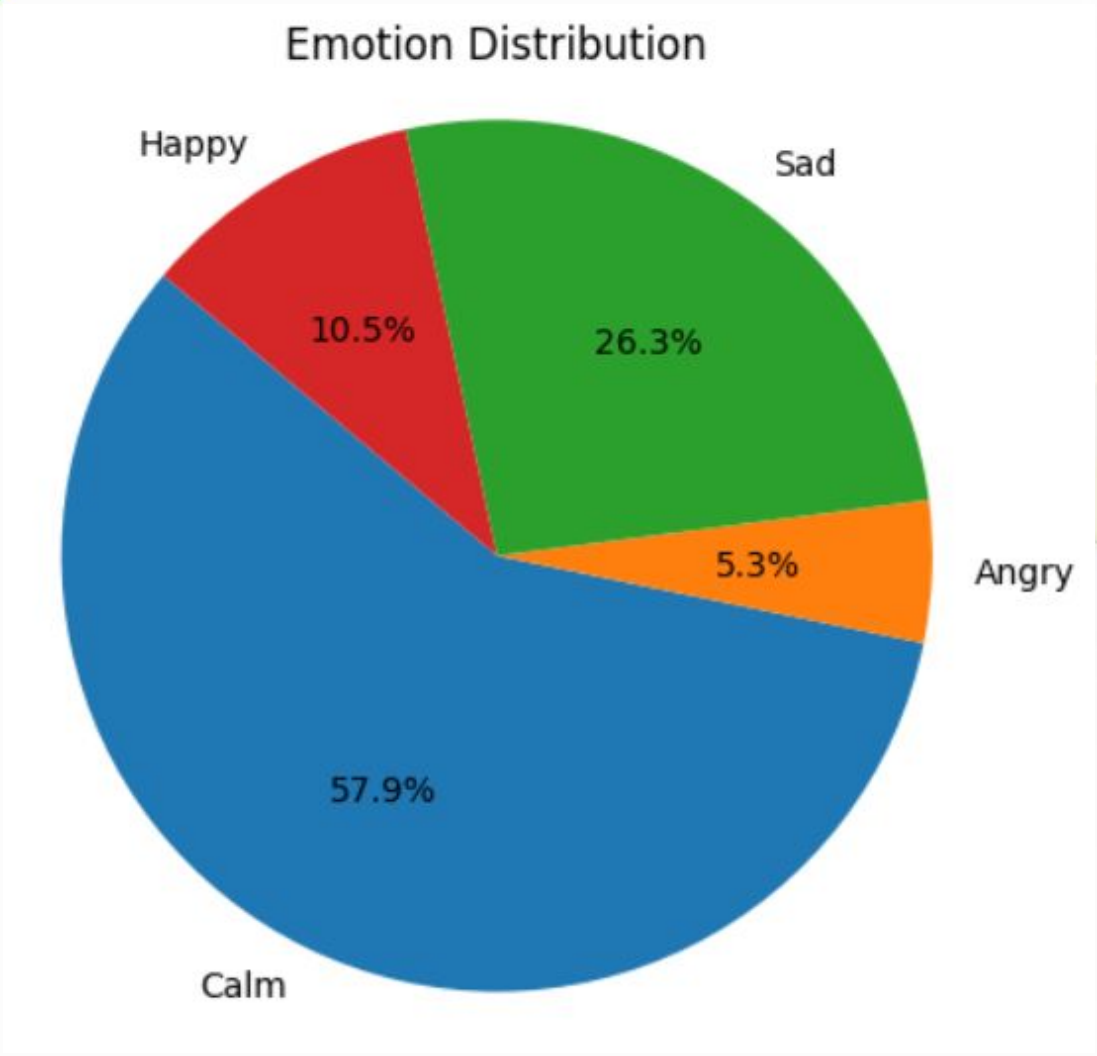
Findings...

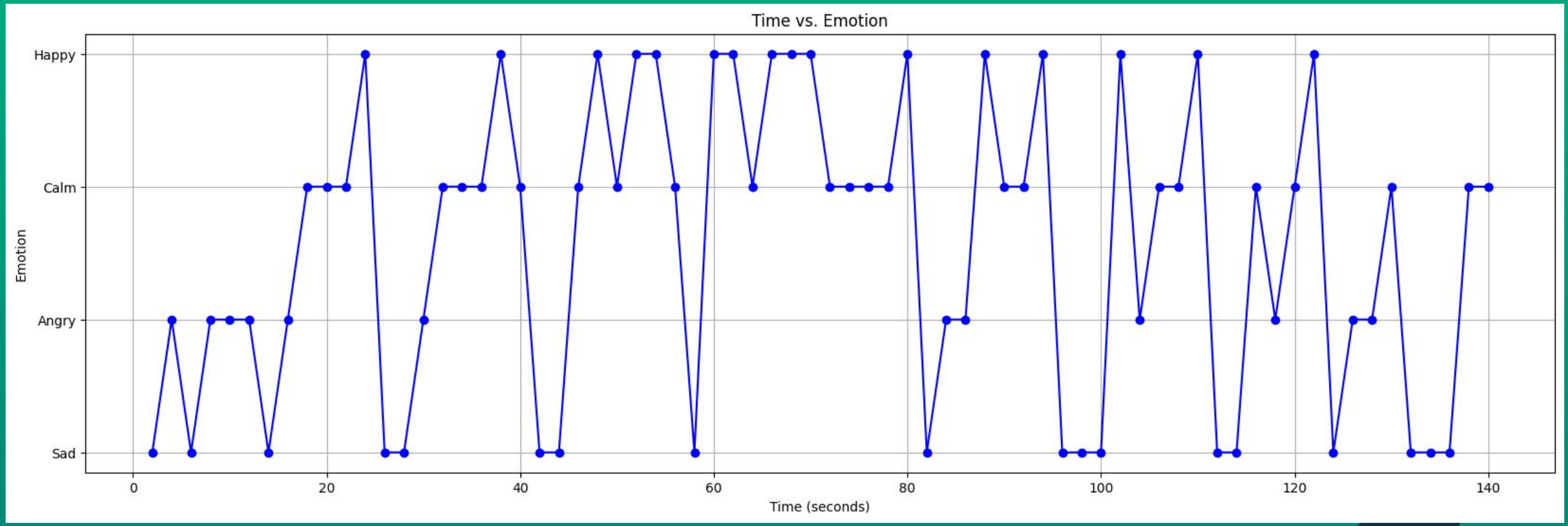
- Systems that use Deep Neural Network, have a higher accuracy compared to the systems that use only the traditional machine learning algorithms.
- Dynamic MER models are more accurate than the static MER models.
- The systems that have used hybrid models rather than a single model display more accuracy.

Demonstration



Shootout - by Izzamuzzic and Julien Marchal


Emotion Distribution
Happy
Sad
10.5%
26.3%
5.3%
Angry
57.9%
Calm


Time vs. Emotion
Happy
Calm
Angry
Sad
Emotion
0
20
40
60
80
100
120
140
Time (seconds)

Impact

Impacts	Dynamic MER	Static MER
• Enhanced Music Understanding	Recognizes changing emotions over the duration of a song.	Assigns emotions to specific moments, might not reflect the entire emotional journey.
• Improved Music Recommendation Systems	Recommendations based on real-time emotional changes.	Recommends based on isolated emotional moments, may not align with overall mood.
• Music Therapy	Adjusts the music as you listen to fit your changing emotions.	Labels emotions in chunks, might miss the continuous emotional flow of the entire song.

THANK YOU