

**Melody Informatics:
Computational Approaches to
Understanding the Relationships
Between Human Affective
Reasoning and Music**

Jessica Sharmin Rahman

A thesis submitted for the degree of
Doctor of Philosophy
The Australian National University

September 2022

© Jessica Sharmin Rahman 2021

Draft Copy – 30 September 2022

Except where otherwise indicated, this thesis is my own original work.

Jessica Sharmin Rahman
30 September 2022

To my parents, Hosne Jahan Begum and Maqbulur Rahman,
thank you for always being there for me

Acknowledgments

Throughout the journey of completing my Ph.D., I have received a great deal of support and encouragement from many people, whom I wish to acknowledge in this section. First, I would also like to extend my deepest gratitude to my primary supervisor Professor Tom Gedeon. He has given me a tremendous amount of support and encouragement throughout my research. Not only that, he has been extremely caring about my health and wellbeing during this time. Whether it is his hardworking attitude, fantastic research ideas, constructive criticism, or occasional lighthearted jokes – all of these have inspired me a lot. I am incredibly grateful for everything I learned from him. His valuable advice and suggestions will stay with me forever.

I want to express my sincere appreciation to my two other supervisors. My thanks to Dr. Sabrina Caldwell, who has been an incredible mentor to me. I am very thankful for her valuable feedback, constant motivation, and all the love and care she has given me. As a fellow woman in STEM, she led by example and encouraged me to inspire others with my work. I also want to thank Dr. Richard Jones for all his guidance and support. His knowledge of the subject matter and insightful facts on music were very useful. I am especially grateful for the kindness and empathy he showed me during the pandemic when I often became uncertain about my progress.

I would like to extend my thanks to Dr. Duncan Stevenson. His insights on the field of HCI have been beneficial in furthering my research. I am also thankful for his feedback on my thesis and his kind reminders to make sure I eat and rest as I write it. I want to thank Dr. Henry Gardner for his guidance, suggestions, and encouragement. I also wish to acknowledge the help and support provided by the technical staff and admin team of the School of Computing at ANU.

Thanks to my colleagues from ANU, who have been very supportive during my candidature. My sincere thanks to Zakir Hossain for the assistance he provided me, especially during the early stages of my Ph.D. I am very thankful to Xuanying Zhu, Nicholas Kuo, and Weiwei Hou for giving me so many research and career-related advice throughout the years. Many thanks to my colleagues Richa Sharma, Zi Jin, Wenbo Ge, Liyuan Zhou, Dehai Zhao, Atiqul Islam, Jieshan Chen, Elliot Catt, Amy Zhang, Guyver Fu, Yang Liu, Zhenyue Qin, and Yue Yao for all the lunchtime hangouts and fun chats about research life.

I am incredibly grateful to all my friends in Australia who have been a constant source of encouragement during this journey. Thanks to all my friends at Project

Beats Dance Studio, who always cheered me up. My special thanks to all my crew members during these years: Abigail Miranda, Anjlika Guglani, Anna McRae, Aradhya Sharma, Chaeyoung Kim, Erin Richman, Gabriella Pirintji, Holly Martin, Kendall Simmons, Larissa El-Khoury, Lucy O'Sullivan, Lucy Quinn, Maddie Ryan, Maree De Marco, Marisa O'Connell, Michelle Arizapa, Nicole Jones, Renee Su, Sarah Quinn, Taylor Manalo, Ushini Attanayake, Yaya Lu and Zohra Yasmeen. They always patiently listened to me whenever I had a stressful day at work and constantly gave me motivation and made me smile. Many thanks to my crew leaders, Chippy Lo and Malissa Huynh, for giving me lots of support and flexibility, which helped me continue my research and hobbies simultaneously. I extend my thanks to all my friends from the Bangladeshi community in Canberra for their support and assistance. My thanks also go to my housemates + friends Katrina Marshall and Zoe Connell, who spent hours motivating me whenever I went through a hard time.

I want to thank my family here in Australia, who have taken care of me over the last few years. I thank my uncle Abul Majumder and aunt Dilara Parvin for constantly checking up on my health and wellbeing during my studies. My uncle provided me with a lot of support and guidance, which helped me during my stay in Australia. My heartfelt thanks to my elder brother Rony Hasinur Rahman, my sister-in-law Nusrat Nowsheen, and my adorable nephew Fardis Nehad Rahman. Their love and support have been a huge source of happiness, and my nephew's smile can instantly lift my mood.

Finally, I would like to thank and dedicate this thesis to my wonderful parents, my mother Hosne Jahan Begum and father Maqbulur Rahman. Their endless love and support have been the biggest reason I was able to complete this journey. Their research and teaching career was my primary inspiration for undertaking this degree. And I was able to complete it, thanks to their constant encouragement and understanding. They taught me to dream big and chase those dreams, which is why I am here today. Mum and dad, this is for you.

Abstract

Music is a powerful and complex medium that allows people to express their emotions, while enhancing focus and creativity. It is a universal medium that can elicit strong emotion in people, regardless of their gender, age or cultural background. Music is all around us, whether it is in the sound of raindrops, birds chirping, or a popular song played as we walk along an aisle in a supermarket. Music can also significantly help us regain focus while doing a number of different tasks.

The relationship between music stimuli and humans has been of particular interest due to music's multifaceted effects on human brain and body. While music can have an anticonvulsant effect on people's bodily signals and act as a therapeutic stimulus, it can also have proconvulsant effects such as triggering epileptic seizures. It is also unclear what types of music can help to improve focus while doing other activities. Although studies have recognised the effects of music in human physiology, research has yet to systematically investigate the effects of different genres of music on human emotion, and how they correlate with their subjective and physiological responses.

The research set out in this thesis takes a human-centric computational approach to understanding how human affective (emotional) reasoning is influenced by sensory input, particularly music. Several user studies are designed in order to collect human physiological data while they interact with different stimuli. Physiological signals considered are: electrodermal activity (EDA), blood volume pulse (BVP), skin temperature (ST), pupil dilation (PD), electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS). Several computational approaches, including traditional machine learning approaches with a combination of feature selection methods are proposed which can effectively identify patterns from small to medium scale physiological feature sets. A novel data visualisation approach called "Gingerbread Animation" is proposed, which allows physiological signals to be converted into images that are compatible with transfer learning methods. A novel stacked ensemble based deep learning model is also proposed to analyse large-scale physiological datasets.

In the beginning of this research, two user studies were designed to collect physiological signals from people interacting with visual stimuli. The computational models showed high efficacy in detecting people's emotional reactions. The results provided motivation to design a third user study, where these visual stimuli were combined with music stimuli. The results from the study showed decline in recognition accuracy comparing to the previous study. These three studies also gave a

key insight that people's physiological response provide a stronger indicator of their emotional state, compared with their verbal statements.

Based on the outcomes of the first three user studies, three more user studies were carried out to look into people's physiological responses to music stimuli alone. Three different music genres were investigated: classical, instrumental and pop music. Results from the studies showed that human emotion has a strong correlation with different types of music, and these can be computationally identified using their physiological response.

Findings from this research could provide motivation to create advanced wearable technologies such as smartwatches or smart headphones that could provide personalised music recommendation based on an individual's physiological state. The computational approaches can be used to distinguish music based on their positive or negative effect on human mental health. The work can enhance existing music therapy techniques and lead to improvements in various medical and affective computing research.

List of Publications

The following publications are based on the research work reported in this thesis. I was the primary contributor in all of these publications. All of the publications are peer reviewed or under consideration in a peer-reviewed journal.

1. **Rahman, J.S.**, Gedeon, T., Caldwell, S., Jones, R., Hossain, M.Z. and Zhu, X., 2019, July. Melodious micro-frissons: detecting music genres from skin response. In *2019 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.
2. **Rahman, J.S.**, Hossain, M.Z. and Gedeon, T., 2019, December. Measuring Observers' EDA Responses to Emotional Videos. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction* (pp. 457-461).
3. **Rahman, J.S.**, Gedeon, T., Caldwell, S. and Jones, R., 2020, July. Brain melody informatics: Analysing effects of music on brainwave patterns. In *2020 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.
4. **Rahman, J.S.**, Hossain, M.Z. and Gedeon, T., 2020, December. Are paired or single stimuli better to recognize genuine and posed smiles from observers' galvanic skin response?. In *32nd Australian Conference on Human-Computer Interaction* (pp. 661-665).
5. **Rahman, J.S.**, Gedeon, T., Caldwell, S.B., Jones, R. and Jin, Z., 2021. Towards Effective Music Therapy for Mental Health Care Using Machine Learning Tools: Human Affective Reasoning and Music Genres. *J. Artif. Intell. Soft Comput. Res.*, 11(1), pp.5-20.
6. **Rahman, J.S.**, Gedeon, T., Caldwell, S. and Jones, R.L., 2021, May. Can Binaural Beats Increase Your Focus? Exploring the Effects of Music in Participants' Conscious and Brain Activity Responses. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-6).

7. **Rahman, J.S.**, Caldwell, S.B., Jones, R. and Gedeon, T., Brain Melody Interaction: Understanding Effects of Music on Cerebral Hemodynamic Responses. *Multimodal Technologies and Interaction*, 6(5), p.35.

The following publications are based on the extension of the research work set out in this thesis. I was one of the contributors in all of the publications. Some of these publications used the datasets I created for the studies of this research, while some of the publications extend some of the techniques proposed in the research. All of the publications are peer reviewed or under consideration in a peer-reviewed journal or conference venue. Please note that research work conducted in these publications do not appear in this thesis.

8. Brewer, M. and **Rahman, J.S.**, 2020, November. Pruning Long Short Term Memory Networks and Convolutional Neural Networks for Music Emotion Recognition. In *International Conference on Neural Information Processing* (pp. 343-352). Springer, Cham.
9. Renkin, M. and **Rahman, J.S.**, 2020, November. Improving the Stability of a Convolutional Neural Network Time-Series Classifier Using SeLU and Tanh. In *International Conference on Neural Information Processing* (pp. 788-795). Springer, Cham.
10. Rostov, M., Hossain, M.Z. and **Rahman, J.S.**, 2021, August. Robotic Emotion Monitoring for Mental Health Applications: Preliminary Outcomes of a Survey. In *IFIP Conference on Human-Computer Interaction* (pp. 481-485). Springer, Cham.
11. Chu, R., **Rahman, J.S.**, Caldwell, S., Zhu, X., and Gedeon, T., 2021. Detecting Lies: Finding the Degree of Falsehood from Observers' Physiological Responses. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 1959-1965). IEEE.
12. Pan, E. and **Rahman, J.S.**, 2021. Explainable Neural Networks in Stress Recognition using EEG Signals and Genetic Algorithm based Feature Selection. In *2021 International Conference on Neural Information Processing* (pp. 136-143). Springer, Cham.
13. Li, X., **Rahman, J.S.** and Hossain, M.Z., Distinguishing Between Real and Posed Smiles from Observers' Accelerometer Data. In *2021 Electrical Information and Communication Technology (EICT)* (pp. 1-5). IEEE.

-
14. Rostov, M., Khan, A.R., **Rahman, J.S.**, Ahmed, K.A. and Hossain, M.Z., Robotic ML models for monitoring mental health: A Systematic Review. *International Journal of Human-Computer Studies* (Under Review).

Contents

Acknowledgments	vii
Abstract	ix
List of Publications	xi
List of Figures	xxiii
List of Tables	xxvi
List of Abbreviations	xxvii
1 Introduction	1
1.1 Motivation	1
1.2 Research Questions and Objectives	4
1.3 Thesis Outline	5
2 Background and Related Work	9
2.1 Emotion Theory and Emotion Model	9
2.2 Measures to Understand Emotion	11
2.2.1 Subjective Measures	12
2.2.2 Physical Measures	12
2.2.2.1 Facial Expression	12
2.2.2.2 Gestures	13
2.2.2.3 Speech	13
2.2.3 Physiological Measures	13
2.2.3.1 Electrodermal Activity	14
2.2.3.2 Blood Volume Pulse	15
2.2.3.3 Heart Rate Variability	16
2.2.3.4 Skin Temperature	16
2.2.3.5 Electroencephalogram	17
2.2.3.6 Functional Near-infrared Spectroscopy	18
2.2.3.7 Pupillary Response	20
2.2.3.8 Other Signals	21
2.2.4 Research on Music and Physiological Signals	21
2.2.4.1 Music as Therapy	22
2.2.4.2 Different Measures for Music Emotion Recognition	23
2.2.4.3 Physiological Signals Based Emotion Recognition	24

2.3	Summary	24
3	Computational Methods to Analyse Human Physiological Signals	27
3.1	Stimuli	27
3.1.1	Music Datasets Used in the Literature	27
3.1.2	Limitations of the Available Datasets	28
3.1.3	Stimuli Used in the Experiments	29
3.1.3.1	Music	29
3.1.3.2	Video	30
3.2	Computational Methods	31
3.2.1	Experiment Design and Data Collection	31
3.2.2	Data Pre-processing	33
3.2.2.1	Filtering	33
3.2.2.2	Normalisation	34
3.2.2.3	Baseline Correction	35
3.2.3	Feature Extraction	35
3.2.4	Feature Selection	37
3.2.4.1	Feature Ranking Methods	37
3.2.4.2	Feature Subset Methods	37
3.2.5	Classification Methods	38
3.2.5.1	K-Nearest Neighbour	38
3.2.5.2	Decision Trees	38
3.2.5.3	Random Forest	39
3.2.5.4	Support Vector Machines	39
3.2.5.5	Bayesian Classifier	39
3.2.5.6	Artificial Neural Network	40
3.2.5.7	Convolutional Neural Networks	40
3.2.5.8	Recurrent Neural Network	41
3.2.5.9	Long Short-term Memory	41
3.2.5.10	Transfer Learning	42
3.3	Statistical Analysis	42
3.4	Evaluation Measures	43
3.5	Summary	45
4	Effects of Different Stimuli in Human Physiological Response	47
4.1	Experiment 1 - Physiological Responses to Emotion Videos	47
4.1.1	Methods	47
4.1.1.1	Participants	47
4.1.1.2	Dataset and Pre-processing	48
4.1.1.3	Features	48
4.1.2	Results	48
4.1.2.1	Mean Analysis	48
4.1.2.2	Classification Results	52
4.1.3	Discussion	53

4.2	Experiment 2 - Physiological Responses Detecting Genuine and Posed Smiles	54
4.2.1	Methods	54
4.2.1.1	Participants	54
4.2.1.2	Dataset and Pre-processing	54
4.2.1.3	Features	55
4.2.2	Results	56
4.2.2.1	Classification Results	56
4.2.2.2	Timeline Analysis	57
4.2.2.3	Comparison with Participants' Verbal Response	59
4.2.3	Discussion	59
4.3	Experiment 3 - Effects of Music in Detecting Genuine and Acted Emotions	60
4.3.1	Methods	60
4.3.1.1	Participants	62
4.3.1.2	Experiment Design	62
4.3.2	Results	64
4.3.2.1	Findings on Music Stimuli	64
4.3.2.2	Verbal and Physiological Data Analysis	66
4.3.3	Discussion	68
4.4	Summary	69
5	Effects of Music in Physiological Response	71
5.1	Experiment Design	71
5.2	Participants	73
5.3	Data Analysis	76
5.3.1	Pre-processing	76
5.3.2	Feature Extraction and Selection	76
5.3.3	Visualisation of the Physiological Signals - Gingerbread Animation	77
5.3.4	Classifiers	79
5.3.4.1	Preliminary analysis using EDA data	79
5.3.4.2	Analysis Using EDA, BVP, ST and PD Data	81
5.4	Results and Discussion	82
5.4.1	Results using EDA signals	82
5.4.2	Results Using EDA, BVP, ST and PD Signals	87
5.4.2.1	Neural Network Performs Best Among All Classifiers	87
5.4.2.2	Feature Selection Produces Best Results for Music Genre Classification	88
5.4.2.3	Statistical Analysis on All Evaluation Measures	90
5.4.2.4	Top Features Selected by Feature Selection Methods	91
5.4.2.5	Best Feature Selection Methods	92
5.4.2.6	Effectiveness of Visualisation	93
5.5	Summary	94

6	Effects of Music on Brainwave Patterns	95
6.1	Experiment Design	95
6.2	Data Analysis	96
6.2.1	Pre-processing	96
6.2.2	Feature Extraction and Selection	96
6.2.3	Classifiers	97
6.3	Results and Discussion	98
6.3.1	Statistical Analysis	98
6.3.2	Best Features	98
6.3.3	Classification Results	100
6.3.4	Verbal Response Analysis	103
6.3.5	Observation of Gamma Levels	106
6.4	Summary	107
7	Effects of Music on Cerebral Hemodynamic Response	109
7.1	Experiment Design	109
7.2	Participants	111
7.3	Data Analysis	112
7.3.1	Pre-processing	112
7.3.2	Feature Extraction	112
7.3.3	Classifiers	112
7.3.3.1	Stacked Ensemble Models	113
7.4	Results and Discussion	115
7.4.1	Classification Results	116
7.4.2	Visual Analysis	119
7.4.3	Music Offset Analysis	122
7.4.4	Verbal Response Analysis	123
7.4.5	Activation Map Analysis	125
7.5	Comparison Between Two Types of Brain Data	127
7.6	Summary	128
8	Conclusion	129
8.1	Summary of Contributions	129
8.1.1	Answering the Research Questions	130
8.1.1.1	RQ1: Do Different Types of Music Generate Different Physiological Response?	131
8.1.1.2	RQ2: Do Other Stimuli (e.g. images, videos) Have Any Impact on These Physiological Responses?	132
8.1.1.3	RQ3: Can Other Stimuli (e.g. images, videos) be Com- bined with Music to Understand Their Combined Ef- fects on Physiological Responses?	132
8.1.1.4	RQ4: Which Physiological Signals Perform Better in Differentiating the Physiological Responses to Differ- ent Stimuli?	133

8.1.1.5	RQ5: Which Computational Methods are Effective to Analyse Physiological Response to Music and Other Stimuli?	134
8.1.2	Applications of the Work	135
8.1.2.1	Personalised Music Recommendations	135
8.1.2.2	Music Therapy and Biofeedback Training	135
8.1.2.3	Portable Device Creation	136
8.1.2.4	Musicogenic Epilepsy Detection and Reduction	137
8.2	Key Limitations of the Work	137
8.2.1	Experiment Design Issues	137
8.2.2	Sample Size	138
8.2.3	Population Bias	138
8.3	Future Work	139
8.3.1	Experiments Conducted In-the-wild	139
8.3.2	Computational Approaches to Detect Emotional Response to Music in Real Time	139
8.3.3	Optimisation Methods to Find Appropriate Hyperparameters	139
8.3.4	Virtual Reality Applications	140
8.3.5	Edge Computing Applications	140
A	Experiment Procedure and General Guidelines	141
B	Experiment Related Documents	143
B.1	Participant Information Sheet	144
B.2	Participant Consent Form	147
B.3	SONA Experiment Sign Up Page	148
B.4	Experiment Steps and Questionnaire	150
B.4.1	Introduction and Instruction Page	150
B.4.2	Pre-experiment Questionnaire Page	150
B.4.3	Playing and Experiment Questionnaire Page	152
B.4.4	Post experiment questionnaire page	153
C	Miscellaneous Materials	155
C.1	External Links for Experimental Materials	155
C.2	Sample Video Link for Gingerbread Animation	155

List of Figures

2.1	Plutchik’s wheel of emotions	10
2.2	Russel’s circumplex model of affect	11
2.3	Empatica E4 device [E4]	14
2.4	NeXus BVP [NeXus BVP]	15
2.5	NeuLog Heart Rate and Pulse logger sensor [NUL-208]	16
2.6	Raw signals collected using Empatica E4 device [E4]	17
2.7	Emotiv EPOC+ device [Emotiv]	18
2.8	Obelab NIRSIT device	19
2.9	Nirsit device channel locations at 30 mm separation	20
2.10	The Eye Tribe device	21
3.1	Summary of methods used in the computational analysis	32
4.1	Mean values of EDA for seven emotion categories (Range 0.3 – 0.5 chosen for better visualisation)	49
4.2	Arousal models of emotion: Standard abstract model	50
4.3	Arousal models of emotion: Data derived model (Neutral as reference)	51
4.4	Arousal models of emotion: Data derived model (Mean as reference)	51
4.5	Classification performance while participants recognised seven emotions from video (Range 90 – 100 displayed for better visualisation)	52
4.6	Sample frames from UvA-NEMO database	55
4.7	Timeline analysis of EDA signals - single condition	58
4.8	Timeline analysis of EDA signals - paired condition	58
4.9	A photo of the experimental setting - participant is listening to different music while watching genuine and acted emotion	62
4.10	Classification results using participants verbal and physiological response in detecting genuine and acted emotions while listening to different music	67
5.1	User interface of The Eye Tribe for calibration process	72
5.2	Two dimensional emotion model by valence and arousal	74
5.3	Experimental setting - participants physiological signals being collected while they listen to music	74
5.4	Physiological signals representation as a graph (Blue = BVP, Red = EDA, Green = ST)	78
5.5	Physiological signals representation in an animation (Red = BVP, Blue = EDA, Green = PD, Grey = ST)	79

5.6	Neural network architecture	80
5.7	Classification accuracy using different hidden node number	81
5.8	Resnet18 architecture	83
5.9	CNN architecture for Gingerbread Animation	84
5.10	Classification based on music genres using EDA signals	84
5.11	Classification results based on subjective rating (<i>depressing</i> → <i>neutral</i> → <i>exciting</i>) using EDA signals	85
5.12	Classification results based on subjective rating (<i>disturbing</i> → <i>neutral</i> → <i>comforting</i>) using EDA signals	86
5.13	Classification result based on subjective rating (<i>tensing</i> → <i>relaxing</i>) using EDA, BVP, ST and PD signals	88
5.14	Neural Network classification result based on music genre using EDA, BVP, ST and PD signals	90
6.1	Emotiv headset channel location and names [Balasubramanian et al., 2018]	96
6.2	Classification results using EEG features based on subjective rating (<i>tensing</i> → <i>neutral</i> → <i>relaxing</i>), range 40-100 chosen for better visualisation	100
6.3	Classification results based on three music genres using EEG features, range 75-100 chosen for better visualisation	101
6.4	Classification accuracy using EEG signals based on participants' subjective response based on three emotion scales, range 84-100 chosen for better visualisation	102
6.5	Word cloud on verbal comments provided by the participants on the twelve music stimuli	103
7.1	Experimental setting - participants fNIRS signals being collected while they listen to music	110
7.2	1D CNN architecture for fNIRS signals classification	114
7.3	Stacked ensemble model architecture for fNIRS signals classification	115
7.4	Classification results using fNIRS signals by grouping participants into three categories (range 40 – 80 chosen for better visualisation)	119
7.5	Timeline analysis of participants HbO2 response to three music genres: points 900 - 1000	120
7.6	Timeline analysis of participants HbO2 response to three music genres: points 2500 - 2600	121
7.7	Classification accuracy with fNIRS signals using different offset length	122
7.8	Word cloud on verbal comments provided by the participants	124
7.9	Frame from activation map video showing changes in HBO2	126
7.10	Frame 417 of P3 listening to Instrumental 1 and Pop 4	126
7.11	Frame 417 of P17 listening to Instrumental 1 and Pop 4	127
B.1	Experiment introduction page	150
B.2	Pre-experiment questionnaire page	151

B.3	Experiment video playing page	152
B.4	Experiment video questionnaire page	152
B.5	Experiment music questionnaire page	153
B.6	Post-experiment questionnaire page	153

List of Tables

3.1	Music stimuli used in the experiment	30
3.2	List of features for physiological signal analysis	36
4.1	Features extracted from participants EDA Signals watching emotional videos	48
4.2	T-test values for all pairs of emotions in identifying seven types of emotional videos	49
4.3	Evaluation measures for classifying seven emotional categories from videos	53
4.4	Features from EDA signals watching genuine and posed smile stimuli .	56
4.5	Classification accuracy differentiating genuine and posed smiles using EDA signals	57
4.6	Type of comments provided by participants on each music stimuli . . .	64
5.1	Participant demographic of the experiment of collecting physiological signals during music listening	75
5.2	Features extracted from participants physiological signals while listening to three genres of music	77
5.3	Classification based on music genres using EDA signals	82
5.4	Classification results based on subjective rating (<i>depressing</i> → <i>neutral</i> → <i>exciting</i>) using EDA signals	86
5.5	Classification results based on subjective rating (<i>disturbing</i> → <i>neutral</i> → <i>comforting</i>) using EDA signals	87
5.6	Classification results based on music genre using EDA, BVP, ST and PD signals	89
5.7	Significance values for all pairs of feature selection methods	91
5.8	Top 12 features (from EDA, BVP, ST and PD signals) selected by all methods	91
6.1	Emotiv channel names, locations and extracted feature list from EEG signals	97
6.2	Top 25 EEG features selected by feature selection methods	99
6.3	Evaluation measures of participants' subjective response using EEG signals based on three emotion scales	102
6.4	Comments provided by participants on the twelve music stimuli	104
7.1	Participant demographic of the experiment exploring the effects of hemodynamic response	111

7.2 Features extracted from fNIRS signals 112

7.3 Evaluation measure results of KNN, RF and 1D CNN using fNIRS
signals 117

7.4 Type of comments provided by participants on each music stimuli . . . 125

List of Abbreviations

1D CNN	One-dimensional Convolutional Neural Network
2D CNN	Two-dimensional Convolutional Neural Network
ACC	Accelerometer
AFEW	Acted Facial Expression In The Wild
AI	Artificial Intelligence
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
ANS	Autonomic Nervous System
ANU	Australian National University
AUC	Area Under Curve
BC	Bayesian Classifier
BCI	Brain-Computer Interfaces
BP	Blood Pressure
BT	Bagged Trees
BVP	Blood Volume Pulse
CNN	Convolutional Neural Network
CONV	Convolutional
DBN	Deep Belief Network

List of Abbreviations

DEAP	Database for Emotion Analysis Using Physiological Signals
DFA	Detrended Fluctuation Analysis
DNN	Deep Neural Networks
DT	Decision Trees
ECG	Electrocardiogram
EDA	Electrodermal Activity
EEG	Electroencephalography
EMG	Electromyography
FC	Fully Connected
FN	False Negative
FNIRS	Functional Near-Infrared Spectroscopy
FP	False Positive
FT	Fourier Transformation
GA	Genetic Algorithm
GSR	Galvanic Skin Response
HbO₂	Oxygenated Hemoglobin
HbR	Deoxygenated Hemoglobin
HbT	Total Hemoglobin
HCI	Human – Computer Interaction
HR	Heart Rate
HRV	Heart Rate Variability
IAPS	International Affective Picture System

List of Abbreviations

IBI	Inter-beat Interval
KNN	K-nearest Neighbor
KS	Kolmogorov-Smirnov
LSTM	Long Short-term Memory Network
MAHNOB	Multimodal Analysis of Human Nonverbal Behaviour in Real World Settings
MER	Music Emotion Recognition
MI	Mutual Information
MMI	M&M Initiative
MRMR	Minimal-Redundancy-Maximum-Relevance
NN	Neural Network
PCA	Principal Component Analysis
PD	Pupil Dilation
PI	Paired Image
PMemo	Dataset for Music Emotion Computing
PPG	Photoplethysmography
PPV	Positive Predictive Value
PSO	Particle Swarm Optimization
PV	Paired Video
QDC	Quadratic Discriminant Classifier
RCT	Randomised Controlled Trial
RF	Random Forest
RNN	Recurrent Neural Networks
RQ	Research Questions
RR	Respiration Rate

List of Abbreviations

RSFS	Random Subset Feature Selection
SC	Skin Conductance
SCL	Skin Conductance Level
SCR	Skin Conductance Response
SD	Statistical Dependency
SDNN	Standard Deviation of Normal to Normal Intervals
SFFS	Sequential Floating Forward Selection
SFS	Sequential Forward Selection
SI	Single Image
SNR	Signal to Noise Ratio
SONA	ANU Research School of Psychology's Psychology Research Participation Scheme
ST	Skin Temperature
SV	Single Video
SVM	Support Vector Machines
TL	Transfer Learning
TN	True Negative
TP	True Positive
UvA-NEMO	UvA-NEMO University of Amsterdam-NEMO
VR	Virtual Reality
WT	Wavelet Transformation

Introduction

This chapter introduces this research by describing music's impact on human body and emotions. It provides an introduction to the area of affective computing relevant to the research. Then, the chapter focuses on the primary research questions and objectives. Finally, the chapter provides the organisation of the thesis, briefly summarising the focus areas of the remaining chapters.

1.1 Motivation

A famous line penned by Stevie Wonder in his song *Sir Duke* goes, "*Music is a world within itself, with a language we all understand*". Music is an art form enjoyed and understood by people all around the world. It is a popular form of entertainment that plays a significant role in our day to day life. Music is also an integral part of a country's culture so it shapes the preferences and emotional responses of its people. Listening to music or playing musical instruments can be an enjoyable experience for anyone. Music also has the power to elicit different emotions in people, which can be reflected in their conscious and unconscious responses.

The correlation between music and emotion is often mysterious and thought-provoking. Let us consider the following scenario. Alice is an employee in a multinational company. She had a terrible day at work, she had a big argument with her colleagues. So she left the office angry, put on her headphones and played the radio. There is an unfamiliar song playing. Suddenly, she is feeling very calm and relaxed, the music is even giving her chills. Then she recognised the artist and realised she always disliked the artist and the song. Thus she is perplexed to see how this particular song is making her feel so calm. Now let us consider another scenario. Bob is a university student. After a productive day of studying, he went to the supermarket to buy some necessary items. After a while, he started to feel irritated. He also started getting a headache, and he was unsure why he felt such way. He left the supermarket and after a few minutes he started feeling better. Only then he realised that, it may have been the music playing in the supermarket which caused him discomfort.

Different types of reactions have been reported in regards to people's reaction to

music. Some of them include: frustration when a particular style of music is played in a shop, sadness in response to a late-night movie soundtrack, nostalgia evoked by a familiar song playing on the radio [Juslin and Sloboda, 2001]. There are benefits from music including increased focus [Huang and Shih, 2011], reduction in stress and anxiety levels [de Witte et al., 2020; Umbrello et al., 2019], and improvement in memory and cognitive function [Innes et al., 2017]. Thus, music has a significant impact on our daily life and activities. Due to such diverse effects and applications of music, it is frequently used as stimuli in a wide range of applications.

Music has been used as an alternate form of medicine to reduce stress and anxiety among people for many years [Umbrello et al., 2019]. Music stimuli are used in therapeutic interventions which have been shown to improve sleep quality [Feng et al., 2018a]. It appears that it can affect the emotional and physiological state of a person, though this is controversial. Experiments have demonstrated that music has the ability to create specific patterns in the autonomic nervous system (ANS) that reflect a relaxing or arousing state [Krumhansl, 1997]. Some studies have demonstrated that music creates specific patterns in heart rate and blood pressure [Kim and André, 2008]. A significant increase in skin conductance was observed in subjects listening to emotionally intense music [Sudheesh and Joseph, 2000] and music evoking fear or happiness [Khalifa et al., 2002]. Skin temperature can also be influenced by listening to music that induces positive emotions [Hu et al., 2018]. Moreover, according to brain anatomy researchers, music can affect brain functions in two ways. First, it can act as a nonverbal medium that can move through the auditory cortex directly to the limbic system, which is a crucial part of the emotional response system. Second, it stimulates release of endorphins and allows these polypeptides to act on specific brain receptors [McCraty et al., 1998].

Due to the power of stimulating different emotional reactions, music therapy has been used to treat different mental disorders such as stress, anxiety, depression. It has also been used to treat epilepsy which is a neurological condition affecting around 65 million people all over the world. This condition affects 1 in a 100 people of the world [Thurman et al., 2011]. While 70 percent of patients with epilepsy can reduce their frequency of seizures with currently available antiepileptic medications, the other 30 percent are diagnosed with medically refractory epilepsy which cannot be helped by drugs [D'Alessandro et al., 2017]. People belonging to this category have a higher risk of death, depression and anxiety [Taylor et al., 2011]. Music therapy has been used to reduce the frequency of epileptic seizures among patients. However, little research has been done to understand exactly how music changes the physiological states of these patients to reduce the frequency of seizures, or how it causes changes in mental state in general.

Human – Computer Interaction (HCI) is a popular research field that focuses in large part on the ability of computing to understand human behaviour, and to a lesser extent their emotions. Emotions are a crucial human aspect to understand

because they have a huge effect on our intelligence and daily social interaction [Petrantonakis and Hadjileontiadis, 2010]. It has led to the emergence of a field that specifically deals with this phenomenon, called affective computing. Shouse [2005] defines affect as the “non-conscious experience of intensity”. Affective computing refers to analysing the physical and physiological reaction to different emotions. According to Picard [2000], affective computing has three types of applications. They are: 1) systems to detect the emotions of a user, 2) systems to act as how human would perceive a certain emotion and 3) systems that could “feel” an emotion. The research reported in this thesis deals with the first application. Research in this area of affective computing typically tries to identify human emotion states from their interaction with different stimuli. The stimuli can be visual such as digital images, text or videos. It can also be auditory such as music pieces, natural sounds or conversation.

Studies in the field of affective computing aim to build computing systems that can accurately understand human emotions. Understanding emotional reactions to music could be beneficial for giving personal music recommendations, which could improve emotional well-being by avoiding inappropriate music. There are different ways to capture data about people’s emotional reactions. The most common methods are self-reports [Dindar et al., 2019] and facial expressions [Shan et al., 2017]. Some other common measures are speech [Huang et al., 2019], pupillary response [Dhall et al., 2020], hand and body gestures [Noroozi et al., 2018]. However, some of these methods can be prone to high individual biases. For instance, people often refrain from showing their true emotions in their facial expression.

Physiological signals are strong measures found in human beings that demonstrate sensitivity to emotional changes. Emotion recognition using physiological signals has become a topic of interest for the last few years. This research has a range of applications such as stress detection [Liao et al., 2018], anxiety measurement [Tarrant et al., 2018], healthcare and so on. Identifying different physiological signal patterns caused by different types of music can help in understanding which music should be used in responding to or even treating different physical and mental disorders. As physical expressions can often hide true emotions, capturing emotional responses using physiological signals is beneficial in such cases as these signals are involuntary and cannot be readily hidden, muted or faked. Studies have also shown that music can induce universal psycho-physiological responses among different groups of people [Egermann et al., 2015]. There are different physiological signals which reflect human emotions. Some of them are: electroencephalography (EEG), galvanic skin response (GSR, also known as skin conductance or electrodermal activity), blood volume pulse (BVP), heart rate (HR), skin temperature (ST), pupil dilation (PD) and functional near-infrared spectroscopy (fNIRS). With the advent of modern wearable technologies, collecting physiological signals is becoming easier day by day.

Terms such as ‘chills’, ‘thrills’ and ‘frissons’ are often used by psychology re-

searchers to describe the psychophysiological moments of musical experience [Harrison and Loui, 2014]. A 'frisson' is 'a sudden strong feeling of excitement' and 'micro-frisson' is a sudden small feeling, which is too small to detect consciously, but is reflected by a person's physiological signals [Rahman et al., 2019]. In particular, chills and micro-frissons are closely related and they reflect the emotional intensity induced by music [Huron and Margulis, 2010]. These sensations are said to be highly reflected in different physiological measures [Guhn et al., 2007; Craig, 2005]. Thus, physiological signals can be very useful in analysing the emotional effects of music. This research takes a human-centric computational approach to understand these effects.

1.2 Research Questions and Objectives

Due to the complex nature of different physiological signals and music's ability to provoke a variety of emotions among people, several research questions can be formulated from this research problem. The main research questions relating to the relationship between music-induced affect and physiological signals examined in this research are:

- RQ1: Do different types of music generate different physiological response?
- RQ2: Do other stimuli (e.g. images, videos) have any impact on these physiological responses?
- RQ3: Can other stimuli (e.g. images, videos) be combined with music to understand their combined effects on physiological responses?
- RQ4: Which physiological signals perform better in differentiating the physiological responses to different stimuli?
- RQ5: Which computational methods are effective to analyse physiological response to music and other stimuli?

Based on these research questions, different computational models can be developed that can classify different music and other stimuli based on participants' physiological signals while they are engaging with the stimuli. In order to answer RQ1, RQ2 and RQ3, a number of different approaches are taken including computational, qualitative, quantitative and visualisation approaches. A total of six user studies are designed to collect a range of different physiological signals from participants when they interact with music and other stimuli. In order to answer RQ4 and RQ5, the key findings from the six user studies are compared. Various data pre-processing, feature selection and classification methods are applied and compared for these purposes.

The different computational approaches that are taken throughout this research work can be used in a range of applications. For example, the computational models can be further extended to develop biofeedback training models that may help

people change their emotional state while listening to a certain piece of music, e.g. reduce stress and anxiety levels, or reduce the frequency of epileptic seizures. These computational models can provide a significant contribution to the field of affective computing as well as medical research.

1.3 Thesis Outline

The chapters forming this thesis aim to answer the research questions outlined in section 1.1. Several experiments are conducted to understand the effects of music and video stimuli in human physiological response. There are a total of eight chapters and three appendices in this thesis. The names of the chapters are given below with a brief overview of the contents.

Chapter 1: Introduction

This chapter provides the motivation behind the work conducted in this research. It proposes the research questions and objectives that the studies conducted in this research aim to address.

Chapter 2: Background and Related Work

The research reported in this thesis lies at the intersection of multiple disciplines. It draws inspiration from emotion theory, physiology, music therapy, signal processing, machine learning and human-computer interaction / affective computing. This chapter provides the necessary background from these different disciplines that builds the foundation of the studies conducted during this research. The chapter primarily focuses on different ways to measure human emotion. It also highlights relevant research in this area.

Chapter 3: Methods

This chapter focuses on the computational background required to understand the different experimental designs and methods described in the latter chapters. It discusses the available dataset related to this research work. It also describes the stimuli used in different studies reported in the thesis. Then the chapter describes step-by-step the different computational approaches taken in the studies reported in chapters 4 to 7. This chapter also highlights some relevant research work related to the computational methods and provides a foundation for the experiments conducted and reported in the following chapters.

Chapter 4: Effects of Different Stimuli in Human Physiological Response

This chapter reports three short user studies that examine the use of video and

music stimuli to evoke physiological response in humans. The first two studies only look at the effects of visual stimuli in human physiological response. The results of these two studies show that the visual stimuli are effective in invoking emotional responses in participants, which can be reflected in their physiological responses. The results of these studies are then used to design an experiment where participants look at visual stimuli with different music stimuli playing in the background. The result of that study shows a decrease in the performance of the computational models compared to the previous two studies. Therefore, in the following chapters, studies are conducted to focus only on understanding the effects of different music in human physiological response.

Chapter 5: Effects of Music in Physiological Response

This chapter focuses on a detailed user study to understand participants' physiological responses to different genres of music. A number of different physiological signals are collected in this experiment. These are pre-processed and a number of features are extracted. Different feature selection techniques are tried along with a neural network model to identify the most effective combinations of features and feature selection methods. A novel visualisation approach named *Gingerbread Animation* is also proposed which is then validated using a transfer learning based computational model. This chapter provides knowledge on the relationship of participants' physiological and subjective response with different genres of music.

Chapter 6: Effects of Music in Brainwave Patterns

This chapter expands the study reported on chapter 5. The data was collected at the same time as the previous experiment, focusing only on participants' brain activity responses collected via EEG signals. It also expands on the computational methods, by comparing the neural network models with two other traditional machine learning models. The chapter also gives some insights into how specific brainwaves get impacted by different music. A qualitative analysis on participants' verbal comments on the music is also reported. The chapter extends the knowledge derived from the results of chapters 4 and 5.

Chapter 7: Effects of Music in Hemodynamic Response

The chapter focuses on another user study where participants' hemodynamic responses are collected via fNIRS signals. Some of the computational approaches that showed satisfactory performance in the previous studies are used here. In addition, a one-dimensional neural network model is explored to draw a comparison between using automatic feature extraction methods and a handcrafted feature extraction method. A qualitative and visual analysis are also conducted on the different aspects of the study. This chapter expands on the knowledge from previous chapters. It shows the efficacy of automatic feature extraction over handcrafted features.

A further comparison is shown between brain activity responses (EEG) and hemodynamic response (fNIRS) and which type of data is a stronger indicator of people's emotional response to music stimuli.

Chapter 8: Conclusion

The thesis concludes in this chapter by highlighting the contributions of this research and addressing the research questions specified in section 1.1. Some limitations and potential future application of the work is also discussed.

Appendix A: Experiment Procedure and General Guidelines

The appendix lists the step by step procedure taken in all of the user studies. The procedures include setting up procedures of each devices and guidelines that are given to the participants of the studies.

Appendix B: Experiment Related Documents

This appendix includes participation information sheet, consent form, questionnaires and SONA experiment sign up page for the user studies. Materials from one user study is attached as the other user studies follow the same pattern.

Appendix C: Miscellaneous Materials

Finally, the last appendix includes some links to relevant material, including a sample video of the Gingerbread Animation.

Background and Related Work

This chapter provides a review of the necessary background that is required to establish the foundation of this research work. It starts with a brief overview of the emotion theory and model, followed by a detailed description of different measures of emotion. Following that discussion, the chapter looks into previous research on the effects of music stimuli in human emotion. Finally, the chapter is concluded with a report on previous research that looks into the relationship between human physiological signals and human emotion. This chapter builds the foundation for the research to computationally understand the relation between music stimuli and human physiological response.

2.1 Emotion Theory and Emotion Model

Emotions are fundamental to human life. Paul Ekman defined the basic emotion theory by introducing six emotions that are known to be universally experienced across all cultures [Ekman, 1992]. These six basic emotions are: anger, disgust, fear, happiness, sadness and surprise. There are some distinct characteristics Ekman proposed regarding basic emotions such as,

- they can be identified through distinctive universal signals like facial expression
- they are associated with distinctive thoughts, memories and images
- they can be found in non-human primates
- they will have rapid onset and short duration
- they are not controlled voluntarily
- they will have physiological correlation and distinct subjective response

An extended version of this basic emotion model is Plutchik's emotion model [Plutchik, 2001]. A wheel model was introduced which included a wide range of emotions including eight basic emotions. These are: anger, anticipation, disgust, fear, joy, sadness, surprise and trust. Figure 2.1 shows Plutchik's model.

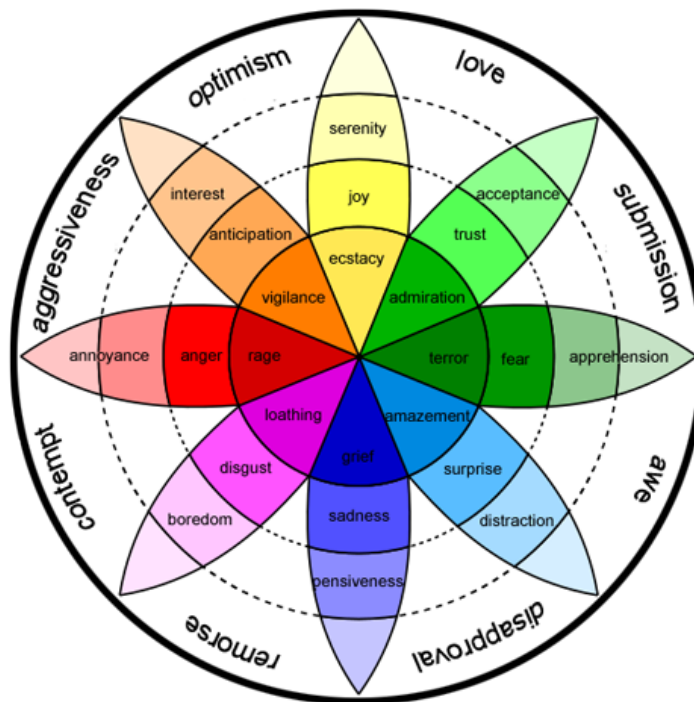


Figure 2.1: Plutchik's wheel of emotions

Researchers of affective computing often use two types of systems to model different emotions. The first one is using discrete labels, Ekman's basic emotion model and Plutchik's model are examples of that. The other one is using multiple dimensions or scales to categorise emotions. The main drawback of discrete labels is that stimuli can contain blended emotions which cannot be fully expressed just with one label [Kim and André, 2008]. Therefore a multidimensional space is more appropriate to express these emotions. The common scales used for this are valence (intrinsic goodness or badness) and arousal (alertness/response readiness). The first model that described this concept is called the "Circumplex Model of Affect". It was created by James Russell in 1980 [Russell, 1980]. The model is two-dimensional, having arousal and valence situated perpendicular to each other. Figure 2.2 shows an example of Russell's model.

Figure 2.2 shows where each emotion is situated in the multidimensional space. For instance, *happy* is considered a high arousal high valence emotion, and *sad* is a low arousal low valence emotion. *Fear* is a high arousal low valence emotion, while *calm* is a low arousal high valence emotion. The goal of the research reported in this thesis is to identify and understand the relationship of different stimuli patterns with some of the emotions shown in the emotion model. This research explores both discrete and continuous emotion models.

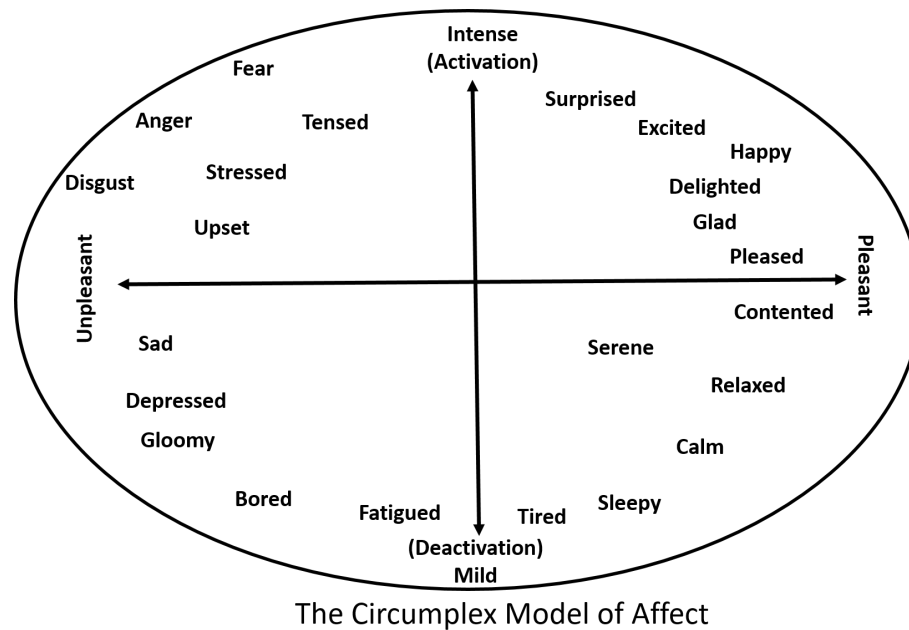


Figure 2.2: Russel's circumplex model of affect

Whether the emotions are described in discrete or continuous scale, they can be reflected in people's facial expressions, body language and physiological reaction. Therefore, the study of emotions has been developed with the usage of different types of stimuli that elicit emotions in humans. The elicitation is measured based on the three types of physical and physiological reaction. While physical reactions such as facial expression and body language can be seen easily, they also have the disadvantage of not necessarily being genuine. Physiological responses are involuntary and therefore can give a more accurate indication of different emotions. In the following sections, different measures to understand emotional response are described in greater detail.

2.2 Measures to Understand Emotion

The measurement of emotions has been researched in many different disciplines. There is still debate on which type of measurement is more suitable to measure different emotions. The commonly used measures are also dependent on the discipline. However, the biological, social and cognitive processes in emotions are interconnected with each other. Therefore to understand the effects extensively, multiple measurements can also be used. The measurements relevant to this research work can broadly be divided in three categories. They are: subjective measures, physiological measures and physical measures. They are described in the sections below.

2.2.1 Subjective Measures

Understanding different emotions from subjective self-assessment measures is a very common approach in affective computing, but it also poses a lot of challenges. Every person has a different perception in assessing their emotions. Different people view emotions differently, which leads to inconsistency in definition [Gooty et al., 2009]. Personality, age, gender, and life experiences all have severe impact on the subjective self-assessment reports for emotion recognition [Hoffmann et al., 2010; Sauter et al., 2010; Thompson and Voyer, 2014]. The most commonly used structured methods to assess subjective feelings are surveys, diaries and interviews. While these data are relatively easy to collect and a large amount of data can be gathered, subjective self-assessment data does not necessarily provide a very comprehensive dataset. Therefore, it is hard to get a deeper understanding just from subjective measures.

There are a number of publicly published assessment scales that are used to collect subjective reports [Ulstein et al., 2007; Harmon-Jones et al., 2016; Gross and John, 2003; Petrides, 2009]. However, researchers can create their own questionnaire based on their research goal to collect subjective data. For evaluation of the subjective ratings the most common approaches used are 5-point and 7-point Likert scale. A 7-point scale is shown to be more reliable than a 5-point scale whereas having a scale with more than 7 ratings is shown to be impractical [Alwin, 1997]. Even though self-assessment of subjects has been used for many years in psychology and medical research, the reliability of these measures are still in question due to the difficulty of generalising the measures. Therefore, physiological and physical measures are also used in the recent literature alongside subjective assessment reports.

2.2.2 Physical Measures

The physical measures discussed in this section are known to prominently show signs related to different emotions. The top physical measures that show efficacy in identifying emotions evoked by different stimuli are described below:

2.2.2.1 Facial Expression

Facial expressions are the most visible indicator of human emotional state, although it is not always accurate. Sometimes people try to hide their true emotional state so facial features do not always reflect their true emotion. To understand the emotion associated with different facial expression, there is a guide available that defines different expression based on the muscles that produce them. This guide is called "Facial Action Coding Units" [Ekman and Friesen, 1976]. However, complex emotions such as depression and surprise can be difficult to understand just with facial expression as these emotions are expressed using multiple action units [Wegrzyn et al., 2017]. Regardless of these complexities, facial expressions have been used widely in the field of affective computing and many computation techniques have achieved good results in predicting basic emotions [Dino and Abdulrazzaq, 2019; Ahmed et al.,

2019; Sun and Lv, 2019]. Combining facial expression along with physiological features can also improve the accuracy of computational models [Zhang, 2020; Huang et al., 2017b].

2.2.2.2 Gestures

Different body gestures are correlated to different emotions, some are easily recognisable by other humans while some might not be very obvious. Dominance and intimacy can be recognised from physical gestures such as head direction and head tilt direction [Mignault and Chaudhuri, 2003; Andersen and Sull, 1985]. Physical gestures such as head nods, smiles, eye contact are proven to result in better outcomes in interviews [Edinger and Patterson, 1983]. Features from body gestures have been used for many years in creating computational models for different emotion and cognitive loads [Kessous et al., 2010; Mitra and Acharya, 2007; Ginns and Kydd, 2019]. Gesture data can be effective in detecting basic emotions such as sadness, joy, anger, and fear [Kapur et al., 2005; Noroozi et al., 2018]. However, gesture data suffers from similar issues to facial expression, because these can be done voluntarily, and therefore may be manipulated.

2.2.2.3 Speech

Speech is a very important and useful physical measure in the field of Human-Computer Interaction. It is the fastest mode of communication between people and can be efficiently understood by machines. Thus, a great deal of previous research in the field of emotion recognition has focused on speech signals. There are different vocal characteristics that reflect emotional states as well. For instance, higher-pitched vocal samples have been shown to reflect high arousal emotions such as joy, fear and anger [Morningstar et al., 2017]. Similarly, low pitch voices can be linked to low arousal emotion such as sadness [Juslin and Laukka, 2003].

Speech signals are often combined with physiological responses to improve accuracy of the system [Greco et al., 2019]. Speech features have shown to identify different characteristics between suicidal and non-suicidal adolescents [Scherer et al., 2013]. A set of features from human speech can be analysed using computational models to detect seven discrete emotional states [Schuller et al., 2004]. Speech features are a popular measure in affective computing [Zhang et al., 2019; Latif et al., 2020], although finding the appropriate set of features is challenging due to different studies showing inconsistency in their feature sets used [El Ayadi et al., 2011].

2.2.3 Physiological Measures

A number of physiological signals have been used over the years to detect different categories of emotions. Some works take only one physiological signal as inputs while some use a combination of signals as inputs. The most common signals used

as raw inputs are described below, along with the devices and software used for them:

2.2.3.1 Electrodermal Activity

Electrodermal activity (EDA), also known as skin conductance (SC) or galvanic skin response (GSR), is a useful physiological signal which is seen to be sensitive to emotional changes [Kim and André, 2008]. The EDA response fluctuates slowly but significantly, reflecting the current emotional state, and has been shown to have a strong correlation with cognitive load [Shi et al., 2007; Lin et al., 2005]. The flow of electricity along the skin increases during stressful tasks, while it decreases during a relaxed state. Due to the reliability of data (less prone to noise) and easy analysis methods, EDA has become one of the most used physiological signals to detect various affective states. The signal can be measured by placing electrodes on the surface of the skin. They are generally placed on the hands; some devices are placed on the wrist while others require electrodes to be placed on the fingers. Some commonly used devices for capturing EDA responses are Empatica E4 [E4], Biopac EDA100C [BIOPAC], Affectiva Q Sensor [Affectiva], BodyBugg [Bodybugg] and BodyMedia Sensewear [Bai et al., 2016]. The experiments conducted during this research uses Empatica E4 device to collect EDA signals at the sampling rate of 4Hz. Figure 2.3 shows an image of Empatica E4 device.



Figure 2.3: Empatica E4 device [E4]

Skin conductance signals can be divided into two categories based on their frequency. One is referred to as skin conductance response (SCR) which shows the rapidly changing peaks in the signal. The other one is called skin conductance level (SCL) which are the slowly changing levels of the signal. Generally for affective computing, SCR signals are analysed.

As mentioned earlier, EDA signals have been extensively used in the field of emotion recognition [Feng et al., 2018b; Shukla et al., 2019; Al Machot et al., 2018; Yu and Sun, 2020; Cecchi et al., 2020]. In addition, research has been done specifically in music emotion recognition, mostly using the DEAP dataset (explained in section 3.1.3.1) [Al Machot et al., 2019; Bota et al., 2020; Ganapathy et al., 2021]. One

well-known study that is often cited in research on analysing people’s EDA response while listening to music was done by Kim and André [2008], achieving 70 percent accuracy for subject-independent classification of arousal and valence. A small number of music pieces were considered, and they were chosen based on participants’ individual preference. However, the work did not consider same music pieces for all of their participants. Therefore, the computational models could not generalise for large scale data involving multiple participants.

2.2.3.2 Blood Volume Pulse

Blood volume pulse (BVP) refers to measurement of the volume of blood that is flowing through the tissues of a particular part of the body. It is usually measured on every pulse. The measurement is done on any part of the body where the pulse can easily be accessed. The commonly used location for BVP data collection are the pads of the fingers or the earlobes. BVP has been shown to have a correlation to emotional state change. For instance, higher stress is said to be reflected by low BVP level and vice versa [Reisman, 1997]. Therefore this signal is often used in biofeedback training for reducing stress and anxiety. The sensors are also less complicated than for other signals, thus BVP is a popular choice for biofeedback based therapy. BVP is generally obtained by a photoplethysmography (PPG) sensor that detects the amount of light reflected by an infra-red light on the skin. This gives the amount of blood present in that certain area. Devices that record EDA such as Empatica E4, BIOPAC can also be used to record BVP. NeXus BVP is another very popular sensor for collecting BVP signals [NeXus BVP]. Figure 2.4 shows an image of NeXus BVP.

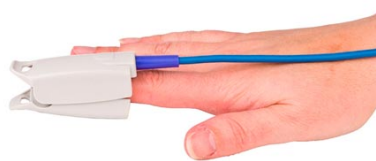


Figure 2.4: NeXus BVP [NeXus BVP]

The studies reported in this thesis also use Empatica E4 to collect BVP signals. They are collected at the rate of 64Hz, which means 64 measurements per second. Along with biofeedback therapy [Speckenbach and Gerber, 1999; Andrasik et al., 2001; Oded, 2018], BVP signals have also shown effective performance in emotion recognition [Handouzi et al., 2014; Nakisa et al., 2020]. Combining BVP signals with EDA has also proven to create a more robust machine learning system [Khan and Lawo, 2016].

2.2.3.3 Heart Rate Variability

Heart rate variability (HRV) is a strongly detectable signal that provides information about the cardiovascular system [Acharya et al., 2007]. These values can show significant increase during emotional arousal. HRV has shown strong association with emotion recognition, particularly social cognition [Quintana et al., 2012]. Combining other physiological signals with HRV values increases the accuracy of recognising emotion using various stimuli [Lee et al., 2006; Kulic and Croft, 2007]. Frameworks have been built using HRV to detect mental stress as well [Bousefsaf et al., 2013]. HR values can also be derived from raw BVP signals. It can be done by calculating the inter-beat interval (IBI) values. Therefore some devices that capture BVP also capture HR data. Both BVP and HRV are strong cardiovascular signals and thus have been frequently used in emotion recognition. HRV data can also be measured using the Empatica E4 device. Another well known device to collect this signal is NeuLog sensor [NUL-208]. Figure 2.5 shows an image of NeuLog Heart Rate and Pulse logger sensor.



Figure 2.5: NeuLog Heart Rate and Pulse logger sensor [NUL-208]

2.2.3.4 Skin Temperature

Skin temperature (ST) is another commonly used physiological measure. Although it is a relatively sluggish indicator, it is still able to show correlation to different emotional states. ST tends to increase during the relaxed state while it decreases during increased stress or anxiety [McFarland, 1985]. ST is measured normally on the surface on the skin, using the same sensor delivery platform as devices that measure skin conductance. The Empatica E4 device is used to collect ST signals studies in this research.

Figure 2.6 shows raw EDA, BVP, ST and HRV signals collected from a participant using Empatica E4 device.



Figure 2.6: Raw signals collected using Empatica E4 device [E4]

2.2.3.5 Electroencephalogram

Electroencephalogram (EEG) is the most common physiological signal used to understand brain activity associated with affective computing [Lin et al., 2010; Nie et al., 2011; Yang et al., 2019b]. It is used to record brain wave patterns. It is very useful in detecting mental conditions such as: epilepsy, sleep disorders, stroke, stress and anxiety [Moore, 2000; Thibodeau et al., 2006; Hou et al., 2015]. The most important data captured by EEG is the different brain waves which can be divided into multiple frequency bands. Each of these bands have different functions in the brain. These brain waves are:

- Delta (δ) waves – These waves are the slowest, having the lowest frequency range of 0.5 – 4 Hz. These waves are not seen in adult brains while they are awake. These waves are generally associated with deep sleep, as well as dis-function such as hypoxia and schizophrenia.
- Theta (θ) waves- Having the frequency of 4 – 8 Hz, theta waves are produced during sleep and drowsiness.
- Alpha (α) waves – Alpha waves have the frequency of 8 – 12 Hz, and are found in almost every part of the brain, but mostly in the occipital lobe. These waves are highly associated with any relaxed state. Alpha waves are often boosted during meditation or any other stress relieving activities.
- Beta (β) waves – Beta waves (12 – 30 Hz) are the most frequently seen brain waves that reflect the active state of the brain. They are mostly associated with increased attention and alertness.
- Gamma (γ) waves – These are the fastest brain waves (> 30 Hz), which are thought to increase cognitive function and boost memory and focus. These waves can also be found in stroke and epileptic patients [Hughes, 2008a].

EEG is typically recorded by placing some electrodes on the scalp. The number of electrodes and what information they capture differs based on the device that is used to capture the signals. The electrodes have distinguishable names, which reflects the placement location on the head. The name consists of a letter and a number, the letter represents the brain lobe while the number represents the position and hemisphere.

There are several devices that are used to record EEG data. The most popular device is the Emotiv EPOC+ headset, which is a 14-channel wireless headset that also has 9-axis motion sensors [Emotiv]. It comes with a software (requires paid subscription) that can be used to record raw EEG data. From the raw data, different brain waves and related information can be extracted. There are more versions of Emotiv, such as Emotiv insight, which has 5 channels and Emotiv EPOC flex, which is a more flexible, head cap system based version of Emotiv EPOC. Muse is another portable headband that is gaining popularity for therapy and attention training because of its free and user-friendly software [Muse]. It also has a free software development kit that can be used to extract raw EEG data. NeuroSky is a single channel, low cost device that captures EEG information from the frontal lobe [NeuroSky].

For this research work, EEG data is collected using the Emotiv EPOC headset device. Data can be collected from the pre-frontal, frontal, temporal and occipital lobes of the brain. Raw data from the device is collected at a sampling rate of 128 Hz, while the band power data is collected at a sampling rate of 8 Hz. Figure 2.7 shows an image of Emotiv EPOC+. The electrode placement of the device follows the International 10-20 System of Electrode Placement [Pastelak-Price, 1983].



Figure 2.7: Emotiv EPOC+ device [Emotiv]

2.2.3.6 Functional Near-infrared Spectroscopy

Functional near-infrared spectroscopy, commonly known as fNIRS, is a wearable, non-invasive means of measuring cerebral hemodynamic responses (blood flow variations) using near-infrared light. FNIRS is highly portable, safe, and less susceptible to noise in comparison to EEG signals. FNIRS has higher spatial resolution but lower

temporal resolution compared to EEG. Another advantage fNIRS has over EEG is that fNIRS does not need any conductive gel to connect with different brain regions, so it greatly reduces setup time and system complexity, and provides ecologically valid measurements [Curtin and Ayaz, 2019]. Recently, it has shown promising performance in measuring mental workload [Midha et al., 2021] and different emotions [Tang et al., 2021]. Hence, despite being a relatively new measurement modality, fNIRS has become a popular choice of physiological signal in brain-computer interaction studies.

fNIRS devices can collect responses from the pre-frontal cortex area. The pre-frontal cortex area of the brain is involved in various functions such as decision making, emotion processing and keeping focus [Ramnani and Owen, 2004; Manelis et al., 2019]. Hemodynamic responses in the brain are measured by changes in two types of blood oxygen conditions, namely oxygenated hemoglobin (HbO₂) and deoxygenated hemoglobin (HbR). An active state of the brain is generally reflected by an increase in HbO₂ and decrease in HbR as the blood supply overcompensates [Pinti et al., 2018]. Therefore, the concentrations of HbO₂ and HbR measured by the fNIRS used in this experiment can provide insight into the subjects' pre-frontal cortex emotion processing functions.

There are many devices that are used to collect fNIRS signals. Some of them are: OEG-16 [oeg16], Brite23 [brite23] and LIGHTNIRS [lightnirs]. In the study described in chapter 7, the NIRSIT device by Obelab [nirsit] has been used. The device is shown in Figure 2.8.



Figure 2.8: Obelab NIRSIT device

NIRSIT has a total of 24 laser diode sources and 32 detectors. The relative changes in hemoglobin concentration are measured by using light attenuation of two different wavelengths: 780 nm and 850 nm. There are 48 primary channels in this device of which 16 are located on the right, 16 in the center and 16 on the left of the pre-frontal cortex. In addition, the device also considers the horizontal, vertical and diagonal connections between channels. Four different distances (15 mm, 21.2 mm, 30 mm

and 33.5 mm) between channels are considered by the device. This results in a total of 204 channels. FNIRS data using this device is collected at the sampling rate of 8.138 Hz. The channel locations are shown in Figure 2.9.

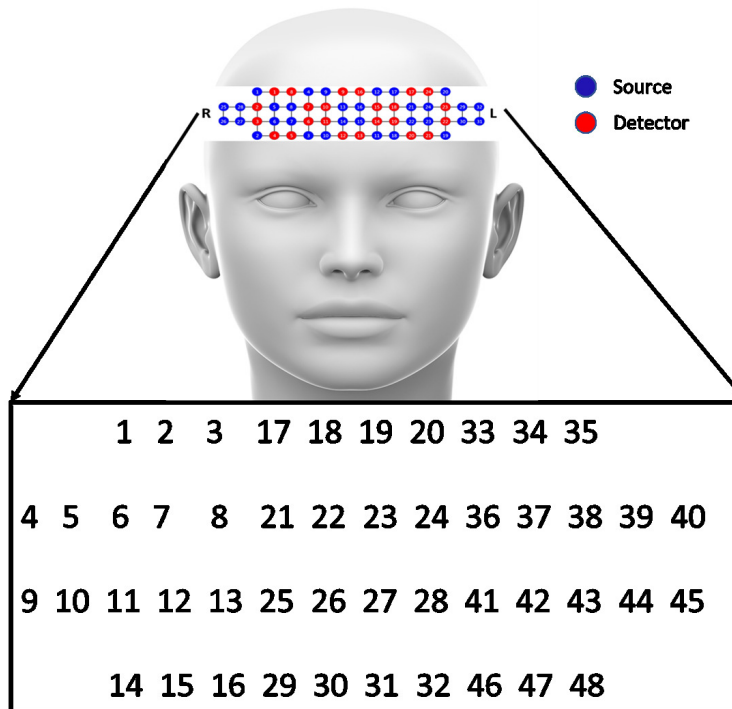


Figure 2.9: Nirx device channel locations at 30 mm separation

2.2.3.7 Pupillary Response

Human eyes provide valuable information on their emotions and current mental state. There are various features that can be derived from the eye such as pupil dilation, blinking rates, eye gaze point, fixation point and saccade and fixation times. Among these, pupil dilation, which is the measurement of pupil size over time, is considered a very effective feature in emotion recognition [Partala and Surakka, 2003]. Pupil diameter changes are said to reflect changes in brain state [Larsen and Waters, 2018]. Various eye-tracking devices are used to collect different features from human eye. Typically in lab based experiments these devices are placed aligning with a computer screen so it can track where a person is looking at the screen.

Some popular devices and software used for eye tracking are The Eye Tribe [Eye-Tribe], FaceLAB [FaceLAB], Pupil Labs [PupilLabs]. The Eye Tribe is used to collect pupillary response in the study reported in chapter 5. Figure 2.10 shows an image of The Eye Tribe.



Figure 2.10: The Eye Tribe device

Eye movements are useful in predicting human reaction to any audio or visual stimuli. Eye gaze has also been used to understand how people perceive manipulation in digital images [Caldwell et al., 2015]. Pupil dilation has been used as an indication of emotional arousal such as stress [Zhai and Barreto, 2006]. Changes in pupil dilation have been seen while subjects listen to familiar music and vocals [Weiss et al., 2016]. Therefore it can be a useful signal to identify the effects of different music.

2.2.3.8 Other Signals

There are various other physiological signals that have been used for emotion recognition related research problems. Some of them are blood pressure (BP), respiration rate, accelerometer (ACC), electromyography (EMG) and electrocardiogram (ECG). Increase in blood pressure can reflect emotional arousal such as fear [Kim et al., 2004] and stress [Vrijlkotte et al., 2000]. In addition to heart rate and blood pressure, positive and negative arousal can be also reflected from an increased rate of respiration [Rigas et al., 2007]. ECG and EMG, along with EEG have been used to detect cognitive load and stress [Katsis et al., 2008; Healey and Picard, 2005]. Accelerometer signals can be useful to detect movements associated with different emotions [Quiroz et al., 2017]. These other physiological signals are not considered for further exploration due to their low effectiveness in identifying emotions elicited by music or video stimuli.

2.2.4 Research on Music and Physiological Signals

Research to understand the effects of different genres of music on human physiological signals is a relatively new area. However, music has been extensively researched to be used as stimuli for therapy. There has been some popular work with music in the area of psychology and sleep research. Music stimuli has also been researched

in the field of emotion recognition. Physiological signals have also been researched a lot in the field of emotion recognition. In the sections below, some well-known and recent works in these areas are discussed briefly. These works form the basis for designing experiments reported in this thesis, that combine music stimuli and physiological signals.

2.2.4.1 Music as Therapy

Music therapy has been a well-known method to reduce many mental health issues and epileptic seizures. In a study conducted by Li and Xiong [2016], 90 students were divided into three groups where one group received music therapy, one group received music therapy along with biofeedback training and the third was the control group. The results demonstrated that music therapy in combination with biofeedback training has a significantly greater effect in reducing anxiety among students. In Yang et al. [2016], 22 psychiatric patients were divided into three groups based on their level of anxiety (mild, moderate and severe). They listened to 20 minutes of music for 10 days and their finger temperature and EEG were measured before and after music intervention. The results showed a significant decrease in anxiety across all three anxiety levels after the music intervention. Lee et al. [2016] performed a randomised controlled trial (RCT) on 64 students to measure effects of music therapy on stress. They found significant differences in blood pressure, diastolic blood pressure, pulse, SDNN, normalised low frequency, normalised high frequency of signals, and subjective stress after music therapy. One study on the effects of music in sleep quality was performed by Huang et al. [2017a]. They did a randomised controlled trial on 71 adults and divided them into control group, music group and music video group. Results showed that the music group had significantly longer subjective total sleep time than the control group and music video group. Coppola et al. [2015] used a set of Mozart's compositions for 2 hours per day for fifteen days on 11 patients with drug-resistant epileptic encephalopathy. They found that 5 out of 11 patients had a 50% reduction in their frequency of seizures. They also reported a significant improvement in the patients sleep and daily behaviour.

While music can be highly influential in reducing seizure frequency in many patients, some reports have demonstrated that music can also *trigger* seizures. One form of epilepsy called musicogenic epilepsy is prevalent in 1 out of 10,000,000 people. It is classified as a rare form of epileptic disorder [Berg et al., 2010]. In this kind of epilepsy, seizures can be provoked by listening to music, playing or even thinking of music [Sutherling et al., 1980; Ogunyemi and Breen, 1993]. According to a review done by Pittau et al. [2008], between 1884 and 2007 there were 110 reports of music-evoked seizures. One third of these cases showed epileptic seizures happened only because of music, while the rest reported other factors as well. Different types of music for instance classical, instrumental, or jazz, or specific instruments or even composers are said to have an impact on these type of seizures. A case study [Brien and Murray, 1984] reported a patient who has seizures while listening to music by

certain singers having a voice with “throaty” and “metallic” quality.

Music seems to have both proconvulsant and anticonvulsant effects on epileptic disorders. However, there is little research to understand these effects on a physiological level. Current clinical studies have not been able to explain why more neutral music (e.g. a specific sound) can invoke seizures in epileptic patients [Wieser et al., 1997]. A review by Hughes [2008b] discusses the presence of gamma brain waves in a majority of the seizures, particularly during ictal (seizure) activity in extratemporal and regional onsets. On the other hand, it is commonly known that increasing gamma waves in the brain can be beneficial as these waves are known to improve focus, cognition and memory formation. This is why many music therapy sessions use music or video stimuli to increase gamma waves in the brain to enhance cognitive ability. These multiple applications of music are fascinating and it is certainly worthwhile to explore how human physiological signals change pattern in response to music stimuli.

2.2.4.2 Different Measures for Music Emotion Recognition

There are many different areas where emotion recognition has been studied using music stimuli to elicit the basic emotions. The emotion labelling of these stimuli has primarily relied on the human participants annotating them. Some of the known examples of such work were done by Turnbull et al. [2008] and Trohidis et al. [2008]. The first work created a dataset of 500 songs with 18 different emotions annotated by three non-expert listeners. In the second work, 593 songs were annotated to six basic emotions by three expert listeners. It can be noticed that these methods can be highly time consuming and only few human participants’ responses can be collected. Therefore, the annotating may not be considered reliable. Another similar approach involves social tagging, which was done by Miller et al. [2008]. They used the social tags from the Last.fm website to get 960,000 free-text tags to annotate a large number of songs. While this approach is beneficial in getting a larger set of data, it suffers heavily due to not having a distinct structure of how the tags were given.

Automatic annotations have also been conducted based on different spectral, temporal and rhythmic features of the music. Features such as high pitch, fast tempo and bright timbre are associated with high arousal, while low pitch, slow tempo and soft timbre are related to low arousal. Similarly, music played in major key or tonal music are related to high valence. In contrast, music played in minor key or having atonal quality corresponds to low valence [Yang and Chen, 2011b]. Features such as song lyrics have also been correlated to emotion values [Yang et al., 2011]. Using musical features are an effective method to capture the music’s emotional content. However, they still do not reflect how humans perceive that music. Thus in this research, the focus is on identifying musical emotional response based on human physiological signals. The following section mentions some works in the literature that showed success in using physiological signals for emotion recognition.

2.2.4.3 Physiological Signals Based Emotion Recognition

A number of researchers have identified different categories of emotion using physiological signals. A variety of audio and visual stimuli have been used to elicit different emotions. Most of the work involve the use of images, text or videos as the stimuli for emotion recognition. Wu et al. [2010] collected skin response signals from participants when they watched six videos inducing the six basic emotions: surprise, fear, disgust, grief, happy and angry. A particle swarm optimization (PSO) based algorithm reached 78.7%, 73.4%, 70.5%, 62.6%, 62.5% and 44.9% respectively in classifying these emotions. In a research conducted by Valenza et al. [2011], three types of physiological signals (ECG, EDA and RR) were collected when participants watched image stimuli from the International Affective Picture System (IAPS). The stimuli contain five levels of arousal, and neutral as reference. A quadratic discriminant classifier (QDC) based model using features from the three signals reached over an average of 90.0% classification accuracy. Sharma and Gedeon [2013] used stress inducing and non-stress inducing texts to collect various physiological and physical signals from subjects. The model achieved 98.0% accuracy. Picard et al. [2001] collected many physiological signals such as heart rate (HR), temperature and skin conductance (SC), and used personalised imagery to evoke emotions in one subject. They achieved an accuracy of 81.0% for eight emotions.

Physiological signals have been used to design experiments to reduce epileptic seizures as well. Nagai et al. [2004] conducted biofeedback training using a series of animated pictures as stimuli to collect galvanic skin response (GSR) signals from 18 patients with drug-refractory epilepsy. Compared to the control group, the biofeedback group showed a correlation between their GSR responses and a reduction in frequency of seizures. In a recent study done by D'Alessandro et al. [2017], Mozart sonata for two pianos in D major, K448, was used on 12 patients with epileptic disorder for six months. They observed an average of 20.5% reduction in their frequency of seizures.

Based on the literature it is evident that there are certain kinds of music that are being used to reduce stress, anxiety and epileptic seizures. However, this intuitive approach has not been empirically explored by experimentation to understand if different music genres have different effects, and what specific physiological signals are beneficent to detect these effects. This research explores this phenomena in greater detail.

2.3 Summary

This chapter presented a survey of different measurements to understand emotions. Based on the previous work done in this area, it is evident that physiological signals can provide the strongest and most accurate measurements, when the emotions are elicited by different stimuli. It also provides evidence that music can be an effective

medium to elicit emotions. The following chapter discusses different computational approaches that can be used to understand the effects of music in physiological signals. Existing music datasets are introduced which have been used to collect physiological responses. Along with music, some visual stimuli are also discussed, as one of the goals of this research is to understand what other types of stimuli can be combined with music to elicit emotions that can be computationally analysed.

Computational Methods to Analyse Human Physiological Signals

This chapter elaborates on the various computational approaches needed to be taken to analyse human physiological signals. It starts by discussing various audio and visual stimuli that are used to invoke emotional reactions. Then the stimuli used in various experiments throughout this thesis are listed. Subsequently, step-by-step descriptions are provided to illustrate the different stages of physiological data analysis and the commonly used methods used to complete these analyses. Relevant research articles from the literature are highlighted where these methods were used to analyse human physiological responses.

3.1 Stimuli

This section starts with brief descriptions of some existing datasets that uses music stimuli. There are several limitations of these datasets which are described in the next subsections. Based on these existing datasets and their limitations, a new set of stimuli is introduced which is used for the experiments of this research. A number of existing datasets involving video stimuli are also introduced. These datasets are used in some of the preliminary experiments of this research.

3.1.1 Music Datasets Used in the Literature

- DEAP [Koelstra et al., 2011] - This dataset contains EEG and peripheral physiological signals collected from 32 participants while they watched 40 one-minute long music videos. Participants' ratings were reported based on arousal, valence, dominance and familiarity levels. Classification using a decision fusion based approach was also applied. This dataset has been widely used in studies on physiological signal based emotion recognition [Únal et al., 2020]. The dataset is publicly available. A drawback of using this database is that it is not possible to understand the effects of music alone.

- PMemo [Zhang et al., 2018a] - PMemo is a dataset for music emotion recognition from different popular music pieces. It contains emotion annotations of 794 songs and EDA signals collected from 457 subjects. The chorus was manually extracted from each music piece, which was listened to by the participants. However, the paper does not address why this method was chosen for the scenario. The stimuli are not suitable for a real world setting where people would listen to a piece of music from the beginning, not just the chorus. Generally someone would listen to a music piece from the beginning for some time in order to develop an emotional response to the music. Along with DEAP, this is the only dataset that contains participants' physiological responses to music. The dataset is currently not available publicly.
- Emotify [Aljanaki et al., 2016] - This dataset contains 400 musical excerpts from four different genres annotated with induced emotion. The participants manually annotated the excerpts to nine different emotion categories (amazement, solemnity, tenderness, nostalgia, calmness, power, joyful activation, tension, sadness). An advantage of this dataset is that the music excerpts are one minute long, which gives a longer time for the participants' emotion to be induced. However, this dataset only contains verbal responses, which does not suit the experiments for research involving physiological signals.
- emoMusic [Eerola and Vuoskoski, 2011] - The dataset contains 110 film music excerpts which were labelled by 116 participants into five discrete emotions (anger, fear, sadness, happiness and tenderness). The music excerpts were 45 seconds long. Analysis was conducted on the valence and arousal level of participants' responses. This dataset also does not contain any physical or physiological responses. The dataset could partially be compared to DEAP as the data has both music and video components. However, it is not suitable for the purposes of this research as the dataset contains film music excerpts.
- MER60 [Yang and Chen, 2011a] - The dataset contains excerpts from 60 English pop songs. A limitation of this dataset is similar to PMemo where chorus excerpts were manually extracted from the song for use as stimuli, which does not replicate real world settings of evoking emotional reactions. The dataset was created for mood regression tasks and only valence and arousal annotation was reported. Another characteristic of the database is that all of the annotators were from a Chinese cultural background. It is unclear whether this created differences in the annotations of the all English songs.

3.1.2 Limitations of the Available Datasets

There are several limitations of the existing music datasets which posed a challenge to their use in the experiments. One of the primary limitations were the genre and du-

ration of music used in the datasets. As reported in section 3.1.1, most of the datasets contained excerpts that were manually extracted from the music pieces. These excerpts are sometimes too short (< 30 seconds) which may not be enough to evoke a strong emotional response. In addition, many of these excerpts were extracted from the middle of the song, e.g. the chorus. In a real world scenario, people usually listen to a music from the very beginning, and often form an emotional reaction during the introduction of the song. The initial lyrics of the song is also an important part to understand the emotion that is being portrayed. Thus, recording the reaction to the beginning of a music is a crucial part of this investigation.

Furthermore, none of the datasets (except DEAP) are currently available publicly, and thus were not able to be used for comparison purposes. DEAP has been extensively studied in the literature. It also contains both music and video stimuli in the same component, which is not suitable for the research objective of this work. Therefore, this dataset was not used.

Music therapy studies have predominantly seen the use of classical music stimuli during therapy. However, using only one type of stimulus is not prudent and limits the ecological validity of the results [Harrison and Loui, 2014]. It is necessary to use a combination of different genres as this is how people listen to music in general. Thus, alongside classical, music stimuli from instrumental and pop music genres were also chosen for this research.

3.1.3 Stimuli Used in the Experiments

Based on the existing literature, datasets and their limitations, the music and video stimuli were selected for the experiments described in chapters 4, 5, 6 and 7. They are described in detail in the following sections.

3.1.3.1 Music

A total of 12 music pieces were used for the experiments which were divided into three categories: classical, instrumental and pop. These music stimuli were chosen based on some specific characteristics. After analysing a number of classical music stimuli, Hughes suggested that music stimuli which have a long lasting periodicity (phrases spanning several bars of music) have a positive influence on the brain [Hughes and Fino, 2000]. Therefore, four classical music stimuli with this feature were selected.

Binaural beats are a type of audio stimulus which can synchronise brainwaves to enhance specific brainwave patterns [McCraty et al., 1998]. These beats have different applications depending on which types of brainwaves are being enhanced. For this research, two different types of binaural beats were chosen: a piece that increases gamma waves in the brain to regain focus and awareness [Gamma, 2016], and a piece

that increases alpha waves in the brain, primarily used for meditation and relaxation [Alpha, 2017]. The other two instrumental piece chosen were used by Hurless et al. to analyse the effects of these stimuli in producing alpha and beta waves on the brain. Finally for the pop stimuli category, four music pieces were selected based on the No. 1 song of the Billboard Hot 100 year-end charts from years 2014-2017 [Billboard]. The names of the 12 music stimuli and their corresponding genres are shown in Table 3.1.

Table 3.1: Music stimuli used in the experiment

Genre and Stimuli No.	Music Stimulus Name
Classical 1	Mozart Sonata K.448 Coppola et al. [2015]
Classical 2	Mozart Sonata K.545 Lin et al. [2013]
Classical 3	F. Chopin's "Funeral March" from Sonata in B flat minor Op. 35/2 Hughes and Fino [2000]
Classical 4	J.S Bach's Suite for Orchestra No. 3 in D "Air" Hughes and Fino [2000].
Instrumental 1	Gamma Brain Energizer Gamma [2016]
Instrumental 2	Serotonin Release Music with Alpha Waves Alpha [2017]
Instrumental 3	"The Feeling of Jazz" by Duke Ellington Hurless et al.
Instrumental 4	"YYZ" by Rush Hurless et al.
Pop 1	"Happy" by Pharrell Williams
Pop 2	"Uptown Funk" by Mark Ronson featuring Bruno Mars
Pop 3	"Love Yourself" by Justin Bieber
Pop 4	"Shape of You" by Ed Sheeran

3.1.3.2 Video

Video datasets used in this experiment were chosen to compliment the music emotion recognition tasks of the experiments. These datasets were used in the experiments described in chapter 4. The datasets are described below:

- Acted Facial Expression In The Wild (AFEW) [Dhall et al., 2011] - This dataset contains clips from 957 videos that contain emotions in six basic emotions: angry, happy, disgust, fear, sad, surprise and neutral. The sequence lengths of the videos are 300 - 5400 ms.
- UvA-NEMO Smile Database [Dibeklioğlu et al., 2012] - This database contains 1240 videos from 400 subjects. The smiles are classified into two classes: genuine and posed.

- MAHNOB-HCI [Soleymani et al., 2011a] - In this dataset, physiological signals from 27 participants were collected while they watched 20 emotional videos. They reported their arousal and valence level for those videos.
- MMI [Pantic et al., 2005] - MMI is a web based database that has facial expression data from 61 participants acting different emotions and 25 participants reacting to those emotional videos.
- Anger dataset [Chen et al., 2017b] - This dataset follows a similar approach to the UvA-NEMO database, the difference being the use of video containing anger instead of smiles. It contains a total of 20 videos that include spontaneous and acted anger.

3.2 Computational Methods

The computational approaches to understand the relationships between music and human affective reasoning involve multiple steps. These steps include: data collection, pre-processing, feature extraction, feature selection and classification. Figure 3.1 shows a summary of the computational techniques used in the later chapters of this thesis.

3.2.1 Experiment Design and Data Collection

In order to analyse human physiological signals, first the signals need to be collected in a real life setting or lab-based experiments. Signals collected during a real life setting are highly prone to noise and often a large number of data need to be discarded because of that. Therefore a majority of the time, signals are collected using a lab based experiment where the environment can be controlled by the researcher more easily. Experiment designs are primarily divided into two types. The first one is between group design, where one group participates in only one experiment condition. So the two groups that are being compared are exposed to different experiment conditions. The second type is within group design, where both the conditions considered for comparison are carried out in the same experiment process. This means that all of participants experience multiple experiment conditions. Depending on the goal set up for the experiment, the appropriate design method is chosen. The studies reported in this thesis follows the within group design.

While collecting data, it is important to measure the effects the researcher is interested to observe and analyse. These are called dependent variables. In order to understand the outcome of the dependent variables, researchers need to control a number of other variables. These are known as independent variables.

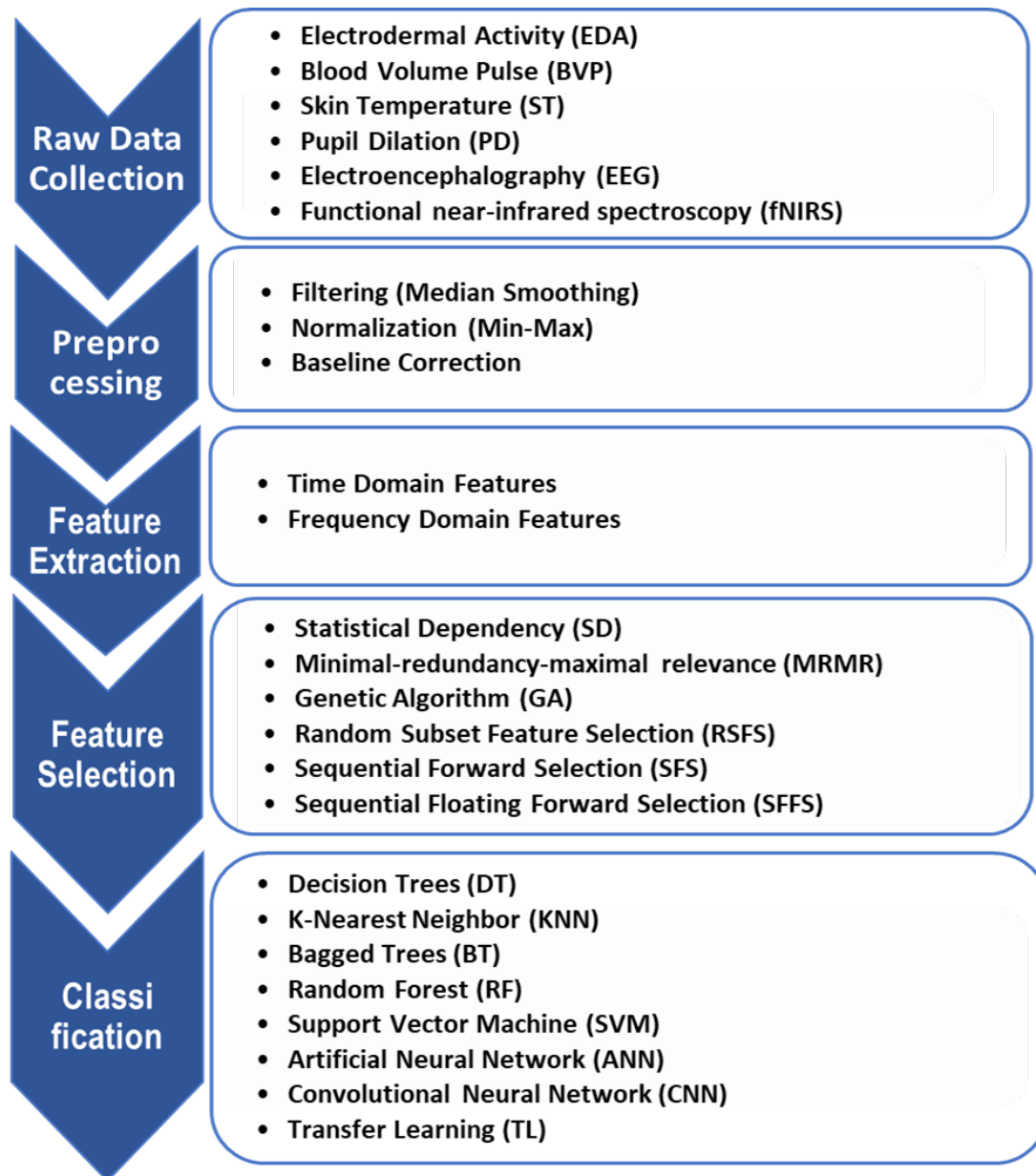


Figure 3.1: Summary of methods used in the computational analysis

In this research, depending on the experiment, the independent variables will be different audio and video stimuli, and the dependent variables will be the changes in participants' physiological responses.

After setting up the dependent and independent variables, a number of external environment variables are considered. These need to be kept as consistent as possible for every experiment so that they do not influence the outcome of dependent variables. Some examples of these environment variables are: location and time of day of the experiment, room temperature, participant age range, gender.

3.2.2 Data Pre-processing

Physiological signals collected from subjects during the experiment are highly prone to artefacts caused by subject movements, such as blinking during eye gaze tracking. In addition, sometimes a few channels of the devices fail to receive a good connection and therefore add noise artefacts to the collected signals. Therefore, it is very important to use some pre-processing techniques to remove these artefacts before doing any further analysis. Several standard pre-processing methods currently exist for physiological data. Some of them are: normalisation, filtering and baseline correction. These steps are necessary to perform in order to remove various artefacts and make the data suitable for next steps such as, feature extraction, feature selection and classification.

3.2.2.1 Filtering

To remove noise caused due to various environmental factors and device issues, it is crucial to apply some filtering techniques on the signals. There are a couple of filtering techniques used to remove artefacts from physiological signals. Standard physiological signal filters used in this research are:

- Median Smoothing Filter – This is a kind of a smoothing technique which replaces each data point with the median of its neighbouring data points. Median filters are frequently used in signal processing as well as image processing.
- Butterworth Filter – This low pass filter is also called a maximally flat filter as it has a flat frequency response in the passband. It uses a cut-off frequency and all frequencies higher than that value are changed to zero. Butterworth filters are often used to filter EEG and skin conductance response signals.
- Band-pass filter – Band-pass filter uses two cut-off frequencies and rejects any frequency points that lie outside of these two frequencies. Some variations of

this filtering techniques are high-pass and low-pass filters, where only one cut-off frequency is used.

- Band-stop filter – A band-stop filter uses two cut-off frequencies and rejects any frequency point that lies between these two frequencies.

3.2.2.2 Normalisation

The values from different physiological signals are subject-dependent, which means they have different ranges of values. So it is necessary to use some normalisation methods on these raw signals to bring all the values within one range. The commonly known techniques for physiological data normalisation are defined below. These methods have been applied to this research's experiment data and the method that leads to better performance has been chosen.

- Min-max normalisation – The min-max normalisation technique converts the collected data to bring them within a range, specified by a minimum and maximum value. The equation for min-max normalisation is,

$$v' = \left(\frac{v - \min_v}{\max_v - \min_v} \right) * (new_max - new_min) + new_min \quad (3.1)$$

Where, v' corresponds to min-max normalised data and v is the range of raw data, \max_v and \min_v are the maximum and minimum value of v respectively.

For example, if $new_min = 0$ and $new_max = 1$, all the values will be normalised to have a value within the range of 0 to 1. This technique is a popular choice for pre-processing physiological signals as it removes the subject specific variance in the response.

- Normalisation by Z-score – This technique converts all the attributes to a common scale with the average of zero and standard deviation of 1. This equation for Z-score normalisation is the following:

$$e' = \frac{e_i - E}{std(E)} \quad (3.2)$$

Here, it is assumed that there are multiple rows of values where each row contains multiple different variables. The above equation gives the normalised

value of data e_i which is located in the i column of row E . Note that *std* stands for standard deviation.

- Decimal Scaling – This is a normalisation technique which moves the decimal point of values in the dataset. The equation for this technique is,

$$v^j = \frac{v^i}{10^j} \quad (3.3)$$

Here v^i represents every value in the dataset and j represents the number of digits in the largest number.

3.2.2.3 Baseline Correction

As previously mentioned, physiological signals are quite sensitive to noise generated by participants' head and body movements. They are also affected by noise from the external environment. These interference effects often result in shifts from the baseline values and fast spikes in the signals. In such scenarios, an additional step of baseline correction is necessary. This step is often needed for spectroscopy signals. Some widely used baseline correction techniques are listed below. Similar to normalisation techniques, both approaches listed below have been applied to this research's experiment data and the method that leads to better baseline correction has been chosen.

- Polynomial fitting - This method removes the noisy signal elements without losing the information of small peaks that may hold important information. In the literature, several algorithms have been proposed using this technique [Lieber and Mahadevan-Jansen, 2003; Zhao et al., 2007; Lan et al., 2007; Zhang et al., 2010].
- Wavelet transforms - This technique is based on decomposing a signal based on mathematically defined functions (wavelets). This technique has gained popularity in image processing as well as physiological data analysis [Bertinetto and Vuorinen, 2014; Shao and Griffiths, 2007]

3.2.3 Feature Extraction

Physiological signals collected using multiple devices provide a large amount of data for each participant. Not only is it difficult to analyse the entire set of recorded data, it is also computationally very expensive. Therefore, a number of features are extracted from the data after finishing the data normalisation and filtering. The extracted features can illustrate different properties of the signals such as statistical trends and

randomness properties. These manually extracted features (also known as hand-crafted features) can broadly be divided into two categories, time domain features and frequency domain features. Time domain features could further be divided into linear and non-linear features.

There are a number of papers in the literature that talk about commonly used features in affective computing [Picard et al., 2001; Valstar et al., 2016; Samara et al., 2016; Acharya et al., 2019; Chowdhury et al., 2013; Katsis et al., 2008; Triwiyanto et al., 2017]. Here are some of the features that can be calculated from the physiological signals, shown in Table 3.2:

Table 3.2: List of features for physiological signal analysis

- | | |
|--|--------------------------|
| • Mean | • Minimum |
| • Maximum | • Standard Deviation |
| • Skewness | • Kurtosis |
| • Interquartile Range | • Variance |
| • Number of peaks for each periodic signals | • Summation |
| • Absolute Summation | • Mean amplitude rate |
| • Average Amplitude Change | • Root Mean Square |
| • Average of the power of signals | • Integrated Signals |
| • Ratio of the maximum and minimum | • Log Detector |
| • Difference Absolute Standard Deviation Value | • Mean rise duration |
| • Mean of the absolute values of the first differences | • Simple Square Integral |
| • Mean of the absolute value of the second differences | • Approximate Entropy |
| • Detrended Fluctuation Analysis | • Fuzzy Entropy |
| • Shannon's Entropy | • Permutation Entropy |
| • Hjorth Parameters | • Hurst Exponent |
| • Power spectral density analysis | |

Depending on the data, time and/or frequency domain features are extracted from it. There are also various techniques that are used to convert from time to frequency domain or vice versa. Some of the commonly used methods are Fourier Transformation, Wavelet Transformation, and Principal Component Analysis. These methods are briefly described below:

- Fourier Transformation (FT): FT is done by taking an N point time domain signal and decomposing it into N frequency domain signals each containing a single point.
- Wavelet Transformation (WT): WT is similar to FT, and is used to convert time

domain features to frequency domain by dividing the signal into frequency components instead of time. The difference between FT and WT is that while FT is good at giving a global frequency information, it does not provide details on the time components associated with the frequency. WT overcomes this limitation by providing the time-frequency localisation.

3.2.4 Feature Selection

The feature selection process is often used before classification in order to reduce the dimension of the feature space. Sometimes the extracted feature set can contain redundant or noisy features. Feature selection algorithms can identify those features and remove them from the set. A reduced number of features can also result in a shorter run time for the classification process, which means it helps create a robust system. Having irrelevant features can significantly decrease the performance of classification models [Kohavi and John, 1997]. Pohjalainen et al. [2015] provided a detailed explanation and comparison of many state-of-the-art feature selection methods. The feature selection process can be done in two ways. One is to rank each of the features and select a fixed number of top ranked features to build the feature set; the other way is to select different subsets of features and classify in order to find the optimal set. Some feature selection methods used for physiological data are briefly described below, all of them are explored in the experiments of this research:

3.2.4.1 Feature Ranking Methods

- Statistical Dependency (SD) - SD ranks all the features by measuring if the values of the features are dependent on their class labels or not.
- Minimal-redundancy-maximal-relevance (MRMR) [Peng et al., 2005] – Features are ranked according to the mutual information between features and their associated class labels.

3.2.4.2 Feature Subset Methods

- Genetic Algorithm (GA) [Man et al., 1996; Yang and Honavar, 1998] - A heuristic optimization method that selects a subset of features having the best fitness value.
- Sequential Forward Selection (SFS) [Whitney, 1971] - This is a greedy search method that starts with an empty feature set and adds features according to their contribution in maximizing the output.

- Sequential Floating Forward Selection (SFFS) [Pudil et al., 1994] - This is an extension of SFS where after each forward step, the method executes backward steps until the objective function increases.
- Random Subset Feature Selection (RSFS) [Räsänen and Pohjalainen, 2013] - This method ranks every feature based on their relevance values and then chooses a subset of features based on that ranking. The relevance values are updated at each iteration until a (locally) optimal set of features is found.

3.2.5 Classification Methods

After extracting and selecting a good set of features from the collected data, the next and final step is to classify them into different categories based on the goal of the study. In this section, some of the state-of-the-art classification techniques commonly used are discussed.

3.2.5.1 K-Nearest Neighbour

The nearest neighbour method selects a particular point's label based on the labels of its neighbouring labels. The most commonly used method is the K-nearest neighbour (KNN) method where the value of k describes the number of neighbours taken into account. An advantage of this method is that it is faster compared to many other machine learning techniques, and it always converges. However, determining the value of k is the most challenging aspect of using this classifier as it has a huge effect on the classification accuracy. With the increasing capability of collecting large scale physiological data, using KNN to create a robust classification model is becoming more challenging. Despite this, KNN is widely used in many emotion recognition papers that are based on physiological and physical signals [McDuff et al., 2012; Palanisamy et al., 2013; Mehmood and Lee, 2015; Shukla and Chaurasiya, 2018; Xie and Xue, 2021].

3.2.5.2 Decision Trees

Decision trees build classification models as a tree structure. Data is broken into smaller subsets to form decision nodes and leaf nodes. Classification labels are represented by the leaf nodes. Decision tree models are quite intuitive and easy to explain. Therefore, several publications have appeared in recent years which used decision trees as their classification model for emotion recognition. Liu et al. [2018] proposed a computational model for speech emotion recognition using decision trees. The model achieved 89.6% accuracy. Salmam et al. [2016] achieved 89.2% accuracy using a decision tree based classifier for facial expression recognition. Bobade and Vani [2020] used decision tree for mood classification using multimodal physiological data which reached 68.2% in ternary classification and 87.6% in binary classification. One

disadvantage of decision tree classification is that it can be more unstable compared to other models.

3.2.5.3 Random Forest

Random forest is another supervised method which overcomes the limitations of decision trees. It is created using an ensemble of decision trees. A “forest” of decision trees is constructed using merging techniques such as bagging or bootstrap aggregating. For each data point, a result is produced by each decision tree and then a final prediction is made by merging all results. Thus it overcomes the instability issue of decision trees and creates a more accurate and stable model. Increasing the number of trees in the model increases the precision of the final prediction. Random forest is a very popular choice among researchers for physiological signal classification. Chaudhuri et al. [2018] implemented the random forest classifier to achieve 92.9% accuracy in identifying binary thermal states. It has also shown promising results in voice recognition [Zvarevashe and Olugbara, 2018]. Islam et al. [2019] used random forest for neurodegenerative disease classification using gait features. All of these techniques rely on efficient feature extraction and feature selection which can also increase computational time and complexity.

3.2.5.4 Support Vector Machines

The support vector machine is another popular supervised learning algorithm that is used in both classification and regression models. It is appropriate for binary classification problems. This method is also found to be appropriate for use with a small dataset with a large dimensionality in the feature space [Sánchez-González et al., 2018]. The algorithm plots all data in an n -dimensional space for n number of features and then classifies every data point by finding the hyper-plane that differentiates two classes. Wang et al. [2011] used SVMs and frequency domain features to achieve an average of 65.0% accuracy from an EEG-based emotion recognition model. Sharma and Gedeon [2014] achieved 98.0% accuracy using SVM in their computational model for stress recognition. Vecchio et al. [2020] used SVM to classify Alzheimer’s disease from EEG biomarkers and achieved 95.0% accuracy for binary classification. Despite its impressive performance, it poses some challenges in multi-class classification problems. Furthermore, it also does not perform well with large scale datasets, which makes it unsuitable for real time data monitoring.

3.2.5.5 Bayesian Classifier

This classifier applies Bayes’ theorem to calculate the prior probability of every class and then classifies a new data point based on that prior probability. Large datasets can be easily classified using a Bayesian classifier. Therefore usage of this classifier is prevalent in fields such as big data, bioinformatics and is now gaining popularity in affective computing as well. Liao et al. [2005] used a large dataset containing facial expression, head and eye movements and other behavioural data and classified them

using a dynamic Bayesian network for stress monitoring system. A similar study using this classifier for stress recognition is done in Barreto et al. [2007]. Another study uses a Bayesian classifier on EEG data to classify different emotions [Chung and Yoon, 2012]. A major disadvantage of the Bayesian classifier is that it assumes all of the features have equal contribution, which is often not the case, and especially so with physiological response data.

3.2.5.6 Artificial Neural Network

Artificial neural networks (ANNs) are one of the primary classification techniques used in the field of machine learning and affective computing. The method is inspired by the structure of the human brain, having multiple different mathematical layers processing the data from one input layer to give useful information in the output layer. Although the functions of ANNs are quite complicated and sometimes hard to interpret, they are used frequently in affective computing. Wilson and Russell [2003] used an ANN to classify real time mental workload using physiological measures achieving the maximum accuracy of 86.0%. Srinivasan et al. [2007] proposed an ANN based automated epileptic EEG detection system. A study predicts reading comprehension scores using ANNs from subjects' eye movements [Copeland et al., 2014]. It has been effective in recognising stress from physical gestures and speech as well [Scherer et al., 2008]. Usage of ANNs are also said to be effective in pharmaceutical research [Agatonovic-Kustrin and Beresford, 2000].

Human brains have the capability of interpreting the context behind complex scenarios, which are difficult for computers to interpret. Since ANNs are created using the concept of how human brain processes information, it has the ability to understand patterns in complex data such as physiological signals. ANNs can perform reasonably well using features from physiological data as well as raw physiological data. Singh et al. [2013] used a set of features from participants' photoplethysmography and galvanic skin response data to classify stress during driving tasks. They tested with seven different combinations of neural networks and achieved a maximum of 89.2% in precision. Pinto et al. [2020] used ECG data and neural networks to classify different emotion states (neutral, fear, happy) and reached a maximum of 77.0% accuracy.

A disadvantage of this method is that it requires a large amount of data to train a robust model. It may be able to achieve high accuracy with a small dataset, but in such instances it does not tend to generalise well with larger datasets.

3.2.5.7 Convolutional Neural Networks

With the advent of modern hardware and large scale datasets, computers became able to do heavy computation with minimum effort and time. This allowed the rise of deep neural networks (DNN), which allow to create more complex network struc-

tures than the ones mentioned in the previous section. DNNs have the capability to automatically extract useful features from the data and make predictions based on the extracted features. One of the most popular deep learning techniques is convolutional neural networks (CNN). CNNs can recognise patterns from one-dimensional (1D), two-dimensional (2D) or three-dimensional (3D) data. Therefore in the recent years, this became popular in physiological signal analysis [Rim et al., 2020; Dar et al., 2020] because these types of data can also be represented in 1D (time-series data), 2D (image) or 3D (video) data. Some recent studies used deep learning methods which automatically extracted features from the raw data, reducing overall computational time. Yang et al. [2019a] conducted a study where they collected fNIRS signals while patients with mild cognitive impairment completed three mental tasks. They applied a CNN on the signals and reached a highest accuracy of 90.6%. Ho et al. [2019] also investigated effects of mental workload signals by applying a combination of deep belief network (DBN) and a CNN. The classification accuracy using DBN and CNN reached 84.3% and 72.8% respectively. CNNs have also shown promising results in classifying six basic emotions [Oh et al., 2020]. Sheykhivand et al. [2020] used a combination of CNN and LSTM to predict emotions evoked by music using participants' EEG signals. This gives a strong motivation to use this method for physiological signal based music emotion recognition. However, detecting automatic biomarkers for different emotions while listening to music still remains a challenge.

3.2.5.8 Recurrent Neural Network

Recurrent neural networks (RNNs) hold sequential information of data so they are a very common choice for time-series data analysis. The way an RNN differs from shallow neural networks or convolutional networks is it has a memory component. This means that the network takes information from the previous input to influence the current input. It is often used in analysis of data that are sequential in nature. Examples of such data are, text, music, movies, speech [Zhang et al., 2021]. Zhang et al. [2018b] applied RNNs to classify emotions from EEG signals and facial expression to achieve the maximum of 89.0% accuracy. It has also been used in healthcare applications such as sleep stage detection [Cheng et al., 2017]. Of all of the variations of neural networks, RNNs are most widely investigated for music emotion recognition [Liu et al., 2019; Zhao et al., 2018; Grekow, 2020; Rajesh and Nalini, 2020]. However, most of these studies are based on detecting emotion from various characteristics of the music such as pitch, key, and do not involve human experimental/physiological data.

3.2.5.9 Long Short-term Memory

Long short-term memory networks (LSTMs) are a variant of recurrent neural networks created to learn long-term dependencies [Hochreiter and Schmidhuber, 1997]. The model is inspired by the concept of logic gates. It adds the mechanism of resetting/forgetting some of the information, and only storing the necessary information.

LSTMs have been used in the affective computing area. Similar to standard RNNs, LSTMs have also gained popularity in physiological data analysis [Dar et al., 2020]. Choi and Kim [2018] used LSTMs for arousal and valence classification using the DEAP dataset. They were able to reach 72.7% and 73.1% accuracy for arousal and valence classification respectively. Ma et al. [2019] also used EEG data from the DEAP dataset and their approach using an LSTM reached 92.87% accuracy classifying arousal and 92.3% accuracy classifying valence. Alhagry et al. [2017] used LSTMs to classify EEG signals, reaching up to 85.7% accuracy. Similar work was done by Liu et al. [2017] where they used LSTMs to classify EEG signals from the Mahnob-HCI database. Their classification model reached up to 73.1% classifying arousal and 74.5% for valence. A major disadvantage of this method is that due to the recurrent nature of the networks, computation can be extremely slow. It is also sensitive to random weight initialisation that causes the models to become unstable.

3.2.5.10 Transfer Learning

As explained in section 3.1, collection of human physiological data often requires large scale experiments. Collecting good quality data from human subjects is an extremely time consuming and challenging task. Therefore, most of the experiments reported are not able to generate a large amount of data to create a robust classification model. Furthermore, all the deep learning models mentioned in this chapter require a large number of training data which are often hard to gather for physiological data. Transfer learning aims to overcome this challenge by using a pre-trained dataset model to train a new model. The pre-trained model is trained to solve a different task. The pre-trained model then transfers some knowledge to the new model in order to accomplish the new task. In such cases, the final few layers of the pre-trained model are fine-tuned to learn patterns based on the new dataset. Using this approach, a more robust model can be created using a lower amount of data [Torrey and Shavlik, 2010]. Some of the popularly used pre-trained model for transfer learning are ImageNet [Deng et al., 2009], AlexNet [Krizhevsky et al., 2012], SqueezeNet [Iandola et al., 2016], ResNet [He et al., 2016], VGG [Simonyan and Zisserman, 2014]. In recent years, transfer learning approaches have been utilised to overcome the scarcity issue of physiological signals. For example, it has been used for driver status detection using a number of physiological signals [Chen et al., 2019]. Transfer learning has also been investigated recently to forecast different health outcomes from physiological responses [Chen et al., 2020]. Although it has been used in physiological data analysis, it has not been investigated much as a method to recognise emotion.

3.3 Statistical Analysis

Statistical analysis helps to identify the general trends and patterns observed in the data collected during any experiment. This also helps to interpret the collected data and justify the analysis results. Based on the normality of collected data type (e.g.

parametric or non-parametric data) different statistical methods can be used. In the field of affective computing, these are some of the frequently used statistical analysis methods:

- T-test – T-test is the most widely used statistical analysis test which compares the means of two different groups. This test can also be divided into two categories. If the two groups that are being compared are unrelated, then the t-test is called an independent samples t-test, whereas if the groups are the same then it is called a paired t-test. The T-test is begun by setting up a hypothesis that there is no significant difference in the means of the two groups. Based on the result of t-test the hypothesis is either accepted or rejected. This test is appropriate to compare two groups of parametric data.
- Analysis of Variance (ANOVA) – ANOVA tests are conducted to compare the means of two or more groups. There are multiple types of ANOVA tests, and depending on the correlation between the groups the appropriate test is chosen. The one way ANOVA test is used when there is one independent variable. If there are two or more independent variables then factorial ANOVA can be used. This method is appropriate for a group of data that follow a normal distribution.
- Chi- square Test – The Chi-square test is a popular non-parametric method primarily used to analyse categorical data and identifying relationships among them. Data is put on a contingency table based on the frequency of variables to analyse their relationships.
- Wilcoxon signed rank test – This is another non-parametric test which has a similar procedure to the paired t-test. Paired t-test is not appropriate for dataset which do not follow normal distribution. In such cases, Wilcoxon signed rank test is more appropriate.
- Kolmogorov–Smirnov test - This a non-parametric test to detect variance between samples.

3.4 Evaluation Measures

After completing the classification process, the predictive power of the classification model needs to be evaluated. In order to do that, different evaluation measures can be used. In order to understand the evaluation measures, the following four terms need to be understood.

- True positive (TP) – A sample that belongs to a certain class and the classification model also predicted it to be in that class.
- False positive (FP) - A sample that does not belong to a certain class but the classification model predicted it to be in that class.

- True Negative (TN) - A sample that does not belong to a certain class and the classification model also predicted it to not be in that class.
- False Negative (FN) - A sample that does not belong to a certain class but the classification model predicted it to be in that class.

The evaluation measures used in this research are listed below:

- Accuracy – This is the most popular evaluation measure which gives the fraction of correct predictions made by the classification model. This measure is useful for datasets with balanced classes, that is, with the same or similar number of members in each class. It is defined by the equation below:

$$accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (3.4)$$

- Specificity – This is also referred to as the true negative rate of the model. The equation for this measure is below:

$$specificity = \frac{TN}{TN+FP} \quad (3.5)$$

- Precision - Precision refers to the fraction of the predicted labels matched. It is also known as the positive predictive value (PPV). The equation to calculate precision is:

$$precision = \frac{TP}{TP+FP} \quad (3.6)$$

- Recall - Recall refers to the fraction of reference labels matched. It is also known as sensitivity. The equation is:

$$recall = \frac{TP}{TP+FN} \quad (3.7)$$

- F-measure - Also known as F-Score or F1 Score, the F-measure is a commonly used evaluation measure represented by the harmonic mean of precision and recall. It is considered a stronger measure than precision and recall as it takes

both these measures into account. It is a better evaluation measure than accuracy, especially for data with imbalanced classes. The equation is below:

$$f - measure = \frac{2 * precision * recall}{precision + recall} \quad (3.8)$$

- Geometric mean – It is also known as the measure of central tendency which measures the separation between the classification performance of the majority and minority classes. The equation of geometric mean is:

$$g - mean = \sqrt{sensitivity * specificity} \quad (3.9)$$

- Area under curve (AUC) – This demonstrates a curve that displays how well separated the probabilities from the positive classes are from the negative classes.

3.5 Summary

This chapter builds the foundation for Chapters 4 to 7 of this thesis. In this chapter, the stimuli and computational methods that are frequently used in the area of affective computing and physiological data analysis were introduced. Some of these stimuli and methods are used to achieve the research objectives of this work. In the following chapters, experiments that use some of the techniques mentioned in this chapter are introduced and their results are discussed.

Effects of Different Stimuli in Human Physiological Response

This chapter focuses on three different experiments which investigate the effects of emotional videos on physiological responses. The first two experiments used two different visual stimuli datasets and looked into the effects on human physiology when they watch the stimuli. In the third experiment, these two datasets were combined with music stimuli to understand the joint effect of these two types of stimuli on human physiology. This chapter builds on the work presented at the 31st Australian Conference on Human-Computer Interaction - OzCHI 2019 [Rahman et al., 2019], 32nd Australian Conference on Human-Computer Interaction - OzCHI 2020 [Rahman et al., 2020b] and the 2021 CHI Conference on Human Factors in Computing Systems - CHI 2021 [Rahman et al., 2021b]. In all of these works, I was the primary contributor.

4.1 Experiment 1 - Physiological Responses to Emotion Videos

In this experiment, participants' EDA activity was computationally analysed to recognise seven emotional categories while watching a total of 80 emotion videos. This experiment was conducted to understand the effects of different emotional videos on participants' physiology and whether patterns can be identified using different features and computational methods.

4.1.1 Methods

4.1.1.1 Participants

Twenty participants (14 female and 6 male) took part voluntarily in this experiment. The mean age was 23 years old with a standard deviation of 5.8. All the participants were asked to sign a written consent form before their voluntary participation in the study. This study was approved by the Human Research Ethics Committee of The Australian National University (ANU).

4.1.1.2 Dataset and Pre-processing

The Acted Facial Expressions In The Wild (AFEW) dataset [Dhall et al., 2011] has been used for the purpose of this experiment. Each participant watched a total of 80 videos which were divided into seven categories. They are: Anger, Disgust, Fear, Happy, Neutral, Sad and Surprise. All the videos were around 2-3 seconds in length. Participants were asked some general demographic questions at the beginning of the experiment. After watching each video, they were asked to rate the genuineness of the video in a 5-point rating scale ('Completely fake', 'Surface acted', 'Don't know', 'Deep acted', and 'Completely real'). They were also asked to rate their confidence level on their answer using a 5-point scale (1 being not confident at all and 5 being very confident). They were also asked if they had seen the video before or not.

EDA data was collected using Empatica E4 wristband with a sampling rate of 4 Hz. Pre-processing was done by normalising the data using min-max normalisation and smoothing using median filter [Jerritta et al., 2011].

4.1.1.3 Features

A total of 16 features (linear and nonlinear) were extracted from the pre-processed data. They are listed in Table 4.1:

Table 4.1: Features extracted from participants EDA Signals watching emotional videos

Feature Type	Feature Names
Linear features	Mean, root mean square, variance, integrated signals, simple square integral, average amplitude change, log detector, difference absolute standard deviation value
Non-linear features	Hjorth parameters (mobility), Hurst exponent, sample entropy, approximate entropy, Shannon's entropy, permutation entropy, fuzzy entropy, detrended fluctuation analysis (DFA)

4.1.2 Results

4.1.2.1 Mean Analysis

From the set of extracted features, mean values of all participants were chosen in every emotion category and some statistical tests on that data were performed. Mean is the most commonly used statistical feature in machine learning models. Figure 4.1 shows the mean values for the seven emotion categories:

Figure 4.1 shows that participants feel higher and lower cognitive load while watching *surprise* and *happy* videos compared to other emotional categories. To find the differences between emotion pairs, a two-tailed permutation test was performed.

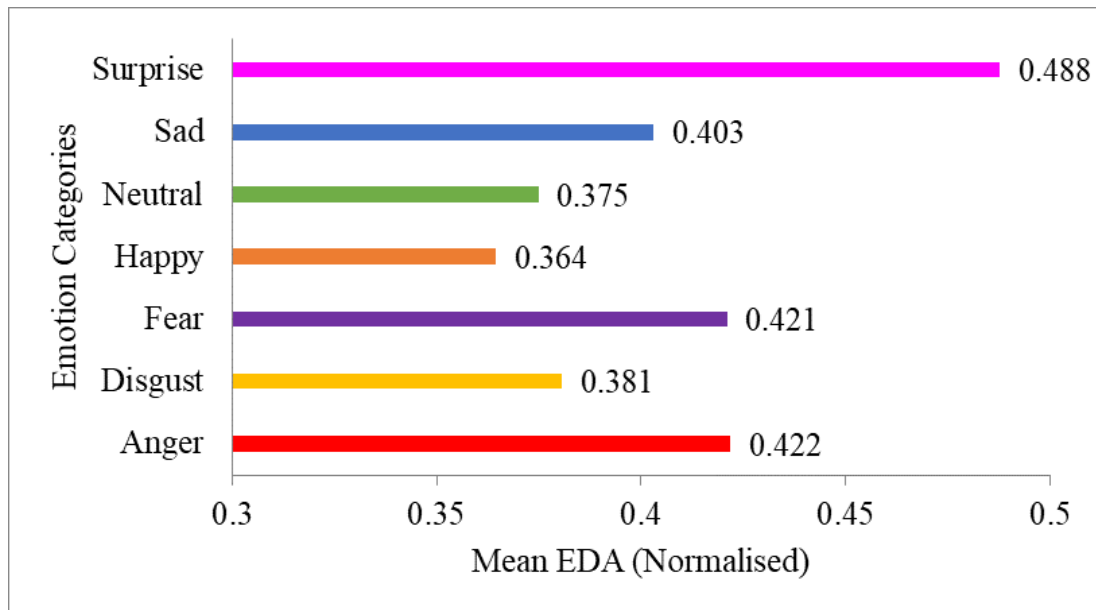


Figure 4.1: Mean values of EDA for seven emotion categories (Range 0.3 – 0.5 chosen for better visualisation)

The test is applied here to identify the time points where EDA is different between two emotions. Over the analysis, four emotion pairs (*disgust-surprise*, *happy-sad*, *happy-surprise*, *neutral-surprise*) were found to significantly differ from one another ($p < 0.05$). Every pair of emotions were also analysed using t-test and the results showed statistical significance ($p < 0.05$) for six pairs of emotions. They are: *happy-sad*, *happy-surprise*, *disgust-surprise*, *fear-surprise*, *neutral-surprise* and *sad-surprise*. Table 4.2 shows the significance value for all pairs of emotions. The numbers in colour and bold are the pairs that show meaningful differences. The red colour shows a significance of $p < 0.05$, while the blue colour shows high significance $p < 0.01$.

Table 4.2: T-test values for all pairs of emotions in identifying seven types of emotional videos

Anger						
Disgust	0.234					
Fear	0.492	0.141				
Happy	0.093	0.203	0.067			
Neutral	0.126	0.407	0.105	0.287		
Sad	0.336	0.157	0.214	0.007	0.157	
Surprise	0.073	0.022	0.029	0.005	0.015	0.021
	Anger	Disgust	Fear	Happy	Neutral	Sad

This relationship can be visualised using an emotion model, which is a two dimensional model based on valence and arousal level of emotions frequently used in the area of affective computing [Russell, 1980]. Valence refers to intrinsic goodness or badness while arousal corresponds to alertness/response readiness. The standard abstract model is shown in Figure 4.2.

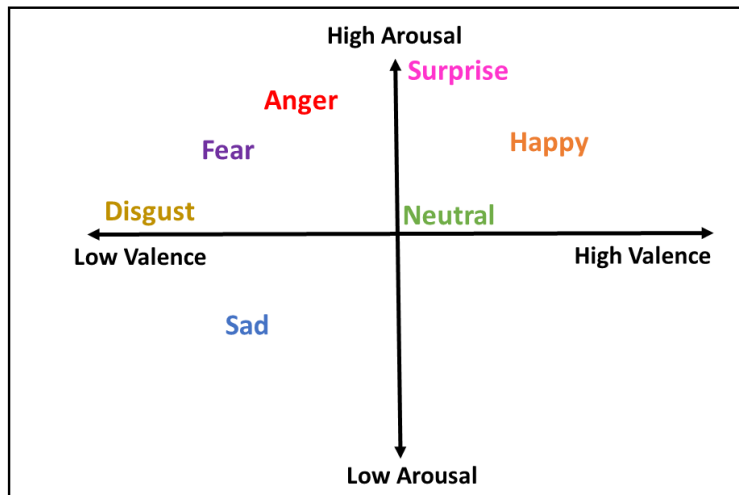


Figure 4.2: Arousal models of emotion: Standard abstract model

The arousal model for the seven emotions using the mean feature values described in 4.1.2.1 is shown in Figures 4.3 and 4.4. These are called arousal modes, because valence values were not considered as a measurement reference. The valence levels are kept the same as in the original model.

From Figures 4.3 and 4.4, it can be seen that *surprise* has a very high arousal value compared to all other emotions. Therefore, it is able to show significant difference with all other emotions except another relatively high arousal emotion of *anger*, and there it is close to being significant ($p = 0.073$). The feature is also able to differentiate between a high valence high arousal emotion (*happy*) and a low valence low arousal emotion (*sad*). Figure 4.3 also shows that if we consider *neutral* as a reference line then *happy* and *sad* are not located in the position proposed for the arousal models widely in the literature. *Happy* is a high arousal high valence emotion and *sad* is a low arousal low valence emotion according to the literature. But based on the data from this study we can see *happy* shows low arousal and *sad* shows high arousal. This is possible from the participants' perspective perhaps because when they are watching sad videos they feel sad, but maybe when watching happy videos (more common than seeing sad videos) they accept them as being normal. Comparing neutral with more emotions in all of the categories will help us to understand this phenomenon in greater detail.

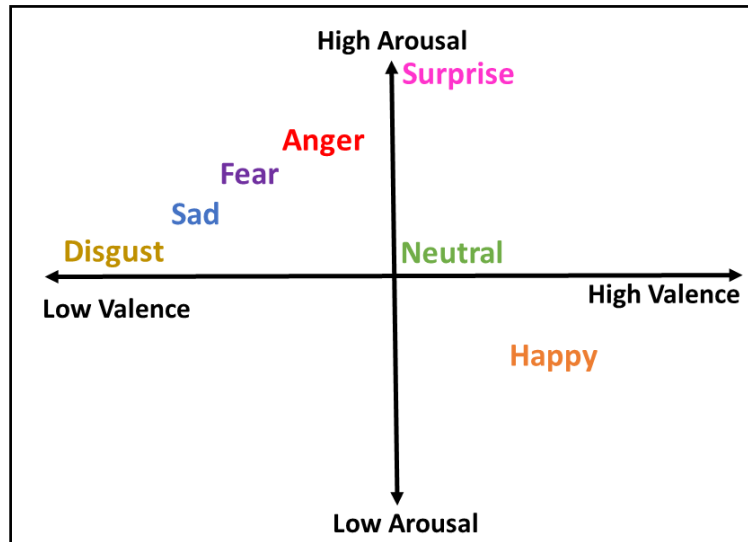


Figure 4.3: Arousal models of emotion: Data derived model (Neutral as reference)

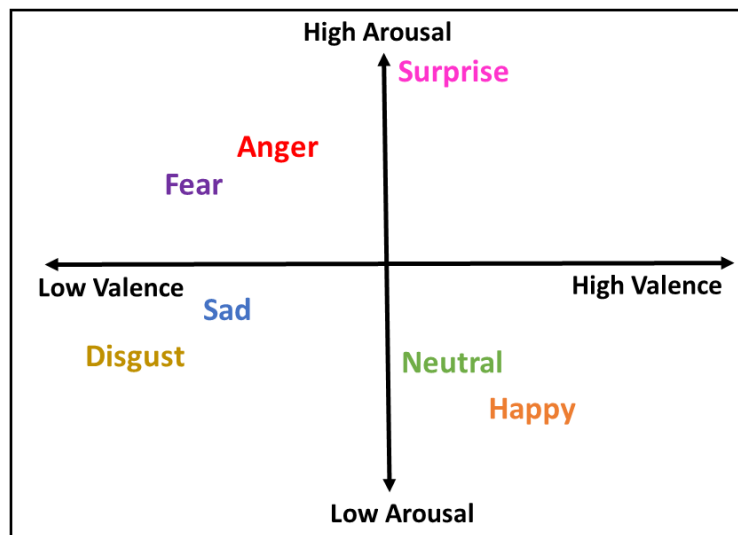


Figure 4.4: Arousal models of emotion: Data derived model (Mean as reference)

4.1.2.2 Classification Results

Classification accuracy of the system is reported as the percentage correctness of the system predicting the video category. Also, the accuracy, f-measure, precision, recall, specificity and geometric mean values are reported. The classification process was done using MATLAB R2018a software with an Intel(R) Core(TM) i7-5200U processor with 3.60 GHz, 16.00 GB of RAM and Microsoft Windows 10 Enterprise 64-bit operating system. The labels were given according to the seven video categories mentioned in the experiment design. A leave-one-participant-out process was performed to distinguish among the seven emotion categories. A simple pattern recognition network was employed, which consisted of one input layer, one hidden layer and one output layer. The hidden layer was constructed using 30 hidden nodes. The model achieves a total of 94.8% accuracy based on the average of 20 runs. Figure 4.5 shows the accuracies of all seven emotional categories.

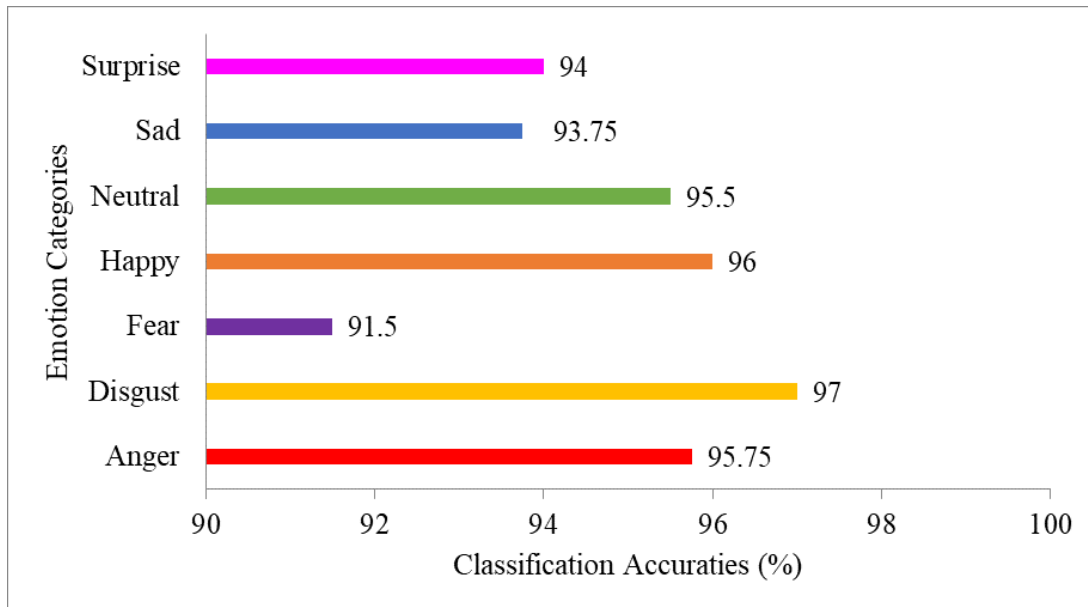


Figure 4.5: Classification performance while participants recognised seven emotions from video (Range 90 – 100 displayed for better visualisation)

All other evaluation measures are shown in Table 4.3. The values are calculated on the average result of all categories. Geometric mean and harmonic mean values are highlighted as they provide more useful information than arithmetic mean when comparing groups having different properties [Hand and Christen, 2018].

From Table 4.3 it is evident that the neural network model achieves high scores in both F-measure and Geometric mean. So this model is effective for this EDA signal based emotion recognition problem.

Table 4.3: Evaluation measures for classifying seven emotional categories from videos

F-measure	Precision	Recall	Specificity	G-mean
0.842	0.753	0.958	0.947	0.952

4.1.3 Discussion

In this preliminary experiment, the effects in participants' EDA activity were investigated while they watched a set of videos comprising of seven different emotional categories. Signals were collected from participants in an experimental setting. Collected signals were normalised, filtered, and then a set of 16 features were extracted. Classification using a simple neural network showed a high accuracy of 94.8% in identifying the seven different emotional video categories. The high accuracy gives motivation to use this dataset in future experiments combined with music stimuli.

The initial analysis further showed some noticeable difference of the data driven arousal model from participants' perspective, when compared to the (abstract) standard models in the literature. The data-derived model with neutral as the baseline is quite similar to the standard abstract model, with the only changes being happy and sad changing sides as low/high arousal. Questions to be answered are how the data upon which these results were obtained differs from the rationale behind the standard models in the literature, and whether the participants were somehow different from the expected population reaction. It is also important to point out that EDA activity can vary according to the difference in stimuli types, participants' age and gender [Gatti et al., 2018]. Arguably, it makes more sense to use the overall average reaction to be the baseline between high and low arousal, which spreads the emotional reactions over a wider range. However, this differs even more from the standard abstract model.

This analysis is crucial to understanding the various issues of determining the ground truth labels of a stimuli. For datasets where ground truth labels are not available, one challenge is to determine the factors that will be used for the labels. Using participants' physiological feature averages could be one possible approach. However, the results show that participants' data driven results can differ from the standard models. Therefore, using participants' physiological features to automatically label the stimuli may not be appropriate for every scenario. A comparison of participants' subjective and physiological response with the ground truth label is necessary to answer this question. This is addressed in the follow-up experiment classifying genuine and posed smiles using participants' physiological responses.

4.2 Experiment 2 - Physiological Responses Detecting Genuine and Posed Smiles

This experiment demonstrates computational techniques to recognise the genuine and posed smiles by sensing participants' EDA activity while watching sets of images and videos. This experiment builds on the findings of the previous experiment and looks into effects of both image and video stimuli. The effects of showing stimuli in pairs versus one at a time is also explored.

4.2.1 Methods

4.2.1.1 Participants

A total of 25 participants volunteered to participate in this study, being 13 male and 12 female. Their mean age was 21.3 with a standard deviation of 2.6. Participants' EDA signal was recorded using the Empatica E4 watch-format device. The signals were recorded at a sampling rate of 4Hz. The study was approved by the Human Research Ethics Committee of the Australian National University (ANU). The participants were asked to sign a written consent form before the experiment, followed by them providing some demographic information. Then they were presented with a series of stimuli and some questions related to the stimuli were asked. All experiments were conducted in the same location in order to minimise any environmental effects. Verbal responses from the participants were collected using an interactive website.

4.2.1.2 Dataset and Pre-processing

The stimuli used in this study were selected from UvA-NEMO database [Dibeklioglu et al., 2012], which consists of stimuli showing genuine and posed smiles. There were a total of four conditions in which participants viewed the stimuli. They are: single image (SI), paired image (PI), single video (SV) and paired video (PV). In SI and SV conditions, participants were presented with a single image or video of smiles and then they were asked if they thought it was genuine or posed. In PI and PV conditions, participants watched images and videos in pairs and they were asked which one of them they thought was genuine. The stimuli that were presented in pairs were from the same smiler. Figure 4.6 shows sample frames from the database in both single and paired conditions.

For both SI and SV conditions 10 stimuli were used, out of which five were genuine and five were posed. For the PI and PV conditions, 20 stimuli were used (10 genuine and 10 posed). Altogether there are 20 stimuli in the single condition and 40 in the paired condition. All participants watched all stimuli. They were order balanced in order to remove sequence bias. The image stimuli were shown for five seconds each while the video stimuli varied from two to seven seconds in length. From the 25 participants' data, four were discarded due to poor or missing EDA

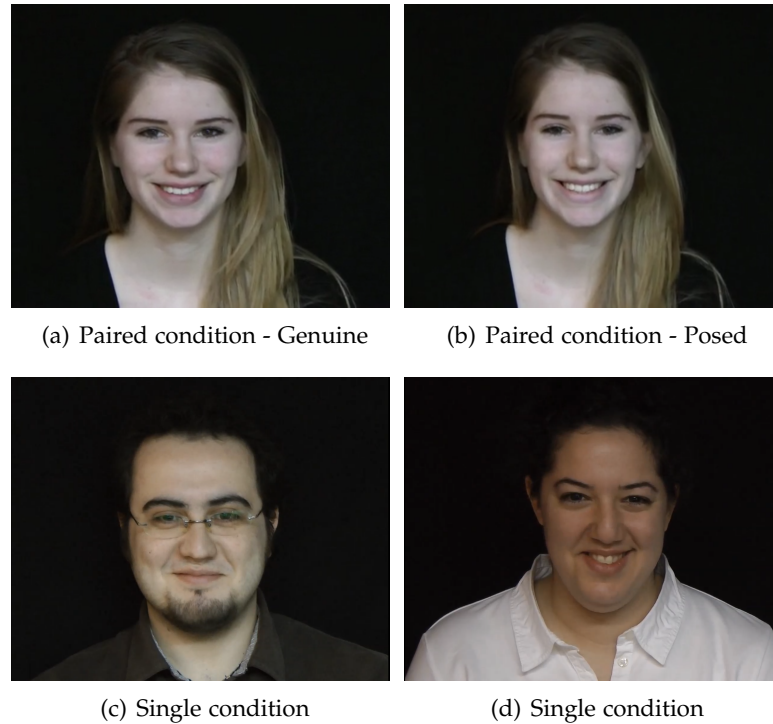


Figure 4.6: Sample frames from UvA-NEMO database

recordings. This sometimes occurred due to the poor connection quality or low charge of the device. Therefore, 21 participants' data were used for further analysis.

Min-max normalisation was used to move the signals within the 0 – 1 range. In order to remove noise from the signals, the median smoothing filter was used [Jerritta et al., 2011]. Then the signals were segmented according to the length of stimuli. For example, the images used in the experiment were shown for five seconds. Thus, the five seconds' data corresponding to a particular image stimulus from the whole data recorded for a single participant were extracted, which is referred to as a 'segment'.

4.2.1.3 Features

A number of time domain and frequency domain features were extracted from the segmented EDA signals. The list of features is shown in Table 4.4.

Twenty five features were initially extracted from each segment of EDA signals. The features were visualised and some features which could not give much insights into the data were removed. For instance, the number of peaks were calculated for each segment. However, due to the small size of each segment, many peaks were not noticed for most segments. Thus, the feature was discarded. Similarly, some features were discarded where the values were mostly zeroes or ones. Furthermore, some

Table 4.4: Features from EDA signals watching genuine and posed smile stimuli

Feature Type	Feature Names
Time domain features	Mean, minimum, maximum, standard deviation, variance, root mean square, summation, absolute summation, simple square integral, mean of first and second difference of normalised signals, Hjorth parameters (mobility), simple square integral, log detector
Frequency domain features	Mean, minimum, maximum of the first 16 points from Welch power spectral density

redundant features were removed (e.g. summation and absolute summation as both yielded positive value), and this resulted in a total of 16 features from each segment which are listed in Table 4.4.

4.2.2 Results

4.2.2.1 Classification Results

Selected features from the segmented EDA signals were then trained using three different classification techniques. They are, decision trees (DT), K-nearest neighbor (KNN) and bagged trees (BT). The methods are described in sections 3.2.5.2, 3.2.5.1 and 3.2.5.3 respectively. Cross validation was done using a leave-one-participant-out method. Training and testing were done in both single and paired conditions. The prediction accuracies were compared with both the labels provided by the database and participants' verbal response labels. In addition to accuracies, precision, recall, specificity and f-measure were also calculated as evaluation measures. The complete data analysis was done using MATLAB® R2020a software with AMD Ryzen 7 3700X 8-Core Processor with 3593 Mhz, 16.00 GB of RAM and Microsoft Windows 10 Home 64-bit operating system.

Classification accuracies obtained from participants EDA signal features using KNN, DT and BT methods are reported in Table 4.5. The first three rows correspond to the accuracies using the labels provided by the database (objective ground truth). The following three rows are results obtained by comparing with participants' verbal response labels (subjective 'ground truth').

Table 4.5 shows that the highest accuracy is achieved by DT in a paired condition (93.6%). Initially the models were trained using the entire dataset and it was seen that the accuracy is higher in the single condition (91.4%) in comparison to paired condition (90.9%). However, it was postulated that this could be due to the imbalance in the dataset. As described in section 4.2.1.2, a total of 20 stimuli in single condition and 40 in the paired condition were used. Therefore, a subset of 20 stim-

Table 4.5: Classification accuracy differentiating genuine and posed smiles using EDA signals

Condition	Label	KNN	DT	BT
Single	UvA-NEMO label	59.0%	91.4%	90.9%
Paired	UvA-NEMO label	68.1%	90.9%	86.1%
Paired (Subset)	UvA-NEMO label	71.9%	93.6%	88.6%
Single	Participant's verbal response	48.3%	56.2%	52.1%
Paired	Participant's verbal response	60.2%	58.6%	62.1%
Paired (Subset)	Participant's verbal response	60.9%	59.8%	65.2%

uli (10 genuine and 10 posed) was randomly chosen from the paired condition to train the classifiers. Results from Table 4.5 demonstrate that the accuracy is generally higher than the single condition for all three classification methods. Thus, the results indicate that the paired condition is better than the single condition for observers to distinguish between genuine and posed smiles.

Classification results of the other evaluation measures were also higher in the paired condition in comparison to the single condition. Precision, recall, specificity and F-measure scores of the DT model in the paired condition were 93.8%, 93.3%, 93.8% and 93.6% respectively. The same evaluation measures in the single condition were 92.2%, 90.5%, 92.4% and 91.3%. The high scores of our model in the evaluation measures confirms the results and suggests that stimuli shown in pairs is a more beneficial approach than showing as a single stimulus.

4.2.2.2 Timeline Analysis

In order to understand how well the participants' EDA signals are distinguishing between genuine and posed smiles, a timeline analysis was performed on the pre-processed EDA signal averages in both single and paired conditions. The signals were reshaped to have the initial value 0.5. This initial value was chosen in order to clearly observe the increasing or decreasing trend of the EDA signals. The values were selected from the first five seconds of each stimuli as this is the length of most stimuli used in the experiment. The results are shown in Figure 4.7 and 4.8. Red shaded area displays EDA signals when participants are watching genuine smiles and blue shaded areas correspond to watching posed smiles.

It can be seen that participants' EDA signals display some differences in range from the first second of watching a single genuine and posed stimulus. The dif-

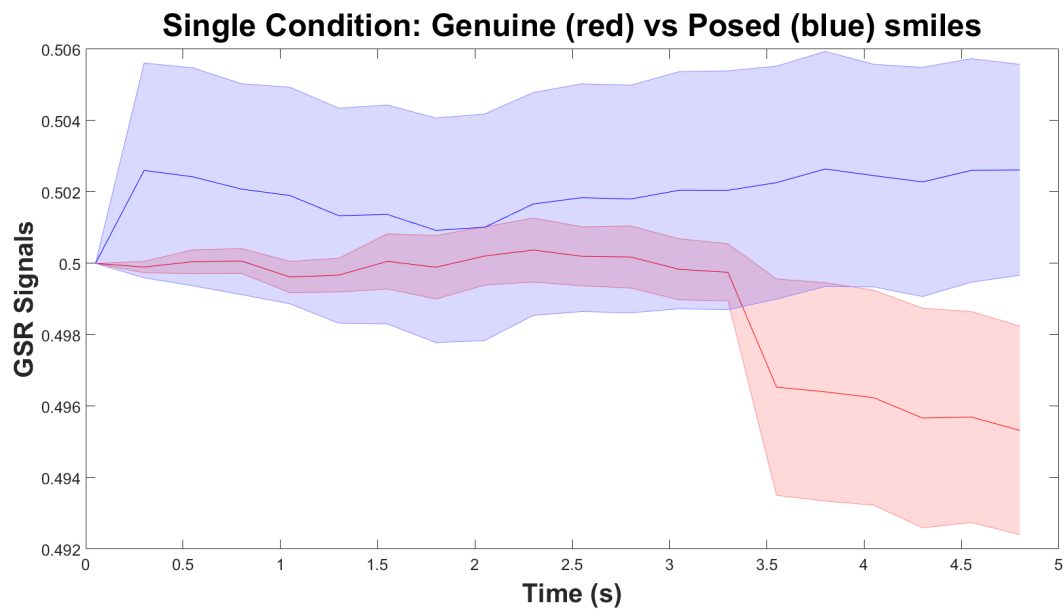


Figure 4.7: Timeline analysis of EDA signals - single condition

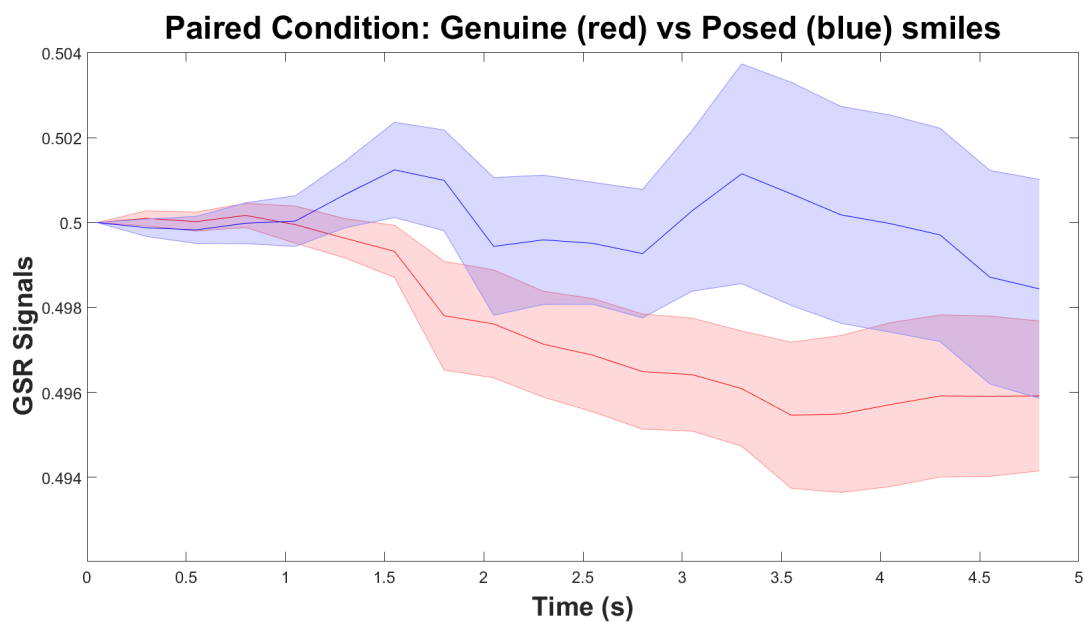


Figure 4.8: Timeline analysis of EDA signals - paired condition

ference becomes larger after 3.5 seconds, which suggests that participants' response becomes clearer after watching the stimuli for a few seconds. In the case of paired stimuli, the EDA signals are quite similar for both genuine and posed smiles in the first second, but starts diverging from each other quickly after that. After three seconds, both signals show larger difference in their values. In both cases, participants' EDA signals are able to distinguish between genuine and posed stimuli after three seconds. This finding is useful in designing future studies in this area, where shorter length stimuli should not be used. It is also found that EDA signals are higher when they are presented posed stimuli, which shows that their subconscious reactions to posed stimuli is stronger compared to genuine stimuli, as found in another study considering pupillary responses [Hossain et al., 2016]. The significance level of EDA signals in both single and paired conditions was further verified using a Kolmogorov–Smirnov (K-S) test. The results showed a statistically significant difference ($p < 0.05$) between genuine and posed smile signals for both conditions.

4.2.2.3 Comparison with Participants' Verbal Response

Although participants' EDA response aligned very well with the objective ground truth information, it did not align well when compared with their individual verbal response. We can further see from Table 4.5 that the highest accuracy obtained using participants' own verbal responses was 65.2% with the BT method, which is above chance only to a limited extent. In the single conditions, the accuracy was above chance using the DT method (56.2%), and lower than chance (48.3%) with the KNN method. This indicates that participants perform poorly in discriminating genuine and posed smiles when they are watching only one stimulus at a time. However, when shown in pairs, they perform slightly better. Overall, the analysis shows that participants' subconscious EDA signals are better in recognising genuine and posed smiles when compared to their *own* verbal responses.

4.2.3 Discussion

This experiment presented a preliminary study to collect and analyse participants' EDA signals while they watched stimuli containing genuine and posed smiles in single and paired conditions. Experimental results showed that participants' EDA signals were 91.4% accurate in differentiating between genuine and posed smiles when they watch a single stimulus, and 93.6% accurate when they watch stimuli in pairs. This study also revealed that participants' verbal responses perform poorly compared to their physiological responses, as their correct response rate was only 56.2% and 59.8% accurate in single and paired conditions respectively. The effect was also evident in that the verbal response labels were harder to classify by any of the methods, with a maximum of 65.2% achieved. It is clear that the paired stimuli helped all techniques: the computational methods on ground truth (UvA-NEMO) labels was helped the least, followed by the verbal response, while the AI tools on the verbal response labels were improved by 9.8% on average in paired condition.

This suggests that the improvement is due to humans finding comparisons easier than absolute judgements.

The disparity between participants' verbal and physiological response is quite interesting, but the collected data is not enough to confirm the low reliability of verbal responses. More data needs to be collected, especially detailed verbal responses from participants in order to understand what difficulties they faced to identify the different smile types. Nevertheless, these preliminary findings provide evidence of the usefulness of participants' subconscious response in differentiating between genuine and posed emotions. This extends the finding of the previous experiment and provides further motivation to use participants' physiological response to automatically label video stimuli, in the absence of ground truth labels.

In addition, this study revealed usage of paired stimuli to be more effective than using single stimuli. This phenomena has not been explored before, even though there have been studies that showed significant differences in emotional and physiological response watching genuine and posed stimuli [Hossain et al., 2016; Aracena et al., 2015]. This finding can be useful in grouping stimuli in future experiments to invoke stronger physiological responses. In the next experiment, the video stimuli used in this experiment and previous experiment is combined with some music stimuli to understand the combined effect of two different stimuli.

4.3 Experiment 3 - Effects of Music in Detecting Genuine and Acted Emotions

In this experiment, the effects of six different music stimuli on people's affective reasoning is explored using their verbal and brain activity responses. Based on the outcomes of the previous two short studies, a broader third user study was designed. A study was conducted where EEG signals from different brain regions were collected when participants listened to these stimuli and identified different emotions from video stimuli. These signals were then processed and analysed using statistical and machine learning techniques to classify the video stimuli emotions. In addition, comments on the different music stimuli were collected from the participants in order to understand how the music influenced their performance on their given task.

4.3.1 Methods

Six different music stimuli were used from the list described in Table 3.1. These stimuli can broadly be divided into three categories. They are: binaural beats, classical music and pop music. The reasons for choosing these six stimuli are briefly described in detail below:

- "Brain Energizer - Gamma Waves for Focus / Concentration / Memory - Binaural Beats - Focus" - This stimulus is said to increase gamma wave activity on the brain. According to the description of the video, this stimulus is highly beneficial for studying and improving focus during work.
- "Serotonin Release Music with Alpha Waves - Binaural Beats Relaxing Music, Happiness Frequency" - This binaural beat stimulus is a relaxing piece which is said to boost alpha wave activity on the brain.
- "F. Chopin's 'Funeral March' from Sonata in B flat minor Op.35/2" - This is one of Chopin's most popular pieces and has a very sombre tone to it.
- "J.S Bach's 'Air' from Suite for Orchestra No. 3 in D" - This is an orchestral music piece with a moderate intensity and can be considered a more relaxing piece.
- "Justin Bieber's 'Love Yourself' from the album Love Yourself" - This music piece was chosen as the top song in the 2017 Billboard Hot 100 year-end chart. It is classified as an acoustic pop song, having a moderate valence level (Spotify valence score 0.515).
- "Ed Sheeran's 'Shape of You' from the album Divide" - This song was chosen as it was the top song according to the Billboard Hot 100 year-end chart of 2018. It is an upbeat and energetic song having a high valence level (Spotify valence score 0.931).

Only the category of music was considered for the computational analysis. The valence scores were not considered as this score was not known for some of the stimuli.

Video stimuli used in this experiment were taken from four different datasets. All of these datasets contain videos of people displaying different types of emotions in genuine or acted form. These are, AFEW, MAHNOB-HCI, MMI, and the Anger dataset. The datasets are explained in detail in section 3.1.3.2. A total of 48 video clips were used in this experiment, 24 of them labelled as genuine emotion, and 24 labelled as acted. The six types of basic emotions were represented in the videos. These emotions were contained in a blended manner in the videos to reflect the blended nature of real world emotions Kim and André [2008]. The video clip lengths ranged from one second to four seconds. However, during analysis, all physiological signal recordings were cropped to the same length for comparison. All the clips were cropped to the same width, height and converted to grayscale. As one of the datasets contained only grayscale videos, the other videos were also converted to eliminate any effects of video colour.

4.3.1.1 Participants

A total of 22 participants (nine female and thirteen male) participated in this study. Their mean age was 20.8 with a standard deviation of 4.8. All participants were university students and they were recruited through the University's voluntary research participation scheme website. They were given participation credit after completion of the study.

4.3.1.2 Experiment Design

The experiment was approved by the Australian National University's Human Research Ethics Committee. After arriving at their scheduled time, the participants were welcomed and given a participation information sheet and consent form. They were briefed on the aim of the study and asked to read through the information sheet. Following their agreement to participate in the experiment (through the signed consent form), they were asked to sit in front of a monitor and asked to adjust the seat so that they had a comfortable view and easy access to the keyboard and mouse. Then the EEG device was fitted. All the participants completed the experiment in the same experiment lab, and the room temperature, lighting were kept consistent for all of them. A photo taken during the experiment is shown in Figure 4.9.



Figure 4.9: A photo of the experimental setting - participant is listening to different music while watching genuine and acted emotion

EEG data was collected using the Emotiv EPOC headset device. In order to set up the device, the electrodes are first hydrated using a conductive gel and then fitted on the participants' head. After ensuring good connectivity between all the electrodes and the scalp, participants' baseline EEG values were collected. This was done by asking participants to keep their eyes open for 15 seconds, then keep their eyes closed for 15 seconds. Raw data from the device was collected at a sampling rate of 128 Hz, while the band power data is collected at a sampling rate of 8 Hz. The data was recorded using the Emotiv Pro Software.

In the final step before starting the experiment, participants were asked to wear a pair of Bose QuietComfort® 20 Acoustic Noise Cancelling™ headphones to ensure participants were not affected by any outside noise. Participants' verbal responses and comments were collected through an interactive website. At the start of the experiment, participants were asked some pre-experiment demographic questions such as their age, gender and music preferences. They were also asked if they had experience in playing musical instruments.

After completing the pre-experiment questionnaire, the experiment began. All the participants listened to the six music pieces. Each music piece was played for two minutes. The music pieces were order balanced to reduce bias which may occur due to the presentation order. While each music piece was playing, participants watched short video clips showing people displaying different emotions. The videos were order balanced as well. Then they were asked the following question, "How does the expression presented in the video look to you?". This was a closed question where the two options below were given:

- *Genuine Emotion* means the dominating emotion this person experienced is genuine
- *Acted Emotion* means this person acted the emotion

These two options were given based on the experience from experiment 4.2, where participants had a lot of difficulty understanding the concept of deep and surface acting (and therefore using these labels while rating). Therefore, these options were removed to reduce further complication. They were further asked whether they had seen this video clip before. Most of the participants said that they had not watched the clip before. Their verbal and physiological responses can be considered to be free from prior bias on the videos. After each music stimulus finished playing, participants were asked an open ended question to provide comments on the music.

4.3.2 Results

4.3.2.1 Findings on Music Stimuli

A qualitative analysis was performed on the comments participants provided on the music stimuli by using a grounded theory approach [Glaser and Strauss, 2017]. NVivo 12 software was used to complete this analysis. Memos on NVivo were used for coding participants comments into higher level themes. The comments were divided based on how participants described what they felt while listening to the music. These codes were then divided into three categories: positive, neutral and negative. During the coding process, frequently appearing words that were considered negative were: "dislike", "depressing", "irritating", "disturbing". Some of the comments highlighted as positive were: "like", "calm", "relax", "soothing". The neutral comments mostly described some features about the music, or whether they heard the song or not, and the comments did not reflect participants' emotions. Some of the common words used for neutral comments were: "slow", "fast", "know" and "familiar". Table 4.6 shows the percentage of participants providing different types of comments on the stimuli.

Table 4.6: Type of comments provided by participants on each music stimuli

Stimuli	Negative	Neutral	Positive
Binaural Beats 1 (Brain Energizer)	45.5%	22.7%	31.8%
Binaural Beats 2 (Serotonin Release)	18.2%	18.2%	63.6%
Classical 1 (Funeral March)	36.3%	22.7%	41.0%
Classical 2 (Air)	18.2%	0%	81.8%
Pop 1 (Love Yourself)	9.1%	13.6%	77.3%
Pop 2 (Shape of You)	13.6%	18.2%	68.2%

All 22 participants provided comments on the six stimuli. A total of 132 comments were analysed to extract some general themes that were prevalent. The analysis of participants' verbal and physiological response (described in section 4.3.2.2) aligns with the themes presented in the points below. Participants are mentioned as P1, P2... P22.

- Binaural beats inducing gamma waves cause discomfort and distraction - The stimulus *Binaural Beats 1* (Brain Energizer) was designed to improve the level of concentration and focus of the brain. Thus the expectation was that it will help participants with the video emotion classification task. However, surprisingly, this stimulus received quite a few negative comments (45.5%) and was said to cause discomfort among participants. One participant (P19) mentioned that they "...disliked the continuous tone underlying, became quite irritating". Another participant (P17) added, "...Background droning noise was off-putting. Sounded

sci-fi like". This was unexpected and also resulted in a lower classification result, which is described in detail in section 4.3.2.2. *Binaural Beats 2* (Serotonin Release Music with Alpha Waves) on the other hand, received the expected reaction from the participants. It was meant to increase alpha waves on the brain and promote relaxation, it often made the participants too calm to focus on the given task. According to participant P14, "...it was a slow piece and made me feel sleepy and hindered with concentration". Another participant (P17) mentioned, "I liked this piece, it was very calming. Sounded like mindfulness/meditation music".

- Classical music having a sombre tone helps increase focus and answer questions - The stimulus *Classical 1* (Funeral March) is played in a minor key and has very slow recurring patterns and accents. Therefore, the stimulus in general should invoke sad emotions. This was aligned with the participants' comments on this stimulus. However, although some participants reported that the music stimuli made them feel sad and depressed, they also thought it had a calming effect and therefore helped them focus in identifying the emotions from the video stimuli. For instance, P20 commented about this stimuli by saying, "...this piece is a bit sad, but it helps while answering questions". P2 said, "...feels heavy to listen but like". This resulted in this stimuli receiving more positive comments than negative (36.3% negative comments, 41.0% positive comments), which was mildly surprising. This theme was also observed in the better classification results of participants' verbal and physiological responses using *Classical 1* stimulus which is described in section 4.3.2.2. In comparison, *Classical 2* (Air) mostly received positive comments and was said to have a very relaxing effect. One participant (P7) mentioned, "...the quiet piece playing in the background makes it nice and easy to hear". Another one (P17) said, "I loved this piece. The music was smooth and rich. The violin was beautiful and very inspiring". Although this piece was generally liked by the participants (81.8%) and induced a relaxing effect, it did not contribute as much to the video emotion classification task compared to the previous stimulus, and had a similar effect as *Binaural Beats 2* on the classification accuracy.
- Pop songs received positive feedback due to familiarity, but that can also be distracting - Both music stimuli in the pop category received mostly positive comments (Stimulus 1 and 2 received 77.3% and 68.2% positive comments respectively). In addition, both stimuli were familiar to the participants (21 out of 22 participants were familiar with at least one of the songs). This may have added certain bias from the stimuli or the artists, which may have affected their task performance. Commenting about *Pop Stimulus 1* (Love Yourself by Justin Bieber), one participant (P11) mentioned this piece as, "...familiar and predictable", while another (P17) said, "I liked this piece because it was soothing and I listen to it frequently, it has a calming vibe". *Pop 2* (Shape of You by Ed Sheeran) generally received favourable comments such as "I like this song, probably one of my

favourites. It makes me somewhat dance to the beat" (P1). Another participant (P14) said, "*...upbeat tune and heard a lot so made me feel comfortable"*. However, the familiarity and biases may have caused some mixed outcomes in the classification tasks and therefore these aspects need to be analysed in greater detail.

4.3.2.2 Verbal and Physiological Data Analysis

The results from participants' verbal and EEG responses showed a strong correlation with the comments they gave on the music stimuli. Video labels provided by the original datasets were used as the ground truth and were compared with participants' verbal and physiological responses. For the analysis reported in this chapter, out of the 14 channels of Emotiv EPOC, data from two channels were chosen, namely F7 and F3. Both of these channels are located on the frontal lobe. They have been connected to emotion processing [The Human Brain and Seizures; Salzman and Fusi, 2010] and decision making [Collins and Koechlin, 2012]. These channels have also been shown to contain the most useful features for classifying different types of music using participants' EEG signals [Rahman et al., 2020]. Emotiv captures data on five frequency bands. They are: theta, alpha, low beta, high beta and gamma. For this study, alpha and gamma band power data were chosen as features.

Unlike the previous two experiments, raw data was used for computational analysis instead of feature data. The data were first normalised using min-max normalisation. Then they were segmented according to the length of videos. Finally, baseline correction using polynomial fitting was applied in order to remove environmental noise from the signals. These segments were then used to classify genuine and acted emotion labels using a one dimensional convolutional neural network (1D CNN) network. The CNN has the advantage of automatic feature extraction, which eliminates the process of handcrafted feature extraction and reduces computational complexity. Due to the size of data being very small, a simple architecture was created with two convolution layers, two maxpool layers, one dropout layer and two dense layers. A 5-fold cross validation approach was used by randomly splitting 80 percent data for training and 20 percent data for testing. For evaluation measures, the classification accuracy and F1-score is reported.

For this analysis, all genuine emotion videos and all acted emotion videos were combined together to feed into the classifier model, instead of classifying the videos individually. This was done because some of the videos were too short in length (less than one second) which resulted in very small numbers of samples. This was not enough for the classification model. Thus, they were only classified into genuine or acted emotions, making the analysis a binary classification problem. The results of participants' verbal and EEG response classifications are shown in Figure 4.10 below:

From Figure 4.10 it can be seen that the highest accuracy using participants' verbal and EEG response were 62.5% and 68.6% respectively and both were achieved

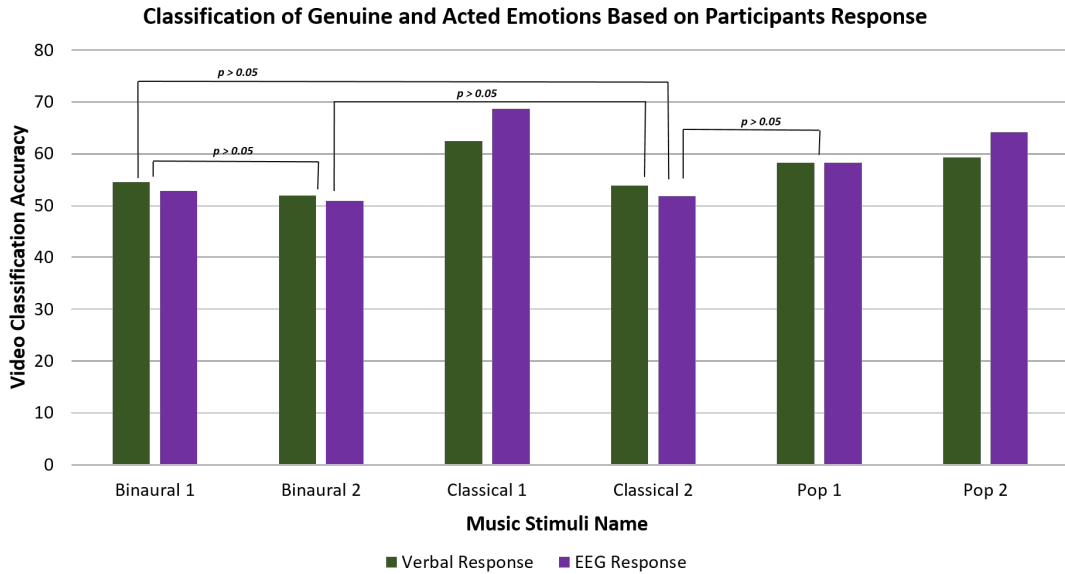


Figure 4.10: Classification results using participants verbal and physiological response in detecting genuine and acted emotions while listening to different music

when participants listened to the *Classical 1* stimulus. The F1-score using *Classical 1* stimulus is 0.69, which is also the highest out of all conditions. This suggests that music stimuli that can invoke a sad emotion also help in centering focus on identifying genuine and acted emotion from videos. The next best performance was achieved using *Pop 2* stimulus, achieving 64.1% accuracy and 0.64 F1-score using participants' EEG response. This stimulus was also liked by participants due to its familiarity and liking for the artist. It is also worth noting that this stimuli has a high valence score (mentioned in section 4.3.1). In contrast, *Binaural Beats 1* achieved low accuracy of 52.8% and 0.53 F1-score in detecting genuine and acted emotions. This result is contrary to what has been widely suggested about gamma music helping with focus. *Binaural beats 2* which induces alpha waves in the brain showed the expected outcome as participants performed poorly when listening to this piece (51.9% and 50.9% accuracy using verbal and EEG response respectively). Other stimuli which also promoted high levels of relaxation such as *Classical music 2* also achieved low accuracy (53.8% and 51.8% accuracy using verbal and EEG response respectively). This is expected as these stimuli were meant to promote relaxation and therefore mostly used for mediation and sleep studies [Vijayalakshmi et al., 2010].

A further observation from the results is that, in three out of the six cases, participants' EEG response performed better than their verbal response in classifying between genuine and acted emotion. Although physiological responses did not perform better in all cases, this aligns with previous studies which showed participants' physiological response performs better than their self-reports in recognising emo-

tions from videos [Soleymani et al., 2011b]. This also aligns with the findings from experiment 4.2 reported in this chapter. It should be noted that two out of the three stimuli where participants EEG response resulted in lower accuracy were *Classical 2* and *Binaural Beats 2*. Both of them had a relaxing effect on the participants, which hindered their task performance. The other stimulus was *Binaural Beats 1*, which caused discomfort in many participants, resulting in poor task performance. The results suggest that three stimuli, *Classical 1*, *Pop 1* and *Pop 2*, helped participants keep their focus during their task, and this was reflected better through their EEG response. A one-way ANOVA test was also conducted among all the cross-validation results across the six stimuli. The result shows high statistical significance ($p < 0.01$). A paired sample t-test was further performed for all pairs of stimuli results. The pairs that were not statistically significant were *Classical 2* with *Binaural 1/Binaural 2/Pop 1* and *Binaural 1 - Binaural 2* (highlighted in Figure 4.10). The other pairs were significant. This provides the motivation to explore this study in greater detail to understand and identify the effects of different music stimuli in participants' physiological response in detecting different type of emotions.

4.3.3 Discussion

In this experiment, a study was conducted to identify what types of music stimuli are beneficial to help improve concentration while identifying genuine and acted emotions from short video clips. Participants' EEG and verbal response were collected and analysed. A grounded theory approach was applied on participants' comments in order to understand their emotional reaction to the music stimuli. Then participants' performance in the experiment task of identifying emotions from videos were analysed and compared using their verbal and EEG responses. An additional analysis was done to compare the outcomes of the grounded theory approach on participants' verbal comments and their performance in the experiment tasks. The results show that classical music possessing a sombre tone increases concentration on the brain and helps participants identify different emotions from video clips. Familiar and popular music with high valence also helps improve participants' focus. The study also reveals a crucial outcome related to binuaral beats which are believed to improve focus on the brain. The experimental results showed that certain binaural beats can also cause discomfort to participants, which results in disrupting focus in participants and thus achieving lower accuracy in detecting emotion veracity. This study further shows that participants' EEG response perform better than their verbal response in identifying emotion from videos, when incorporated with a music stimuli that increases their focus on the task.

The study revealed several limitations of the stimuli and computational methods that were considered so far. Due to the limited number of participants, the dataset was quite small and therefore the collected signals were not sufficient to train a suitably deep network. Another limitation was the length of the video stimuli, which were too short to evoke a strong reaction from participants, especially combined with

music stimuli. Both of these caused lower accuracy compared to the previous two experiments. Participants also reported this when we asked if they had any general comments on the experiment. P1 said, "...especially the videos were extremely short, it made it more difficult to focus and evaluate the emotions as to whether they're genuine or acted". This aligns with the findings reported in section 4.2.2.2, where we saw that it takes a few seconds for participants' physiological response to show differences while watching genuine and acted stimuli. As some of the video stimuli used in this experiment were shorter than that, participants could not show a strong reaction differentiating the underlying emotion in the short stimulus videos. Thus, we can conclude that the chosen video stimuli were not appropriate to combine with music stimuli. It is also important to consider whether the underlying emotion of the video aligns with the emotion of the music stimuli. If they were widely different from each other, it may confuse the participant, which may reflect in their physiological response. This was reported by some of the participants' in their verbal response. For instance, P15 mentioned, "...when the tension of music and video did not match each other it made me confused". For similar experiments in the future, the emotion and length of the chosen video stimuli need to be considered carefully.

Nevertheless, the results are promising and show the usefulness of different types of music stimuli in facilitating emotion recognition tasks. The use of binaural beats, specifically gamma inducing beats deviated from the expected outcome. This demands a further investigation into similar types of music stimuli and the use of physiological signals to identify which music stimuli are truly beneficial to improve concentration on various kinds of tasks, rather than just assumed to be.

4.4 Summary

In this chapter, three different experimental results were reported which investigated participants' physiological reactions to image, video and music stimuli. A number of different computational models were also explored, ranging from traditional machine learning methods to a shallow and deep neural network. The results showed promising results using image and video stimuli. However, combining them with music stimuli did not result in impressive outcomes using computational methods. This is more likely due to the music stimuli and given task not aligning properly which confused the participants. In the following chapters, experimental analysis are conducted using only music stimuli to see whether patterns can be identified in participants' physiological responses when they listen to different types of music. The results of the three experiments discussed in this chapter further showed that deep learning techniques performed relatively worse with the (small) amount of data that were collected. For experiments where the number of collected samples are smaller, different feature extraction and feature selection techniques are explored in greater detail and are combined with traditional machine learning methods.

Effects of Music in Physiological Response

This chapter explores the effects on human physiological signals while listening to three different genres of music. The focus of this chapter is to investigate the effects of music stimuli alone, and collecting a broader range of physiological signals (EDA, BVP, ST, PD, EEG), to gather more samples and create robust models. In this chapter, effects on participants' EDA, BVP, ST and PD signals are discussed. Effects on EEG responses are discussed in chapter 6. More demographic and qualitative information were collected to do further analysis on the data. This chapter builds on the results published in the 2019 International Joint Conference on Neural Networks – IJCNN 2019 [Rahman et al., 2019] and Journal of Artificial Intelligence and Soft Computing Research - JAISCR [Rahman et al., 2021a]. In both publications, I was the primary contributor to the work.

5.1 Experiment Design

All participants were recruited through the ANU Research School of Psychology's "Psychology Research Participation Scheme" website SONA. After arrival at the lab, they were briefed about the experiment procedure and handed a participation information sheet with detailed instructions. These documents are attached in appendix B. After they understood the procedure and agreed to participate in the experiment by signing a written consent form, they were asked to sit comfortably in a chair in front of a 17.1 inch monitor. Next, participants were fitted with a wrist borne Empatica E4 device which collects EDA, BVP and ST data. Participants were asked to wear the device on their non-dominant hand. After turning the device on, the first 40 seconds of data were considered as the baseline value before the experiment recording began. Both EDA and ST data were collected at a sampling rate of 4 Hz, while BVP data was collected at 64 Hz. Next, the Eye Tribe device was placed in front of the participants to capture their eye movements and pupil diameter. Sometimes the device needed to be moved around to identify the optimal position as it varied based on participant's height and distance from monitor. Then, calibration was done using the device's software development kit. Participants were asked to follow a series of

dots that appeared on the screen. After the calibration, PD data was collected at a sampling rate of 60 Hz. Figure 5.1 shows a photo of the calibration process.

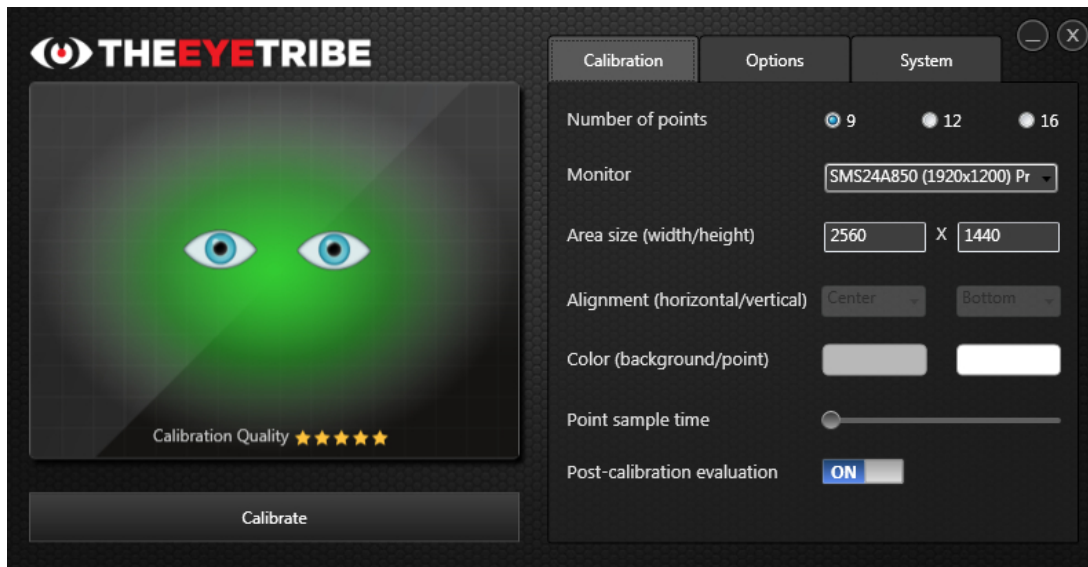


Figure 5.1: User interface of The Eye Tribe for calibration process

Due to the device being sensitive to external movements, all participants were asked to limit any unnecessary movement during the experiment in order to avoid adding artefacts to the signals. They were also asked to wear noise cancelling headphones (Bose QuietComfort® 20 Acoustic Noise Cancelling™) which helped remove any effects from outside noise during the experiment. The entire experiment was conducted through an interactive website prepared for this purpose. The website was created using Python’s Django web framework.

The stimuli used in this experiment are all the twelve music pieces outlined in Table 3.1. Participants answered some basic demographic questions at the beginning of the experiment. These questions included their name, age, gender, ethnicity, musical preferences and whether they had migraine or severe headache problems. Then the participants listened to each piece of music and gave a series of ratings to the music based on six different emotion scales. The scales were decided based on the work done by Walker [1977]. These scales are i) *sad* → *happy*, ii) *disturbing* → *comforting*, iii) *depressing* → *exciting*, iv) *unpleasant* → *pleasant*, v) *irritating* → *soothing*, and vi) *tensing* → *relaxing*. The first four ratings were to find participants’ general impression about the music itself, and the other two asked about the participants’ feelings while listening to that piece of music. The subjective ratings were based on a 7-point Likert scale, chosen as this is considered the most appropriate number for Likert [Alwin, 1997]. At the end of the ratings, participants provided some general comments about the music piece they just listened to.

All the music stimuli were played for the length of the piece itself. Each participant listened to two out of the three genres of music. As each genre contained four pieces of music stimulus, each participant listened to a total of eight pieces of music. The decision to use eight pieces of music instead of all twelve is due to some observations from a pilot study conducted prior to this experiment. Three participants did a pilot study where they listened to all 12 pieces of music. This caused the overall experiment participation time to be close to 120 minutes. The long duration of sitting in one position wearing multiple devices caused fatigue and headache to the participants. Therefore, it was decided to shorten the experiment by reducing four music stimuli (one genre) presented from the dataset for each participant.

The music stimuli were order balanced based on genre using the Latin square method, and within the genre, the music stimuli were played in a fixed order. When participants listened to each pieces of music, they were also given a short article to read from the New Scientist magazine [NewScientist]. This was done so that participants did not get bored or distract themselves thinking about other things and remained concentrated on the experiment. However, there were no tasks involved regarding the content of the text.

In order to analyse the emotional ratings provided by the participants, the ratings were visualised based on their valence-arousal level in a two-dimensional emotion model. The original model was proposed by Russell which contained a wider list of emotions (shown in Figure 2.2). Based on that model, an updated model was created with the six emotion scales used in this study. This is a more effective approach than modeling the emotions with discrete labels because real world stimuli induce blended emotions, and they can be expressed better in a multidimensional space [Kim and André, 2008]. Figure 5.2 shows the emotion model of this experiment.

A photo of the experimental setup is shown in Figure 5.3. The experiment ran for approximately 60 - 90 minutes, which includes device setup time.

5.2 Participants

Demographics of the participants of this study are shown in Table 5.1.

Thirteen male and eleven female students (24 in total) participated voluntarily in this experiment. The mean age was 21 years old with a standard deviation of 4.6. Among the participants 19 were undergraduates while five were postgraduate students. Some of the students had experience in playing different instruments, but none of them were professional musicians or music students. Participants were also asked about their music preference. Their responses varied widely from classical, pop, instrumental, rock, hiphop, metal and folk. The study was approved by the

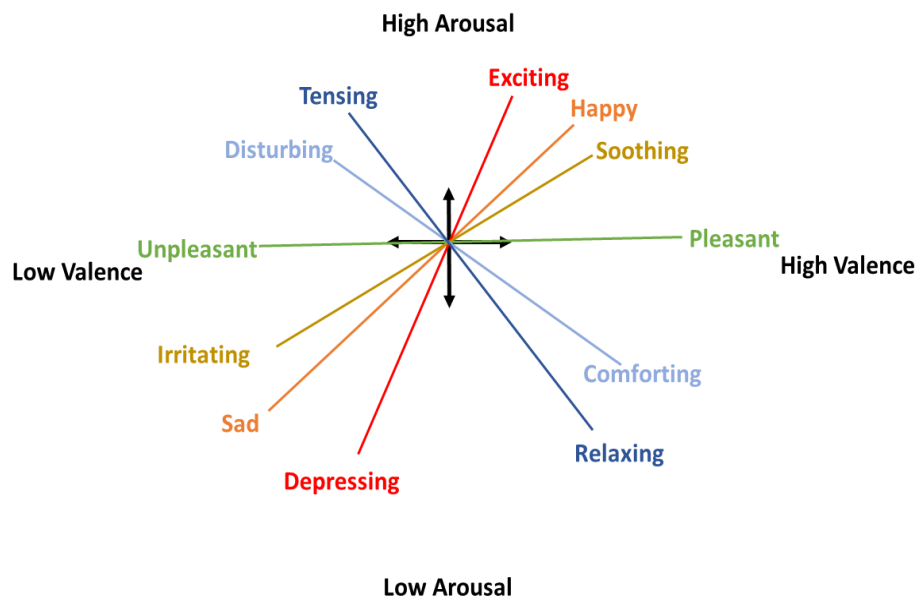


Figure 5.2: Two dimensional emotion model by valence and arousal

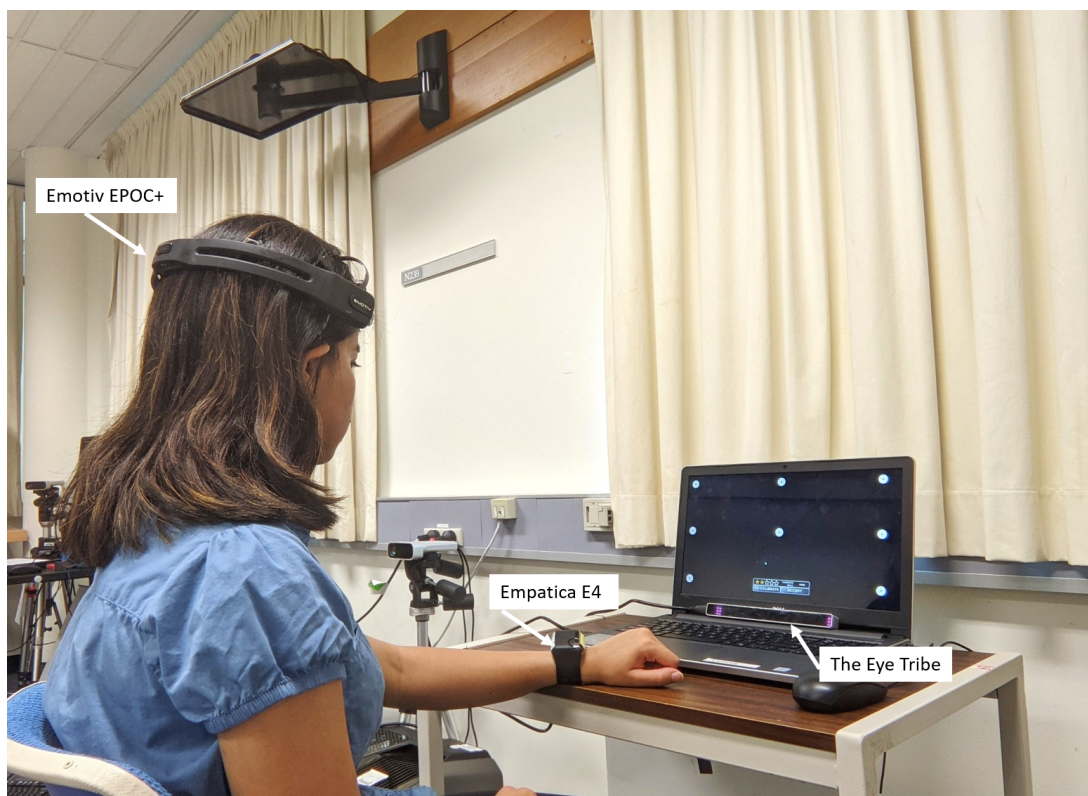


Figure 5.3: Experimental setting - participants physiological signals being collected while they listen to music

Table 5.1: Participant demographic of the experiment of collecting physiological signals during music listening

Subject	Age	Gender	Ethnicity	Education	Music Genre Preference
1	23	Female	Asian	Postgraduate	Classical
2	29	Male	Asian	Undergraduate	Classical
3	22	Male	Asian	Undergraduate	Rap
4	19	Male	Caucasian	Undergraduate	Classical
5	18	Female	Caucasian	Undergraduate	Modern Hip Hop
6	19	Female	Caucasian	Undergraduate	No Specific Type
7	20	Male	Caucasian	Undergraduate	Pop
8	18	Male	Caucasian	Undergraduate	Indie Rock
9	18	Female	Asian	Undergraduate	Pop
10	18	Female	Caucasian	Undergraduate	Hip Hop
11	18	Female	Caucasian	Undergraduate	Metal
12	26	Male	Asian	Postgraduate	Light Music
13	18	Male	Caucasian	Undergraduate	Classical
14	22	Male	Caucasian	Undergraduate	Rock/Musicals
15	19	Female	Caucasian	Undergraduate	House, Electronic, Pop
16	18	Female	Caucasian	Undergraduate	Pop
17	19	Male	Asian	Undergraduate	Pop, Folk
18	18	Female	Asian	Undergraduate	Pop
19	21	Male	Caucasian	Undergraduate	Chill Hip Hop
20	22	Male	Asian	Postgraduate	Contemporary Pop
21	25	Male	Asian	Undergraduate	Indie Rock
22	37	Female	Asian	Postgraduate	Rock
23	24	Female	Caucasian/Other	Undergraduate	Light Music
24	25	Male	Asian	Postgraduate	Instrumental

Australian National University's Human Research Ethics Committee.

5.3 Data Analysis

Data analysis for this experiment was conducted in two phases. In the first phase, only EDA signals were considered with a small subset of features to build the classifier models. After the efficiency of the models were tested, a larger set of features along with EDA, BVP, ST and PD data were analysed in phase two. In the subsequent analysis, these phases will be mentioned to differentiate the two analyses.

5.3.1 Pre-processing

Physiological signals collected during the experiment were first normalised to remove subject dependency from the signals. Min-max normalisation was used to normalise the signals between range 0 to 1. The EDA, BVP, ST and PD signals from each participant were normalised individually. After normalising the data, filtering was done to remove artefacts caused by subject movements and additional environmental noise. The median smoothing technique was used for this purpose. Choosing a high value as the parameter for filtering might cause the loss of valuable data, on the other hand a low value will result in the data remaining too noisy. Based on previous literature, a 10 point median filter was chosen in order to avoid the loss of too much data [Stone, 1995]. These pre-processing steps were the same in both phases of analysis.

For PD data, an additional pre-processing step was performed because several data points were empty due to blinking by the participants. In this case, linear interpolation was applied to generate those data points. Finally all the signals were segmented to the length of each music stimulus prior to further analysis.

5.3.2 Feature Extraction and Selection

A number of features were extracted from both time and frequency domains using all of the physiological signals collected. The complete list is set out in 3.2.3. Table 5.2 lists the features used in the two phases of this data analysis. In phase one, a smaller subset of features was extracted only from EDA data, while phase two used all features extracted from all signals.

The features were extracted from both normalised and filtered signals. Afterwards, some redundant features and features with skewed values were removed. In total, 14 features were extracted from EDA signals in phase one, while 34 features were extracted from each signal in phase two.

For feature selection, all six methods described in 3.2.4 were used in both the preliminary and detailed analysis phases. The methods are statistical dependency (SD),

Table 5.2: Features extracted from participants physiological signals while listening to three genres of music

Feature Type	Phase	Feature Names
Frequency domain features	Both	Mean, minimum, maximum of the first 16 points from Welch power spectral density
Time domain features	Both	Mean, minimum, maximum, standard deviation, interquartile range, variance, kurtosis, number of peaks, mean of first and second difference of the signals
Time domain features	Two	Root mean square, average amplitude change, log detector, difference absolute standard deviation value, detrended fluctuation analysis (DFA), Hjorth parameters (mobility only), Hurst exponent, sample entropy, approximate entropy, Shannon's entropy, permutation entropy, fuzzy entropy

minimal-redundancy-maximal relevance (MRMR), genetic algorithm (GA), sequential forward selection (SFS), sequential floating forward selection (SFFS) and random subset feature selection (RSFS).

5.3.3 Visualisation of the Physiological Signals - Gingerbread Animation

Data visualisation provides an effective method of identifying patterns in complex data, such as physiological signals. Thus, many researchers in the area of computer vision have utilised different visualisation techniques to build classifier models. The effects of analysing physiological signals using different visualisation techniques were explored using the data collected in this experiment. As a preliminary exploration, the physiological signals were visualised in a 2D graph. Each participant's data were segmented according to the music stimuli length. EDA, BVP and ST values were represented in red, blue and green colours respectively. PD data values were not used in this preliminary experiment. Figure 5.4 shows a sample graph image used in this analysis.

When using physiological signals, in particular many physiological signals over a longer period, it is difficult to visualise the data. Therefore in the next stage, an approach was devised called *Gingerbread Animation* which uses a stylised 2D representation of a human body and visually represents the time series of physiological signals propagating on that 2D surface, which can be presented as a video.

In the Gingerbread Animation, BVP, PD, EDA and ST signals were used, which can be represented by red, green, blue and grey colours respectively. The locations of data representation also reflect the locations where these signals are generated

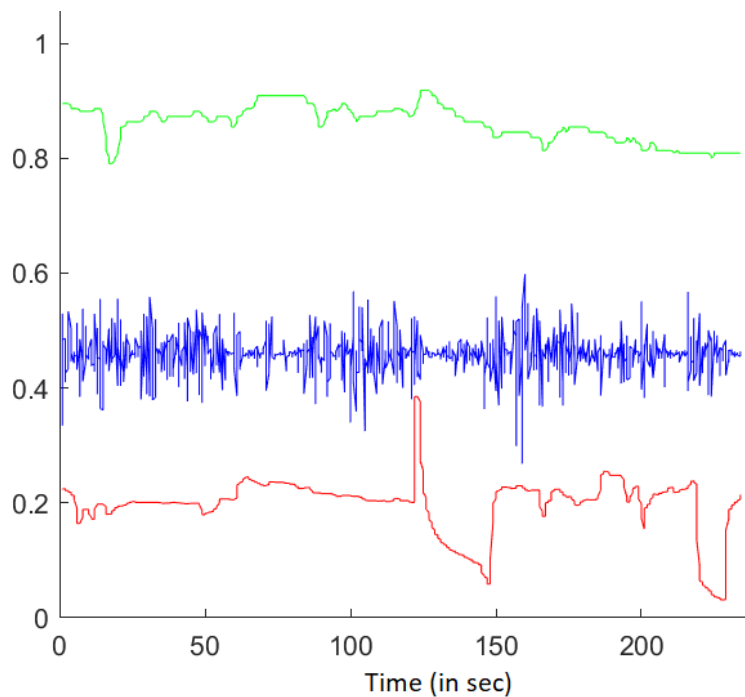


Figure 5.4: Physiological signals representation as a graph (Blue = BVP, Red = EDA, Green = ST)

where possible. The PD, BVP, EDA and ST signals are displayed in right eye, heart, left wrist and right foot area respectively. These colours can combine and create mixed colours on the surface. Thus, a sequence of images are produced (forming a video) representing each experimental trial, and retaining a representation of each signal – the colours mix, but the RGB values are not lost. This representation leads to an additional benefit; it can make use of the highly advanced computer vision techniques available for images to classify and predict based on the new video data.

Figure 5.5 shows some representative images, the first one being a few seconds into a music stimulus, second one during the middle and third one during the end of the stimulus. Each datum is represented as a ring with a fixed maximum width in the animation. The latest data appears in the middle of the circle, for each type of signal, up to 40 time steps of data are showing at the same time. These 40 concentric rings constitute an entire circle. The older the data, the closer to the outside edge the circle it moves to, which simulates the effect of data rippling out.

Physiological data is mapped to the RGB model in the animation, in which $(0, 0, 0)$ is black and $(255, 255, 255)$ is white. To make the visualisation more in line with human intuition, the background is set as white, so as to highlight stronger signals that appear darker due to lower RGB values. As the data spreads, the intensity continues to decay until it drops to 0, which is represented by 255 in the RGB model. For example, when a BVP datum is 0.8, it is represented as $(51.2, 0, 0)$ in the RGB model in

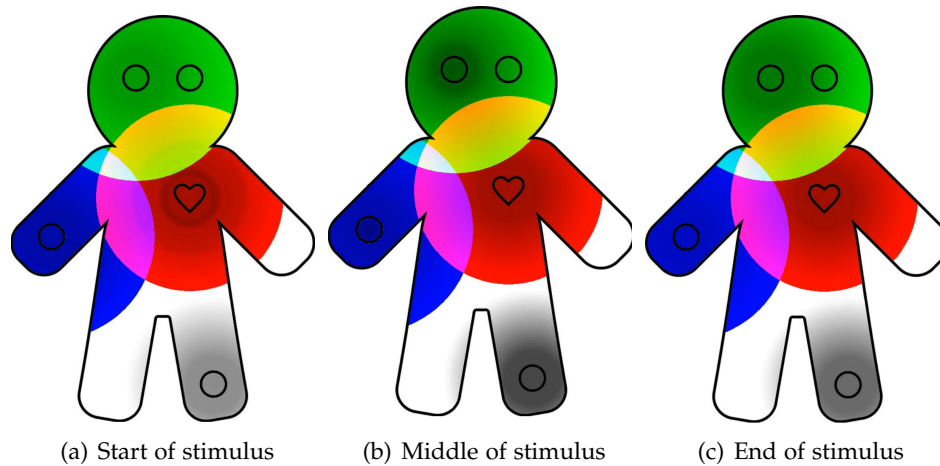


Figure 5.5: Physiological signals representation in an animation (Red = BVP, Blue = EDA, Green = PD, Grey = ST)

the middle of the circle when it first appears and then after spreading out, it begins to decay slowly and ends up as (255,0,0) which is seen as a bright red color in the animation.

In areas where multiple types of signal overlap, the overlapped RGB value is added by the RGB values of each signal. For example, a BVP datum of 0.8 (i.e. (51.2, 0, 0) in RGB) meets a wrist datum of 0.5 (i.e. (0, 0, 128) in RGB), and the resulting output is (51.2, 0, 128) in RGB. The figure also shows that in some parts of the image the amplitudes of the original signals can still be easily seen as the colours have not yet begun to mix. It is observed that the BVP signals vary rhythmically, while the ST varies in a much smoother manner, while the EDA is not rhythmic in this fashion. Finally, the figure further shows some regions where the colour has begun to mix, and produce visually pleasing complex patterns related to the data. Both of these visualisation techniques were analysed along with the feature data mentioned in the previous section.

5.3.4 Classifiers

5.3.4.1 Preliminary analysis using EDA data

In the phase one analysis using only EDA signals, a neural network was created for predictive modelling. All the music stimuli were labelled as one of the following three categories: classical, instrumental and modern pop. A leave-one-participant-out process was performed using a three class classifier to distinguish among the three music categories. The leave-one-participant-out process was done 20 times for all methods and the average results are shown. Classification was done using MATLAB R2017a software with an Intel(R) Core(TM) i7-5200U processor with 3.60 GHz, 16.00 GB of RAM and Microsoft Windows 10 Enterprise 64-bit operating system. For

the classification process, a pattern recognition network was constructed with one input layer, one hidden layer and one output layer. The hidden layer consisted of 30 nodes. Figure 5.6 shows a diagram of the neural network model.

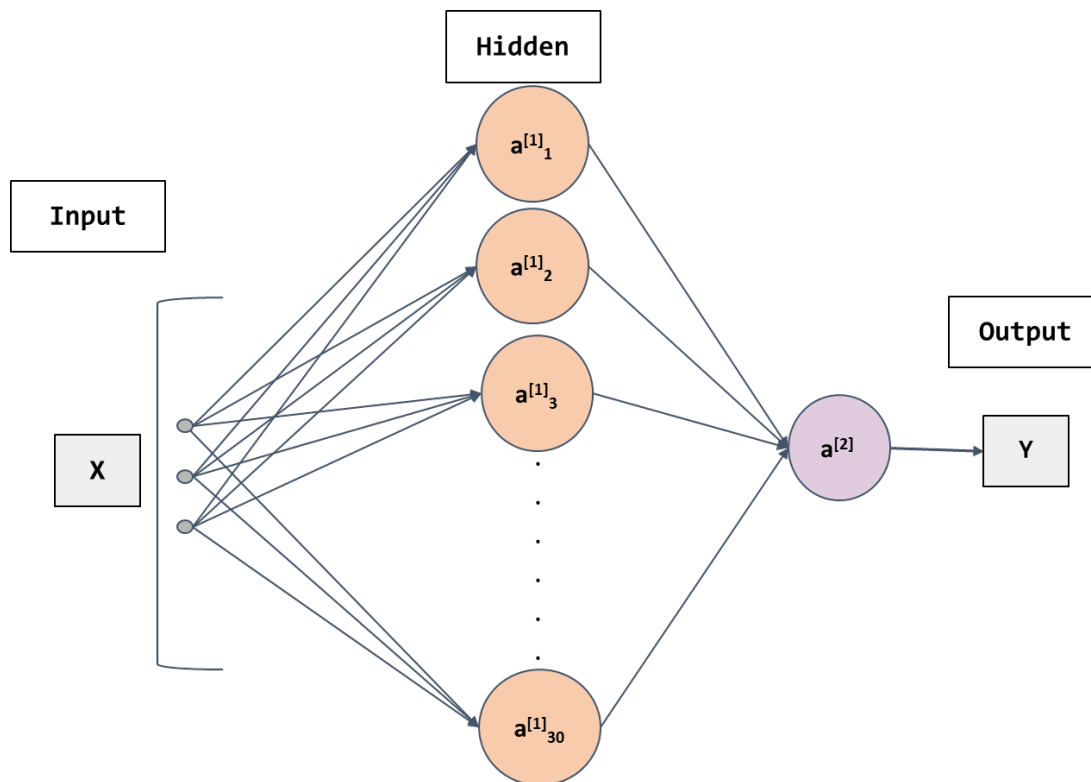


Figure 5.6: Neural network architecture

Choosing the optimum number of hidden nodes is the most crucial task in building the neural network. Too many neurons in the hidden layer may result in overfitting, while too few neurons may cause underfitting. During the phase one study, the small subset of EDA features was analysed to determine certain parameters. In that analysis, a neural network using hidden node numbers from 5 - 50 were analysed and their classification accuracy were compared. The result is shown in Fig. 5.7

We can observe that from the hidden node number of 30 the network produces a reasonably stable result in terms of accuracy. Therefore, the hidden node number of 30 was chosen as the optimum number for all the shallow neural network analyses. Other parameters of the network were: Levenberg—Marquardt [Levenberg, 1944; Marquardt, 1963] methods as network training function and mean squared normalised error as performance function. Multiple neural network models were created based on the features selected by the six feature selection methods. For the two feature selection methods SD and MRMR, the top twelve features were chosen as input for the classification model.

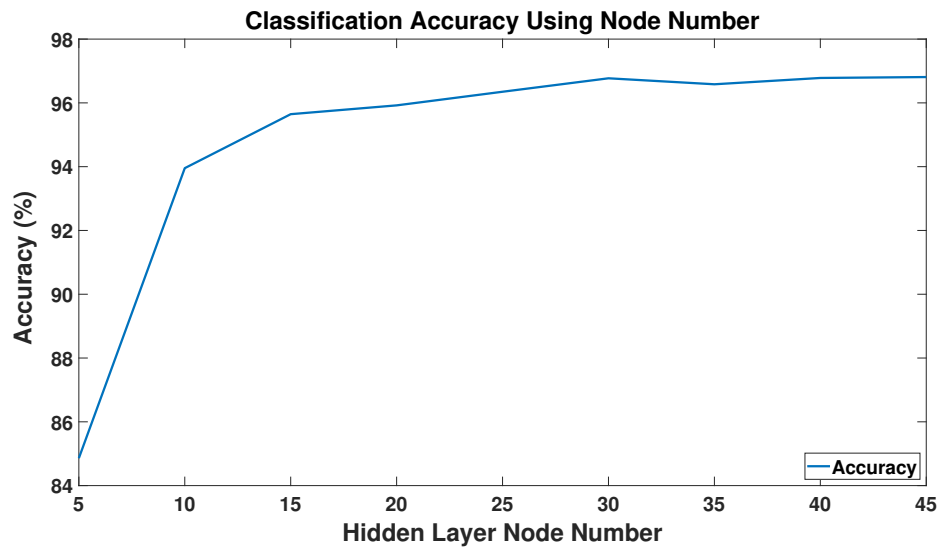


Figure 5.7: Classification accuracy using different hidden node number

5.3.4.2 Analysis Using EDA, BVP, ST and PD Data

Classification using EDA, BVP, ST and PD features was conducted with three techniques. They are: Neural Network (NN), K-Nearest Neighbor (KNN) and Support Vector Machine (SVM). The methods are described in sections 3.2.5.6, 3.2.5.1 and 3.2.5.4 respectively. Using these methods, classification was performed in five different conditions. EDA, BVP, ST and PD features were used individually for classification, and also feature-level fusion was done using all four signals. The entire process was done using all the features and also features selected by the six feature selection methods. For the two feature ranking methods SD and MRMR, the top 12 features were chosen to use in the classification process. A leave-one-participant-out process was performed as the validation approach. Classification was performed using MATLAB R2018a software with an Intel(R) Core(TM) i7-5200U processor with 3.60 GHz, 16.00 GB of RAM and Microsoft Windows 10 Enterprise 64-bit operating system. Similar to the preliminary analysis, classification was done based on both genre and participants' subjective response.

Classification using neural network was done using the same parameters used in the previous section. The classification process was done 20 times and the average of those results were selected. For KNN, the size of K was determined using experimentation. The size of K was tried ranging from 3 to 30 to choose the best results. K= 5 or 7 resulted in best outputs for all cases. For distance metric, Minkowski distance was chosen. The multiclass SVM chosen for this study used tree learner and one-versus-all coding design.

To classify the graphs constructed from physiological signals, a pre-trained con-

volutional neural network (CNN) resnet18 was used and the final layer was modified in order to train (fine-tune) the model using the graph images. Resnet introduced skip connections which help resolve the vanishing gradient issue [He et al., 2016]. Figure 5.8 shows the resnet18 architecture.

To classify the images obtained from the Gingerbread Animation, a CNN was constructed using stochastic gradient descent with momentum (SGDM) with an initial learning rate of 0.005, mini batch size of 32. A variation of the classic Lenet-5 architecture [LeCun et al., 1998] containing three convolutional layers, two max pooling layers, a fully-connected layer and a softmax classifier was used. Figure 5.9 shows the CNN architecture.

5.4 Results and Discussion

In this section the results from the preliminary analysis in phase one are reported, followed by the complete analysis from phase two.

5.4.1 Results using EDA signals

Figure 5.10 shows the classification results for three music genres using the EDA feature data extracted during phase one analysis.

From Figure 5.10 it can be observed that classification using a neural network along with GA feature selection method can give a high accuracy of 96.8% for three different music genres. This implies that EDA can be a good measure in classifying music categories. A neural network with GA gives best results in terms of all six evaluation measures. Table 5.3 gives a comparison of GA results with SD/MRMR (which performed moderately) and SFFS (which performed the worst). It can be seen that the difference between GA and SFFS method is quite significant (around 15.0% and 13.0% for precision and F measure respectively).

Table 5.3: Classification based on music genres using EDA signals

	GA	SD/MRMR	SFFS
Accuracy	0.968	0.961	0.875
Precision	0.929	0.918	0.775
Recall	0.979	0.97	0.885
Specificity	0.962	0.957	0.869
F Measure	0.953	0.943	0.825
Geometric Mean	0.971	0.963	0.877

As mentioned earlier, all 24 participants of this study provided subjective ratings

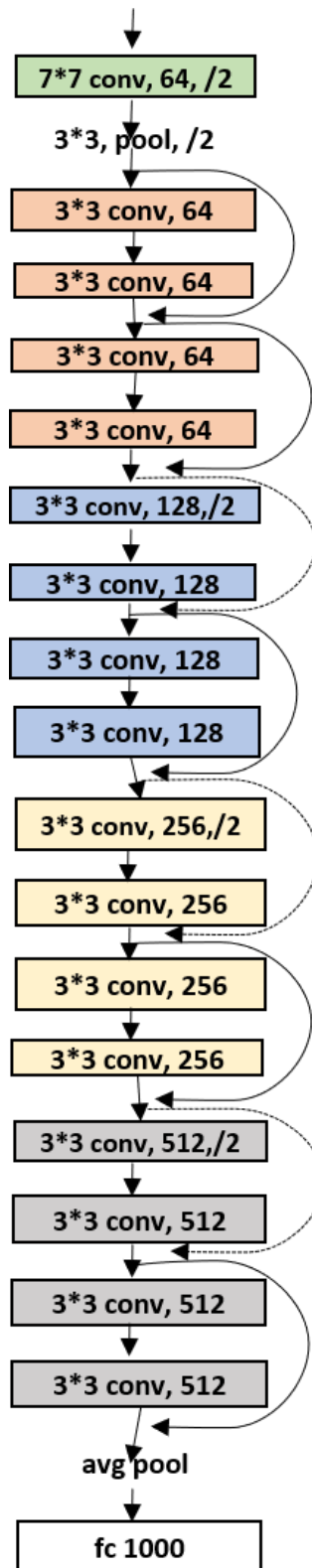


Figure 5.8: Resnet18 architecture

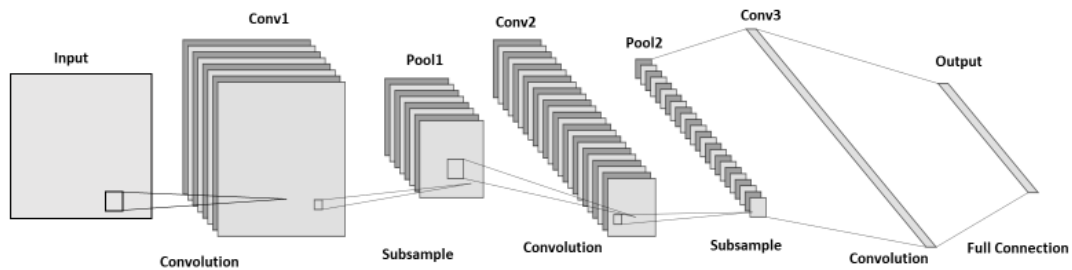


Figure 5.9: CNN architecture for Gingerbread Animation

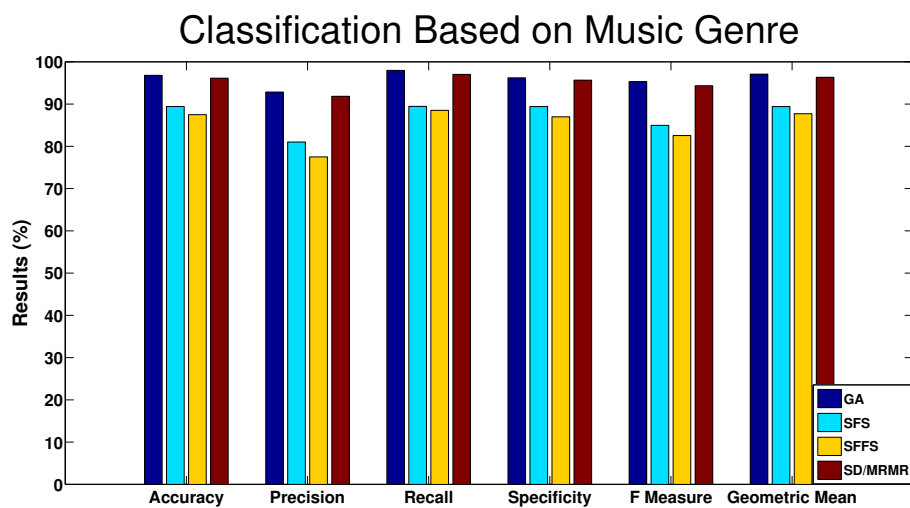


Figure 5.10: Classification based on music genres using EDA signals

about the music stimuli on a 7-point likert scale. These ratings were then converted to three ranges, and the music stimuli were labelled based on that. Thus the labels for the 6 questions were: i) *sad* → *neutral* → *happy*, ii) *disturbing* → *neutral* → *comforting*, iii) *depressing* → *neutral* → *exciting*, iv) *unpleasant* → *neutral* → *pleasant*, v) *irritating* → *neutral* → *soothing*, and vi) *relaxing* → *neutral* → *tensing*. These labels were used to perform the classification using the three class classifier similar to the genre classification problem.

Subjective ratings provided by the participants were analysed using the analysis of variance (ANOVA) test. The significance level of each music category was computed based on the six questions about emotional response to the music stimuli. The results show statistical significance ($p < 0.05$) for two emotions (*sad* → *happy*, *depressing* → *exciting*), but it did not show significant differences for other emotions. Based on the significance results, the classification results are reported below based on the one emotion that showed significance (*depressing* → *neutral* → *exciting*) and one that did not (*disturbing* → *neutral* → *comforting*).

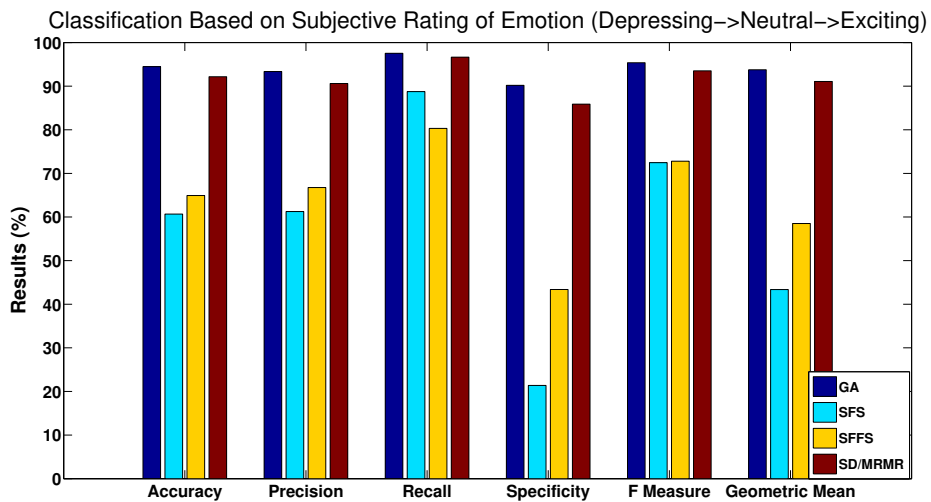


Figure 5.11: Classification results based on subjective rating (*depressing* → *neutral* → *exciting*) using EDA signals

Figure 5.11 shows the evaluation results of the subjective rating based on the emotion *depressing* → *neutral* → *exciting*. Similar to the genre based classification, all evaluation measures are calculated from an average of 20 runs. It can be observed that GA again performs better than all other feature selection methods. A comparison between the evaluation measure value of GA and two other methods is given in Table 5.4. Similar to genre based classification, it can be seen that GA performs well. Compared to the worst performing method SFS, the improvement is around 35% difference in accuracy. Similar outcomes are observed for classification based on

two other emotions (*sad* → *neutral* → *happy*, and *irritating* → *neutral* → *soothing*). In these three cases GA performs better than all other feature selection methods. The average accuracy based on these two emotions are 93.8% and 95.1% respectively.

Table 5.4: Classification results based on subjective rating (*depressing* → *neutral* → *exciting*) using EDA signals

	GA	SD/MRMR	SFS
Accuracy	0.945	0.922	0.607
Precision	0.933	0.906	0.612
Recall	0.975	0.967	0.888
Specificity	0.902	0.859	0.214
F Measure	0.954	0.935	0.725
Geometric Mean	0.938	0.911	0.434

However, different patterns are observed for the other three emotion based classifications (*disturbing* → *neutral* → *comforting*, *unpleasant* → *neutral* → *pleasant* and *relaxing* → *neutral* → *tensing*). Figure 5.12 shows the results of six evaluation measures based on the emotion *disturbing* → *neutral* → *comforting*.

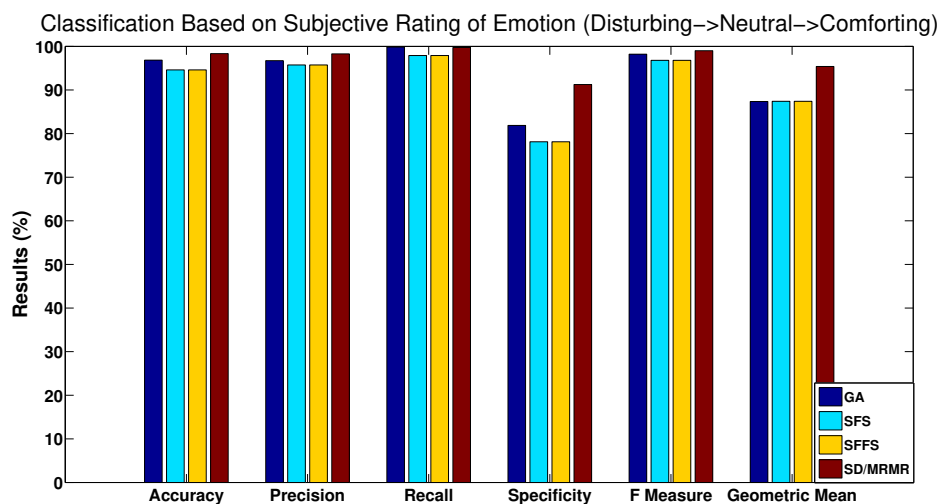


Figure 5.12: Classification results based on subjective rating (*disturbing* → *neutral* → *comforting*) using EDA signals

As seen in Figure 5.12, the SD/MRMR feature selection method gives the best results in all measures. Compared with the worst performing method (both SFS and SFFS in this case as they have selected the same set of features), the difference is high

in terms of specificity (around 13.0%). The values are given in Table 5.5.

Table 5.5: Classification results based on subjective rating (*disturbing* → *neutral* → *comforting*) using EDA signals

	GA	SD/MRMR	SFS/SFFS
Accuracy	0.968	0.983	0.946
Precision	0.967	0.983	0.957
Recall	0.998	0.998	0.979
Specificity	0.819	0.913	0.781
F Measure	0.982	0.99	0.968
Geometric Mean	0.874	0.954	0.874

As for the other two emotions (*unpleasant* → *neutral* → *pleasant* and *relaxing* → *neutral* → *tensing*), the best results for all evaluation measures are also found by using SD/MRMR feature selection method with the average accuracy of 98.0% and 96.4% respectively. Although the values of GA and SD/MRMR are quite similar, t-test analysis shows a significant difference in terms of accuracy for *depressing* → *neutral* → *exciting* ($p < 0.01$), but not for the case of *disturbing* → *neutral* → *comforting*. Values of other evaluation measures also show the same pattern for both cases.

From the emotion model in Figure 5.2, we can observe that for the three emotions that have a positive slope (*depressing* → *neutral* → *exciting*, *sad* → *neutral* → *happy* and *irritating* → *neutral* → *soothing*) GA feature selection methods work the best. But for the emotions that have a slope of 0 or a negative value (*disturbing* → *neutral* → *comforting*, *relaxing* → *neutral* → *tensing* and *unpleasant* → *neutral* → *pleasant*) SD/MRMR method works the best. Further experimentation and analysis are required to understand this phenomenon. Based on the current results it is evident that there is a correlation between human emotions and their physiological signals which can be differentiated by different feature selection methods during classification.

5.4.2 Results Using EDA, BVP, ST and PD Signals

Based on the preliminary results in phase one using EDA signals, an extended classification was done using all four signals (EDA, BVP, ST, PD). Certain patterns are observed from the classification results which are described below.

5.4.2.1 Neural Network Performs Best Among All Classifiers

The results of all five classification approaches show that in every case NN performed significantly better than KNN and SVM. Figure 5.13 shows the accuracy

results based on the participants' subjective rating in respect of the emotion scale *tensing* → *relaxing*, using all features.

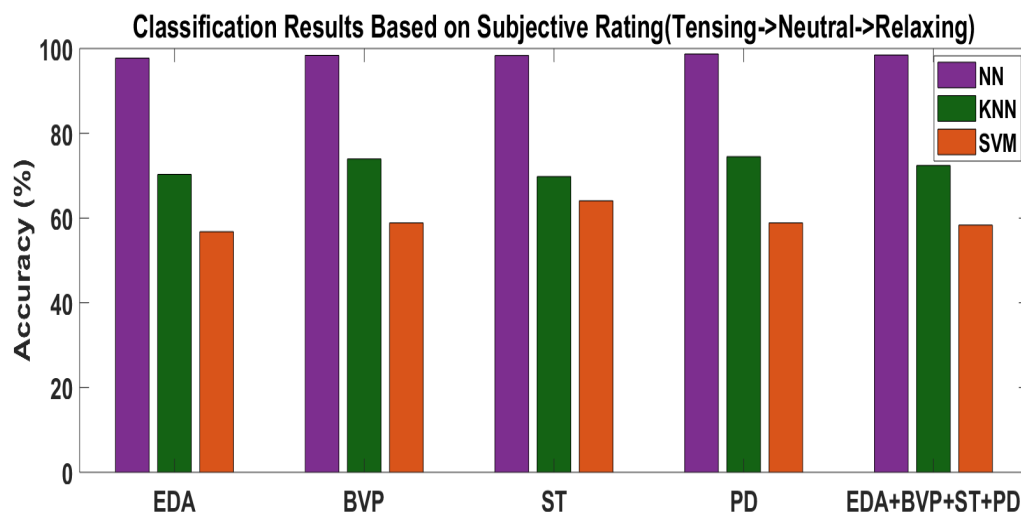


Figure 5.13: Classification result based on subjective rating (*tensing* → *relaxing*) using EDA, BVP, ST and PD signals

From Figure 5.13 it is evident that NN performs best in terms of accuracy in all 5 combinations of features. Neural network gives the accuracy of 97.7%, 98.3%, 98.3%, 98.7% and 98.5% accuracy using EDA, BVP, ST, PD and EDA+BVP+ST+PD features respectively. In comparison, KNN gives 70.3%, 73.9%, 69.8%, 74.5%, 77.6% and SVM gives 56.8%, 58.9%, 64.1%, 58.9%, 62.5% accuracy. This pattern prevails in classification using features from the feature selection methods as well. This study solidifies the results from the phase one analysis in showing that a simple NN can be a strong system in classifying physiological signals.

5.4.2.2 Feature Selection Produces Best Results for Music Genre Classification

Classification accuracy results of different feature selection methods using neural network for music genre classification were compared and the results are shown in Figure 5.14.

It is observed that for the EDA, BVP, ST, PD and EDA+BVP+ST+PD feature combinations, the RSFS, MRMR and SD methods result in the best NN accuracy. Both KNN and SVM also produce their best results using feature selection methods. The results show similar patterns across all evaluation measures. Table 5.6 shows the results of all six evaluation measures for NN classification of music genres. The table does not include results using GA and SFFS as feature selection methods because they do not achieve the highest values in any of the evaluation measures.

Table 5.6: Classification results based on music genre using EDA, BVP, ST and PD signals

		All	SD	MRMR	RSFS	SFS
EDA	Accuracy	0.981	0.975	0.971	0.984	0.908
	Precision	0.982	0.976	0.979	0.983	0.875
	Recall	0.959	0.947	0.934	0.969	0.847
	Specificity	0.991	0.988	0.989	0.992	0.939
	F-Measure	0.971	0.961	0.956	0.976	0.860
	G-mean	0.975	0.967	0.961	0.980	0.892
BVP		All	SD	MRMR	RSFS	SFS
	Accuracy	0.975	0.992	0.993	0.983	0.989
	Precision	0.957	0.992	0.994	0.974	0.984
	Recall	0.969	0.983	0.984	0.977	0.983
	Specificity	0.978	0.996	0.997	0.987	0.992
	F-Measure	0.963	0.988	0.989	0.975	0.984
G-mean	0.973	0.989	0.991	0.982	0.988	
ST		All	SD	MRMR	RSFS	SFS
	Accuracy	0.982	0.989	0.986	0.975	0.944
	Precision	0.959	0.983	0.979	0.962	0.925
	Recall	0.989	0.985	0.979	0.963	0.906
	Specificity	0.979	0.991	0.989	0.981	0.963
	F-Measure	0.974	0.984	0.979	0.963	0.915
G-mean	0.984	0.988	0.984	0.972	0.934	
PD		All	SD	MRMR	RSFS	SFS
	Accuracy	0.983	0.981	0.984	0.979	0.941
	Precision	0.984	0.979	0.985	0.981	0.924
	Recall	0.968	0.963	0.966	0.959	0.898
	Specificity	0.992	0.989	0.993	0.991	0.963
	F-Measure	0.975	0.971	0.975	0.969	0.91
G-mean	0.979	0.976	0.979	0.974	0.929	
EDA+ BVP+ ST+ PD		All	SD	MRMR	RSFS	SFS
	Accuracy	0.97	0.978	0.977	0.972	0.958
	Precision	0.952	0.951	0.948	0.957	0.937
	Recall	0.96	0.984	0.986	0.96	0.936
	Specificity	0.975	0.975	0.973	0.979	0.969
	F-Measure	0.956	0.967	0.966	0.959	0.936
G-mean	0.968	0.979	0.979	0.969	0.952	

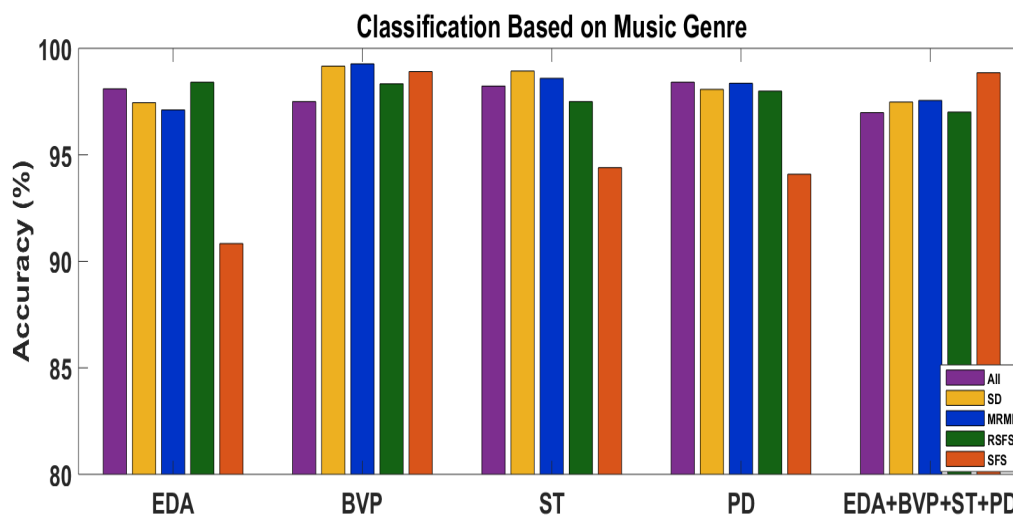


Figure 5.14: Neural Network classification result based on music genre using EDA, BVP, ST and PD signals

Table 5.6 shows that the high score for all evaluation measures is reached by a feature selection method. A few exceptions can be observed, such as the condition using ST features where the highest recall score is achieved by using all features. Furthermore, using PD features, the highest recall and specificity is achieved using the full set of features. However, we can see that in those combinations the highest F-measure is reached by the SD and MRMR methods respectively, aligning with the other measure values. F-measure is the harmonic mean of precision and recall, which takes both false positive and false negative values into account. Recall does not consider false positive values, therefore the F-measure is a stronger measure for evaluating a model, compared to just precision or recall. This result suggests that using a smaller subset of the features not only reduces the computational time, but also increases accuracy of NN models in classifying different music genres.

This result also improves the phase one study which used a smaller subset of features. That analysis reached the highest accuracy of 96.8% for music genre classification. In the extended analysis, the accuracy using EDA features increased to 98.4%, and overall the highest accuracy reached 99.3%, using BVP features. The improved set of features identified through this analysis has resulted in the improved accuracy of the classification models.

5.4.2.3 Statistical Analysis on All Evaluation Measures

Results of the six evaluation measures for NN classification across all five conditions were analysed using analysis of variance (ANOVA) test. A one-way ANOVA test showed high statistical significance ($p < 0.01$) for all of the evaluation measures. The accuracy results for all pairs of feature selection methods were compared for

statistical significance. The results are shown in Table 5.7.

Table 5.7: Significance values for all pairs of feature selection methods

All						
SD	0.00003					
MRMR	0.00009	0.386				
GA	0.0002	0.029	0.085			
RSFS	0.006	0.139	0.119	0.782		
SFS	0.000002	0.00002	0.000009	0.000006	0.00004	
SFFS	0.000009	0.00008	0.00006	0.00003	0.0003	0.335
	All	SD	MRMR	GA	RSFS	SFS

In Table 5.7, the numbers in colour and bold are the pairs that show meaningful differences. Red colour shows a significance of $p < 0.05$, while blue colour shows significance $p < 0.01$ and teal colour shows significance at threshold $p < 0.001$. A further observation is that both SFS and SFFS reached high significance values in comparison with other selection methods. This is reflected in the number and type of features chosen by these methods as well. It can be clearly seen from the table that different combinations of features in the model result in significant differences in model accuracy. Therefore, in the next section, some of the features that were shown to be useful for the classification models are discussed.

5.4.2.4 Top Features Selected by Feature Selection Methods

The number of times each feature was chosen by the six feature selection methods was counted for all classification models. Based on that, the top 12 features are reported from all the features that were extracted in Table 5.8. Unless specifically mentioned, most of the features were extracted from the filtered signals.

Table 5.8: Top 12 features (from EDA, BVP, ST and PD signals) selected by all methods

Feature Names	Feature Type
Number of peaks (both normalised and filtered), variance, sum, absolute sum, simple square integral	Linear features from time domain
Mean, minimum, maximum of the first 16 data points from Welch power spectrum density analysis	Linear features from frequency domain
Sample and approximate entropy, Hjorth parameters (mobility)	Non-linear features from time domain

The list gives us some interesting insights into which types of features are best to represent the four physiological signals' changes. The number of peaks for both normalised and filtered values were selected the most times by all methods. These peaks

are thus the most valuable feature that reflects the SCR occurrences (rapidly changing states). SCR occurrences are considered to be most useful in reflecting autonomic arousal [Bos et al., 2013]. Although the normalised and filtered signal features are quite similar, they clearly do not add redundancy to the system. With some signals, useful peaks might be removed due to the filtering process. In those cases, peaks in the normalised signals proved to be more useful. We also notice that the three features extracted from the Welch power spectrum density analysis appeared in the top features list. This shows that the frequency domain features can be very useful to identify patterns in these signals.

Some of the other interesting features are entropies and mobility. All of these features represent the level of complexity of the signals. Features like entropies can effectively capture short range correlations and thus, they are effective in identifying transient emotional state changes [Jerritta et al., 2013].

5.4.2.5 Best Feature Selection Methods

NN accuracy results for all feature selection methods were further analysed and the methods were ranked based on how many times that method achieved highest accuracy. The list below shows the rank of the feature selection methods and their frequency of achieving the highest accuracy.

- GA - 11
- MRMR - 7
- RSFS - 5
- SD - 4
- SFS - 1
- SFFS - 0

It should be noted that GA was not able to achieve the highest accuracy for music genre classification in any combination. But it was able to achieve the highest accuracy in most combinations for the six emotion based classifications. For the cases where GA was not able to reach the highest accuracy, it was still able to achieve close to the highest. In the preliminary analysis with only EDA signals, it was shown that for the three emotions that have a negative slope (*depressing* → *neutral* → *exciting*, *sad* → *neutral* → *happy* and *irritating* → *neutral* → *soothing*) in the emotion model (shown in Figure 5.2), GA feature selection methods performed the best. For the emotions that have a slope of 0 or a positive value (*disturbing* → *neutral* → *comforting*, *relaxing* → *neutral* → *tensing* and *unpleasant* → *neutral* → *pleasant*) SD/MRMR methods work the best. However, further analysis using more physiological signals and a wider set of features showed that GA is able to select a robust set of features for all six emotion based classifications. Therefore, using the GA feature selection

method is most suitable for classification of music based on different emotion ratings.

5.4.2.6 Effectiveness of Visualisation

Two different subjective ratings of emotions (*sad* → *neutral* → *happy* and *tensing* → *neutral* → *relaxing*) were used to demonstrate classification using the graph and animation images. A leave-n-participants-out cross validation approach was used to validate the accuracy of the network. A total of 16 participants' data were randomly chosen for training and eight participants' data for testing. The graph images trained using a pre-trained CNN achieved 61.9% accuracy for the emotions *sad* → *neutral* → *happy* and 73.4% accuracy for *tensing* → *neutral* → *relaxing*. In comparison, the animation images reached 68.1% and 74.8% for the same emotion pairs. It should be noted that the comparisons are not exact; the graph visualisations show 250 sec of data with 3 physiological signals, while the Gingerbread Animation in the comparison presents much less information, being only 10 sec of data (40 time steps at 4 Hz) in each frame for four physiological signals. The better results with less data suggest that the Gingerbread Animation can be both a visually attractive and effective approach to identify emotions from human physiology using state-of-the-art machine learning methods. This merits further investigation in the future.

Another observation is that humans may interpret their feelings in response to some music stimuli differently to what their underlying physiological signals indicate. To initially label the emotions according to subjective rating, a majority voting approach was used to label each music stimuli. Afterwards each music stimulus was labelled based on each participant's individual subjective response. This resulted in the accuracy dropping from 62.0% to 50.0% for *sad* → *neutral* → *happy* and from 73.4% to 47.4% for *tensing* → *neutral* → *relaxing*. Therefore, we can see that some participants are different (compared to the overall population of this experiment) in rating their emotions listening to the music stimuli. However, on average the participants' responses correlate with their physiological response. Thus, considering the overall view of the population, each person's emotional reaction to the music stimuli was as expected. However, their own conscious view was often not supported by their physiological responses.

There is scope for improvement in both the Gingerbread Animation and the computational models using that data. In particular, it should be emphasised that results using the simple neural network are based on substantial work in pre-processing. However, using a pre-trained CNN, the notable results were achieved using just the raw data. The visualisation approach also makes the data compatible with both traditional machine learning and deep learning models, where the models primarily take image data as input.

5.5 Summary

In this chapter, results were reported based on the effects of different music stimuli in four different physiological signals. Signals were pre-processed and different numbers of features were extracted for experimentation. Six different feature selection methods were employed to identify useful features. Analysis using three different classification methods (NN, KNN and SVM) were performed and evaluated using six different measures. All the results were compared using features from a specific signal and also the combination of all signals. Neural networks achieved the highest accuracy across all different conditions with the highest accuracy of 99.2% and 98.5% in classifying music based on genre type and human emotions respectively. Furthermore, the GA feature selection method has shown to be best for classifying music based on subjective emotion ratings by participants. A novel animation technique was introduced to both visualise physiological signals and to make them accessible to computer vision classifiers. Preliminary results using a CNN achieved up to 74.8% accuracy in identifying different music based on the subjective rating of participants' emotion.

There are certain limitations to the work introduced in this chapter. The number of samples is not very large for higher power deep learning models. More samples need to be collected in order to use more state-of-the-art deep learning methods. The visualisation methods used in this chapter leverages some deep learning techniques. However, these need to be trained longer in order to create more robust models. This was not possible due to the current processing power of the system (the processing power issue is improved in chapter 7). In the next chapter, some of these methods will be used to analyse participants' brainwave signals.

Effects of Music on Brainwave Patterns

This chapter explores the impact of three different types of music stimuli on human brain activity using EEG. Several signals from different brain regions were investigated to identify which features provide useful information regarding music type and emotion processing. Three different classifiers were used to recognise the three music genres based on the selected brain activity features. The subjective responses provided by the participants related to the music were also classified using a similar approach. The chapter is based on the results published in 2020 International Joint Conference on Neural Networks – IJCNN 2020 [Rahman et al., 2020a], where I was the primary contributor.

6.1 Experiment Design

The data collected in this analysis was from the same experiment discussed in section 5.1. After participants received all the initial information on the experiment and were fitted with the other wearable devices such as Empatica E4, they were fitted with the Emotiv EPOC headset. The Emotiv EPOC headset is a 14-channel wireless headset that also has 9-axis motion sensors. Emotiv also provides software that can be used to record raw EEG data, from which different brain waves and related information can be extracted. Figure 6.1 shows the channels' names and locations of Emotiv EPOC electrodes.

The headset electrodes were properly hydrated to achieve good connectivity prior to the calibration process. After participants put on the headset, the calibration process began. Participants were asked to keep their eyes open for 15 seconds and keep their eyes closed for another 15 seconds to record the baseline data. After that the calibration was completed. Then the data collection process began at the sampling rate of 128 Hz. Band power data was collected at 8 Hz.

Participants in this experiment were the same as the experiment described in chapter 5. The detailed participant demographic is given in Table 5.1.

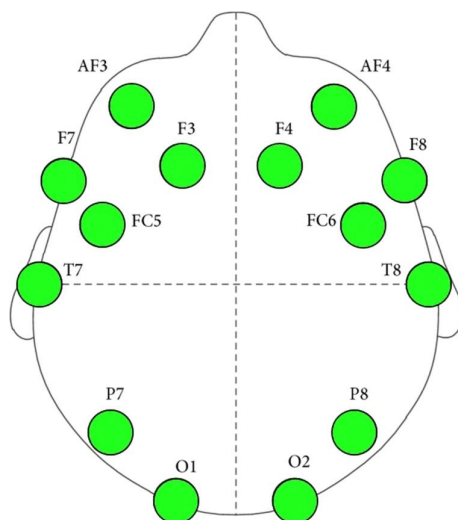


Figure 6.1: Emotiv headset channel location and names [Balasubramanian et al., 2018]

6.2 Data Analysis

6.2.1 Pre-processing

Raw EEG signals collected from participants are very sensitive to subject movements. In addition, sometimes a few channels could not obtain a good connection and added noise artefacts to the collected signals. To make the findings plausible for real world application, the use of head immobilisation or chin rests were not used. Therefore, multiple filtering methods were applied to the raw signals. A median smoothing filtering was used to smooth out the noisy signals. Then the EEG data was band-pass-filtered between 3 to 60 Hz. This was done primarily to separate the band frequency ranges of interest, which are: Alpha [8 – 13 Hz], Beta [14 – 30 Hz] and Gamma [31 – 50 Hz] bands. Then the data was segmented into the lengths of the music pieces for feature extraction.

6.2.2 Feature Extraction and Selection

EEG signals were collected using all 14 channels at the sampling rate of 128 Hz. A total of 26 linear and non-linear statistical features were extracted from the pre-processed data. The features were extracted from the three chosen band frequency ranges. Table 6.1 shows the 26 linear and non-linear features extracted from every participant's music segments. The process was done in the same manner for all 14 channels. The channel names and locations are also noted in Table 6.1. Channel names follow the convention of the International 10 – 20 locations system [Pastelak-Price, 1983].

All six feature selection methods described in section 3.2.4 were used on the list of extracted features to find the optimum set of features. The methods are, statistical

Table 6.1: Emotiv channel names, locations and extracted feature list from EEG signals

Channels	Location	Names
	Pre-Frontal Lobe	AF3, AF4
	Frontal Lobe	F3, F4, F7, F8, FC5, FC6
	Temporal Lobe	T7, T8, P7, P8
	Occipital Lobe	O1, O2
Features	Type	Names
	Linear	Mean, maximum, minimum, standard deviation, interquartile range, sum, variance, skewness, kurtosis, root mean square, average of the power of signals, peaks in periodic signals, integrated signals, simple square integral, means of the absolute values of the first and second differences, log detector, average amplitude change, difference absolute standard deviation value
	Non-Linear	Detrended fluctuation analysis (DFA), approximate entropy, fuzzy entropy, Shannon's entropy, permutation entropy, Hjorth parameters (mobility only), Hurst exponent

dependency (SD), minimal-redundancy-maximal relevance (MRMR), genetic algorithm (GA), sequential forward selection (SFS), sequential floating forward selection (SFFS) and random subset feature selection (RSFS).

6.2.3 Classifiers

Three different classification methods were used to classify the EEG signals. They are: neural network (NN), k-nearest neighbor (KNN) and support vector machine (SVM). The methods are described in sections 3.2.5.6, 3.2.5.1 and 3.2.5.4 respectively. For the two feature ranking methods SD and MRMR, the top 150 features were chosen for use in the classification process. This number was chosen because the feature subset selection methods generally resulted in around 100-180 features. The number 150 was chosen as an optimum level to lead to good classification performance and not be too computationally heavy. A leave-one-observer-out process was performed as the validation approach.

For the neural network, a pattern recognition network was constructed with one input layer, one hidden layer and one output layer. The hidden layer consisted of 30 nodes, based on the analysis done in section 5.3.4.1. Other parameters of the network were: Levenberg-Marquardt method [Levenberg, 1944; Marquardt, 1963] as network training function and mean squared normalised error as performance

function. The classification process was done 20 times and the average of those results were selected. For KNN, experimentation was conducted using K sizes 3 to 30 to choose the best results. K = 9 resulted in best outputs for most cases. Minkowski was used as the distance metric. The multi-class SVM chosen for this study uses tree learner and one-versus-all coding design. The evaluation measures used in this study were classification accuracy, precision, recall, specificity and f-measure. All of these are described in section 3.4.

6.3 Results and Discussion

The classification was performed using MATLAB® R2018a software with an Intel® Core™ i7-5200U processor with 3.60 GHz, 16.00 GB of RAM and Microsoft Windows 10 Enterprise 64-bit operating system. The sections below highlight the key findings of this study.

6.3.1 Statistical Analysis

The statistical analysis was conducted using analysis of variance (ANOVA). The classification accuracy using NN for all feature selection combinations was analysed. The results show high statistical significance ($p < 0.01$) across all the selection methods. However, there was no statistical significance observed for classifications using KNN and SVM. Thus, different feature selection methods have significant impacts only on the NN models of this study. In the later sections the optimal feature selection methods will be discussed further.

6.3.2 Best Features

The frequency of every feature chosen by each feature selection method was counted for all seven classification processes. Table 6.2 shows the list of top 25 features in decreasing order of frequency.

The table gives two types of useful information. Firstly, we can identify which extracted features are providing useful information as derived by a number of feature selection models. Secondly, it tells us which channels (signals from parts of brain regions) are useful in the classification process. From the top 25 features, 10 come from the channels F3 and F7, both located in the frontal lobe of the brain. Most of the other features were also from the channels located in the frontal and pre-frontal region of the brain (except four of them which were features from the temporal lobe). This shows that the frontal lobe can reveal important information related to music processing in the brain. The frontal and pre-frontal lobes are considered to be the emotional control centres of the brain [Salzman and Fusi, 2010]. Frontal lobes are also involved in decision making [Collins and Koehlin, 2012]. These observations align with the literature where high activity in the frontal lobe has been seen during various activities. Khushaba et al. [2012] reported high delta and theta activity in F3

Table 6.2: Top 25 EEG features selected by feature selection methods

Channel	Feature Name
F3	Standard Deviation
FC5	Permutation Entropy
P8	Permutation Entropy
F3	Maximum
F8	Permutation Entropy
F7	Shannon's Entropy
AF3	Skewness
AF3	Shannon's Entropy
P7	Permutation Entropy
F4	Permutation Entropy
FC5	Skewness
T7	Skewness
F3	Mean of the First Difference
F7	Approximate Entropy
T7	Permutation Entropy
F7	Hurst Exponent
AF3	Maximum
F7	Skewness
F7	Kurtosis
FC6	Root Mean Square
P8	Approximate Entropy
FC6	Permutation Entropy
AF4	Permutation Entropy
F3	Hurst Exponent
F3	Mean

and F4 region during decision making. This finding can also be beneficial for future research in making wearable devices to capture EEG. An observations while conducting the experiment was that participants often felt uncomfortable wearing the 14 channel headset for a period longer than an hour. This often hampered their concentration in listening to the music and answering questions. A comfortable wearable device which captures data only from the frontal region of the brain, requiring less points of pressure on the head, may be beneficial for longer experiments in such cases.

Another observation from this feature list is the usefulness of the entropy features. From this list it can be seen that permutation entropy of eight different channels appeared in the top features list. Furthermore, entropies cover 12 out of the top 25 features. Entropies in general reflect the randomness and complexity properties of physiological signals. Permutation entropy analyses various permutation patterns

of these signals to identify the complexity level [Bandt and Pompe, 2002]. These features highlight useful properties from non-stationary signals like EEG. Entropies have also been shown to be effective features for building models for epileptic seizure detection [Chen et al., 2015]. Using these features and relevant channel data can significantly reduce the computational cost of the systems without compromising its predictive power.

Finally, the list of best features from this study is quite different from the best features using EDA, BVP, ST and PD data (described in Table 5.8). This shows the importance of the feature selection step when the features are extracted.

6.3.3 Classification Results

Classification using NN, KNN and SVM was done based on the three music genre and participants' subjective ratings on emotions. The labelling approach follow the same approach as the study reported in chapter 5. Participants' subjective ratings were recorded on six emotion scales described in section 5.1.

In general, for all cases, NN performed significantly better than KNN and SVM. Figure 6.2 shows the classification accuracy of all three models using all six feature selection methods based on the ratings on emotion *tensing* → *neutral* → *relaxing*.

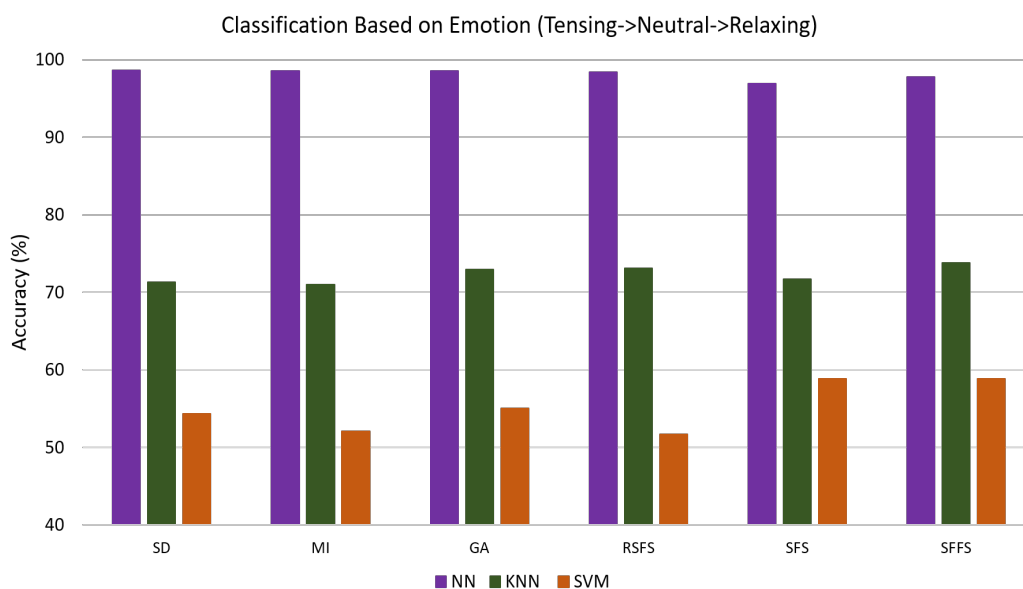


Figure 6.2: Classification results using EEG features based on subjective rating (*tensing* → *neutral* → *relaxing*), range 40-100 chosen for better visualisation

Figure 6.2 shows that NN can reach the highest accuracy of 98.6% based on the average of 20 runs, whereas KNN and SVM reached 73.8% and 58.9% respectively.

Similar patterns are observed in other emotion scales as well across all evaluation measures. Figure 6.3 shows the six evaluation measures for classification based on the music genres using NN. It can be observed that NN achieves a high accuracy of 97.5% and 96.3% in F-measure. For KNN and SVM, even though the models achieve reasonable results in terms of accuracy, it often gets a low score ($< 40.0\%$) for F-measure. Therefore, NN should be considered a more effective model compared to KNN and SVM, as it achieves high scores for all evaluation measures.

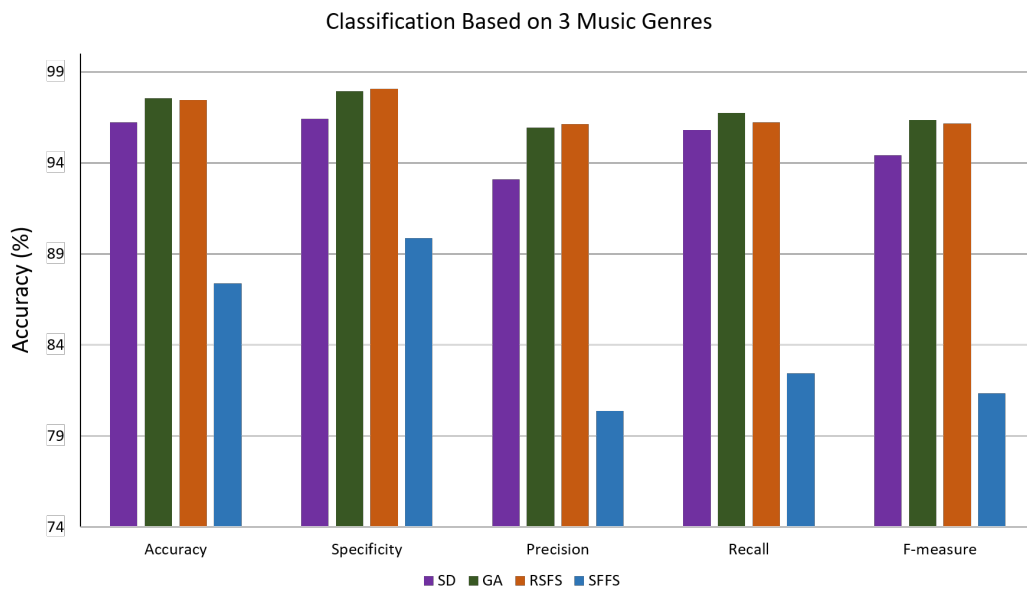


Figure 6.3: Classification results based on three music genres using EEG features, range 75-100 chosen for better visualisation

A further investigation was done to identify which feature selection methods are most suitable to use for the classification models. Figure 6.4 shows the NN accuracy results of three emotion scales using the SD, GA, RSFS and SFFS. Similar patterns are observed for other emotion scales as well.

It can be seen in Figure 6.4 that the feature selection methods achieve very close results in terms of accuracy. But when compared with other measures, the results show that GA and RSFS achieve the highest results in all evaluation measures for most cases. Table 6.3 shows the results of all evaluation measures for the same combinations shown in Figure 6.4.

The results are statistically significant ($p < 0.001$). It should also be mentioned that both these methods are feature subset selection algorithms, and they produced better results than feature ranking algorithms. Although the feature ranking algorithms get the highest accuracy in some cases, they do not consistently achieve high scores in other measures such as F1. One of the challenges of feature ranking meth-

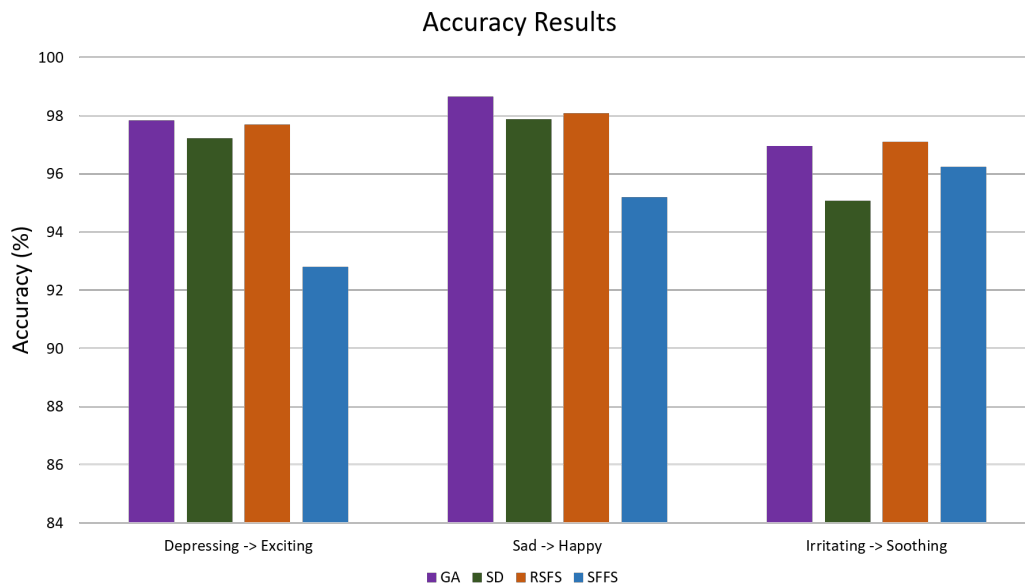


Figure 6.4: Classification accuracy using EEG signals based on participants' subjective response based on three emotion scales, range 84-100 chosen for better visualisation

Table 6.3: Evaluation measures of participants' subjective response using EEG signals based on three emotion scales

		SD	GA	RSFS	SFFS
<i>Depressing</i> → <i>Exciting</i>	Accuracy	0.972	0.978	0.976	0.928
	Precision	0.879	0.899	0.905	0.758
	Recall	0.968	0.981	0.964	0.865
	Specificity	0.973	0.977	0.979	0.941
	F-Measure	0.958	0.938	0.933	0.849
<i>Sad</i> → <i>Happy</i>		SD	GA	RSFS	SFFS
	Accuracy	0.979	0.987	0.981	0.952
	Precision	0.911	0.954	0.924	0.824
	Recall	0.968	0.967	0.965	0.909
	Specificity	0.981	0.991	0.984	0.961
F-Measure	0.939	0.96	0.944	0.863	
<i>Irritating</i> → <i>Soothing</i>		SD	GA	RSFS	SFFS
	Accuracy	0.951	0.969	0.971	0.963
	Precision	0.875	0.918	0.926	0.907
	Recall	0.915	0.965	0.961	0.948
	Specificity	0.963	0.971	0.974	0.967
F-Measure	0.941	0.941	0.943	0.927	

ods is choosing the optimum number of features. Feature subset methods are more beneficial in such cases.

6.3.4 Verbal Response Analysis

To identify whether participants' brain wave activity aligned with their verbal comments on the music pieces, a qualitative analysis was performed using a grounded theory approach [Glaser and Strauss, 2017] on the open-ended responses provided for each music stimulus. The responses were coded into higher level themes based on participant descriptions of the emotions they felt while listening to a particular stimulus. These codes were then divided into three categories: positive, negative, and neutral. Even though the questions were specifically related to the music stimuli, some comments were not relevant to the music pieces and only reflected the text reading experience. Thus, those words were discarded from further analysis. However, there were some comments that discussed both the texts and music pieces together, those were kept for further analysis. Figure 6.5 shows the word cloud created using the verbal comments provided by the participants.



Figure 6.5: Word cloud on verbal comments provided by the participants on the twelve music stimuli

During the coding process, frequently appearing words that were considered negative were: "dislike", "sad", "depressing" and "irritating". Some of the comments

highlighted as positive were: *"like"*, *"relaxing/relaxed"*, *"soothing"* and *"calming"*. The neutral comments mostly described some features about the music, or whether they had previously heard the song or not. These comments did not reflect participants' emotions, therefore they were considered neutral comments. Some of the common words used for neutral comments were: *"familiar"*, *"loud"*, *"slow"*, *"fast"* and *"upbeat"*. The analysis was completed using NVivo 12 software. Table 6.4 demonstrates the percentage of participants providing different categories of comments on each stimulus.

Table 6.4: Comments provided by participants on the twelve music stimuli

Stimuli	Negative Comments	Neutral Comments	Positive Comments
Classical 1	6.3%	37.5%	56.2%
Classical 2	18.7%	50.0%	31.3%
Classical 3	12.4%	43.8%	43.8%
Classical 4	12.4%	18.8%	68.8%
Instrumental 1	25.0%	50.0%	25.0%
Instrumental 2	6.3%	18.7%	75.0%
Instrumental 3	18.8%	18.7%	62.5%
Instrumental 4	18.8%	68.7%	12.5%
Pop 1	0%	68.7%	31.3%
Pop 2	12.5%	56.2%	31.3%
Pop 3	0%	62.5%	37.5%
Pop 4	18.7%	50.0%	31.3%

The results from 6.4 and verbal comments from participants give some useful insights. However, overall it did not demonstrate any distinguishable patterns. Below some findings are reported along with some participant comments. Participants will be referred as P1, P2, P3, . . . , P24.

Table 6.4 shows that eight out of the twelve music pieces received a majority of neutral comments. This means that the comments did not reflect their liking or emotional response to the music. One reason behind it could be the way the question was asked was not clear. Another reason could be the lack of clarity on the experiment design. Although the primary focus of this study was to look for effects of different music, this was not communicated to the participants. Participants were asked to listen to the music and read the text. But they were not told about the primary goal of this experiment. This was done intentionally so that participants do not avoid reading the text. However, many participants mistook the purpose of the experiment and assumed they will be asked questions about the text. Therefore, they aimed to mem-

orise parts of the texts, and music listening in general restrained them from doing so.

A number of the comments from participants reflected that the music did not allow them to give enough attention to reading the text. This was especially true for the pop music pieces used in this study. All four music pieces received a majority of neutral comments, which primarily focused on the reading distraction. P9's comments about *Pop 1* was: *"It was a little bit harder to read when the music was so catchy. Because it's such a popular song I knew all the words and was thinking them along with it"*. P15 described *Pop 4* saying, *"The underlying beat made me read at a faster rate, but again the lyrics make it somewhat difficult to concentrate on the text as well"*. Comparing the findings with the ones in section 4.3.2.1, we see that the pop song genre had mixed outcomes for the task of identifying emotions in videos (some participants reported that familiarity of the music helped them relax and focus on the task of watching videos). However, these music pieces certainly distracted participants from reading texts. The verbal comments, in general, did not reflect participants' emotional reaction to the music. This can be understood better through the subjective ratings participants provided on the music pieces, which were used for the classifier models. The distraction effect of music also needs to be investigated further using participants' EGG response.

Some participants also assumed the music and texts were aligned to evoke the same emotional response, which was not the purpose. Although the music were chosen to evoke specific emotional response, the accompanying texts appeared at random. For instance, P14 commented on *Classical 3* saying, *"It did fit with the tone of the text fairly well, and is an enjoyable piece to listen to when you're in a sombre mood. Not so much for when you're in a good mood though"*. Although the response was a bit unexpected based on the question, it aligned with the findings of another experiment reported in section 4.3.2.1 when music with sombre tone helped doing the task of identifying video emotions.

Another finding which was similar to section 4.3.2.1 was the effects of gamma wave inducing binaural beats. The music piece was used in this study as well (named *Instrumental 1*) and received a majority of neutral comments. Although the purpose of this music piece was to induce gamma waves in the brain and help with focus, it caused distraction to participants. P8 commented on saying, *"I don't like the music piece since it has some weird sound in it. Those sounds interrupted my attention while reading"*. The piece also received negative comments such as *"... don't like it as it is very sad piece and make me feel unpleasant"* (P1). Based on the results of two studies it is evident that careful consideration needs to be taken in choosing gamma wave inducing music to help with focus and attention.

Although this section provided insights on the effects some music had on text reading, it did not provide much information on participants' emotional response to music. The experiment design needs to be modified in the future to provide more

clarity on the questions.

6.3.5 Observation of Gamma Levels

As explained in chapter 2, gamma wave activity in the brain is crucial in various activities such as focus, attention, and epileptic seizures. The frequency band data collected by the EmotivPro software were further analysed for this purpose. The gamma level of every participant was observed when they listened to different music pieces. The songs were labelled based on the gamma levels seen in participants' brain activity while they were listening to a particular music piece. Then the pieces were divided into high, mid and low gamma levels. This division was made by averaging the gamma level score for every participant listening to every piece of music. This procedure was repeated for all 14 channels' gamma level information. A majority voting was performed among all channel's data to finally label the music piece. The results were the following:

- Low Gamma – Instrumental 1, instrumental 3, pop 1, pop 3, pop 4 (mostly pop)
- Mid Gamma – Classical 1, classical 2, classical 3, classical 4 (all classical)
- High Gamma – Instrumental 2, instrumental 4, pop 2 (mostly instrumental)

This division was very closely aligned to the different genres, with some interesting differences. It confirms some assumptions regarding the brain wave activity associated with the music pieces. For instance, instrumental 1 and 2 (both are binaural beats) were picked from YouTube and they were said to be inducing gamma waves and alpha waves in the brain respectively. The gamma level observation confirms this, as music instrumental 2 appears in the low gamma category (the piece was meant to be used for relaxation so low gamma level would be expected). Instrumental 1 appears in the high gamma category which also matches the description of the music. Both the binaural beats were able to induce the expected brain waves. Another observation was that all four of the classical music pieces appeared in the mid gamma level category. These music pieces are frequently used in music therapy as classical music pieces are said to be beneficial to reduce stress, anxiety and improve sleep patterns [Crawford et al., 2013; Thoma et al., 2013; Huang et al., 2017a]. However, they might not be very relaxing for all people. Pieces like binaural beats that induce more alpha waves can be of higher benefit in these cases. On the other hand, binaural beats that increase gamma levels can contribute to epileptiform activity. A detailed review on musicogenic epilepsy by Maguire mentioned that it has been hard to understand why neutral music like a specific sound triggers seizures, reported by some clinical studies [Maguire, 2015; Wieser et al., 1997]. The findings of this study may contribute to understanding this effect in the future by identifying music stimuli that demonstrates gamma wave activity associated with epileptic seizures.

A notable observation relates to the pop music pieces chosen for this study. Out of the four pop music, only one appeared in the high gamma category and the other

three appeared in the low gamma category. The assumption was that all of them would be in the mid or high gamma range as these music pieces contain a lot of lyrics and instrument usage and thus would require more concentration (usage of beta and gamma waves) while listening. One possibility might be the fact that these music pieces were all very popular in recent times, and most of the participants had listened to these pieces before (all four pieces in the pop category were known to the majority of the participants, as reported in the questionnaire). The fact that these pieces were already in their memory might have caused them to not concentrate as much while listening to the pieces. It has been reported before that there is correlation between high gamma activity and memory in the temporal locations of the brain [Gamma and Memory]. This was tested by observing the gamma activity in the temporal locations (channel P7, P8, T7 and T8). The results align with the literature (e.g. all four pop songs induce high gamma activity in P7 and mid gamma activity in P8). However, channels in the other locations do not follow the same patterns. It should also be noted that both temporal and frontal lobes have been shown to be regions where most epileptic seizures occur, especially in children [Epilepsy and Seizures; Childhood Epilepsy: The Brain]. Thus any music that reflects or induces these patterns in the brain of epileptic patients should be avoided. Further analysis using features from these regions can reveal the potential of identifying brain regions and music pieces that contribute to musicogenic epilepsy.

To observe if the division of the music pieces based on participants' brain wave level can be reflected computationally, classification using NN using all 26 features from every channel was performed. The labels were given according to the gamma levels of the music pieces. The model achieved the highest accuracy of 91.4% using the features from channel F3. This also aligns with the observation in Table 6.2 where some of the features extracted from channel F3 data were chosen a high number of times by all feature selection methods. These results were also compared based on all six evaluation measures from all channels using ANOVA test and the results show very high statistical significance ($p < 0.001$). Therefore, it can be concluded that signals obtained from specific channels could provide more valuable contribution to the computational models, compared to other channels. Using signals from more useful channels can greatly reduce the data collection and computational cost, and further increase efficacy of predictive models.

6.4 Summary

This chapter reported on a study that collected participants' brain activity via EEG signals while they listened to three different categories of music. Signals were collected using a 14-channel wearable headset, the Emotiv EPOC. Raw signals were first pre-processed by filtering them and dividing them into frequency bands alpha, beta and gamma. Then a number of linear and non-linear features were extracted from the frequency bands of all channels. A total of six feature selection methods were ap-

plied to select a feature set which were then used in NN, KNN and SVM classifiers. Analysis of the data showed that an NN model reached a high accuracy of 97.5% in classifying the music pieces based on genre and 98.6% in classifying the pieces based on the subjective rating of emotions given by the participants. The analysis also revealed that most of the useful features selected were from the pre-frontal and frontal regions of the brain. Thus in the following chapter, a study is reported where physiological signals are collected from the frontal region of the brain. An analysis is conducted using more advanced computational methods, and the limitations of the other experiments will be considered to improve the experiment design.

Effects of Music on Cerebral Hemodynamic Response

This chapter presents the outcomes of a study of the effects of three different music genres on people's cerebral hemodynamic responses. Participants' fNIRS signals were recorded while they listened to three different genres of music. Three commonly used machine learning and deep learning methods were applied to classify the physiological responses into the three genres. Classification was also performed based on the subjective responses of the participants. The contribution of this study is to analyse the effects of different types of traditional and popular music in participants' hemodynamic responses in the pre-frontal cortex using computational techniques. A comparison was also done with the brain wave responses analysed in the previous chapter. The chapter builds on the work submitted to the International Journal of Human – Computer Studies, where I was the primary contributor.

7.1 Experiment Design

The study was approved by the Human Research Ethics Committee of the Australian National University (ANU). After arriving at the scheduled time, participants were given an information sheet that included the description and requirements for the experiment. The document also highlighted potential risks, and how the data would be stored and used. Participants were given a consent form which they were required to sign before proceeding further in the experiment. The documents were similar to the documents shown in appendix B. Figure 7.1 shows a photo of the experimental setup.

In the first step of the experiment, participants sat in a chair in front of a 15.6 inch laptop where they were fitted with an Obelab NIRSIT device. The device was placed on the forehead of the participants. Participants were asked to move any hair from the forehead area in order to ensure good recordings. The calibration process began by first checking in the associated tablet application that all the points of the device connected properly and the application was able to visualise the blood flow in the participant's pre-frontal cortex. Then, participants were asked to move their head slightly in order to measure the baseline. The baseline signals were recorded



Figure 7.1: Experimental setting - participants fNIRS signals being collected while they listen to music

for about 50 seconds.

Participants answered some pre-experiment demographic questions on the laptop prior to the start of data collection. They also wore a pair of Bose QuietComfort® 20 Acoustic Noise Cancelling™ earphones to avoid any outside noise that might occur during the experiment. All the participants listened to all 12 pieces of music mentioned in Table 3.1. The genres were order balanced using the Latin square method to remove any ordering bias.

As fNIRS is a slow modality physiological signal [Peck et al., 2013], each music piece was played for two minutes in order to ensure opportunity for changes in participants' hemodynamic response during each song. Based on the previous experiments, it was noticed that two minutes of each music stimulus is enough for the participants to show a response to it; longer than that may cause boredom and distraction. After participants finished listening to one music piece, they were asked to give ratings to the music based on their general impression and their feelings while listening, using the same six emotion scales mentioned in the previous experiments. The entire experiment was conducted through an interactive website created using a Python Django web framework prepared for this purpose. The experiment took

approximately one hour including device setup and participation.

7.2 Participants

A total of 27 participants (17 female and 10 male) were recruited for voluntary participation in this experiment. Similar to the previous experiments, participants were recruited through the ANU SONA website. Their mean age was 19.4 with a standard deviation of 1.5 (range: 18–24 years). Most of the participants were undergraduate students at the Australian National University (ANU). A minority of them were post-graduate, diploma or high school students. Demographics of the participants of this study are shown in Table 7.1.

Table 7.1: Participant demographic of the experiment exploring the effects of hemodynamic response

Participant No.	Age	Gender	Ethnicity	Education
1	18	Male	Asian	Undergraduate
2	19	Female	Caucasian	Undergraduate
3	18	Female	Other	Other
4	20	Male	Asian	Undergraduate
5	20	Female	Asian	Undergraduate
6	21	Female	Asian	Undergraduate
7	18	Female	Asian	Undergraduate
8	19	Female	Asian	Other
9	21	Male	Asian	Undergraduate
10	18	Male	Asian	Undergraduate
11	18	Female	Asian	Undergraduate
12	20	Female	Asian	Undergraduate
13	19	Female	Caucasian	Undergraduate
14	18	Female	Caucasian	Undergraduate
15	24	Female	Asian	Postgraduate
16	22	Male	Asian	Postgraduate
17	18	Female	Asian	Undergraduate
18	20	Female	Asian	Other
19	21	Male	Asian	Other
20	19	Female	Asian	Undergraduate
21	19	Male	Asian	Undergraduate
22	21	Female	Caucasian	Other
23	18	Male	Asian	Undergraduate
24	20	Female	Asian	Undergraduate
25	18	Female	Asian	Other
26	19	Male	Caucasian	Undergraduate
27	19	Male	Caucasian	Other

7.3 Data Analysis

7.3.1 Pre-processing

fNIRS data using a NIRSIT device was collected at the sampling rate of 8.138 Hz. A number of pre-processing steps were done on the raw signals collected from the device. In selecting the device configuration for subsequent analysis, 750 nm wavelength and 30mm separation between channels were chosen as this is standard for many fNIRS-BCI studies [Shin et al., 2017]. From the 204 channels of the device, the 48 primary channels were used for further analysis. The raw signals were first low-pass filtered at 0.1 Hz and high-pass filtered at 0.005 Hz. Then some noisy channels were rejected based on their signal to noise ratio (SNR). Afterwards, the signals were filtered using the Modified Beer-Lambert law [Delpy et al., 1988]. This method converts the near-infrared signals to HbO₂, HbR and HbT (total hemoglobin) data and normalises the signals. This resulted in all values to be normalised within the range of -1 to 1 . Only HbO₂ and HbR values for each channel were used for further analysis. All pre-processing steps were done using the Matlab NIRSIT Analysis Tool. Finally, the signals were segmented into two minute lengths to identify the effects of each music piece.

7.3.2 Feature Extraction

A number of features were extracted from the pre-processed HbO₂ and HBR signals to be used in the machine learning methods applied. The features used in this study are listed in Table 7.2.

Table 7.2: Features extracted from fNIRS signals

Feature Type	Feature Names
Time Domain (Linear)	Mean, maximum, minimum, standard deviation, interquartile range, variance, summation, skewness, kurtosis, number of peaks, root mean square, absolute summation, difference absolute standard deviation value, simple square integral, average amplitude change, means of the absolute values of the first and second differences
Time Domain (Non-Linear)	Hjorth parameters (mobility), Hurst exponent
Frequency Domain	Mean, minimum and maximum of the first 16 points from Welch's power spectrum

7.3.3 Classifiers

Features extracted from the signals were further analysed using two commonly used classification methods, k-nearest neighbor (KNN) and random forest (RF). The meth-

ods are described in sections 3.2.5.1 and 3.2.5.3 respectively. Different values of parameters were experimented with and suitable parameters which led to optimum results were picked. For the KNN method, $k = 5$ and the Chebyshev distance metric was chosen. For the RF method, the number of trees chosen was 1000 with maximum depth of 20. A leave-one-participant-out approach was used to evaluate the models.

Biomedical signals such as fNIRS can be represented in two formats, one-dimensional (1D) and two-dimensional (2D) data. In this study, a 1D convolutional neural network (CNN) was created which is used for classifying time-series data.

7.3.3.1 Stacked Ensemble Models

The one-dimensional CNN (1D CNN) network used the pre-processed time-series data obtained after completing the steps in section 7.3.1. As this model takes in the time-series fNIRS signals as the input (without any handcrafted features), it introduces some additional challenges. Every participant's neural structure is different, which results in high variance in their physiological signals. Even after pre-processing, there remain differences in individuals' responses. Therefore, the classifiers need to be trained on a per individual basis to identify useful features from each participant.

During the pre-processing stage, it was found that each participant had different numbers of channels that recorded good quality data. After removing some channels based on low signal to noise ratio, each participant was left with different numbers of channel data. Thus, the sample size of each participant was different. This produced an additional challenge for the dataset. If all the participants' data are used together to train the model, some participants who had lower amounts of data would experience low training accuracy and this would have a significant impact on the final prediction.

In order to overcome these challenges and combine each participants' output into the final output, an ensemble approach based model was created. Ensemble methods are used where a new model learns the best approach to combine predictions from multiple sub-models to determine the final prediction result. This provides better generalisation and often results in better accuracy compared to using a single model. Ensemble models have been used in traditional machine learning techniques for quite some time. Recently, deep ensemble models have gained popularity as they combine the advantages of deep learning models and ensemble models. There are different techniques of creating ensemble models. Some of the techniques include bagging, boosting, and stacking. Stacked ensemble based deep learning methods have been used in studies where time-series sequences were used [Palangi et al., 2014]. It has most commonly been used in speech recognition [Deng and Platt, 2014; Deng et al., 2012; Tur et al., 2012] and speech emotion recognition [Zvarevashe and Olugbara, 2020]. Stacked approaches have also been used in music emotion recognition [Malik

et al., 2017]. Furthermore, stacked ensemble approaches recently achieved impressive results classifying physiological signals from the DEAP dataset, which contains EEG and EMG data [Bagherzadeh et al., 2018]. Therefore, in this study, a novel stacked ensemble model was created using participants' fNIRS signals.

There are multiple ways to create stacked ensemble models. Different models in the ensemble can be created using different techniques (e.g. KNN, SVM, NN), which is called model based fusion. Another way is to combine the weights of multiple neural networks having the same structure, called decision based fusion. The latter approach was adopted for this study.

In the stacked ensemble based approach, each sub-model provides a contribution to obtain the final prediction output. The model consists of two stages. In the first stage, a model is trained on each participant's data to create each sub-model. In the second stage, a meta-learner model is created based on the outputs from the sub-models in the first stage. The meta-learner model is then validated on a new participant's data to make a final prediction. In this scenario, a subject independent k-fold cross validation approach to validate the model. This approach is also used in similar analysis using EEG data [Jiang et al., 2021].

The 1D CNN model in the first stage was created as follows. It has two convolutional layers, one max pooling layer, one fully-connected dense layer, two dropout layers and a softmax classifier. In both convolutional and dense layers, a rectified linear unit (ReLU) was used as an activation function. The dropout layers were used after the convolutional layers and the dense layer to perform better regularization. Mean squared error was used as the loss function. For the optimisation algorithm, a stochastic gradient descent (SGD) with a momentum of 0.9 and a decaying learning rate was used, with an initial learning rate of 0.01, and a mini batch size of 64. The maximum epoch number was set to 200. The schematic diagram of the 1D CNN model is shown in Figure 7.2.

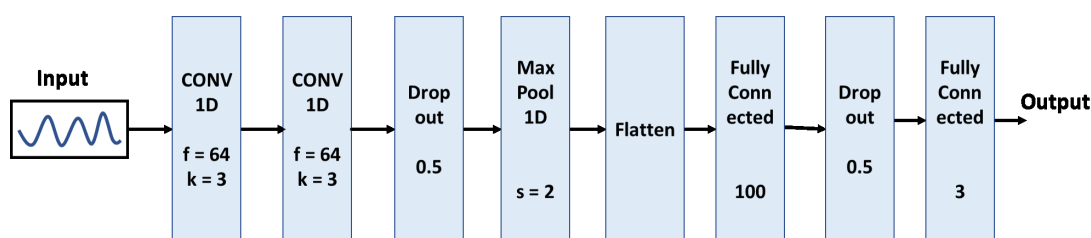


Figure 7.2: 1D CNN architecture for fNIRS signals classification

In the meta-learning stage, the output of the sub-models were fed into a shallow neural network with one dense layer and one softmax classifier. The schematic diagram of the overall ensemble model is shown in Figure 7.3.

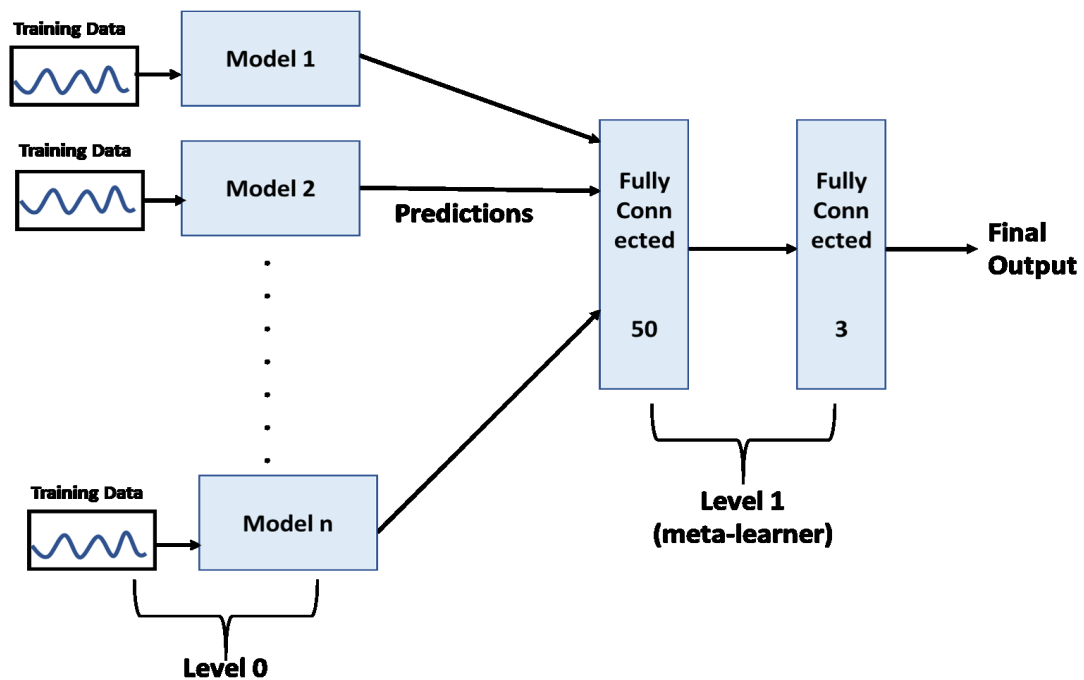


Figure 7.3: Stacked ensemble model architecture for fNIRS signals classification

For all of the classification tasks, four evaluation measures are reported. They are: classification accuracy, precision, recall and f-measure. Classification was done using the TensorFlow framework with the Python Keras library. The system specifications were an AMD Ryzen 7 3700X 8-core processor with 3.59 GHz, NVIDIA GeForce GTX 1660 SUPER GPU, 16.00 GB of RAM and Microsoft Windows 10 Enterprise 64-bit operating system.

7.4 Results and Discussion

During data pre-processing, it was found that three participants' fNIRS data were incomplete. Therefore, those participants' data were discarded, and classification was performed using data from the remaining 24 participants. For all the subsequent computational analysis, two types of classification are reported using traditional machine learning and deep learning techniques. The first is classification by music genre, where the three genres provided the three classification labels. The other is classification based on the subjective rating of participants' emotions, where the six different emotion ratings given by the participants were used as labels. The ratings were the same as described in the studies reported in chapter 5 and 6. All of these were converted into three class problems. As a reminder, the 7-point Likert scale responses for all emotion scales were converted into three categories (positive, negative

and neutral). A majority voting method was applied to determine the final label for each music stimulus. However, for three emotion scales (*disturbing* → *comforting*, *depressing* → *exciting*, *irritating* → *soothing*), only two out of the three categories received votes by the participants. The votes were either in the positive or neutral category. Thus, those three emotion scales were converted into binary classification tasks, while the other three (*sad* → *neutral* → *happy*, *unpleasant* → *neutral* → *pleasant*, *tensing* → *neutral* → *relaxing*) remained ternary classification tasks. It is also important to note that while the genre based classification had the same number of samples in each class, the subjective rating based classification had uneven numbers of samples in each class, leading to an imbalanced dataset. The other evaluation measures (precision, recall and f1-score) are useful in such cases as they account for the weight of each class.

In the following subsections, the key findings derived from qualitative, quantitative, visual and computational analysis conducted on participants' fNIRS and subjective response data are reported.

7.4.1 Classification Results

Table 7.3 reports the four evaluation measures using participants' fNIRS signals in three different combination (HbO₂, HbR and both).

Table 7.3 shows the four evaluation measures for all seven (one genre based and six subjective rating based) classification problems using KNN, RF and 1D CNN model. It shows that the highest evaluation measures in all seven classification problems were achieved by the 1D CNN model. The classification accuracies of the 1D CNN model in classifying three genres using HbO₂, HbR and a combination of both signals are 69.6%, 61.4% and 73.4% respectively. The other evaluation measures also achieved highest scores using a combination of both signals (0.762 precision, 0.734 recall and 0.731 f1-score). Classification using participants' subjective responses in a three class category achieved up to 77.4% accuracy in classifying *sad* → *neutral* → *happy* emotion. For the binary classification, the accuracy reached 80.5% in classifying *irritating* → *soothing* emotion. Compared to the 1D CNN model, the traditional machine learning techniques achieved 59.4% accuracy in ternary classification and 74.9% accuracy in binary classification. Both were achieved using all of the extracted features and the RF method. A one-way ANOVA test on the accuracy results of the three methods showed high statistical significance ($p < 0.001$).

It is important to note that in all seven cases the highest accuracy was achieved by using both HbO₂ and HbR signals together, followed by only HbO₂ signals and only HbR signals. Therefore, it can be suggested that using the combination of both hemoglobin concentration values is more beneficial in building a robust computational model. If it is not possible to collect both types of data, collecting only HbO₂ data would be more useful than collecting HbR data. The outcome is similar to some

Table 7.3: Evaluation measure results of KNN, RF and 1D CNN using fNIRS signals

Label	Signal	KNN				RF				1D CNN			
		Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
<i>Classical</i> →	HbO2	0.342	0.34	0.342	0.334	0.327	0.326	0.327	0.321	0.696	0.724	0.696	0.689
<i>Instrumental</i>	HbR	0.339	0.336	0.339	0.33	0.341	0.342	0.34	0.336	0.614	0.649	0.614	0.602
→ <i>Pop</i>	HbO2+HbR	0.371	0.378	0.371	0.369	0.376	0.374	0.376	0.368	0.734	0.762	0.734	0.731
<i>sad</i> →	HbO2	0.495	0.455	0.495	0.467	0.553	0.449	0.553	0.461	0.74	0.758	0.74	0.707
<i>neutral</i> →	HbR	0.49	0.452	0.491	0.466	0.56	0.46	0.56	0.466	0.67	0.66	0.67	0.614
<i>happy</i>	HbO2+HbR	0.541	0.512	0.541	0.521	0.594	0.538	0.593	0.519	0.774	0.786	0.774	0.749
<i>unpleasant</i> →	HbO2	0.437	0.424	0.437	0.427	0.464	0.422	0.464	0.428	0.694	0.716	0.694	0.668
<i>neutral</i> →	HbR	0.451	0.433	0.451	0.438	0.486	0.451	0.486	0.446	0.587	0.589	0.587	0.523
<i>pleasant</i>	HbO2+HbR	0.489	0.476	0.489	0.479	0.517	0.478	0.517	0.481	0.734	0.748	0.734	0.717
<i>tensing</i> →	HbO2	0.452	0.439	0.452	0.442	0.476	0.444	0.476	0.446	0.697	0.719	0.697	0.682
<i>neutral</i> →	HbR	0.442	0.425	0.442	0.429	0.472	0.437	0.472	0.435	0.619	0.638	0.619	0.597
<i>relaxing</i>	HbO2+HbR	0.479	0.463	0.479	0.466	0.491	0.451	0.491	0.457	0.719	0.741	0.719	0.708
<i>disturbing</i> →	HbO2	0.517	0.512	0.516	0.513	0.541	0.505	0.54	0.493	0.708	0.734	0.708	0.674
<i>neutral</i> →	HbR	0.518	0.509	0.518	0.512	0.549	0.513	0.549	0.499	0.626	0.644	0.626	0.546
<i>comforting</i>	HbO2+HbR	0.539	0.533	0.538	0.533	0.544	0.511	0.544	0.496	0.718	0.743	0.718	0.684
<i>depressing</i> →	HbO2	0.596	0.559	0.595	0.57	0.649	0.555	0.649	0.557	0.749	0.747	0.749	0.697
<i>neutral</i> →	HbR	0.595	0.556	0.595	0.568	0.651	0.544	0.651	0.55	0.692	0.674	0.692	0.605
<i>exciting</i>	HbO2+HbR	0.648	0.633	0.658	0.637	0.684	0.668	0.683	0.618	0.77	0.772	0.77	0.731
<i>irritating</i> →	HbO2	0.706	0.638	0.706	0.659	0.744	0.631	0.742	0.653	0.794	0.791	0.794	0.734
<i>neutral</i> →	HbR	0.7	0.629	0.7	0.652	0.74	0.627	0.74	0.651	0.769	0.726	0.769	0.69
<i>soothing</i>	HbO2+HbR	0.748	0.727	0.747	0.73	0.769	0.749	0.768	0.702	0.805	0.799	0.805	0.753

papers in the literature where oxyhemoglobin features were shown to be more useful than deoxyhemoglobin and total hemoglobin features [Bauernfeind et al., 2014; Pathan et al., 2019]. The improved performance of the 1D CNN model over KNN and RF models highlights the benefit of using deep learning techniques over traditional machine learning techniques in physiological signal analysis. In traditional machine learning methods, identifying the useful features to extract is a difficult and time consuming step. Useful features also vary for different physiological signals. An additional step of feature selection may also be required to identify the useful set of features. Automatic feature extraction in 1D CNN removes the requirement of these steps and thus significantly reduces the time and complexity of the process.

A limitation of the stacked ensemble 1D CNN method is that it assumes every participant's model provides a useful contribution to the final model. However, during training it was noticed that some participants' models constantly achieved lower training accuracy compared to others. One of the reasons is likely to be the lower number of samples in some participants' data. Upon further investigation, it was found that some participants' had lower training accuracy, despite having higher number of samples. For instance, participant 21 and participant 24 had the same number of samples (2172 samples). However, training accuracy results of these two participants show that, the model of participant 24 is able to reach above 90.0% accuracy, whereas participant 21 can only reach training accuracy in the range of 60.0%. This requires further investigation to understand why certain participants' models reach low training accuracy although they had large number of samples. It could be that these participants' responses are different due to having different musical experience or preferences.

In order to understand the effects of different participants' models in the final ensemble results, participants' models were grouped into three categories based on their training performance. The groups are: group 1 (high), group 2 (mid) and group 3 (low). These different groups of models were used separately to build the final stacked ensemble models. The results are shown in Figure 7.4.

Figure 7.4 shows the classification accuracy of three groups using all three combination of signals. It is clear that there is difference in the results of these three groups, group 1 accuracy being the highest. Group 1 models reached 60.1%, 54.2% and 63.1% accuracy using HbO₂, HbR and combination of both signals respectively. In comparison, group 3 models reach only 55.7%, 50.7% and 56.2%. As the results of the stacked 1D CNN models which are mentioned in Table 7.3 includes all of these participants' models, it is assumed that some participants from group 2 and 3 would contribute to lowering the testing performance of the final stacked model. Therefore, it is necessary to discard lower quality data and collect a larger dataset to contribute to the final model.

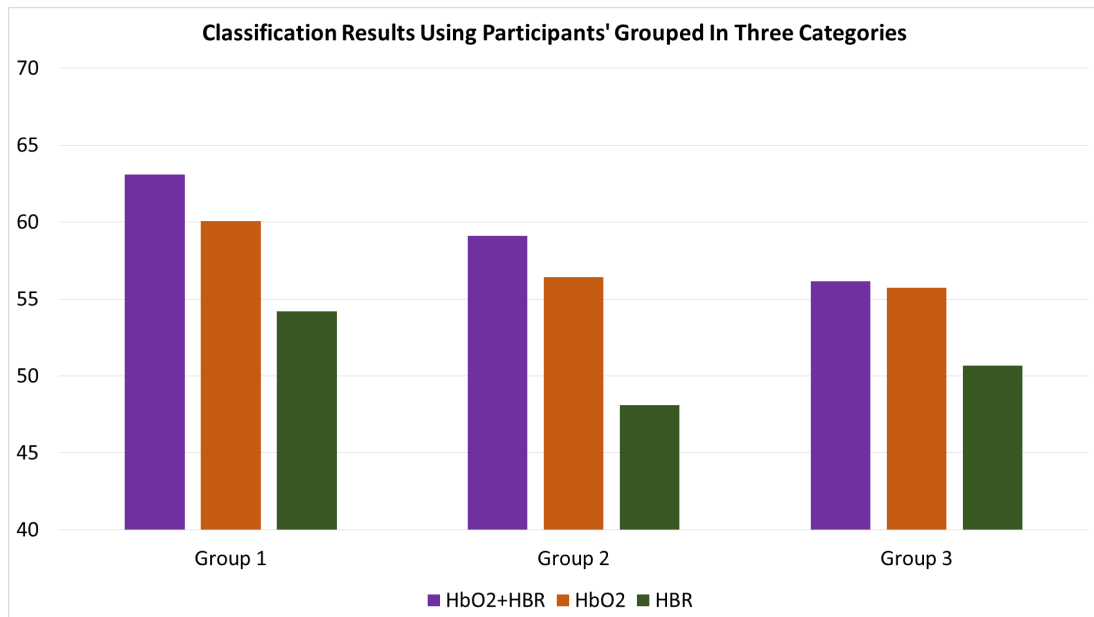
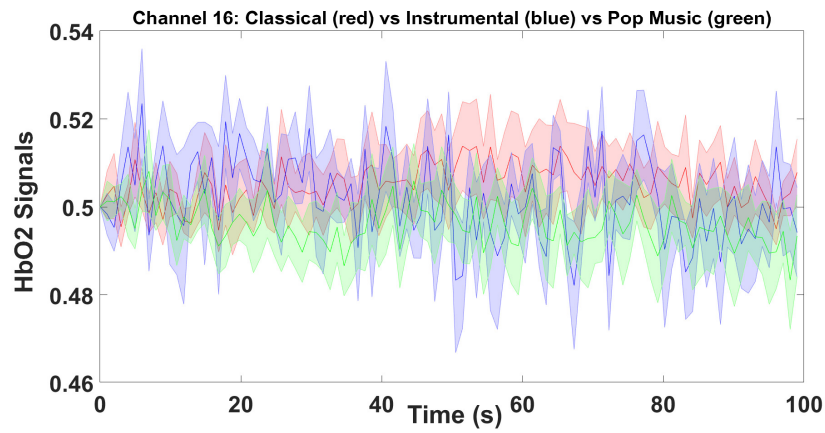


Figure 7.4: Classification results using fNIRS signals by grouping participants into three categories (range 40 – 80 chosen for better visualisation)

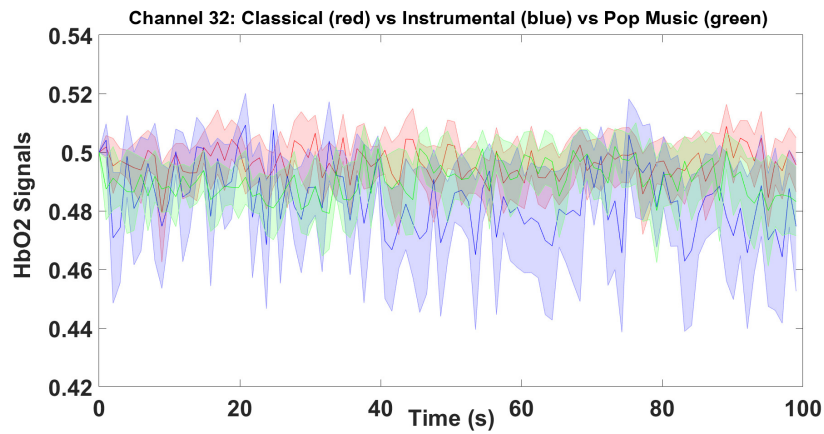
7.4.2 Visual Analysis

A visual analysis was conducted using the HbO₂ signals to understand how well the signals can differentiate the three music genres. A timeline analysis was performed on 100 seconds of signals recorded in two different stages of the experiment. The first stage signals were taken from point 900 to 1000 starting from the beginning of one genre, which is about 100 seconds after the start of presentation of music from that genre. The second stage was from point 2500 to 2600, which is about 300 seconds into listening to one genre. The analysis was performed on the average of all participants pre-processed HbO₂ signals from three different channels. Channel no. 16, 32 and 46 were selected from the left, mid and right side of the pre-frontal cortex respectively. These channels were chosen based on their overall good quality of data. The signals were reshaped to the initial value 0.5. This value was chosen so that the increasing or decreasing trend of fNIRS response could be seen in a clear manner. The result of the timeline analysis is shown in Figure 7.5 and Figure 7.6. The red shaded area shows participants' fNIRS responses to classical music, while the blue and green shaded area shows responses to instrumental and pop music respectively.

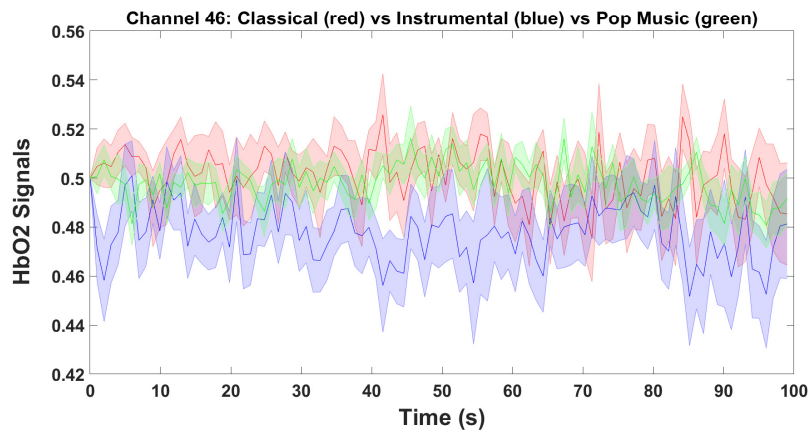
From Figure 7.5, it can be seen that participants' oxyhemoglobin response did not show much difference while they were listening to the three genres of music. These signals were captured while participants were listening to the first stimulus of each genre. However, participants' responses were more distinguishable in Figure 7.6, with stronger response seen during classical and instrumental music listening in



(a) Channel 16 : 900 - 1000 points

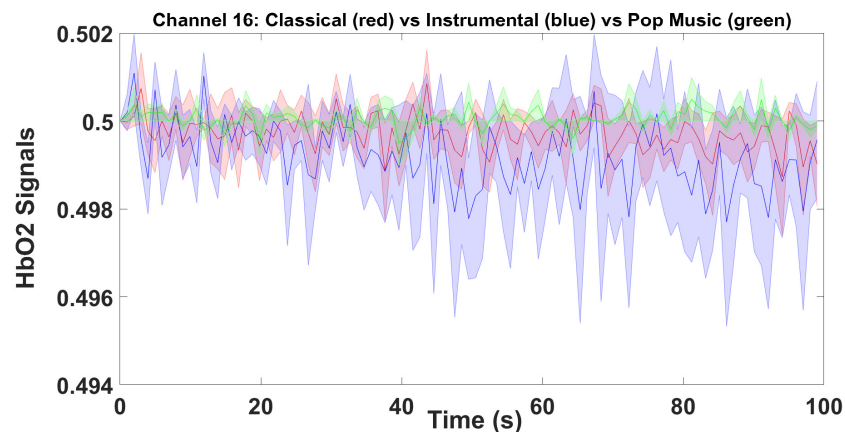


(b) Channel 32 : 900 - 1000 points

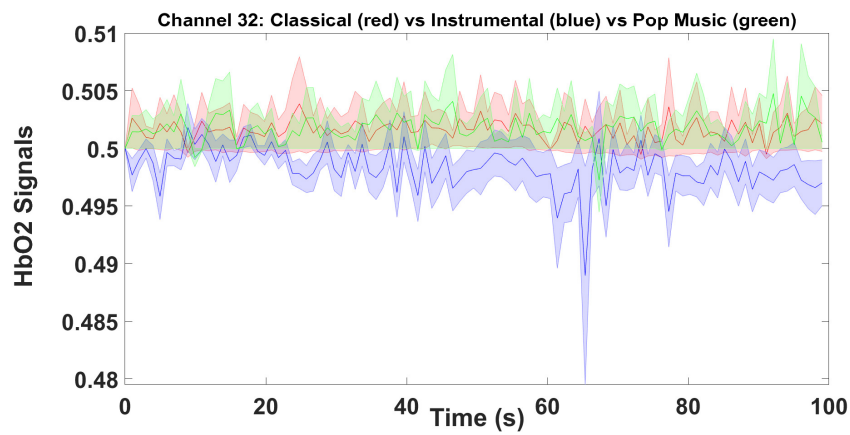


(c) Channel 46 : 900 - 1000 points

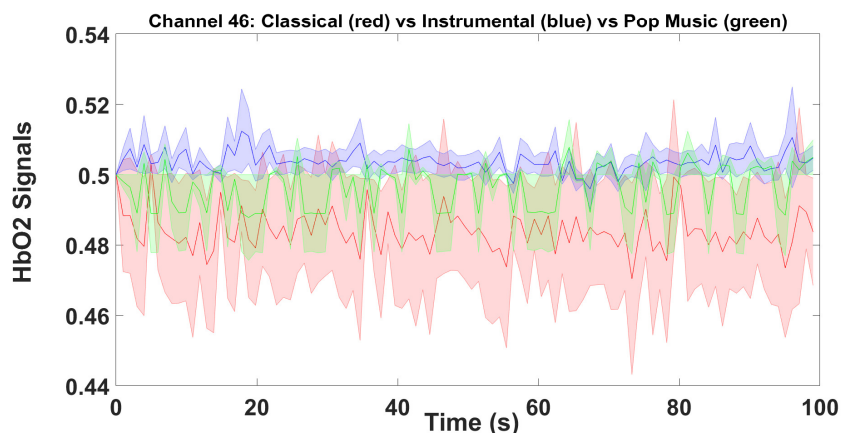
Figure 7.5: Timeline analysis of participants HbO₂ response to three music genres: points 900 - 1000



(a) Channel 16 : 2500 - 2600 points



(b) Channel 32 : 2500 - 2600 points



(c) Channel 32 : 2500 - 2600 points

Figure 7.6: Timeline analysis of participants HbO2 response to three music genres: points 2500 - 2600

the mid and right pre-frontal cortex. These signals represent the responses elicited during the third stimulus of each genre.

In summary, the figures show that the fNIRS signals provide a slow response in differentiating three genres. However, the responses become more prominent after the first few minutes, and show a more distinct range for the different genres, especially in the mid and right pre-frontal cortex. The mid region of the pre-frontal cortex is known for decision making and maintaining emotional information within working memory [Euston et al., 2012; Smith et al., 2018]. The right pre-frontal cortex is associated with self-evaluation of the face and episodic memory [Morita et al., 2008; Henson et al., 1999].

7.4.3 Music Offset Analysis

The 1D CNN model was further trained without the data from the first music track of every genre. This resulted in an increase in the classification accuracy to 75.7% using both HbO2 and HbR signals, 73.1% accuracy using only HbO2 signals, and 63.9% accuracy using only HbR signals. This could be due to the fact that fNIRS is a slow modality signal, so the effects of listening to a specific genre require time to be reflected in the signals recorded. Since the effect is seen in a delayed manner, it can be assumed that the effect of listening to one genre may be reflected after the playback was finished for one genre. Therefore, a further exploration was done by training the model with varied offset lengths of the final stimuli in every genre. The classification result in differentiating three genres is shown in Figure 7.7.

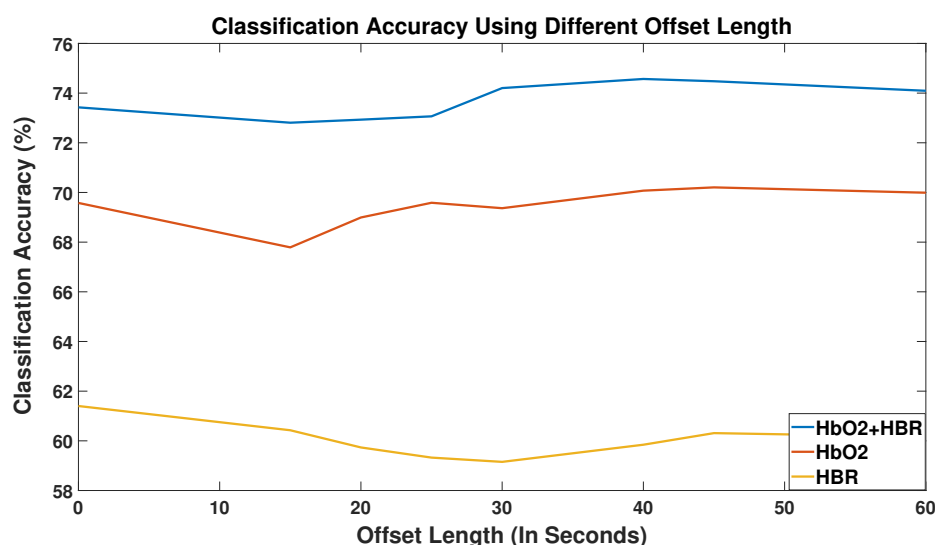


Figure 7.7: Classification accuracy with fNIRS signals using different offset length

Figure 7.7 shows that the classification performance decreases from the initial value of 73.4% (two minutes segment without any offset for any stimuli and not discarding first stimuli) to 72.8% using the offset length of 20 seconds. After that point, the accuracy starts increasing again and reaches 74.6% with the offset length of 40 seconds. Looking at the experiment participation of each subject, it was identified that this is the time period when they were completing the post experiment questionnaire. In particular, they were answering the open-ended question of providing any comments about the music they listened to. It could be that this question triggered the participants' memory of listening to the music and feeling the same emotion they felt while listening to it. Thus, this effect can be seen in their hemodynamic response. The same trend can be seen using only HbO2 or HbR signals. A one-way ANOVA test among the classification accuracy values showed high statistical significance ($p < 0.001$).

This result tells us that there is a lingering effect on brain patterns while reliving the experience of listening to music from each genre. Similar findings were reported by Chen et al. [2017a] where they noticed similar neural activity when participants watched and described the events of a TV show episode. The results also align with the results in the previous section where it is shown that the responses to different music genres became more prominent on different brain regions after the first few minutes of listening to the stimuli.

7.4.4 Verbal Response Analysis

Some expected and unexpected responses were observed for some of the stimuli, which were reflected in both participants' verbal and physiological signals. For instance, the stimulus *Instrumental 1* is a binaural beat designed to enhance gamma waves on the brain, thus increasing focus and concentration on tasks. However, the majority of the votes by participants leaned towards "sad", "unpleasant" and "tensed" rating for the three respective categories, and received neutral votes in the rest. This is contrary to expectations, as the assumption is that this stimulus would have a positive impact on participants' emotion. In addition, all of the classical music mostly received neutral votes from the participants. Depending on the stimuli, three out of the four stimuli were expected to evoke a positive response, and a negative response by the fourth one. The stimulus *Classical 3* is a piece played in a minor key and a very sombre tone. This piece has been used in funerals. However, participants mostly voted towards neutral or positive emotions for this piece. Although this aligns with some studies that mention sad music inducing pleasant emotions [Kawakami et al., 2013], the findings were still surprising and interesting.

To understand this effect in a greater detail, a qualitative analysis was performed using a grounded theory approach [Glaser and Strauss, 2017] on participants' comments provided for each music stimulus. The comments were coded into higher level themes based on participant descriptions of the emotions they felt while listening to

Table 7.4: Type of comments provided by participants on each music stimuli

Stimuli	Negative Comments	Neutral Comments	Positive Comments
Classical 1	7.4%	22.2%	70.4%
Classical 2	14.8%	22.3%	62.9%
Classical 3	33.3%	22.3%	44.4%
Classical 4	11.1%	18.5%	70.4%
Instrumental 1	70.4%	11.1%	18.5%
Instrumental 2	29.6%	11.1%	59.3%
Instrumental 3	18.5%	3.7%	77.8%
Instrumental 4	33.3%	29.6%	37.1%
Pop 1	7.4%	37%	55.6%
Pop 2	3.7%	40.7%	55.6%
Pop 3	11.1%	29.6%	59.3%
Pop 4	0%	44.4%	55.6%

discomfort in participants, which is likely to cause distraction and reduced focus. Similar outcomes were seen on a different set of participants (described in chapter 4). The pop music pieces received a mix of neutral and positive comments. However, both of these types of comments were influenced by the fact that these music tracks were more familiar (all of the participants were familiar with at least one stimulus in this category). This suggests that music stimuli invoking sad emotions or familiar music invoking positive emotion may both perform better in improving focus rather than binaural beats. This also aligns with the outcome described in chapter 4, where results showed that listening to this music helped improved focus in the task of detecting emotions from videos.

7.4.5 Activation Map Analysis

While some of the stimuli received a different emotion label than expected, the verbal response correlated with participants' hemodynamic responses. In order to analyse this, the image frames generated from the activation map videos by the Matlab NIR-SIT Analysis Tool were visually analysed. The activation map shows the changes of HBO2 and HbR in the pre-frontal cortex over time. The colorful areas show which areas in the pre-frontal cortex were activated at a given time, and the color intensity represents the value. The images were extracted at 25 frames/second and segmented according to the song length. Figure 7.9 shows a sample frame from the activation map videos.

The activation maps demonstrated a higher HbO2 response listening to stimuli

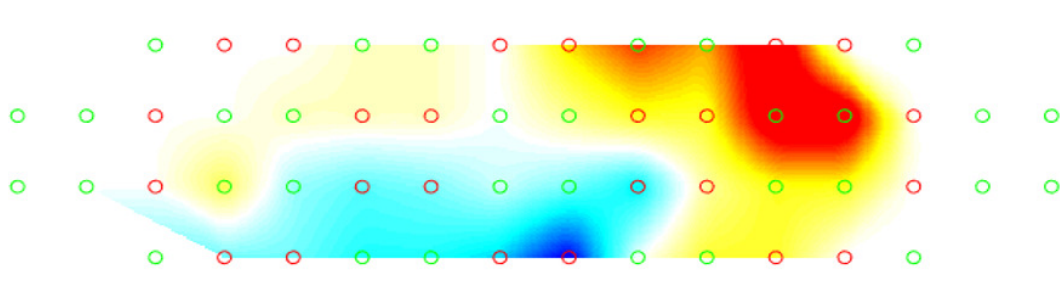


Figure 7.9: Frame from activation map video showing changes in HbO2

that were labelled sad compared to happy ones. Figure 7.10 and 7.11 shows sample frames of two participants listening to *Instrumental 1* (received mostly negative ratings) and *Pop 4* (received mostly positive ratings).

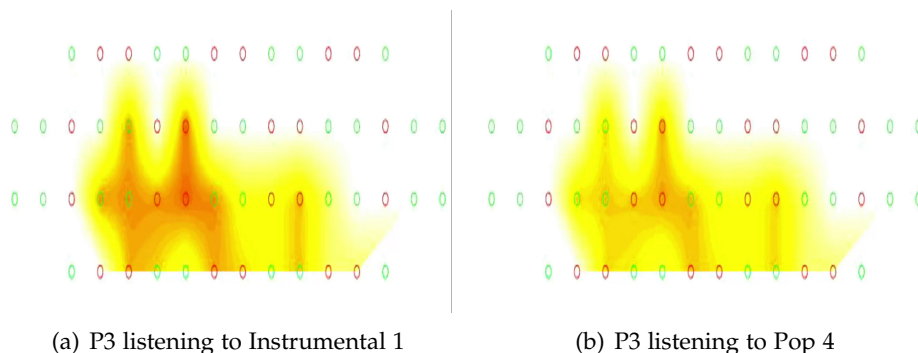


Figure 7.10: Frame 417 of P3 listening to Instrumental 1 and Pop 4

Figures 7.10 (a) and 7.11 (a) show that for both participants, there is a higher level of HbO2 activation in the mid pre-frontal area while listening to the piece *Instrumental 1*. This stimulus was voted by the majority of the participants as being in either neutral or negative categories such as "tensing" and "unpleasant". Figures 7.10 (b) and 7.11 (b) show lower HbO2 activation listening to Pop 4, which was voted in the positive categories such as "exciting" and "soothing" by the participants. The trend is observed for other participants as well. This is similar to the work by Moghimi et al. [2012] where they found larger peaks in HbO2 responses in negative emotion inducing music pieces. The findings of this study suggests that participants' hemodynamic responses are correlated with their emotional reaction. Therefore, these signals can reliably be used to build computational models to provide music recommendations based on human emotional states.

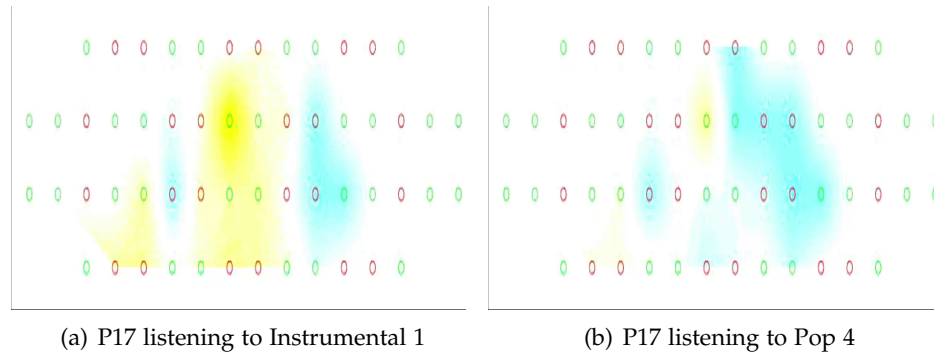


Figure 7.11: Frame 417 of P17 listening to Instrumental 1 and Pop 4

7.5 Comparison Between Two Types of Brain Data

Chapters 6 and 7 both explored the effects of music in people’s brain activity. While chapter 6 explored and analysed raw EEG and alpha, beta, gamma brain wave response, chapter 7 explored people’s hemodynamic response by analysing their HbO₂ and HbR data from fNIRS signals. The computational techniques used with these two types of data were also different. Both types of signals went through a number of pre-processing steps to remove noisy data. EEG data was analysed by extracting features from the brain wave data. This required an additional step of feature selection to identify the useful feature set. For fNIRS signals, the features extraction step was considered for two classification models and in addition, the pre-processed HbO₂ and HbR signals were directly used for classification using a deep learning model, skipping the steps of feature extraction and selection. In terms of classification models, EEG signals were classified using KNN, SVM and NN, and fNIRS signals were classified using KNN, RF and 1D CNN model.

The results show that using selected EEG features, a shallow NN model reached a highest level of 98.6% accuracy, whereas the highest accuracy using pre-processed fNIRS signals in a stacked ensemble 1D CNN model reached a maximum of 80.5% accuracy. The accuracy results might suggest that EEG signals perform better in classifying musical emotion response. However, this assumption is not necessarily correct. The sample size of data used for both these models vary widely (approximately 8,000 samples for EEG features and 24,000 samples for fNIRS signals). Therefore, although the accuracy of fNIRS models were lower compared to EEG models, they were analysed using more robust models, which means the results are expected to be more reliable. The shallow neural network may perform worse with a larger set of datasets, which makes the model difficult to apply in many real world applications. An additional advantage of the 1D CNN models is the removal of feature extraction and selection steps. This requires significant computational time and deeper understanding of the data to decide what features to extract. The 1D CNN model automatically extracts useful features from the data and removes the manual labor

of extracting features. Finally, it was reported in Table 6.2 that the most useful features for music emotion processing using EEG data were found in the frontal and pre-frontal region of the brain. The data collected using fNIRS signals covers that region and expands on it, therefore creating a larger corpus of data for emotional response to music listening.

It should be noted that for both these data considered only the temporal information, not the spatial information. As mentioned earlier in section 2.2.3.6, fNIRS data has higher spatial resolution, but lower temporal resolution than EEG. This has also contributed to the lower performance of fNIRS models. Combining spatial information of the data can lead to a more robust outcome for the computation models. The research solidifies the design recommendation of using fNIRS data over EEG for research involving emotion processing and decision making. For research that involves investigating biomarkers for musicogenic epilepsy, fNIRS signals have not been considered in the literature as yet. Future work can involve more extensive usage of fNIRS signals in this area. If fNIRS signals are not possible to collect, using brain wave data from EEG signals is recommended.

7.6 Summary

This chapter described an experiment that collected participants' brain activity response via fNIRS signals while they listened to three different genres of music. Signals were first pre-processed using different techniques to convert the raw signals into oxyhemoglobin (HbO₂) and deoxyhemoglobin (HbR) responses. Three well known machine learning and deep learning methods (KNN, RF and 1D CNN) were applied to classify the signals. Results from the analysis show that the deep learning models achieved higher accuracy in classifying different music based on their genres and participants' subjective rating of emotions. A 1D CNN model achieved 73.4% accuracy in classifying three music genres and 80.5% accuracy in classifying subjective rating of emotions based on the fNIRS data.

This study also identified the usefulness of combining HbO₂ and HbR signals to construct effective fusion based models. The study revealed that human brains process different genres of music differently and that can be seen in their fNIRS signals. It also revealed the strength of fNIRS signals' alignment with participants' emotional states. The results from all parts of section 7.4 indicate that participants' hemodynamic responses are a strong indicator of their emotional responses to music. As fNIRS is a highly portable and non-invasive wearable technology, multiple prospects from this study can be identified which could benefit future affective computing research.

Conclusion

This chapter concludes the thesis by highlighting the key contributions of the research. The aim of this chapter is to revisit the research questions outlined in section 1.2 of the introduction and summarise the findings of this research work in response to those questions. There are several current and future applications that can be identified from this research work. These are discussed in brief. Finally, some limitations and challenges of the experiment design and computational methods are discussed along with recommendations for future work to overcome these challenges.

8.1 Summary of Contributions

This research explored the effects of different types of music stimuli using a number of physiological signals collected from human participants through different experiments. Six different user studies were conducted and analysed for this purpose. The research work conducted can broadly be divided into three stages, with key contributions found in each stage.

In the first stage, some well-known databases of visual stimuli were used to evoke emotional reactions from human participants. Two preliminary studies were conducted to build robust computational models for analysing physiological signals. In addition, these works were aligned with the state-of-the-art research conducted in the area of affective computing. Thus these studies were designed to inform the third user study in this phase, which was to combine these studies and use music stimuli as a secondary stimuli for the experiment. Key contributions of this phase of the research are:

- Efficient computational models to identify different emotions from human participants when they watch emotional images and videos.
- Comparative analysis to show benefits of using multiple stimuli instead of single stimuli in human-centred experiments.
- Qualitative and quantitative analysis to understand the effects of different types of music when doing a task of identifying emotions from visual stimuli.

Based on the results of the studies in stage one, two more studies were designed in stage two which looked into the effects of only music stimuli in human physiological response. Five different types of physiological signals were recorded when participants listened to twelve different music stimuli, divided into three categories. A wide range of computational techniques were experimented with in this stage of the research. Key contributions of this stage are:

- Robust feature set extracted from physiological signals to use as input for classification models and comparative analysis of different feature selection techniques.
- Neural network based efficient classification models to differentiate music based on genres and participants' subjective responses using their physiological response.
- A novel visualisation technique using physiological signals which could leverage state-of-the-art deep learning based approaches with limited physiological data.

In the final stage of this research, a larger set of data was collected from participants when they listened to different music stimuli. The aim of this stage was to build deep neural networks using a larger dataset which would leverage the benefits of automatic feature extraction to reduce computation complexity in building robust classifiers. The contributions of this stage of the research are:

- Novel deep learning based model using model-based fusion to understand the effects of music in human brain activity.
- Qualitative, quantitative and visual analysis investigating the effects of music in cerebral hemodynamic responses.
- Offset analysis to provide design recommendations for human-centred experiments involving fNIRS signals.
- Comparative analysis of EEG and fNIRS signals and design recommendations for experiments involving these signals.

8.1.1 Answering the Research Questions

As outlined in section 1.2, this thesis set out to explore five research questions, which will be revisited in the sections below:

8.1.1.1 RQ1: Do Different Types of Music Generate Different Physiological Response?

In this research, studies were conducted to computationally explore the effects of three different music genres on human affective reasoning through a range of different bodily signals. Results from the computational analysis indicate that peoples' bodily signals are strong indicators which differentiate what genre of music they are listening to.

Chapter 5 dealt with a range of physiological signals such as EDA, BVP, ST and PD, while chapters 6 and 7 discussed analysis using EEG and fNIRS signals respectively. Chapter 5 showed that using an optimal set of features, a shallow neural network achieved the highest accuracy of 99.2% in classifying classical, instrumental and pop music genres. Computational analysis on the alpha, beta and gamma brain wave data from EEG signals reported in chapter 6 showed that a neural network model reached a high accuracy of 97.5% in classifying the music pieces based on three genres. Finally chapter 7 showed that oxyhemoglobin (HbO₂) and deoxyhemoglobin (HbR) responses derived from fNIRS signals are also a good physiological indicator in differentiating music genres. A one-dimensional CNN model achieved 73.4% accuracy in classifying three music genres. All of the studies reveal that people process different genres of music differently and that can be seen in their bodily reactions as measured by their physiological signals.

In addition to differentiating music genres, this research also identified a strong correlation between people's subjective and physiological responses to music. The results from different analyses conducted in this work showed that physiological responses provide high accuracy in binary and ternary classification of emotion. Chapter 5 showed that a high accuracy of 98.3% was achieved in classifying the *disturbing* → *comforting* emotion pair using only EDA signals. EEG signals also showed impressive results in identifying *sad* → *happy* emotions, with an accuracy of 98.7%. FNIRS signals can identify the same three emotions at 80.5% accuracy rate, as reported in chapter 7.

The studies reported in this thesis using music stimuli provide a key contribution to the area of affective computing, which have previously been dominated by the use of visual stimuli such as images and videos. These results provide a strong motivation for using physiological responses to measure various emotional responses evoked by musical stimuli. It can have a wide range of applications in medical and affective computing. Some potential application area for these studies will be highlighted in section 8.1.2.

8.1.1.2 RQ2: Do Other Stimuli (e.g. images, videos) Have Any Impact on These Physiological Responses?

Two short studies were reported in this thesis which investigated the effects of visual stimuli on human physiological responses. The results showed that computational models can effectively differentiate people's emotional reaction to these visual stimuli based on their physiological response.

Prior to looking into the effects of different types of music stimuli, research was also conducted to investigate effects of different emotional video and image stimuli. These were done as preliminary studies to explore the usability of these stimuli in combination with music stimuli to evoke emotional response. Two studies were conducted, both looking into people's EDA activity. While the first experiment looked into EDA response as participants only watched emotional videos, the second experiment focused on the effects while participants watched both video and image stimuli. Computational analysis of the data collected from these experiments showed that participants' EDA responses can be distinguished very clearly when they watch different types of image and video stimuli. In the first experiment, classification using a simple neural network resulted in a high accuracy of 94.8% in identifying the seven different emotional video categories. The experiment also revealed the efficacy of physiological data driven emotional modelling over standard emotion models. In the second experiment, participants' EDA responses were analysed using traditional machine learning methods. A standard decision tree based model using an optimal feature set was able to achieve up to 93.6% accuracy in classifying different types of smiles, shown in video and image format. Both experiments extended the previous studies in the area where computational models using physiological signals showed excellent performance in identifying basic emotions from visual stimuli.

Contrary to the relationship between participants' subjective and physiological responses to music stimuli, the studies involving image and video stimuli demonstrated a different type of relationship. It was observed that participants' intuitive response to identifying smiles from visual stimuli is not as aligned with their physiological response. Results from the study revealed that participants' verbal responses perform poorly compared to their physiological responses, as their correct response rate was only 59.8% accurate in comparison to their physiological response, which was 93.6% accurate. The result provides even further motivation to use physiological signals as the primary indicator of different emotional reactions in people.

8.1.1.3 RQ3: Can Other Stimuli (e.g. images, videos) be Combined with Music to Understand Their Combined Effects on Physiological Responses?

The third study in chapter 4 reported on an experiment that examined the combined effects of visual and audio stimuli. The outcomes of the study showed a decline in the classification performance of the computation models. The results suggest that when using music stimuli as secondary stimuli in combination with visual stimuli,

more experiments are needed to identify the appropriate music stimuli that assists in the primary tasks given to the participants.

In the study reported in section 4.3, six music stimuli were played to the participants while they performed the task of identifying emotions from videos, which was an extension of the second study reported in section 4.2. The results from this study showed a decline in the computation models' predictive power, with the highest accuracy of 68.6% using their EEG signals. The study also leveraged the use of automatic feature extraction of a deep convolutional network. The study showed that careful consideration needs to be taken in choosing the music stimuli, when they are being used to increase focus on another task. This study provided a key insight that commonly used music known to increase focus during educational activities resulted in declining outcome for the task of identifying emotions from videos. Therefore, this research work revealed that different music stimuli need to be explored based on the type of tasks.

This study however, consolidated the previous outcome where participants' physiological response performed better than their verbal response in identifying emotions from videos. In comparison to the 68.6% accuracy using EEG signals, participants' verbal response was correct only 62.5% times. Therefore, it can be concluded that physiological signals can provide a confident estimate of people's state when they interact with different types of stimuli.

8.1.1.4 RQ4: Which Physiological Signals Perform Better in Differentiating the Physiological Responses to Different Stimuli?

All six studies reported in this thesis looked into effects of music and visual stimuli on a number of different physiological signals. Based on the results of these studies, there is not one kind of physiological signal that has been found most appropriate in differentiating responses to varied stimuli.

A range of physiological signals was considered as input for the computational models designed for this research work. The physiological signals that were considered are: electrodermal activity (EDA), blood volume pulse (BVP), skin temperature (ST), pupil dilation (PD), electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS). Due to the portability and easy access of data (i.e. does not require additional subscription for the analysis software), EDA data have been collected in three out of the six user studies. In addition, efficacy of EDA data in recognising human emotions have been reported widely in the literature. Depending on the scenario, EDA, EEG and fNIRS signals all have the capability to capture participants' response to different types of stimuli. Results from this research work indicate that EDA responses show excellent performance in distinguishing responses to different visual stimuli such as images and videos containing different emotions. In addition, brain signals such as EEG and fNIRS signals have also showed impressive perfor-

mance in distinguishing effects of different music stimuli. Brain signals, fNIRS in particular, also provide additional insights into human decision making process, as reported in sections 7.4.2, 7.4.3 and 7.4.5. These signals can also be collected using easily portable devices so collecting these type of data are highly recommended to understand effects of music stimuli.

The duration of the stimuli used to provoke reaction also plays an important role in analysis of these signals. EDA and EEG are fast modality signals, which means distinguishable patterns can be identified within 1-2 seconds of interacting with the stimuli. This is the reason EDA data performed well with the video stimuli, as the majority of the stimuli were 1-3 seconds long. However, slow modality signals like fNIRS can be useful for experiments where the effects of the stimuli needs to be observed for a longer period. The music stimuli used in different experiments ranged from 2-4 minutes. To observe the response to the music stimuli over the period of entire music duration, fNIRS signals may be more beneficial.

However, whenever possible, the primary recommendation is to collect multiple physiological signals as fusion based models have shown better results than one signal based models. This is discussed in detail in response to the next research question.

8.1.1.5 RQ5: Which Computational Methods are Effective to Analyse Physiological Response to Music and Other Stimuli?

Various computational methods created and analysed in this research work indicate that neural networks using a robust set of physiological features and fusion based machine learning methods perform the best in analysing human physiological responses to music and other stimuli.

There are several computational models that were built and tested to understand the effects of music and other stimuli. These methods range from traditional machine learning approaches such as k-nearest neighbours (KNN), support vector machines (SVM), decision trees (DT), random forest (RF), to artificial neural network based approaches such as shallow neural networks (NN) and one-dimensional convolution neural networks (1D CNN). Based on all the studies, it is evident that neural network based approaches outperform traditional machine learning methods in analysing different physiological signals. With shallow neural networks, an optimal set of features is necessary to be extracted and selected in order to create a robust model. This requires deep understanding of the data type and characteristics. On the other hand, 1D CNN performs automatic feature extraction from the data which could be a great way to reduce computational complexity. However, one limitation with this approach is that it requires a large amount of data, which is often difficult to collect from a large number of human participants. In the absence of large amounts of data, a transfer learning based approach can also prove beneficial in analysing the physio-

logical signals. As shown in chapter 5, physiological signals can be visualised using the novel approach proposed in this thesis called Gingerbread Animation, which achieved an accuracy of 74.8% using a deep learning model with a small amount of data. This approach will allow researchers to leverage state-of-the-art computer vision approaches in analysing multiple physiological signals collected during affective experiments with high effectiveness.

It is also important to note that in the studies where multiple physiological signal were collected and analysed, a fusion of multiple signals performed much better in comparison to only one signal. The high accuracy in classifying three music genres reported in chapter 5 was achieved with a feature fusion based model using features from four types of physiological response, namely EDA, BVP, ST and PD signals. This result highlights the usefulness of fusion based approach in building classification models that require feature engineering of the data. Chapter 7 also demonstrated that combining both HbO₂ and HbR signals outperform the models that only used either HbO₂ or HbR data from the fNIRS signals. Furthermore, results also showed that a model-level based fusion in the 1D CNN model outperformed the other classification models. Therefore, it can be concluded that, fusion based computational approaches perform the best in analysing physiological response to music and visual stimuli.

8.1.2 Applications of the Work

There are several application areas where this research work would be useful. Some of these areas are music emotion recognition, music therapy, biofeedback training, wearable technology, epileptic seizure detection / reduction / avoidance. In the subsections below, some applications of this work are suggested.

8.1.2.1 Personalised Music Recommendations

Personalised music recommendation is currently one of the most popular research areas in the field of machine learning. This research has gained considerable interest among music platforms and service providers. Currently, the research in this field is dominated by recommender systems using music features such as pitch, tempo, lyrics, voice and social media tags [Su et al., 2013; Chang et al., 2018]. However, these features fail to identify a user's current physiological and emotional state, which are crucial for personalised music recommendation. The stacked ensemble based method proposed in section 7.3.3.1 could be used to create personalised models for every person, which will reflect their current emotional state. Based on that, appropriate music can be recommended.

8.1.2.2 Music Therapy and Biofeedback Training

Section 2.2.4.1 elaborated on many recent works where music stimuli have been used for therapy. Applications of these works include sleep quality improvement, stress

and anxiety reduction, managing depression, Parkinson's and Alzheimer's disease. One limitation of existing work is that only a handful of music stimuli are used in therapy. While those music pieces can show positive effects, they are often not the music pieces that are preferred by the participants. Music preferences depend on a variety of characteristics such as a person's age, gender, cultural background and emotional state. There can be many other music stimuli that can evoke the same or similar reactions in a person's response as the regularly used stimuli. Looking at the physiological response may be a more accurate way in determining which stimuli to use. This research work can be of use for this purpose.

One of the approaches where this research can be applied is in categorising relaxing music pieces for music therapy based on participants physiological responses. As previously mentioned, music pieces that induce more alpha waves or less gamma waves on the brain are more appropriate for music therapy. Results from the computational models using EEG signals showed that (as reported in section 6.3.5), the models can appropriately categorise music based on the gamma wave level. These models can also be expanded to categorise music based on alpha and beta wave levels. This is a more efficient way of choosing appropriate music for therapy, rather than just choosing from a limited set of classical or instrumental pieces.

The work can further be used in biofeedback training. In biofeedback training, users get real time updates on their physiological state so that they can control it to reduce stress and anxiety. The computational models created for this research can be used to provide users with real time updates on their emotional state while listening to a particular piece of music or watching a video. Users can then use this information to make informed decisions regarding their interaction with the stimuli.

8.1.2.3 Portable Device Creation

With the advent of modern wearable technology, collecting physiological signals from different parts of the body have become easier and cheaper. However, wearable devices that collect brain data are not as portable and comfortable as the devices that collect signals from skin or heart. In all the experiments conducted for this research that needed to collect brain data, participants reported discomfort due to the device after 45-60 minutes of the studies. This finding makes it difficult for these devices to be used for longer experiments, e.g. in-the-wild experiments. Therefore, it is important to create devices that are comfortable to wear long term. The experiments reported in chapters 6 and 7 showed that there are specific areas of the brain which are more active in processing emotional music and visual stimuli. Creating devices that collect data from only these regions can be more cost effective and comfortable for users. Two recommendations are given regarding future brain-worn devices based on this research work.

- Creation of wearable devices that collect EEG signals from only the pre-frontal and frontal region of the brain which can then be worn more comfortably for

longer duration experiments.

- Creation of wearable devices that collect fNIRS signals from only the medial pre-frontal cortex region of the forehead which can be used for longer duration experiments and continuous measurements.

Most of the current wearable devices that collect brain data are made for medical and research purposes. The above outlined approaches will make brain sensing devices more portable and easily accessible to users to use in daily life.

8.1.2.4 Musicogenic Epilepsy Detection and Reduction

The literature review presented in chapter 2 indicated that musicogenic epilepsy reduction still remains a mysterious area of research. Based on the musical features alone, it is unclear what type of music in particular are responsible for triggering seizures in epileptic patients. One crucial information regarding this is that the gamma waves in the brain have an important relationship with musicogenic epilepsy, which can be seen in their EEG response [Tayah et al., 2006]. This research work can be used to categorise music that can potentially trigger seizures and thus should be avoided by musicogenic epilepsy patients. Categorising music via genre alone is insufficient to distinguish the best pieces for music therapy; the brain wave activity induced by a specific piece of music may potentially trigger seizures. The computational models can indicate when someone is experiencing high gamma activity due to a music stimuli, and it can provide potential warning when gamma wave reaches a certain threshold.

8.2 Key Limitations of the Work

There are several challenges regarding the experiment design and data analysis conducted in this research. Some of them are described below:

8.2.1 Experiment Design Issues

Several issues regarding the experimental design of the studies were identified throughout the duration of this research. These need to be carefully considered for future experiments.

- Rest periods between stimuli were not considered carefully in the experiment design. The experiments showed that different physiological signals used in this study take different amounts of time to come back to the baseline. However, the time period between stimuli were kept consistent regardless of the type of physiological signal collected. This needs to be changed in future studies to consider appropriate rest times depending on what signals are being collected.

- Issues regarding device connectivity were not considered in depth. For EEG signals, generalised methods were applied in the pre-processing stage. However, it was not taken into account whether different participants had different connectivity levels for the channels (Emotiv EPOC reports four levels of connectivity). The device is sensitive to movement and might show poor connections in some channels during the experiment. These need to be analysed further. The issue has been improved on the next experiment where fNIRS signals were collected. A number of channel data were removed based on the SNR values. However, this resulted in a lot of data being discarded. A more efficient approach could be to apply different weights to channel values based on their connectivity levels. Using this approach, the channels with better connectivity will have a higher influence in the model, without the relatively poor quality data being discarded.
- In regards to the choice of pop music stimuli, there were no clear annotations that mapped the songs to a certain emotion. So in order to come to a resolution about what pop music should be chosen, popularity was used as the determining factor. As it is not based on emotion, this could have potentially biased the choice of stimuli.

8.2.2 Sample Size

Due to the difficulty of finding participants and collecting data, the number of samples was not very large for the majority of the experiments. The participant number of each experiment ranged from 20-27. According to a physiological data analysis study conducted by Hossain et al. [2018], the minimal suitable number of participants required to train a machine learning model is nine. In comparison, the participant numbers of the studies of this research can be considered reasonable. However, this number is also dependent on the sample size of the signals collected. In chapter 7, the stacked ensemble based deep learning approach could be applied on data collected from only 27 participants because the sampling rate and number of channels were much larger than the data collected in previous studies. Due to the limited number of collected samples, the more robust deep learning based approach could not be tested during all studies. So it can be argued that the number of participants is not enough to generalise the physiological activity of humans at scale. A larger number of participants need to be observed to see if patterns emerging from the current studies remain consistent in large scale studies.

8.2.3 Population Bias

It should be acknowledged that the demography of the participants may have introduced some biases in the studies. The participants of all of the six studies were students from ANU. The majority of the students were recruited from computer science and psychology. Therefore, their age range, study and music interests may have been similar. It has been reported that physiological activity can vary according to

the difference in stimuli types, participants' age and gender. Therefore, the models need to be validated using data from different age and occupation groups to confidently show the models are capable of generalising to wider populations.

8.3 Future Work

This research work opens the door to a wide range of future research directions. These can not only enrich the area of medical and affective computing, but also expand the area of virtual reality, edge computing and so on. There are also areas where computational approaches can be expanded and improved. Below are some suggestions for future research directions, broadly categorised to four points:

8.3.1 Experiments Conducted In-the-wild

The experiments designed for this research were conducted in controlled laboratory environments. This allowed the data collection process to be smoother due to limited movements. Future work should include similar experiments conducted in-the-wild. This means that participants will wear devices and interact with different stimuli during their normal daily activities, when different types of physiological signals will be collected. Further pre-processing techniques will need to be investigated which will look into movement signals due to walking or other activities.

8.3.2 Computational Approaches to Detect Emotional Response to Music in Real Time

The computational models built for this research were all offline classification methods. In the future, these models can be extended to provide feedback to the users in real time, such as indicating a high alpha, high gamma activity or high level of arousal. The real time feedback can help users make informed decisions on whether to continue engaging with the stimuli. The computational times of these models will also be explored to identify the the best models to be used for real time feedback.

8.3.3 Optimisation Methods to Find Appropriate Hyperparameters

Selecting the optimal hyperparameters of machine learning models is a challenging task. In this research, all the hyperparameters were chosen either based on experimentation or previous literature. Future work in this area could be to apply various optimisation methods such as grid search, random search, evolutionary optimisation and early-stopping based optimisation to select the optimal hyperparameters for the classifiers.

An additional limitation of the stacked ensemble 1D CNN method proposed in section 7.3.3.1 is that it assumes every participant's model provides a useful contribution to the final model. However, there were some participants whose models

resulted in poor training accuracy due to many noisy channels or low number of samples. Future work should include grid search and majority voting based methods to identify the best set of models for decision fusion.

8.3.4 Virtual Reality Applications

The computational models that uses EEG and fNIRS signals can further be translated into many virtual reality based applications. There are a few wearable virtual reality headsets recently created to capture brain wave signals such as Looxid Link[LooxidLink] and Next Mind [NextMind]. The current research work can be extended to analyse brain signals in virtual reality settings. Some of the potential works are:

- Exploring the effects of different background music in virtual reality games based on participants' brain activity response.
- Investigating the efficacy of virtual reality based remote/hybrid communication methods based on brain signals.
- Biofeedback therapy using VR based immersive visualisation.

8.3.5 Edge Computing Applications

Edge computing systems have recently showed tremendous success in collecting and processing a large volume of signals very fast and in real time [Sittón-Candanedo et al., 2019; Chen and Ran, 2019]. The current research work can further be expanded to build edge computing models that can conduct real-time and ubiquitous processing of large scale physiological data. This will have many applications in medical and e-health research.

Experiment Procedure and General Guidelines

This appendix lists some general guidelines and procedures that were followed during all six studies. It includes step-by-step procedures taken in each study. It also includes general guidelines regarding the usage of each equipment that were used to collect data.

1. When the participant arrives, they are greeted and the participation information sheet and consent form are handed to them. Then they are briefly explained what the experiment is about. A reminder is given that they have to answer all the questions before moving on to the next steps. Participants are also asked if they have any questions or concerns regarding the experiment.
2. The eye tribe server is started. The UI is launched and in the settings the sampling rate is changed to 60 Hz. Then the server is restarted. The participants are then asked to sit in a position where they are comfortable typing. Then the position of the eye tribe is changed accordingly to do the calibration.
3. The participants are then asked to wear the E4 on their non-dominant hand. The E4 is then started, it takes about 40 seconds to setup.
4. (Only for collecting EEG) The Emotiv headset is started and the positions of the sensors are moved to get better connectivity. Ideally all the sensor should be green, but the aim is to get good connectivity on all the sensors without spending too much time (if some stay orange that's okay). After the sensors are all connected, data recording is started (tick the 'include baseline' checkbox) and calibration process is done.
5. (Only for collecting fNIRS) The fNIRS headset is started and put on the participant's head (making sure they do not have any hair on their forehead). There

are 3 dots on the device, the middle one should be in line with the participant's nose. Then the software is started, and then 'Monitoring' is chosen. Then participants are handed the tablet (where the software is installed) and asked to follow the instructions for the calibration process.

6. The eye tribe UI is launched again to start recording the data. The participants are asked to try to not move too much during the experiment and not bring their hands in front of their face as it will disrupt the eye gaze collection process.
7. The website is started. Participants are asked to put the earphones on and start the noise cancelling button.
8. The E4 is tagged once right before the experiment to indicate the start. Then it is tagged every time a participant starts listening to a new music stimulus.
9. After the experiment is finished all recordings are stopped and data is saved according to specific naming conventions. E4 saves each data according to the start time of the experiment, and they are saved directly to the empatica account associated with the device. For the other devices, the following convention is used to save the data:

DateMonthYear_SubjectNoForThatDay(e.g.130519_1)

10. Finally, devices are removed from the participants and they are given a debrief on the experiment.

Experiment Related Documents

This appendix includes participation information sheet, consent form, questionnaires and SONA experiment sign up page for the study reported in study 3 of chapter 4. The documents for other studies reported in chapter 5, chapter 6 and chapter 7 follow similar pattern. Therefore, they are not included in this thesis.

B.1 Participant Information Sheet



Participant Information Sheet

Project Title: Music and Emotion

Experiment:

In this experiment, you will watch a selection of video while listening to some music pieces from different genres. Your task is to answer some questions about those videos and music pieces. During the experiment, you will wear a headset and a wrist worn device with an aim to collect your electroencephalogram (EEG), functional imaging of brain activities, Heart Rate Variability (HRV), Blood Volume Pulse (BVP) and Galvanic Skin Response (GSR). Your eye gaze and pupil dilation will be tracked by an eye tracking device placed in front of you.

Devices

In this experiment, your physiological signals will be captured by three devices:

1. Emotiv EPOC+ device, which is a black round headset collecting raw EEG data.
2. Obelab Nirsit device, which is a white round headset collecting functional images of brain activities
3. Empatica E4 device, which is a black roundish watch with a button on it, aiming to collect HRV, BVP and GSR;
4. TheEyeTribe eye-tracking device, which is a black bar tracking eye gaze point and pupil dilation.

Tasks

In this experiment, you will watch a series of videos while listening to music from the computer screen and answer questions involving those videos and music pieces. After completion of tasks, all sensors will be removed. Only during the observation and identification of the experiment, the equipment will be recording your biometrics.

At the start of the experiment, you will:

- i) Fill demographic information.
- ii) Read instructions and press the **Next** button.
- iii) Watch the videos while listening to some music and answer the questions. Press **Next** to continue.
- iv) Repeat step v) until all the texts have been displayed.
- v) Fill in the post-experiment questionnaires and press **Submit**.
- vi) Read the Thank you page and wait patiently for the experimenter to collect the form, and give you any further instructions.

Use of Data and Feedback

The data collected will be used to draw conclusions about certain interaction techniques and the nature of the tasks. Any data collected, either raw or processed, may be used research and publications. The data will be made unidentifiable so that no participant will be able to be identified from any data collected.

Voluntary Participation & Withdrawal



This usability experiment is completely voluntary. You may end the test session or ask for a break at any time. You may request that any or all data collected from you be destroyed. You have the right to completely withdraw from the experiment at any point with no explanation to the researcher. In this case, your data and personal information will be destroyed in accordance with the ANU Code of Research Conduct. You can ask that your name be deleted from our contact list for future testing at any time.

What does participation in the research request of you?

The main purpose of the user study is to collect data to enable useful information to be gained on the interface, the interaction techniques, and tasks. We will give you a pre- and post-task questionnaire that may contain some questions of an identifying nature. You do not need to complete these or any of the other questions if you have any objections to them. The task carried during the session will involve recording of EEG, HRV, BVP, GSR, brain imaging, eye gaze and pupil dilation data.

Location and Duration

The study will take place in N239, Level 2, Computer Science and Information Technology (CSIT) Building 108 on the ANU Campus. The time needed to complete this user study will be about 90-120 minutes in one standalone session. This time will include an introduction to the tasks, setup, and completion of the tasks mentioned above.

Incentives

No incentives are provided. Participants signing up via the SONA system gain course credits.

Risks

As the study is conducted in a carefully designed lab environment, all care will be taken to make participants as comfortable as possible, given the nature of the interaction tasks. Some physical discomfort such as wrist and muscle strain may occur with some people *including, in rare cases, motion sickness. Participants are free to request that your participation in the user study cease at any stage without explanation.*

Confidentiality

The data from the experiment will be made unidentifiable so that no participant will be able to be identified from any data collected. All results published will be in regards to the overall findings from the cohort of participants and not on an individual basis. Until that time, if you give your permission, your contact details will be retained for follow-up testing. The data may be used in follow-up research by researchers not listed on this form. All researchers that will gain access to the data collected in this research will be listed under the same human ethics protocol as the current researcher.

Data Storage

The data from the research will be stored securely at the CSIT Building, ANU. The data from the experiment will be made unidentifiable to retain privacy of each participant. The lookup for the unidentifiable data will be kept in a separate secure location so that participants information can be found in the case of their wanting access to their data or destruction of their data.



Australian
National
University

In accordance with the ANU Code of Research Conduct all data collected for the research will be stored for at minimum 5 years from the date of publication. After this period the data will be archived for follow-up research. The data will be kept in secure storage at the Research School of Computer Science, ANU.

Queries and Concerns:

If you have any further requests for information or queries regarding the study please contact on the contact information given below,

Jessica Sharmin Rahman
Office: N320, CSIT Building, ANU
Email: jessica.rahman@anu.edu.au

Prof Tom Gedeon
Office: N332 / N331, CSIT Building, ANU
Email: tom@cs.anu.edu.au
Phone: +61 2 6125 1052

Ethics Committee Clearance:

The ethical aspects of this research have been approved by the ANU Human Research Ethics Committee. If you have any concerns or complaints about how this research has been conducted, please contact:

Ethics Manager
The ANU Human Research Ethics Committee
The Australian National University
Telephone: +61 2 6125 3427
Email: Human.Ethics.Officer@anu.edu.au

B.2 Participant Consent Form

Consent Form

1. I consent to take part in the research study. I have read the information sheet for this research project and understand its contents. The information provided explains the nature and purpose of the research project, so far as it affects me, to my satisfaction. My consent is freely given.
2. I understand that if I agree to participate in the research project I will be required to perform some text reading and music listening tasks and that I may be asked to answer some questions regarding before and after to determine my experience of the environments. The process will require approximately one and a half to two hours of my time.
3. I understand that the user study is for the purpose of research. It may or may not be of direct benefit to me.
4. I understand that information gained during the research project may be published in this and subsequent research, and that my personal details will remain confidential. My name will not be used in relation to any of the information I have provided, unless I explicitly consent in writing to be identified.
5. I understand that personal information, such as my name, will be kept confidential so far as the law allows. This form and any other identifying materials will be safeguarded.
6. I understand that I may withdraw from the research project at any stage without providing any reason and that this will not have any adverse consequences for me. If I withdraw, the information I provide will not be used by the project.
7. I understand that my participation, withdrawal or non-participation will not directly affect assessment in any course, including at the Australian National University, and my participation is completely voluntary.
8. I understand that it is sometimes essential for the validity of research results not to reveal the true purpose of the research to participants. If this occurs, I understand that I will be debriefed as soon as is practicable after my participation and, at that time, given the opportunity to withdraw from the research and have records of my participation erased.

Please list any Special Considerations (e.g. any medical conditions) you have which you would like to bring to the attention of the user study supervisor

Name of Participant: _____

Signature: _____

Date: _____

B.3 SONA Experiment Sign Up Page

12/10/2021, 14:18

Study Information




Research School of Psychology Psychology Research Participation Scheme

Rahman Jessica (Researcher)

⚙️ Study Menu ▾

Study Information

Study Name	Music and Emotion 3
Study Type	 <p>Standard (lab) study This is a standard lab study. To participate, sign up, and go to the specified location at the chosen time.</p>
Study Status	<p>Visible to participants : Approved</p> <p>Active study : Appears on list of available studies</p>
Duration	60 minutes
Credits	1 Credits
Abstract	In this experiment, you will watch a selection of videos while listening to some music pieces from different genres. Your task is to answer some questions about those videos and music pieces while wearing several physiological-signal-capturing devices.
Description	In this experiment, you will watch a selection of videos while listening to some music pieces from different genres. Your task is to answer some questions about those videos and music pieces. During the experiment, you will wear a headset and a wrist worn device with an aim to collect your electroencephalogram (EEG), functional imaging of brain activities, Heart Rate Variability (HRV), Blood Volume Pulse (BVP) and Galvanic Skin Response (GSR). Your eye gaze and pupil dilation will be tracked by an eye tracking device placed in front of you.
Eligibility	Age: 18-40. Participants must have normal or corrected-to-normal vision and must be

https://anupsych.sona-systems.com/exp_info.aspx?experiment_id=650

1/3

12/10/2021, 14:18

Study Information

Requirements	comfortable listening to music wearing earphones
Preparation	It is advised to avoid using hair styling products (wax, gel, etc.) prior to the experiment as you will be wearing several headsets.

Restrictions ▼

Sign-Up Restrictions	<p>Must NOT have signed up or completed ANY of these studies:</p> <ul style="list-style-type: none"> • Music and Emotion • Music and Emotion 2
-----------------------------	---

Additional Study Information ▼

Participant Sign-Up Deadline	24 hours before the study is to occur
Participant Cancellation Deadline	24 hours before the study is to occur
HREC Approval Code	2018/489 (expires 13 March 2024)
Direct Study Link	<div style="border: 1px solid #ccc; padding: 2px; margin-bottom: 5px;"> https://anupsych.sona-systems.com/default.aspx?p_return_experiment_id </div> <p>This is a direct URL for participants to access the study. You may use this in an email or study advertisement.</p>
Date Created	29 May 2019

Researcher Information ▼

Researcher	Rahman Jessica ✉
-------------------	---

Study Menu

- [👁 View/Administer Time Slots](#)

- [📊 Timeslot Usage Summary](#)

- [⬇ Download Participant List](#)

- [✉ Contact Participants](#)

- [📧 View Bulk Mail Summary](#)

https://anupsych.sona-systems.com/exp_info.aspx?experiment_id=650

2/3

B.4 Experiment Steps and Questionnaire

The experiment website was designed using Python's Django web framework. After all the devices were properly connected and calibrated, participants ran the experiment website using Chrome web browser. Below are the sequence of pages and their contents summarised:

B.4.1 Introduction and Instruction Page

This page gives participants instructions on what they need to do in the experiment.

Instructions:

The main task is to determine whether the Emotions presented in these videos are real or acted.

*The types of emotion include **fear, surprise, anger, happy(smile), sadness and disgust***

Note that you might recognise some characters appearing in the videos, as the video sources include movies, TV shows and other possibly known video media. There is no guarantee they are expressing acted or genuine emotions.

After viewing a video, you will answer the following question:

How does this expression presented in the video look to you?

Genuine Emotion means the dominating emotion this person experienced is genuine.

Acted Emotion means this person acted the emotion.

After submission, you will be taken to the next page to answer the following questions:

How do you rate your confidence level?

- 1 means you are **not confident**.
- 2 means you are **a little confident**.
- 3 means you are **average confident**.
- 4 means you are **confident**.
- 5 means you are **very confident**.

Have you seen this video before?

Yes or No

click [here](#) for next page.

Figure B.1: Experiment introduction page

B.4.2 Pre-experiment Questionnaire Page

This page is used to collect some basic demographic questions. Below is a list of the questions.

Pre-experiment Questionnaire

Please answer all of the following demographic questions:

Age:

Gender:

Are you currently wearing glasses or contacts?:

What level of education have you completed (or currently completing)?:

How would you classify yourself?:

- Asian
- Hispanic
- Latino
- Indigenous
- Caucasian/White
- Black
- Other
- I would prefer not to say

What is/was your major/field of study?:

What is your native language?:

Do you have any history of Migraine/Severe Headache?:

How much time do you spend listening to music every day?:

What type of music do you mostly listen to?:

Do you play any instrument? If yes then please mention the instruments you play.:

Figure B.2: Pre-experiment questionnaire page

B.4.3 Playing and Experiment Questionnaire Page

The music stimulus is played in background, eight videos are displayed during each music duration. Participants are required to answer three questions for each videos and six questions for each pieces of music. Below are some screenshots of the website page that shows these steps:

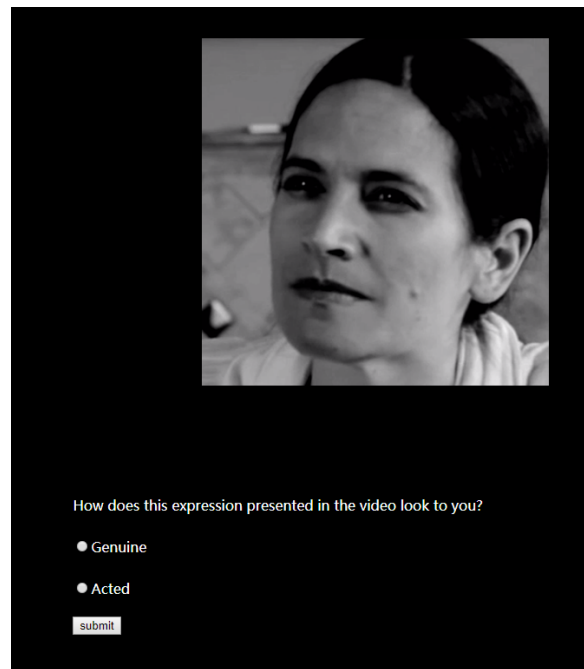


Figure B.3: Experiment video playing page

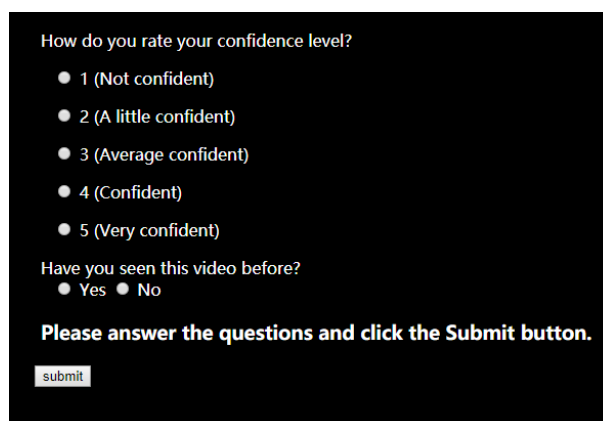


Figure B.4: Experiment video questionnaire page

Q1: Have you heard this piece before?:

Q2: How would you describe the music you have just heard?:

Comforting Moderately Comforting Somewhat Comforting Neutral Somewhat Disturbing Moderately Disturbing Disturbing

Depressing Moderately Depressing Somewhat Depressing Neutral Somewhat Exciting Moderately Exciting Exciting

Happy Moderately Happy Somewhat Happy Neutral Somewhat Sad Moderately Sad Sad

Unpleasant Moderately Unpleasant Somewhat Unpleasant Neutral Somewhat Pleasant Moderately Pleasant Pleasant

Q3: While the music was playing, how much attention did you give to the music?:

No Attention Very Low Attention Low Attention Neutral Some Attention Moderate Attention Listened Attentively

Q4: You found the music--:

Irritating Moderately Irritating Somewhat Irritating Neutral Somewhat Soothing Moderately Soothing Soothing

Q5: The music made you feel more--:

Relaxed Moderately Relaxed Somewhat Relaxed Neutral Somewhat Tensed Moderately Tensed Tensed

Q6: Why did you like/dislike this piece?:

Figure B.5: Experiment music questionnaire page

B.4.4 Post experiment questionnaire page

This is the final page of the experiment that asks a post-experiment question.

Post-experiment Questionnaire

Please answer the last question:

If any of the video or music pieces made you feel uncomfortable, why did you feel so?

[Finish the experiment](#)

Figure B.6: Post-experiment questionnaire page

Miscellaneous Materials

C.1 External Links for Experimental Materials

The experiment materials for the first and second study reported in chapter 4 can be found in the appendix chapter of the following document: Zakir Hossain ANU Thesis 2019.

C.2 Sample Video Link for Gingerbread Animation

A sample of Gingerbread Animation video can be found in the following link: [Gingerbread_Animation_Demo](#)

Bibliography

- ACHARYA, U. R.; HAGIWARA, Y.; DESHPANDE, S. N.; SUREN, S.; KOH, J. E. W.; OH, S. L.; ARUNKUMAR, N.; CIACCIO, E. J.; AND LIM, C. M., 2019. Characterization of focal eeg signals: a review. *Future Generation Computer Systems*, 91 (2019), 290–299. (cited on page 36)
- ACHARYA, U. R.; JOSEPH, K. P.; KANNATHAL, N.; MIN, L. C.; AND SURJ, J. S., 2007. Heart rate variability. In *Advances in cardiac signal processing*, 121–165. Springer. (cited on page 16)
- AFFECTIVA. Affectiva. q sensor. <http://www.affectiva.com/q-sensor/>. (cited on page 14)
- AGATONOVIC-KUSTRIN, S. AND BERESFORD, R., 2000. Basic concepts of artificial neural network (ann) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis*, 22, 5 (2000), 717–727. doi:[https://doi.org/10.1016/S0731-7085\(99\)00272-1](https://doi.org/10.1016/S0731-7085(99)00272-1). <https://www.sciencedirect.com/science/article/pii/S0731708599002721>. (cited on page 40)
- AHMED, T. U.; HOSSAIN, S.; HOSSAIN, M. S.; UL ISLAM, R.; AND ANDERSSON, K., 2019. Facial expression recognition using convolutional neural network with data augmentation. In *2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, 336–341. IEEE. (cited on page 12)
- AL MACHOT, F.; ALI, M.; RANASINGHE, S.; MOSA, A. H.; AND KYANDOGHERE, K., 2018. Improving subject-independent human emotion recognition using electrodermal activity sensors for active and assisted living. In *Proceedings of the 11th Pervasive Technologies Related to Assistive Environments Conference*, 222–228. (cited on page 14)
- AL MACHOT, F.; ELMACHOT, A.; ALI, M.; AL MACHOT, E.; AND KYAMAKYA, K., 2019. A deep-learning model for subject-independent human emotion recognition using electrodermal activity sensors. *Sensors*, 19, 7 (2019), 1659. (cited on page 14)
- ALHAGRY, S.; FAHMY, A. A.; AND EL-KHORIBI, R. A., 2017. Emotion recognition based on eeg using lstm recurrent neural network. *Emotion*, 8, 10 (2017), 355–358. (cited on page 42)
- ALJANAKI, A.; WIERING, F.; AND VELTKAMP, R. C., 2016. Studying emotion induced by music through a crowdsourcing game. *Information Processing & Management*, 52, 1 (2016), 115–128. (cited on page 28)

- ALPHA, 2017. Serotonin release music with alpha waves - binaural beats relaxing music, happiness frequency. <https://www.youtube.com/watch?v=9TPSs16DwbA>. (cited on page 30)
- ALWIN, D. F., 1997. Feeling thermometers versus 7-point scales: Which are better? *Sociological Methods & Research*, 25, 3 (1997), 318–340. (cited on pages 12 and 72)
- ANDERSEN, P. A. AND SULL, K. K., 1985. Out of touch, out of reach: Tactile predispositions as predictors of interpersonal distance. *Western Journal of Communication (includes Communication Reports)*, 49, 1 (1985), 57–72. (cited on page 13)
- ANDRASIK, F.; LARSSON, B.; AND GRAZZI, L., 2001. Biofeedback treatment of recurrent headaches in children and adolescents. In *Headache and migraine in childhood and adolescence*, 327–342. CRC Press. (cited on page 15)
- ARACENA, C.; BASTERRECH, S.; SNÁEL, V.; AND VELÁSQUEZ, J., 2015. Neural networks for emotion recognition based on eye tracking data. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, 2632–2637. IEEE. (cited on page 60)
- BAGHERZADEH, S.; MAGHOOLI, K.; FARHADI, J.; AND ZANGENEH SOROUSH, M., 2018. Emotion recognition from physiological signals using parallel stacked autoencoders. *Neurophysiology*, 50, 6 (2018). (cited on page 114)
- BAI, Y.; WELK, G. J.; NAM, Y. H.; LEE, J. A.; LEE, J.-M.; KIM, Y.; MEIER, N. F.; AND DIXON, P. M., 2016. Comparison of consumer and research monitors under semistructured settings. *Med Sci Sports Exerc*, 48, 1 (2016), 151–158. (cited on page 14)
- BALASUBRAMANIAN, S.; GULLAPURAM, S. S.; AND SHUKLA, A., 2018. Engagement estimation in advertisement videos with eeg. *arXiv preprint arXiv:1812.03364*, (2018). (cited on pages xxii and 96)
- BANDT, C. AND POMPE, B., 2002. Permutation entropy: a natural complexity measure for time series. *Physical review letters*, 88, 17 (2002), 174102. (cited on page 100)
- BARRETO, A.; ZHAI, J.; AND ADJOUADI, M., 2007. Non-intrusive physiological monitoring for automated stress detection in human-computer interaction. In *Human-Computer Interaction*, 29–38. Springer Berlin Heidelberg, Berlin, Heidelberg. (cited on page 40)
- BAUERNFEIND, G.; STEYRL, D.; BRUNNER, C.; AND MÜLLER-PUTZ, G. R., 2014. Single trial classification of fnirs-based brain-computer interface mental arithmetic data: a comparison between different classifiers. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2004–2007. IEEE. (cited on page 118)
- BERG, A. T.; BERKOVIC, S. F.; BRODIE, M. J.; BUCHHALTER, J.; CROSS, J. H.; VAN EMDE BOAS, W.; ENGEL, J.; FRENCH, J.; GLAUSER, T. A.; MATHERN, G. W.; ET AL.,

2010. Revised terminology and concepts for organization of seizures and epilepsies: report of the ilae commission on classification and terminology, 2005–2009. *Epilepsia*, 51, 4 (2010), 676–685. (cited on page 22)
- BERTINETTO, C. G. AND VUORINEN, T., 2014. Automatic baseline recognition for the correction of large sets of spectra using continuous wavelet transform and iterative fitting. *Applied Spectroscopy*, 68, 2 (2014), 155–164. (cited on page 35)
- BILLBOARD. Billboard year end chart. <https://www.billboard.com/charts/year-end>. (cited on page 30)
- BIOPAC. Eda electrodermal activity amplifier. <https://www.biopac.com/product/eda-electrodermal-activity-amplifier/>. (cited on page 14)
- BOBADE, P. AND VANI, M., 2020. Stress detection with machine learning and deep learning using multimodal physiological data. In *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, 51–57. IEEE. (cited on page 38)
- BODYBUGG. Apex-fitness. bodybugg. <http://www.bodybugg.com/>. (cited on page 14)
- BOS, M. G.; JENTGENS, P.; BECKERS, T.; AND KINDT, M., 2013. Psychophysiological response patterns to affective film stimuli. *PloS one*, 8, 4 (2013), e62661. (cited on page 92)
- BOTA, P.; WANG, C.; FRED, A.; AND SILVA, H., 2020. Emotion assessment using feature fusion and decision fusion classification based on physiological data: Are we there yet? *Sensors*, 20, 17 (2020), 4723. (cited on page 14)
- BOUSEFSAF, F.; MAAOUI, C.; AND PRUSKI, A., 2013. Remote assessment of the heart rate variability to detect mental stress. In *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, 348–351. IEEE. (cited on page 16)
- BRIEN, S. AND MURRAY, T., 1984. Musicogenic epilepsy. *Canadian Medical Association Journal*, 131, 10 (1984), 1255. (cited on page 22)
- BRITE23. Brite23 - artinis medical systems | fnirs and nirs devices-blog. <https://www.artinis.com/blogpost-all/category/Brite23>. (cited on page 19)
- CALDWELL, S.; GEDEON, T.; JONES, R.; AND COPELAND, L., 2015. Imperfect understandings: a grounded theory and eye gaze investigation of human perceptions of manipulated and unmanipulated digital images. In *Proceedings of the World Congress on Electrical Engineering and Computer Systems and Science*, vol. 308. (cited on page 21)
- CECCHI, S.; PIERSANTI, A.; POLI, A.; AND SPINSANTE, S., 2020. Physical stimuli and emotions: Eda features analysis from a wrist-worn measurement sensor. In *2020 IEEE 25th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, 1–6. IEEE. (cited on page 14)

- CHANG, S.-H.; ABDUL, A.; CHEN, J.; AND LIAO, H.-Y., 2018. A personalized music recommendation system using convolutional neural networks approach. In *2018 IEEE International Conference on Applied System Invention (ICASI)*, 47–49. doi:10.1109/ICASI.2018.8394293. (cited on page 135)
- CHAUDHURI, T.; ZHAI, D.; SOH, Y. C.; LI, H.; AND XIE, L., 2018. Random forest based thermal comfort prediction from gender-specific physiological parameters using wearable sensing technology. *Energy and Buildings*, 166 (2018), 391–406. doi:<https://doi.org/10.1016/j.enbuild.2018.02.035>. <https://www.sciencedirect.com/science/article/pii/S0378778817338446>. (cited on page 39)
- CHEN, H.; LUNDBERG, S.; ERION, G.; KIM, J. H.; AND LEE, S.-I., 2020. Deep transfer learning for physiological signals. *arXiv preprint arXiv:2002.04770*, (2020). (cited on page 42)
- CHEN, J.; LEONG, Y. C.; HONEY, C. J.; YONG, C. H.; NORMAN, K. A.; AND HASSON, U., 2017a. Shared memories reveal shared structure in neural activity across individuals. *Nature neuroscience*, 20, 1 (2017), 115–125. (cited on page 123)
- CHEN, J. AND RAN, X., 2019. Deep learning with edge computing: A review. *Proc. IEEE*, 107, 8 (2019), 1655–1674. (cited on page 140)
- CHEN, L.; GEDEON, T.; HOSSAIN, M. Z.; AND CALDWELL, S., 2017b. Are you really angry? detecting emotion veracity as a proposed tool for interaction. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction, OZCHI '17* (Brisbane, Queensland, Australia, 2017), 412–416. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3152771.3156147. <https://doi.org/10.1145/3152771.3156147>. (cited on page 31)
- CHEN, L.-L.; ZHANG, A.; AND LOU, X.-G., 2019. Cross-subject driver status detection from physiological signals based on hybrid feature selection and transfer learning. *Expert Systems with Applications*, 137 (2019), 266–280. (cited on page 42)
- CHEN, L.-L.; ZHAO, Y.; ZHANG, J.; AND ZOU, J.-z., 2015. Automatic detection of alertness/drowsiness from physiological signals using wavelet-based nonlinear features and machine learning. *Expert Systems with Applications*, 42, 21 (2015), 7344–7355. (cited on page 100)
- CHENG, M.; SORI, W. J.; JIANG, F.; KHAN, A.; AND LIU, S., 2017. Recurrent neural network based classification of ecg signal features for obstruction of sleep apnea detection. In *2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, vol. 2, 199–202. IEEE. (cited on page 41)
- CHILDHOOD EPILEPSY: THE BRAIN. Childhood epilepsy: The brain. <https://www.massgeneral.org/children/epilepsy/education/the-brain>. (cited on page 107)

- CHOI, E. J. AND KIM, D. K., 2018. Arousal and valence classification model based on long short-term memory and deep data for mental healthcare management. *Healthcare informatics research*, 24, 4 (2018), 309–316. (cited on page 42)
- CHOWDHURY, R. H.; REAZ, M. B.; ALI, M. A. B. M.; BAKAR, A. A.; CHELLAPPAN, K.; AND CHANG, T. G., 2013. Surface electromyography signal processing and classification techniques. *Sensors*, 13, 9 (2013), 12431–12466. (cited on page 36)
- CHUNG, S. Y. AND YOON, H. J., 2012. Affective classification using bayesian classifier and supervised learning. In *2012 12th International Conference on Control, Automation and Systems*, 1768–1771. (cited on page 40)
- COLLINS, A. AND KOEHLIN, E., 2012. Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS biology*, 10, 3 (2012), e1001293. (cited on pages 66 and 98)
- COPELAND, L.; GEDEON, T.; AND MENDIS, S., 2014. Fuzzy output error as the performance function for training artificial neural networks to predict reading comprehension from eye gaze. In *Neural Information Processing*, 586–593. Springer International Publishing, Cham. (cited on page 40)
- COPPOLA, G.; TORO, A.; OPERTO, F. F.; FERRARIOLI, G.; PISANO, S.; VIGGIANO, A.; AND VERROTTI, A., 2015. Mozart’s music in children with drug-refractory epileptic encephalopathies. *Epilepsy & Behavior*, 50 (2015), 18–22. (cited on pages 22 and 30)
- CRAIG, D. G., 2005. An exploratory study of physiological changes during “chills” induced by music. *Musicae scientiae*, 9, 2 (2005), 273–287. (cited on page 4)
- CRAWFORD, I.; HOGAN, T.; AND SILVERMAN, M. J., 2013. Effects of music therapy on perception of stress, relaxation, mood, and side effects in patients on a solid organ transplant unit: A randomized effectiveness study. *The arts in psychotherapy*, 40, 2 (2013), 224–229. (cited on page 106)
- CURTIN, A. AND AYAZ, H., 2019. Chapter 22 - neural efficiency metrics in neuroergonomics: Theory and applications. In *Neuroergonomics* (Eds. H. AYAZ AND F. DEHAIS), 133 – 140. Academic Press. ISBN 978-0-12-811926-6. doi:<https://doi.org/10.1016/B978-0-12-811926-6.00022-1>. <http://www.sciencedirect.com/science/article/pii/B9780128119266000221>. (cited on page 19)
- DAR, M. N.; AKRAM, M. U.; KHAWAJA, S. G.; AND PUJARI, A. N., 2020. Cnn and lstm-based emotion charting using physiological signals. *Sensors*, 20, 16 (2020), 4551. (cited on pages 41 and 42)
- DE WITTE, M.; PINHO, A. D. S.; STAMS, G.-J.; MOONEN, X.; BOS, A. E.; AND VAN HOOREN, S., 2020. Music therapy for stress reduction: a systematic review and meta-analysis. *Health Psychology Review*, (2020), 1–26. (cited on page 2)

- DELPY, D. T.; COPE, M.; VAN DER ZEE, P.; ARRIDGE, S.; WRAY, S.; AND WYATT, J., 1988. Estimation of optical pathlength through tissue from direct time of flight measurement. *Physics in Medicine & Biology*, 33, 12 (1988), 1433. (cited on page 112)
- DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; AND FEI-FEI, L., 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255. Ieee. (cited on page 42)
- DENG, L. AND PLATT, J. C., 2014. Ensemble deep learning for speech recognition. In *Fifteenth annual conference of the international speech communication association*. (cited on page 113)
- DENG, L.; TUR, G.; HE, X.; AND HAKKANI-TUR, D., 2012. Use of kernel deep convex networks and end-to-end learning for spoken language understanding. In *2012 IEEE Spoken Language Technology Workshop (SLT)*, 210–215. IEEE. (cited on page 113)
- DHALL, A.; GOECKE, R.; LUCEY, S.; AND GEDEON, T., 2011. Acted facial expressions in the wild database. *Australian National University, Canberra, Australia, Technical Report TR-CS-11*, 2 (2011), 1. (cited on pages 30 and 48)
- DHALL, A.; SHARMA, G.; GOECKE, R.; AND GEDEON, T., 2020. Emotiw 2020: Driver gaze, group emotion, student engagement and physiological signal based challenges. In *Proceedings of the 2020 International Conference on Multimodal Interaction*, 784–789. (cited on page 3)
- DIBEKLIOĞLU, H.; SALAH, A. A.; AND GEVERS, T., 2012. Are you really smiling at me? spontaneous versus posed enjoyment smiles. In *European Conference on Computer Vision*, 525–538. Springer. (cited on pages 30 and 54)
- DINDAR, M.; MALMBERG, J.; JÄRVELÄ, S.; HAATAJA, E.; AND KIRSCHNER, P. A., 2019. Matching self-reports with electrodermal activity data: Investigating temporal changes in self-regulated learning. *Education and Information Technologies*, (2019), 1–18. (cited on page 3)
- DINO, H. I. AND ABDULRAZZAQ, M. B., 2019. Facial expression classification based on svm, knn and mlp classifiers. In *2019 International Conference on Advanced Science and Engineering (ICOASE)*, 70–75. IEEE. (cited on page 12)
- D’ALESSANDRO, P.; GIUGLIETTI, M.; BAGLIONI, A.; VERDOLINI, N.; MURGIA, N.; PICCIRILLI, M.; AND ELISEI, S., 2017. Effects of music on seizure frequency in institutionalized subjects with severe/profound intellectual disability and drug-resistant epilepsy. *Psychiatr. Danub*, 29 (2017), 399–404. (cited on pages 2 and 24)
- E4. E4 wristband from empatica. <https://www.empatica.com/research/e4/>. (cited on pages xxi, 14, and 17)

-
- EDINGER, J. A. AND PATTERSON, M. L., 1983. Nonverbal involvement and social control. *Psychological bulletin*, 93, 1 (1983), 30. (cited on page 13)
- EEROLA, T. AND VUOSKOSKI, J. K., 2011. A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, 39, 1 (2011), 18–49. (cited on page 28)
- EGERMANN, H.; FERNANDO, N.; CHUEN, L.; AND McADAMS, S., 2015. Music induces universal emotion-related psychophysiological responses: comparing canadian listeners to congolese pygmies. *Frontiers in psychology*, 5 (2015), 1341. (cited on page 3)
- EKMAN, P., 1992. An argument for basic emotions. *Cognition & emotion*, 6, 3-4 (1992), 169–200. (cited on page 9)
- EKMAN, P. AND FRIESEN, W. V., 1976. Measuring facial movement. *Environmental psychology and nonverbal behavior*, 1, 1 (1976), 56–75. (cited on page 12)
- EL AYADI, M.; KAMEL, M. S.; AND KARRAY, F., 2011. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern recognition*, 44, 3 (2011), 572–587. (cited on page 13)
- EMOTIV. Emotiv pro academic license. <https://www.emotiv.com/product/emotivpro/>. (cited on pages xxi and 18)
- EPILEPSY AND SEIZURES. Epilepsy and seizures. <http://www.columbianeurology.org/neurology/staywell/document.php?id=33912>. (cited on page 107)
- EUSTON, D. R.; GRUBER, A. J.; AND McNAUGHTON, B. L., 2012. The role of medial prefrontal cortex in memory and decision making. *Neuron*, 76, 6 (2012), 1057–1070. (cited on page 122)
- EYETRIBE. The eye tribe. <http://theeyetribe.com/theeyetribe.com/about/index.html/>. (cited on page 20)
- FACELAB. Facelab. <http://www.seeingmachines.com/product/facelab/>. (cited on page 20)
- FENG, F.; ZHANG, Y.; HOU, J.; CAI, J.; JIANG, Q.; LI, X.; ZHAO, Q.; AND LI, B.-A., 2018a. Can music improve sleep quality in adults with primary insomnia? a systematic review and network meta-analysis. *International journal of nursing studies*, 77 (2018), 189–196. (cited on page 2)
- FENG, H.; GOLSHAN, H. M.; AND MAHOOR, M. H., 2018b. A wavelet-based approach to emotion classification using eda signals. *Expert Systems with Applications*, 112 (2018), 77–86. (cited on page 14)
- GAMMA, 2016. Gamma brain energizer - 40 hz - clean mental energy - focus music - binaural beats. <https://www.youtube.com/watch?v=9wrFk5vuOsk>. (cited on pages 29 and 30)

- GAMMA AND MEMORY. Gamma wave - an overview. <https://www.sciencedirect.com/topics/neuroscience/gamma-wave>. (cited on page 107)
- GANAPATHY, N.; VEERANKI, Y. R.; KUMAR, H.; AND SWAMINATHAN, R., 2021. Emotion recognition using electrodermal activity signals and multiscale deep convolutional neural network. *Journal of Medical Systems*, 45, 4 (2021), 1–10. (cited on page 14)
- GATTI, E.; CALZOLARI, E.; MAGGIONI, E.; AND OBRIST, M., 2018. Emotional ratings and skin conductance response to visual, auditory and haptic stimuli. *Scientific data*, 5, 1 (2018), 1–12. (cited on page 53)
- GINNS, P. AND KYDD, A., 2019. Learning human physiology by pointing and tracing: A cognitive load approach. In *Advances in Cognitive Load Theory*, 119–129. Routledge. (cited on page 13)
- GLASER, B. G. AND STRAUSS, A. L., 2017. *Discovery of grounded theory: Strategies for qualitative research*. Routledge. (cited on pages 64, 103, and 123)
- GOOTY, J.; GAVIN, M.; AND ASHKANASY, N. M., 2009. Emotions research in ob: The challenges that lie ahead. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 30, 6 (2009), 833–838. (cited on page 12)
- GRECO, A.; MARZI, C.; LANATA, A.; SCILINGO, E. P.; AND VANELLO, N., 2019. Combining electrodermal activity and speech analysis towards a more accurate emotion recognition system. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 229–232. IEEE. (cited on page 13)
- GREKOW, J., 2020. Static music emotion recognition using recurrent neural networks. In *International Symposium on Methodologies for Intelligent Systems*, 150–160. Springer. (cited on page 41)
- GROSS, J. J. AND JOHN, O. P., 2003. Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *Journal of personality and social psychology*, 85, 2 (2003), 348. (cited on page 12)
- GUHN, M.; HAMM, A.; AND ZENTNER, M., 2007. Physiological and musico-acoustic correlates of the chill response. *Music Perception*, 24, 5 (2007), 473–484. (cited on page 4)
- HAND, D. AND CHRISTEN, P., 2018. A note on using the f-measure for evaluating record linkage algorithms. *Statistics and Computing*, 28, 3 (2018), 539–547. (cited on page 52)
- HANDOUZI, W.; MAAOUI, C.; PRUSKI, A.; AND MOUSSAOUI, A., 2014. Objective model assessment for short-term anxiety recognition from blood volume pulse signal. *Biomedical Signal Processing and Control*, 14 (2014), 217–227. (cited on page 15)

-
- HARMON-JONES, C.; BASTIAN, B.; AND HARMON-JONES, E., 2016. The discrete emotions questionnaire: A new tool for measuring state self-reported emotions. *PloS one*, 11, 8 (2016), e0159915. (cited on page 12)
- HARRISON, L. AND LOUI, P., 2014. Thrills, chills, frissons, and skin orgasms: toward an integrative model of transcendent psychophysiological experiences in music. *Frontiers in Psychology*, 5 (2014), 790. (cited on pages 4 and 29)
- HE, K.; ZHANG, X.; REN, S.; AND SUN, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778. (cited on pages 42 and 82)
- HEALEY, J. A. AND PICARD, R. W., 2005. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on intelligent transportation systems*, 6, 2 (2005), 156–166. (cited on page 21)
- HENSON, R.; SHALLICE, T.; AND DOLAN, R. J., 1999. Right prefrontal cortex and episodic memory retrieval: a functional mri test of the monitoring hypothesis. *Brain*, 122, 7 (1999), 1367–1381. (cited on page 122)
- HO, T. K. K.; GWAK, J.; PARK, C. M.; AND SONG, J.-I., 2019. Discrimination of mental workload levels from multi-channel fnirs using deep leaning-based approaches. *IEEE Access*, 7 (2019), 24392–24403. (cited on page 41)
- HOCHREITER, S. AND SCHMIDHUBER, J., 1997. Long short-term memory. *Neural computation*, 9, 8 (1997), 1735–1780. (cited on page 41)
- HOFFMANN, H.; KESSLER, H.; EPEL, T.; RUKAVINA, S.; AND TRAUER, H. C., 2010. Expression intensity, gender and facial emotion recognition: Women recognize only subtle facial emotions better than men. *Acta psychologica*, 135, 3 (2010), 278–283. (cited on page 12)
- HOSSAIN, M.; GEDEON, T.; SANKARANARAYANA, R.; APHORP, D.; AND DAWEL, A., 2016. Pupillary responses of asian observers in discriminating real from fake smiles: A preliminary study. In *Measuring Behavior*, 170–176. (cited on pages 59 and 60)
- HOSSAIN, M. Z.; GEDEON, T.; AND SANKARANARAYANA, R., 2018. Using temporal features of observers' physiological measures to distinguish between genuine and fake smiles. *IEEE Transactions on Affective Computing*, 11, 1 (2018), 163–173. (cited on page 138)
- HOU, X.; LIU, Y.; SOURINA, O.; TAN, Y. R. E.; WANG, L.; AND MUELLER-WITTIG, W., 2015. Eeg based stress monitoring. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, 3110–3115. IEEE. (cited on page 17)
- HU, X.; LI, F.; AND NG, T.-D. J., 2018. On the relationships between music-induced emotion and physiological signals. In *ISMIR*, 362–369. (cited on page 2)

- HUANG, C.-Y.; CHANG, E.-T.; HSIEH, Y.-M.; AND LAI, H.-L., 2017a. Effects of music and music video interventions on sleep quality: A randomized controlled trial in adults with sleep disturbances. *Complementary therapies in medicine*, 34 (2017), 116–122. (cited on pages 22 and 106)
- HUANG, K.-Y.; WU, C.-H.; HONG, Q.-B.; SU, M.-H.; AND CHEN, Y.-H., 2019. Speech emotion recognition using deep neural network considering verbal and nonverbal speech sounds. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5866–5870. IEEE. (cited on page 3)
- HUANG, R.-H. AND SHIH, Y.-N., 2011. Effects of background music on concentration of workers. *Work*, 38, 4 (2011), 383–387. (cited on page 2)
- HUANG, Y.; YANG, J.; LIAO, P.; AND PAN, J., 2017b. Fusion of facial expressions and eeg for multimodal emotion recognition. *Computational intelligence and neuroscience*, 2017 (2017). (cited on page 13)
- HUGHES, J. R., 2008a. Gamma, fast, and ultrafast waves of the brain: their relationships with epilepsy and behavior. *Epilepsy & Behavior*, 13, 1 (2008), 25–31. (cited on page 17)
- HUGHES, J. R., 2008b. Gamma, fast, and ultrafast waves of the brain: Their relationships with epilepsy and behavior. *Epilepsy Behavior*, 13, 1 (2008), 25 – 31. doi:<https://doi.org/10.1016/j.yebeh.2008.01.011>. <http://www.sciencedirect.com/science/article/pii/S1525505008000127>. (cited on page 23)
- HUGHES, J. R. AND FINO, J. J., 2000. The mozart effect: distinctive aspects of the music—a clue to brain coding? *Clinical Electroencephalography*, 31, 2 (2000), 94–103. (cited on pages 29 and 30)
- HURLESS, N.; MEKIC, A.; PENA, S.; HUMPHRIES, E.; GENTRY, H.; AND NICHOLS, D. Music genre preference and tempo alter alpha and beta waves in human non-musicians. (cited on page 30)
- HURON, D. AND MARGULIS, E. H., 2010. Musical expectancy and thrills. (2010). (cited on page 4)
- IANDOLA, F. N.; HAN, S.; MOSKEWICZ, M. W.; ASHRAF, K.; DALLY, W. J.; AND KEUTZER, K., 2016. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, (2016). (cited on page 42)
- INNES, K. E.; SELFE, T. K.; KHALSA, D. S.; AND KANDATI, S., 2017. Meditation and music improve memory and cognitive function in adults with subjective cognitive decline: a pilot randomized controlled trial. *Journal of Alzheimer's disease*, 56, 3 (2017), 899–916. (cited on page 2)

-
- ISLAM, M. R.; PAVEL, M. S. R.; AND TUNAZ, S. A., 2019. Neurodegenerative disease classification using gait signal features and random forest classifier. In *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*, 1–4. doi:10.1109/EICT48899.2019.9068822. (cited on page 39)
- JERRITTA, S.; MURUGAPPAN, M.; NAGARAJAN, R.; AND WAN, K., 2011. Physiological signals based human emotion recognition: a review. In *2011 IEEE 7th international colloquium on signal processing and its applications*, 410–415. IEEE. (cited on pages 48 and 55)
- JERRITTA, S.; MURUGAPPAN, M.; WAN, K.; AND YAACOB, S., 2013. Emotion detection from qrs complex of ecg signals using hurst exponent for different age groups. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, 849–854. IEEE. (cited on page 92)
- JIANG, C.; LI, Y.; TANG, Y.; AND GUAN, C., 2021. Enhancing eeg-based classification of depression patients using spatial information. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29 (2021), 566–575. (cited on page 114)
- JUSLIN, P. N. AND LAUKKA, P., 2003. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological bulletin*, 129, 5 (2003), 770. (cited on page 13)
- JUSLIN, P. N. AND SLOBODA, J. A., 2001. *Music and emotion: Theory and research*. Oxford University Press. (cited on page 2)
- KAPUR, A.; KAPUR, A.; VIRJI-BABUL, N.; TZANETAKIS, G.; AND DRIESSEN, P. F., 2005. Gesture-based affective computing on motion capture data. In *International conference on affective computing and intelligent interaction*, 1–7. Springer. (cited on page 13)
- KATSIS, C. D.; KATERTSIDIS, N.; GANIATSAS, G.; AND FOTIADIS, D. I., 2008. Toward emotion recognition in car-racing drivers: A biosignal processing approach. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 38, 3 (2008), 502–512. (cited on pages 21 and 36)
- KAWAKAMI, A.; FURUKAWA, K.; KATAHIRA, K.; AND OKANOYA, K., 2013. Sad music induces pleasant emotion. *Frontiers in psychology*, 4 (2013), 311. (cited on page 123)
- KESSOUS, L.; CASTELLANO, G.; AND CARIDAKIS, G., 2010. Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis. *Journal on Multimodal User Interfaces*, 3, 1 (2010), 33–48. (cited on page 13)
- KHALFA, S.; ISABELLE, P.; JEAN-PIERRE, B.; AND MANON, R., 2002. Event-related skin conductance responses to musical emotions in humans. *Neuroscience letters*, 328, 2 (2002), 145–149. (cited on page 2)

- KHAN, A. M. AND LAWU, M., 2016. Recognizing emotion from blood volume pulse and skin conductance sensor using machine learning algorithms. In *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016*, 1297–1303. Springer. (cited on page 15)
- KHUSHABA, R. N.; GREENACRE, L.; KODAGODA, S.; LOUVIERE, J.; BURKE, S.; AND DISANAYAKE, G., 2012. Choice modeling and the brain: A study on the electroencephalogram (eeg) of preferences. *Expert Systems with Applications*, 39, 16 (2012), 12378–12388. (cited on page 98)
- KIM, J. AND ANDRÉ, E., 2008. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30, 12 (Dec 2008), 2067–2083. doi:10.1109/TPAMI.2008.26. (cited on pages 2, 10, 14, 15, 61, and 73)
- KIM, K. H.; BANG, S. W.; AND KIM, S. R., 2004. Emotion recognition system using short-term monitoring of physiological signals. *Medical and biological engineering and computing*, 42, 3 (2004), 419–427. (cited on page 21)
- KOELSTRA, S.; MUHL, C.; SOLEYMANI, M.; LEE, J.-S.; YAZDANI, A.; EBRAHIMI, T.; PUN, T.; NIJHOLT, A.; AND PATRAS, I., 2011. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3, 1 (2011), 18–31. (cited on page 27)
- KOHAVI, R. AND JOHN, G. H., 1997. Wrappers for feature subset selection. *Artificial intelligence*, 97, 1-2 (1997), 273–324. (cited on page 37)
- KRIZHEVSKY, A.; SUTSKEVER, I.; AND HINTON, G. E., 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25 (2012), 1097–1105. (cited on page 42)
- KRUMHANSL, C. L., 1997. An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 51, 4 (1997), 336. (cited on page 2)
- KULIC, D. AND CROFT, E. A., 2007. Affective state estimation for human–robot interaction. *IEEE Transactions on Robotics*, 23, 5 (2007), 991–1000. (cited on page 16)
- LAN, T.; FANG, Y.; XIONG, W.; AND KONG, C., 2007. Automatic baseline correction of infrared spectra. *Chinese Optics Letters*, 5, 10 (2007), 613–616. (cited on page 35)
- LARSEN, R. S. AND WATERS, J., 2018. Neuromodulatory correlates of pupil dilation. *Frontiers in neural circuits*, 12 (2018), 21. (cited on page 20)
- LATIF, S.; RANA, R.; KHALIFA, S.; JURDAK, R.; EPPS, J.; AND SCHULLER, B. W., 2020. Multi-task semi-supervised adversarial autoencoding for speech emotion recognition. *IEEE Transactions on Affective Computing*, (2020). (cited on page 13)

-
- LECUN, Y.; BOTTOU, L.; BENGIO, Y.; AND HAFFNER, P., 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86, 11 (1998), 2278–2324. (cited on page 82)
- LEE, C.; YOO, S.; PARK, Y.; KIM, N.; JEONG, K.; AND LEE, B., 2006. Using neural network to recognize human emotions from heart rate variability and skin resistance. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, 5523–5525. IEEE. (cited on page 16)
- LEE, K. S.; JEONG, H. C.; YIM, J. E.; AND JEON, M. Y., 2016. Effects of music therapy on the cardiovascular and autonomic nervous system in stress-induced university students: a randomized controlled trial. *The Journal of Alternative and Complementary Medicine*, 22, 1 (2016), 59–65. (cited on page 22)
- LEVENBERG, K., 1944. A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2, 2 (1944), 164–168. (cited on pages 80 and 97)
- LI, F. AND XIONG, Y., 2016. Application of music therapy combined with computer biofeedback in the treatment of anxiety disorders. In *Information Technology in Medicine and Education (ITME), 2016 8th International Conference on*, 90–93. IEEE. (cited on page 22)
- LIAO, C.-Y.; CHEN, R.-C.; AND TAI, S.-K., 2018. Emotion stress detection using eeg signal and deep learning technologies. In *2018 IEEE International Conference on Applied System Invention (ICASI)*, 90–93. IEEE. (cited on page 3)
- LIAO, W.; ZHANG, W.; ZHU, Z.; AND JI, Q., 2005. A real-time human stress monitoring system using dynamic bayesian network. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, 70–70. doi:10.1109/CVPR.2005.394. (cited on page 39)
- LIEBER, C. A. AND MAHADEVAN-JANSEN, A., 2003. Automated method for subtraction of fluorescence from biological raman spectra. *Applied spectroscopy*, 57, 11 (2003), 1363–1367. (cited on page 35)
- LIGHTNIRS. Lightnirs | shimadzu europa - shimadzu europe. <https://www.shimadzu.eu/lightnirs>. (cited on page 19)
- LIN, L.-C.; CHIANG, C.-T.; LEE, M.-W.; MOK, H.-K.; YANG, Y.-H.; WU, H.-C.; TSAI, C.-L.; AND YANG, R.-C., 2013. Parasympathetic activation is involved in reducing epileptiform discharges when listening to mozart music. *Clinical Neurophysiology*, 124, 8 (2013), 1528–1535. (cited on page 30)
- LIN, T.; OMATA, M.; HU, W.; AND IMAMIYA, A., 2005. Do physiological data relate to traditional usability indexes? In *Proceedings of the 17th Australia conference on computer-human interaction: Citizens online: Considerations for today and the future*, 1–10. Citeseer. (cited on page 14)

- LIN, Y.-P.; WANG, C.-H.; JUNG, T.-P.; WU, T.-L.; JENG, S.-K.; DUANN, J.-R.; AND CHEN, J.-H., 2010. Eeg-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering*, 57, 7 (2010), 1798–1806. (cited on page 17)
- LIU, H.; FANG, Y.; AND HUANG, Q., 2019. Music emotion recognition using a variant of recurrent neural network. In *2018 International Conference on Mathematics, Modeling, Simulation and Statistics Application (MMSSA 2018)*. Atlantis Press. (cited on page 41)
- LIU, J.; SU, Y.; AND LIU, Y., 2017. Multi-modal emotion recognition with temporal-band attention based on lstm-rnn. In *Pacific Rim Conference on Multimedia*, 194–204. Springer. (cited on page 42)
- LIU, Z.-T.; WU, M.; CAO, W.-H.; MAO, J.-W.; XU, J.-P.; AND TAN, G.-Z., 2018. Speech emotion recognition based on feature selection and extreme learning machine decision tree. *Neurocomputing*, 273 (2018), 271–280. doi:<https://doi.org/10.1016/j.neucom.2017.07.050>. <https://www.sciencedirect.com/science/article/pii/S0925231217313565>. (cited on page 38)
- LOOXIDLINK. Looxid link – connect your mind to vr. <https://looxidlink.looxidlabs.com/>. (cited on page 140)
- MA, J.; TANG, H.; ZHENG, W.-L.; AND LU, B.-L., 2019. Emotion recognition using multimodal residual lstm network. In *Proceedings of the 27th ACM international conference on multimedia*, 176–183. (cited on page 42)
- MAGUIRE, M., 2015. Music and its association with epileptic disorders. In *Progress in brain research*, vol. 217, 107–127. Elsevier. (cited on page 106)
- MALIK, M.; ADAVANNE, S.; DROSSOS, K.; VIRTANEN, T.; TICHA, D.; AND JARINA, R., 2017. Stacked convolutional and recurrent neural networks for music emotion recognition. *arXiv preprint arXiv:1706.02292*, (2017). (cited on page 113)
- MAN, K.-F.; TANG, K.-S.; AND KWONG, S., 1996. Genetic algorithms: concepts and applications [in engineering design]. *IEEE transactions on Industrial Electronics*, 43, 5 (1996), 519–534. (cited on page 37)
- MANELIS, A.; HUPPERT, T. J.; RODGERS, E.; SWARTZ, H. A.; AND PHILLIPS, M. L., 2019. The role of the right prefrontal cortex in recognition of facial emotional expressions in depressed individuals: fnirs study. *Journal of affective disorders*, 258 (2019), 151–158. (cited on page 19)
- MARQUARDT, D. W., 1963. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11, 2 (1963), 431–441. (cited on pages 80 and 97)
- MCCRATY, R.; BARRIOS-CHOPLIN, B.; ATKINSON, M.; AND TOMASINO, D., 1998. The effects of different types of music on mood, tension, and mental clarity. *Alternative therapies in health and medicine*, 4, 1 (1998), 75–84. (cited on pages 2 and 29)

-
- MCDUFF, D.; KARLSON, A.; KAPOOR, A.; ROSEWAY, A.; AND CZERWINSKI, M., 2012. Affectaura: an intelligent system for emotional memory. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 849–858. (cited on page 38)
- McFARLAND, R. A., 1985. Relationship of skin temperature changes to the emotions accompanying music. *Biofeedback and Self-regulation*, 10, 3 (1985), 255–267. (cited on page 16)
- MEHMOOD, R. M. AND LEE, H. J., 2015. Emotion classification of eeg brain signal using svm and knn. In *2015 IEEE international conference on multimedia & expo workshops (ICMEW)*, 1–5. IEEE. (cited on page 38)
- MIDHA, S.; MAIOR, H. A.; WILSON, M. L.; AND SHARPLES, S., 2021. Measuring mental workload variations in office work tasks using fnirs. *International Journal of Human-Computer Studies*, 147 (2021), 102580. (cited on page 19)
- MIGNAULT, A. AND CHAUDHURI, A., 2003. The many faces of a neutral face: Head tilt and perception of dominance and emotion. *Journal of nonverbal behavior*, 27, 2 (2003), 111–132. (cited on page 13)
- MILLER, F.; STIKSEL, M.; AND JONES, R., 2008. Last. fm in numbers. *Last. fm press material*, (2008). (cited on page 23)
- MITRA, S. AND ACHARYA, T., 2007. Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37, 3 (2007), 311–324. (cited on page 13)
- MOGHIMI, S.; KUSHKI, A.; GUERGUERIAN, A. M.; AND CHAU, T., 2012. Characterizing emotional response to music in the prefrontal cortex using near infrared spectroscopy. *Neuroscience Letters*, 525, 1 (2012), 7–11. (cited on page 126)
- MOORE, N. C., 2000. A review of eeg biofeedback treatment of anxiety disorders. *Clinical electroencephalography*, 31, 1 (2000), 1–6. (cited on page 17)
- MORITA, T.; ITAKURA, S.; SAITO, D. N.; NAKASHITA, S.; HARADA, T.; KOCHIYAMA, T.; AND SADATO, N., 2008. The role of the right prefrontal cortex in self-evaluation of the face: a functional magnetic resonance imaging study. *Journal of cognitive neuroscience*, 20, 2 (2008), 342–355. (cited on page 122)
- MORNINGSTAR, M.; DIRKS, M. A.; AND HUANG, S., 2017. Vocal cues underlying youth and adult portrayals of socio-emotional expressions. *Journal of Nonverbal Behavior*, 41, 2 (2017), 155–183. (cited on page 13)
- MUSE. Muse -technology enhanced education. <https://choosemuse.com/muse/>. (cited on page 18)
- NAGAI, Y.; GOLDSTEIN, L. H.; FENWICK, P. B.; AND TRIMBLE, M. R., 2004. Clinical efficacy of galvanic skin response biofeedback training in reducing seizures in adult epilepsy: a preliminary randomized controlled study. *Epilepsy & Behavior*, 5, 2 (2004), 216–223. (cited on page 24)

- NAKISA, B.; RASTGOO, M. N.; RAKOTONIRAINY, A.; MAIRE, F.; AND CHANDRAN, V., 2020. Automatic emotion recognition using temporal multimodal deep learning. *IEEE Access*, 8 (2020), 225463–225474. doi:10.1109/ACCESS.2020.3027026. (cited on page 15)
- NEUROSKY. Neurosky eeg biosensors. <http://neurosky.com/biosensors/eeg-sensor/>. (cited on page 18)
- NEWSIDENTIST. New scientist | science news and science articles from new scientist. <https://www.newscientist.com/>. (cited on page 73)
- NEXTMIND. Nextmind - let your mind take control. <https://www.next-mind.com/>. (cited on page 140)
- NEXUS BVP. Nexus blood volume pulse sensor (bvp). <https://www.biofeedback-tech.com/biofeedback-shop/nexus-blood-volume-pulse-sensor-bvp/>. (cited on pages xxi and 15)
- NIE, D.; WANG, X.-W.; SHI, L.-C.; AND LU, B.-L., 2011. Eeg-based emotion recognition during watching movies. In *2011 5th International IEEE/EMBS Conference on Neural Engineering*, 667–670. IEEE. (cited on page 17)
- NIRSIT. Obelab - fnirs devices. <https://www.obelab.com/>. (cited on page 19)
- NOROOZI, F.; KAMINSKA, D.; CORNEANU, C.; SAPINSKI, T.; ESCALERA, S.; AND ANBARJAFARI, G., 2018. Survey on emotional body gesture recognition. *IEEE transactions on affective computing*, (2018). (cited on pages 3 and 13)
- NUL-208. Heart rate pulse logger sensor nul-208. <https://neulog.com/heart-rate-pulse/>. (cited on pages xxi and 16)
- ODED, Y., 2018. Integrating mindfulness and biofeedback in the treatment of post-traumatic stress disorder. *Biofeedback*, 46, 2 (2018), 37–47. (cited on page 15)
- OEG16. Oeg-16 product / spectratech. <https://www.spectratech.co.jp/En/product/productOeg16En.html>. (cited on page 19)
- OGUNYEMI, A. AND BREEN, H., 1993. Seizures induced by music. *Behavioural neurology*, 6, 4 (1993), 215–219. (cited on page 22)
- OH, S.; LEE, J.-Y.; AND KIM, D. K., 2020. The design of cnn architectures for optimal six basic emotion classification using multiple physiological signals. *Sensors*, 20, 3 (2020), 866. (cited on page 41)
- PALANGI, H.; DENG, L.; AND WARD, R. K., 2014. Recurrent deep-stacking networks for sequence classification. In *2014 IEEE China Summit & International Conference on Signal and Information Processing (ChinaSIP)*, 510–514. IEEE. (cited on page 113)

-
- PALANISAMY, K.; MURUGAPPAN, M.; AND YAACOB, S., 2013. Multiple physiological signal-based human stress identification using non-linear classifiers. *Elektronika ir elektrotechnika*, 19, 7 (2013), 80–85. (cited on page 38)
- PANTIC, M.; VALSTAR, M.; RADEMAKER, R.; AND MAAT, L., 2005. Web-based database for facial expression analysis. In *2005 IEEE international conference on multimedia and Expo*, 5–pp. IEEE. (cited on page 31)
- PARTALA, T. AND SURAKKA, V., 2003. Pupil size variation as an indication of affective processing. *International journal of human-computer studies*, 59, 1-2 (2003), 185–198. (cited on page 20)
- PASTELAK-PRICE, C., 1983. The international 10-20-system of electrode placement: Its rationale and a practical guide to measuring procedures and electrode placement. (cited on pages 18 and 96)
- PATHAN, N. S.; FOYSAL, M.; AND ALAM, M. M., 2019. Efficient mental arithmetic task classification using wavelet domain statistical features and svm classifier. In *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 1–5. IEEE. (cited on page 118)
- PECK, E. M. M.; YUKSEL, B. F.; OTTLEY, A.; JACOB, R. J.; AND CHANG, R., 2013. Using fnirs brain sensing to evaluate information visualization interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 473–482. (cited on page 110)
- PENG, H.; LONG, F.; AND DING, C., 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27, 8 (2005), 1226–1238. (cited on page 37)
- PETRANTONAKIS, P. C. AND HADJILEONTIADIS, L. J., 2010. Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis. *IEEE Transactions on affective computing*, 1, 2 (2010), 81–97. (cited on page 3)
- PETRIDES, K. V., 2009. Psychometric properties of the trait emotional intelligence questionnaire (teique). In *Assessing emotional intelligence*, 85–101. Springer. (cited on page 12)
- PICARD, R. W., 2000. *Affective computing*. MIT press. (cited on page 3)
- PICARD, R. W.; VYZAS, E.; AND HEALEY, J., 2001. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE transactions on pattern analysis and machine intelligence*, 23, 10 (2001), 1175–1191. (cited on pages 24 and 36)
- PINTI, P.; AICHELBURG, C.; GILBERT, S.; HAMILTON, A.; HIRSCH, J.; BURGESS, P.; AND TACHTSIDIS, I., 2018. A review on the use of wearable functional near-infrared spectroscopy in naturalistic environments. *Japanese Psychological Research*, 60, 4 (2018), 347–373. (cited on page 19)

- PINTO, G.; CARVALHO, J. M.; BARROS, F.; SOARES, S. C.; PINHO, A. J.; AND BRÁS, S., 2020. Multimodal emotion evaluation: A physiological model for cost-effective emotion classification. *Sensors*, 20, 12 (2020), 3510. (cited on page 40)
- PITTAU, F.; TINUPER, P.; BISULLI, F.; NALDI, I.; CORTELLI, P.; BISULLI, A.; STIPA, C.; CEVOLANI, D.; AGATI, R.; LEONARDI, M.; ET AL., 2008. Videopolygraphic and functional mri study of musicogenic epilepsy. a case report and literature review. *Epilepsy & Behavior*, 13, 4 (2008), 685–692. (cited on page 22)
- PLUTCHIK, R., 2001. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, 89, 4 (2001), 344–350. (cited on page 9)
- POHJALAINEN, J.; RÄSÄNEN, O.; AND KADIOGLU, S., 2015. Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits. *Computer Speech & Language*, 29, 1 (2015), 145–171. (cited on page 37)
- PUDIL, P.; NOVOTIČOVÁ, J.; AND KITTLER, J., 1994. Floating search methods in feature selection. *Pattern recognition letters*, 15, 11 (1994), 1119–1125. (cited on page 38)
- PUPILLABS. Pupil labs. <https://pupil-labs.com/>. (cited on page 20)
- QUINTANA, D. S.; GUASTELLA, A. J.; OUTHRED, T.; HICKIE, I. B.; AND KEMP, A. H., 2012. Heart rate variability is associated with emotion recognition: direct evidence for a relationship between the autonomic nervous system and social cognition. *International journal of psychophysiology*, 86, 2 (2012), 168–172. (cited on page 16)
- QUIROZ, J. C.; YONG, M. H.; AND GEANGU, E., 2017. Emotion-recognition using smart watch accelerometer data: Preliminary findings. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, 805–812. (cited on page 21)
- RAHMAN, J. S.; GEDEON, T.; CALDWELL, S.; AND JONES, R., 2020. Brain melody informatics: Analysing effects of music on brainwave patterns. In *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–8. doi:10.1109/IJCNN48605.2020.9207392. (cited on page 66)
- RAHMAN, J. S.; GEDEON, T.; CALDWELL, S.; AND JONES, R., 2020a. Brain melody informatics: Analysing effects of music on brainwave patterns. In *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–8. IEEE. (cited on page 95)
- RAHMAN, J. S.; GEDEON, T.; CALDWELL, S.; JONES, R.; HOSSAIN, M. Z.; AND ZHU, X., 2019. Melodious micro-frissons: Detecting music genres from skin response. In *2019 International Joint Conference on Neural Networks (IJCNN)*, 1–8. (cited on pages 4 and 71)

-
- RAHMAN, J. S.; GEDEON, T.; CALDWELL, S.; JONES, R.; AND JIN, Z., 2021a. Towards effective music therapy for mental health care using machine learning tools: Human affective reasoning and music genres. *Journal of Artificial Intelligence and Soft Computing Research*, 11, 1 (2021), 5–20. (cited on page 71)
- RAHMAN, J. S.; GEDEON, T.; CALDWELL, S.; AND JONES, R. L., 2021b. Can binaural beats increase your focus? exploring the effects of music in participants' conscious and brain activity responses. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–6. (cited on page 47)
- RAHMAN, J. S.; HOSSAIN, M. Z.; AND GEDEON, T., 2019. Measuring observers' eda responses to emotional videos. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*, 457–461. (cited on page 47)
- RAHMAN, J. S.; HOSSAIN, M. Z.; AND GEDEON, T., 2020b. Are paired or single stimuli better to recognize genuine and posed smiles from observers' galvanic skin response? In *32nd Australian Conference on Human-Computer Interaction*, 661–665. (cited on page 47)
- RAJESH, S. AND NALINI, N., 2020. Musical instrument emotion recognition using deep recurrent neural network. *Procedia Computer Science*, 167 (2020), 16–25. (cited on page 41)
- RAMNANI, N. AND OWEN, A. M., 2004. Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nature reviews neuroscience*, 5, 3 (2004), 184–194. (cited on page 19)
- RÄSÄNEN, O. AND POHJALAINEN, J., 2013. Random subset feature selection in automatic recognition of developmental disorders, affective states, and level of conflict from speech. In *Interspeech*, 210–214. (cited on page 38)
- REISMAN, S., 1997. Measurement of physiological stress. *Proceedings of the IEEE 23rd Northeast Bioengineering Conference*, (1997), 21–23. (cited on page 15)
- RIGAS, G.; KATSIAS, C. D.; GANIATSAS, G.; AND FOTIADIS, D. I., 2007. A user independent, biosignal based, emotion recognition method. In *International Conference on User Modeling*, 314–318. Springer. (cited on page 21)
- RIM, B.; SUNG, N.-J.; MIN, S.; AND HONG, M., 2020. Deep learning in physiological signal data: A survey. *Sensors*, 20, 4 (2020). doi:10.3390/s20040969. <https://www.mdpi.com/1424-8220/20/4/969>. (cited on page 41)
- RUSSELL, J. A., 1980. A circumplex model of affect. *Journal of personality and social psychology*, 39, 6 (1980), 1161. (cited on pages 10 and 50)
- SALMAM, F. Z.; MADANI, A.; AND KISSI, M., 2016. Facial expression recognition using decision trees. In *2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV)*, 125–130. doi:10.1109/CGiV.2016.33. (cited on page 38)

- SALZMAN, C. D. AND FUSI, S., 2010. Emotion, cognition, and mental state representation in amygdala and prefrontal cortex. *Annual review of neuroscience*, 33 (2010), 173–202. (cited on pages 66 and 98)
- SAMARA, A.; MENEZES, M. L. R.; AND GALWAY, L., 2016. Feature extraction for emotion recognition and modelling using neurophysiological data. In *2016 15th international conference on ubiquitous computing and communications and 2016 international symposium on cyberspace and security (IUCC-CSS)*, 138–144. IEEE. (cited on page 36)
- SÁNCHEZ-GONZÁLEZ, B.; BARJA, I.; PIÑEIRO, A.; HERNÁNDEZ-GONZÁLEZ, M. C.; SILVÁN, G.; ILLERA, J. C.; AND LATORRE, R., 2018. Support vector machines for explaining physiological stress response in wood mice (*apodemus sylvaticus*). *Scientific reports*, 8, 1 (2018), 1–14. (cited on page 39)
- SAUTER, D. A.; EISNER, F.; CALDER, A. J.; AND SCOTT, S. K., 2010. Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology*, 63, 11 (2010), 2251–2272. (cited on page 12)
- SCHERER, S.; HOFMANN, H.; LAMPMANN, M.; PFEIL, M.; RHINOW, S.; SCHWENKER, F.; AND PALM, G., 2008. Emotion recognition from speech: Stress experiment. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. European Language Resources Association (ELRA), Marrakech, Morocco. http://www.lrec-conf.org/proceedings/lrec2008/pdf/336_paper.pdf. (cited on page 40)
- SCHERER, S.; PESTIAN, J.; AND MORENCY, L.-P., 2013. Investigating the speech characteristics of suicidal adolescents. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 709–713. IEEE. (cited on page 13)
- SCHULLER, B.; RIGOLL, G.; AND LANG, M., 2004. Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture. In *2004 IEEE international conference on acoustics, speech, and signal processing*, vol. 1, I–577. IEEE. (cited on page 13)
- SHAN, K.; GUO, J.; YOU, W.; LU, D.; AND BIE, R., 2017. Automatic facial expression recognition based on a deep convolutional-neural-network structure. In *2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*, 123–128. IEEE. (cited on page 3)
- SHAO, L. AND GRIFFITHS, P. R., 2007. Automatic baseline correction by wavelet transform for quantitative open-path fourier transform infrared spectroscopy. *Environmental science & technology*, 41, 20 (2007), 7054–7059. (cited on page 35)
- SHARMA, N. AND GEDEON, T., 2013. Computational models of stress in reading using physiological and physical sensor data. In *Advances in Knowledge Discovery and Data Mining*, 111–122. Springer Berlin Heidelberg, Berlin, Heidelberg. (cited on page 24)

-
- SHARMA, N. AND GEDEON, T., 2014. Modeling observer stress for typical real environments. *Expert Systems with Applications*, 41, 5 (2014), 2231–2238. (cited on page 39)
- SHEYKHIVAND, S.; MOUSAVI, Z.; REZAI, T. Y.; AND FARZAMNIA, A., 2020. Recognizing emotions evoked by music using cnn-lstm networks on eeg signals. *IEEE Access*, 8 (2020), 139332–139345. doi:10.1109/ACCESS.2020.3011882. (cited on page 41)
- SHI, Y.; RUIZ, N.; TAIB, R.; CHOI, E.; AND CHEN, F., 2007. Galvanic skin response (gsr) as an index of cognitive load. In *CHI'07 extended abstracts on Human factors in computing systems*, 2651–2656. (cited on page 14)
- SHIN, J.; KWON, J.; CHOI, J.; AND IM, C.-H., 2017. Performance enhancement of a brain-computer interface using high-density multi-distance nirs. *Scientific reports*, 7, 1 (2017), 1–10. (cited on page 112)
- SHOUSE, E., 2005. Feeling, emotion, affect. *M/c journal*, 8, 6 (2005). (cited on page 3)
- SHUKLA, J.; BARREDA-ANGELES, M.; OLIVER, J.; NANDI, G.; AND PUIG, D., 2019. Feature extraction and selection for emotion recognition from electrodermal activity. *IEEE Transactions on Affective Computing*, (2019). (cited on page 14)
- SHUKLA, S. AND CHAURASIYA, R. K., 2018. Emotion analysis through eeg and peripheral physiological signals using knn classifier. In *International Conference on ISMAC in Computational Vision and Bio-Engineering*, 97–106. Springer. (cited on page 38)
- SIMONYAN, K. AND ZISSERMAN, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, (2014). (cited on page 42)
- SINGH, R. R.; CONJETI, S.; AND BANERJEE, R., 2013. A comparative evaluation of neural network classifiers for stress level analysis of automotive drivers using physiological signals. *Biomedical Signal Processing and Control*, 8, 6 (2013), 740–754. (cited on page 40)
- SITTÓN-CANDANEDO, I.; ALONSO, R. S.; CORCHADO, J. M.; RODRÍGUEZ-GONZÁLEZ, S.; AND CASADO-VARA, R., 2019. A review of edge computing reference architectures and a new global edge proposal. *Future Generation Computer Systems*, 99 (2019), 278–294. (cited on page 140)
- SMITH, R.; LANE, R. D.; ALKOZEI, A.; BAO, J.; SMITH, C.; SANOVA, A.; NETTLES, M.; AND KILLGORE, W. D., 2018. The role of medial prefrontal cortex in the working memory maintenance of one's own emotional responses. *Scientific reports*, 8, 1 (2018), 1–15. (cited on page 122)
- SOLEYMANI, M.; LICHTENAUER, J.; PUN, T.; AND PANTIC, M., 2011a. A multimodal database for affect recognition and implicit tagging. *IEEE transactions on affective computing*, 3, 1 (2011), 42–55. (cited on page 31)

- SOLEYMANI, M.; PANTIC, M.; AND PUN, T., 2011b. Multimodal emotion recognition in response to videos. *IEEE transactions on affective computing*, 3, 2 (2011), 211–223. (cited on page 68)
- SPECKENBACH, U. AND GERBER, W., 1999. Reliability of infrared plethysmography in bvp biofeedback therapy and the relevance for clinical application. *Applied psychophysiology and biofeedback*, 24, 4 (1999), 261–265. (cited on page 15)
- SRINIVASAN, V.; ESWARAN, C.; AND SRIRAAM, N., 2007. Approximate entropy-based epileptic eeg detection using artificial neural networks. *IEEE Transactions on Information Technology in Biomedicine*, 11, 3 (2007), 288–295. doi:10.1109/TITB.2006.884369. (cited on page 40)
- STONE, D. C., 1995. Application of median filtering to noisy data. *Canadian Journal of chemistry*, 73, 10 (1995), 1573–1581. (cited on page 76)
- SU, J.-H.; CHANG, W.-Y.; AND TSENG, V. S., 2013. Personalized music recommendation by mining social media tags. *Procedia Computer Science*, 22 (2013), 303–312. (cited on page 135)
- SUDHEESH, N. AND JOSEPH, K., 2000. Investigation into the effects of music and meditation on galvanic skin response. *ITBM-RBM*, 21, 3 (2000), 158–163. (cited on page 2)
- SUN, X. AND LV, M., 2019. Facial expression recognition based on a hybrid model combining deep and shallow features. *Cognitive Computation*, 11, 4 (2019), 587–597. (cited on page 13)
- SUTHERLING, W. W.; HERSHMAN, L. M.; MILLER, J. Q.; AND LEE, S. I., 1980. Seizures induced by playing music. *Neurology*, 30, 9 (1980), 1001–1001. (cited on page 22)
- TANG, T. B.; CHONG, J. S.; KIGUCHI, M.; FUNANE, T.; AND LU, C.-K., 2021. Detection of emotional sensitivity using fnirs based dynamic functional connectivity. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29 (2021), 894–904. (cited on page 19)
- TARRANT, J.; VICZKO, J.; AND COPE, H., 2018. Virtual reality for anxiety reduction demonstrated by quantitative eeg: a pilot study. *Frontiers in psychology*, 9 (2018), 1280. (cited on page 3)
- TAYAH, T. F.; ABOU-KHALIL, B.; GILLIAM, F. G.; KNOWLTON, R. C.; WUSHENSKY, C. A.; AND GALLAGHER, M. J., 2006. Musicogenic seizures can arise from multiple temporal lobe foci: intracranial eeg analyses of three patients. *Epilepsia*, 47, 8 (2006), 1402–1406. (cited on page 137)
- TAYLOR, R. S.; SANDER, J. W.; TAYLOR, R. J.; AND BAKER, G. A., 2011. Predictors of health-related quality of life and costs in adults with epilepsy: a systematic review. *Epilepsia*, 52, 12 (2011), 2168–2180. (cited on page 2)

- THE HUMAN BRAIN AND SEIZURES. The human brain and seizures. <https://www.epilepsy.org.au/about-epilepsy/understanding-epilepsy/the-human-brain-and-seizures/>. (cited on page 66)
- THIBODEAU, R.; JORGENSEN, R. S.; AND KIM, S., 2006. Depression, anxiety, and resting frontal eeg asymmetry: a meta-analytic review. *Journal of abnormal psychology*, 115, 4 (2006), 715. (cited on page 17)
- THOMA, M. V.; LA MARCA, R.; BRÖNNIMANN, R.; FINKEL, L.; EHLERT, U.; AND NATER, U. M., 2013. The effect of music on the human stress response. *PloS one*, 8, 8 (2013), e70156. (cited on page 106)
- THOMPSON, A. E. AND VOYER, D., 2014. Sex differences in the ability to recognise non-verbal displays of emotion: A meta-analysis. *Cognition and Emotion*, 28, 7 (2014), 1164–1195. (cited on page 12)
- THURMAN, D. J.; BEGHI, E.; BEGLEY, C. E.; BERG, A. T.; BUCHHALTER, J. R.; DING, D.; HESDORFFER, D. C.; HAUSER, W. A.; KAZIS, L.; KOBAYASHI, R.; ET AL., 2011. Standards for epidemiologic studies and surveillance of epilepsy. *Epilepsia*, 52 (2011), 2–26. (cited on page 2)
- TORREY, L. AND SHAVLIK, J., 2010. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, 242–264. IGI global. (cited on page 42)
- TRIWYANTO, T.; WAHYUNGGORO, O.; NUGROHO, H. A.; AND HERIANTO, H., 2017. An investigation into time domain features of surface electromyography to estimate the elbow joint angle. *Advances in Electrical and Electronic Engineering*, 15, 3 (2017), 448–458. (cited on page 36)
- TROHIDIS, K.; TSOUMAKAS, G.; KALLIRIS, G.; VLAHAVAS, I. P.; ET AL., 2008. Multi-label classification of music into emotions. In *ISMIR*, vol. 8, 325–330. (cited on page 23)
- TUR, G.; DENG, L.; HAKKANI-TÜR, D.; AND HE, X., 2012. Towards deeper understanding: Deep convex networks for semantic utterance classification. In *2012 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 5045–5048. IEEE. (cited on page 113)
- TURNBULL, D.; BARRINGTON, L.; TORRES, D.; AND LANCKRIET, G., 2008. Semantic annotation and retrieval of music and sound effects. *IEEE Transactions on Audio, Speech, and Language Processing*, 16, 2 (2008), 467–476. (cited on page 23)
- ULSTEIN, I.; WYLLER, T. B.; AND ENGEDAL, K., 2007. High score on the relative stress scale, a marker of possible psychiatric disorder in family carers of patients with dementia. *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences*, 22, 3 (2007), 195–202. (cited on page 12)

- UMBRELLO, M.; SORRENTI, T.; MISTRALETTI, G.; FORMENTI, P.; CHIUMELLO, D.; AND TERZONI, S., 2019. Music therapy reduces stress and anxiety in critically ill patients: a systematic review of randomized clinical trials. (2019). (cited on page 2)
- ÜNAL, H. P.; GÖKMEN, G.; AND YUMURTACI, M., 2020. Emotion classification with deep dataset: Survey. In *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 1–6. IEEE. (cited on page 27)
- VALENZA, G.; LANATA, A.; AND SCILINGO, E. P., 2011. The role of nonlinear dynamics in affective valence and arousal recognition. *IEEE transactions on affective computing*, 3, 2 (2011), 237–249. (cited on page 24)
- VALSTAR, M.; GRATCH, J.; SCHULLER, B.; RINGEVAL, F.; LALANNE, D.; TORRES TORRES, M.; SCHERER, S.; STRATOU, G.; COWIE, R.; AND PANTIC, M., 2016. Avec 2016: Depression, mood, and emotion recognition workshop and challenge. In *Proceedings of the 6th international workshop on audio/visual emotion challenge*, 3–10. (cited on page 36)
- VECCHIO, F.; MIRAGLIA, F.; ALÙ, F.; MENNA, M.; JUDICA, E.; COTELLI, M.; AND ROSSINI, P. M., 2020. Classification of alzheimer’s disease with respect to physiological aging with innovative eeg biomarkers in a machine learning implementation. *Journal of Alzheimer’s Disease*, 75, 4 (2020), 1253–1261. (cited on page 39)
- VIJAYALAKSHMI, K.; SRIDHAR, S.; AND KHANWANI, P., 2010. Estimation of effects of alpha music on eeg components by time and frequency domain analysis. In *International Conference on Computer and Communication Engineering (ICCCE’10)*, 1–5. IEEE. (cited on page 67)
- VRIJKOTTE, T. G.; VAN DOORNEN, L. J.; AND DE GEUS, E. J., 2000. Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. *Hypertension*, 35, 4 (2000), 880–886. (cited on page 21)
- WALKER, J. L., 1977. Subjective reactions to music and brainwave rhythms. *Physiological Psychology*, 5, 4 (1977), 483–489. (cited on page 72)
- WANG, X.-W.; NIE, D.; AND LU, B.-L., 2011. Eeg-based emotion recognition using frequency domain features and support vector machines. In *International conference on neural information processing*, 734–743. Springer. (cited on page 39)
- WEGRZYN, M.; VOGT, M.; KIRECLIOGLU, B.; SCHNEIDER, J.; AND KISSLER, J., 2017. Mapping the emotional face. how individual face parts contribute to successful emotion recognition. *PLoS one*, 12, 5 (2017), e0177239. (cited on page 12)
- WEISS, M. W.; TREHUB, S. E.; SCHELLENBERG, E. G.; AND HABASHI, P., 2016. Pupils dilate for vocal or familiar music. *Journal of Experimental Psychology: Human Perception and Performance*, 42, 8 (2016), 1061. (cited on page 21)
- WHITNEY, A. W., 1971. A direct method of nonparametric measurement selection. *IEEE Transactions on Computers*, 100, 9 (1971), 1100–1103. (cited on page 37)

- WIESER, H. G.; HUNGERBÖHLER, H.; SIEGEL, A. M.; AND BUCK, A., 1997. Musicogenic epilepsy: review of the literature and case report with ictal single photon emission computed tomography. *Epilepsia*, 38, 2 (1997), 200–207. (cited on pages 23 and 106)
- WILSON, G. F. AND RUSSELL, C. A., 2003. Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Human Factors*, 45, 4 (2003), 635–644. doi:10.1518/hfes.45.4.635.27088. https://doi.org/10.1518/hfes.45.4.635.27088. PMID: 15055460. (cited on page 40)
- WU, G.; LIU, G.; AND HAO, M., 2010. The analysis of emotion recognition from gsr based on pso. In *2010 International symposium on intelligence information processing and trusted computing*, 360–363. IEEE. (cited on page 24)
- XIE, W. AND XUE, W., 2021. Wb-knn for emotion recognition from physiological signals. *Optoelectronics Letters*, 17, 7 (2021), 444–448. (cited on page 38)
- YANG, C.-Y.; MIAO, N.-F.; LEE, T.-Y.; TSAI, J.-C.; YANG, H.-L.; CHEN, W.-C.; CHUNG, M.-H.; LIAO, Y.-M.; AND CHOU, K.-R., 2016. The effect of a researcher designated music intervention on hospitalised psychiatric patients with different levels of anxiety. *Journal of clinical nursing*, 25, 5-6 (2016), 777–787. (cited on page 22)
- YANG, D.; CHEN, X.; AND ZHAO, Y., 2011. A lda-based approach to lyric emotion regression. In *Knowledge Engineering and Management*, 331–340. Springer. (cited on page 23)
- YANG, D.; YOO, S.-H.; KIM, C.-S.; AND HONG, K.-S., 2019a. Evaluation of neural degeneration biomarkers in the prefrontal cortex for early identification of patients with mild cognitive impairment: an fnirs study. *Frontiers in human neuroscience*, 13 (2019), 317. (cited on page 41)
- YANG, F.; ZHAO, X.; JIANG, W.; GAO, P.; AND LIU, G., 2019b. Multi-method fusion of cross-subject emotion recognition based on high-dimensional eeg features. *Frontiers in computational neuroscience*, 13 (2019), 53. (cited on page 17)
- YANG, J. AND HONAVAR, V., 1998. Feature subset selection using a genetic algorithm. In *Feature extraction, construction and selection*, 117–136. Springer. (cited on page 37)
- YANG, Y. AND CHEN, H., 2011a. Predicting the distribution of perceived emotions of a music signal for content retrieval. In *IEEE Trans. Audio, Speech and Lang. Proc.*, vol. 19, 2184–2196. (cited on page 28)
- YANG, Y.-H. AND CHEN, H. H., 2011b. *Music emotion recognition*. CRC Press. (cited on page 23)
- YU, D. AND SUN, S., 2020. A systematic exploration of deep neural networks for eda-based emotion recognition. *Information*, 11, 4 (2020), 212. (cited on page 14)
- ZHAI, J. AND BARRETO, A., 2006. Stress recognition using non-invasive technology. In *FLAIRS Conference*, 395–401. (cited on page 21)

- ZHANG, A.; LIPTON, Z. C.; LI, M.; AND SMOLA, A. J., 2021. Dive into deep learning. *arXiv preprint arXiv:2106.11342*, (2021). (cited on page 41)
- ZHANG, H., 2020. Expression-eeg based collaborative multimodal emotion recognition using deep autoencoder. *IEEE Access*, 8 (2020), 164130–164143. (cited on page 13)
- ZHANG, K.; ZHANG, H.; LI, S.; YANG, C.; AND SUN, L., 2018a. The pmemo dataset for music emotion recognition. In *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval, ICMR '18* (Yokohama, Japan, 2018), 135–142. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3206025.3206037. <https://doi.org/10.1145/3206025.3206037>. (cited on page 28)
- ZHANG, S.; ZHAO, X.; AND TIAN, Q., 2019. Spontaneous speech emotion recognition using multiscale deep convolutional lstm. *IEEE Transactions on Affective Computing*, (2019). (cited on page 13)
- ZHANG, T.; ZHENG, W.; CUI, Z.; ZONG, Y.; AND LI, Y., 2018b. Spatial–temporal recurrent neural network for emotion recognition. *IEEE transactions on cybernetics*, 49, 3 (2018), 839–847. (cited on page 41)
- ZHANG, Z.-M.; CHEN, S.; AND LIANG, Y.-Z., 2010. Baseline correction using adaptive iteratively reweighted penalized least squares. *Analyst*, 135, 5 (2010), 1138–1146. (cited on page 35)
- ZHAO, J.; LUI, H.; MCLEAN, D. I.; AND ZENG, H., 2007. Automated autofluorescence background subtraction algorithm for biomedical raman spectroscopy. *Applied spectroscopy*, 61, 11 (2007), 1225–1232. (cited on page 35)
- ZHAO, W.; ZHOU, Y.; TIE, Y.; AND ZHAO, Y., 2018. Recurrent neural network for MIDI music emotion classification. In *2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 2596–2600. IEEE. (cited on page 41)
- ZVAREVASHE, K. AND OLUGBARA, O. O., 2018. Gender voice recognition using random forest recursive feature elimination with gradient boosting machines. In *2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD)*, 1–6. doi:10.1109/ICABCD.2018.8465466. (cited on page 39)
- ZVAREVASHE, K. AND OLUGBARA, O. O., 2020. Recognition of cross-language acoustic emotional valence using stacked ensemble learning. *Algorithms*, 13, 10 (2020), 246. (cited on page 113)