

**VARIANTS OF CONVOLUTIONAL NEURAL
NETWORKS FOR CLASSIFICATION OF
MULTICHANNEL EEG SIGNALS:
A STUDY BASED ON INFLUENCE OF MUSIC
AND EMOTION ON HUMAN BRAIN**

CHEAH KIT HWA

**MASTER OF
ENGINEERING SCIENCE**

**FACULTY OF
ENGINEERING AND GREEN TECHNOLOGY**

UNIVERSITI TUNKU ABDUL RAHMAN

2021

**VARIANTS OF CONVOLUTIONAL NEURAL NETWORKS FOR
CLASSIFICATION OF MULTICHANNEL EEG SIGNALS:
A STUDY BASED ON INFLUENCE OF MUSIC AND EMOTION ON
HUMAN BRAIN**

By

CHEAH KIT HWA

**A dissertation submitted to
Faculty of Engineering and Green Technology,
Universiti Tunku Abdul Rahman,
in partial fulfillment of the requirements for the degree of
Master of Engineering Science
2021**

ABSTRACT

Electroencephalography (EEG) records the electrical potential fields generated by the neuronal activities at various parts of the brain. With increasing popularity and interest from the research community of different disciplinary background, the applicability of EEG is getting more promising in many different areas from the research settings to the clinical neurology for diagnosis and treatment monitoring. Nonetheless, efficiently identifying and extracting the highly representative EEG signal features for a particular scenario is crucial for the success of the classification task. Convolutional neural network (CNN) which is specialized in processing the data structures with grid-like topology can be helpful in achieving automated extraction of key representative features from the multichannel EEG signals.

While EEG and images both have grid-like topology, their data format are differently organized in the grid. This project which consists of three studies aims at developing CNN classifiers that better fit for the processing of EEG signals and identifying the factors that influence the performance of the CNN classifiers, based on the EEG data obtained from the experiments studying the influence of music and emotion on human brain.

In Study 1 which is based on the influence of music on the brain, the impacts of various architectural aspects of CNN on the classification performance, the importance of spatial-dimension convolution in EEG data classification, and the computational resource-efficiency between CNN with 2D

and 1D convolution kernels are evaluated. In Study 2 which deals with emotion recognition, the possibility of reducing the internal parameters of the model by using double-path convolution with kernels of different dilation factors is investigated. Study 3, which is also an emotion recognition study, investigates the applicability of the CNN models originally developed for image processing for EEG data classification and further explores the architectural changes that can help in performance improvement in EEG processing.

This project has also revealed the non-uniform or lateralized influence of music and emotion on the human brain, based on the discrepancy in classification accuracies between EEG subsets from different brain regions. For the classification of EEG listening to different pieces of music, the test accuracy achieved using the EEG channels from the left cerebral hemisphere (88.91%) is approximately 5% higher than the accuracy achieved with the right hemisphere (84.12%). The test accuracy discrepancy in music-EEG classification is even higher (10% difference) between EEG channels from the frontal cerebral lobes (84.93%) and EEG channels from the temporal, parietal and occipital lobes combined (74.69%). For emotion classification using EEG in Study 3, emotion classification accuracy achieved with EEG from the temporal region (83.84%) is approximately 7% higher than that achieved using EEG from the frontal (76.90%) and parietal (76.78%) regions. There is 5.1% accuracy discrepancy in emotion classification using the EEG channels from the left (88.48%) and right (83.38%) cerebral hemispheres.

While music appears to be more greatly affecting the frontal cerebral lobes than the other (temporal, parietal and occipital) lobes, emotion is more greatly reflected on the EEG obtained near the temporal lobes. In addition, both the music and emotion have greater influence on EEG of the left cerebral hemisphere than the right cerebral hemisphere. The neurological findings relevant to the influence of music and emotion on human brain are potentially helpful in selecting a smaller subset of EEG channels for the particular classification application.

ACKNOWLEDGEMENT

I would like to express my sincere appreciation to my supervisor, Dr. Humaira Nisar, for her guidance, encouragement, inspiration, and patience throughout my Master study.

I would like to convey my earnest gratitude to Universiti Tunku Abdul Rahman (UTAR) and the Centre for Healthcare Science and Technology (CHST) of UTAR for the scholarship, research grant and the conference financial support offered. The Master project is financially supported by the UTAR Research Fund (UTARRF) with Grant No.: IPSR/RMC/UTARRF/2018-C1/H03, and the Excellent Research Center Award Fund of the CHST of UTAR.

I would also like to thank all the lecturers and staff members of the Department of Electronic Engineering and the Faculty of Engineering and Green Technology of UTAR for their technical assistance and experience shared.

I am also thankful to my research laboratory mates for their company, guidance and advice, with special mention to Rab Nawaz and Thee Kang Wei.

I am forever indebted to my parents who have been so loving, kind and supportive. Also, I wish to express my appreciation to all my family members and friends.

PERMISSION SHEET

FACULTY OF ENGINEERING AND GREEN TECHNOLOGY

UNIVERSITI TUNKU ABDUL RAHMAN

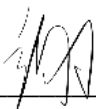
Date: 13 January 2021

SUBMISSION OF DISSERTATION

It is hereby certified that **CHEAH KIT HWA** (ID No: **18AGM06367**) has completed this dissertation entitled “**VARIANTS OF CONVOLUTIONAL NEURAL NETWORKS FOR CLASSIFICATION OF MULTICHANNEL EEG SIGNALS: A STUDY BASED ON INFLUENCE OF MUSIC AND EMOTION ON HUMAN BRAIN**” under the supervision of Assoc. Prof. Dr. HUMAIRA NISAR (Supervisor) from the Department of Electronic Engineering, Faculty of Engineering and Green Technology, Universiti Tunku Abdul Rahman, Assoc. Prof. Dr. YAP VOOI VOON (Co-Supervisor) from the Department of Electronic Engineering, Faculty of Engineering and Green Technology, Universiti Tunku Abdul Rahman, and Prof. Dr. LEE CHEN-YI (Co-Supervisor) from the Department of Electronics Engineering, Institute of Electronics, National Chiao Tung University, Hsinchu.

I understand that the University will upload softcopy of my dissertation in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

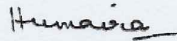


(CHEAH KIT HWA)

APPROVAL SHEET

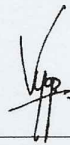
This dissertation entitled “**VARIANTS OF CONVOLUTIONAL NEURAL NETWORKS FOR CLASSIFICATION OF MULTICHANNEL EEG SIGNALS: A STUDY BASED ON INFLUENCE OF MUSIC AND EMOTION ON HUMAN BRAIN**” was prepared by CHEAH KIT HWA and submitted as partial fulfillment of the requirements for the degree of Master of Engineering Science at Universiti Tunku Abdul Rahman.

Approved by:



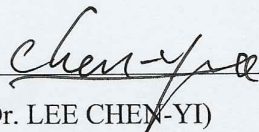
(Assoc. Prof. Dr. HUMAIRA NISAR)
Supervisor
Department of Electronic Engineering
Faculty of Engineering and Green Technology
Universiti Tunku Abdul Rahman, Kampar.

Date: ...13.01.2021.....



(Assoc. Prof. Dr. YAP VOOI VOON)
Co-supervisor
Department of Electronic Engineering
Faculty of Engineering and Green Technology
Universiti Tunku Abdul Rahman, Kampar.

Date: ...13/01/2021.....



(Prof. Dr. LEE CHEN-YI)
Co-supervisor
Department of Electronics Engineering
Institute of Electronics
National Chiao Tung University, Hsinchu.

Date: Jan. 14, 2021

LIST OF TABLES

Table		Page
2.1	Literatures Published on Valence and Arousal Classification using the DEAP Dataset	14
2.2	Recent Research on SEED Dataset	16
3.1	Allocated amount of EEG segments for training, validation (trained-model selection), and performance testing for the binary classification on the short-term music experiment data	24
3.2	Allocated amount of EEG segments for training, validation (trained-model selection), and performance testing for the three-class classification on the short-term music experiment data	25
3.3	The Manually Extracted Features from EEG Signals for the Training of SVM Classifier	39
3.4	Details of the Architecture of the Single-path CNN Model	48
3.5	Details of the Architecture of the Double-path CNN Model with Dilated Convolution in One of the Two Parallel Convolution Paths	49
3.6	Film Clips in SEED Dataset	53
3.7	Emotion Class Labelling of the Video Clips in the SEED Dataset and the EEG Recordings Segmented for Study 3	56
4.1	Performance Comparison across Different CNN Architectures and SVM classifiers with Different Input Features across Ten Folds of Cross Validation Process	73
4.2	Performance Comparison between the 2D-kernel Spatial-Temporal CNN and 1D-kernel Spatial-Temporal CNN over 10-fold Cross-Validation based on 3-class Classification on the Short-Term Music Experiment Dataset	76
4.3	Confusion Matrices of the Cross-Validation Fold with the Lowest Accuracy in Table 4.2 and Figure 4.6 by (a) 2D-kernel Spatial-Temporal CNN and (b) 1D-kernel Spatial-Temporal CNN	78
4.4	Detailed Comparison of the Trainable Parameters in the Two Different Spatial-Temporal CNN Classifiers	80
4.5	Training Time and Prediction (Test) Time of Different EEG Classifiers for the Short-term Music Experiment Binary Classification	81

4.6	Three-class Classification Performance (Accuracy and Cross-Entropy Loss) Achieved with Left vs. Right Cerebral Hemispheric EEG Signals	85
4.7	Three-class Classification Performance (Accuracy and Cross-Entropy Loss) Achieved with Frontal-lobe EEG vs. TPO-lobe EEG	87
4.8	Average four-fold Cross-Validation Accuracies for the Three-Class Emotional Valence Classification	89
4.9	Averaged 4-fold Cross-Validation Accuracies for the Three-Class Emotion Arousal Level Classification	90
4.10	Convergence Time Needed during Model Training for the <i>ResNet</i> and <i>VGG</i>	100

LIST OF FIGURES

Figure		Page
3.1	Steps for the collection of EEG datasets for (a) short-term music experiment and (b) long-term music experiment	20
3.2	Methodological Steps from Model Construction to Model Training, Validation and Testing	28
3.3	Architectures of Temporal-dimension CNN with no Spatial-dimension Convolution	36
3.4	Architectures of Spatial-Temporal CNN constructed with (a) 2-dimensional kernels and (b) only 1-dimensional kernels	37
3.5	Layout of the 32 EEG Electrodes in DEAP Dataset	41
3.6	Emotional Valence and Arousal Scaling of SAM and the Three Classes used in this Study	42
3.7	Segmentation and Allocation of the Full-length EEG Recording into Training, Validation and Test Datasets for Four-fold Cross-validation	45
3.8	Illustration of the sliding window approach for the extraction of overlapping one-second sub-segments from every ten-second EEG segment	45
3.9	Session Flow of Data Collection Process of SEED Experiment	54
3.10	Placement Layout of EEG Channels in SEED Experiment	55
3.11	Architectural details of (a) the original ResNet18 and its modified variants (b, c) for EEG signal processing	59
3.12	Architectural details of (a) VGG16 with 1D kernels, (b) VGG14 with 1D kernels and (c) VGG14 without batch normalization	60
4.1	Performance Log of the Model Training-Validation Process using Different Amounts of Convolution Kernels Working on Short-term Music Experiment Binary Classification	66
4.2	Performance Log of the Model Training-Validation Process with Different Widths and Depths of Fully-Connected (FC) Perceptron Networks	68
4.3	Performance Log for the Model Training-Validation Process with/without Pooling Mechanism on Short-term Music Experiment Binary Classification	69

4.4	Performance Log for the Model Training-Validation Process of Spatial-Temporal CNN vs. Pure-Temporal-CNN without Spatial Convolution	71
4.5	Graphical Performance Comparison across Different CNN Architectures and SVM classifiers with Different Input Features across Ten Folds of Cross Validation Process	72
4.6	Graphical illustration of the Performance Comparison between the 2D-kernel Spatial-Temporal CNN and 1D-kernel Spatial-Temporal CNN over 10-fold Cross-Validation based on 3-class Classification on the Short-Term Music Experiment Dataset	77
4.7	Graphical Comparison of Training Time and Prediction (Test) Time of Different EEG Classifiers on the Short-term Music Experiment Binary Classification	81
4.8	Graphical Comparison of Three-class Classification Performance Achieved with Left vs. Right Cerebral Hemispheric EEG Signals	85
4.9	Graphical Three-class Classification Performance Achieved with Frontal-lobe EEG vs. TPO-lobe EEG	87
4.10	SEED 3-class Emotion Recognition Accuracy by Variants of ResNet18 using Different Subsets of EEG Channels	95
4.11	Training-validation Performance Log of Variants of ResNet18-1D	96
4.12	Classification Accuracy of Emotion-labelled EEG by ResNet18-1D and VGG16 Variants	99
4.13	SEED 3-class Emotion Recognition Accuracy using Different Subsets of EEG Channels along the Nasion-Inion Axis	102
4.14	SEED 3-class Emotion Recognition Accuracy Comparison using Left and Right Hemispheric EEG Channels	104

LIST OF ABBREVIATIONS

AgCl	: Argentum Chloride
AI	: Artificial Intelligence
ApEn	: Approximate Entropy
BN	: Batch Normalization
CE	: Cross Entropy
CNN	: Convolutional Neural Network
Conv	: Convolution
CPSD	: Cross Power Spectral Density
DE	: Differential Entropy
DEAP	: a Dataset for Emotion Analysis using EEG, Physiological and video signals
DNN	: Deep Neural Network
DWT	: Discrete Wavelet Transform
ECoG	: Electrocorticography
EEG	: Electroencephalography
FB-CSP	: Filter Bank Common Spatial Patterns
FC-MLP	: Fully-Connected Multilayer Perceptrons
FFT	: Fast Fourier Transform
FIR	: Finite Impulse Response
GELM	: Graph-regularized Extreme Learning Machine
GPU	: Graphical Processing Unit
HOC	: Higher-Order Crossings
HOS	: Higher-Order Statistics
KNN	: K-Nearest Neighbours
LFCC	: Linear-Frequency Cepstral Coefficients
LOO	: Leave-One-subject-Out validation
MEG	: Magnetoencephalography
PCA	: Principal Component Analysis
PSD	: Power Spectral Density
RAM	: Random Access Memory

RASM	: Rational Asymmetry
ReLU	: Rectified Linear Unit
ResNet	: Residual Network
ResNet18-1D-(S-then-T)	: ResNet with 1D kernels initiated with Spatial convolutions followed by Temporal convolutions
ResNet18-1D-(T-then-S)	: ResNet with 1D kernels initiated with Temporal convolutions followed by Spatial convolutions
ResNet18-1D-(S-T-alter)	: 1D-kernel ResNet with alternating spatial and temporal convolution initiated with spatial convolution
ResNet18-1D-(T-S-alter)	: 1D-kernel ResNet with alternating spatial and temporal convolution initiated with temporal convolution
rLDA	: regularized Linear Discriminant Analysis
RNN	: Recurrent Neural Network
SampEn	: Sampling Entropy
SAM	: Self-Assessment Manikin
SEED	: Shanghai-Jiao-Tong-University (SJTU) Emotion EEG Dataset
SNR	: Signal-to-noise ratio
std	: standard deviation
SVM	: Support Vector Machine
var	: variance
WE	: Wavelet Entropy

TABLE OF CONTENTS

	Page
ABSTRACT	ii
ACKNOWLEDGEMENTS	v
PERMISSION SHEET	vi
APPROVAL SHEET	vii
LIST OF TABLES	viii
LIST OF FIGURES	x
LIST OF ABBREVIATIONS	xii
CHAPTER	
1.0 INTRODUCTION	1
1.1 Background	1
1.1.1 Electroencephalography	1
1.1.2 Challenges Faced in EEG Signal Classification	4
1.1.3 Deep Learning and Convolutional Neural Network	5
1.2 Problem Statement	8
1.3 Objectives	9
2.0 LITERATURE REVIEW	10
2.1 Visual and Auditory Stimuli EEG Classification	10
2.2 Emotion Classification	11
2.2.1 DEAP Dataset Literature Review	12
2.2.2 SEED Dataset Literature Review	15

3.0	METHODOLOGY	18
3.1	Music-Listening EEG Classification (Study 1)	18
3.1.1	Experiment Description and EEG Signal Recording	18
3.1.2	Allocation of EEG Segments into Training, Validation, and Test Datasets	22
3.1.2.1	EEG Dataset for Short-term Music Experiment Binary Classification	22
3.1.2.2	EEG Dataset for Short-term Music Experiment Three-class Classification	24
3.1.3	Computing Environment Settings	26
3.1.4	Methodological Steps of Model Training, Validation, and Testing	27
3.1.5	Design of Classifiers	29
3.1.5.1	CNN Models	29
3.1.5.2	SVM Classifier	38
3.2	CNN for Personalized Emotion Classification (Study 2)	40
3.2.1	DEAP Dataset	40
3.2.2	Emotion-Class Relabelling	42
3.2.3	EEG Allocation into Training, Validation, and Test Set Allocation	43
3.2.4	Architectural Details of the CNN	46
3.2.5	Dilated Convolution	50
3.2.6	Training, Validation, and Testing of the CNN Models	51
3.3	Residual Network and VGG for Emotion EEG Classification (Study 3)	53
3.3.1	SEED Dataset	53
3.3.2	EEG Data Preprocessing	55
3.3.3	Optimizing <i>ResNet</i> & <i>VGG</i> for EEG Signals	56
3.3.3.1	<i>ResNet</i> Optimization	56
3.3.3.2	<i>VGG</i> Optimization	61
3.3.4	Model Training	63

4.0	RESULTS & DISCUSSION	64
4.1	Music-Listening EEG Classification (Study 1)	64
4.1.1	Adjusting for the Suitable Hyperparameters and Constituent Components in the CNN Architecture	64
4.1.2	The Importance of Convolution across the Spatial Dimension for EEG Signal Classification	70
4.1.3	Ten-fold Cross-Validation Comparing the CNN with SVM	71
4.1.4	Three-class Classification by Spatial-Temporal CNN with 1D vs. 2D Kernels	74
4.1.5	Comparing the Computational Efficiency of Different Classifiers in Terms of Size of Model & Computational Time	79
	4.1.5.1 Model Size	79
	4.1.5.2 Computational Time	81
4.1.6	Brain State Classification based on EEG Signals from Different Brain Lobes	83
	4.1.6.1 Left Hemisphere vs. Right Hemisphere of the Cerebra	83
	4.1.6.2 Significance of Frontal-lobe Signals vs. Temporal-Parietal-Occipital (TPO) EEG Signals	86
4.2	CNN for Personalized Emotion Classification (Study 2)	88
4.3	Residual Network and VGG for Emotion EEG Classification (Study 3)	91
4.3.1	Performance of Variants of <i>ResNet18</i>	91
4.3.2	Performance of <i>ResNet</i> vs. <i>VGG</i>	96
4.3.3	EEG Channel Significance for Emotion Recognition	100
	4.3.3.1 Electrode Distance from the Midline	101
	4.3.3.2 Significance Along the Nasion-Inion Axis	102
	4.3.3.3 Cerebral Lateralization of Emotion	103
	4.3.3.4 Comparing across all the EEG Channel Subsets	105

5.0	CONCLUSION	107
5.1	CNN for EEG Classification	107
5.2	Non-uniform Influence of Music on Regional Brain Waves	110
5.3	Emotion-Relevance of Different EEG Channels	111
5.4	Recommendation for Future Work	111
	REFERENCES/BIBLIOGRAPHY	113
	APPENDICES	

CHAPTER 1

INTRODUCTION

1.1 Background

1.1.1 Electroencephalography

Electroencephalography, commonly abbreviated as EEG, is a neurological measurement technique that detects and records the electric potentials generated by the neurological activity of the brain, with the recording electrodes being placed in non-invasive direct contact with the scalp, as contrasted with electrocorticography (ECoG) which has the recording electrodes placed directly in contact with the surface of the brain cortex for the measurement of electrical activity of the particular region.

Neuronal processes of the brain produce trans-membrane (that spans across the cellular membrane) electrical currents which are also detectable in the extracellular medium. These electric currents generated from the active cellular activities within the neighbouring regions of the brain superimpose and generate a field of electrical potentials. The generated field of electrical potentials can also be monitored with extra-cellular electrodes (Buzsáki *et al.*,

2012). These measurement electrodes typically function over hundred hertz of sampling frequency, producing sub-millisecond time resolution. Each of the recording electrode generates a different channel of electrical signal. The multi-channel electrical signals recorded at fine temporal resolution are useful for interpreting many aspects of neuronal communication and states of the brain.

This multi-channel array of brain signals, when recorded from the scalp, has been conventionally known as the electroencephalogram. When recorded from the intra-cranial sub-dural electrodes, the signals are known as the electrocorticogram (ECoG). In contrast, the magnetoencephalogram (MEG) refers to the recording of magnetic field generated by the same neuronal processes.

With increasing popularity and focus from research community of different disciplinary background, EEG has advanced to the current state of being widely applied in clinical neurology for diagnostic, treatment monitoring, prognostic, and research purposes.

EEG is valuable in assisting the diagnostic process. EEG can be helpful in identifying the type of seizure in epileptic patients which in turn determine the choice of medication (Smith, 2005a). For instance, EEG is especially crucial for diagnosing nonconvulsive seizure and nonconvulsive status epilepticus (Kaplan, 2007). EEG can also serve in identifying encephalopathies of various origins (e.g. tumour, trauma/stroke, encephalitis, or deranged

metabolic state), and as an adjunct test for confirmation of brain death in persistent coma (Smith, 2005b; St. Louis and Frey, 2016).

EEG can be used in guiding medical treatment. As the compromised cerebral perfusion is associated with changes in electrical cortical activity, EEG can be used as an informative adjunct indicator to monitor the cerebral blood flow during surgery (Foreman and Claassen, 2012; Kreitzer *et al.*, 2018). Pharmaco-EEG is also a potential emerging trend where the EEG is used in the assessment and guidance of therapeutic drug administration, as well as in the assessment of efficacy-toxicity of therapeutic agents (Swatzyna *et al.*, 2015; Höller *et al.*, 2018). Besides, EEG shows potential in aiding the prognosis of postanoxic coma, cardiac arrest, and epilepsy (Smith, 2005a; Hofmeijer and van Putten, 2016; Muhlhofer and Szaflarski, 2018).

The involvement of EEG in neuroscience research is also gaining momentum. For example, EEG has been used to study normal sleep pattern and sleep anomaly, and various other psychiatric conditions such as schizophrenia, attention deficit hyperactive disorder and depression. Besides, EEG is used in various cognition studies such as for mapping out the functional brain topology for the cognitive task of interest.

1.1.2 Challenges Faced in EEG Signal Classification

Notwithstanding being loaded with various pathological and physiological information regarding the neurological activities of the cerebra, the EEG recordings have rather low signal-to-noise ratio (SNR). The reason for the low SNR of EEG lies innately in the non-invasive nature of the EEG recording method. With the recording electrodes of the EEG being placed in the scalp which is centimeters apart from the underlying brain structures, the detectible electrical signals truly generated from cerebral cortices is typically only at the range of microvolts. Therefore, the electrical activities of many other sources such as the muscular activities of blinking the eyelids, rotation of the eyeballs, and other muscular contractions especially of those in the face and neck can be easily picked up by the recording electrodes as the electrical noises. In addition, environment electric noises such as those emitted from the electrical lines from the proximity can also be undesirably recorded by the highly sensitive EEG electrodes.

Being an organ packed with heavily-interconnected electrically-excitabile neurons, the brain by itself has constantly multiple different lobes or regions being activated and deactivated beyond voluntary control. The electrical potentials generated from various regions of the brain can be potentially interfering with and masking the electrical potentials generated by the brain region of interest. Compounding to the above-mentioned difficulty in getting high spatial resolution, the non-invasive nature of EEG requires its measuring electrodes to be located extracranially at a considerable distance from the

cerebral cortices. Therefore, the EEG technique innately has low spatial resolution with respect to the source of neurological activities in the brain, in spite of the development of high-density EEG recordings headset which can accommodate over two hundred of recording electrodes.

To manually decode the information in the EEG signals and identify the key features of EEG signals is still a research challenge. Meanwhile, the performance of traditional feature-based EEG classifiers is highly dependent on the discriminative quality of the EEG features or the relevance of the feature set to the particular activity of interest. In the multi-channel time series of EEG data, every different EEG channel being measured at different location of the scalp can be embedded with different relevant features, which adds to the complexity of manual identification of useful EEG features. The manual critical feature identification from the raw EEG signals is thus a very time-consuming and effort-consuming process. Many useful EEG signal features may yet be beyond the current knowledge collection of the research community.

1.1.3 Deep Learning and Convolutional Neural Network

The recent AI techniques that enable machine to perform deep hierarchical concept learning (or commonly deep learning) have successfully automated the difficult task of representative feature/concept extraction from different domains of raw data which include the image processing, the audio signal processing, video semantic processing, and language semantic processing

(Graves *et al.*, 2008; Graves *et al.*, 2013; Ng *et al.*, 2015). Deep learning has enabled the end-to-end execution of many different classification and regression scenarios in those domains with complicated raw data.

Deep learning enables computers to accumulate experience and understanding from past encounters of data. Deep learning models allow computers to understand the data in the forms of hierarchies of concepts. Each of the concepts is defined as a function of simpler concepts or concepts of lower hierarchical level (Goodfellow *et al.*, 2016a). This hierarchy of concepts or levels of information abstraction allows computers to efficiently extract meanings from raw data. As this machine understanding of data is built upon many layers of concepts and is granted with greater learning capability with more layers or greater depth of concepts or abstraction, this family of approach of machine learning is hence termed deep learning.

Deep learning is thus a promising tool in overcoming the long-experienced difficulties in manual decoding of EEG signals. Various architectures of deep learning models have been trained and tested by previous studies for the analysis of EEG signals. They have consistently reported better performance in comparison with the methods using manual feature extraction for EEG classification (Ren and Wu, 2014; Behncke *et al.*, 2018; Schirrneister *et al.*, 2017). Nonetheless, the analysis of EEG using deep learning is a new research area that requires further performance improvement for higher reliability in the application in practical settings.

Convolutional neural networks (CNN) are a specific family of neural networks employing the mathematical convolution operation for the specialized processing of data structures with grid-like topology (Goodfellow *et al.*, 2016b). While images are typical example of data with grid-like topology, time-series data (especially multichannel time-series signals) like the EEG recordings are also organized in grid-like pattern which fits the specialized operation of convolutional neural networks very much.

There are three important characteristics of CNN that enable the performance improvement of a machine learning system, namely the sparse interactions of the computational nodes, parameter sharing, and translation-equivariant representation (Goodfellow *et al.*, 2016b). As opposed to the traditional densely connected neural network layers where every output node of the layer interacts with every input unit of the particular layer, the CNN have sparse interaction between the output nodes and input nodes of each of its convolutional layers which is achieved by using convolution kernels with sizes much smaller than the size of input data. Because of the much lower parameter counts in the small-size kernels, the memory requirement and statistical efficiency of the model can be improved significantly. With the property of parameter sharing where each of the constructed kernels is applied at every location of the input data rather than learning a different set of parameters for every location, the CNN has also the property of being equivariant to translation which can be very useful for time-series data such as the EEG. Although the

CNN is not innately equivariant to other kinds of transformations such as the rotation, EEG signals do not typically experience rotational transformation either. Also, the CNN has exemplary ability at exploiting and discovering the spatial or temporal correlation in the input data (Khan *et al.*, 2020).

1.2 Problem Statement

EEG are signals packed with multidimensional information. Identifying and extracting the true representative EEG features is crucial for the success of the tasks of EEG classification. The types of EEG features reported to be relevant in a particular application domain may not be as effectively applicable to the other domains. For instance, EEG frequency band powers as a set of useful features for sleep stages classification may not carry the same effectiveness in emotion state classification. On top of that, many of the EEG features reported to be accurate at intra-personal mental state classification does not scale well into cross-personal classification or the cross-database classification.

Meanwhile, deep convolutional neural networks with automatic representative feature identification has repeatedly been reported with better performance at EEG classification compared with other state-of-art EEG feature extraction and classification methods (Ren and Wu, 2014; Behncke *et al.*, 2018; Schirrmester *et al.*, 2017). Nevertheless, in many of the application domains, deep learning models working on plain EEG signals has not yet seen

performance superiority over the models based on pre-extracted EEG features. Although the potential of deep learning model as both the feature extractor and classifier working on plain EEG signals could likely be limited by the current size of available EEG data pool, the increasing amount of publicly available research databases of EEG signals should warrant the study into the application of very deep networks for plain EEG signals.

Further performance improvement and more understanding of how the architectural aspects of the deep CNN affect its EEG classification performance is in need, in line with the increasing trend of EEG data availability.

1.3 Objectives

The objectives of this research project are:

- i) To design and develop CNN models for EEG signals for mental state classification
- ii) To analyze the effectiveness of variants of CNN models in multichannel EEG signal classification and to identify factors that influence the performance of the CNN classifiers
- iii) To compare the performance between the CNN classifiers and other non-neural-network classifiers in EEG signal classification tasks

CHAPTER 2

LITERATURE REVIEW

2.1 Visual and Auditory Stimuli EEG Classification

The human brain actively responds to all kinds of sensory perception inputs such as the auditory input through the auditory nerves, visual input through the optic nerves, and the input through cutaneous or somatic senses. The accurate classification of the EEG signals recorded from the brain while the subject is receiving different kinds of sensory inputs can be helpful in myriad forms of practical EEG applications. They can serve as a useful guide for the biofeedback therapy, in particular the development of neurofeedback system.

In the work by [Behncke et al. \(2018\)](#), deep CNN models and other non-neural-network classifiers had been used to classify the EEG signals recorded from human subjects visually observing different kinds of robotic actions. The different scenarios of robotic actions include successful robotic operation versus robotic failure in object grasping and pouring tasks. The non-neural network classifiers include the regularized Linear Discriminant Analysis (rLDA) and the filter bank common spatial patterns (FB-CSP) combined with rLDA. In their study, deep CNN has attained accuracy of $75\pm 9\%$, which is significantly better than the other two non-neural-network EEG classifiers. The rLDA classifier has

achieved accuracy of $65\pm 10\%$, while the FB-CSP combined with rLDA has the accuracy of $63\pm 6\%$.

Moinnereau *et al.* (2018) had worked on the classification of EEG signals recorded from human subjects while they were listening to different auditory stimuli (specifically in their case, the English vowels ‘*a*,’ ‘*i*’ and ‘*u*’). The recurrent neural network in their study had attained the average accuracy of 83.2% using sixty-four EEG electrodes and the accuracy of 81.7% using ten selected EEG electrodes.

Also working on the task of EEG classification based on auditory stimulus, **Stober *et al.* (2014)** had constructed convolutional neural networks for the classification of EEG signals recorded during their rhythm perception study. The subjects in their study listened to a wide variety of musical rhythms encompassing twelve Western and twelve East African rhythmic stimuli. For a 24-class EEG classification task, an average classification accuracy of 24.4% is achieved, which is significantly better than the random chance accuracy level of 4.2%.

2.2 Emotion Classification

Two popular emotion recognition databases are used in the Study 2 and Study 3 of this project, namely the DEAP dataset (**Koelstra *et al.*, 2012**) and

the SEED dataset (Duan *et al.*, 2013; Zheng and Lu, 2015). These are both open-access databases available upon request for research purpose.

2.2.1. DEAP Dataset Literature Review

The previous studies working on DEAP dataset for emotion classification is recorded in Table 2.1. The emotional aspects reviewed are the emotional valence and the emotional arousal. “Emotional valence describes the extent to which an emotion is positive or negative, whereas arousal refers to its intensity, i.e., the strength of the associated emotional state.” (Citron *et al.*, 2014) The overview of the DEAP dataset and the measurement of emotional valence and arousal are presented in Section Methodology 3.2.1.

Vast majority of the previous studies working on the emotional valence and arousal level classification had their classifiers constructed based on the manually extracted EEG features, instead of pure EEG signals.

Signal features manually computed from the EEG recording are reported to have different degree of correlation to the emotional measurement. Numerous studies including Rayatdoost and Soleymani (2018), Zheng *et al.* (2017), Yang *et al.* (2017), and Zheng *et al.* (2015) had reported the differential entropy (DE) of different EEG frequency bands as a category of highly emotion-relevant signal features of EEG data. Meanwhile, Petrantonakis and Hadjileontiadis (2010) had recommended EEG features under the category of higher order

crossings (HOC) for EEG-based emotion classification. [Jenke et al. \(2014\)](#) had also reported the HOC-based EEG features as the most useful signal features in the time domain for emotion recognition, among the set of features they had investigated. Linear-frequency cepstral coefficients (LFCC) had also been suggested as a crucial set of features for EEG-based emotion recognition ([Liu et al., 2018](#)).

Among the studies that had focused on manually computing the EEG features for emotion classification, a recent study by [Liu et al. \(2018\)](#) had made use of the EEG features extracted automatically by pre-optimized residual network (*ResNet*). They had achieved the much higher binary emotion valence-level and arousal-level classification accuracy. This indicates that convolutional neural networks (with *ResNet* being a form of convolutional neural network) can be optimized to extract EEG signal features that are relevant for emotion classification.

Table 2.1: Literatures Published on Valence and Arousal Classification using the DEAP Dataset (Cheah *et al.*, 2019b)

Research	Method	Accuracy		Task	
		Valence	Arousal	Valence	Arousal
Yoon & Chung (2013)	Bayes classifier	70.9%	70.1%	2-class	2-class
		53.4%	51.0%	3-class	3-class
Rozgić <i>et al.</i> (2013)	PCA + SVM	76.9%	68.4%	2-class	2-class
Zhang <i>et al.</i> (2013)	EEG + Ontology Reasoning	75.19%	81.74%	2-class	2-class
				8 selected subjects	
Li <i>et al.</i> (2016)	C-RNN	72.06%	74.12%	2-class	2-class
Al-Nafjan <i>et al.</i> (2017)	PSD+DNN	82%	82%	2-class	2-class
Liu <i>et al.</i> (2016)	Multimodal Deep Learning	85.20%	80.50%	2-class	2-class
Tripathi <i>et al.</i> (2017)	CNN	81.41%	73.36%	2-class (LOO)	2-class (LOO)
		66.79%	57.58%	3-class (LOO)	3-class (LOO)
Zheng <i>et al.</i> (2017)	DE + GELM	69.67%		4-class (Valence-arousal space)	
Rayatdoost & Soleymani (2018)	PSD + DE + HOC + HOS + Random Forest	60.86%	58.08%	2-class	2-class
Liu <i>et al.</i> (2018)	ResNet + LFCC + KNN	90.39%	89.06%	2-class	2-class
		61.55%	54.53%	2-class (LOO)	2-class (LOO)

*LOO denotes the cross-validation method of Leaving-One-(subject)-Out

2.2.2. SEED Dataset Literature Review

Increasingly significant research attention has been given to emotion recognition using EEG in the recent years. The research working on SEED dataset in the recent three years (2018-2020) was reviewed and summarized in Table 2.2. Although many of the research works were using one or another kind of neural network classifier, almost all of the attention had been placed on pre-extracted EEG features, instead of plain EEG signals. The experiment design and overview of SEED dataset are presented in Section Methodology 3.3.1.

2.2.3. Algorithms/Methods used in Emotion Classification

The EEG features commonly used in emotion classification include the power spectrum density (PSD), the rational asymmetry (RASM), sampling entropy (SampEn), wavelet entropy (WE), differential entropy (DE), standard deviation (std) of signal, and the Hjorth features (e.g. Hjorth activity, mobility and complexity). Among the EEG features investigated, differential entropy (DE) has consistently been reported as the most emotion-relevant feature type (Song *et al.*, 2018; Li *et al.*, 2019; Wang *et al.*, 2019).

The algorithms used in the research community for emotion classification vary from the conventional machine learning algorithms to the more recent families of artificial neural networks. The conventional machine learning algorithms used in this research area include the Bayesian classifier, Random Forest classifier, K-nearest neighbour (KNN) classifier, logistic regression classifier, support vector machine (SVM) classifier, and their

modified variants such as the graph-regularized sparse linear regression and the sequential backward selection SVM. The families of artificial neural networks reported in the emotion recognition research studies include the CNN, RNN and their modified variants such as the dynamic graph CNN, bidirectional LSTM, and spiking neural network.

Regardless of the classification algorithm reported being a conventional machine learning classifier or a neural-network based model, almost all of them were trained based on the manually-calculated EEG features, instead of the EEG signal itself. The applicability of these feature-based algorithms at overcoming cross-database variation is yet impractical, reported at below 50% accuracy (Lan *et al.*, 2018).

Using plain EEG signals as the input data to the emotion classifiers has currently received relatively much lower research attention. Although the amount of currently available public EEG research database may not yet be sufficiently representative of the general population, the trend of increasing number of publicly available EEG databases shall warrant more research works into the application of very-deep neural networks on plain EEG signals.

Table 2.2: Recent Research on SEED Dataset

Classifier Algorithm / Year	Data Input	Accuracy (%)
Dynamic Graph CNN (Song <i>et al.</i> , 2018)	DE	79.95
Logistic Regression Classifier (Lan <i>et al.</i> , 2018)	DE	72.47
GRSLR (Graph regularized sparse linear regression) (Li <i>et al.</i> , 2019)	DE, Hjorth features	88.41
Bidirectional LSTM (Wang <i>et al.</i> , 2019)	DE / PSD	94.96 / 86.27
Graph convolutional broad network (GCBN) (Zhang <i>et al.</i> , 2019)	DE	94.24
CNN + LSTM (Hwang <i>et al.</i> , 2019)	DE	89.88
Variational Pathway Reasoning (VPR) (Zhang <i>et al.</i> , 2020)	DE	94.3
Sequential Backward Selection SVM (Yang <i>et al.</i> , 2019)	Hjorth features, std, SampEn, WE	89
Spiking NN (Luo <i>et al.</i> , 2020)	DWT, FFT, var	96.67

In line with this, the focus of Study 3 in this project is on eliciting the architectural modification on the originally image-oriented *ResNet* and *VGG* that results in vast improvement of their performance on plain EEG signal. On top of the above focus on *ResNet* and *VGG* architectural improvement and comparison, we have also proposed the location of EEG channels that are most useful in emotion recognition.

CHAPTER 3

METHODOLOGY

3.1. Music-Listening EEG Classification (Study 1)

3.1.1. Experiment Description and EEG Signal Recording

The EEG dataset used in this study is the EEG signals collected for a previous study (Phneah and Nisar, 2017) which had assessed the effect of music on mood improvement. The EEG dataset collected in Phneah and Nisar (2017) is composed of a long-term music experiment and a short-term music experiment.

Figure 3.1(a) illustrates the experiment design of the study by Phneah and Nisar (2017). The short-term experiment of the above-mentioned study had participation of a total of thirty-three subjects (twenty-seven males and five females, with an average of 24.9 ± 7.6 years of age). Three different sets of EEG signals were collected from every participant. Corresponding to Figure 3.1(a), the three sets of EEG are respectively a resting baseline 3-minute EEG recording with eye opened, another open-eye 3-minute EEG recorded while each subject was listening to own favourite music, and another open-eye 3-

minute EEG collected while each subject was listening to the relaxing audio clip of alpha binaural beats.

The long-term experiment lasted for two weeks and had the participation of ten subjects. Five of the ten subjects were allocated into the control group, while the remaining five were assigned into the alpha-binaural-beats treatment group. Throughout the 2-week duration, each participant in the alpha-binaural-beats treatment group listened to the audio clip of alpha binaural beats for thirty minutes daily. On the other hand, the control group was not listening to alpha binaural beats for the 2-week duration. The EEG signals of each participant from both groups were collected three times throughout the 2-week duration, i.e. at the start of the 2-week experiment, at the end of the first week, and at the end of the second week.

For the alpha-binaural-beats treatment group of the long-term experiment, [Nawaz et al. \(2018\)](#) had reported significant changes (with ANOVA p-value test) in the manually extracted EEG features (i.e. absolute alpha-band power, approximate entropy and sample entropy) before and after the experiment. In contrast, no statistically significant difference was reported by [Nawaz, et al. \(2018\)](#) in the above-mentioned EEG signal features manually extracted from three different categories of EEG signals of the short-term music experiment. Therefore, the EEG data of the short-term music experiment in [Phneah and Nisar \(2017\)](#) is set as the target dataset for the classification task in this study.



Figure 3.1: Steps for the collection of EEG datasets for (a) short-term music experiment and (b) long-term music experiment (Nawaz *et al.*, 2018)

Emotive Epoc wireless headset with 14 recording channels was used to record the EEG data. The signal sampling frequency of the recording process was 128 Hz. The placement of the fourteen active electrodes was according to the international 10-20 system, at the specific locations over the scalp, namely “AF3, AF4, F3, F4, F7, F8, FC5, FC6, P7, P8, T7, T8, O1 and O2” ([Headset](#)

Comparison Chart: Technical Specification, [n.d.]. Nomenclature of the EEG channel closely follows the underlying brain lobe over which the recording electrode is placed:

- “AF” overlies the antero-frontal regions.
- “F” overlies the frontal lobes.
- “FC” overlies the fronto-central regions.
- “P” overlies the parietal lobes.
- “T” overlies the temporal lobes.
- “O” overlies the occipital lobes.

Bandpass FIR filter with passband of 1-60 Hz was used to filter the recorded raw EEG signals. *EEGlab* (Delorme and Makeig, 2004) was used to remove the recording artefacts in the EEG signals to obtain the cleaned/pre-processed EEG dataset.

Out of the total of thirty-two subjects partook in the short-term music experiment, only the pre-processed EEG signals of twenty-eight participants were taken in this study for analysis because there are EEG channels with severe measuring errors in the recordings of the remaining four of the subjects.

3.1.2. Allocation of EEG Segments into Training, Validation, and Test Datasets

In this study, binary classification and 3-class classification tasks are performed on the EEG dataset of the short-term experiment.

Only two categories (i.e. the baseline eye-open resting EEG and the alpha-binaural-beat eye-open EEG) of the above-mentioned three categories of EEG signals are used in the binary classification tasks. All the three classes of EEG recordings (i.e. the baseline eye-open resting EEG, respective self-favourite music eye-open EEG, and the alpha-binaural-beat eye-open EEG) are used in the 3-class classification task. Each full recording of the pre-processed EEG is segmented along the temporal(time)-dimension into non-overlapping short segments of equal lengths. The length of every short segment is two seconds, which is equivalent to 256 sampling points.

3.1.2.1. EEG Dataset for Short-term Music Experiment Binary Classification

There are 3,850 two-second EEG segments in the short-term music experiment binary-classification dataset. None of these EEG segments has any degree of overlapping with another. Out of the total of 3,850 segments of EEG, 1,839 segments are the baseline eye-open resting EEG before listening to any

piece of music and the remaining 2,011 segments of EEG signals were collected while the subjects were listening to the audio clip of alpha-binaural beats. All of the 3850 two-second segments of EEG are shuffled randomly with the *sklearn.utils.shuffle()* Python function and then divided into training-validation data pool and the test data pool at 10:1 ratio using the *sklearn.model_selection.train_test_split()* Python function. This gives 3,500 segments of EEG in the training-validation data pool and 350 segments of EEG in the test data pool.

The data pool for training and validation which contains 3500 EEG segments is further split into ten non-intersecting sub-sets for the implementation of ten-fold cross-validation. Every fold of the ten-fold cross-validation process uses a different EEG data sub-sets as the validation data pool (for the purpose of model selection). The other nine sub-sets of EEG data are used as the model training pool.

Table 3.1 presents the EEG data distribution in each cycle of model training and model validation. The double asterisks (**) in Table 3.1 are to indicate the number of EEG segments which is approximately half of the total of the training set. Analogously, the triple asterisks (***) are to indicate the number of EEG segments which is approximately half of the total of the validation set. These numbers are not constant for every training-validation fold because the EEG segments from the baseline category and the alpha-binaural-

beat category are mixed and shuffled randomly before being split into ten non-intersecting sub-sets for the ten-fold cross-validation process.

Table 3.1: Allocated amount of EEG segments for training, validation (trained-model selection), and performance testing for the binary classification on the short-term music experiment data (Cheah *et al.*, 2019a)

EEG Dataset	Training set	Validation set	Test set	Total
Baseline (no music)	✱	✱✱	171 (4.44%)	1839 (47.77%)
Alpha binaural beats	✱	✱✱	179 (4.65%)	2011 (52.23%)
Total	3150 (81.82%)	350 (9.09%)	350 (9.09%)	3850 (100%)

3.1.2.2. EEG Dataset for Short-term Music Experiment Three-class Classification

The three-class classification dataset contains 5,841 two-second EEG segments, among which there is no overlapping with each another. Out of the total 5,841 EEG segments, there are 1,839 baseline eye-open resting EEG segments without listening to music, 1,983 EEG segments recorded while the subjects were listening to self-favourite music, and 2,019 segments of EEG signals collected while the subjects were listening to the audio clip of alpha-binaural beats.

For the task of three-class EEG signal classification, all of the 5,841 non-overlapping segments of EEG signals are divided randomly into training-validation data set and test data set in the proportion of 10:1. The training-validation data set is further split randomly into ten non-intersecting sub-sets for ten-fold cross-validation, as explained in Section 3.1.2.1 above. Table 3.2 presents the above-discussed distribution of EEG data sets for the three-class classification task.

Table 3.2: Allocated amount of EEG segments for training, validation (trained-model selection), and performance testing for the three-class classification on the short-term music experiment data (Cheah *et al.*, 2019a)

EEG Dataset	Training set	Validation set	Test set	Total
Baseline (no music)	*	**	161 (2.76%)	1839 (31.48%)
Own favourite music	*	**	174 (2.98%)	1983 (33.95%)
Alpha binaural beats	*	**	196 (3.35%)	2019 (34.57%)
Total	4779 (81.82%)	531 (9.09%)	531 (9.09%)	5841 (100%)

As in Table 3.1, the double asterisks (*) and triple asterisks (***) in Table 3.2 respectively indicate approximately equivalent values for every fold of cross-validation. The values do not stay constant because of the randomized sub-sets division.

3.1.3. Computing Environment Settings

The computing hardware system used in this study consists of the following hardware components:

- central processing unit : Intel-Core-i5-7300HQ,
- graphical processing unit : NVIDIA-Geforce-GTX-1050 with 4GB dedicated graphic RAM,
- random access memory : 12GB DDR4

Of the above hardware components, the graphical processing unit (GPU) is of particular importance for the training of deep neural network which involves large number of parallel computations. The architectural design of GPU which consists of a mass of parallel computing units can substantially speed up the training process of the deep neural networks.

The software programming environment for this study is set up using the virtual environment management system by Anaconda distribution of Python language. The construction, training, and validation of deep neural networks in this study are performed with *TensorFlow* library. The loading/reading of EEG data and the handling the loaded EEG data in Python environment is achieved with the *MNE-Python* library. The signal processing functions from the *scipy* library are used for the extraction of EEG signal features which are needed as the input data of the support vector machine (SVM) classifier. The machine learning functions in the *Scikit-learn* library are used for

the management of EEG data pools and the construction of SVM classifier. Other important auxiliary Python libraries used in this study include the *numpy* for the management of EEG in the form of array data structure, the *os* library and *re* library for the convenient navigation through the system directory and for the efficient access to the target EEG files.

3.1.4. Methodological Steps of Model Training, Validation, and Testing

Figure 3.2 illustrates the work flow of training-validation cycle for the deep neural networks. The neural network models are trained iteratively using randomized combination of mini batches of EEG segments instead of using the grand sum of the training pool for every training iteration. By splitting the training data pool into mini batches and conducting the model training with these mini batches of training data, significantly less dedicated GPU memory is required. This at the same time allows the construction of neural networks of greater complexity and learning capacity. Furthermore, using randomized combinations of mini batches (instead of going through all the training iterations with the same combination and permutation of training data batches) may reduce the likelihood of the optimization process falling into the trap of local minima of the objective function.

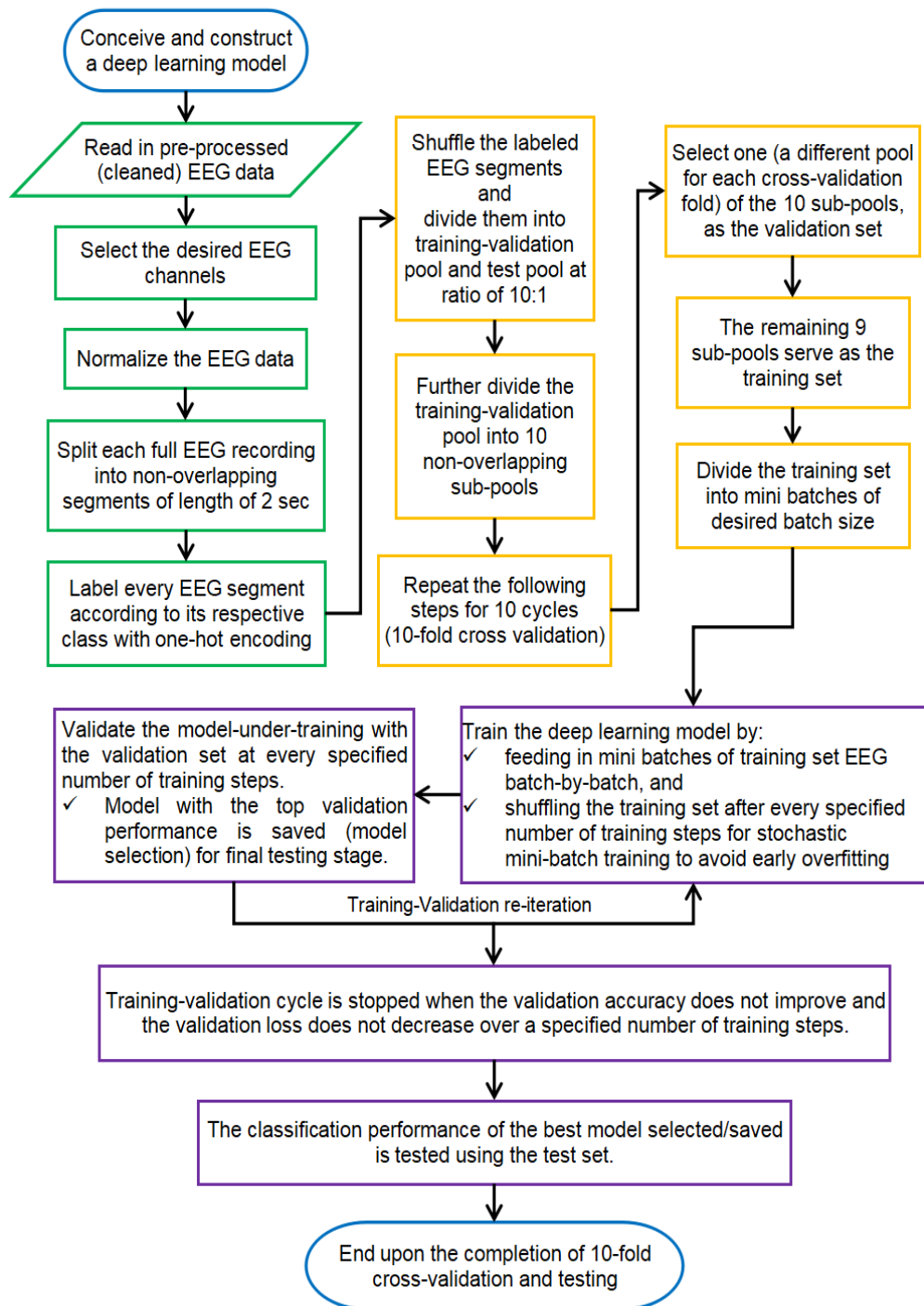


Figure 3.2: Methodological Steps from Model Construction to Model Training, Validation and Testing (Cheah *et al.*, 2019a)

3.1.5. Design of Classifiers

Convolutional neural network (CNN) classifiers of different architectures are constructed for performance comparison. In addition, the performance of the CNN is also compared with that of the SVM, which is one of the most capable machine learning classifiers that are not neural network based.

3.1.5.1. CNN Models

Figure 3.3 and Figure 3.4 illustrate the architectural designs of four CNN classifiers constructed in this study.

Figure 3.3 presents two CNN classifiers with the whole convolutional path containing only temporal (time-dimension) convolution. The CNN model in Figure 3.3(a) has three convolution blocks operating in serial order. Every convolution block is composed of the convolution layer, the rectified linear unit (ReLU) as the activation layer, and the max pooling layer. The shape and size of the convolution kernels and pooling filters are specified accordingly in the figure. Figure 3.3(b) presents a CNN model with relatively greater depth, containing six convolution blocks in serial. For the model in Figure 3.3(b), only the first three convolution blocks contain the max pooling operation. The remaining convolution blocks are constructed with no pooling operation.

On the other hand, Figure 3.4 presents two different CNN classifiers composed of both temporal and spatial dimension convolutions. Spatial dimension convolution refers to the convolution across different EEG channels, as each EEG channel represents a specific location over the scalp. The initial three convolutional blocks in Figure 3.4(a) are identical to that of the CNN models in Figure 3.3. Every of the subsequent convolution blocks of Figure 3.4(a) CNN classifier is composed of two-dimensional (spatial-temporal) convolution kernels.

The CNN classifier in Figure 3.4(b) is composed of nine convolution blocks in serial, with the initial three convolution blocks being identical to that of Figure 3.4(a). The remaining six convolution blocks of Figure 3.4(b) are designed by splitting the two-dimensional spatial-temporal convolution blocks in Figure 3.4(a) into three one-dimensional temporal convolution blocks and three one-dimensional spatial convolution blocks. As a result, the CNN classifier in Figure 3.4(b) is constructed with only one-dimensional convolution kernels (i.e. initial six blocks being purely temporal convolution, while the last three blocks being purely spatial convolution).

The convolutional mechanism can be represented with Equation (1) (Goodfellow [et al.](#), 2016b).

$$Z_k^l = W_k^l * A^{l-1} + b_k^l \quad (1)$$

The capital letters (Z, W, A) symbolize matrices, with Z being a two-dimensional matrix while W and A being three-dimensional matrices. The lowercase letter b symbolizes a scalar value. With the superscript and subscript, W_k^l symbolizes the matrix containing the weights of the k^{th} convolution kernel of the l^{th} convolution block. A^{l-1} symbolizes the three-dimensional aggregation of all the activated feature maps generated from the immediately preceding $(l - 1)^{th}$ convolution block. Similarly, b_k^l symbolizes the scalar bias value of the k^{th} convolution kernel of the l^{th} convolution block. Likewise, Z_k^l is the k^{th} feature map generated by the k^{th} convolution kernel of the l^{th} convolution block operated on the $(l - 1)^{th}$ activated feature maps.

The two-dimensional matrix of a particular feature map (Z_k^l) is derived by moving the convolution kernel (W_k^l) across whole width (X) and height (Y) of the feature maps collection A^{l-1} (Goodfellow *et al.*, 2016c), as in Equation (2).

$$z_k^{l(x,y)} = W_k^l \cdot A^{[l-1](x,y)} + b_k^l \quad (2)$$

The lowercase $z_k^{l(x,y)}$ with the superscripts and subscript now symbolizes a scalar point at (x, y) position within the feature map Z_k^l . Similarly, $A^{[l-1](x,y)}$ symbolizes the sub-matrix of A^{l-1} of the size of the convolution kernel W_k^l centered at position (x, y) of the matrix A^{l-1} .

The collection of all the feature map matrices of the l^{th} layer (denoted as Z^l) will then be activated by the activation function g^l , resulting in the 3D collection of activated feature maps A^l , which is equivalently conveyed in the equation $A^l = g^l(Z^l)$. With k number of independent convolution kernels, the particular convolution layer will generate k number of feature maps.

All of the convolution operations in every CNN models presented in this study are executed using

- stride length of convolution kernels of 1 unit in both the temporal and spatial directions,
- kernel dilation factor of 1 (not dilated) in all directions, and
- the “SAME” padding mode in TensorFlow setting, which maintains the dimensional sizes of the input data and the convolutional output.

For every of the CNN models presented, the 3D feature map output of the last convolution block is dimensionally flattened into a 1D array (a vector) before being passed to the input layer of fully-connected multilayer-perceptron (FC-MLP) network. The network of FC-MLP with two hidden layers of perceptrons (64 nodes and 32 nodes respectively in the first and the second hidden layers) has given the best classification result among the fully connected networks examined in this study.

In contrast to the convolution mechanism, the densely connected FC-MLP network (Goodfellow *et al.*, 2016c) is defined in Equation (3).

$$a_j^l = g_j^l(\sum_{k=1}^K w_{jk}^l a_k^{l-1} + b_j^l) \quad (3)$$

The symbol a_j^l represents the activated output of the j^{th} perceptron of the l^{th} layer of the densely-connected network, while the symbol g_j^l represents the activation function for the j^{th} perceptron. Similarly, a_k^{l-1} represents the activated output of the k^{th} perceptron of the immediately previous layer in the densely-connected network which has a total of K number of perceptrons. The symbol w_{jk}^l represents the weight parameter connecting the j^{th} perceptron of the l^{th} layer to the k^{th} perceptron of the $(l-1)^{th}$ layer. Likewise, b_j^l represents the bias parameter of the j^{th} perceptron of the l^{th} layer.

The output layer of the CNN binary classifier contains two perceptrons activated by the softmax function, while the output layer of the three-class classifier contains three softmax-activated perceptrons. Softmax function (Goodfellow *et al.*, 2016c) computes the probability distributed over a specific number of possible outcomes (two or three outcomes as in this study). Softmax function is defined in Equation (4), with the capital letter C denotes the total number of possible outcomes, while the lower-case i denotes a specific class of outcome.

$$\sigma(y_j) = \frac{e^{y_j}}{\sum_{i=1}^C e^{y_i}}, \text{ for } i = 1, \dots, C. \quad (4)$$

Other computational nodes of the convolution layers and the densely-connected network have ReLU as their activation function. For the output error backpropagation and internal parameter optimization, Adam optimizer (Kingma and Ba, 2015) is used with the softmax cross-entropy loss function as the objective function and the learning rate of 0.001.

Softmax cross-entropy loss is used as the loss function for the CNN internal parameter optimization. Cross-entropy (CE) loss (Good, 1956) is defined in Equation (5).

$$CE(s) = -\sum_j^C t_j \log(s_j) \quad (5)$$

The capital letter C denotes total number of possible classes. The symbol t_j denotes ground-truth score of the j^{th} class, while s_j denotes the predicted score of the j^{th} class estimated by the CNN.

The loss function adopted is the softmax cross-entropy function, $CE(s = \sigma)$, which computes the cross-entropy value using the outputs of softmax activation function as the predicted score s_j , resulting in Equation (6).

$$CE(\sigma(y)) = -\sum_j^C t_j \log(\sigma(y_j)) = -\sum_j^C t_j \log\left(\frac{e^{y_j}}{\sum_i^C e^{y_i}}\right) \quad (6)$$

Using one-hot encoding, the target class (p) has the ground-truth score (t_j) of one ($t_{j=p} = 1$), while all the rest of the non-target classes have ground-truth

score of zero ($t_{j \neq p} = 0$). Therefore, the softmax cross-entropy loss function in equation (6) can be simplified to that in Equation (7).

$$CE(\sigma(y)) = -\log\left(\frac{e^{y_p}}{\sum_i^C e^{y_i}}\right) \quad (7)$$

Generally, the accuracy for a classification task is equal to $\frac{\text{number of correctly predicted cases}}{\text{total number of cases}} \times 100\%$. For binary classification task, the accuracy score will be equal to $\frac{TP+TN}{TP+TN+FP+FN} \times 100\%$, with *TP* denoting *true positives*, *TN* denoting *true negatives*, *FP* denoting *false positives*, and *FN* denoting *false negatives*. For any classification task of even higher number of classes, the accuracy score of the classifier will generally be $\frac{\text{the sum of cases along the main diagonal of confusion matrix}}{\text{total cases in the confusion matrix}} \times 100\%$.

Overfitting of CNN model to the training data is a common problem during the model optimization process. The overfitted model will have reduced accuracy at representing the general population. To avoid the overfitting of CNN models to the training data, dropout mechanism (Srivastava *et al.*, 2014) is implemented during the model training process serving as the model regularization technique. Dropout mechanism with the rate of 0.4 (40%) is implemented empirically during the parameter-optimization process of CNN model. Dropout mechanism blocks the stipulated percentage of perceptrons (with different randomized combination of perceptron at every training iteration) from being involved in the forward predictive computation as well as the

backpropagation of prediction error. Consequently, under the dropout regularization, every training iteration will have a portion of the full network composed of different perceptron connections serving as the predicting model.

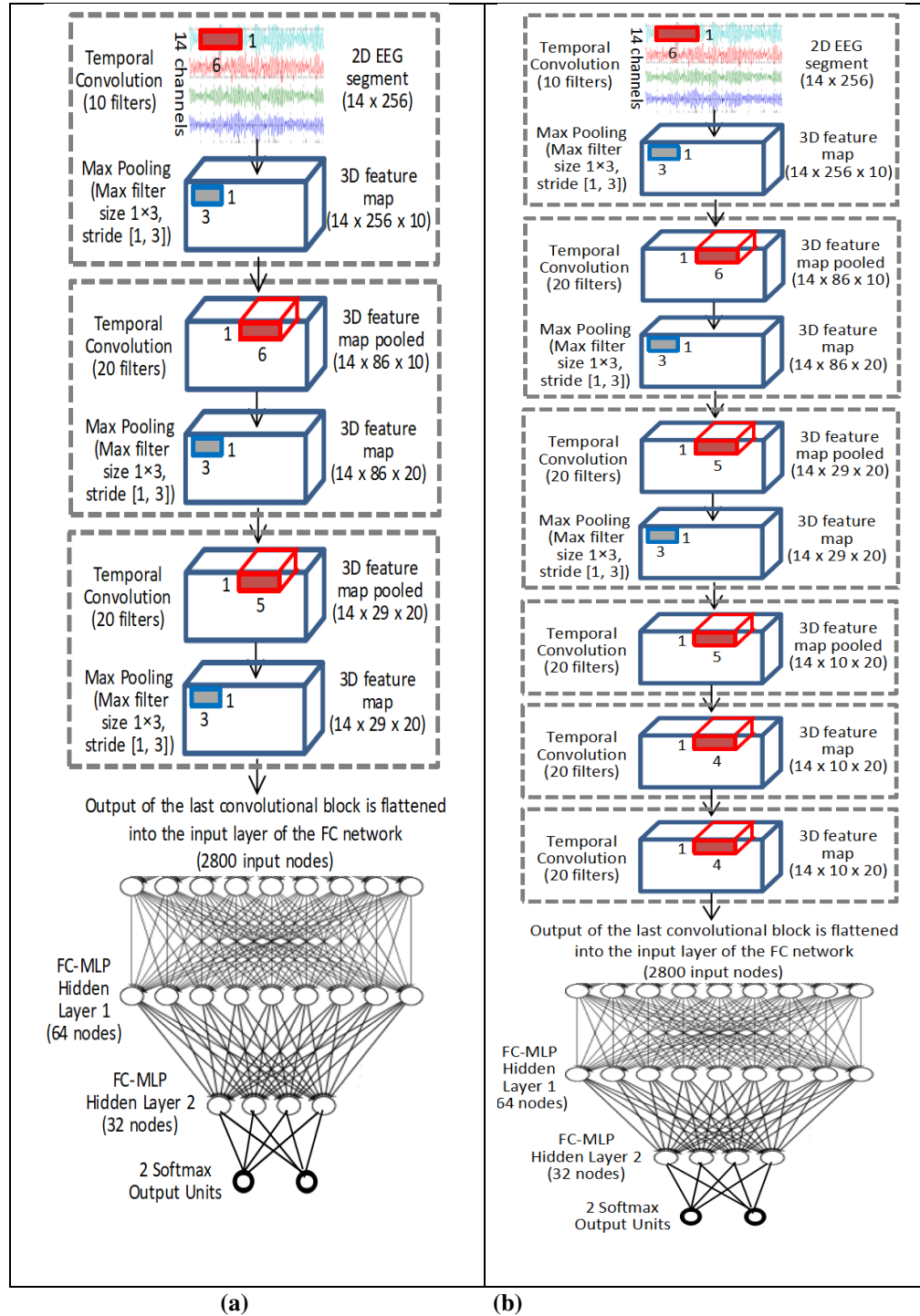


Figure 3.3: Architectures of Temporal-dimension CNN with no Spatial-dimension Convolution (Cheah *et al.*, 2019a)

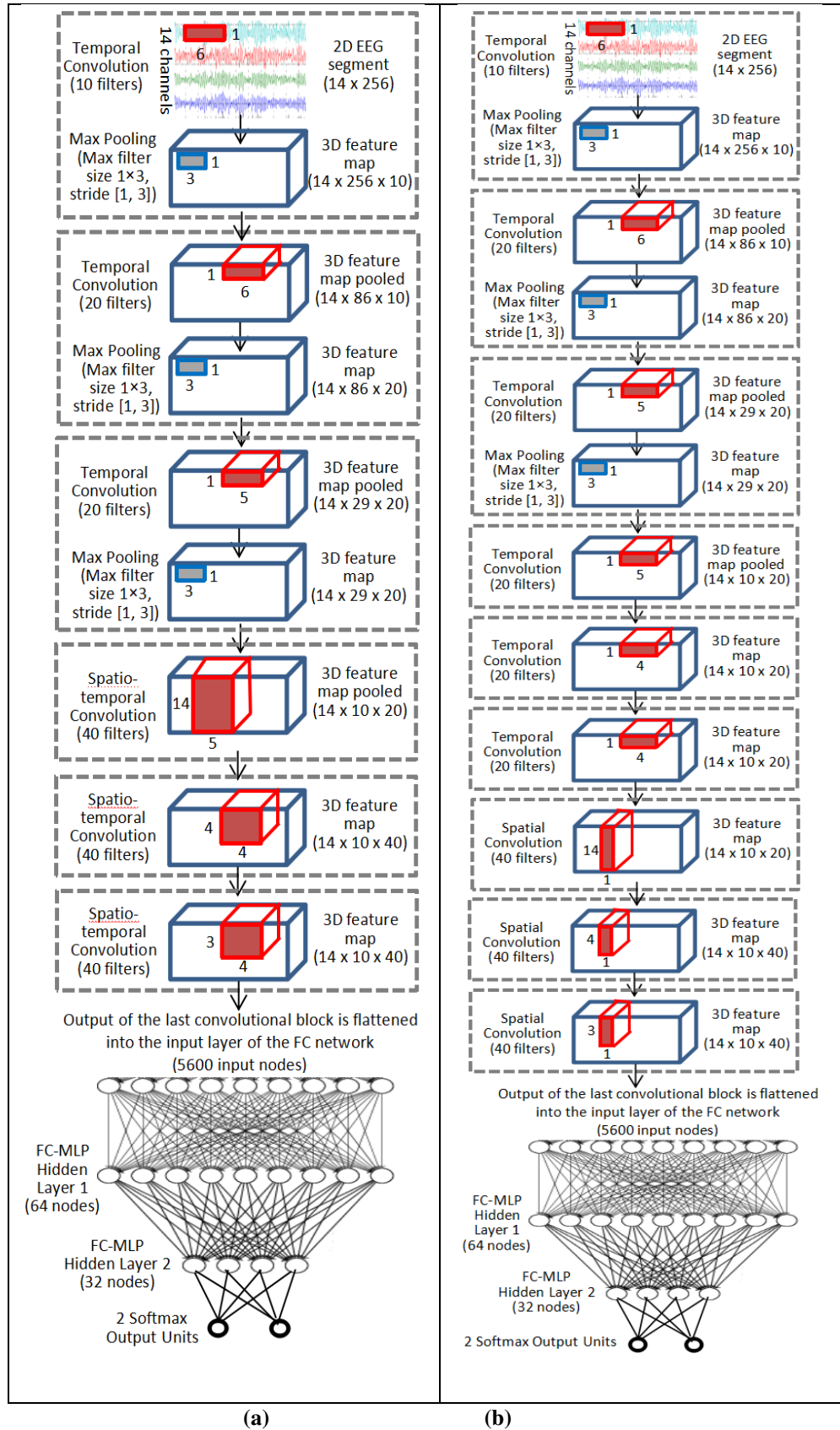


Figure 3.4: Architectures of Spatial-Temporal CNN constructed with (a) 2-dimensional kernels and (b) only 1-dimensional kernels (Cheah *et al.*, 2019a)

3.1.5.2. SVM Classifier

The support vector machine (SVM) classifiers ([Hsu et al., 2016](#)) have also been constructed in this study for the purpose of performance comparison with the CNN classifiers. The SVM classifiers are designed and executed with the `sklearn.svm.svc` package from the popular *Python* machine learning library *Scikit-learn*. The `sklearn.svm.svc` package is written with the technical basis of LIBSVM ([Chang and Lin, 2011](#)).

The kernel used in the SVM classifier in this work is the radial basis function (RBF) kernel. The RBF kernel fitting parameters C and gamma, γ , are respectively assigned with the values of 10^6 and 10^{-5} . Since there is no definitive direction for estimating suitable values of C and γ for every different problem, according to practical guides such as the [Hsu et al. \(2016\)](#), the values of those parameters are decided through grid searching.

A total of 161 EEG features are passed into the SVM classifier. These features are

- the powers of EEG signal of 4 different frequency bands (namely the delta band: 1-4 Hz, theta band: 4-8 Hz, alpha band: 8-13 Hz, and beta band: 13-30 Hz) of every EEG channel,
- the peak-power frequency of each of the four frequency bands for all channels,

- the approximate entropy (ApEn) and the sample entropy (SampEn) of each EEG channel, and
- the real value, the imaginary value, and the absolute value of cross power spectral density (CPSD) computed from the pairs of corresponding left and right hemispheric channels.

Prior to the extraction of the EEG features, the full-length EEG signals are split into non-overlapping signals of two-second length. Table 3.3 presents the number of features in every feature category. These EEG features are computed from every two-second EEG segment.

Table 3.3: The Manually Extracted Features from EEG Signals for the Training of SVM Classifier (Cheah *et al.*, 2019a)

Measure	Features	Number of sub-features
Power spectrum	Band power	$4 \text{ bands} \times 14 \text{ channels} = 56$
	Peak band power frequency	$4 \text{ bands} \times 14 \text{ channels} = 56$
Signal regularity and complexity	Approximate entropy	$14 \text{ channels} = 14$
	Sample entropy	$14 \text{ channels} = 14$
Cross spectral analysis	Left-right-EEG-channel cross power spectral density	$3 \times (14 \text{ channels} / 2) = 21$
Total number of sub-features :		161

3.2 CNN for Personalized Emotion Classification (Study 2)

Low capacity CNN models with low number of convolutional kernels are constructed in this study for personalized emotion classification. The emotion classification task in this study is based on the DEAP dataset collectively prepared by the researchers from Queen Mary University of London, University of Twente, University of Geneva, and the EPFL research institute in Switzerland (Koelstra *et al.*, 2012).

3.2.1 DEAP Dataset

DEAP dataset used in this study is a multimodal dataset for emotion recognition containing the EEG, peripheral physiological signals, and facial video recordings. The EEG and physiological signals were obtained from thirty-two healthy participants with the age range of 19-37 years, while each of the participants were watching and listening to forty different musical video excerpts. Every video has the length of 60 seconds.

In line with the project objectives, only the EEG signal is used in this emotion recognition study. The EEG signals of DEAP dataset are obtained from thirty-two recording electrodes, as in Figure 3.5. The placement of the recording electrodes over the scalp follows the format of the International 10-20 system. The EEG signals were recorded at sampling frequency of 512 Hz, which was then down-sampled to 128 Hz during the signal pre-processing. Also, the

electrooculography-related artifacts were removed and the EEG signals were filtered with bandpass frequency of 4.0 to 45.0 Hz. Each EEG signal has the length of 63 seconds, which includes baseline recording of three seconds preceding the video length of sixty seconds.

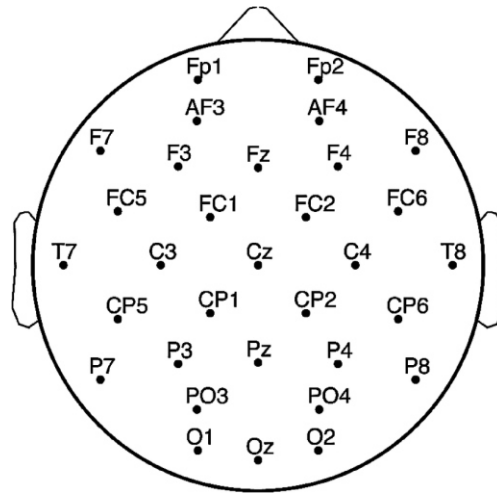


Figure 3.5: Layout of the 32 EEG Electrodes in DEAP Dataset (Yazdani *et al.*, 2012)

Corresponding to each EEG recording, the participants had subjectively rated their own emotional aspects (i.e. emotional valence, emotional arousal, emotional dominance, and liking) experienced by themselves while watching the particular video excerpt. The subjective ratings of the emotions were made using the self-assessment manikin (SAM) (Bradley and Lang, 1994) with the scoring scale ranging from 1 to 9. In terms of the valence SAM scale, the lower scores correspond to the negative emotions (unhappy/sad) while the higher scores correspond to the positive emotions (happy/joyful). On the other hand, the arousal SAM scale which also ranges from 1 to 9, implies how heightened

the valence element is. Figure 3.6 shows the SAM scaling for emotional valence and arousal rating.

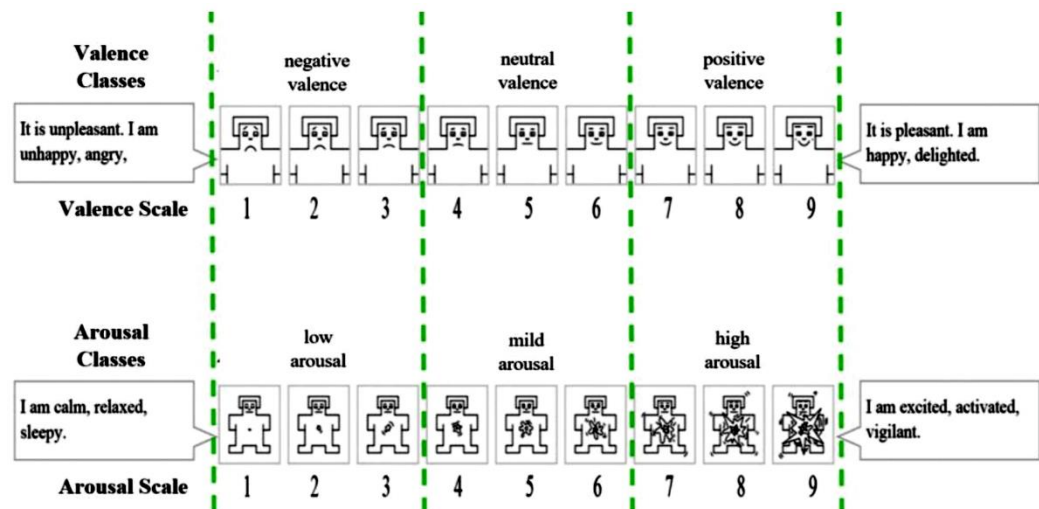


Figure 3.6: Emotional Valence and Arousal Scaling of SAM (Bartosova et al., 2019) and the Three Classes used in this Study

3.2.2 Emotion-Class Relabeling

In this study, each of the emotional aspects (i.e. valence and arousal) is re-labeled into three discrete classes based on the original continuous values between 1 and 9. For the emotional valence, SAM scaling values between 1 and 4 is labeled as the negative valence class. The neutral valence class takes the SAM scaling values between 4 and 6, while the positive valence class consists of the values between 6 and 9.

Correspondingly, for the emotional arousal aspect, the continuous SAM ratings between 1 and 4 are categorized as the low arousal status, SAM ratings between 4 and 6 categorized as the mild arousal status, while the ratings from between 6 and 9 as the high arousal mental state. The emotional valence and arousal classes are indicated by the green dotted lines in Figure 3.6 above.

3.2.3 EEG Allocation into Training, Validation, and Test Set

For personalized emotion recognition, the CNN classifier is optimized and validated subject-by-subject. For every EEG recording, the initial thirteen seconds of the total signal length of sixty-three seconds are discarded. It is because the 1st three seconds are the baseline resting EEG recorded before the start of the video excerpt and the subsequent ten-second segment is the duration that is allowed for the emotion induced by watching the musical video to set in.

The remaining fifty-second EEG signal is divided into five non-overlapping segments, with each segment of the length of ten seconds. The last ten seconds of the signals are used as the test dataset. The other four non-overlapping ten-second signal segments are used in the four-fold cross-validation training of the CNN classifier. Each validation fold has one of the ten-second signal segments as the validation dataset and the remaining signal segments as the training dataset. This has ensured that the signal samples in the training, validation and testing dataset have no degree of overlap.

For each of the ten-second segments, the sliding window with window length of one-second and sliding step overlap of approximately 90% (0.90625 seconds) is used to further generate EEG sub-segments of one second duration. This produces higher number of training examples covering greater signal variation for the classifiers to learn from for picking up key features. Nevertheless, none of the one-second EEG segments in the test dataset or the validation dataset has any degree of overlapping with the segments in the training dataset. Similarly, the EEG segments in the test dataset have not overlapped with the segments in the validation dataset at all.

With the above-described extraction operation of sliding window, each ten-second EEG segment gives off ninety-seven sub-segments of one-second duration, where $(1280 - 128) \div (128 - 116) + 1 = 97$ sub-segments. This number of EEG sub-segments generated by the sliding-window operation follows the pattern in Equation (8).

$$\begin{aligned} & \text{number of segments generated by sliding window} \\ & = \frac{(\text{total segment length} - \text{sliding window length})}{(\text{sliding window length} - \text{overlap length})} + 1 \end{aligned} \quad (8)$$

Therefore, there are a total of 11,640 one-second EEG segments in the training dataset ($40 \text{ videos} \times 3 \text{ ten-second segments/video} \times 97 \text{ sub-segments/ten-second segments}$). Validation dataset and test dataset each has 3,880 one-second EEG segments ($40 \text{ videos} \times 1 \text{ ten-second segment/video} \times 97 \text{ sub-segments/ten-second segment}$).

Figure 3.7 demonstrates the segmentation of the sixty-three-second full-length EEG signal into training dataset, validation dataset and test dataset for 4-fold cross validation. Subsequently, Figure 3.8 shows the operation of the sliding window for incrementing the number and thus augmenting the variation of EEG segments to be presented to the CNN classifiers during the training process.

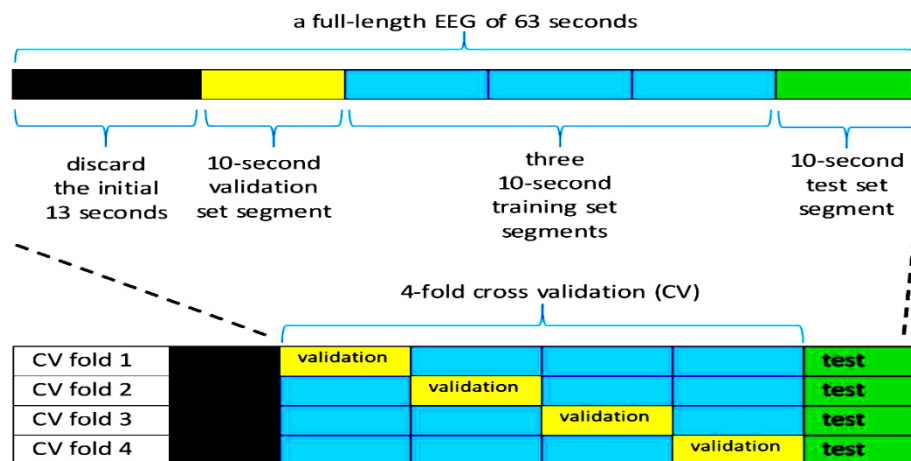


Figure 3.7: Segmentation and Allocation of the Full-length EEG Recording into Training, Validation and Test Datasets for Four-fold Cross-validation (Cheah *et al.*, 2019b)

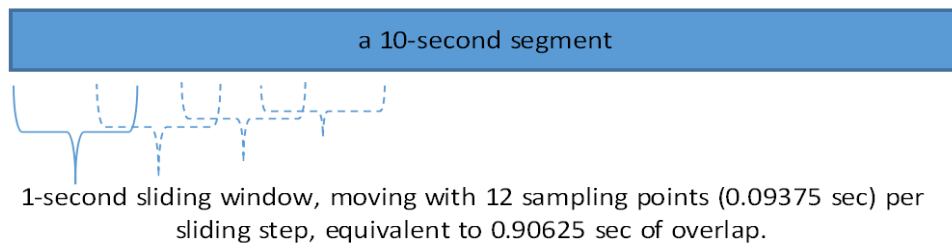


Figure 3.8: Illustration of the sliding window approach for the extraction of overlapping one-second sub-segments from every ten-second EEG segment (Cheah *et al.*, 2019b)

3.2.4 Architectural Details of the CNN

The two CNN models trained and validated in this study are a single-path CNN classifier and a double-path CNN classifier with dilated convolution. The information of the architecture of the two CNN classifiers are documented in Table 3.4 and Table 3.5.

The CNN classifiers read in the thirty-two channels of normalized EEG segment and perform the emotion classification with no requirement for manually extracted features from EEG signals. The single-path CNN classifier contains eight successive convolution layers, which are composed of five temporal convolution layers and three spatial convolution layers. The feature maps output by the last spatial convolution layer is passed to the FC-MLP network. The double-path CNN classifier is designed with two convolution pathways operating in parallel. Each of the two parallel convolution paths has eight successive convolution layers, which are five temporal convolution layers followed by three spatial convolution layers. The two parallel convolution paths each respectively has the temporal convolution operating at dilation factor of 1 (without dilation) and 2. The feature maps output by the final convolution layers of both parallel paths are flattened and concatenated before being fed into the FC-MLP network.

The FC-MLP networks of both CNN classifiers are each composed of 3 hidden fully-connected layers. The first hidden fully-connected layer contains

ninety-six perceptrons which read in and process the flattened feature maps generated by the final layer of convolution operation. The second hidden fully-connected layer contains thirty-two perceptrons, connecting the first and the third fully-connected hidden layer which contains sixteen perceptrons. The third fully-connected layer is in turn connected to the final output layer of the FC-MLP network of the CNN classifier. The three output nodes of the FC-MLP have softmax function as their activation function.

All of the hidden fully-connected layers have the ReLU as their activation function. The temporal convolutions are performed using the “*SAME*” Tensorflow padding method which maintains length of the temporal dimension. Meanwhile, the spatial convolutions are performed with the “*VALID*” padding method which is essentially striding the convolution kernel without padding the particular dimension. The first two temporal convolution layers are the only sections constructed with pooling operation, with max pooling on the feature maps before passing onto the subsequent layer.

Table 3.4: Details of the Architecture of the Single-path CNN Model (Cheah *et al.*, 2019b)

Single-path Convolution (reading in plain EEG data of size 32x128)			
Layer (type)	Filter size (dilation factor) // stride // padding // # of filters	Output size of feature map	# of trainable parameters
Conv_A1 (1D temporal conv)	1 x 8 (1) // 1,1 // SAME // 6	32 x 128 x 6	54
Pooling_A1 (1D max pooling)	1 x 4 // 1,4 // SAME // 1	32 x 32 x 6	0
Conv_A2 (1D temporal conv)	1 x 6 (1) // 1,1 // SAME // 6	32 x 32 x 6	222
Pooling_A2 (1D max pooling)	1 x 2 // 1,2 // SAME // 1	32 x 16 x 6	0
Conv_A3 (1D temporal conv)	1 x 4 (1) // 1,1 // SAME // 6	32 x 16 x 6	150
Conv_A4 (1D temporal conv)	1 x 3 (1) // 1,1 // SAME // 6	32 x 16 x 6	114
Conv_A5 (1D temporal conv)	1 x 3 (1) // 1,1 // SAME // 6	32 x 16 x 6	114
Conv_A6 (1D spatial conv)	13 x 1 (1) // 1,1 // VALID // 6	20 x 16 x 6	474
Conv_A7 (1D spatial conv)	11 x 1 (1) // 1,1 // VALID // 6	10 x 16 x 6	402
Conv_A8 (1D spatial conv)	8 x 1 (1) // 1,1 // VALID // 6 (40% dropout)	3 x 16 x 6	294
FC-MLP (reading in the flattened final feature maps of the single-path convolution)			
Layer (type)	Number of nodes // Activation function // Dropout rate	Output size	# of trainable parameters
FC hidden layer 1 (dense)	96 // ReLU // 40%	96	27744
FC hidden layer 2 (dense)	32 // ReLU // 40%	32	3104
FC hidden layer 3 (dense)	16 // ReLU // 40%	16	528
Prediction output layer (dense)	3 // Softmax // 0%	3	50

Table 3.5: Details of the Architecture of the Double-path CNN Model with Dilated Convolution in One of the Two Parallel Convolution Paths (Cheah *et al.*, 2019b)

Convolutional Path A (reading in EEG data dimension of 32x128; running parallel with Path B)			
Layer (type)	Filter size (dilation factor) // stride // padding // # of filters	Output size of feature map	# of trainable parameters
Conv_A1 (1D temporal conv)	1 x 8 (1) // 1,1 // SAME // 3	32 x 128 x 3	27
Pooling_A1 (1D max pooling)	1 x 4 // 1,4 // SAME // 1	32 x 32 x 3	0
Conv_A2 (1D temporal conv)	1 x 6 (1) // 1,1 // SAME // 3	32 x 32 x 3	57
Pooling_A2 (1D max pooling)	1 x 2 // 1,2 // SAME // 1	32 x 16 x 3	0
Conv_A3 (1D temporal conv)	1 x 4 (1) // 1,1 // SAME // 3	32 x 16 x 3	39
Conv_A4 (1D temporal conv)	1 x 3 (1) // 1,1 // SAME // 3	32 x 16 x 3	30
Conv_A5 (1D temporal conv)	1 x 3 (1) // 1,1 // SAME // 3	32 x 16 x 3	30
Conv_A6 (1D spatial conv)	13 x 1 (1) // 1,1 // VALID // 3	20 x 16 x 3	120
Conv_A7 (1D spatial conv)	11 x 1 (1) // 1,1 // VALID // 3	10 x 16 x 3	102
Conv_A8 (1D spatial conv)	8 x 1 (1) // 1,1 // VALID // 3 (40% dropout)	3 x 16 x 3	75
Convolutional Path B (reading in EEG data dimension of 32x128; running parallel with Path A)			
Layer (type)	Filter size (dilation factor) // stride // padding // # of filters	Output size of feature map	# of trainable parameters
Conv_B1 (1D temporal conv)	1 x 8 (2) // 1,1 // SAME // 3	32 x 128 x 3	27
Pooling_B1 (1D max pooling)	1 x 4 // 1,4 // SAME // 1	32 x 32 x 3	0
Conv_B2 (1D temporal conv)	1 x 6 (2) // 1,1 // SAME // 3	32 x 32 x 3	57
Pooling_B2 (1D max pooling)	1 x 2 // 1,2 // SAME // 1	32 x 16 x 3	0
Conv_B3 (1D temporal conv)	1 x 4 (2) // 1,1 // SAME // 3	32 x 16 x 3	39
Conv_B4 (1D temporal conv)	1 x 3 (2) // 1,1 // SAME // 3	32 x 16 x 3	30
Conv_B5 (1D temporal conv)	1 x 3 (2) // 1,1 // SAME // 3	32 x 16 x 3	30
Conv_B6 (1D spatial conv)	13 x 1 (1) // 1,1 // VALID // 3	20 x 16 x 3	120
Conv_B7 (1D spatial conv)	11 x 1 (1) // 1,1 // VALID // 3	10 x 16 x 3	102
Conv_B8 (1D spatial conv)	8 x 1 (1) // 1,1 // VALID // 3 (40% dropout)	3 x 16 x 3	75
FC-MLP (reading in the flattened final feature maps of both convolutional paths A and B)			
Layer (type)	Number of nodes // Activation function // Dropout rate	Output size	# of trainable parameters
FC hidden layer 1 (dense)	96 // ReLU // 40%	96	27744
FC hidden layer 2 (dense)	32 // ReLU // 40%	32	3104
FC hidden layer 3 (dense)	16 // ReLU // 40%	16	528
Prediction output layer (dense)	3 // Softmax // 0%	3	50

3.2.5 Dilated Convolution

In the double-path CNN classifier, one of the two convolution paths is implemented completely with dilated convolution with the dilation factor of 2. **Yu and Koltun (2016)** had mathematically described the dilated convolution as simply the execution of the usual discrete convolution mechanism with the dilated convolution operator, as in Equation (10).

For \mathbb{Z} symbolizes the integer-number set and \mathbb{R} symbolizes the real-number set, the two-dimensional dilated convolution can be represented mathematically as below.

Let F be a discrete function such that $F : \mathbb{Z}^2 \rightarrow \mathbb{Z}$, and k be a discrete filter such that $k : \Omega_r \rightarrow \mathbb{R}$ of size $(2r+1)^2$ with $\Omega_r = [-r, r]^2 \cap \mathbb{Z}^2$. The usual discrete convolution operator $*$ can be expressed as in Equation (9) (**Yu and Koltun, 2016**).

$$(F * k)(p) = \sum_{s+t=p} F(s) \cdot k(t) \quad (9)$$

The above-mentioned discrete convolution operator $*$ can be extrapolated into the discrete dilated convolution operator $*_l$, where l specifies the value of dilation factor, as described in Equation (10) (**Yu and Koltun, 2016**).

$$(F *_l k)(p) = \sum_{s+l=t=p} F(s) \cdot k(t) \quad (10)$$

The above concept of dilated convolution applies to the one-dimensional convolution in the CNN classifier in this study, in which case, the filter size corresponds to single dimension $(2r+1)$ instead of $(2r+1)^2$. All the superscripts that denote the two-dimensional vector space is set to one (which is solely along the time dimension).

3.2.6 Training, Validation, and Testing of the CNN Models

The objective function for model parameter optimization is the cross-entropy loss of softmax outputs. Adam optimizer (Kingma and Ba, 2015) is used as the optimization method for minimizing the loss function. The optimization progress is executed at learning rate of 0.004.

To serve the purpose of model regularization, the CNN classifiers in this study are trained using the dropout technique with the dropout rate of 40%. The dropout technique is applied to the output of the last convolution layer and all the hidden FC-MLP layers.

The model training process is conducted with mini batches with 800 EEG segments per training-iteration. The training dataset is randomly shuffled

after every three complete training epochs. There are 15 mini-batch training iterations for a complete epoch.

Along the progress of the model training stage, the classification performance of the model is validated using the validation dataset. While the training dataset plays a role in optimization of the model's parameters, the classification performance using the validation dataset serves to guide the model selection process. The model under iterative training with highest validation accuracy will be selected for the final performance testing using the test dataset.

3.3 Residual Network and VGG for Emotion EEG Classification (Study 3)

3.3.1 SEED Dataset

The SEED dataset (Duan *et al.*, 2013; Zheng and Lu, 2015) is a publicly available emotion related EEG dataset for research purpose provided by the Shanghai Jiao Tong University (SJTU). The stimuli in the SEED experiment were 15 film clips carefully chosen such that each elicits a single desired target emotion. Each film clip lasts about 4 minutes and is coherent to either positive, neutral, or negative valence emotion as described in Table 3.6.

Table 3.6: Film Clips in SEED Dataset

Source Film Name	Emotion	Number of Clips
Tangshan Earthquake	Negative	2
Back to 1942	Negative	3
Lost in Thailand	Positive	2
Flirting Scholar	Positive	1
Just Another Pandora's Box	Positive	2
World Heritage in China	Neutral	5

SEED experiment had 15 participants. Every participant underwent 3 sessions of experiment, with at least one week interval in between every 2

sessions. Each experiment session contained 15 trials, each playing one of the 15 film clips followed by self-assessment and a short rest. The play sequence of the film clips was arranged such that no two consecutive trials carried the clips of the same emotion category. Figure 3.9 shows the structure of the experiment session.

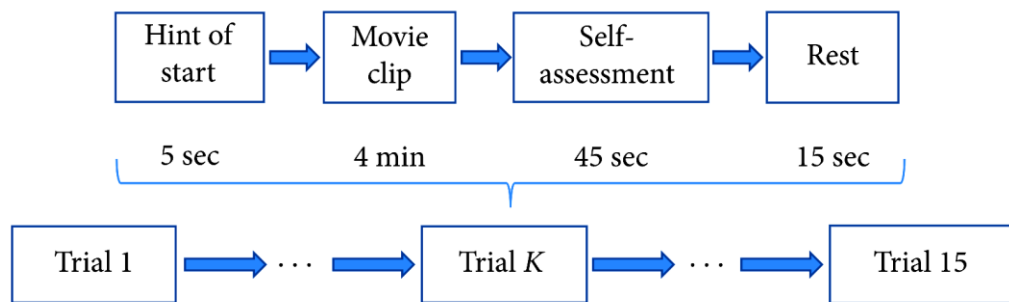


Figure 3.9: Session Flow of Data Collection Process of SEED Experiment (Zheng and Lu, 2015)

The EEG signals were recorded with 62 active AgCl electrodes of the ESI NeuroScan System at sampling frequency of 1000 Hz. The electrode placement was based on the international 10-20 system as shown in Figure 3.10. The recorded EEG signals were then downsampled to 200 Hz and bandpass frequency filter of 0.5 Hz to 70 Hz was applied.

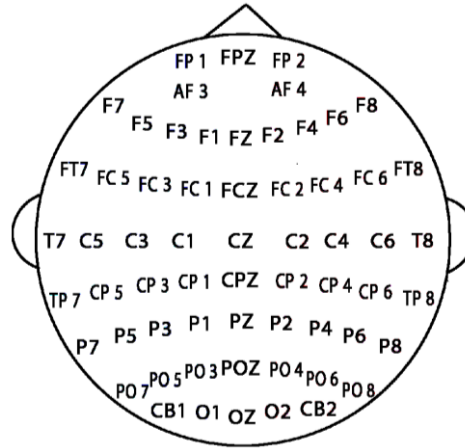


Figure 3.10: Placement Layout of EEG Channels in SEED Experiment (Zheng and Lu, 2015)

3.3.2 EEG Data Preprocessing

As the target emotion caused by watching the film clip would not likely be successfully induced immediately at the start of film clip, we have set a buffering period of 90 seconds for the emotion establishment. Therefore, the initial 90 seconds of each of the 4-minute EEG trial were discarded. The remaining EEG recording is split into 2-second non-overlapping segments, with each EEG segment assuming the length of 400 sampling points for the sampling frequency of 200Hz. Table 3.7 presents the emotion labelling of each video clips and the number of two-second EEG sub-segments generated from each of the EEG recording corresponding to the particular video clip.

Each of the non-overlapping segments is then normalized along the time axis respectively. All the generated EEG segments are split into five sub-pools for 5-fold cross validation of the model performance.

Table 3.7: Emotion Class Labelling of the Video Clips in the SEED Dataset and the EEG Recording Segmented for Study 3

Emotion Class	Trial (Video Clip) Number	EEG Recording Length Retrieved (sampling points)	Number of 2-second EEG subsegments
negative	3	23200	58
	4	29600	74
	7	29200	73
	12	28400	71
	15	23200	58
neutral	2	28400	71
	5	18800	47
	8	25200	63
	11	28800	72
	13	28800	72
positive	1	28800	72
	6	20800	52
	9	34800	87
	10	29200	73
	14	29600	74

3.3.3 Optimizing *ResNet* & *VGG* for EEG Signals

Figure 3.11 and Figure 3.12 respectively illustrate the architectural details of different versions of *ResNet18* and *VGG16* examined in this study.

3.3.3.1. *ResNet* Optimization

The original architecture of *ResNet18* (He *et al.*, 2016) consisting of 17 convolutional layers and 1 layer of fully-connected network is depicted in Figure 4(a). As the original *ResNet18* is designed for image processing, the convolutional kernels within the model are all 2-dimensional kernels. It has 3-by-3 kernels throughout its convolutional path, except for the very first convolutional layer (*Conv 0*) which has 7-by-7 kernels.

The colour coding of Figure 3 denotes the major convolutional blocks of the *ResNet*. The convolutional layers of the same colour has the same number of kernels (e.g. orange for 64 kernels, yellow for 128 kernels, green for 256 kernels, and blue for 512 kernels). The darker colour layers are convolutional layers, while the lighter layers are the other functional layers in the block, such as the batch normalization (BN) function, the Rectified Linear Unit (ReLU) activation function, the summation (Sum) of the by-passed feature map and the main convolution feature map, and the adaptive average pooling (*AvgPool*). The adaptive *AvgPool* layer before the fully connected (FC) layer allows the model to process EEG signals of different numbers of channels without the need to reassign the number of connections in the FC network.

The last layer of the *ResNet18* is a single layer of fully connected (FC) network with three output nodes, corresponding to the three emotion classes.

There are two types of bypass connection in the *ResNet*, i.e. the identity bypass and the downsampling bypass. The identity bypass has its feature map being passed on, skipping two convolutional layers without any further processing before the summation function. The downsampling bypass happens at the initial stage of every major convolutional block, where the input feature maps will have their map size reduced due to kernel stride and the number of feature maps will increase due to the increment of convolutional kernels. Therefore, the downsampling bypass is necessary in order to have the dimension

of the shortcut data matching the data dimension of the main convolutional path. While the identity bypass performs no additional processing on the data passed onwards, the downsampling bypass has 1-by-1 convolutional kernels which introduce an additional small number of trainable parameters as reported in Figure 3.11.

In this study, three variants of the original *ResNet18* were constructed and investigated. Two of the three *ResNet18* variants are illustrated in Figure 3.11(b) and 3.11(c). The 2D kernels of the *ResNet* were all restructured into 1D kernels along either the temporal(time)-dimension or the spatial(channel)-dimension.

The variant in Figure 3.11(b) has alternating temporal and spatial-dimension convolution. [Eckart and Young \(1936\)](#) and [Maji and Mullins \(2018\)](#) have reported that the matrix such as the convolution filters can be well approximated with an arbitrary number of lower rank matrices. [Maji and Mullins \(2018\)](#) had also demonstrated the feasibility of separating the 2D kernels of the well-established CNNs (e.g. *AlexNet*, *VGG-16*, *Inception-v1*, *ResNet-152*) into alternating 1D vertical and horizontal kernels, achieving near baseline accuracy for image classification with significant speedup of training.

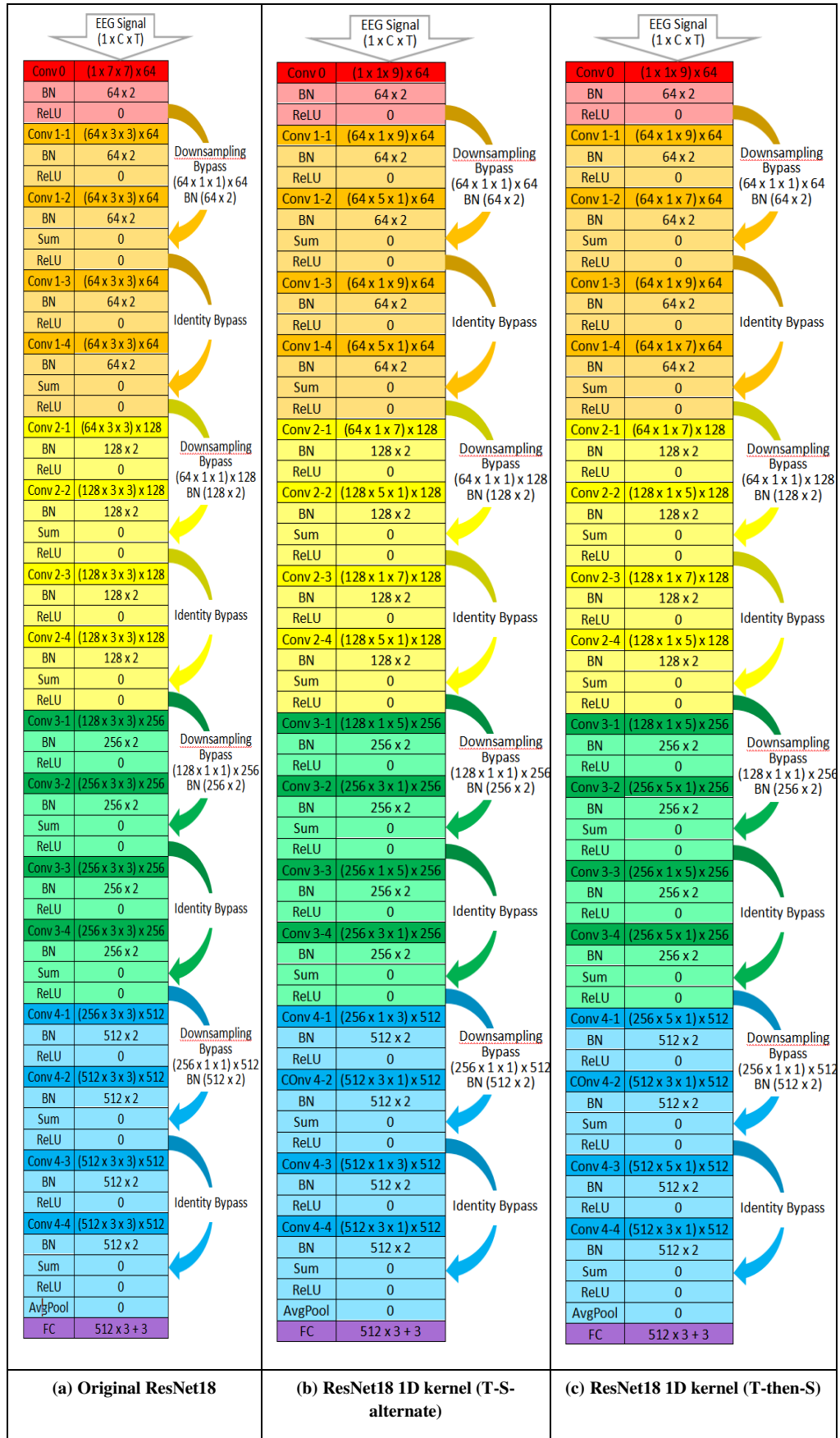


Figure 3.11: Architectural details of (a) the original ResNet18 and its modified variants (b, c) for EEG signal processing.

EEG Signal (1 x C x T)		EEG Signal (1 x C x T)		EEG Signal (1 x C x T)	
Conv 1-1	(1 x 1 x 7) x 64	Conv 1-1	(1 x 1 x 7) x 64	Conv 1-1	(1 x 1 x 7) x 64
BN	64 x 2	BN	64 x 2	ReLU	0
ReLU	0	ReLU	0	Conv 1-2	(64 x 1 x 7) x 64
Conv 1-2	(64 x 1 x 7) x 64	Conv 1-2	(64 x 1 x 7) x 64	BN	64 x 2
BN	64 x 2	ReLU	0	MaxPool	0
ReLU	0	MaxPool	0	Conv 2-1	(64 x 1 x 7) x 128
MaxPool	0	Conv 2-1	(64 x 1 x 7) x 128	BN	128 x 2
Conv 2-1	(64 x 1 x 7) x 128	BN	128 x 2	ReLU	0
BN	128 x 2	ReLU	0	Conv 2-2	(128 x 1 x 7) x 128
ReLU	0	Conv 2-2	(128 x 1 x 7) x 128	BN	128 x 2
Conv 2-2	(128 x 1 x 7) x 128	BN	128 x 2	ReLU	0
BN	128 x 2	ReLU	0	MaxPool	0
ReLU	0	MaxPool	0	Conv 3-1	(128 x 1 x 7) x 256
MaxPool	0	Conv 3-1	(128 x 1 x 7) x 256	BN	256 x 2
Conv 3-1	(128 x 1 x 7) x 256	ReLU	0	ReLU	0
BN	256 x 2	Conv 3-2	(256 x 1 x 5) x 256	BN	256 x 2
ReLU	0	BN	256 x 2	ReLU	0
Conv 3-2	(256 x 1 x 5) x 256	ReLU	0	MaxPool	0
BN	256 x 2	Conv 3-3	(256 x 1 x 5) x 256	BN	256 x 2
ReLU	0	BN	256 x 2	ReLU	0
MaxPool	0	ReLU	0	MaxPool	0
Conv 3-3	(256 x 1 x 5) x 256	Conv 3-3	(256 x 1 x 5) x 256	Conv 4-1	(256 x 1 x 5) x 512
BN	256 x 2	BN	256 x 2	BN	512 x 2
ReLU	0	ReLU	0	ReLU	0
MaxPool	0	MaxPool	0	Conv 4-2	(512 x 1 x 5) x 512
Conv 4-1	(256 x 1 x 5) x 512	Conv 4-1	(256 x 1 x 5) x 512	BN	512 x 2
BN	512 x 2	BN	512 x 2	ReLU	0
ReLU	0	ReLU	0	MaxPool	0
Conv 4-2	(512 x 1 x 5) x 512	Conv 4-2	(512 x 1 x 5) x 512	Conv 4-3	(512 x 3 x 1) x 512
BN	512 x 2	BN	512 x 2	BN	512 x 2
ReLU	0	ReLU	0	ReLU	0
Conv 4-3	(512 x 3 x 1) x 512	MaxPool	0	Conv 5-1	(512 x 3 x 1) x 512
BN	512 x 2	Conv 5-1	(512 x 3 x 1) x 512	BN	512 x 2
ReLU	0	BN	512 x 2	ReLU	0
MaxPool	0	ReLU	0	Conv 5-2	(512 x 3 x 1) x 512
Conv 5-1	(512 x 3 x 1) x 512	Conv 5-2	(512 x 3 x 1) x 512	BN	512 x 2
BN	512 x 2	BN	512 x 2	ReLU	0
ReLU	0	ReLU	0	Conv 5-3	(512 x 3 x 1) x 512
Conv 5-2	(512 x 3 x 1) x 512	Conv 5-3	(512 x 3 x 1) x 512	BN	512 x 2
BN	512 x 2	BN	512 x 2	ReLU	0
ReLU	0	ReLU	0	AvgPool	0
Conv 5-3	(512 x 3 x 1) x 512	AvgPool	0	FC 1	512 x 3 + 3
BN	512 x 2	FC 1	512 x 3 + 3		
ReLU	0				
AvgPool	0				
FC 1	512 x 4096 + 4096				
FC 2	4096 x 4096 + 4096				
FC 3	512 x 3 + 3				
(a) VGG16 1D kernel		(b) VGG14 1D kernel		(c) VGG14 1D kernel (no batch norm)	

Figure 3.12: Architectural details of (a) VGG16 with 1D kernels, (b) VGG14 with 1D kernels and (c) VGG14 without batch normalization.

Nevertheless, given the different format and nature of EEG signals from the images, the alternating arrangement of 1D horizontal (time-dimension) kernel and 1D vertical (spatial-dimension) kernel may not be the optimal design

for EEG signal processing. Therefore, we have constructed another variant of *ResNet18* (Figure 3.11(c)) with the initial two major convolutional blocks (all the nine initial convolutional layers) operating purely in the temporal dimension before introducing the spatial convolution kernels. Spatial-dimension convolution of this *ResNet* variant appears only in the final two convolutional blocks.

In addition, we have investigated the effect of initializing the convolutional path with spatial-dimension convolution, by making only a single change in the initial layer (*Conv 0*) of *ResNet18-1D-kernel-(T-S-alternate)* in Figure 3.11(b), from time-dimension convolution into spatial-dimension convolution. We have name-coded this variant as *ResNet18-1D-kernel-(S-T-alternate)*, such as for comparison with the model in Figure 3.11(b) in order to highlight the effect of the above-mentioned single change on the model's performance which is presented in Figure 4.10.

The right columns of the Figures 3.11(a), (b) & (c) indicate the number of trainable parameters in each architectural layer of the *ResNet* variants.

3.3.3.2. VGG Optimization

As illustrated in Figure 3.12, variants of *VGG16* are also constructed for performance comparison with the variant of *ResNet18*. The *VGG* models (Liu

and Deng, 2015) have classical convolutional pathway without data bypassing. The *VGG16* has five major convolutional blocks, with two convolutional layers in each of its first two major convolutional blocks and three convolutional layers in each of its last three convolutional blocks. These thirteen convolutional layers together with the final three FC layers have made up the 16 main functional layers in the *VGG16*.

Figure 3.12(a) shows the structure of the *VGG16* with all the original 2D kernels being modified into 1D kernels along either the temporal or spatial dimension. The model in Figure 3.12(b) is named *VGG14-1D* with the removal of the two hidden FC layers from the *VGG16-1D*, such that the fully-connected network more closely resembles and is comparable to that of the *ResNet18*.

The *VGG* architectures in Figure 3.12 are also colour-coded such that the transition between different colour blocks is preceded by max pooling (*MaxPool*) operation along the dimension of the previous convolution operation. The adaptive *AvgPool* layer placed before the FC networks in the *VGG* is for the same purpose as described for the *ResNet18*.

We have also investigated the importance of batch normalization in CNN for EEG processing by removing the BN layers of the *VGG16* as in Figure 3.12(c). The performance impact of the BN function is discussed in the Results Section 4.3.2.

3.3.4 Model Training

The objective function for model optimization during training was set as the cross-entropy loss of the CNN outputs. Adam optimizer was used to update the trainable parameters of the CNN at the learning rate of 0.001, based on the backpropagated error from the output cross-entropy loss.

The model training process was conducted with stochastic mini-batches, with the size of each mini-batch being one 200th of the total training pool. Thus, one complete training epoch consists of 200 training iterations. The training data pool will be re-shuffled after every complete training epoch to ensure different combination of mini-batch samples in the subsequent training epochs. Stochastic mini-batch training serves to prevent the training process from being stuck at local minima of the objective function.

CHAPTER 4

RESULTS & DISCUSSION

4.1. Music-Listening EEG Classification (Study 1)

4.1.1. Adjusting for the Suitable Hyperparameters and Constituent Components in the CNN Architecture

Different specification of the architectural details of the CNN can have considerable impacts on the model's classification performance. With this respect, the impacts of the following architectural aspects on the binary EEG classification are investigated based on the pure-temporal shallow CNN model presented in Figure 3.3(a):

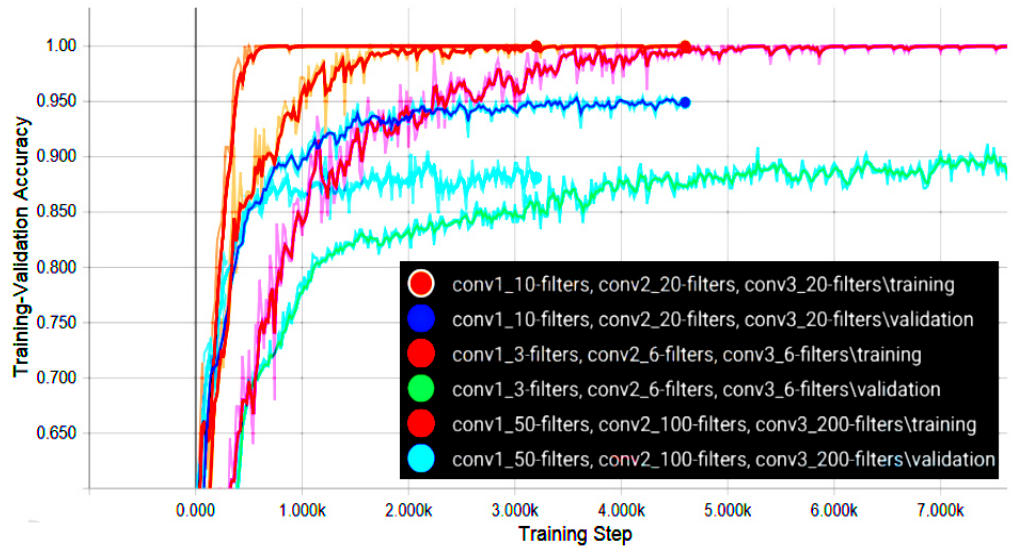
- the amount of convolution channels (i.e. the amount of convolution filters),
- the pooling operation after the convolution, and
- the presence of hidden perceptron layers in the FC-MLP network.

The classification performance of the pure-temporal shallow CNN with different amounts of convolution channels on the task of binary music-EEG classification is presented in Figure 4.1. Among the examined variations, the

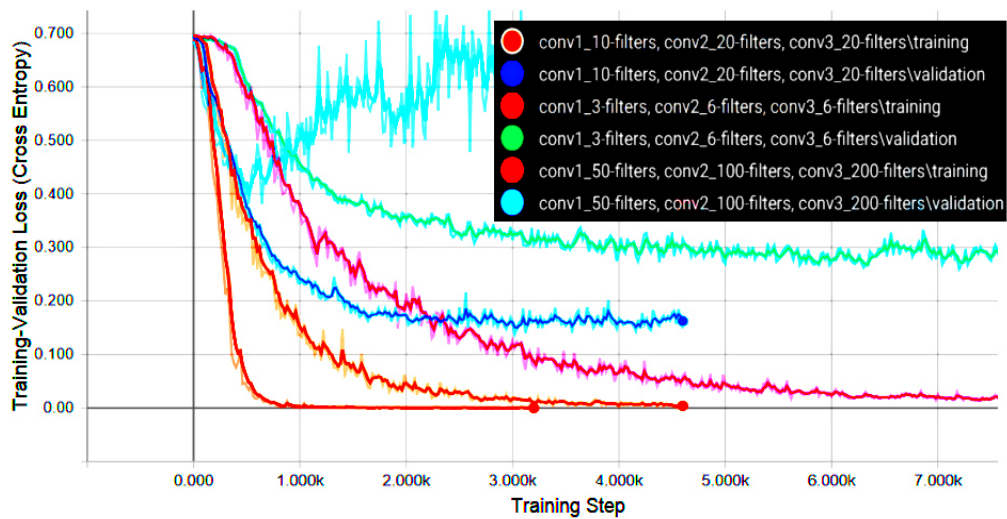
model with ten convolution kernels in the first layer and twenty kernels each in the second and third layers has achieved the highest validation accuracy and the lowest validation loss. The model with low number of convolution kernels with three in the first layer and six each in the second and third layers has experienced significantly lower model learning speed. Low number of convolution kernels has also resulted in eventual lower validation accuracy and higher validation loss, potentially due to insufficient capacity of the CNN to assume the sufficiently complex representation for the data domain. Meanwhile, the model with a very high number of convolutional channels (fifty in the first layer, one hundred in the second layer and two hundred in the third layer) has attained the fastest learning progression at the expense of experiencing early overfitting to the training dataset, which is indicated by the pair of red training curve and the light blue validation curve in Figure 4.1.

Figure 4.2 presents the training-validation log which reflects the impact of FC-MLP on the performance of binary EEG classification of the CNN model. Much resembling the effect of large number of convolution channels, a wider and deeper FC-MLP network with higher capacity (with the depth of three hidden FC layers of 2048, 512, 64 perceptrons respectively at the first, second, and third layers) has attained the fastest learning progress, at the cost of undesirable early overfitting to the training dataset. On the contrary, the CNN classifier with two hidden fully-connected layers of 64 and 32 perceptrons has its learning progress slower than the model with deeper-wider FC-MLP network, it has managed to attain a better validation accuracy and a lower validation loss. The CNN classifier which has no hidden layer in the fully-connected network

has a significantly slower training progress in comparison to its other two CNN counterparts.



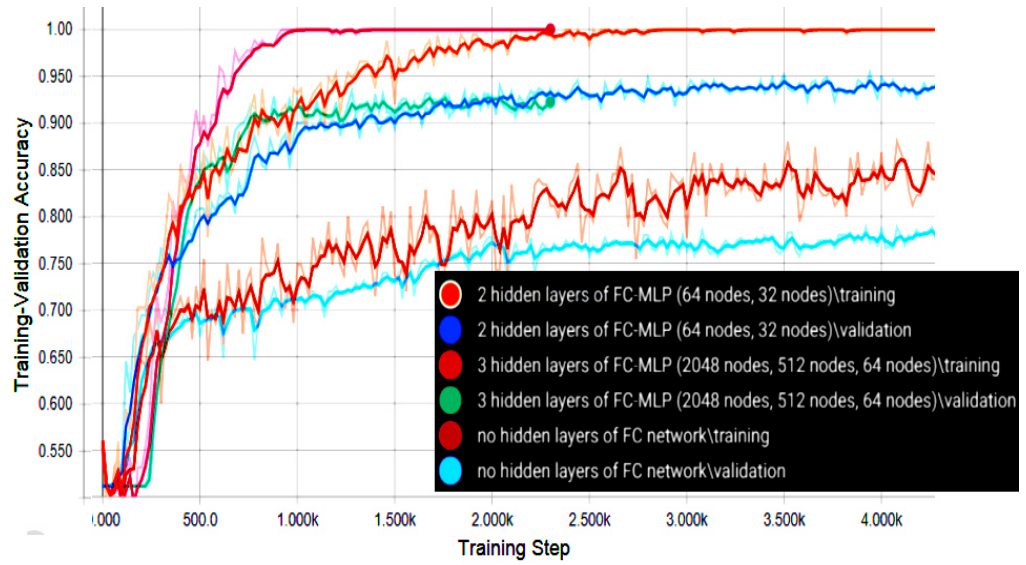
(a)



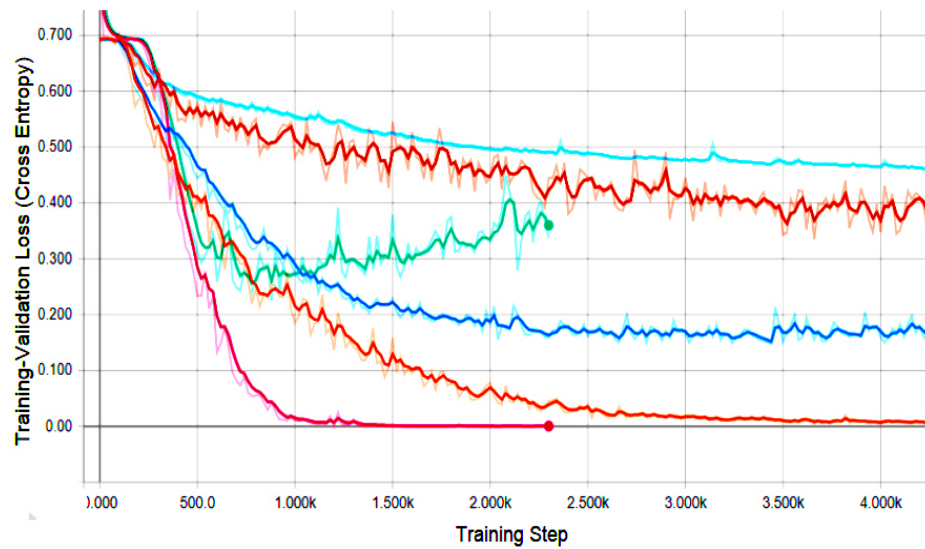
(b)

Figure 4.1: Performance Log of the Model Training-Validation Process using Different Amounts of Convolution Kernels Working on Short-term Music Experiment Binary Classification (Cheah *et al.*, 2019a)

Figure 4.3 presents the performance log of the CNN models with and without pooling mechanism. The performance of the CNN without pooling mechanism is very susceptible to the problem of overfitting which can be easily read from the pair of red and light-blue curves in Figure 4.3(b). The excessive overfitting problem does not occur in the CNN with max pooling mechanism. This indicates that pooling mechanism does not function merely as a method for decreasing the data size of the feature maps to save processing power and to improve computational efficiency, but also significantly improves the classification performance of the CNN model.

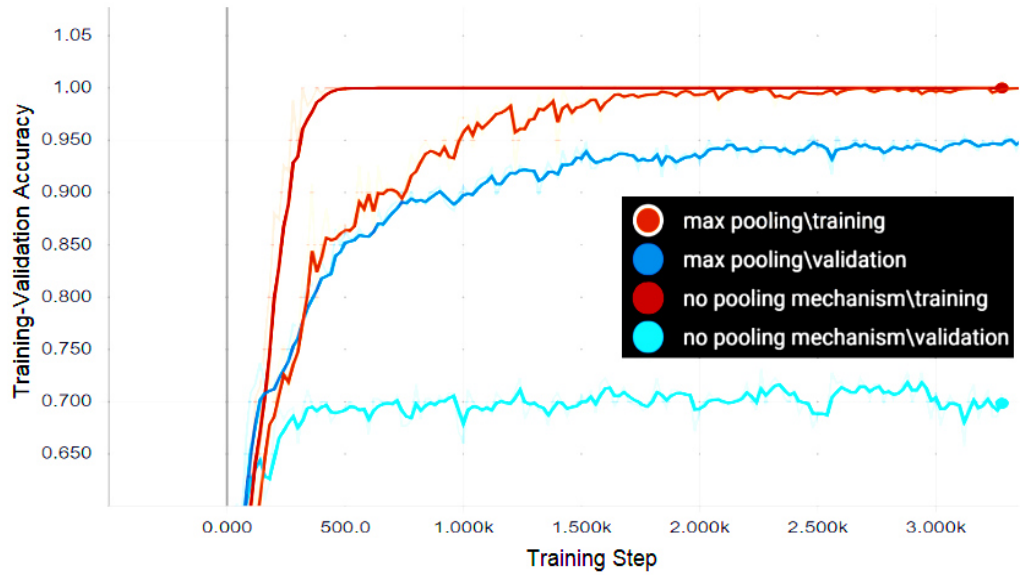


(a)

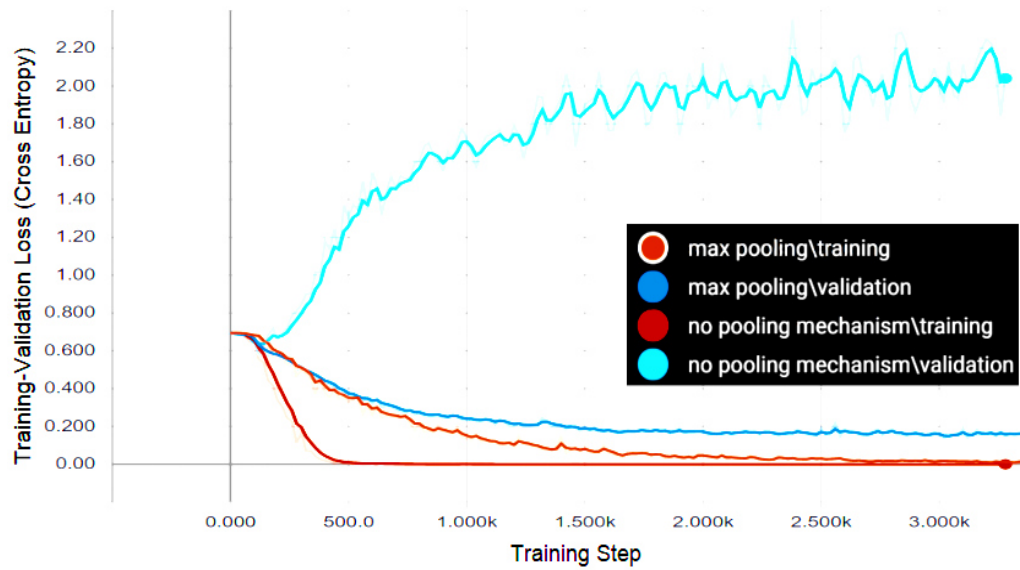


(b)

Figure 4.2: Performance Log of the Model Training-Validation Process with Different Widths and Depths of Fully-Connected (FC) Perceptron Networks (Cheah *et al.*, 2019a)



(a)



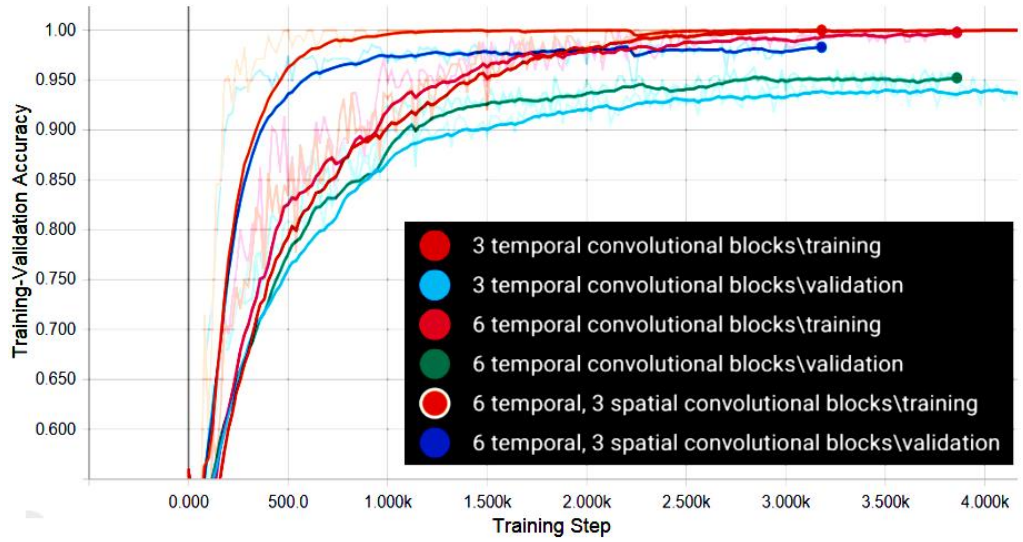
(b)

Figure 4.3: Performance Log for the Model Training-Validation Process with/without Pooling Mechanism on Short-term Music Experiment Binary Classification (Cheah *et al.*, 2019a)

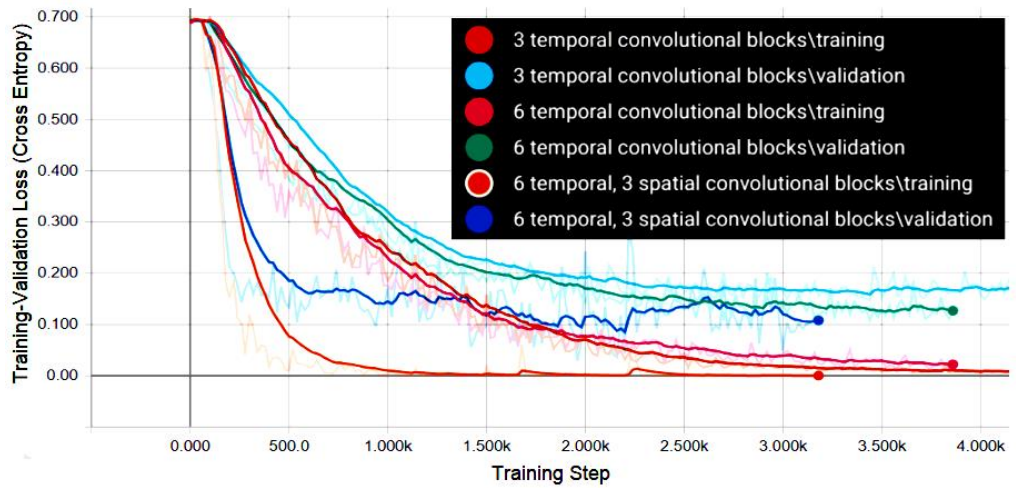
4.1.2. The Importance of Convolution across the Spatial Dimension for EEG Signal Classification

For investigating the significance of spatial-dimension convolution across the EEG channels, Figure 4.4 presents the comparison of the EEG signal classification performance of the three different CNN models illustrated in Figure 3.3(a), Figure 3.3(b), and Figure 3.4(b), where both the models in Figure 3.3 are of pure-temporal convolution and the CNN in Figure 3.4(b) is designed with temporal-spatial convolution.

Both of the CNN models with purely temporal convolution have shown approximately the same classification performance, with the deep temporal CNN with six temporal convolution layers slightly outperforming the shallow temporal CNN model with three temporal convolution layers. With the inclusion of spatial convolution on top of the temporal convolution, the CNN model as illustrated in Figure 3.4(b) has attained considerable performance improvement with higher the validation accuracy, smaller the validation loss, and also much reduced number of training iterations needed to reach the optimally trained state.



(a)



(b)

Figure 4.4: Performance Log for the Model Training-Validation Process of Spatial-Temporal CNN vs. Pure-Temporal-CNN without Spatial Convolution (Cheah *et al.*, 2019a)

4.1.3. Ten-fold Cross-Validation Comparing the CNN with SVM

Instead of single cycle of training performance log as in Figure 4.4, the Table 4.1 and Figure 4.5 together further present the classification performance

of the three CNN models illustrated in Figure 3.3(a), Figure 3.3(b), Figure 3.4(b) as well as the performance of SVM classifiers, over ten folds of cross-validation. The results presented in Figure 4.5 and Table 4.1 are from the CNN and SVM classifiers performing the binary EEG signal classification using the data from the short-term experiment.

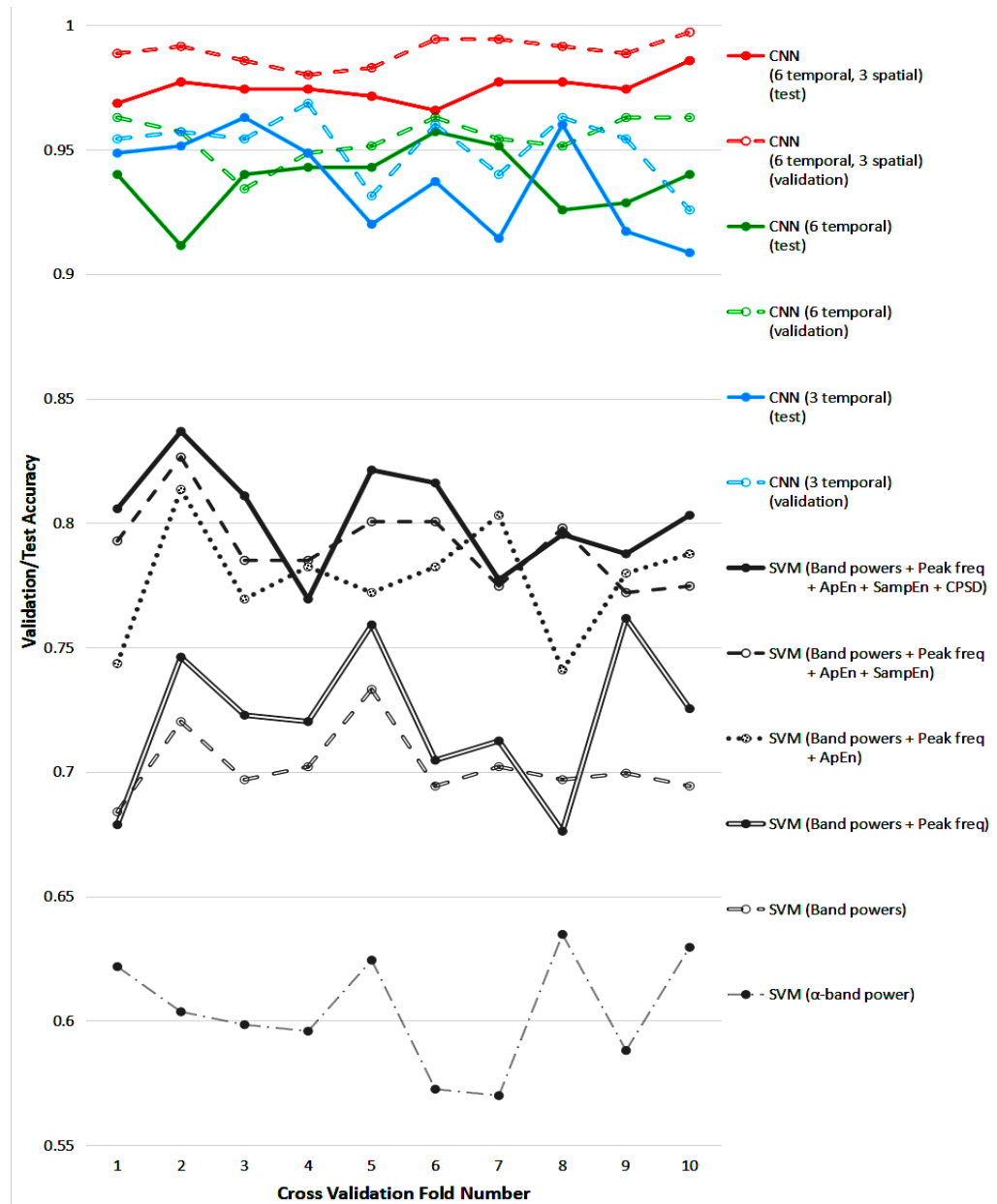


Figure 4.5: Graphical Performance Comparison across Different CNN Architectures and SVM classifiers with Different Input Features across Ten Folds of Cross Validation Process (Cheah *et al.*, 2019a)

Table 4.1: Performance Comparison across Different CNN Architectures and SVM classifiers with Different Input Features across Ten Folds of Cross Validation Process (Cheah *et al.*, 2019a)

Classifier		10-fold Cross-Validation Accuracy (%)										10-fold Average (%)
		Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Fold-6	Fold-7	Fold-8	Fold-9	Fold-10	
CNN (6 temporal, 3 spatial)	Test	96.86	97.71	97.43	97.43	97.14	96.57	97.71	97.71	97.43	98.57	97.46
	Validation	98.86	99.14	98.57	98	98.29	99.43	99.43	99.14	98.86	99.71	98.94
CNN (6 temporal)	Test	94	91.14	94	94.29	94.29	95.71	95.14	92.57	92.86	94	93.8
	Validation	96.29	95.71	93.43	94.86	95.14	96.29	95.43	95.14	96.29	96.29	95.49
CNN (3 temporal)	Test	94.86	95.14	96.29	94.86	92	93.71	91.43	96	91.71	90.86	93.69
	Validation	95.43	95.71	95.43	96.86	93.14	96	94	96.29	95.43	92.57	95.09
SVM (Band powers + Peak freq + ApEn + SampEn + CPSD)		80.57	83.68	81.09	76.94	82.12	81.61	77.72	79.53	78.76	80.31	80.23
SVM (Band powers + Peak freq + ApEn + SampEn)		79.27	82.64	78.5	78.5	80.05	80.05	77.46	79.79	77.2	77.46	79.09
SVM (Band powers + Peak freq + ApEn)		74.35	81.35	76.94	78.24	77.2	78.24	80.31	74.09	77.98	78.76	77.75
SVM (Band powers + Peak freq)		67.88	74.61	72.28	72.02	75.91	70.47	71.24	67.62	76.17	72.54	72.07
SVM (Band powers)		68.39	72.02	69.69	70.21	73.32	69.43	70.21	69.69	69.95	69.43	70.23
SVM (α -band power)		62.18	60.36	59.84	59.59	62.44	57.25	56.99	63.47	58.81	62.95	60.39

The CNN classifiers in general have significantly outperformed the SVM classifier. The CNN model with six temporal and three spatial convolution layers has achieved the averaged validation accuracy of 98.94% over the ten-

fold cross-validation. The optimized temporal-spatial CNN models selected using the top validation performance have an averaged test accuracy of 97.46%. The ten-fold cross-validation classification performance of the deep pure temporal (6 layers) and shallow pure temporal (3 layers) CNN models are close, with the mean test accuracy of 93.8% and 93.69% respectively. The top performing SVM classifier has attained the ten-fold cross-validation mean accuracy of 80.23%.

4.1.4. Three-class Classification by Spatial-Temporal CNN with 1D vs. 2D Kernels

Both of the spatial-temporal CNN models illustrated in Figure 3.4(a) and Figure 3.4(b) are trained for the three-class EEG classification task using the short-term music experiment dataset as shown in Table 3.2. The CNN model in Figure 3.4(a) performs its spatial-temporal convolution with 2D kernels, while the model in Figure 3.4(b) consists of only 1D kernels along either the temporal dimension or the spatial dimension. The validation and test performance of both spatial-temporal CNN models are recorded in Table 4.2 and Figure 4.6. Both of the above-mentioned CNN classifiers have attained very similar accuracy and cross-entropy loss on both the validation and test datasets. Both have achieved the averaged test accuracy of over 95% for the task of three-class EEG classification over the ten-fold cross-validation. On top of being highly capable at differentiating the EEG signals of the mental states listening to music from the EEG of a baseline resting mind, both of the spatial-temporal CNN models

are also very accurate at classifying the mental states of listening to own favorite music or of listening to the alpha binaural beats.

Table 4.3(a) presents the confusion matrix of the three-class classification performance of the 2D-kernel spatial-temporal CNN with the lowest test accuracy (at the 4th fold of the ten-fold cross validation). Correspondingly, Table 4.3(b) presents the confusion matrix of three-class classification performance of the 1D-kernel spatial-temporal CNN with the lowest test accuracy (at the 5th fold out of the ten-fold cross validation).

Both CNN classifiers are slightly more accurate at recognizing the baseline resting EEG segments than at task of identifying EEG segments from the two sub-classes with music, indicated by the consistent highest per-class test accuracy in both of the confusion matrices.

Table 4.2: Performance Comparison between the 2D-kernel Spatial-Temporal CNN and 1D-kernel Spatial-Temporal CNN over 10-fold Cross-Validation based on 3-class Classification on the Short-Term Music Experiment Dataset (Cheah *et al.*, 2019a)

Classifier		10-fold Cross-Validation Accuracy (%)										10-fold Average (%)
		Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Fold-6	Fold-7	Fold-8	Fold-9	Fold-10	
Temporal-Spatial CNN (2D conv kernels)	Test	96.23	96.42	94.92	93.41	95.1	94.73	96.05	95.48	96.05	96.42	95.48
	Validation	97.74	96.99	96.05	96.61	96.99	98.12	98.87	97.55	96.61	97.36	97.29
Temporal-Spatial CNN (Separate 1D conv kernels)	Test	95.67	96.23	95.86	95.29	94.73	96.05	94.92	96.61	96.23	95.48	95.71
	Validation	98.31	96.42	97.36	98.31	97.36	97.36	98.68	97.55	97.74	97.74	97.68
		10-fold Cross-Validation Loss (Cross Entropy)										10-fold Average
		Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Fold-6	Fold-7	Fold-8	Fold-9	Fold-10	
Temporal-Spatial CNN (2D conv kernels)	Test	0.171	0.2274	0.2472	0.2504	0.2449	0.31	0.3986	0.1645	0.1595	0.1732	0.2347
	Validation	0.1154	0.123	0.1541	0.1155	0.1002	0.0895	0.0473	0.121	0.2293	0.1238	0.1219
Temporal-Spatial CNN (Separate 1D conv kernels)	Test	0.2809	0.2202	0.2599	0.3325	0.2427	0.2683	0.3869	0.2407	0.2569	0.2512	0.274
	Validation	0.0581	0.1359	0.1033	0.0687	0.099	0.1168	0.0516	0.0946	0.1317	0.1373	0.0997

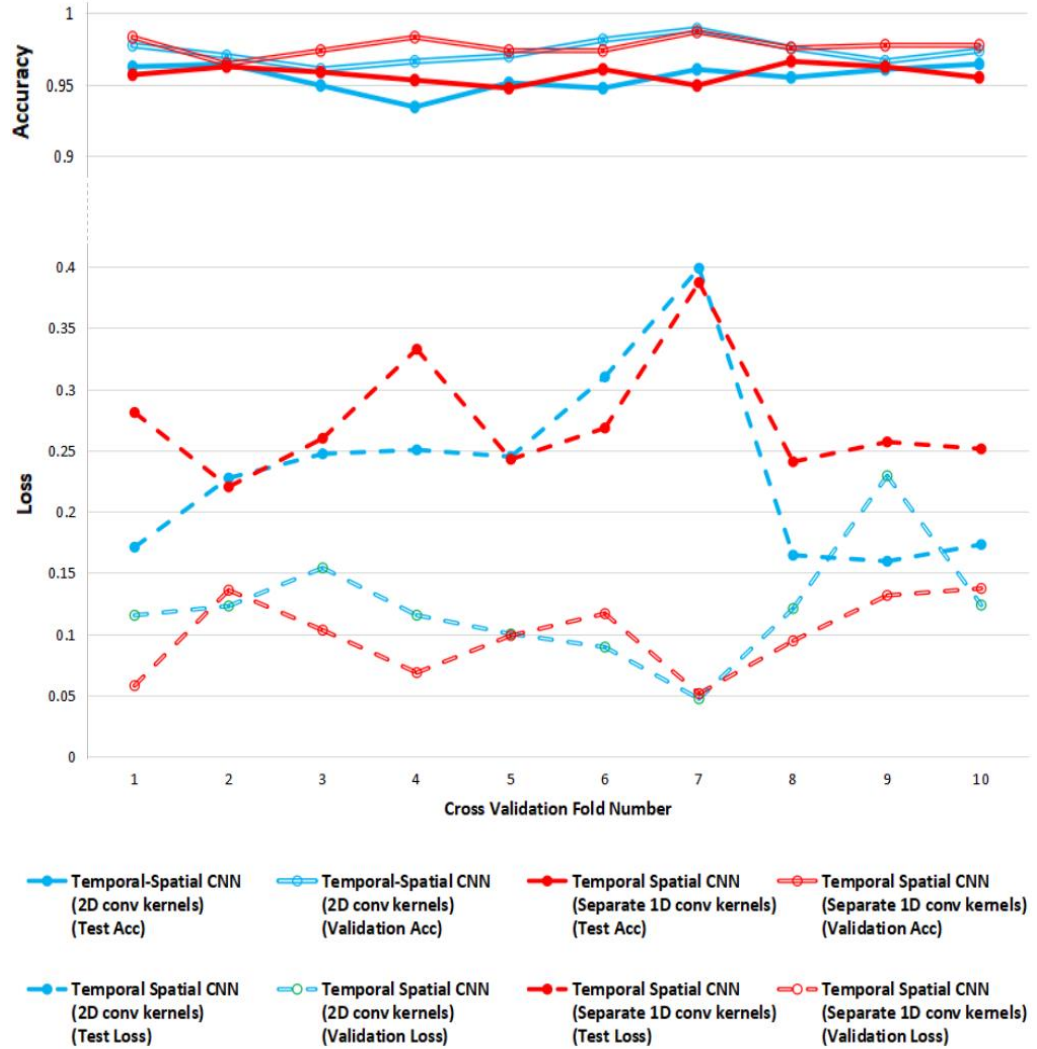


Figure 4.6: Graphical illustration of the Performance Comparison between the 2D-kernel Spatial-Temporal CNN and 1D-kernel Spatial-Temporal CNN over 10-fold Cross-Validation based on 3-class Classification on the Short-Term Music Experiment Dataset (Cheah *et al.*, 2019a)

**Table 4.3: Confusion Matrices of the Cross-Validation Fold with the Lowest Accuracy in Table 4.2 and Figure 4.6 by
(a) 2D-kernel Spatial-Temporal CNN and
(b) 1D-kernel Spatial-Temporal CNN (Cheah *et al.*, 2019a)**

		True Cases (Test dataset)		
		Before Music (161 EEG segments)	Favourite Music (174 EEG segments)	α -Binaural Beats (196 EEG segments)
CNN Predicted Cases	Before Music	152	4	7
	Favourite Music	5	161	6
	α -Binaural Beats	4	9	183
Per-class Test Accuracy:		94.41 %	92.53 %	93.37 %
Overall Test Accuracy:		93.41 %		

(a)

		True Cases (Test dataset)		
		Before Music (161 EEG segments)	Favourite Music (174 EEG segments)	α -Binaural Beats (196 EEG segments)
CNN Predicted Cases	Before Music	154	1	4
	Favourite Music	5	165	8
	α -Binaural Beats	2	8	184
Per-class Test Accuracy:		95.65 %	94.83 %	93.88 %
Overall Test Accuracy:		94.73 %		

(b)

4.1.5. Comparing the Computational Efficiency of Different Classifiers in Terms of Size of Model & Computational Time

4.1.5.1. Size of Model

Table 4.4 presents the detailed calculation of the number of trainable parameters in the two different spatial-temporal CNN classifiers illustrated in Figure 3.4. With similar EEG classification accuracy of both models, the 1D-kernel spatial-temporal CNN model actually operates on a lower amount of trainable parameters than the 2D-kernel spatial-temporal CNN model. The lower number of trainable parameters is achieved by avoiding the use of 2D convolution kernels. As the 1D-kernel spatial-temporal CNN model contains a much lower number of trainable parameters, its storage requires less memory on disc (4,918 kilobytes) in comparison to the 2D-kernel CNN model which requires 5,776 kilobytes of memory. Such reduction in the required disc memory for the model storage can be essential for memory-critical applications such as the embedded systems working on EEG signal processing.

“The trainable parameter calculation presented in Table 4.4 is on the following conceptual basis:

- | |
|---|
| the number of trainable parameters in a convolution layer (no bias parameter)
= (size of conv kernel) × (number of in - channels) × (number of conv kernels) |
|---|
- | |
|--|
| the number of trainable parameters in a fully connected layer
= (number of incoming nodes + 1) × (number of nodes in current layer) |
|--|
- Operation of pooling layer will result in the reduction of data points in the particular dimension, by a reduction factor equivalent to the length of pooling filter along the

dimension (e.g. three successive layers with 1x3 pooling filter reduces the data length by a factor of $3^3 = 27$) ” (Cheah *et al.*, 2019a)

With the adoption of 1D convolution kernels, the amount of trainable parameters in the convolution network/section of the CNN is greatly reduced from 104060 parameters in the 2D-kernel model to as low as 30860 in the 1D-kernel model. This is a parameter reduction of 70% in the convolution section without affecting the classification performance.

Table 4.4: Detailed Comparison of the Trainable Parameters in the Two Different Spatial-Temporal CNN Classifiers (Cheah *et al.*, 2019a)

		Spatio-temporal CNN with 2D-kernels			Spatio-temporal CNN with 1D-kernels		
Trainable Components in the CNN Model	Convolution	Temporal Convolutional Layers	conv-1	$1 \times 6 \times 1 \times 10 = 60$	Temporal Convolutional Layers	conv-1	$1 \times 6 \times 1 \times 10 = 60$
			conv-2	$1 \times 6 \times 10 \times 20 = 1200$		conv-2	$1 \times 6 \times 10 \times 20 = 1200$
			conv-3	$1 \times 5 \times 20 \times 20 = 2000$		conv-3	$1 \times 5 \times 20 \times 20 = 2000$
		Spatio-temporal Convolutional Layers	conv-4	$14 \times 5 \times 20 \times 40 = 56000$	Temporal Convolutional Layers	conv-4	$1 \times 5 \times 20 \times 20 = 2000$
			conv-5	$4 \times 4 \times 40 \times 40 = 25600$		conv-5	$1 \times 4 \times 20 \times 20 = 1600$
			conv-6	$3 \times 4 \times 40 \times 40 = 19200$		conv-6	$1 \times 4 \times 20 \times 20 = 1600$
	N.A.	N.A.	N.A.	Spatial Convolutional Layers	conv-7	$14 \times 1 \times 20 \times 40 = 11200$	
	N.A.	N.A.	N.A.		conv-8	$4 \times 1 \times 40 \times 40 = 6400$	
	N.A.	N.A.	N.A.		conv-9	$3 \times 1 \times 40 \times 40 = 4800$	
	Fully-connected Network	Fully-connected Hidden Layers	fc-1	$(14 \times \text{Round_Up}(256/27) \times 40 + 1) \times 64 = 358464$	Fully-connected Hidden Layers	fc-1	$(14 \times \text{Round_Up}(256/27) \times 40 + 1) \times 64 = 358464$
fc-2			$(64 + 1) \times 32 = 2080$	fc-2		$(64 + 1) \times 32 = 2080$	
		Total Trainable Parameters:	$104060 + 360544 = 464604$	Total Trainable Parameters:	$30860 + 360544 = 391404$		

4.1.5.2. Computational Time

The efficiency of classifiers in terms of computational time is an important aspect of performance measurement of a brain-computer interface (BCI) application (Jiao *et al.*, 2019; Jin *et al.*, 2020; Zhang *et al.* 2016). The computational times of the SVM classifier and the various CNN classifiers constructed in this study are compared in terms of both the training and prediction (test) time which are presented in Table 4.5 and Figure 4.7.

Table 4.5: Training Time and Prediction (Test) Time of Different EEG Classifiers for the Short-term Music Experiment Binary Classification (Cheah *et al.*, 2019a)

	Training time (sec)	Prediction time (millisec)
SVM (all 161 features)	193.7	123.41
Temporal CNN (3 temporal)	201.2	56.04
Temporal CNN (6 temporal)	226.2	63.83
1D-kernel Spatio-temporal CNN	211.5	74.41
2D-kernel Spatio-temporal CNN	118.0	74.00

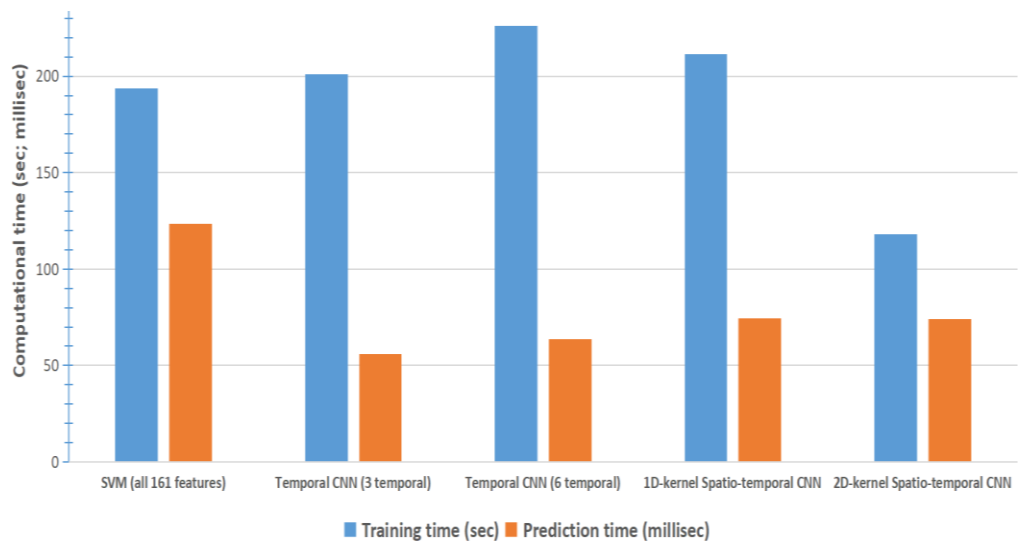


Figure 4.7: Graphical Comparison of Training Time and Prediction (Test) Time of Different EEG Classifiers on the Short-term Music Experiment Binary Classification (Cheah *et al.*, 2019a)

The recorded training time is the average time taken for one of the ten folds of cross validation. The presented prediction (test) time is based on the task of binary EEG classification for 350 EEG segments. The amount of training data is as presented in Table 3.1 in Section 3.1.2.1. With 3150 samples of two-second EEG segments for each fold of training-validation cycle, the time required to optimally train the CNN classifiers and SVM classifier do not differ much, ranging from around 3 to 4 minutes, except for the 2D-kernel spatial-temporal CNN. The time needed for optimizing the SVM classifier will increase exponentially with larger training datasets. On the other hand, the time needed for the training of CNN classifier depends substantially on the training specifications such as the mini-batch size presented per iteration, the learning rate of optimization algorithm, and the validation-based stop criteria. The spatial-temporal CNN with 2D-kernels is able to reach the optimized state of minimal cross-entropy loss in a much shorter training time than the other CNN classifiers with 1D kernels.

Nonetheless, instead of the training time, the prediction time is of greater importance for the deployment and applied execution of the model. The prediction times needed by the spatial-temporal CNN with pure 1D-kernels and the spatial-temporal CNN with 2D-kernels are almost identical, requiring 74.41 and 74.00 milliseconds respectively.

4.1.6. Brain State Classification based on EEG Signals from Different Brain Lobes

4.1.6.1. Left Hemisphere vs. Right Hemisphere of the Cerebra

Table 4.6 and Figure 4.8 show the validation and test accuracy of the three-class music-EEG classification over ten-fold cross-validation, using only either the left or the right cerebral hemispheric EEG channels. On average, the classification accuracy achieved using the EEG channels from the left cerebral hemisphere [“AF3, F3, F7, FC5, T7, P7, O1” ([Headset Comparison Chart: Technical Specification, \[n.d.\]](#))] is approximately 5% better than the result achieved using the EEG channels from the right cerebral hemisphere [“AF4, F4, F8, FC6, T8, P8, O2” ([Headset Comparison Chart: Technical Specification, \[n.d.\]](#))]. The discrepancy in the above classification accuracy suggests that the left cerebral hemispheric EEG signals elicit a greater difference between the resting baseline state and the brain state under different kinds of music stimuli, in comparison to that of the right cerebral hemisphere.

The discrepancy in the left-vs-right cerebral hemispheric EEG classification accuracies, under the same auditory stimulus in this study is probably because of the fundamental lateralization of cerebral hemispheric functions. While the self-favorite music stimulus and the alpha binaural music rhythm have induced clearly differentiable EEG signals in the right cerebral hemisphere (distinguishable by the CNN classifier with the averaged accuracy of 84.12%) which has a dominant role in handling the emotional functions, these auditory stimuli have induced an even greater difference in EEG signals

generated by the left cerebral hemisphere (distinguishable by the CNN classifier with the averaged accuracy of 88.91%) which has a more important role in linguistic processing and logical functions.

The innate differences in the structures and functions of the left and the right cerebra have resulted in their different ways of responding to all kinds of external or internal stimuli. This is in-fact a well-known widespread primal property among the humans as well as the other animals (Corballis, 2014), termed as the cerebral lateralization or cerebral asymmetry. Lateralization of cerebral functions has been recognized as an indication of successful and efficient neurological development (Liu *et al.*, 2009). On the contrary, a number of studies had reported the diminished degree of cerebral lateralization or higher degree of ambidexterity to be positively correlated with neuropsychological dysfunctions such as stuttering, deficiency in academic skills, difficulty in maintaining mental health, and even schizophrenia (Crow *et al.*, 1998; Kushner, 2011; Orr *et al.*, 1999; Rodriguez *et al.*, 2010).

Table 4.6: Three-class Classification Performance (Accuracy and Cross-Entropy Loss) Achieved with Left vs. Right Cerebral Hemispheric EEG Signals (Cheah *et al.*, 2019a)

EEG Channels		10-fold Cross-Validation Accuracy (%)										10-fold Average (%)
		Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Fold-6	Fold-7	Fold-8	Fold-9	Fold-10	
Left Hemisphere Channels	Validation	90.58	90.02	90.02	89.83	88.89	89.64	91.9	93.97	89.27	92.66	90.68
	Test	87.01	88.7	88.32	89.64	88.89	87.57	89.45	89.64	89.83	90.02	88.91
Right Hemisphere Channels	Validation	87.57	85.12	85.88	82.49	86.82	86.44	86.63	85.5	83.24	86.63	85.63
	Test	83.05	85.69	82.49	82.67	83.99	85.31	84.75	85.31	82.49	85.5	84.12
		10-fold Cross-Validation Loss (Cross Entropy)										10-fold Average
		Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Fold-6	Fold-7	Fold-8	Fold-9	Fold-10	
Left Hemisphere Channels	Validation	0.2997	0.3525	0.36	0.3527	0.3175	0.3823	0.2537	0.2392	0.4305	0.2743	0.3263
	Test	0.4972	0.7013	0.583	0.4369	0.349	0.6072	0.3493	0.4678	0.7225	0.3179	0.5032
Right Hemisphere Channels	Validation	0.3432	0.4688	0.4352	0.5028	0.3777	0.4677	0.4369	0.4439	0.4714	0.4975	0.4445
	Test	1.0055	0.5865	0.8259	0.6505	0.8273	0.9729	0.5784	0.9382	0.8294	0.724	0.7939

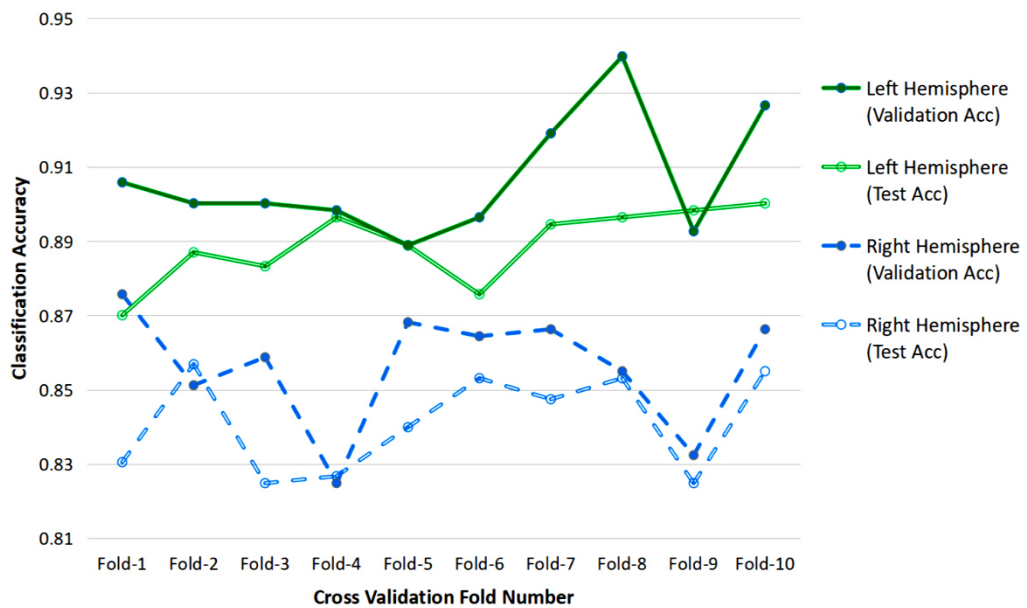


Figure 4.8: Graphical Comparison of Three-class Classification Performance Achieved with Left vs. Right Cerebral Hemispheric EEG Signals (Cheah *et al.*, 2019a)

4.1.6.2. Significance of Frontal-lobe Signals vs. Temporal-Parietal-Occipital (TPO) EEG Signals

Table 4.7 and Figure 4.9 present the validation and test accuracy of the three-class music-EEG classification over ten-fold cross-validation, using only either the “six frontal-lobe EEG channels (AF3, AF4, F3, F4, F7, F8)” (Cheah *et al.*, 2019a) or the other “six EEG channels from the temporal, parietal and occipital lobes (T7, T8, P7, P8, O1, O2)” (Cheah *et al.*, 2019a). The classification performance discrepancy between using six frontal lobe channels and using the six EEG channels from the temporal, parietal and occipital (TPO) lobes is even greater than the discrepancy elicited between the left-vs-right cerebral hemispheres. On average, the classification test accuracy achieved with six TPO channels is barely 74.69%, which is over 10% worse than the accuracy achieved with six frontal-lobe channels of 84.93%. This accuracy discrepancy indicates that neural activity of the frontal cerebrum is more strongly activated and influenced by the music stimulus than the other cerebral cortices (the temporal, parietal and occipital lobes).

The frontal cerebrum has particularly dominant duty in the integrative, imaginative, and executive functions of the mind. These include the integration of different modalities related to perception or sensory inputs, integration of memory, prospection or future projection, execution planning, and emotional mediation (Siddiqui *et al.*, 2008). With this respect, musical stimuli could potentially have significant effects on these neurological functions of the frontal lobes.

Table 4.7: Three-class Classification Performance (Accuracy and Cross-Entropy Loss) Achieved with Frontal-lobe EEG vs. TPO-lobe EEG (Cheah *et al.*, 2019a)

EEG Channels		10-fold Cross-Validation Accuracy (%)										10-fold Average (%)
		Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Fold-6	Fold-7	Fold-8	Fold-9	Fold-10	
Frontal Channels (AF3, AF4, F3, F4, F7, F8)	Validation	88.51	83.62	88.89	86.63	87.76	88.51	87.76	91.34	88.14	87.57	87.87
	Test	84.93	83.99	84.37	84.56	86.44	84.93	84.18	85.88	86.82	83.24	84.93
TPO Channels (T7, T8, P7, P8, O1, O2)	Validation	80.6	78.53	79.47	79.66	84.56	80.41	81.36	79.66	78.72	81.36	80.43
	Test	74.01	77.78	74.01	77.78	75.52	72.13	71.56	73.63	75.71	74.76	74.69
		10-fold Cross-Validation Loss (Cross Entropy)										10-fold Average
		Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Fold-6	Fold-7	Fold-8	Fold-9	Fold-10	
Frontal Channels (AF3, AF4, F3, F4, F7, F8)	Validation	0.3811	0.4609	0.3595	0.392	0.4332	0.3642	0.3862	0.358	0.421	0.4192	0.3975
	Test	0.9445	0.5366	0.5351	0.7084	0.5467	0.5216	0.4408	1.2594	0.648	0.52	0.6661
TPO Channels (T7, T8, P7, P8, O1, O2)	Validation	0.5447	0.6057	0.64	0.694	0.5213	0.5563	0.5409	0.6632	0.5852	0.5289	0.588
	Test	0.7743	0.8931	0.72	0.7086	1.0357	0.8955	0.7434	0.7273	0.7566	1.5063	0.8761

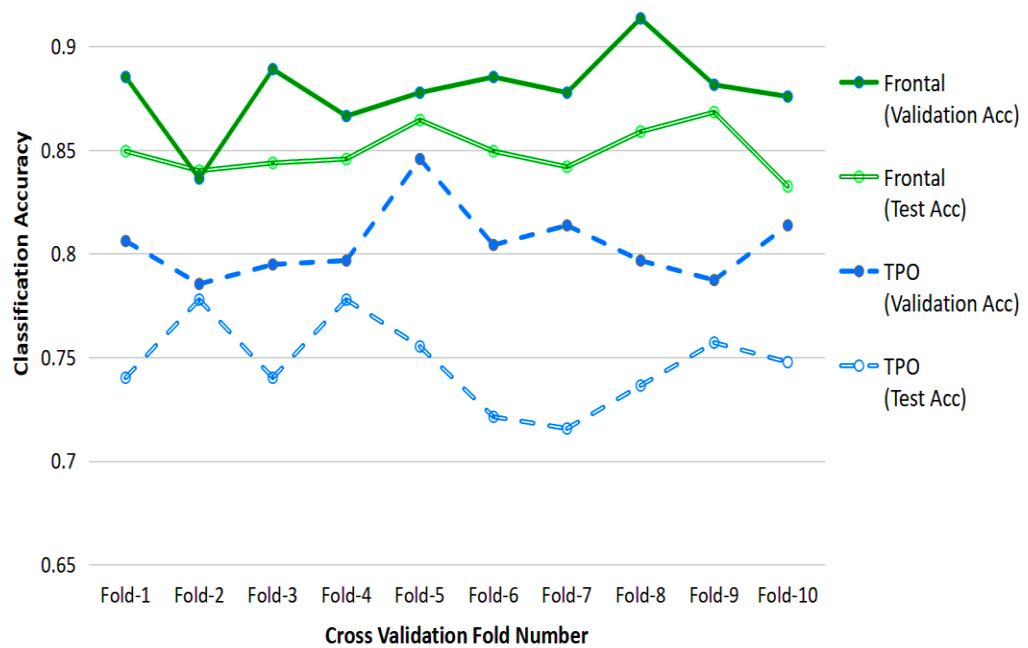


Figure 4.9: Graphical Three-class Classification Performance Achieved with Frontal-lobe EEG vs. TPO-lobe EEG (Cheah *et al.*, 2019a)

4.2. CNN for Personalized Emotion Classification (Study 2)

The classification accuracies of the CNN classifiers in Study 2 (in Table 3.4 and Table 3.5) for the tasks of three-class valence level recognition and three-class arousal level recognition are recorded in Table 4.8 and Table 4.9. The classification accuracy tabulated is the average of the four-fold cross validation for each individual participant.

The performance of emotion classification of the single-path CNN and the double-path CNN classifiers are closely equivalent to each other. The single-path CNN classifier attains test accuracy of 97.59% and 98.48% respectively for the valence and arousal level recognition. The double-path CNN classifier attains the corresponding accuracy of 98.75% and 97.58%.

Although the two CNN classifiers have achieved closely identical accuracies, the convolutional networks of the CNN models contain vastly different amounts of trainable parameters. The convolution network of the double-path CNN model is designed with three kernels in every layer along each convolution path, aggregating to a sum of 960 trainable parameters (with 480 in one convolution path) in the whole convolution network. On the other hand, the single-path CNN model is equipped with six convolution kernels per layer, aggregating to a sum of 1824 trainable parameters, which is 1.9 times as many as the parameters in the convolution network of the double-path CNN. Meanwhile, the FC-MLP sections of both classifiers are designed with exactly identical number of hidden perceptrons.

The CNN models in this study are designed with low number of convolutional kernels for the purpose of testing the performance of models with low convolutional capacity in emotion recognition. Also, with lower number of kernels in the convolution layers, both the convolutional filters and the output feature maps are more interpretable for the potential identification of new useful EEG feature. With lower number of convolution kernels, the CNN classifier will also be mannered to identify only the highly relevant signal features, which in turn can be helpful in preventing the model from overfitting to the training dataset. Meanwhile, with different dilation factors for the operation of the convolution kernels, the CNN network is mannered into picking up the signal features at different frequencies.

Table 4.8: Average four-fold Cross-Validation and Test Accuracies for the Three-Class Emotional Valence Classification (Cheah *et al.*, 2019b)

Single-path CNN accuracy (%)						2-path CNN accuracy (%)					
Subject	Validation	Test	Subject	Validation	Test	Subject	Validation	Test	Subject	Validation	Test
1	98.79	97.46	17	98.09	95.05	1	99.19	98.89	17	99.42	97.65
2	99.36	98.93	18	97.82	96.53	2	94.12	92.41	18	97.80	95.80
3	99.61	99.37	19	97.86	97.82	3	97.89	97.36	19	97.02	97.36
4	99.48	97.58	20	98.99	98.35	4	94.07	91.61	20	99.32	99.37
5	98.70	98.93	21	99.41	98.58	5	93.99	94.95	21	99.24	97.65
6	99.68	99.56	22	98.89	97.40	6	98.93	98.27	22	99.24	98.13
7	99.92	99.85	23	99.03	98.09	7	98.25	97.58	23	99.12	96.80
8	95.95	89.99	24	98.93	99.32	8	96.48	90.67	24	99.60	99.64
9	99.50	99.56	25	98.27	98.87	9	99.66	99.37	25	99.36	98.83
10	98.09	95.94	26	98.26	99.09	10	95.18	92.67	26	97.23	98.20
11	97.64	97.36	27	99.81	99.88	11	97.78	98.23	27	97.85	98.23
12	99.27	99.24	28	99.06	97.78	12	95.76	95.34	28	99.57	98.36
13	98.79	98.70	29	99.27	98.97	13	99.96	99.91	29	98.89	98.90
14	99.21	97.84	30	98.34	96.20	14	98.16	97.27	30	99.28	96.79
15	93.92	92.91	31	99.52	99.20	15	94.81	93.49	31	99.76	99.77
16	98.98	97.90	32	95.17	90.52	16	95.28	92.87	32	96.49	93.32
			Overall	98.55	97.59				Overall	97.77	98.75

Table 4.9: Averaged 4-fold Cross-Validation and Test Accuracies for the Three-Class Emotion Arousal Level Classification (Cheah *et al.*, 2019b)

Single-path CNN						2-path CNN					
Subject	Validation	Test	Subject	Validation	Test	Subject	Validation	Test	Subject	Validation	Test
1	99.56	99.72	17	98.25	95.77	1	99.85	99.74	17	97.93	94.51
2	99.36	98.75	18	98.45	98.20	2	98.54	97.68	18	97.76	93.41
3	99.66	98.41	19	99.43	99.68	3	98.65	97.29	19	98.96	99.14
4	98.97	98.62	20	99.48	99.21	4	97.49	96.47	20	98.92	98.70
5	99.02	98.69	21	99.66	99.50	5	96.48	98.18	21	99.52	99.51
6	99.88	99.64	22	98.02	96.73	6	99.61	99.59	22	99.05	97.93
7	99.66	99.63	23	99.02	99.03	7	96.87	97.28	23	98.07	97.55
8	98.16	98.29	24	99.70	99.61	8	97.86	94.55	24	98.75	98.90
9	99.29	99.03	25	98.00	97.87	9	99.63	99.74	25	97.98	97.59
10	98.72	98.29	26	98.39	99.09	10	98.47	97.46	26	98.57	99.02
11	97.22	96.06	27	99.06	99.19	11	96.39	93.99	27	97.96	98.05
12	99.30	99.38	28	99.27	98.47	12	98.26	98.09	28	97.05	95.23
13	99.19	98.75	29	99.48	99.47	13	97.50	97.59	29	98.70	98.35
14	99.55	98.87	30	94.24	92.02	14	99.61	99.37	30	98.85	97.56
15	99.38	99.45	31	99.38	98.90	15	97.96	98.56	31	98.00	97.02
16	98.53	98.07	32	99.59	99.02	16	97.72	96.10	32	99.21	98.51
			Overall	98.90	98.48				Overall	98.32	97.58

4.3. Residual Network and VGG for Emotion EEG Classification (Study 3)

4.3.1. Performance of Variants of *ResNet18*

Figure 4.10 presents the averaged 5-fold cross-validation classification accuracy of the *ResNet* variants, using different subsets of EEG channels as their data input.

The *ResNet* variant with 1D kernels has generally outperformed the original *ResNet18*, particularly at the scenario of using lower number of EEG channels (10 channels for each subset). Not only has the classification improved with the *ResNet18* architectural restructuring from 2D-kernel convolution to 1D-kernel convolution, the total number of trainable parameters (obtainable by summing up the layer-wise parameters in Figure 3.11) in the *ResNet18* has also seen a reduction of more than 50% from the original 11.17 million parameters down to the range of 4.27 to 5.34 million parameters.

As described above in Section 3.3.3.1, the models *ResNet18-1D-(S-T-alternate)* and *ResNet18-1D-(T-S-alternate)* differ in only their very first convolutional layer (the *Conv-0* layer of Figure 3.11(b)), where the *ResNet18-1D-(T-S-alternate)* model has the layer *Conv-0* as temporal convolution while the *ResNet18-1D-(S-T-alternate)* model has its *Conv-0* layer as spatial convolution. Although this single change in the *Conv-0* layer has resulted in the

difference in parameter count by only 256 ($(1 \times 9 - 5 \times 1) \times 64 = 256$), the performance in EEG signal classification has seen substantial improvement by about 10% elevation (using either all 62 channels, the outermost 10 channels, or outer 10 channels), as presented in Figure 4.10. This strongly indicates that the convolution operation on plain EEG signal should not be initiated with spatial(channel)-dimension convolution.

Some other previous works that used CNN for plain EEG signals processing had also forced the convolution process to operate only along either the temporal or spatial dimension for every single convolutional layer. Most of the works ([Chambon et al., 2017](#); [Kwak et al., 2017](#); [Manor and Geva, 2015](#); [Zafar, Dass and Malik, 2017](#)) applying 1D-kernel CNN on EEG signals had initiated the convolutional path with temporal convolution. However, they had not justified the reason for design nor had they provided the performance comparison with the models that did otherwise, as we highlighted in this study.

We took a further step of increasing the number of layers of pure temporal convolution before starting spatial convolutional operation, as in the architecture of *ResNet18-1D-(T-then-S)* in Figure 3.11(c). The *ResNet18-1D-(T-then-S)* model has outperformed all the other *ResNet18* variants substantially, in every classification scenario as reported in Figure 4.10.

This supports that constructing multiple consecutive layers of temporal convolution before starting spatial convolution is beneficial for extracting distinctive information from the EEG signals. Although *ResNet* had been reported with inferior performance than the typical CNN at EEG classification in [Schirrneister et al. \(2017\)](#), their *ResNet* architecture was however designed with spatial convolution very early on as the second convolutional layer. If more temporal convolutional layers were introduced before the spatial convolution, the *ResNet* presented in [Schirrneister et al. \(2017\)](#) may perhaps have significant performance improvement.

With the presence of multiple consecutive temporal convolutional layers before spatial convolution, higher hierarchical features within each EEG channel could be extracted before comparing across different channels. Direct cross-channel convolution of rudimentary EEG voltages may not carry as much distinctive information as that of the higher hierarchical features.

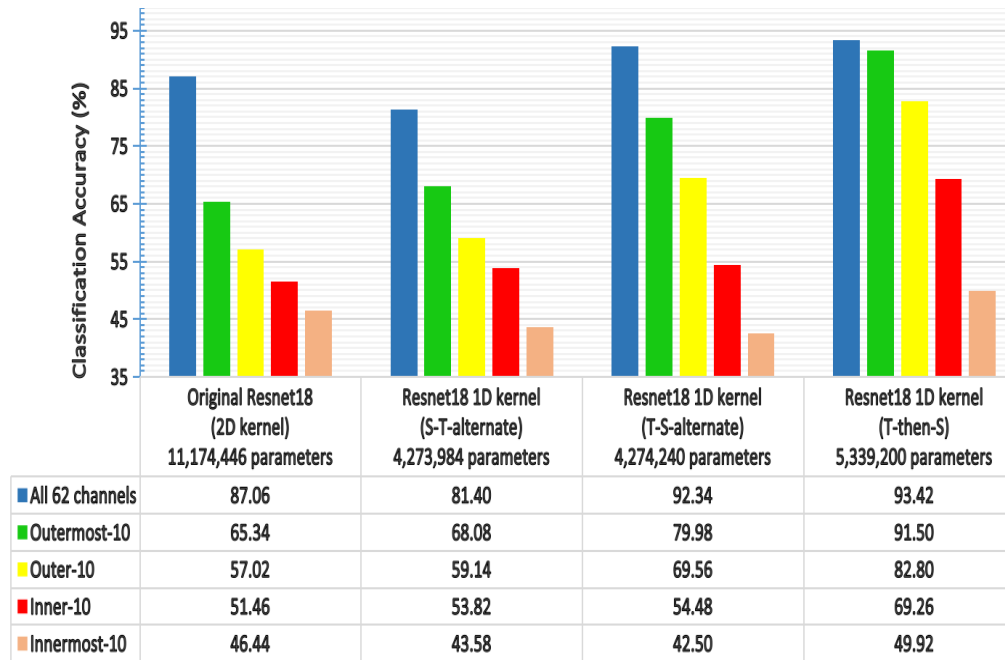
Plain EEG signals carry only voltage levels measured over the scalp. Every single sampling point of the voltage level in an EEG channel is not as meaningful as a sequence of sampling points along the channel. Excessively short receptive field over a single channel is susceptible to recording artifacts and other non-essential signal variations.

Therefore, with multiple consecutive temporal convolutional layers, the initial stages of the model can cover a larger receptive field over the raw signal,

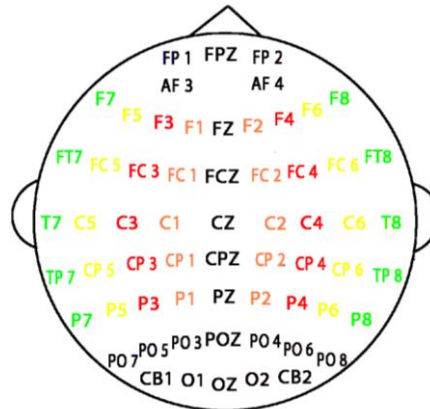
at the same time extracting features of higher level of abstraction from the particular channel. Comparing the rudimentary EEG signal sampling-point by sampling-point across the channels may have taken into account a considerable amount of the undesired meaningless voltage variations, resulting in lower classification accuracy in the *ResNet18-ID-(S-T-alternate)* model.

We have also constructed and examined another variant of *ResNet18-ID-(S-then-T)* model with its several initial convolutional layers all being spatial-dimension convolution then only followed by temporal convolution. This model which was not presented in Figure 3.11 had presented worse performance than even the *ResNet18-ID-(S-T-alternate)* model, which further supports the discussion above that EEG signal convolution for emotion recognition should ideally be started with temporal-dimension convolution.

Figure 4.11 reports the training-validation performance log of the four variants of *ResNet-ID*, using the 10 outermost channels. Based on the training-validation cross-entropy loss plot, the *ResNet18-ID-(T-then-S)* model which had outperformed all the rest, was clearly less susceptible to overfitting. The other three *ResNet18-ID* models had started to experience overfitting after around eight to ten training epochs, with the models *ResNet18-ID-(S-then-T)* and *ResNet18-ID-(S-T-alternate)* experiencing the greatest degree of overfitting.



(a) Classification accuracy & total number of model parameters



(b) Different subsets EEG channels

Figure 4.10: SEED 3-class Emotion Recognition Accuracy by Variants of ResNet18 using Different Subsets of EEG Channels

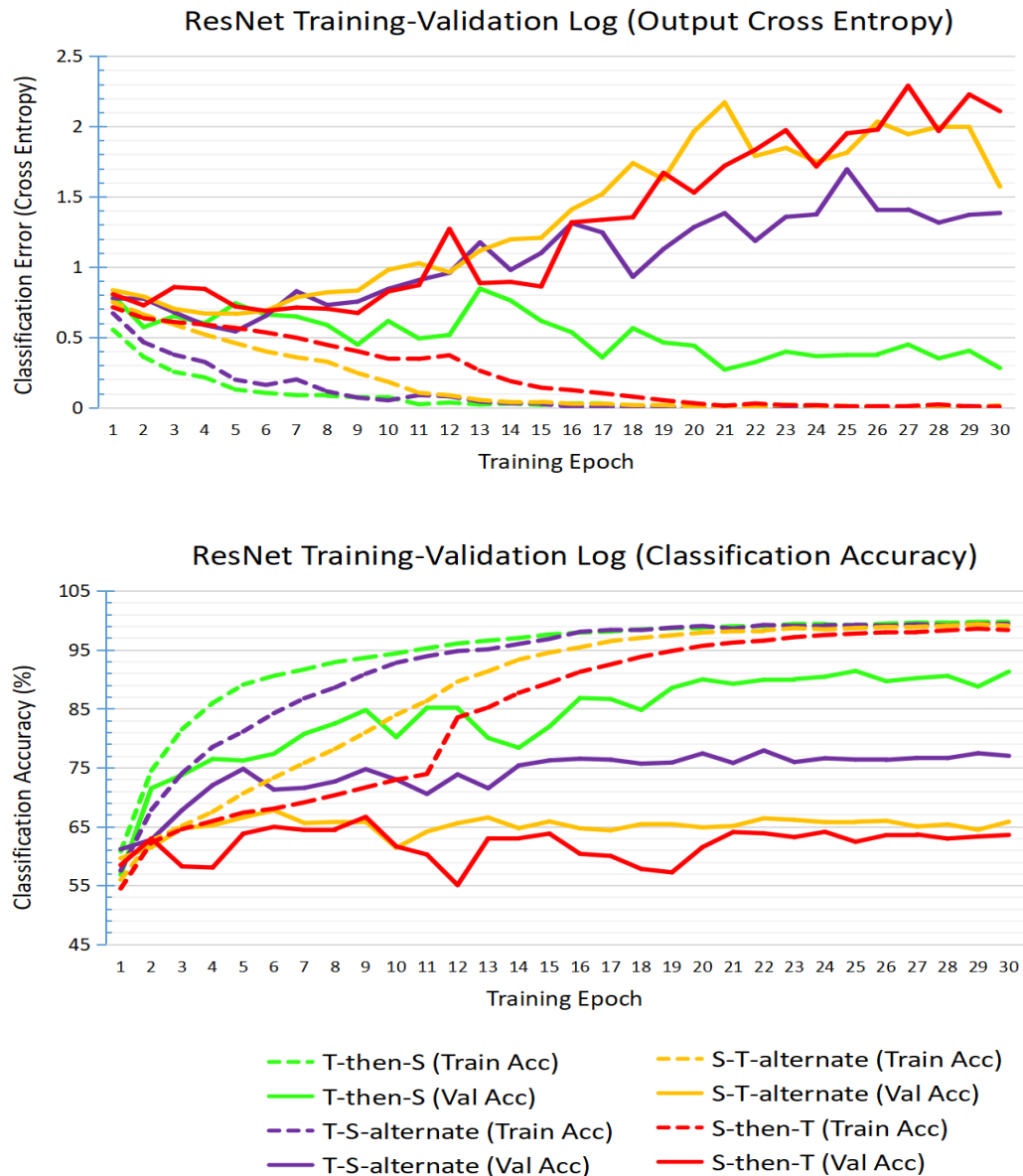


Figure 4.11: Training-validation Performance Log of Variants of *ResNet18-1D*

4.3.2. Performance of *ResNet* vs. *VGG*

We have compared the performance of *ResNet18* with the more classical CNN architecture (the *VGG16*) from the aspects of classification accuracy, number of trainable parameters, and the model training convergence speed.

Figure 4.12 shows that the classification accuracy achieved by *ResNet18-1D(T-then-S)*, *VGG14-1D*, and *VGG16-1D* models are very close to each other. The *ResNet18-1D(T-then-S)* achieves 93.42% classification accuracy, outperforming the *VGG* at using all 62 EEG channels. The *VGG* models have achieved higher accuracy at the less significant subsets of EEG channels (e.g. using the innermost 10 channels).

Given the almost negligible difference in the classification accuracy, the *ResNet18-1D(T-then-S)* model contains only 5.34 million parameters, which is only about 36.3% of that in the *VGG14-1D* model which has 14.72 million of parameters. The *VGG16-1D* has an even staggering greater number of parameters (at 46.18 million) due to the large number of fully-connected perceptrons in its original 3-layer FC networks. This densely connected FC network containing over 31 million parameters does not appear to be essential to the classification accuracy.

Another aspect of performance measurement investigated is the convergence speed of the model under training. With reference to Table 4.10, using all 62 EEG channels, the *ResNet18-1D(T-then-S)* and the *VGG14-1D* models are able to converge to above 95% training accuracy in 11 epochs and 10 epochs, respectively. The *VGG16-1D* requires a greater number of training epochs (14 complete rounds) to reach its training accuracy of 95%. The lower convergence speed of *VGG16-1D* is likely due to its complex FC network.

The *ResNet18-1D(T-then-S)* model completes a training epoch with $(1665/11 \approx 151)$ seconds, while the *VGG* models require much greater amount of time to complete a training epoch (*VGG14-1D* using about 249 seconds, and *VGG16-1D* using about 250 seconds).

Similarly, the *ResNet18-1D(T-then-S)* uses only about 38 seconds for a complete training epoch with 10 EEG channels, while the two *VGG* models use about 50 seconds for completing a training epoch.

The *VGG14-1D* illustrated in Figure 3.12(c) is the version of *VGG14-1D* without the batch normalization function after every convolutional layer. This model without the batch normalization had failed to progress well even its training phase. The training accuracy of this model had stayed at around 35%, with the training loss staying at around the initial value. The failure of the *VGG14-1D* without batch normalization has indicated the importance of batch normalization in training deep CNN on EEG signals, even with the EEG signals being pre-normalized before being passed into the CNN model. All the *ResNet18* variants in Figure 3.11 are also equipped with batch normalization at the output of their convolutional layers.

In our model, each layer of batch normalization function introduces two additional trainable parameters per feature map. The dimension of the feature

map depends on the number of convolutional kernels immediately preceding the batch norm function.

The short EEG segments being passed into the classifier may contain large signal amplitude variations from segment to segment. Different batches of the EEG segments may also encounter the problem of large internal covariate shift (Ioffe and Szegedy, 2015) which is a notorious reason for diverging loss during model optimization (Bjorck et al., 2018).

This does not only slow down the training speed by demanding very low learning rate, but also potentially disrupt altogether the convergence of the model optimization process as experienced in our model (Figure 3.12(c)) without batch normalization.

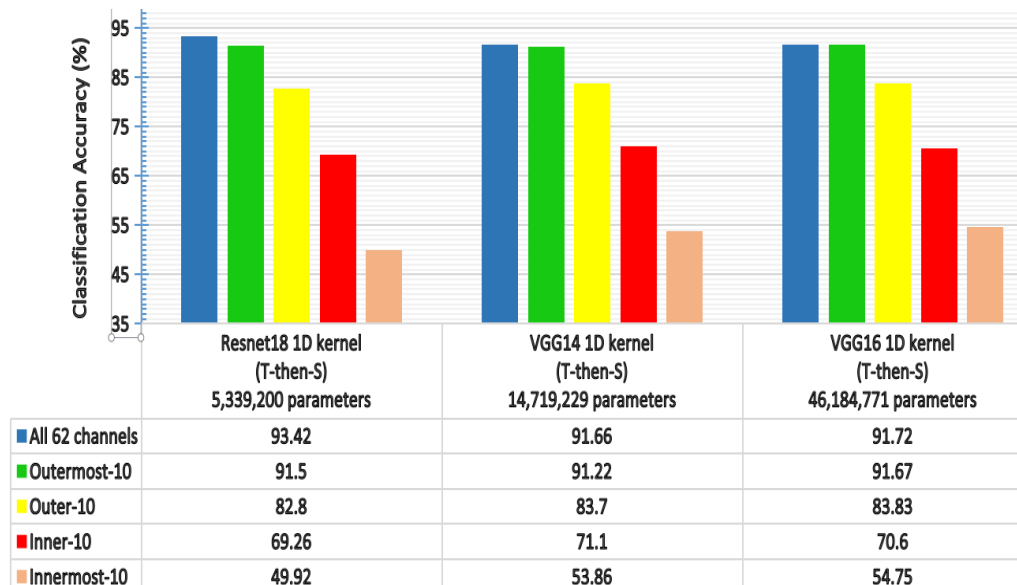


Figure 4.12: Classification Accuracy of Emotion-labelled EEG by *ResNet18-1D* and *VGG16* Variants

Table 4.10: Convergence Time Needed during Model Training for the *ResNet* and *VGG*

	Training length to reach 95% training accuracy (epochs // seconds)	
	All 62 channels	Outermost 10 channels
ResNet18-1D (T-then-S)	11 // 1665	11 // 416
VGG14-1D (T-then-S)	10 // 2488	10 // 503
VGG16-1D (T-then-S)	14 // 3505	12 // 622

4.3.3. EEG Channel Significance for Emotion Recognition

Identifying the most critical subsets of EEG channels can reduce the input data redundancy and ease the design and mounting of portable consumer-friendly EEG recording hardware. Therefore, previous works ([Ansari-Asl *et al.*, 2007](#); [Ozerdem and Polat, 2017](#); [Zheng and Lu, 2015](#)) had tried to identify the subsets of EEG channels that are most crucial for emotion recognition. In line with the purpose, we have look into the emotion-EEG channel significance with regard to lateral-medial placement, along the nasion-inion axis, and in terms of left-vs-right hemispheric discrepancy.

4.3.3.1. Electrode Distance from the Midline

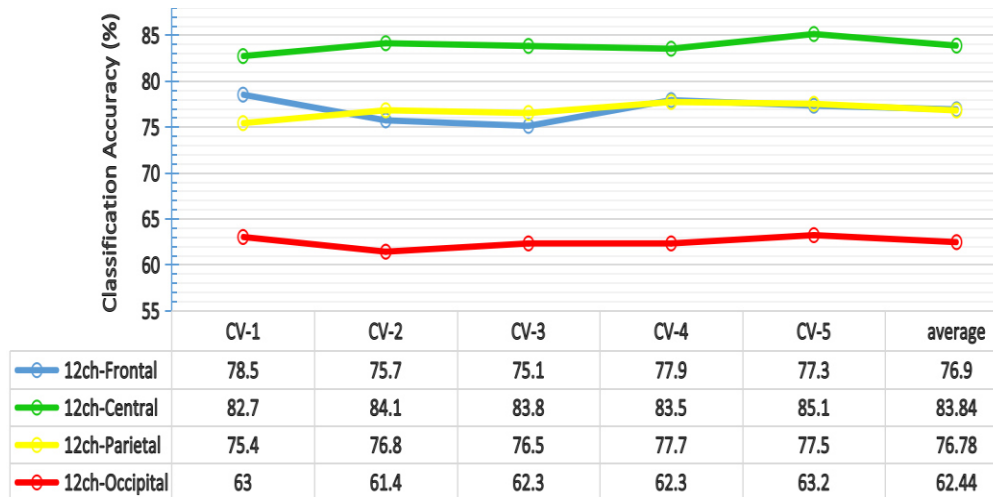
With reference to Figure 4.10(a) & 4.10(b), the relevance of different subsets of EEG channels for emotion recognition is investigated, with respect to the distance of the EEG channels from the midline.

The classification accuracies as reported in Figure 4.10 follow the trend that the more laterally-placed the EEG channels are, the higher the classification accuracy they deliver. This implies that more emotionally-distinctive information is carried in the laterally-placed (farther away from midline) EEG channels than the medially-placed channels.

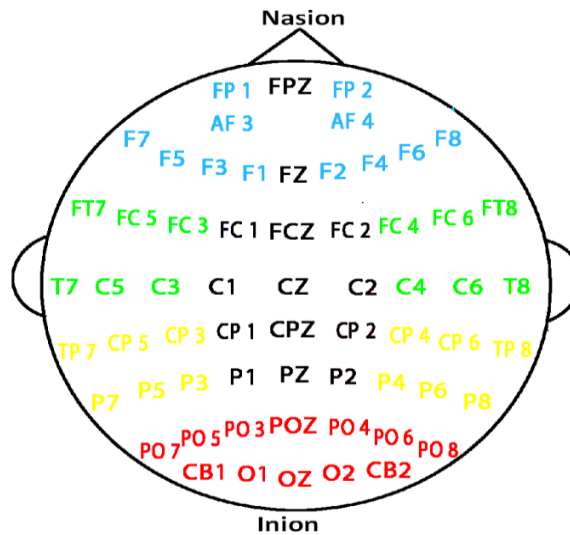
The possible reason for this channel significance distribution pattern is that the lateral channels are in fact placed over or close to the temporal region above the ears on the scalp. These electrode locations are closer to the brain structures that are highly involved in emotion response. These structures such as the anterior temporal pole, the insular cortex, the amygdala and the hippocampus (Dolan *et al.*, 2000; Dolcos *et al.*, 2005; Iidaka *et al.*, 2002) are either part of the temporal lobe itself or lying at just the medial side of the temporal lobe. Hence, the more medially-placed EEG electrodes are located higher up on top of the scalp and are hence farther away from these emotionally important brain structures.

4.3.3.2. Significance Along the Nasion-Inion Axis

Figure 4.13 shows the 5-fold cross-validated emotion classification accuracy of *ResNet18-1D(T-then-S)* model, using four different subsets of EEG channels along the nasion-inion axis.



(a) 5-fold cross validation classification accuracy



(b) Electrode placement

Figure 4.13: SEED 3-class Emotion Recognition Accuracy using Different Subsets of EEG Channels along the Nasion-Inion Axis

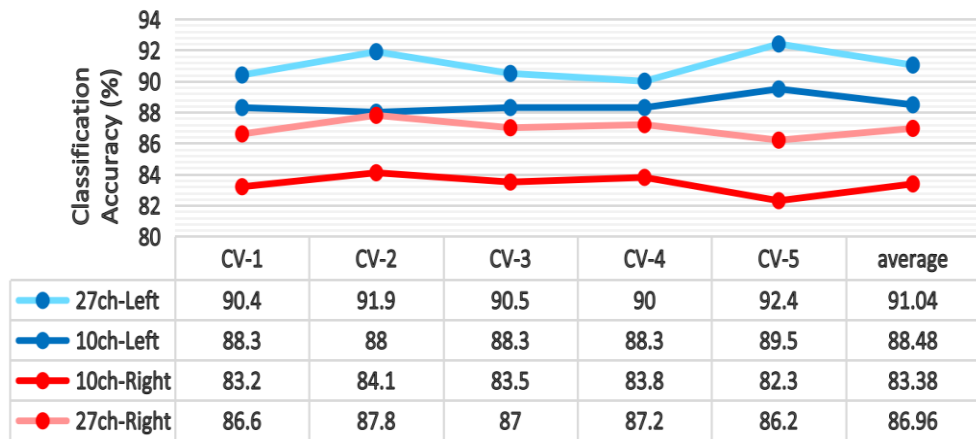
As indicated by Figure 4.13(b), these subsets of EEG channels respectively cover the frontal region (blue), centro-temporal region (green), centro-parietal region (yellow), and the parieto-occipital region (red).

In coherence with the distribution of emotionally important brain structures (e.g. the anterior temporal pole, the insular cortex and the amygdala) discussed above, the three different emotion classes are best classified with the twelve centro-temporal channels (green colour coded) because these twelve channels are located nearest to these structures, relative to the other three subsets.

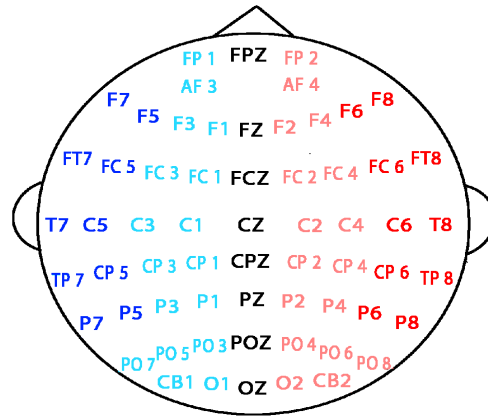
The twelve frontal channels give the same accuracy as that achieved by the twelve parietal channels. The occipital channels are the least emotionally-correlated set of EEG channels.

4.3.3.3. Cerebral Lateralization of Emotion

Figure 4.14 shows the 5-fold cross-validation accuracy using EEG channels of the left hemisphere versus right hemisphere. The left channels present around 4–5% higher accuracy than the right channels. Using only 10 lateral channels of the left hemisphere has resulted in 88.48% average accuracy which is still even better than using all 27 right hemispheric channels which gives 86.96%.



(a) 5-fold cross validation classification accuracy



(b) Electrode placement

Figure 4.14: SEED 3-class Emotion Recognition Accuracy Comparison using Left and Right Hemispheric EEG Channels

This lateralized significance of EEG channels in emotion recognition can be due to the fundamental cerebral lateralization (Corballis, 2014; Liu *et al.*, 2009) or simply because of the nature of SEED experiment design.

The stimuli of the SEED experiment were movie clips and the mode of content delivery of movies can be heavily verbal or language-based. The center

of language processing and understanding is located exactly in the lateral side of the left temporal lobe, known as Wernicke's area (DeWitt and Rauschecker, 2013). Therefore, the imbalanced activation of the Wernicke's area in comparison to its right hemispheric counterpart area can be a compounding factor resulting in the classification accuracy discrepancy.

4.3.3.4. Comparing across all the EEG Channel Subsets

Reviewing the classification results using various EEG channel subsets presented in Figure 4.10, Figure 4.13 and Figure 4.14, the ten lateral-most left-and-right combination of EEG channels in Figure 4.10 have achieved the highest accuracy (91.5%), compared to using the ten lateral left channels in Figure 4.14 which has achieved 88.48% recognition accuracy and the twelve centro-temporal channels in Figure 4.13 which has achieved 83.84% accuracy.

With comparable number of channels used in the subsets, the above result implies that there is additional distinctive information for emotion recognition retrievable from the left-vs-right channel feature cross-correlation, in view of the pairing of 10 left-and-right channels in Figure 4.10 gives better classification result than the 10 lateral-most left channels in Figure 4.14.

The highly emotion-correlated subsets of EEG channels identified by this work is close to the "12-channel (FT7, FT8, T7, T8, C5, C6, TP7, TP8, CP5,

CP6, P7, P8)” subsets used by [Zheng and Lu \(2015\)](#) which was reported to have achieved even higher emotion recognition accuracy than using all 62 channels. The results presented by [Zheng and Lu \(2015\)](#) were based on SVM classifier taking signal differential entropy of EEG as the input.

CHAPTER 5

CONCLUSION

5.1 CNN for EEG Classification

The CNN models in Study 1 have accurately classified the EEG data of the short-term music experiment which was reported to have insignificant statistical difference in a range of EEG signal features manually extracted by previous research. The temporal-spatial CNN model with 1D kernels has the average ten-fold cross-validation test accuracy of 97.46% in the binary classification task and average ten-fold cross-validation test accuracy of 95.71% in the three-class classification task.

The performance of the CNN for EEG classification is substantially affected by the constituent aspects of the model which include the amount of convolution kernels, the pooling method, the depth and width of the densely-connected perceptron network, and batch normalization.

With insufficient amount of convolution channels, the learning capacity of the CNN classifier is adversely affected, resulting in less accurate model representation of the target data domain and thus lower classification accuracy

as discussed in Section 4.1.1. On the other hand, the model with excessively large amount of convolutional kernels is prone to overfitting to the training dataset. The depth and width of the densely-connected perceptron network deliver a similar impact to the performance of the model where an excessively wide and deep network causes early overfitting and a densely-connected network without hidden layer faces difficulty in becoming a properly optimized representative of the data presented.

Pooling mechanism has beneficial impacts on both the computational efficiency and the classification performance of the CNN model. The pooling layers lessen the computational load by decreasing the data size to be processed, the time-dimensional pooling in our model has elevated the EEG classification validation accuracy from 70% to approximately 95%. The CNN classifier designed with no pooling layer in between the convolutional blocks has experienced substantial overfitting issue.

Regularization techniques that are adopted in this study have helped the training progression of the CNN models on EEG classification, which include the dropout mechanism and the batch normalization layers. Batch normalization layers have shown their essential role in handling the issue of convergence failure during model optimization process which is probably due to the internal covariate shift.

Replacing the 2D convolution kernels with separate 1D-temporal and 1D-spatial convolution kernels significantly reduce the trainable parameters in the CNN model and hence disc memory requirement for the model storage (as presented in both Study 1 and Study 3), yet preserving or even improving the performance of the CNN classifier on EEG signal classification. The reduction in the model size and disc memory demand can be crucial in the applications that are memory critical.

The CNN conventionally designed for processing image data (such as the *ResNet18* and *VGG16*) are not as effective in the task of EEG signal convolution. With kernel shape and arrangement adapted for the nature and format of EEG signals by replacing the 2D convolution kernels with separate 1D-temporal and 1D-spatial convolution kernels at desirable sequence, the CNN models in Study 1 as well as the modified variants of *ResNet18* in Study 3 have presented substantial performance elevation in the aspects of both the EEG classification accuracy and the disc memory requirement for the model storage due to the reduction in model parameters.

The modified versions of *ResNet18* have presented faster training convergence process than the modified *VGG16* models. The sequence of convolutional dimension arrangement within the CNN is also proven to be crucial for the performance of the classifier on EEG signal processing. The results obtained in Study 3 has advised against initiating the sequence of convolutional operation along spatial-dimension. Instead, inserting multiple

consecutive layers pure temporal-dimension convolution prior to the start of spatial-dimension convolution is highly recommended based on the performance comparison presented in Figure 4.10 and Figure 4.11 in Study 3. Working on the SEED public dataset, the top performing model (*ResNet18-1D-(T-then-S)*) among our CNNs for emotion recognition has attained the three-class emotion classification accuracy of 93.42%.

5.2 Non-uniform Influence of Music on Regional Brain Waves

Deducing from the discrepancy in the classification accuracies achieved using different partial sets of EEG channels from either the left or the right brain, the EEG signals emitted from the left cerebral hemisphere show greater difference with and without music, than the EEG signals emitted from the right cerebral hemisphere. The left-vs-right cerebral hemispheric asymmetry in the influence of music recorded in this project agrees with the concept of cerebral functional lateralization.

In addition, the EEG signals generated by the frontal lobes are even more distinctive with and without music, than that generated by the other cerebral (the temporal, parietal and occipital) regions.

5.3 Emotion-Relevance of Different EEG Channels

The significance of the subsets of EEG channels for the task of emotion recognition has also been investigated. The laterally-distributed EEG channels deliver higher emotion recognition accuracy compared to the medially-distributed channels. The EEG channels located near to the temporal lobes exhibit greater importance compared to the channels over other regions. Emotion classification using the left-sided EEG channels has also resulted in higher accuracy than using the channels from the right cerebral hemisphere.

On the whole, the laterally-distributed EEG channels over the temporal lobe are of the highest importance. This agrees with the well-established fact that many emotion-related brain structures are located within or close to the temporal lobes.

5.4 Recommendation for Future Work

The high classification accuracy achieved in the emotion recognition studies in this project is subject-dependent. The emotion-recognition CNN presented may not perform well in cross-subject and cross-database validation. The cross-database performance of the CNN is likely affected by the limited amount of EEG data available with insufficient signal variation representative of that in other databases. Future work with further collection of emotion-labelled EEG data using standardized recording setting is highly recommended. This contribution to the quantity of publicly available EEG datasets is highly

valuable to the development of CNN models with plain EEG signals as the input data.

Further investigation into the architectural improvement of CNN for EEG classification in the future may include the development of three-dimensional CNN which accommodates for not only the temporal dimension and 1D spatial dimension. With higher-density EEG recording headsets, the EEG-channel space could be organized as a two-dimensional space which allow for the construction of 3D data input for the CNN. In addition, the frequency dimension can be constructed as an addition dimension which is potentially useful for further performance improvement.

REFERENCES

- Al-Nafjan, A., Hosny, M., Al-Wabil, A. and Al-Ohali, Y., 2017. Classification of Human Emotions from Electroencephalogram (EEG) Signal using Deep Neural Network. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 8(9).
- Ansari-Asl, K., Chanel, G. and Pun, T., 2007. A Channel Selection Method For EEG Classification in Emotion Assessment Based on Synchronization Likelihood. *15th European Signal Processing Conference (EUSIPCO)*, 3-7 September 2007, Poznan, Poland. IEEE, pp. 1241-1245.
- Bartosova, M. et al., 2019. Emotional stimuli candidates for behavioural intervention in the prevention of early childhood caries: a pilot study. *BMC Oral Health*, 19, pp. 33.
- Behncke, J., Schirrmeyer, R.T., Burgard, W., Ball, T., 2018. The signature of robot action success in EEG signals of a human observer: Decoding and visualization using deep convolutional neural networks. *2018 6th International Conference on Brain-Computer Interface (BCI)*, 15-17 January 2018, Gangwon, South Korea. IEEE, pp. 1-6.
- Bjorck, J., Gomes, C., Selman, B. and Weinberger, K.Q., 2018. Understanding Batch Normalization. *ArXiv: 1806.02375v4 [cs.LG]*.
- Bradley, M.M. and Lang, P. J., 1994. Measuring Emotion: The Self Assessment Manikin and the Semantic Differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), pp. 49-59.
- Buzsáki, G., Anastassiou, C.A. and Koch, C., 2012. The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience*, 13(6), pp. 407-420.
- Chambon, S., Galtier, M.N., Arnal, P.J., Wainrib, G. and Gramfort, A., 2017. A Deep Learning Architecture for Temporal Sleep Stage Classification Using Multivariate and Multimodal Time Series. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(4), pp. 758-769.
- Chang, C.C. and Lin, C.J., 2011. {LIBSVM}: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3), pp. 27:1-27:27.

Cheah, K.H., Nisar, H., Yap, V.V. and Lee C-Y., 2019a. Convolutional neural networks for classification of music-listening EEG: comparing 1D convolutional kernels with 2D kernels and cerebral laterality of musical influence. *Neural Computing and Applications*, 32, pp. 8867-8891.

Cheah, K.H., Nisar, H., Yap, V.V. and Lee C-Y., 2019b. Short-time-span EEG-based personalized emotion recognition with deep convolutional neural network. *2019 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, 17-19 September 2019, Kuala Lumpur, Malaysia. IEEE, pp. 78-83.

Citron, F.M.M., Gray, M.A., Critchley, H.D., Weekes, B.S. and Ferstl, E.C., 2014. Emotional valence and arousal affect reading in an interactive way: Neuroimaging evidence for an approach-withdrawal framework. *Neuropsychologia*, 56(100), pp. 79-89.

Corballis, M.C., 2014. Left brain, right brain: facts and fantasies. *PLoS Biology*, 12(1), pp. e1001767.

Crow, T.J., Crow, L.R., Done, D.J. and Leask, S., 1998. Relative hand skill predicts academic ability: global deficits at the point of hemispheric indecision. *Neuropsychologia*, 36(12), pp. 1275-1282.

Delorme, A. and Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*, 134, pp. 9-21.

DeWitt, I. and Rauschecker, J.P., 2013. Wernicke's area revisited: Parallel streams and word processing. *Brain and Language*, 127(2), pp. 181-191.

Dolan, R., Lane, R., Chua, P. and Fletcher, P., 2000. Dissociable Temporal Lobe Activations during Emotional Episodic Memory Retrieval. *NeuroImage*, 11(3), pp. 203-209.

Dolcos, F., LaBar, K.S. and Cabeza, R., 2005. Remembering one year later: Role of the amygdala and the medial temporal lobe memory system in retrieving emotional memories. *Proceedings of the National Academy of Sciences (PNAS)*, 102(7), pp. 2626-2631.

Duan, R.-N., Zhu, J.-Y. and Lu, B.-L., 2013. Differential Entropy Feature for EEG-based Emotion Classification. *Proceeding of the 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, 6-8 November 2013, California, USA. IEEE, pp. 81-84.

Eckart, C. and Young, G., 1936. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3), pp. 211-218.

Foreman, B. and Claassen, J., 2012. Quantitative EEG for the detection of brain ischemia. *Critical Care*, 16, pp. 216.

Good, I.J., 1956. Some terminology and notation in information theory. *Proceedings of the IEE Part C: Monographs*, 103(3), pp. 200.

Goodfellow, I., Bengio, Y. and Courville, A., 2016a. Introduction. In: Goodfellow, I., Bengio, Y. and Courville, A. (eds.). *Deep Learning*. Cambridge: MIT Press, pp. 1-26.

Goodfellow, I., Bengio, Y. and Courville, A., 2016b. Convolutional Networks. In: Goodfellow, I., Bengio, Y. and Courville, A. (eds.). *Deep Learning*. Cambridge: MIT Press, pp. 326-366.

Goodfellow, I., Bengio, Y. and Courville, A., 2016c. Deep Feedforward Networks. In: Goodfellow, I., Bengio, Y. and Courville, A. (eds.). *Deep Learning*. Cambridge, MIT Press, pp. 164-223.

Graves, A., Fernandez, S., Liwicki, M., Bunke, H. and Schmidhuber, J., 2008. Unconstrained online handwriting recognition with recurrent neural networks. *NIPS'07: Proceedings of the 20th International Conference on Neural Information Processing Systems*, 3-5 December 2007, Vancouver, Canada. New York: Curran Associates Inc., pp. 577-584.

Graves, A., Mohamed, A. and Hinton, G., 2013. Speech recognition with deep recurrent neural networks. *2013 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 26-31 May 2013, Vancouver, Canada. IEEE, pp. 6645-6649.

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 27-30 June 2016, Las Vegas, USA. IEEE, pp. 770-778.

Headset Comparison Chart: Technical Specification n.d., viewed 20 July 2020 <<https://www.emotiv.com/comparison/>>.

Hofmeijer, J. and van Putten, M.J., 2016. EEG in postanoxic coma: Prognostic and diagnostic value. *Clinical Neurophysiology*, 127(4), pp. 2047-2055.

Höller, Y, Helmstaedter, C. and Lehnertz, K., 2018. Quantitative Pharmacoelectroencephalography in Antiepileptic Drug Research. *CNS Drugs*, 32(9), pp. 839-848.

Hsu, C.W., Chang, C.C. and Lin, C.J., 2016. *A practical guide to support vector classification*. NTU Department of Computer Science and Information Engineering Web, Department of Computer Science of National Taiwan University, viewed 10 Aug 2018, <<https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>>.

Hwang, S., Hong, K., Son, G. and Byun, H., 2019. Learning CNN features ,from DE features for EEG-based emotion recognition. *Pattern Analysis and Applications*, 23, pp. 1323-1335.

Iidaka, T. et al., 2002. Age-related differences in the medial temporal lobe responses to emotional faces as revealed by fMRI. *Hippocampus*, 12(3), pp. 352-362.

Ioffe, S. and Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *ArXiv*: 1502.03167 [cs.LG].

Jenke, R., Peer, A. and Buss, M., 2014. Feature extraction and selection for emotion recognition from EEG. *IEEE Transactions on Affective Computing*, 5(3), pp. 327-339.

Jiao, Y. et al., 2019. Sparse group representation model for motor imagery EEG classification. *IEEE Journal of Biomedical and Health Informatics*, 23(2), pp. 631-641.

Jin, Z.C., Zhou, G.X., Gao, D.Q. and Zhang, Y., 2020. EEG classification using sparse Bayesian extreme learning machine for brain-computer interface. *Neural Computing and Applications*, 32, pp. 6601-6609.

Kaplan, P.W., 2007. EEG criteria for nonconvulsive status epilepticus. *Epilepsia*, 48(s8), pp. 39-41.

Khan, A., Sohail, A., Zahoora, U. and Qureshi, A.S., 2020. A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, 53, pp. 5455–5516.

Kingma, D.P. and Ba, J.L., 2015. Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations (ICLR)*, 7-9 May 2015, California, USA. ArXiv: 1412.6980

Koelstra, S. et al., 2012. DEAP: A Database for Emotion Analysis ;Using Physiological Signals. *IEEE Transactions on Affective Computing*, 3(1), pp. 18-31.

Kreitzer, N, Huynh, M. and Foreman, B., 2018. Blood Flow and Continuous EEG Changes during Symptomatic Plateau Waves. *Brain Sciences*, 8(1), pp. 14.

Kushner, H.I., 2011. Retraining left-handers and the aetiology of stuttering: the rise and fall of an intriguing theory. *Laterality: Asymmetries of Body, Brain and Cognition*, 17(6), pp. 673-693.

Kwak, N.S., Müller, K.R. and Lee, S.W., 2017. A convolutional neural network for steady state visual evoked potential classification under ambulatory environment. *PLOS ONE*, 12(2), pp. e0172578.

Lan, Z., Sourina, O., Wang, L., Scherer, R. and Müller-Putz, G.R., 2018. Domain Adaptation Techniques for EEG-based Emotion Recognition: A Comparative Study on Two Public Datasets. *IEEE Transactions on Cognitive and Developmental Systems*, 11(1), pp. 85-94.

Li, X. et al., 2016. Emotion recognition from multi-channel eeg data through convolutional recurrent neural network. *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 15-18 Dec. 2016, Shenzhen, China. IEEE, pp. 352-359.

Li, Y., Zheng, W., Cui, Z., Zong, Y. and Ge, S., 2019. EEG Emotion Recognition Based on Graph Regularized Sparse Linear Regression. *Neural Processing Letters*, 49, pp. 555-571.

Liu, H., Stufflebeam, S.M., Sepulcre, J., Hedden, T. and Buckner, R.L., 2009. Evidence from intrinsic activity that asymmetry of the human brain is controlled by multiple factors. *Proceedings of the National Academy of Sciences (PNAS)*, 106(48), pp. 20499-20503.

Liu, N. et al., 2018. Multiple feature fusion for automatic emotion recognition using EEG signals. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 15-20 April 2018, Calgary, Canada. IEEE, pp. 896-900.

Liu, S. and Deng, W., 2015. Very deep convolutional neural network based image classification using small training sample size. *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, 3-6 November 2015, Kuala Lumpur, Malaysia. IEEE, pp. 730-734.

Liu, W., Zheng, W.L. and Lu, B. L., 2016. Emotion recognition using multimodal deep learning. In: Hirose, A. et al. (eds.). *Neural Information Processing. ICONIP 2016. Lecture Notes in Computer Science*, vol 9948. Cham, Switzerland: Springer, pp. 521-529.

Luo, Y. et al., 2020. EEG-Based Emotion Classification Using Spiking Neural Networks. *IEEE Access*, 8, pp. 46007-46016.

Maji, P. and Mullins, R., 2018. On the Reduction of Computational Complexity of Deep Convolutional Neural Networks. *Entropy*, 20(4), pp. 305.

Manor, R. and Geva, A.B., 2015. Convolutional Neural Network for Multi-Category Rapid Serial Visual Presentation BCI. *Frontiers in Computational Neuroscience*, 9, pp. 146.

Moinnereau, M. et al., 2018. Classification of auditory stimuli from EEG signals with a regulated recurrent neural network reservoir. *ArXiv*: 1804.10322

Muhlhofer, W. and Szaflarski, J.P., 2018. Prognostic Value of EEG in Patients after Cardiac Arrest-An Updated Review. *Current Neurology and Neuroscience Reports*, 18(4), pp. 16.

Nawaz, R., Nisar, H. and Yap, V.V., 2018. The effect of music on human brain: frequency domain and time series analysis using electroencephalogram. *IEEE Access*, 6:45191–45205.

Ng, J.Y. et al., 2015. Beyond short snippets: deep networks for video classification. *2015 IEEE conference on computer vision and pattern recognition (CVPR)*, 7-12 June 2015, Boston, USA. IEEE, pp. 4694-4702.

Orr, K.G.D., Cannon, M., Gilvarry, C.M., Jones, P.B., Murray, R.M., 1999. Schizophrenic patients and their first-degree relatives show an excess of mixed-handedness. *Schizophrenia Research*, 39(3), pp. 167-176.

Ozerdem, M.S. and Polat, H., 2017. Emotion recognition based on EEG features in movie clips with channel selection. *Brain Informatics*, 4(4), pp. 241-252.

Petrantonakis, P.C. and Hadjileontiadis, L. J., 2010. Emotion recognition from EEG using higher order crossings. *IEEE Transactions on Information Technology in Biomedicine*, 14(2), pp. 186-197.

Phneah, S.W. and Nisar, H., 2017. EEG-based alpha neurofeedback training for mood enhancement. *Australasian Physical and Engineering Sciences in Medicine*, 40(2), pp. 325–336.

Rayatdoost, S. and Soleymani, M., 2018. 2018 IEEE 28th Cross-corpus EEG-based emotion recognition. *2018 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 17-20 September 2018, Aalborg, Denmark. IEEE, pp. 1-6.

Ren, Y. and Wu, Y., 2014. Convolutional deep belief networks for feature extraction of EEG signal. *2014 International Joint Conference on Neural Networks (IJCNN)*, 6-11 July 2014, Beijing, China. IEEE, pp. 2850-2853.

Rodriguez, A. et al., 2010. Mixed-handedness is linked to mental health problems in children and adolescents. *Pediatrics*, 125(2), pp. e340-e348.

Rozgić, V., Vitaladevuni, S.N. and Prasad, R., 2013. Robust EEG emotion classification using segment level decision fusion. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 26-31 May 2013, Vancouver, Canada. IEEE, pp. 1286-1290.

Schirrneister, R.T. et al., 2017. Deep learning with convolutional neural networks for EEG decoding and visualization. *Humam Brain Mapping*, 38(11), pp. 5391-5420.

Siddiqui, S.V., Chatterjee, U., Kumar, D., Siddiqui, A. and Goyal, N., 2008. Neuropsychology of prefrontal cortex. *Indian Journal of Psychiatry*, 50(3), pp. 202-208.

Smith, S.J.M., 2005a. EEG in the diagnosis, classification, and management of patients with epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry*, 76, pp. ii2-ii7.

Smith, S.J.M., 2005b. EEG in neurological conditions other than epilepsy: when does it help, what does it add? *Journal of Neurology, Neurosurgery & Psychiatry*, 76, pp. ii8-ii12.

Song, T., Zheng, W., Song, P. and Cui, Z., 2018. EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks. *IEEE Transactions on Affective Computing*, 11(3), pp. 532-541.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(2014), pp. 1929-1958.

St. Louis, E.K. and Frey, L.C., 2016. *Electroencephalography: An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants* [Internet]. Chicago: American Epilepsy Society.

Stober, S., Cameron, D.J. and Grahn, J.A., 2014. Using convolutional neural networks to recognize rhythm stimuli from electroencephalography recordings. *NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems*, 8-13 December 2014, Montréal, Canada. Cambridge: MIT Press, pp. 1449-1457.

Swatzyna, R.J., Kozlowski, G.P. and Tarnow, J.D., 2015. Pharmaco-EEG: A Study of Individualized Medicine in Clinical Practice. *Clinical EEG & Neuroscience*, 46(3), pp. 192-196.

Tripathi, S., Acharya, S., Sharma, R.D., Mittal, S. and Bhattacharya, S., 2017. Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset. *AAAI'17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 4-9 February 2017, California, USA. AAAI Press, pp. 4746-4752.

Wang, Y. et al., 2019. EEG-Based Emotion Recognition with Similarity Learning Network. *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 23-27 July 2019, Berlin, Germany. IEEE, pp. 1209-1212.

Yang, F., Zhao, X., Jiang, W., Gao, P. and Liu, G., 2019. Multi-method Fusion of Cross-Subject Emotion Recognition Based on High-Dimensional EEG Features. *Frontiers in Computational Neuroscience*, 13, pp. 53.

Yang, Y., Wu, Q.M.J., Zheng, W.-L. and Lu, B.-L., 2017. EEG-based emotion recognition using hierarchical network with subnetwork nodes. *IEEE Transactions on Cognitive and Developmental Systems*, 10(2), pp. 408-419.

Yazdani, A., Lee, J-S., Vesin, J-M. and Ebrahimi, T., 2012. Affect Recognition Based on Physiological Changes During the Watching of Music Video. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2(1), pp. 7.

Yoon, H.J. and Chung, S. Y., 2013. EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm. *Computers in Biology and Medicine*, 43(12), pp. 2230-2237.

Yu, F. and Koltun, V., 2016. Multi-scale context aggregation by dilated convolutions. *4th International Conference on Learning Representations (ICLR 2016)*, 2-4 May 2016, San Juan, Puerto Rico. ArXiv: 1511.07122

Zafar, R., Dass, S.C. and Malik, A.S., 2017. Electroencephalogram-based decoding cognitive states using convolutional neural network and likelihood ratio based score fusion. *PLOS ONE*, 12(5), pp. e0178410.

Zhang, T., Cui, Z., Xu, C., Zheng, W. and Yang, J., 2020. Variational Pathway Reasoning for EEG Emotion Recognition. *34th AAAI Conference on Artificial Intelligence*, 7-12 February 2020, New York, USA. New York: AAAI Press, pp. 2709-2716.

Zhang, T., Wang, X., Xu, X. and Chen, C.L.P., 2019. GCB-Net: Graph Convolutional Broad Network and Its Application in Emotion Recognition. *IEEE Transactions on Affective Computing*. [Early Access] doi:10.1109/taffc.2019.2937768

Zhang, X., Hu, B., Chen, J. and Moore, P., 2013. Ontology-based context modeling for emotion recognition in an intelligent web. *World Wide Web*, 16(4), pp. 497-513.

Zhang, Y. et al., 2016. Sparse Bayesian classification of EEG for brain-computer interface. *IEEE Transactions on Neural Networks and Learning Systems*, 27(11), pp. 2256-2267.

Zheng, W.-L. and Lu, B.-L., 2015. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development (IEEE TAMD)*, 7(3), pp. 162-175.

Zheng, W., Zhu, J., and Lu, B., 2017. Identifying Stable Patterns over Time for Emotion Recognition from EEG. *IEEE Transactions on Affective Computing*, 10(3), pp. 417-429.