

# Kent Academic Repository

## Full text document (pdf)

### Citation for published version

MZURIKWAO, DEOGRATIAS (2020) Application of Deep Neural Network in Healthcare data. Doctor of Philosophy (PhD) thesis, University of Kent,.

### DOI

### Link to record in KAR

<https://kar.kent.ac.uk/87075/>

### Document Version

UNSPECIFIED

#### Copyright & reuse

Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

#### Versions of research

The version in the Kent Academic Repository may differ from the final published version.

Users are advised to check <http://kar.kent.ac.uk> for the status of the paper. **Users should always cite the published version of record.**

#### Enquiries

For any further enquiries regarding the licence status of this document, please contact:

[researchsupport@kent.ac.uk](mailto:researchsupport@kent.ac.uk)

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at <http://kar.kent.ac.uk/contact.html>

# **Application of Deep Neural Network in Healthcare data**

A thesis submitted for the degree of

Doctor of Philosophy

By

**Deogratias Mzurikwao**



University of Kent

School of Engineering and Digital Arts

United Kingdom

16/02/2021

Word Count: 50532

# Abstract

Biomedical data analysis has been playing an important role in healthcare provision services. For decades, medical practitioners and researchers have been extracting and analyse biomedical data to derive different health-related information. Recently, there has been a significant rise in the amount of biomedical data collection. This is due to the availability of biomedical devices for the extraction of biomedical data which are more portable, easy to use and affordable, as an effect technology advancement. As the amount of biomedical data produced every day increases, the risk of human making analytical and diagnostic mistakes also increases. For example, there are approximately 40 million diagnostic errors involving medical imaging annually worldwide, hence rise a need for the development of fast, accurate, reliable and automatic means for analysis of biomedical data. Conventional machine learning has been used to assist in the analysis and interpretation of biomedical data automatically, but always limited with the need for feature extraction process to train the built models.

To achieve this, three studies have been conducted. Two studies were conducted by using EEG signals and one study by using microscopic images of cancer cells. In the first study with EEG signals, our method managed to interpret motor imaginary activities from a 64 channels EEG device with 99% classification accuracy when all the 64 channels were used and 91.5% classification when the number of channels was selected to eight (8) channels. In a second study which involved steady-state visual evoked potential form of EEG signals, our method achieved an average of 94% classification accuracy by using two channels, skin like EEG sensor. In the third study for authentication of cancer cell lines by using microscopic images, our method managed to attain an average of 0.91 F1-score in the authentication of eight classes of cancer cell lines.

Studies reported in this thesis, significantly shows that CNN can play a major role in the

development of a computerised way in the analysis of biomedical data. Towards provision of better healthcare by using CNN in analysis of different formats of biomedical data, this thesis has three major contributions, i) introduction of a new method for EEG channels selection towards development of portable EEG sensors for real-life application, and ii) introduction of a method for cancer cell lines authentication in the laboratory environment towards development of anti-cancer drugs, and iii) Introduction of a method for authentication of isogenic cancer cell lines.



# Acknowledgements

I would like to express my sincere thanks to my supervisor, Dr. Chee Siang Ang, for your advice, support, patience, and encouragement throughout my Ph.D. journey. Reaching this stage has been a long road, and your inspiration and guidance throughout this time have been invaluable. The transformation you have made on me is of great value, I will forever be grateful. I would also like to express my thanks to the whole supervisory team for their guidance throughout this Ph.D. studies, Dr. Christos Efstratiou and Dr. Gianluca Marcelli, at different points, I have learned a lot from you.


Above all, I would like to thank my family. My wife, Naomi who has been extremely supportive of me throughout this entire process and has made countless sacrifices to help me get to this point. My children, Abigail and Xavier, who have been a source of motivation and continually provided the requisite breaks from philosophy, and for keeping me sane. But most of all, thank you for being my best friends. I owe you everything. Special thanks to my mum, Maria, my wife's family, and those we have lost especially my late dad, Bonaventure, and my late son, Asahel.

Finally, Special thanks to all my collaborators, Prof Samuel Oluwarotimi Williams, Mr. Musa Mahmood, and Mr. Muhammad Usman Khan. Also, I would like to thank my friends and colleagues from the School of Engineering and Digital Arts at the University of Kent for providing a friendly work environment: James Alexander Lee, Jon Baker, Israel Mensah, Ben Nicholls, Jitrapol Intarasirisawat and Luma Tabbaa. Without forgetting Church members at the umbrella Centre, Canterbury, Pastor Goodwill, and his lovely family and madam Hilda with her great family. Lastly, thanks to Mr. Thomas Neumark for proofreading some of the chapters of this thesis, you have been helpful.

# Declaration

I declare that this thesis titled: ‘Application of Deep Neural Networks in Healthcare data’ and the work presented within is, unless otherwise stated, my own. I confirm that this work was done wholly or mainly while in candidature for a research degree at the University of Kent, in the United Kingdom. No part of this thesis has previously been submitted for any other academic award, at this institution or otherwise. I declare that when consulting and referring to the works of others, I have always clearly attributed said works by specific reference. Where this thesis is based on work done by myself in collaboration with others, I have made it clear within chapters of this thesis.

Author: Deogratias Mzurikwao

Signed: 

Date: 16/02/2021

## Table of Contents

<b>Abstract</b> .....	<b>2</b>
<b>Acknowledgements</b> .....	<b>4</b>
<b>Declaration</b> .....	<b>5</b>
Contents .....	<b>Error! Bookmark not defined.</b>
<b>List of Figures</b> .....	<b>9</b>
<b>List of Tables</b> .....	<b>10</b>
<b>List of Acronyms</b> .....	<b>10</b>
<b>Chapter 1</b> .....	<b>12</b>
<b>Introduction</b> .....	<b>12</b>
1.1 Conventional machine learning in BCI applications.....	16
1.2 Conventional machine learning in biomedical images.....	17
1.3 Deep learning with biomedical data.....	19
1.4 Research challenges of biomedical data analysis.....	20
1.4.1 Challenges of EEG based BCI.....	20
1.4.2 Challenges of availability of medical images .....	22
1.5 Research questions.....	22
1.6 Proposed solutions for applications of CNN for analysis of biomedical data.....	23
1.6.1 CNN on the EEG channel selection.....	23
1.6.2 Possible solutions on the lack of enough biomedical images.....	24
1.7 Thesis structure.....	24
1.8 Contribution.....	26
1.9 Publications.....	26
1.10 Thesis structure .....	27
<b>Chapter 2</b> .....	<b>30</b>
<b>Literature review</b> .....	<b>30</b>
2.1 Time series Biomedical data.....	33
2.1.1 Background of Electroencephalography (EEG).....	33
2.1.2 EEG-based Brain to Computer Interfaces (BCI):.....	35
2.1.2.1 Introduction to BCI .....	35
2.1.2.2 EEG for BCI .....	36
2.1.2.3 Paradigms of EEG for BCI applications.....	37
2.1.2.4 Motor Imaginary (MI) Activities for BCI .....	38
2.1.2.5 Steady-state visual evoked potential (SSVEP) .....	40
2.2 Biomedical images.....	41
2.2.1 Background of biomedical images.....	42
2.2.2 Categories of biomedical images.....	43
2.2.2.1 Medical imaging .....	43
I. Radioactive medical imaging .....	44
II. Non-radioactive medical imaging .....	44
2.2.2.2 Microscopic imaging.....	44
2.2.3 Limitations of manual biomedical images analysis.....	44
2.3 Biomedical data analysis with conventional machine learning .....	46
2.3.1 Introduction to conventional machine learning.....	46
2.3.2 EEG with conventional machine learning for BCI applications .....	47
2.3.3 Conventional machine learning in Microscopic images .....	48

2.3.4	Limitations of computerised biomedical data analysis .....	50
2.4	Limitations of current existing EEG devices for real-life applications of BCI .....	52
2.5	EEG Channels selection .....	53
2.5.1	Filtering method .....	53
2.5.2	Wrapper method .....	53
2.5.3	Hybrid method .....	54
2.5.4	Embedded method .....	54
2.6	Deep learning .....	55
2.6.1	Introduction to Convolutional Neural Network (CNN).....	56
2.6.1.1	Background of CNN.....	56
2.6.1.2	CNN structure.....	56
2.7	CNN in biomedical data analysis.....	59
2.7.1	CNN with EEG for BCI applications.....	59
2.7.2	CNN for EEG channels selection .....	60
2.7.3	CNN in analysis of biomedical images.....	61
2.7.4	Challenges of using CNN in biomedical data analysis .....	61
2.7.5	Possible solutions to the limitation of data.....	62
	Summary .....	63

**Chapter 3..... 65**

***Convolutional Neural Network for interpretation of EEG signals and channels selection for BCI applications..... 65***

3.1	Introduction .....	65
3.2	Study 1: CNN in interpretation Motor Imaginary activities for BCI.....	66
3.2.1	EEG Data from Amputees.....	67
3.2.1.1	Study description.....	67
3.2.1.2	Equipment setup and data acquisition .....	67
3.2.2	Data Analysis.....	69
3.2.2.1	Data pre-processing.....	69
3.2.2.2	CNN architecture.....	70
3.2.3	Classification results .....	71
3.2.4	Weight analysis.....	73
3.2.4.1	Topographic maps .....	73
3.2.4.2	Channel reduction and performance analysis.....	74
3.2.5	Evaluation of our method on the publicly available data set .....	77
3.3	Study 2: CNN in interpretation SSVEP EEG signals for BCI.....	78
3.3.1	Determining significant Channels and scalp locations with a 32 channels EEG device .....	79
3.3.1.1	Data collection .....	79
3.3.1.2	Methodology.....	80
I.	Equipment setup.....	80
II.	Data acquisition .....	80
III.	Model training.....	81
3.3.1.3	Weights analysis.....	81
3.3.2	2-Chanel EEG device development.....	82
3.3.2.1	2 Chanel EEG data acquisition.....	84
3.2.2.2	Data pre-processing and model training.....	85
3.4	Discussions and Conclusion .....	87

**Chapter 4..... 90**

***Image-based Cancer Cell Lines Authentication Using Deep Neural Networks..... 90***

4.1	Introduction .....	91
4.2	Study description.....	93
4.2.1	MobileNet and InceptionResnet.....	93
4.2.2	Data sets .....	94
4.2.2.1	Cancer Cell Lines dataset .....	95
4.2.2.2	Breast Cancer Cells dataset.....	96

4.2.3	Data pre-processing.....	97
4.2.4	Pilot classification task.....	99
4.2.4.1	Training strategies.....	100
4.2.4.2	Performance measure metric.....	101
4.2.4.3	Pilot authentication results.....	102
4.2.5	Hyper-parameters tuning of the optimal Model.....	107
4.2.5.1	Resizing and cropping.....	107
4.2.5.2	Fully connected layers.....	109
4.2.5.3	Batch size.....	110
4.2.5.4	Multi-stage Transfer learning.....	111
4.3	Authentication using the selected model.....	113
4.3.1	Authentication stages.....	114
4.3.1.1	Four classes authentication.....	114
4.3.1.2	Eight classes authentication.....	117
4.3.1.3	Per class performance.....	117
4.3.1.4	Two-classes authentication task.....	118
4.3.2	Effects of sample size.....	120
4.3.4	Further investigation on EFO-21 and EFO-27.....	122
4.4	Summary.....	123
<b>Chapter 5.....</b>		<b>125</b>
<b>Discussion and Conclusion.....</b>		<b>125</b>
5.1	Discussion.....	125
5.1.1	CNN in computerised biomedical data analysis.....	126
5.1.1.1	The strength of CNN in automatic feature extraction.....	127
5.1.1.2	Addressing the “big data” problem in specialised domains.....	132
5.1.1.3	Role of CNN in complementing current practice and allowing new applications not possible previously...	136
5.1.2	Contributions and implications.....	137
5.1.2.1	Contributions.....	137
I.	CNN towards the implementation of real-world BCI applications.....	137
II.	Introducing a computerised and automatic way of authentication of cancer cell lines.....	138
III.	Additional parameters in the authentication of cancer cell lines.....	139
5.1.2.2	Implications to other research.....	140
5.1.2.3	Implications for practitioners.....	143
5.2	Conclusion.....	144
5.2.1	Limitations.....	144
5.2.2	Future research directions.....	146
5.2.3	Concluding remarks.....	148
<b>References.....</b>		<b>149</b>

# List of Figures

<b>Figure 1.1:</b> Thesis structure.....	25
<b>Figure 2.1:</b> Literature flow structure.....	32
<b>Figure 2.2:</b> Different techniques to acquire brain signal.....	36
<b>Figure 2.3:</b> EEG based BCI application system.....	37
<b>Figure 2.4:</b> Motor cortex region of the brain, marked in red.....	39
<b>Figure 2.5:</b> A laboratory set up to trigger motor imaginary potential.....	40
<b>Figure 2.6:</b> The occipital lobe region of the brain.....	41
<b>Figure 2.7:</b> A complete set up of SSVEP based EEG for BCI application.....	41
<b>Figure 2.8:</b> Process of training a conventional machine learning.....	50
<b>Figure 2.9:</b> Components of current existing EEG devices.....	52
<b>Figure 2.10:</b> An example of a CNN architecture.....	57
<b>Figure 3.1:</b> Experimental setup.....	68
<b>Figure 3.2:</b> EEG sample record using a 64 channel EEG device.....	69
<b>Figure 3.3:</b> A 2-layer CNN, with one fully connected layer and five class outputs.....	71
<b>Figure 3.4:</b> Learning curve for 0.5 second window length.....	72
<b>Figure 3.5:</b> Learning curve for 1-second window length.....	72
<b>Figure 3.6:</b> Learning curve for 2 seconds window length.....	72
<b>Figure 3.7:</b> Topographic maps.....	76
<b>Figure 3.8:</b> A study participant wearing 32 channels EEG devices.....	80
<b>Figure 3.9:</b> The topographic maps.....	82
<b>Figure 3.10:</b> 2-channels EEG system.....	84
<b>Figure 3.11:</b> Confusion matrix.....	86
<b>Figure 3.12:</b> Online confusing matrix.....	87
<b>Figure 4.1:</b> Cancer Cell Lines sample images.....	96
<b>Figure 4.2:</b> Breast cancer sample images.....	97
<b>Figure 4.3:</b> Image augmentation samples.....	98
<b>Figure 4.4:</b> Pilot authentication task.....	100
<b>Figure 4.5:</b> Drug treated cancer cell lines classification.....	104
<b>Figure 4.7:</b> Combined cancer cell lines classification model performance.....	105
<b>Figure 4.6:</b> Parental cancer cells classification model performance.....	105
<b>Figure 4.8:</b> Learning curves for model comparisons.....	106
<b>Figure 4.9:</b> Samples of resizing.....	109
<b>Figure 4.10:</b> Batch size learning curves.....	111
<b>Figure 4.11:</b> Single-stage and multi-stage training strategies.....	112
<b>Figure 4.12:</b> Confusion matrices of parental cancer cell lines.....	115
<b>Figure 4.13:</b> Confusion matrices of drug-treated cancer cell lines.....	116
<b>Figure 4.14:</b> ROC curves of combined cancer cell lines.....	119
<b>Figure 4.15:</b> Confusion matrices of combined cancer cell lines.....	120

## List of Tables

<b>Table 1.1:</b> Table of publications .....	27
<b>Table 3.1:</b> Model performance on different windows.....	71
<b>Table 3.2:</b> Conventional machine learning model’s performance.....	73
<b>Table 3.3:</b> Channel selection based on our acquired dataset.....	76
<b>Table 3.4:</b> Channel selection based on Physionet data .....	78
<b>Table 3.5:</b> Testing of different window sizes for each subject.....	86
<b>Table 3.6:</b> CNN comparison to SVM.....	87
<b>Table 4.1:</b> Number of images per cell line.....	95
<b>Table 4.2:</b> Training strategies selection process .....	101
<b>Table 4.3:</b> Model comparison.....	106
<b>Table 4.4:</b> Resizing and cropping .....	106
<b>Table 4.5:</b> Different architecture.....	110
<b>Table 4.6:</b> Batch size f1-score .....	111
<b>Table 4.7:</b> Four class authentication .....	116
<b>Table 4.8:</b> Eight classes authentication .....	117
<b>Table 4.10:</b> Two classes authentication.....	118
<b>Table 4.9:</b> Model performance per class.....	118
<b>Table 4.11:</b> 10 folds cross-validation with the training sample size drop.....	121
<b>Table 4.12:</b> Model confidence in classifying different classes.....	122
<b>Table 4.13:</b> Investigation on Efo-21 and Efo-27 .....	123
<b>Table 5.1:</b> CNN performance compared to conventional machine learning models.....	129
<b>Table 5.2:</b> CNN performance compared to conventional machine learning models.....	130

## List of Acronyms

Amyotrophic Lateral Sclerosis: (ALS) .....	16
Area Under the Curve: (AUC) .....	117
Artificial Intelligence: (AI) .....	14
Artificial Neural network: (ANN) .....	18
Brain to Computer Interaction: (BCI).....	15
Canonical Correlation Analysis: (CCA) .....	129
Class Activation Map: (CAM) .....	142
Computed Tomography: (CT).....	12
Convolutional Encoder Network: (CEN).....	20
Convolutional Neural Networks: (CNN) .....	16
Cross-Spectral Density Analysis: (CSDA) .....	87
Crystal Display: (LCD) .....	48
Deep Boltzmann: Machine (DBM) .....	131
Electrocardiogram: (ECG) .....	12

Electroencephalogram: (EEG).....	12
Electromyography: (EMG).....	12
Electrooculography: (EOG) .....	12, 36
Event-Related Desynchronisation: (ERD).....	41
Event-Related Potential: (ERP).....	41
Fast Fourier Transforms: (FFT).....	15
Fisher Linear Discriminant: (FLD).....	17
Functional Magnetic Resonance Imaging: (fMRI).....	39
Functional Near-Infrared Spectroscopy: (fNIRS) .....	39
General Data Protection Regulation: (GDPR) .....	133
Graphene-based epidermal sensor system: (GESS).....	14
Gray Level Co-Occurrence Matrix: (GLCM) .....	53
Joint Photographic Experts Group: (JPEG).....	96
K-Nearest Neighbour: (KNN) .....	52
Large Scale Visual Recognition Competition: (ILSVRC).....	56
Linear Discriminant Analysis: (LDA) .....	73
Long Short-Term Memory: (LSTM).....	55
Magnetic Resonance Imaging: (MRI).....	42
Magnetoencephalography: (MEG) .....	35
Motor Imaginary: (MI).....	36
Naive Bayes: (NB).....	48
Signal to Noise Ratio: (SNR).....	141
Phantom Limb Pain: (PLP) .....	142
Positron Electron Tomography: (PET) .....	39
Power Spectral Density: (PSD) .....	15
Power Spectral Density Analysis: (PSDA).....	87
Principle Component Analysis: (PCA).....	51
Random Forest: (RF) .....	20
Recursive Neural Network: (RNN) .....	61
Resistant Cancer Cell Line: (RCCL) .....	92
Restricted Boltzmann Machine: (RBM) .....	61
Short-Tandem Repeat: (STR).....	27
Slow Cortical Potential: (SCP).....	41
Standard Deviation: (SD).....	15
Steady-State Visual Evoked Potential: (SSVEP) .....	41
Support Vector Machine: (SVM) .....	17
Tuberculosis: (TB).....	13
Wavelet Transform: (WT).....	15
White Matter Hyperintensities: (WMH).....	20
World Health Organisation: (WHO).....	144



# Chapter 1

## Introduction

Modern healthcare services include all kinds of efforts made to maintain or restore physical, mental, or emotional well-being especially by trained and licensed healthcare professionals [1]. Efforts towards the improvement of healthcare can be delivered in different ways. This can be done by provision of effective interventions, early detection of diseases, through drug development for therapeutic or management purposes or by provision of quality care to the people in need like people with disabilities. Biomedical data, such as images and human physiological signals, plays an important role in healthcare provision services. Medical practitioners and researchers have been extracting and analysing human biomedical data to derive different health-related information. Analysis of human biomedical data has been essential in understanding physiological phenomena of a person [2], understanding how different organs work [3], the development of drugs and in the diagnosis of different diseases [4]. Human biomedical data can be extracted from both healthy people and patients, by using transducers in capturing physiological signals [5] or by using imaging devices such as medical imaging devices and microscope in capturing biomedical images [6].

The extracted biomedical data can be continuous signals like physiological signals such as Electroencephalogram (EEG), Electrocardiogram (ECG), Electrooculography (EOG) and Electromyography (EMG) or can be in the form of images like Magnetic Resonance Imaging (MRI), X-rays image, Computed Tomography (CT scans), microscopic images or ultrasound images. EEG signal is an electric signal produced by the brain when performing different activities as the brain cells communicate and mostly used to monitor the brain as it is non-invasive [7], MRI medical images are produced by using strong magnetic fields and waves [6],

while microscopic images are produced by using microscope devices [8].

The obtained biomedical data are conventionally analysed by medical practitioners or researchers to extract relevant and meaningful information such as diagnosis of diseases, its causes and a possible treatment [9]. It is through simple and rapid methods such as microscopic images analysis which has been used in the diagnosis of common tropical diseases like malaria in conventional hospital laboratories [10], to complex methods of medical imaging like MRI images which have been used in the detection of complex cases like brain injuries and signs of stroke [11]. Analysis of physiological signals like EEG signals has been significant in the diagnosis of brain-related diseases like epilepsy, dementia, stroke or brain tumour [12], or EMG (which is an electrical signal produced by muscular movement [13]) which has played a major role in identify neuromuscular diseases or disorders of motor control [14]. Conditions like bone fracture, some tumours, Tuberculosis (TB) [15], infections such as pneumonia and blood vessel blockages, have been detected by radiologists by the analysis of X-ray images [16]. Unlike X-ray which has known side effects, ultrasound imaging technology with no known side effects for over thirty (30) years of its application, has been used to find sources of diseases by visualising internal body structures such as blood vessels, muscles, joints and internal organs [17]. Although a doctor can screen a disease based on someone's symptoms, it is through analysis of the taken biomedical data like chest-X-ray which can distinguish between a lung infection and pneumonia which have similar symptoms of coughing [18]. Indeed, biomedical data analysis has been essential in the diagnosis of diseases, treatment of diseases and also follow up treatments [9].

However, the analysis of biomedical data by human analysts is time-consuming and highly dependent on the experience of the person analysing the data and therefore subjected to human errors [19], [20]. Recently, the capacity to collect biomedical data has been growing significantly due to the technological advancement and availability of cost-effective sensors and medical devices for production and recording of biomedical data in different formats

[21], [22]. For instance, there have been invention of more convenient, cheap, and stretchable skin-like EEG sensors such as the Graphene-based Epidermal Sensor System (GESS)[23], pocket size and affordable ultrasound machines [24], and advancement in MRI technology which produces high-quality images at lower costs [25]. The advancement in technologies in biomedical devices has resulted in massive availability of biomedical data which needs real time and accurate analysis and interpretation methods. In addition, the analysis and interpretation of biomedical data by human analysts has always been affected by sample sparsity, high dimensional feature space and noises associated with biomedical data [26]. Techniques introduced in this thesis can address those limitations. For example, by using data augmentation and transfer learning, sample sparsity issues can be addressed, use of CNN which doesn't need manual feature extraction can address the dimensionality nature of biomedical data features and the associated noises. Towards the provision of better healthcare services, there is a high demand and potential in the development of fast and accurate ways of interpretation and analysis of the obtained biomedical data due to the amount and nature of the biomedical data captured with latest technologies.

Thanks to the recent advancement in computing power, Artificial Intelligence (AI) has shown great potentials in providing a computerised way for interpretation and analysis of biomedical data. Machine learning, a subset of Artificial intelligence, is the ability of a computer program to learn from the provided data without being explicitly programmed [27]. AI has found applications in a variety of applications areas such as image-based object recognition [28], voice recognition [29] and computerised credit scoring system, and through these has revolutionised different fields such as transport, security and finance, just to mention a few. For a long time, the applications of AI in healthcare has been gaining interests, increasing recently due to advancement in computing power. In healthcare, researchers have explored the use of machine learning algorithms in the diagnosis of diseases, recommendations of treatment and in the analysis of electronic medical record [30]. To train

conventional machine learning algorithms, features need to be extracted from the training data manually. A manual process of extracting features from the training data is called feature engineering [31]. Depending on the application areas, different features have been studied and extracted from the raw data to train conventional machine learning models. For example, features like Standard Deviation (SD), variance, Fast Fourier Transforms (FFT) [32], Wavelet Transform (WT) [33], and Power Spectral Density (PSD) [34] have been extracted from EEG signals and used to train conventional machine learning models for Brain to Computer Interface (BCI) applications. For object recognition using images, features such as colour, texture, shape of the objects in the image have been extracted from the training image samples [35]. The feature extraction process can reduce machine's efforts in building variable combinations (features), hence facilitate the speed of learning and generalisation steps in the machine learning process.

Feature engineering is tiresome, complex, time-consuming and highly dependent on human experience in the field to be able to extract relevant and useful features [36]. A new class of machine learning algorithm called deep learning, which does not need manual feature engineering, provides an alternative to be used in the interpretation and analysis of biomedical data. Deep learning is a subset of the broader family of machine learning using artificial neural networks. A deep learning model has the capabilities of finding its own features from the provided training data. By training deep learning models with the obtained biomedical data, the models will be able to find and detect the learnt relevant features in the unseen data of a similar problem. This thesis explores the applications of deep Convolutional Neural Networks (CNN), a popular deep learning algorithm, in the healthcare domain (Chapter 3, Chapter 4 for details). We demonstrate the application of CNN in two separate case studies, with a focus on two most common forms of data in healthcare, the time series and image data, these studies are:

1. Application of CNN in interpretation electroencephalography (EEG) signals for

the brain-computer interface (BCI) applications to provide a communication pathway for people with disabilities (Chapter 3).

2. Application of CNN in the authentication of cancer cell lines using microscopic images, to understand resistant behaviour of cancer cell lines to anti-cancer drugs (Chapter 4).

## **1.1 Conventional machine learning in BCI applications.**

For individuals who cannot communicate via conventional means, either verbally or physically, as a result of severe motor disabilities (like spinal cord injuries or Amyotrophic Lateral Sclerosis (ALS) disease) or loss of limbs due to accidents or diseases, Brain to Computer Interface (BCI) may offer a possible means of re-establishing their communication capabilities. BCI is a technology that allows direct communication between the human brain and external devices or prosthesis [37].

For BCI applications using EEG signals, researchers have been interpreting the variations in the generated EEG signals. During the performance of different activities by a human, like the imaginary or actual performance of activities [38], exposure to different environmental conditions like humidity and temperature [39], or external stimuli like light or a sound [40], can lead to variations in the EEG signals generated by someone's brain. By interpreting variations on the generated EEG signal, one can extract relevant information in the EEG signal which corresponds with the actual event/action taking place. The interpreted activities from EEG signals can be sent to control external devices like prosthetic arm [5], wheelchair [41], computer games [42] in BCI applications. For interpretation the patterns of the complex EEG signals, different techniques have been explored, including machine learning.

It was until 1970s were EEG signals found application in BCI technologies by using sophisticated signal processing techniques for translation of the incoming signal into output command [43], since its invention in 1929 [38]. For many years, machine learning has

become a core part of BCI for translation of the EEG signal [44]. Conventional machine learning has been used to interpret and extract meaningful and useful information from the captured EEG signals for different applications such as detection of emotion status of an individual [45], BCI applications like control of exoskeleton to support heavy lifting and for people with disabilities [22], and seizure detection [46]. Patterns can be found in the recorded EEG signals which are relevant to the activity taking place during the recording. Several studies have used machine learning models to interpret EEG signals [47]. For instance, machine learning algorithms including Fisher Linear Discriminant (FLD) and Support Vector Machine (SVM), to interpret and classify motor imagination based on EEG signals to control a humanoid robot [7]. Apart from the limitations of having to manually extract features from raw training data to train a conventional machine learning, noise interference from the functioning of other human organs activities like eye movement has been affecting EEG signals in particular and makes feature extraction process more challenging.

## **1.2 Conventional machine learning in biomedical images.**

Another important form of data in healthcare is images, which involves both medical imaging (such as MRI, MRA, Ultrasound and X-rays) and microscopic images. The process of medical imaging seeks to visualise the internal parts of the body hidden under the skin, bones or other organs. It is also used to reveal details which can't be seen by normal human eyes due to their sizes or complicated patterns through microscopic images on medical samples or specimen in the laboratory environment [48]. This thesis focuses specifically on microscopic images. The microscope is a common imaging device in the medical field and play a crucial role in medical research and testing [10]. Microscopy of clinical specimens is a rapid and inexpensive method for diagnosis of diseases such as malaria [49], different types of cancer like leukaemia [36][50].

Although the analysis of biomedical images can be done by human experts in the laboratory,

there are considerable human errors associated with the process [20][51]. According to studies, there are approximately 40 million diagnostic errors involving imaging annually worldwide [52]. In a study of diagnostic error in medical images, researchers found that, out of the 54% of the errors which arose from radiotherapy, CT images analysis contributed majority of the errors with 23% of the errors, followed by MR reading with 9% [53]. Nuclear medicine contributed 7% and ultrasound 1% of the errors [53]. The inexperience of human operators of microscope has been cited to be the major source of microscopic diagnostic errors [54]. These errors can lead to an increased radiation dose which may lead to other diseases such as cancer, misdiagnosis, and clinical mismanagement, unwanted costs or even death [55]. Besides the problem of human errors, laboratory tests are also time-consuming and expensive, and there are risks of contamination of the samples under examination as well as a risk of infection for people undertaking the procedures [56].

The computerised automation using AI in the analysis of medical images offers important benefits, as it will reduce the possibility of human errors and subsequently increase the analysis accuracy, protect human analysts from the risk of coming in contact with the samples under examination for a long time and reduce the analysis time. Conventional machine learning models such as Artificial Neural network (ANN), Support Vector Machine (SVM), Naïve Bayes and Decision Tree, have been used in the analysis and classification of medical images in the prediction of mortality among cancer patients [57], lesion detection such as that of brain haemorrhage, or lung nodules [58]. In medical images, the use of machine learning to perform classification can be complex, particularly feature engineering when pre-existing knowledge is needed, can be challenging [70].

In a study of different feature selections as applied in images classification, there are no universal features that can work well in a range of different applications, so researchers still have to extract and test several features manually in each case [59].

In one study, only 20 features out of 4059 features extracted from histopathology images of heart tissue to predict heart attack were good enough to train a machine learning model [60]. If feature extractions are not performed well, it may result in consistently low performance and non-robust conventional machine learning model [61]. Indeed, the performance of a conventional machine learning model is greatly depending on how features are extracted [62]. In addition, the application of machine learning in the healthcare domain is highly affected by the problem of the imbalanced dataset. In AI world, imbalanced dataset refers to a situation where the training dataset contains unequal distribution of the number of observations in all classes [63]. In some domains like healthcare, imbalanced data problems are quite common. Taking an example of rare diseases diagnosis, the number of negative cases will dominate in the dataset compared to the positive cases [64]. As a consequence, machine learning algorithms trained with imbalanced dataset tend to favour the class with the largest proportion of observations (known as majority class), which may lead to misleading output [65].

### **1.3 Deep learning with biomedical data**

Deep learning offers huge potential as an alternative to the conventional machine learning for the development of computerised methods of interpreting and analysis of biomedical data as it does not need manual feature engineering. Hence it becomes an attractive option for the fast-growing amount of biomedical data. Deep learning models have the capability to search for their own features and patterns in the training data set and avoid the complexity of feature engineering as required in conventional machine learning. For example, deep learning approaches have achieved a remarkable performance in complex tasks such as speech recognition [66], natural language processing [67], and image classification [58], by training on raw data. Deep learning has been reported to have capabilities of detecting necessary features despite the external/unwanted signals interference which is common in biomedical data like EEG signals recording [68], hence it is becoming a promising technique for classification of



biomedical data. In a study of heart failure prediction [60], CNN outperformed conventional machine learning trained with 4059 features and also outperformed two experienced pathologists by over 20% classification accuracy. In a study of comparison of conventional machine learning and deep learning models in White Matter Hyperintensities (WMH) segmentation on brain MRI with mild or no vascular pathology, it was concluded that, deep learning models perform better than conventional machine learning [69]. In that study, authors used SVM and Random Forest (RF) for conventional machine learning models and Deep Boltzmann Machine (DBM), Convolutional Encoder Network (CEN) and patch-wise convolutional neural network (patch-CNN) [69] which are deep learning models.

In this thesis, we explore the application of Convolutional Neural Network (CNN), a popular deep learning algorithm. The algorithm is created with the concepts of artificial neural network which mimics how the human brain's neural network work [70]. CNN has been popular with image data, specifically for object recognition. Its various architectures like AlexNet [71], VGGNet [71] and InceptionResNet [72] have won global competition for general object recognition, the ImageNet [73]. ImageNet is a popular large-scale image database, contain more than 14 million images of 1000 classes of general objects [73]. Recently, CNN models have been successfully applied in healthcare domain for analysis of biomedical data such as diagnosis of breast cancer [74], brain tumour segmentation [56], early detection of seizure [75] and BCI applications for people with disabilities.

## **1.4 Research challenges of biomedical data analysis.**

Although deep learning can be a potential substitute for conventional machine learning models in the analysis of biomedical data, there are research challenges that need to be resolved.

### **1.4.1 Challenges of EEG based BCI**

It should be noted that most of the currently available EEG devices used to capture EEG signals

from the human brain are bulky, with a lot of channels which can easily capture noise and time-consuming to set up. Furthermore, the current existing EEG devices require the application of gel between the electrodes and the scalp to provide enough contact. The quality of the recorded signal falls as the gel dries out. In addition, to have a universal BCI system is still a challenge. In addition, a plenty of training trials are needed from the targeted user, causing the calibration period to be unacceptably long for a realistic model[76], apart from other challenges like psychophysiological and neurological challenges [77], variations in precise electrode placement has a major contributions in the limitations of universal EEG bases BCI applications. A slight change in the electrode placement between users may significantly affect the models accuracy [78], which has resulted in the need of calibration . These factors make them hard to use in daily life applications.

The challenges like the costs, high error rate, low response time and the large size of the current existing EEG devices have limited the application of EEG based real-life BCI applications [79]. Most of the challenges can be reduced by using small portable EEG devices with few electrodes as can lower the costs of the devices, reduce the size of the device and reduce the noises captured by the device hence decrease the error rate. Furthermore, the use of small and portable EEG devices can increase the information transfer rate which will result in the high response time. Although there are attempts to reduce the number of EEG channels with intentions to reduce the size of the current existing EEG devices, the methods used have been investigated to be inefficient as they are based in trial and errors of different channels combination, which is a hard task and consumes a lot of time and computationally expensive [5][80][81]. Channel reduction is a challenging task because when fewer channels are used, less spatial information is captured, making it difficult to detect artefacts as spatial information obtained in electrode placement is significant for accurate detection of artefacts in the collected EEG data [82]. In addition, when portable EEG devices are used, a method for precise placement of the devices on the scalp is needed. For these portable devices to work

practically in everyday environment, we need to restrict the number of channels significantly, to reduce costs, reduce setup time, and to improve usability and user acceptance. Also, the introduction of a method for precise channel placement on the scalp, can reduce the probably of inter personal variations and open the doors for creation of universal BCI applications.

### **1.4.2 Challenges of availability of medical images**

Applications of deep learning is limited by the ability to get enough training dataset, a common problem in medical images. Generally, the performance of deep learning models largely depends on the amount of training data available. Conventional machine learning performs better than deep learning approaches when only a small data set is available [83]. The problem of obtaining a big enough and balanced training data samples to train a deep learning model is significant in medical images. Development of large medical imaging data is quite challenging as annotation requires extensive time from medical experts especially it requires multiple expert opinion to overcome the human error [84]. The manual annotation is highly affected by the absence of qualified medical experts especially in developing countries [85]. In addition, unbalancing of data is very common in health sector especially dealing with rare diseases. By virtue of being rare, are underrepresented in the data sets. Training a deep learning model with unbalanced dataset can result in a biased model [84].

## **1.5 Research questions**

This thesis focuses on investigating the application of CNN in interpretation, analysis, authentication and classification of biomedical data for healthcare applications. Through two case studies, this thesis aims to address the challenges in section 1.4 through in the following research questions,

**Case study 1:** Application of CNN in interpretation and analysis of EEG signals for BCI applications. The research questions for this case study are:

1. How can CNN be applied without the need of feature engineering in interpretation, classification and analysis of raw EEG signals for BCI applications.
2. How CNN can be applied for EEG channels selection to reduce the number of channels for EEG device, towards the development of portable and more efficient EEG devices for real-life application of BCI.
3. How can CNN be applied to determine the location on the scalp for placement of portable few channels EEG devices on the human scalp for BCI applications.

**Case study 2:** Application of CNN in the authentication of cancer cell lines using microscopic images. The research questions for this case study are:

1. How CNN can be used towards the development of fast, easy to use, relatively cheap and computerised ways for authentication of the cancer cell line in the laboratory environment using microscopic images, without the need of feature extraction.
2. What additional authentications of cancer cell lines which can be provided by the application of CNN in the authentication process, to complement the current existing authentication method.
3. What techniques can be applied to take advantage of CNN while addressing the limited training data, a common problem in medical images.

## **1.6 Proposed solutions for applications of CNN for analysis of biomedical data.**

Deep learning, specifically CNN, can offer solutions to the above-mentioned research questions. This is explained below,

### **1.6.1 CNN on the EEG channel selection.**

When training a CNN model from scratch, the weights of the neurons are randomly initialised. These randomly initialised weights will be learned during the learning process. Through learning process of the model, the randomly initialised weights will be updated with a focus

of reducing the prediction error through a back-propagation process. The significant features (channels in our case) to the output related class will be assigned higher weights. This phenomenal of CNN in which significant channels being assigned higher weights gives an opportunity of knowing significant channels by extracting the weights of a trained CNN model. By using the same concept of CNN assigning significant channels with higher weights, a precise significant location on the scalp for placement of portable EEG devices can be identified. This can be done by training a CNN model with EEG data set obtained by using EEG devices with the large number of channels which cover the whole scalp. By using the weights extracted from a trained CNN model, a heat map can be drawn on the scalp and the significant location can be located.

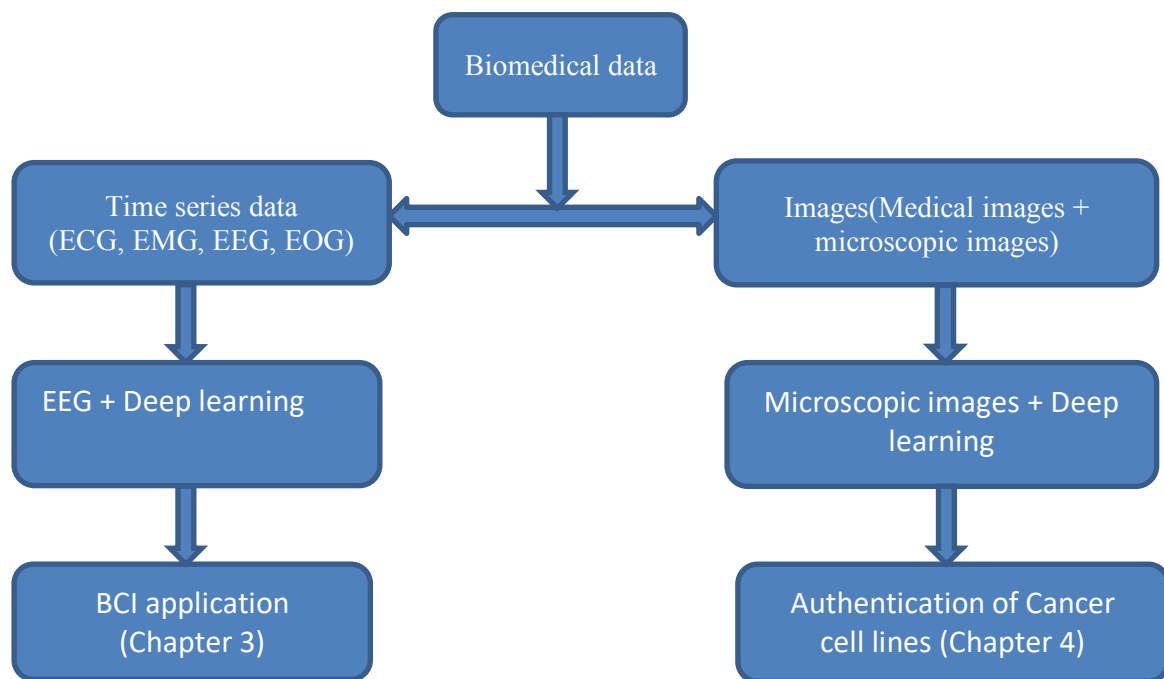
### **1.6.2 Possible solutions on the lack of enough biomedical images.**

When learning on a limited amount of data, data augmentation and transfer learning are two methods that can be used to improve the performances of deep learning models. Data augmentation is the process of artificially increasing the training data sample size by performing multiple operations, such as zooming, rotation and flipping on the available training data [86]. Transfer learning involves the use of pre-trained models, mostly pre-trained on an easily accessible dataset, the weights learned transferred to a task which is hard to get enough training data set [87]. Most of the available pre-trained models are trained on ImageNet [88], a dataset currently containing more than 14 million images of more than 20,000 categories of general objects like kites and balloons [89].

## **1.7 Thesis structure**

By addressing the challenges in analysis and interpretation of biomedical data, this thesis will focus on the application of CNN towards interpretation of EEG signals for BCI applications without the need of feature engineering and in EEG channel selection. This thesis will also

focus in providing a computerised way of authentication of cancer cell lines for anti-cancer drugs development research and in early detection of signs of a stroke on children with sickle cell diseases, by addressing the limitation of availability of getting enough training data on the former. Techniques such as data augmentation to artificially increase the training sample size and transfer learning are employed to solve the problem of scarcity of biomedical images for CNN applications. Generally, in this thesis, we demonstrate the application of CNN in the analysis of the two most common forms of data in healthcare, time series and images data. Time series data are those data that varies with time, like physiological signals such as EEG, ECG, EMG and EOG. For images, we used microscopic images. This thesis can be summarised as shown in Figure 1.1 below, the figure shows the path of all studies reported in this thesis. Studies which demonstrate the application of CNN in EEG based BCI are presented in chapter 3, while studies which demonstrate the application of CNN in biomedical images, for cancer cell lines images are presented chapter 4.



**Figure 1. 1:** Thesis structure showing how this thesis addresses the two most common forms of biomedical data, time series data which are mostly physiological signals and the images. The structure further shows how the dataset analysis were approached by using deep learning, in which we used EEG data for BCI applications and Microscopic images for cancer cells authentication.

## **1.8 Contribution**

By the research questions in section 1.5, the novel contributions of this thesis can be outlined in the following two key areas,

1. Real-life and real-time applications of BCI technologies by introducing portable, durable and user-friendly skin-like sensor for EEG signals with just 2-channels. The design and development of the sensor are achieved through a new method for EEG channels selection by using CNN introduced in this thesis. For precise placement of portable few channels EEG devices on the scalp, this thesis contributes a method by using CNN for determining the location on the scalp for placement of the newly developed few channels sensor. During real-time demonstrations, the developed 2-channels skin like EEG sensor was able to capture and send the signals via Bluetooth to our developed CNN model running on a mobile device for interpretation. The interpreted output was used to control a wheelchair, a robotic car and a PowerPoint presentation in real-time.
2. This thesis contributes a new and computerised way for authentication of cancer cell lines, an important step in the development of anti-cancer drugs. Apart from the method introduced by this thesis been relatively fast and easy to use in the laboratory environment compared to the current authentication methods such as STR, it provides more authentication parameters like the authentication of samples from the same genetic origin which cannot be achieved with current existing authentication. Most of the applications of CNN with biomedical images have been on the diagnosis of different diseases, this thesis introduces the application of CNN in an important and novel area but understudied, cancer cell line authentication.

## **1.9 Publications**

The contributions made in this thesis have been published in several numbers of peer-reviewed publications conferences and journals as highlighted in the table 1.1 below,

**Table 1. 1:** List of publications and contributions of the work in this thesis

S. No	Publication title	Journal/conference	Contribution
1	"Efficient Channel Selection Approach for Motor Imaginary Classification based on Convolutional Neural Network."	IEEE International Conference on Cyborg and Bionic Systems (CBS), pp. 418-421. IEEE, 2018.	Introducing a method for EEG channels selection by using CNN.
2	"A Channel Selection Approach Based on Convolutional Neural Network for Multi-channel EEG Motor Imagery Decoding."	IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), pp. 195-202. IEEE, 2019	Introducing a method for determining scalp location for placement of EEG electrodes
3	"Fully portable and wireless universal brain-machine interfaces enabled by flexible scalp electronics and deep-learning algorithm."	Nature Machine Intelligence (2019).	Contributions made in the first two papers were put into real world demonstration
4	"Towards Image-based Cancer Cell Lines Authentication Using Deep Neural Networks"	Nature scientific report	1.Introduction of a method for computerised authentication of cancer cell lines by using CNN.  2. Additional authentication parameters which can't be done with the current authentication methods

## 1.10 Thesis structure

Apart from the introduction chapter and conclusion chapter which introduces and concludes the work done in this thesis, there are other four main chapters which describe all the studies and work done reported in this thesis, with literature coming in the second chapter and they are organised as follows.

**Chapter 2:** This chapter describes the literature review on biomedical data production and their essential in healthcare. The importance of analysis and authentication of biomedical data, both



times series as physiological signals and medical images, the microscopic and MRI are introduced here. Different ways for analysis of biomedical data, dominantly machine learning and its weakness in analysis and classification is also introduced. Deep learning as a solution for the replacement of conventional machine learning is introduced and discussed in this chapter.

**Chapter 3:** Chapter three contains the major area of this thesis which is in the brain to machine interface for healthcare application. It covers the two studies which involve both motor imagery and steady-state visual evoked potential for triggering EEG signals for application in the brain to computer interface for healthcare applications. The first part of the chapter contains a study for motor imaginary and the second part contain a study for steady-state visual evoked potential. In both two studies, channels selection by using deep learning is extensively applied and the location of the dominant region on the scalp of the human is also located. For the second part, a study which involved the real application of the work done in this chapter is introduced and explained with real-life application of the findings of the two studies mentioned here.

**Chapter 4:** Chapter four contains a study for the application of deep learning in the authentication of cancer cell lines. As it is difficult to get enough medical images for training deep learning as deep learning requires a lot of data for training, different techniques for taking the advantage and power of deep learning while having small training data set are practised in this chapter. Techniques such as data augmentation and transfer learning are introduced. Furthermore, this chapter introduces different essential authentication parameters which cannot be offered by the current state of the art cancer cell lines authentication method, the STR. Finally, in this chapter, a visualisation technique is introduced to point out where the trained deep learning model is looking at to come into the conclusion in the authentication process, through a classification process.

**Chapter 5** is the final major chapter in this thesis work. This chapter discusses all the work

done in this thesis and concluded on the findings and point out research direction for further research work towards the development of better healthcare provision.

# Chapter 2

## Literature review

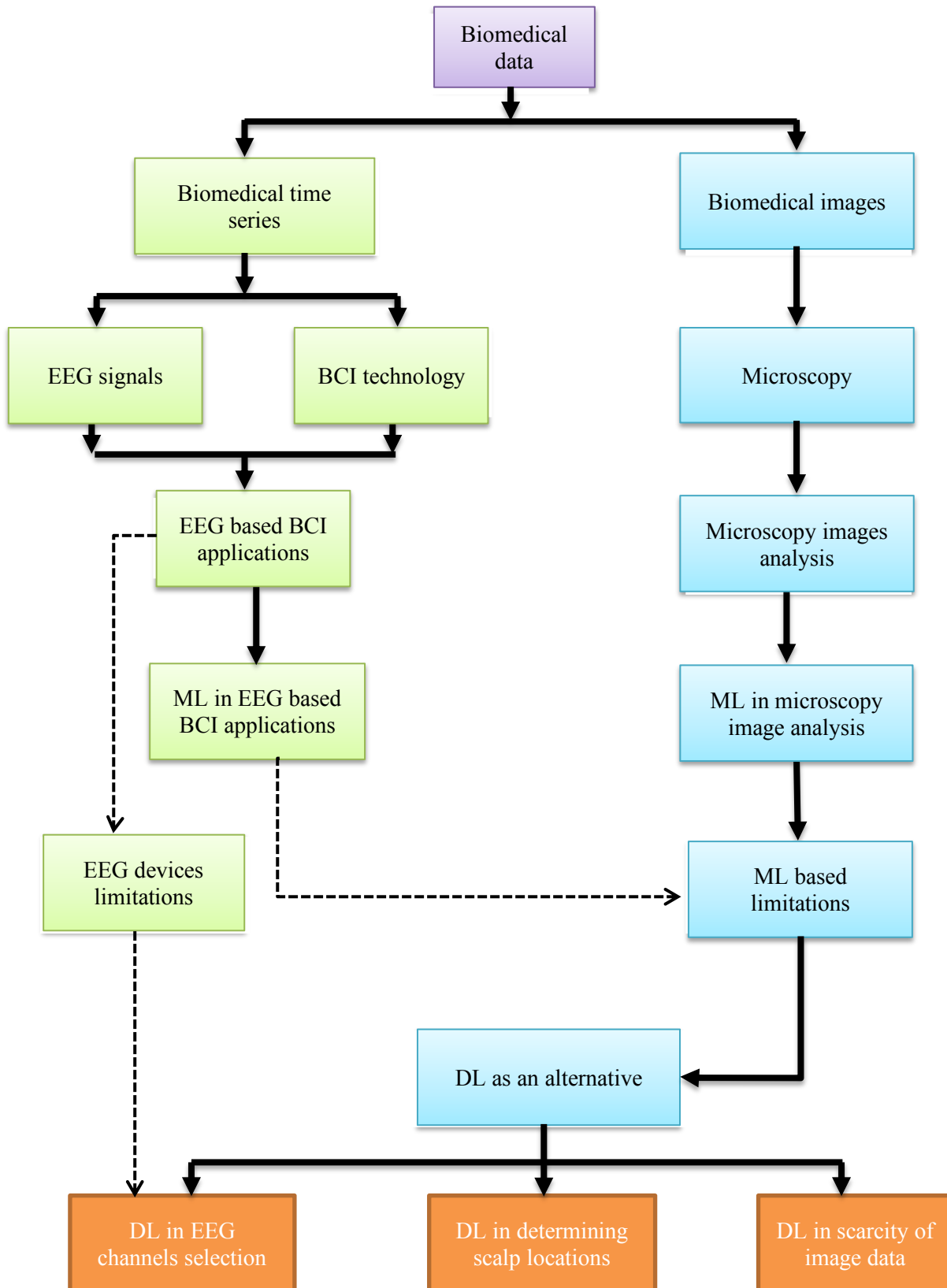
This chapter presents the literature background covering from the role of traditional, manual human biomedical data analysis to the need and potential use of automatic computerised methods. In regard to the latter, this chapter discusses conventional machine learning, showing how its limitations can be addressed by the use of deep learning, with a focus on Convolutional Neural Network (CNN). The literature provided is built around the EEG and microscopic images in which the studies reported in this thesis are based on.

Traditionally, in the process of diagnosis of diseases, genetic studies [90], pharmaceutical researches [4], and biotechnologies all rely on biomedical data analysis to extract health and biological information for treatment plans and research purposes. Analysis of biomedical data by human experts is a difficult and tiresome task, requires long training and is highly affected by human experience and can be affected by human errors. Recently, due to the advancement of technology in biomedical equipment, we have seen an exponential increase in the availability and access of biomedical data. This increase of the production of biomedical data adds work load and more room for human error, hence raising the need for the development of computerised methods for the analysis of biomedical data.

Supported by the advancement in computing technologies, Artificial Intelligence (AI) can assist humans in providing computerised methods for the automatic analysis of biomedical data. AI is concerned with the development of computers able to engage in human-like thought processes such as learning, reasoning, and self-correction [91]. Machine learning, a subset of AI, has been applied for the automatic analysis of biomedical data with high efficiency. Machine learning is the ability of a system to learn without being explicitly programmed [92]. Conventional machine learning has largely been limited by the need for manual feature extractions to train the model, a complex process that takes time and needs knowledge in the

domain of the dataset from which features are extracted (see section 2.3.4 for details). On the other hand, deep learning, a subset of machine learning, has offered an alternative to automated data analysis without the need for feature extraction. However, deep learning requires large training data sets that are of high quality and annotated. Therefore, these techniques have largely been applied in fields where a large amounts of training data are relatively easy to produce, such as voice recognition [93] and general object recognition [94]. The lack of data in the healthcare sector has been a major limitation in its application in this domain. Nevertheless, the data limitation can be somewhat addressed by techniques like data augmentation [95] and transfer learning [96] as discussed in section 2.6.5. As in most cases, the extracted biomedical data are in the form of time-series data [2] or images [97], studies in this thesis addresses both forms.

This chapter is structured as follow: Section 2.1 introduces the analysis of time series biomedical data, with a focus on EEG data and its application in BCI technology in which the first two main studies of this thesis are based on. Section 2.2 focuses on image form of biomedical data, with more emphasis given on Microscopic images in which the third study of this thesis is based on. An introduction to conventional machine learning for computerised automatic analysis of biomedical data is provided in Section 2.3, as an alternative to the manual analysis of biomedical data by human experts. Both the limitations of conventional machine learning in the analysis of biomedical data in general and the limitations of the current existing EEG devices for real-life BCI applications are also discussed in section 2.3. Section 2.5 discusses deep learning and how it addresses the limitations of conventional machine learning. It further explains the structure of CNN, a now widely used deep learning model. The section also discusses the limitations of deep learning. The applications of CNN in interpreting biomedical data are introduced and reviewed in section 2.6 and its limitations are addressed. The flow of this thesis chapter is shown in the literature review structure, figure 2.1.



**Figure 2. 1:** Literature flow structure. For simplicity of the figure, ML stands for Machine Learning and DL stands for Deep Learning.

## **2.1 Time series Biomedical data**

Time series biomedical signals are one of the most common form of biomedical data. Time series data are data which constantly vary with time [98]. A time-series data of variables  $x$  of length  $n$  is an ordered sequence of observations recorded at equispaced instants and can be denoted by  $x_1, x_2, x_3, \dots, x_n$ . This data can be, for instance, an hourly reading of patient's temperature or physiological signals which are continuously generated by the human body, sampled at a certain frequency for a particular period [2]. Human physiological signals have been widely used for research studies, the diagnosis of diseases and for different healthcare conditions [99]. Some of the common physiological signals are Electroencephalography (EEG), Electrocardiography (ECG), electromyography (EMG), and Electrooculography (EOG). EEG signal is an electric signal produced by the brain when performing different activities as the brain cells communicate and mostly used to monitor different functions of the brain [7]. ECG is an electric signal produced by the heart while performing its activities [100], EOG is an electric signal produced due to eye movement and EMG is an electrical signal produced by muscular movement [13]. Parameters like pain intensity [101], emotions [27], or early detection of diseases like seizure [102] and Brain to Computer Interface (BCI) applications [103] have been possible through the analysis of physiological signals.

### **2.1.1 Background of Electroencephalography (EEG)**

The first two main studies of this thesis focus on the application of EEG signals for BCI applications (see chapter 3). This section provides a general literature background of EEG, how it has been applied in BCI technologies and the limitation of the current existing EEG devices in real life BCI applications.

EEG signals are measured in small electrical values, namely microvolts ( $\mu\text{V}$ ), which can be recorded from the scalp. They are commonly sinusoidal in shape [104]. The existence of

the electrical activities in the human brain was discovered more than 100 years ago from the exposed brain of rabbits and monkeys by Richard Caton [104]. Most of the recent medical and non-medical applications of EEG signals were established by Hans Berger in the 20<sup>th</sup> century [105]. He discovered that EEG signals can be detected by amplifying the EEG signal using ordinary radio equipment in the laboratory environment [104], removing the need to open the skull. During the execution of voluntary and involuntary activities by the human, EEG signals with varied frequency bands and amplitude which describe the undergoing activity will be generated by the brain [106]. The generated EEG signals can generally be categorised into four different frequency bands: alpha (8-13 Hz, 30-50  $\mu$ V amplitude), beta (13-30 Hz, 5-30  $\mu$ V amplitude), gamma ( $\geq 30$  Hz), delta (0.5-4 Hz), and theta (4-7 Hz,  $\geq 20\mu$ V amplitude) [107]. The generated EEG signals vary during the execution of different kinds of actions like the movement of arms and legs [108], visualisation [109], problem-solving [110], counting, or even just by imagining performing a certain activity. Previously, EEG signals have been mostly used in medical applications in areas of diagnosis of different diseases like seizure [111], post-surgery monitoring of brain surgeries [112] and other brain related medical researches. Current advanced research in EEG has made these technologies more accessible to people outside the medical field, resulting in the development of cost-effective EEG devices and opening the door for different applications of EEG signals in areas such as neuromarketing [113], social interaction studies [114] and studies of cognitive capabilities [38] and BCI applications [115].

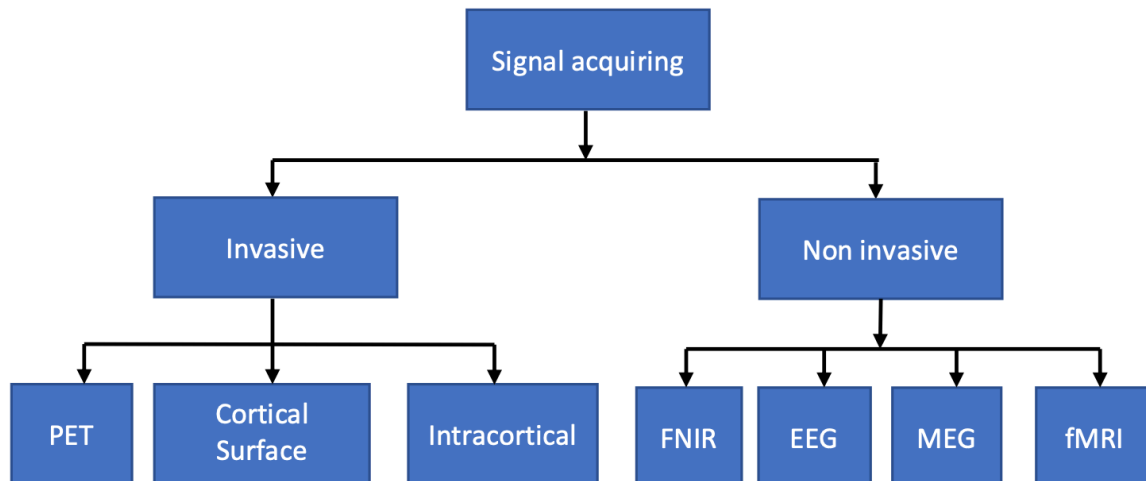
## **2.1.2 EEG-based Brain to Computer Interfaces (BCI):**

A relatively new field, but gaining attention for EEG applications, is brain to computer interfaces (BCI). BCI is a technology that allows a direct communication pathway between the human brain and an external device [37]. BCIs bypass the conventional neuromuscular output pathways and apply computational models (mostly machine learning) to translate brain signals into action [116]. This section provides a background to BCI technologies, and presents background literature on the application of EEG signals in BCI technology which is most commonly used in BCI systems.

### **2.1.2.1 Introduction to BCI**

Typically, a BCI system has a hardware component that includes sensors for the acquisition of signals generated from brain activities and software component that is composed of the computational model such as a machine learning for translation/ interpretation of the acquired signal. Although EEG is dominant, there are other several methods of acquiring brain signals for BCI applications. These methods can be grouped into invasive and non-invasive methods, shown in figure 2.2. Invasive methods involve the insertion of sensors or other materials inside the brain to capture the brain activities. Examples of invasive methods are Positron Electron Tomography (PET) [117], Cortical surface (ECoG) [118] and Intracortical [119]. Non-invasive methods capture the brain signals on top of the scalp without the need for a direct contact with the brain. Examples are Functional Near-Infrared Spectroscopy (fNIRS), EEG [114], Magnetoencephalography (MEG) [120] and Functional Magnetic Resonance Imaging (fMRI) [38].

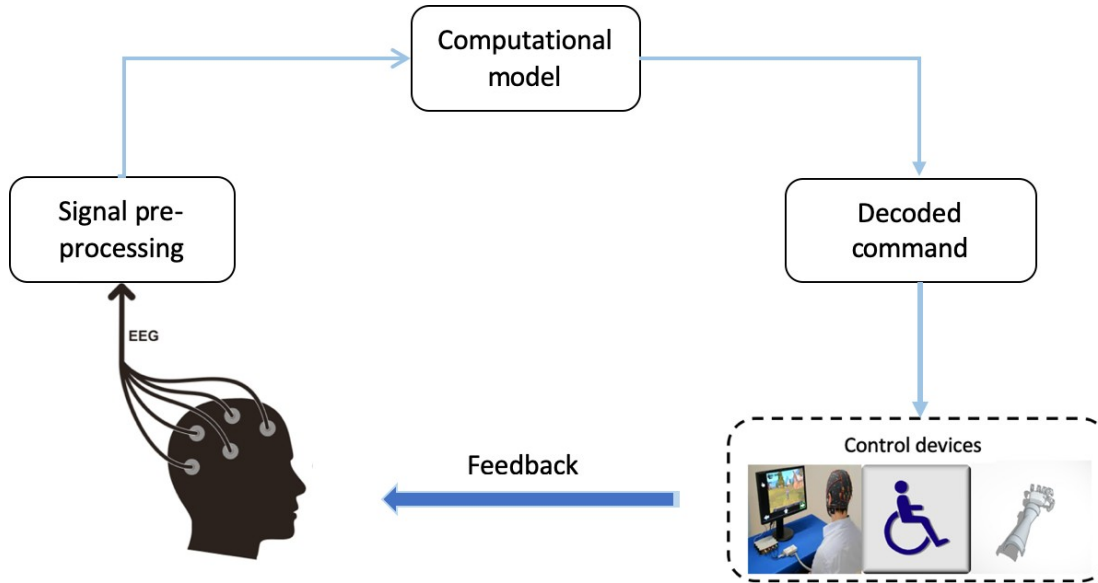




**Figure 2. 2:** Different techniques to acquire brain signal. In which EEG method is one of the non-invasive methods [47].

### 2.1.2.2 EEG for BCI

Naturally, EEG signals are continuously produced by the brain. Traditionally, EEG signal has been widely used in medical applications, such as the diagnosis of neurophysiological disorders like epilepsy, dementia, autism. Until recently, due to the advancement in computing technology, EEG signals have started being used in BCI applications. To apply EEG signals for BCI applications, an internal or external stimulus is required to trigger signal variations [121]. The variations are then detected and translated into a required command. Internal stimuli can be voluntary or involuntary intention to perform a particular activity for motor imaginary activities [42]. External stimuli can be visual sources, such as light signals [149] or audio signals [122]. All BCI applications follow the same trend, i.e. the signal needs to be triggered by an internal or external stimulus, captured by the EEG device and interpreted by a computational model such as a machine learning model. The interpreted output from the model will be sent to control the external device. The whole architecture of the EEG based BCI application is shown in figure 2.3. The feedback component helps the user to modulate the brain activity and to display the interpretation results [123].



**Figure 2. 3:** EEG based BCI application system [124].

### 2.1.2.3 Paradigms of EEG for BCI applications

For BCI applications, variations should be triggered in the EEG signals. These variations are related to particular events or stimuli. In brain activity patterns, based on how the EEG signal variations is triggered for BCI applications, EEG signals can be categorized into three major paradigms. These are: Event-Related Desynchronisation (ERD)/synchronisation (ERS), Event-Related Potential (ERP) and Slow Cortical Potential (SCP).

Event-related desynchronisation (ERD) describes a short-lived and localised amplitude attenuation or blocking of rhythms within the alpha band of the EEG signal frequency bands [125]. Event-related synchronization (ERS) defines a temporary amplitude enhancement in the EEG signals. Both ERD and ERS are frequency bound and can be quantified in space and time with a couple of movement experiments, hence are ideal for BCI applications [126]. The most common form of EEG in this category is Motor Imaginary activity (MI), which is applied in the first study of this thesis in EEG based BCI application (reported in Chapter 3). MI way of triggering EEG variations and its concept is further explained in section 2.1.2.4 of this chapter. The second paradigm of EEG for BCI applications is Event-related potentials (ERPs), which

are very small voltages generated in the brain structures in response to specific events or stimuli [127]. For the ERPs to be studied as a result of external stimuli, several repetitions of the event or exposure to the stimuli should be done for a computational model to be able to detect the changes in the recorded EEG signals [90]. The stimuli can be of different forms, such as visual which will result in Steady-State Visual Evoked Potential (SSVEP) [128]. The second study of this thesis (also reported in Chapter 3) in EEG based BCI application utilizes the SSVEP form of EEG, which is further explained in section 2.1.2.5. The last type is Slow Cortical Potential (SCP), which are EEG signals of less than 1Hz and are related to sensory, motor, and cognitive processes of the brain [129]. The activations of SCP will result in a slow direct current shift related to the event [130].

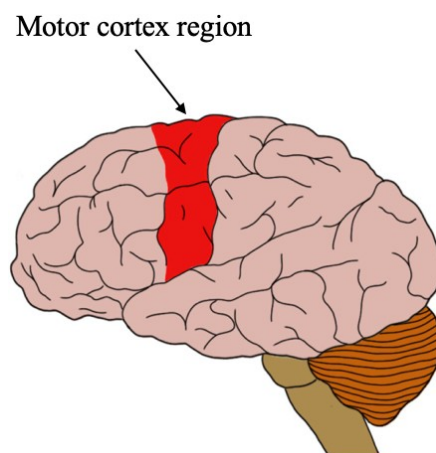
Among all types, the ERD/ERS, which contains the Motor Imaginary (MI) activities, the ERPs which contains SSVEP, and their hybrids have attracted more interest from the researchers in the field of BCI [131]. Generally, based on BCI applications, EEG can be divided into two main categories according to how they are generated, which are evoked and spontaneous [105]. For the evoked form of EEG, an external stimulus is required to trigger the variations of the EEG signals. These external stimuli can be visual, auditory, or sensory stimulation. For spontaneous EEG, the EEG signal is generated with different mental activities and no external stimuli are required [42]. The first two major studies of this thesis apply MI and SSVEP, the two types are further explained in the following sections.

#### **2.1.2.4 Motor Imaginary (MI) Activities for BCI**

Motor imaginary (MI) is a common example of spontaneous EEG in which no external stimuli are required to trigger their generation [42], and it belongs to ERD/ERS type of EEG signals for BCI applications. The EEG signals related to MI activities are naturally produced when performing a certain movement [132]. All human beings' voluntary movement starts with an intention in the brain to make such a movement. The brain will generate small electrical

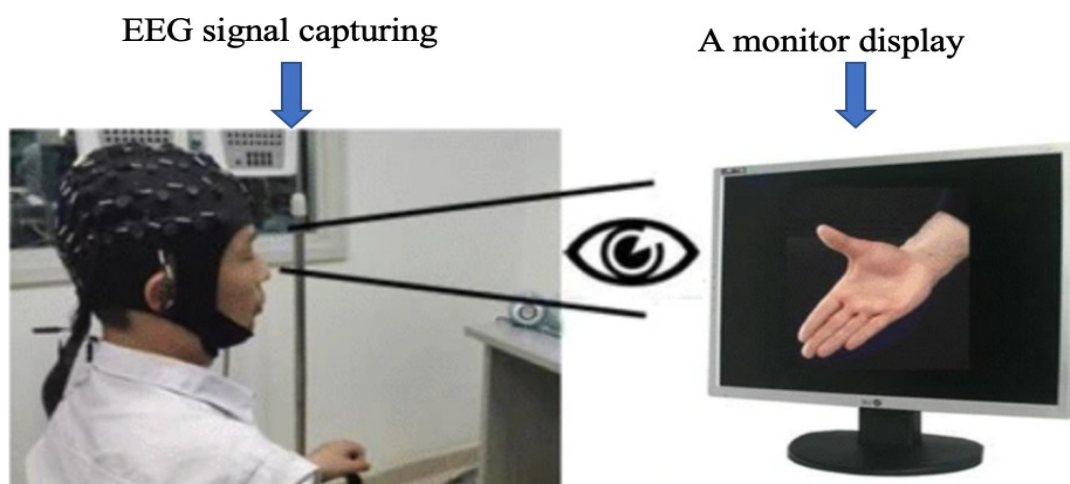
signals (EEG) related to the intended activity and pass the signal through the nervous system to the muscles that generate displacements or movement of the body part intended [133]. The same process happens during the execution of involuntary actions. Involuntary actions such as breathing and heartbeat are also controlled by the brain as the brain realises the need of that particular action to happen and send the electrical signals through the nerve system to the muscle of the particular organ to execute the required involuntary action need [134].

During the execution of the real or imaginary motor task, there are always variations in a particular region of the brain (figure 2.4), such as an increase and decrease of power spectrum and amplitude of the generated EEG signals related to the activity. The motor imaginary activities are more dominant in the motor cortex region of the brain, shown in figure 2.5. These variations mostly occur in the alpha (8Hz-13Hz), sigma (14Hz-18Hz), and beta (18Hz-30Hz) frequency bands of the recorded EEG signals [135]. Individuals with a high-level of amputation or neuromuscular disorder lack sufficient residual muscles or body part to control for execution of the intended action. In this case, the EEG signals generated by their brain for the particular movement intention will be lost as there are no muscles or weak muscles to respond to the generated EEG signals or absence of body part to be controlled. By interpreting the obtained EEG signals related to motor imaginary activities, one can interpret an intention of movement or imaginary activity patterns that are hidden in the produced EEG signals.



**Figure 2. 4:** Motor cortex region of the brain, marked in red [136].

To trigger EEG signals related to MI tasks in the lab environment, a subject needs to imagine performing a certain motor activity, like the movement of the limb [137]. To collect EEG data related to motor imaginary activities in the laboratory environment, a prompt should be used to alert a person under study to start imagining they are performing a particular activity, while a marker is also included in the recorded EEG signals to mark the start and end of the particular activity, as shown in the figure 2.5 [62]. To be able to get enough EEG data for application in BCI technologies which uses machine learning as a computational model for interpretation the EEG data, the experiment needs to be repeated several times, with a predefined time of execution of each of the motor imaginary activity. MI related EEG signals have been playing a major role in BCI technology [28]. A research project [104] managed to record EEG signals related to five different motor imaginary activities for BCI applications in the lab environment.

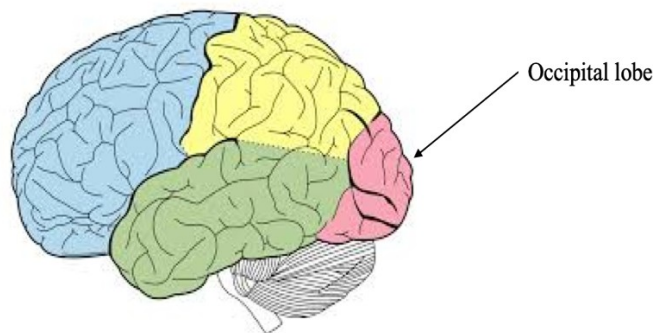


**Figure 2. 5:** A laboratory set up to trigger motor imaginary potential generation in the human brain. An action is displayed on the monitor to inform a participant which action to imagine in doing, while the EEG signals is recorded on the scalp [138].

### 2.1.2.5 Steady-state visual evoked potential (SSVEP)

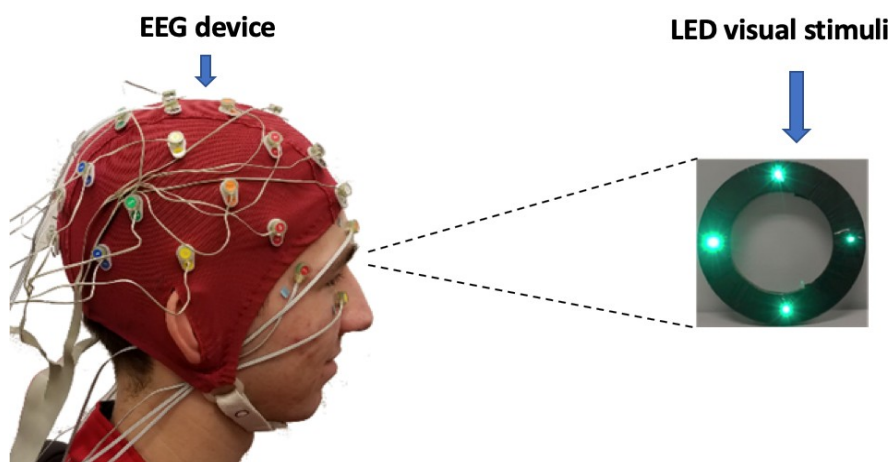
Steady-state visual evoked potential, or SSVEP, is an electrical signal generated by the brain, as a resonance phenomenon evoked when a subject looks at a light source flickering at a

constant frequency [132]. It is an evoked form of EEG as it requires an external stimulus. The rise of EEG amplitude has been reported to be observed when a person focuses on a light source flickering at a constant frequency [132]. To trigger visual evoked potential related EEG signals, it needs an external flickering virtual stimulus light of above 6Hz [139]. The virtual evoked EEG signals are more dominant in the virtual cortex region of the brain [92], which is found in the occipital lobe region of the brain as shown in figure 2.6 [80].



**Figure 2. 6:** The occipital lobe region of the brain. This is where the virtual cortex region of the brain is found. It is in the virtual cortex region where the SSVEP form of EEG are dominant [140].

The source of the flickering light can be programmed light-emitting diodes or programmed flickers on a monitor of a computer screen [141]. A whole set up for SSVEP application of someone using light-emitting diodes is as shown in the figure 2.7.



**Figure 2. 7:** A complete set up of SSVEP based EEG for BCI application. The stimuli source is a programmed light emitting diode flickering at different frequencies.

## 2.2 Biomedical images

Apart from time series data explained above, Biomedical images is another common form of biomedical data. Biomedical imaging involves the process of capturing, processing, and visualising medical images of living objects, systems, or biological samples. The main purpose of biomedical imaging is to monitor the functions of different body organs, diagnose of diseases or other medical conditions like bone fracture [141] and also for medical research [142]. Biomedical imaging spans from simple and common technologies like microscopy to produce microscopic images in medical laboratories, to more complex imaging technologies like magnetic resonance imaging (MRI). Most of biomedical imaging techniques are non-invasive and provide uniquely valuable information about tissue composition, morphology and function, as well as quantitative descriptions of many fundamental biological processes and play a crucial role in diagnostic and therapeutic purposes [143].

### **2.2.1 Background of biomedical images**

Since the discovery of X-rays more than 120 years ago, the world has seen different discoveries of biomedical imaging devices. Now it is possible to image the interior parts of the human body in intricate detail using advanced biomedical imaging devices like Computed Tomography (CT) scans, Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), ultrasound, nuclear scans and other numerous modalities [92]. X-ray is a form of high energy electromagnetic radiation, it finds applications in the medical field in producing images of tissues and structure inside the body [144]. CT scan uses a combination of multiple X-rays taken from different angles to produce cross-sectional (tomographic) images of a specific area of an object under scanning [145], and the method is widely used in the diagnosis of diseases and injuries in the medical field [146]. Ultrasound uses high-frequency sound waves to produce live images from inside of the body and has been used for obstetrical purposes, abdominal scanning and diagnosis of pulmonary and renal diseases [147]. Another popular imaging

technique, MRI, uses strong magnetic fields and radio waves to produce images of organs in the medical field [148] and has been used in the diagnosis of muscle, spinal and brain disease [149]. The most common biomedical imaging is the microscope. A microscope applies light concept and is used in the visualisation of small and abstracts structures that cannot be seen with naked eyes, and is used in many fields, mostly in biology laboratories [150].

Biomedical images are typically analysed by human analysts (e.g. a radiologist or pathologist) to obtain meaningful insights into the patient or function of different organs and systems, as well as in medical research, such as in discoveries of new drugs. It is through pathological samples that common fungal and parasitic common diseases such as Malaria [151] can be diagnosed by using a microscope in the laboratory, and full-body scans of highly imaging technologies like CT scans that can identify complex diseases like cancer and internal injuries [177].

## **2.2.2 Categories of biomedical images**

Although most forms of biomedical images are medical images which produce internal visualisation of the body, microscopic images are common form for magnification and visualisation of samples in the laboratory. Microscopic images are produced by using a microscope device, a common laboratory device for magnification and visualisation of biomedical samples. This section describes the two forms of biomedical imaging.

### **2.2.2.1 Medical imaging**

Medical imaging is the method of producing visual representations of the interior of a body for clinical and medical analysis and intervention, as well as visual representation of the function of some organs or tissues (physiology). Lately, different ways of medical imaging have been developed and largely applied in healthcare and medical researches. Based on the operation mechanism, medical imaging can be classified into the following:



## **I. Radioactive medical imaging**

This is the form of medical imaging that involves the use of ionising radiation. Ionising radiation is the part of the electromagnetic spectrum with sufficient energy to pass through matter and physically extricate orbital electrons to form ions. Although they are known for producing high-quality images, both X-rays and CT scans use ionising radiations and have limited applications on the patients due to their radioactive side effects [ 1 5 2 ] .

## **II. Non-radioactive medical imaging**

On the other hand, non-radioactive imaging does not use any form of ionising radiation. Examples of non-radioactive medical imaging include ultrasounds that produce images by using high-frequency sounds and MRI which produces images by using a strong magnetic field [17]. Most of the non-radioactive imaging, such as the ultrasound, have no known side effect to the users [147].

### **2.2.2.2 Microscopic imaging**

One of the oldest and most common imaging devices is microscopy. It uses the concept of light refraction to produce a magnification of the sample under investigation. In the medical laboratory it is used in visualisation of pathological samples [153]. Recently, due to advancements in technology, electronic microscopes have been developed with the ability to produce high quality microscopic images [154].

### **2.2.3 Limitations of manual biomedical images analysis**

Human-based biomedical data analysis is error-prone. Globally, it is estimated that 94,000-142,000 people died between 1990 and 2013 as a result of diagnostic errors [155]. Approximately 850,000 diagnostic errors are reported yearly in developed countries, which results in both health and economic impacts [156]. A recent study from Johns Hopkins suggests that medical errors are now the third leading cause of death in the US [157], surpassing stroke,

Alzheimer's and diabetes. Delays or misdiagnoses have been identified as major contributors of diagnostic errors in healthcare [33], in which human errors are one of the major reasons. Another study categorised human errors related to diagnosis errors into knowledge-based, rule-based, skill-based, or 'other' (such as violations or failures by deliberate deviations from rules or procedures) [158].

In biomedical images in particular, although factors related to the imaging modality such as image resolution, intrinsic or extrinsic contrast, and noises, can lead to diagnosis errors, most missed radiologic diagnoses are attributable to image interpretation errors by human analysts [159]. Intrinsic contrast factors are those that cannot be changed as they are inherent to the body's tissues, while the extrinsic contrast parameters are those that can be changed because they are under our control. An eight-year study by Kim et al [53] found 1279 errors examining 656 exams with delayed diagnoses. The same study [53], reported that 54% of the errors arose from the interpretation of general radiography, with CT and MRI reading errors accounted for 23% and 9% of the errors respectively. Nuclear medicine and ultrasound contributed 7% and 1% respectively. Under reading was the main cause of the errors, contributing to 42%, while the failure of radiologists to thoroughly examine anatomy after seeing abnormalities relevant to a primary diagnosis, was the second most common error [145]. The errors in biomedical imaging can be classified into two major groups: perceptual errors (those in which an important abnormality is simply not seen by human on the images) and cognitive errors (those in which the abnormality is visually detected but the meaning or importance of the finding is not correctly understood or appreciated) [159]. These errors can lead to increased radiation dose, misdiagnosis, clinical mismanagement, unwanted costs, or even death, and can affect public trust.

The advance of different techniques of biomedical imaging has resulted in the production of large data sets of biomedical images which need to be analysed by human analysts [153]. Healthcare data amounted to about 500 petabytes in 2012 globally [160], and this figure is

projected to grow at 36% annual rate through 2025 [161]. As the availability of the number of biomedical data increases, the workload on human analysts to analyse these biomedical data also increase and hence error rates grow as well [94]. This necessitates the development of more accurate and quicker methods for the analysis of biomedical images.

## **2.3 Biomedical data analysis with conventional machine learning**

During the last few decades, due to the advancement in computing technologies and digitisation of biomedical data, there has been the development of different computerised ways for the analysis of biomedical images. Due to the need of development of computerised methods for automatic analysis of biomedical data, this section introduces machine learning as a potential computerised method to assist human analysts. This section presents literature background on machine learning, and how it has been applied in the automatic analysis of biomedical data, with a focus on EEG for BCI applications (in section 2.3.2) and biomedical imaging (in section 2.3.3), as well as its potential drawbacks.

### **2.3.1 Introduction to conventional machine learning**

The whole concept of machine learning can be categorised into three major tasks, which are: classification, regression and clustering. Based on how the model is trained, machine learning can be categorised into two major groups: supervised training and unsupervised learning. Supervised learning is when a machine learning model is trained with a data set that contains labels. Unsupervised learning is when a machine learning model is provided with a training dataset that does not contain labels, so the model will find the patterns in the dataset and group the dataset based on the patterns found [162]. There is another category which falls between the two, Semi-supervised learning. Semi-supervised learning is an approach to machine learning that combines a small amount of labelled data with a large amount of unlabelled data during training [163]. Classification is a supervised learning which category the data set into classes [164]. This involves tasks like handwriting recognition [165], facial

recognition [166] and voice recognition [167]. In healthcare, machine learning has been used in the classification of cancer cells into benign and malignant by using biomedical images [51], classification of motor imaginary activities [168], and visual evoked potentials [169] for BCI applications by using EEG signals. Regression is also a supervised learning technique, in which the output variable is real and continuous and has been used in predicting genetic sequences [170] and risk of cardiovascular disease. Clustering involves the grouping of data points, in which data points in the same group are more similar to each other than data points in the other group [171]. Clustering techniques have been used in segmenting patients to provide them with effective treatments [172]. Some of the most popular conventional machine learning models include Artificial Neural Network (ANN), Support Vector Machines (SVM), Linear and Logistic regression, Random Forest, and Decision Trees.

The process of training a conventional machine learning model involves pre-processing of the data and manual extraction of features from the training data [173]. The data set is usually divided into training set, validation set and the test set. The training set will be used to train the model, the validation set will be used to monitor the model performance during training, and the test set will be used to test the model for generalisation. After the model been trained, optimised and tested, the model will be deployed for use. The prediction process of the model involves acquiring the data on which the prediction must be made. A pre-trained model will be loaded and features extracted from the incoming data and fed into the model to make predictions. The pre-trained model will make a prediction on the new incoming data based on the knowledge gained through the patterns learned from the training data [174].

### **2.3.2 EEG with conventional machine learning for BCI applications**

Nearly all BCI systems use a machine learning algorithm as a computational model to interpret the captured EEG signals [44]. With a focus on Motor Imaginary (MI) and Steady State

Visual Evoked Potential (SSVEP), this section explains how conventional machine learning has been used to interpret the commands found in raw EEG signals, for BCI applications.

Conventional machine learning has been successfully applied in finding patterns in the obtained EEG signal related to someone's intention to move a limb or attention on a flickering light from the recorded EEG signals. BCI systems which apply machine learning as a computational model have been successfully applied in controlling prosthetic arms [42], wheelchair [47], and computer applications. EEG signals as other biomedical data, have tens of thousands of features that can be studied and interpreted to produce deep information about different physiological activities [175]. Over the past decades, several handcraft techniques have been used for EEG feature extraction, such as standard deviation (SD), variance, Fast Fourier transforms (FFT) [32], Wavelet Transform (WT) [83], and Power Spectral Density (PSD) [34] to mention a few. Neurophysiology prior knowledge plays a major role in determining significant features that will provide more discriminative information among EEG classes [44].

Linear Discriminant Analysis (LDA), K-Nearest Neighbour (KNN), Artificial Neural Network (ANN), and Naive Bayes (NB) are among the most common conventional machine learning models which have been used for interpretation of EEG signals for BCI applications [176]. Wang et al. used Fisher Linear Discriminant (FLD) and Support Vector Machine (SVM), to interpret MI-based EEG signals to control a humanoid robot [103]. ANN and SVM were again used to interpret SSVEP based EEG triggered by four different flickering frequencies displayed on a Liquid Crystal Display (LCD) for BCI applications, with SVM attaining 88.5% classification accuracy [177]. The performance of a machine in classifying the patterns largely depends on the features extracted to train the model and the model itself [178].

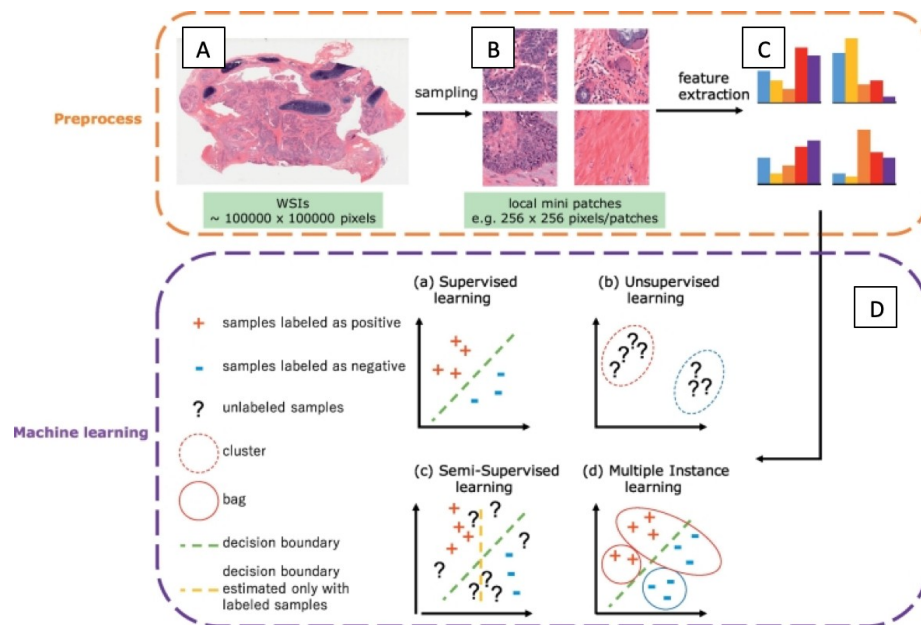
### **2.3.3 Conventional machine learning in Microscopic images**

In biomedical images, conventional machine learning has been used in tasks such as the classification and segmentation of biomedical images. Low-level features such as colour, texture, and shape [179], as well as high-level features such as intensity histogram features and Gray Level Co-Occurrence Matrix (GLCM)[179], have been extracted from biomedical images and used in training conventional machine learning models. For instance, HS Sheshadri *et.al* [180] extracted statistical features such as mean, standard deviation, smoothness, third moment, uniformity, and entropy from digital mammograms of breast tissue for early breast cancer detection.

Conventional machine learning models such as Artificial Neural Network (ANN), Support Vector Machine (SVM), Naïve Bayes and Decision Tree, have been used in biomedical images classification [74], performing segmentation of areas of interest on biomedical images [181], identification of biomedical image samples and anomalies detection [182]. A study by C.Qin *et.al* [183], analysed several conventional machine learning models in their analysis of Chest X-ray images to detect different chest diseases like pulmonary nodules, tuberculosis (TB), and interstitial lung diseases. In another study, Smita *et.al* [183], found that Support Vector Machines (SVM) performed the best compared to other conventional machine learning models at classifying cancer images. Another study introduced an automatic feature extraction method from biomedical images by using the spatial orientation of high-frequency textural features of the processed image, and trained SVM binary classifier [184].

The microscope is one of the common and simple imaging devices, commonly found in medical laboratories for diagnosis of disease and bioscience researches. Features extracted from microscopic images of cancer cells were used to train conventional machine learning models such as Artificial Neural network (ANN), Support Vector Machine (SVM), Naïve Bayes and Decision tree, to classify cancer cells into benign and malignant [96]. In addition, 2D microscopic images of the parasite were used successfully in 3D reconstruction of small organisms like parasite monogenean by using a Machine learning models like ANN. 3D

modelling is essential in studying structural and functional morphology of small organisms as some are not visible in the 2D image [185]. Figure 2.8 below shows a whole process on how machine learning is used to perform different analysis tasks with microscopic images.



**Figure 2. 8:** Process of training a conventional machine learning with pathology samples of microscopic images, as modified from [186]. Part A shows an image of a tissue sample, which patches of interesting areas will be produced as shown on part B where features will be extracted from the patches as shown in part C and the conventional machine learning model will be trained for different tasks in part D.

### 2.3.4 Limitations of computerised biomedical data analysis

Although there are some real-life applications of computerised biomedical data analysis based on conventional machine learning, the practice has always faced some challenges for the real-life applications [187]. This section discusses the limitations of conventional machine learning in the development of computerised biomedical data analysis.

The major limitation of conventional machine learning is the need for manual feature extraction. This section discusses how manual feature extraction affects the application of conventional machine learning in both EEG signals and biomedical images. Biomedical data are known to be heterogeneous, temporal dependent, sparse, irregular, noisy and of high dimension, hence becomes too complex to perform feature extraction to train conventional machine learning model [188]. Determining the most representative feature set is a major

challenge as it needs someone who has vast knowledge and experience on the field in which features are extracted. Furthermore, the task of extracting robust and accurate features is always affected by noises and interference from other signals, a common limiting factor in biomedical data.

In EEG signals, feature extraction process is always affected by factors like motion artefacts [189], electrooculography (EoG) [190] and EMG. EOG is a small electrical signal produced by the eye's movement, and EMG is an electrical signal produced by muscular movement. In MI classification tasks particularly, the imagined movement is often lost in this mixture of other signals such as EOG and EMG [82].

In biomedical images, pre-processing of the obtained images data is needed first to improve the quality and specifying the region of interest before feature extraction as mentioned in section 2.3.3.2. For example, in breast cancer feature extraction, tumour in breast tissue is always surrounded with calcification [191]. Calcification is the deposit of Calcium on tissues, which is the marker of the underlying process occurring in the breast tissue including cancer [191]. To extract relevant features in situations like this, pre-processing of the image is needed to isolate the region of interest. Manually identifying and extracting features from medical images such as mammograms, using good discriminative features to train a conventional machine-learning model, is time consuming, complex and highly affected by the experience in the field of the person extracting the features.

Generally, biomedical data have high dimensions with tens of thousands of features, hence it becomes challenging to extract relevant features due to dimensionality factors. Although several techniques for reduction of feature dimensions like Principle Component Analysis (PCA) have been applied, it is a challenging task to link the features from original feature space to new features [192] as PCA takes the principle components are a linear combination of the original features [193]. The efficiency of machine learning models that are trained on features extracted from acquired biomedical signals are always affected by redundant features



extracted [194].

## 2.4 Limitations of current existing EEG devices for real-life applications of BCI

Most of the current available EEG devices are still dense, bulky, and hard to wear, which makes them difficult to use in daily life [63], as shown in figure 2.9 below, most applications are found in research labs only. These factors limit the development of real-life EEG based BCI applications. The real-life BCI applications need to be conducted in a way that makes it easy for subject to go about their normal life activities. To apply EEG devices in everyday environments, we need to restrict the number of channels significantly, to reduce their costs, reduce setup time, and to improve usability and user acceptance. Hence, rise it is necessary to study the EEG channels reduction towards development and efficient use of small, portable, and easy to use EEG devices.



**Figure 2. 9:** Components of current existing EEG devices which are hard to use, takes a lot of time to wear, not user-friendly, and almost impossible to use real-life applications of EEG based BCI applications [195].

The reduction of EEG channels will result in the reduced size of EEG devices, reducing computational costs as now only a few relevant channels with significant information will be used [196]. The core functions for EEG channel selection process can be categorised into three

major areas: i) To minimise computational expenses in analysis and interpretation of EEG signals;

ii) To improve the machine learning model performance by training the model with only significant channel with less noise; and iii) to reduce the setup time and complexity when wearing the EEG device [197]. An improvement on the generalisation of the classifier was observed when EEG channels selection was conducted by selecting six (6) significant channels out of thirty-two (32) in a BCI application.

## **2.5 EEG Channels selection**

Channel reduction is a challenging task. There are several methods already in the literature for EEG channels selection, many of them have been complex and takes a lot of time to implement and are computational expensive [81]. The most known channel selection methods include the following.

### **2.5.1 Filtering method**

In the filtering method for channel selection, a search algorithm will create a subset combination of the selected EEG channels [198]. The possible combinations generated will be autonomously assessed in terms of distance measure, correlation, dependency measure, and information measure among the channels of the same combination. Apart from the advantages of high speed, independence from the classifier, and scalability, filtering methods are known to be of low accuracy [199]. This is because of their inability to consider the combinations of different channels [81].

### **2.5.2 Wrapper method**

The wrapper method uses a prediction model and its output as an objective function to calculate the subset of channels [81]. Generally, wrappers technique achieves better recognition rates than filters since they are tuned to the specific interactions between the model and the dataset.

The wrapper has a mechanism to avoid over-fitting due to the use of cross-validation measures of predictive accuracy. The disadvantages of wrapper technique are slow execution and lack of generality. The solution lacks generality since it is tied to the bias of the classifier used in the evaluation function. The “optimal” feature subset will be specific to the classifier under consideration [80].

### **2.5.3 Hybrid method**

The hybrid method is the combination of the above two techniques, which eliminates the pre-specification of a stopping criterion. The hybrid model was developed to deal with large datasets and avoid the drawbacks of filtering and wrapping techniques while keeping their advantages by using evolution techniques [200].

### **2.5.4 Embedded method**

In the embedded techniques, the channel selection depends upon the criteria created during the learning process of a specific classifier because the selection model is incorporated into the classifier construction [81]. Embedded techniques reduce the computation time taken up in reclassifying different subset, which is required in wrapper methods [187]. They are less prone to over-fitting and require low computational complexity.

A study [81], surveyed several existing channels selection techniques acknowledge their several limitations such as trial and error selection of channel subset, complexity of the process and the need of experience to select the significant channels for different applications. The study argues for a need of development of more efficient methods for channels selection.

Deep learning, a subset of machine learning, can find its features from a training data set without the need for feature engineering and poses an alternative to conventional machine learning in the analysis of biomedical data.

## 2.6 Deep learning

To address the limitations of conventional machine learning towards development of computerised ways for biomedical data analysis, this thesis applies deep learning, in particular CNN. Deep learning has the ability to learn from data which is unstructured by using multiple layers of neural network in processing the data [201]. This section introduces deep learning concepts, with a focus on CNN, the most popular deep learning model, which has been applied in studies reported in this thesis in Chapter 3, Chapter 4 and Chapter 5. The section also explains concepts of CNN and how it can be applied in EEG channels selections towards development of portable EEG devices for real life BCI applications, set out in section 2.6.2.

In a review study of the application of deep learning in healthcare, it was suggested that the use of deep learning models in biomedical data is useful due to their ability of end-to-end learning from complex data like biomedical data [188]. Deep learning models have been reported to perform better than conventional machine learning models in several cases. For example, in a challenge to predict 5-year occurrence of stroke by using Electronic Medical Claims (EMCs) data of more than 800,000 patients, CNN outperformed conventional machine learning models, including SVM and Logistic regression models.

There are several deep learning algorithms like Convolutional Neural Network (CNN), Recursive Neural Network (RNN), Restricted Boltzmann Machine (RBM) and Long Short-Term Memory (LSTM) to name a few. Originally, CNN was designed for image data as it has the ability to detect spatial information on data, but it has recently successfully been applied to other forms of data, such as in time series data including physiological signals [202] and voice [203]. RNN and LSTM have been applied in time series data due to their ability to remember what the net learned the previous time, a useful feature for stock prediction [204] and voice recognition [205].

As this thesis covers both time series and image forms of biomedical data, CNN's ability

to deal with both forms becomes an ideal deep learning model for our aims. This section provides a literature background for CNN and how it has been applied in the healthcare domain with biomedical data.

## **2.6.1 Introduction to Convolutional Neural Network (CNN)**

CNN is the most popular deep learning algorithm and has become dominant in various computer vision applications including in healthcare. The model has captured attention since its astonishing results in a general object recognition competition, the ImageNet Large Scale Visual Recognition Competition (ILSVRC) in 2012 [71]. CNN is designed to automatically extract spatial features through convolution and pooling layers and adaptively learn the extracted features through backpropagation and fully connected layers attached before the output layer.

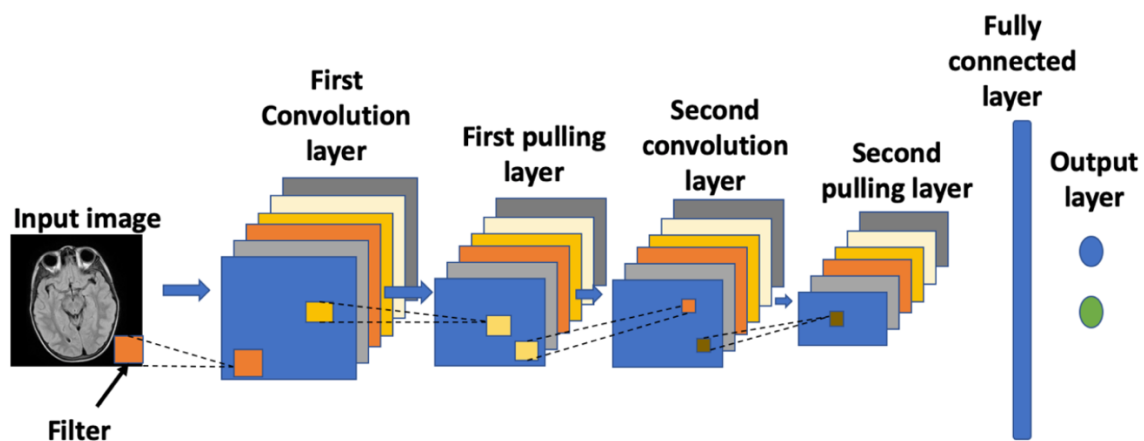
### **2.6.1.1 Background of CNN**

Today, CNN, or some close variant, is used in most neural networks for image recognition in the healthcare domain – such as in the diagnosis of diseases like cancer and TB by using X-ray image [206]. Although CNN has become popular and more applicable in image data, in recent years it has attracted attention in non-image data and has shown remarkable advantage in modelling time-series data like physiological signals [82]. Since 2006 when Hinton and Salakhutdinov applied deep learning in dimensionality reduction in big data [207], learning high-level features using deep architectures like CNN from raw data has become a huge wave of new machine learning paradigms [208].

### **2.6.1.2 CNN structure**

Generally, CNN, as with other deep learning algorithms, is composed of an input layer, hidden layers and output layers. In hidden layers, it contains multiple convolution, pooling and fully

connected layers [209], shown in figure 2.10 below. Each layer contains multiple numbers of neurons. The number of neurons in the input layer matches the shape of the input data, while the number of neurons at the output layer should be the same as the number of output classes. Each specific neuron receives several inputs features and then takes a weighted sum over them, where it passes it through an activation function and responds with an output that will be sent as an input to the neurons of the next layer.



**Figure 2. 10:** An example of a CNN architecture with two hidden layers, one fully connected layer and a 2 classes output.

CNN uses three basic ideas to extract features: local receptive fields, shared weights and pooling. The local receptive fields (patches, also known as filters), are small sliding window with a predefined shape which moves across the entire input data. This happens in the convolution layer, where most of the computational heavy lifting occurs [210]. A patch convolves over the input data, resulting in feature maps, which are the modified versions of the input. Mathematically, convolution is an operation of two functions to produce a third function, which is viewed as the modified version of one of the first two functions [211]. Weights and biases between layers, except the ones between convolution and pooling layers, are acquired by learning after being randomly initialised at the beginning of training. Unlike in fully connected neural networks, a convolution layer is advantageous because it considers the spatial structure of the features [34].

The function of the pooling layer is to down sample the feature maps obtained from the convolutional layers. At the pooling layer, these feature maps will be down sampled, only feature maps with maximum weights (max-pooling) or the average weight (average pooling) will be calculated and taken to the output layer for classification [212]. The resulting down sampled feature maps from the pooling layer become more robust to changes in the position of the feature in the input data [212]. The features extracted in convolutional and pooling layers are then used to train fully connected layers which are connected to the output layers for predictions.

At the output layer, the features maps will be flattened to a single column, and contains an activation function such as softmax or sigmoid to classify the output. The error value will be calculated between the actual output and the predicted output by using a loss function. A loss function can be defined in many different ways but a common one is MSE (mean squared error), expressed in equation below,

$$E_{total} = \sum \frac{1}{2} (target - output)^2 \dots\dots\dots(i)$$

Where  $E_{total}$  stands for the total calculated error.

The whole process from the input to the output when the error is calculated is called feed forward propagation. In order to reduce the prediction error, the network will try to update its weight by using gradient descent techniques such as Stochastic Gradient Descent (SGD), Adam or Root Mean Square (RMS). This error will be propagated back inside the model to update the weights of the neurons in the hidden layers, and that of the filters in each convolutional layer through a process in which significant features will be assigned higher weights [213]. Dense layers are updated to help the net to classify. Convolution layers get updated to let the network learn the features itself. The process will be repeated in rounds of epochs until the error value is minimised to the smallest possible value and a local minima point is reached through a gradient descent process [214]. Local minima in neural networks where the weight adjustments for one or more training patterns simply offset the adjustments

performed for a previously trained pattern One epoch is when an entire training dataset is passed through a CNN model both forward and backward only once [215]. To find the maximum number of iterations of training our models we used early stopping criterion. Early stopping is a method that allows you to specify an arbitrary large number of training epochs and stop training once the model performance stops improving on a validation dataset

## **2.7 CNN in biomedical data analysis**

CNN approaches have achieved remarkable performances in tasks such as speech recognition [93], natural language processing [216], and image classification [217]. In healthcare, CNN has been reported to have capabilities of detecting necessary features automatically despite external/unwanted signals interference [218], hence it is becoming a favourable technique for classification of biomedical data which typically contains a lot of interference. This section looks into how CNN has been applied in biomedical data analysis, with a focus on EEG for BCI applications and biomedical images, which are the major focus areas of studies of this thesis.

Thodorof *et.al* [102] used Recurrent Convolutional Neural Networks, a modified version of CNN for automatic seizures detection by capturing spatial, spectral, and temporal information from raw EEG signals. CNN-based EEG analysis has also found its application in biometrics, where Mao *et.al* [219] successfully performed biometric identification from EEG signals. In biomedical images, Angel *et.al* [220] and Jinhua *et.al* [221] both used CNN for detecting the presence of invasive tumor cells in breast cancer tissue samples.

### **2.7.1 CNN with EEG for BCI applications**

Inspired by the success of CNN on image data since winning the hand-written character recognition competition, the ImageNet [71], researchers have been adopting CNN in time series analysis [222]. The ability of CNN to learn spatially and temporal invariant features [223], a common artefact in time series data, has made them applicable in time series data



[224]. For BCI applications in particular, CNN has shown high capabilities in interpreting EEG signals. Apart from the nature of EEG signals being highly compromised with noise interference, CNN has been applied without the need for manual feature engineering [112][223], researchers managed to interpret motor imaginary activities intended or executed by using CNN, for BCI applications. In most BCI applications, the BCI system is calibrated per user. Without user specific training (re-calibration), a study managed to interpret SSVEP by using CNN and created a universal BCI system without the need of feature engineering [225]. Their method was tested on a seven (7) classes and twelve (12) classes publicly available EEG dataset, attaining 79% and 81% classification accuracies, respectively. For a while, most of the machine learning and deep learning models trained to interpret EEG signals for BCI applications, have been trained and applied for a particular BCI paradigm. With CNN, researchers managed to create a CNN based algorithm, named EEGNet, which was able to be use in across different EEG based BCI paradigms. Their CNN based algorithm was tested for both within-subject and cross-subject classification [226]. This position CNN model particularly fit for EEG based BCI applications.

### **2.7.2 CNN for EEG channels selection**

During training of a CNN model, initially the model weights are randomly initialised. During learning, these random weights will be fine-tuned through a back-propagation process as explained in section 2.6.1.2. The features with high significance and more influence to the output label will be assigned higher weights. By extracting the weights of a training model, we can be able to know the significant channels and drop the less significant channels based on the assigned weights [227]. After the non-significant channels are dropped, only few remaining significant channels can be used to design and develop portable EEG devices. For a portable EEG device to work, determining the location on the scalp is of high importance as it will be used as a guidance for channels placement. Using the weights extracted from a trained deep learning model, a heatmap can be drawn on the scalp to determine the location

were EEG signals for a particular activity are dominant, hence a portable EEG device can be placed.

### **2.7.3 CNN in analysis of biomedical images**

CNN has been increasingly applied in biomedical image analysis due to its ability to automatically extract distinguishing features which can be missed by humans. This ability makes CNN a potential tool aiding the reduction of human related diagnosis errors in biomedical images. Due to factors like inter- and intra-reader variability in reading pathology images, a common issue in biomedical imaging, researchers have turned to CNN to detect the presence of invasive tumor cells in breast cancer tissue [37]. For the diagnosis of breast cancer, CNN achieved an accuracy of 87.3%, compared to 85.8% by a Support Vector Machine working on the same task [39]. Again, another study showed a reduction of human error by 85% by combining CNN models with a professional pathologist in detection of metastatic breast cancer [228]. In another application, researchers managed to predict cardiovascular risk factors from retinal fundus photographs by using CNN [229]. CNN was also applied for the classification of Alzheimer's disease versus mild cognitive impairment and normal controls on the Alzheimer from the dataset of 3D structural MRI brain scans [14]. Indeed, the performance of CNN models in biomedical images shows its use as a potential tool for the development of computerised methods in the analysis of biomedical data.

### **2.7.4 Challenges of using CNN in biomedical data analysis**

Biomedical data are generally complex, heterogeneous, poorly annotated and always unstructured. Furthermore, biomedical data are unbalanced in most cases as it is hard to get balanced data especially when dealing with rare diseases. Unlike conventional machine learning, all deep learning models including CNN need a lot of data to be trained with. In most cases, deep CNN models contain millions of parameters to be trained, which makes them

require a large amount of data in order to generalise well [209]. For example, some of the popular CNN architecture, the MobileNet, has more than 4 million parameters [230] and VGG16 has 138 million parameters [231]. Conventional machine learning performs better than deep learning when only small training data set is available [69]. CNN has been successful in other domains like voice recognition [93] or general object recognition [232] due to the easy access to large training data, unlike in the healthcare domain. In addition, training a deep learning model is computationally expensive and time-consuming [233]. Although, due to the development of computer technologies, high computing devices have been available, not all research labs can afford the devices. The next section introduces possible solutions to deal with the data limitations of CNN.

### **2.7.5 Possible solutions to the limitation of data**

When there is a limited amount of data, data augmentation and transfer learning are two methods that can be used to improve the performance of deep learning models. Data augmentation is a technique which is used to artificially increase the diversity and size of the training sample. This is done by performing operations like shift, zoom, flip, and rotation on the available training data. Study [232] reported how data augmentation contributes to the performance of deep learning models in the classification of medical images. The authors further analysed which data augmentation techniques retain the properties of the original image as it contributes to the model performance. Another study [234], has surveyed the significance of different data augmentation techniques in fields that are difficult to get enough training data sets, including healthcare.

Transfer learning, on the other hand, involves the use of pre-trained models designed to perform a particular task in one context. Their top layers are removed and new layers attached, to be trained to perform a new task, in a different context. Since only the last few layers of the model will be retrained, computation costs and training time are reduced and only a smaller

amount of data is needed. There are several pre-trained models available, such as GoogLeNet [235], VGG [236], MobileNet [230], and InceptionResNet V2 [72]. Most of the existing pre-trained models for image classification are trained on the Large-Scale Visual Recognition Challenge (ILSVRC) ImageNet dataset [237], a general object recognition dataset that includes 1000 classes [238]. For instance, researchers compared three popular models – VGG 16, VGG19 and ResNet50 – for the classification of breast cancer into benign and malignant, using both transfer learning and training from scratch. It was found that with transfer learning, both models significantly outperformed the non-transfer learning method. The transfer learning scores were VGG16(AUC-95.65%) and VGG19(AUC-91.85%), while the non-transfer learning scores were VGG16 (AUC-75.00%) and VGG19 (AUC-65.40%). Transfer learning was again used to classify lung cancer types in a study by Nicolas et al. [220] with a high level of accuracy (0.97 AUC- Area under the Receiver Operating Characteristic Curve). Using Inception V3, a pre-trained model on ImageNet, it was also possible to predict the ten most mutated genes in Lung Adenocarcinoma (LUAD)[220], using the same model [220]. Mark et al. [233] explored several deep learning architectures (VGG16, Inception and ResNet50), to detect glaucomatous optic neuropathy in fundus photography with transfer learning outperforming training the models from scratch. J. Huang et. al [239] proposed InceptionResNet V2 and MobileNet as a feature extractor together with the Faster Regions Convolutional Neural network (R-CNN) as a meta-architecture for the image identification and detection. InceptionResNet V2 has provided a good accuracy at the cost of more computational time, whereas MobileNet’s accuracy is comparable to the VGG model but has 1/30 of the computational cost. Google’s Inception V3, a deep learning model pre-trained on ImageNet, was used to perform the classification of histopathological images of breast cancer [96].

## **Summary**

This chapter has highlighted a need for the development of computerised ways to analyse of biomedical data. More specifically, the discussion focused on EEG and microscopic images, which are the focus of the main studies reported in the thesis. Conventional machine learning has been reviewed as a possible tool, and its limitations analysed in section 2.3. The limitation towards realising BCI applications in real life due to the size of the current existing EEG devices has also been discussed. To address the limitations of conventional machine learning, deep learning, particularly CNN was suggested in section 2.5 above. The major challenge of applying deep learning in biomedical data has been highlighted and possible solutions given. The next chapters demonstrate the application of CNN in EEG signals for BCI applications (Chapter 3) and biomedical images (Microscopic images in Chapter 4).

# Chapter 3

## Convolutional Neural Network for interpretation of EEG signals and channels selection for BCI applications

### 3.1 Introduction

This chapter reports the first two main studies of this thesis, both of which focus on the analysis of EEG signals in Brain to Computer Interface (BCI) applications, with the aim to provide answers to research questions of case study one as highlighted in section 1.5 of chapter 1. Case study one comprises of two studies, using two common EEG triggering mechanism for BCI application, the Motor Imaginary (MI) and the Steady-State Visual Evoked Potential (SSVEP). By using Motor Imagery (MI) or Steady-State Visual Evoked Potential (SSVEP) EEG paradigms, a computational algorithm can be developed to interpret an individual's movement intention or visual stimulation. In BCI applications, the interpreted output is then used to help an individual to interact with his/her environment without having to make any physical movement. By using Convolutional Neural Network (CNN) as a computational model, the studies reported in this chapter address three major challenges to the implementation of real-

life BCI applications. These challenges are:

a) The limitation of using conventional machine learning due to the need for feature engineering by using CNN with minimal data pre-processing;

b) The limitation of current existing EEG devices due to their size and the complexity required in using them by using CNN to perform channels selections; and

c) Determining locations on the scalp for the placement of fewer electrodes of portable EEG devices by analysing the weights of a trained CNN model.

This chapter is divided into two main sections, Sections 3.2 and 3.3, which report on the application of CNN on interpreting MI activities and SSVEP, respectively.

MI and SSVEP are among the most common paradigms of EEG for BCI applications. In the first EEG study (section 3.2), we first investigate the use of CNN in interpretation of MI activities from raw EEG signals without the need of feature extraction (section 3.2.3). Then, a method for channels selection by using the weights learned by a CNN model was extensively studied and reported in section 3.2.4. To test the consistency of our method for channel selection by using the weights learned by a trained CNN model, the technique was further tested on a publicly available MI EEG data set as reported in section 3.2.5. Findings in this study were subsequently implemented in the second EEG study.

In the second study (section 3.3), we performed EEG channels selections and determining significant points on the scalp for fewer electrodes' placement (section 3.3.1) by using a 32-channel EEG device. Then, a 2-channels, portable and skin-like EEG sensor was designed and fabricated (section 3.3.2). The real-life application of BCI by using the newly fabricated 2-channels EEG device was demonstrated in controlling a wheelchair, robotic car, and a PowerPoint presentation, as demonstrated in section 3.3.3. The conclusion of the two studies is reported in section 3.4.

## **3.2 Study 1: CNN in interpretation Motor Imaginary activities for BCI**

Motor imaginary (MI) is one of the most common ways of triggering EEG variations for BCI applications. In this study, a Convolutional Neural Network (CNN), a deep learning algorithm, was implemented to interpret five (5) classes of imagined upper limb movement from raw EEG signals acquired from transhumeral amputees without handcrafted features. This is among the few datasets obtained from real amputees. The possibility of applying CNN methods for real-time application was investigated by experimenting with smaller window sizes of input EEG data. Subsequently, channel selection was extensively explored using the

developed CNN model to reduce the number of EEG channels to the lowest possible number while maintaining good classification accuracy. The significant locations for fewer channels' placement on the scalp of the participants corresponding to the multi-class motor imagery tasks were also identified and reported. The channel selection method was then tested on a widely used publicly available motor imagery dataset, commonly known as the physionet [240] to further test the effectiveness of our proposed method.

This was a collaborative study with Shenzhen Institute of Advanced Technology (SIAT), in Shenzhen, China. Previously, researchers at SIAT performed interpretation of the obtained EEG signals by using conventional machine learning methods and their results are reported in Table

3.3 of this chapter and published in [138], in comparison to our results by using CNN as reported in this section.

### **3.2.1 EEG Data from Amputees**

This section provides the study description and the data acquisition process,

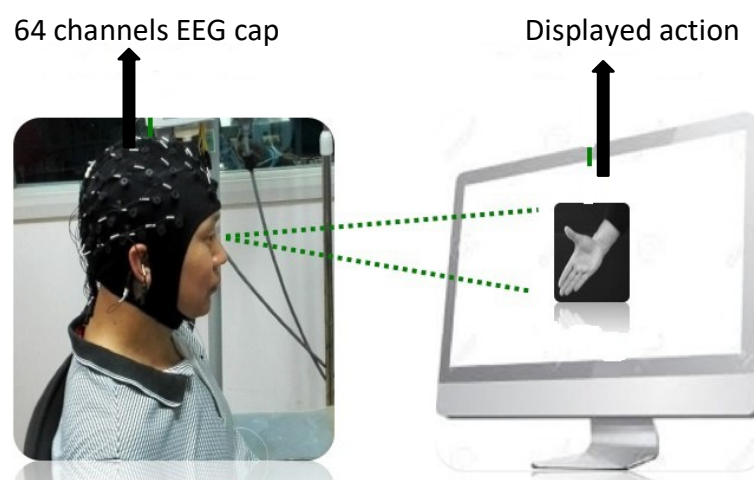
#### **3.2.1.1 Study description**

Four male transhumeral amputees (with an average age of  $41.50 \pm 7.05$  years and mean residual limb of  $25.50 \pm 4.20$  cm as measured from the shoulder blade downwards) were recruited. The participants willingly permitted the publication of their photographs/data for scientific and educational purposes in a written approval. Data collected was carried out by our collaborators from Shenzhen Institute of Advanced Technology. The Institutional Review Board of Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China, approved the study protocol.

#### **3.2.1.2 Equipment setup and data acquisition**



For data collection, we utilized a commercial EEG signal acquisition system (EasyCap, Herrsching, Germany) incorporated with the Neuroscan software (version 4.3), sampled at 1000Hz. The equipment has 64-channels of Al-AgCl electrodes on a cap. The electrodes were distributed over the scalp of a participant based on the 10-20 international system standard. Before the placement of the electrode cap, each participant's hair was properly washed to ensure that high signal quality is acquired. Besides, the impedance between each electrode and the scalp was maintained below 10 k $\Omega$ . After setting up the signal acquisition system, participants were asked to sit straight facing a computer screen where five different motor imagery (MI) tasks were displayed. The five MI tasks include hand close (HC), hand open (HO), wrist pronation (WP), wrist supination (WS), and no movement (NM). Each MI task was displayed for 5 seconds. Participants were asked to perform the action continues throughout the 5 seconds. To avoid mental fatigue, which often affects the quality of the EEG signals, a rest session of 5 seconds was introduced between any two consecutive MI tasks. To acquire a significant amount of dataset for training, validation, and testing of the proposed CNN model, each participant accomplished five experimental trials. In every trial, each MI task was repeated ten times thus producing 50 seconds of data recordings per MI task per trial. This results into 51200 data points for each trial. Figure 3.1 shows the experiment setup.



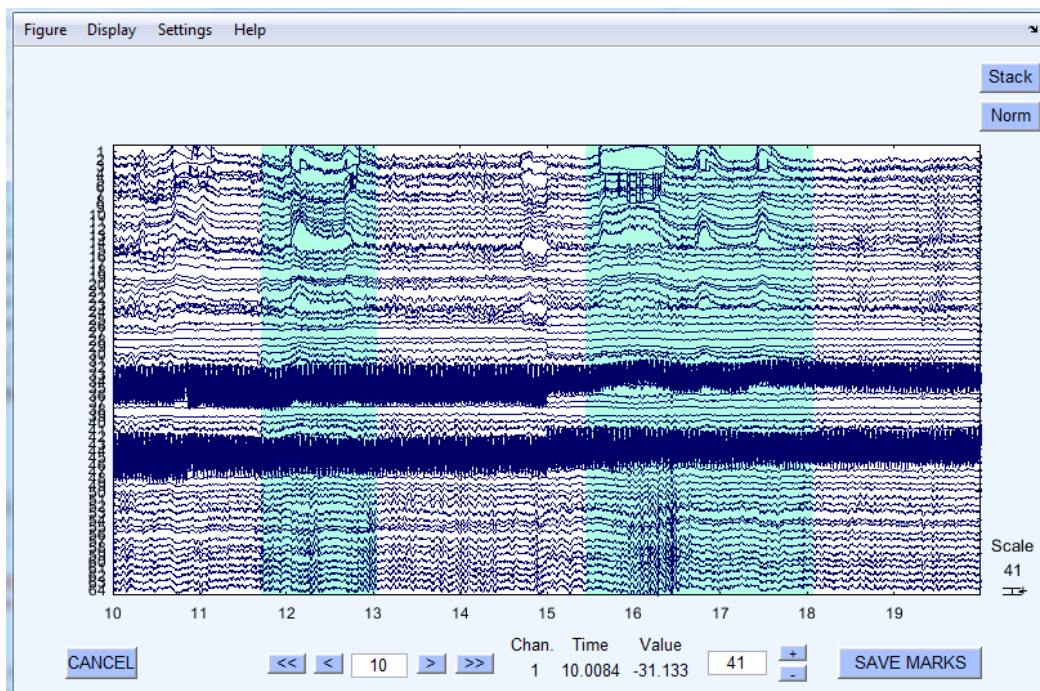
**Figure 3. 1:** Experimental setup. A subject sit in front of a monitor display, asked to perform different motor imaginary activities as displayed on the monitor[138].

## 3.2.2 Data Analysis

This section we describe all the pre-processing done on the data, the CNN model building and the training of the model,

### 3.2.2.1 Data pre-processing

The dataset is in the time domain and considering each participant's data, a moving window (non-overlapping) was used to slice the input resulting in samples of dimension  $(W_t, 64)$ . Different window length  $W_t$ , was tested to imitate the possibilities of real-time classification. 2-D image like data, where the width (64) represents the spatial structure whilst the height ( $W_t$ ) represents the temporal structure [101]. For every trail which produces 51,200 data points, each trial was divided with the window length to produce multiple windows of data. The used windows were the 0.5 Second, 1 Second and the 2 Seconds. 1 second window contained 1000 data points. For training purposes for each window size, 80% of the dataset was used (of the windows produced), while the remaining 20% were used for testing. Batches of 128 were fed into the CNN model. Figure 3.2 shows EEG data sample record.



**Figure 3. 2:** EEG sample record using a 64 channel EEG device, the visualisation was produced by using EEGLAB Open Source MATLAB Toolbox for Physiological Research.

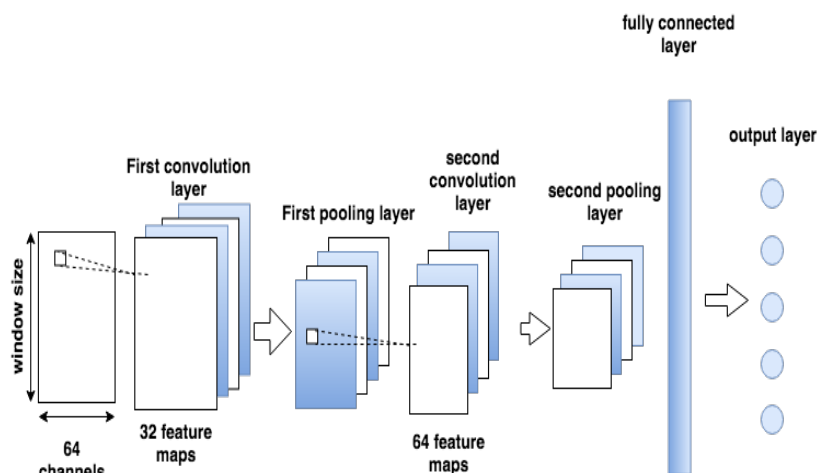
### 3.2.2.2 CNN architecture

Generally, CNN has two main parts that include the feature extractor and the trainable component. The feature extractor part contains multiple layers of convolution and pooling. The convolution layer is able to learn and extract features from the raw data automatically and the pooling layer is used for down sampling. The trainable part contains a fully connected multilayer perceptron, which performs classification based on the features learned in the feature extractor part.

The developed CNN model was trained, validated, and tested with each participant's data separately. As the real-time application is one of the main aims of this study, small and simple architecture was preferred out of other tested models as smaller models are computationally affordable to run on low-cost smart devices during real-time applications. The network had two convolutional layers, and each layer had filters of size (5, 5), with a stride of 1, to form feature maps after dot product between the filter and the input. A single layer CNN model was also tested with poor results. We opted for a simple CNN model with low computational expenses for real time application considerations. These filters are aimed to capture different local spatial and temporal patterns/features related to imagined motor activity. 32 and 64 filters were used in convolutional layer 1 and convolutional layer 2 respectively, hence 32 and 64 feature maps were formed in the first and second convolutional layers respectively. After every convolution layer, the *ReLU* activation function was applied. Max pooling was used after every convolutional layer as it is considered to better improve the network performance [241].

One fully connected layer was used to perform classification. The output layer had five nodes with softmax nonlinearities and the output class was assigned to the one with maximum among the five (argmax). The fully connected layer combined all the feature maps at the output of the last convolution layer. To avoid overfitting, a dropout regularization technique was

applied to the fully connected layer. 75% dropout was used as the optimal dropout after a trial of several dropouts. Figure 3.3 shows the CNN architecture.



**Figure 3. 3:** A 2-layer CNN, with one fully connected layer and five class outputs.

### 3.2.3 Classification results

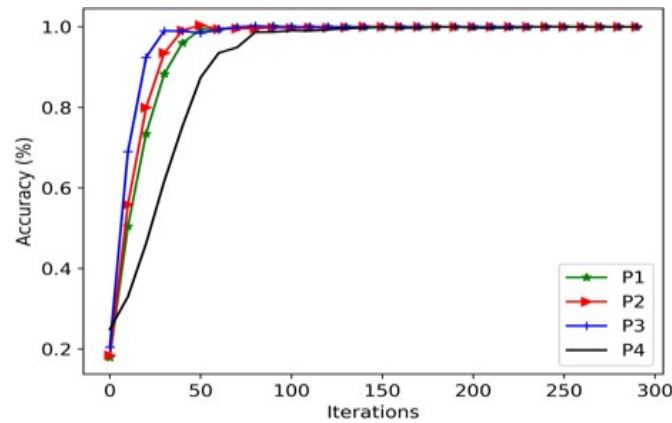
With the proposed CNN model, an overall average accuracy of 99.7% was recorded with a 0.5 second moving window (no overlap), as the performance metrics shown in Table 3.1 across all participants. Accuracy was measured as a percentage of the total number of samples correctly classified over the total number of samples classified. Different window sizes were used to test the possibilities of real-time classification. As shown in Table 3.1, one second window and two seconds window were tested, and the implemented model maintained a highly classification accuracy for both windows.

**Table 3. 1:** Model performance on different window sizes

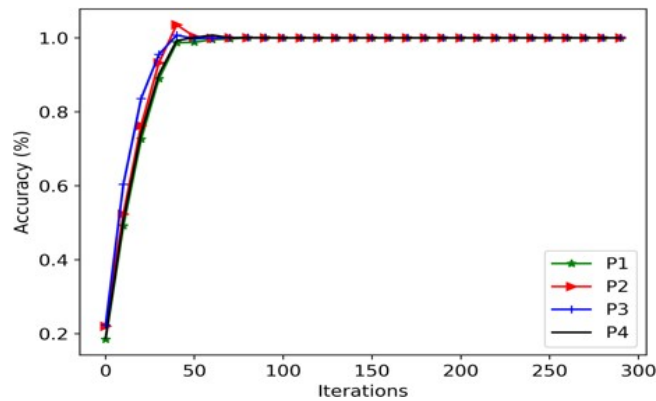
Window size (seconds)	Average accuracy (%)	Average precision (%)	Average recall (%)	Average F1 score (%)
0.5	99.75	99.20	100.00	99.49
1.0	99.23	98.47	99.29	98.49
2.0	99.00	100.00	100.00	100.00

The learning curves for all four participants, for 0.5 Second, 1 Second, and 2-seconds window sizes are shown in Figures 3.4, 3.5, and 3.6 respectively. It took a little bit longer for the model

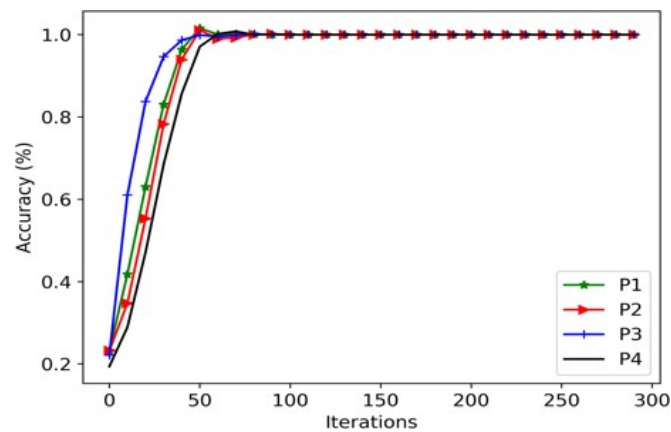
to converge when a smaller window was used as it can be seen in Figures 3.4-3.6. One possible reason for this observed phenomenon is that our dataset is larger for smaller window hence the minibatch corrections might oscillate the learning. P1, P2, P3 and P4 stands for Participants one, two, three and four respectively, all denoted with different colours.



**Figure 3. 4:** Learning curve for 0.5 second window length



**Figure 3. 5:** Learning curve for 1-second window length



**Figure 3. 6:** Learning curve for 2 seconds window length

Table 3.2. Shows the performance of the conventional machine learning models as reported on the previous work on the same dataset [138], in which spectral and time domain features extracted from a 2 seconds window of EEG recordings in frequency domain were used to train conventional machine learning classifiers, the linear discriminant analysis (LDA), artificial neural network (ANN) and k-nearest neighbors (kNN).

**Table 3. 2:** Conventional machine learning model’s performance

	LDA	ANN	kNN
Average accuracy (%)	97.81	96.44	96.92

### 3.2.4 Weight analysis

To be able to perform channels selection and generating topographic, we extracted weights from a trained CNN model and performed analysis,

#### 3.2.4.1 Topographic maps

To identify the best discriminant channels for EEG classification, we extracted weights from the trained CNN model for each participant and constructed the topographic maps. With reference to EEG data, topographic map is the map of the distribution of EEG electrodes on the skull of an individual. For weight extraction, to avoid mixing up features contained in different channels, we retrained our model with the same architecture used above but we changed the size of the filters from a matrix of 5\*5 to a vector of 20\*1 and stride of (1,1) for the filter to convolve along a single channel at a time. This modification was done only on the first convolutional layer and before the first pooling layer. We then extracted the weights from the feature maps formed in the first convolution layer. The Sum of weights of all the feature maps was obtained as we have 32 scalars in the first convolution layer. That is, the 20\*1 filter convolving on the 100\*64 input window to form 32 features of the dimension as the input window. We used the first convolution layer as the feature maps are still in the same

shape since the dimension of the channels are still preserved as the original dimension of the input before maxpooling. The weights were calculated using the following formula in (2).  $l$  stands for window length, which is 500 for this study,  $p$  stands for the number of channels which is 64 maximum, and  $q$  stands for the number of filters/ feature maps which is 32 in the first convolutional layer,

$$W_p = \sum_{q=0}^{q=W_l} W_{(p,q)} \quad (2)$$

Where  $0 \leq p < 64$  and  $0 \leq q < W_l$

### 3.2.4.2 Channel reduction and performance analysis

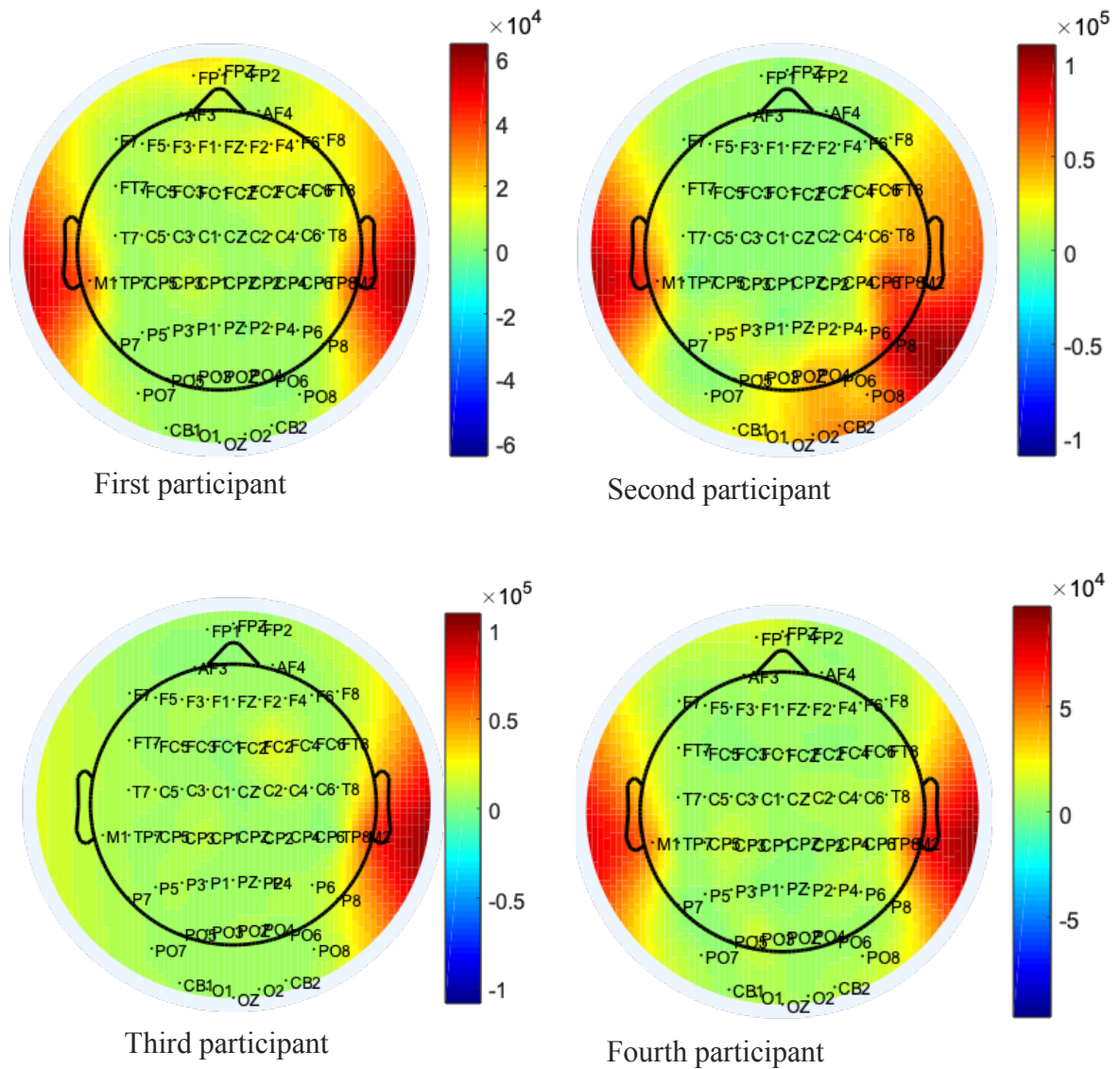
Training a model with all available channels takes time and hence affects the inference speed of the model as it learns from a large number of channels, some of which are not related to the motor activity of the brain. Identifying the necessary channels will help in the design and manufacturing of smaller, more portable, and cheaper EEG devices with fewer channels. Using these portable EEG devices with fewer channels, we will need precise placement of the channels on the scalp. From the topographic maps generated based on the extracted weights from the trained CNN model, it can be seen that the channels with the highest weights are around the motor cortex region of the brain, indicating that these channels are the most discriminating ones.

To perform channel reduction, the summed weights obtained from each channel were first arranged in descending order. Then the first twenty (20) channels with the highest weights were selected for each participant. The number 20 was randomly selected based on the topographic shown in figure 3.8 below. For each participant, the network was then retrained using these 20 channels. We then carried out a systematic analysis, reducing the number of channels down to two (2) in ascending order of the weights across all the participants. Table 3.3 shows the accuracy results of channel reduction for all participants, based on a 0.5 second window using

the two hidden CNN model reported in this study. The model trained with fewer channels maintained reasonable accuracy when fewer electrodes up to eight channels with higher weights are selected. This is because the models learn from less but significant channels' data. When fewer channels are used, the training time was significantly reduced as now the model learns from fewer but significant channels. Significant channels are the channels that carries most of the EEG variations related to the performed activity. A small variation of the significant channels was observed, this might be the case of variation during electrode placement. Our results show that by selecting fewer channels located around the motor cortex region we can still obtain a reasonably high classification accuracy in comparison to using all the 64 channels. It worth noting that, the channels with higher weights were not identical among the participants. This might be due to variation in electrodes placement variations among participants.

The topographic maps created from the weights extracted from the trained model show that the regions with the highest weight are common among participants and they are situated around the motor cortex region of the human brain, as seen in Fig 3.7. It should be noted that the third amputee reported Phantom Limb Pain (PLP) when he was performing a motor imagery task during data collection. Since the left motor cortex region of the human brain controls the right side of the body and his amputated arm was on the right side, we believe that the participant was not performing enough motor imagination due to the pain reported [88]. This can be seen in Figure 3.8 for the third participant, which shows high weight values on only one side of his brain. And it has been previously reported that amputees with PLP generally have worse motor control over their phantom hand [242], [243].





**Figure 3. 7:** Topographic maps showing the region of the brain with the most contributively channels of the four participants.

**Table 3. 3:** Channel selection based on our acquired dataset

Participants	Number of Channels and Accuracy (%)						
	64	20	16	12	8	4	2
<b>P1</b>	100.00	99.60	99.12	98.01	92.20	66.18	47.20
<b>P2</b>	99.10	100.00	98.75	97.40	93.18	71.23	86.45
<b>P3</b>	99.15	99.35	99.34	98.28	92.52	91.26	44.40
<b>P4</b>	99.23	98.67	95.67	96.45	89.34	69.47	62.36

### **3.2.5 Evaluation of our method on the publicly available data set**

This performance of the proposed CNN model for channel selection was further evaluated on a different EEG dataset which is publicly available from the PhysioNet eegmmidb (EEG motor movement/imagery database) database [243]. The data set was collected using a 64 channels BCI200 EEG system at 160Hz sampling rate, with 109 participants. Each participant performed four tasks with a rest between each task. The participants performed three trials while sitting in front of a computer screen. A resting-state was included as a class in our case. To perform the tasks, the participants were instructed as explained in the PhysioBank Automated Teller Machine [240]. For our experiment, four participants were randomly selected to match the number of participants we had in our own dataset. 32,000 labelled instances of EEG data were used from each participant.

Accuracy was used as a performance measure for evaluating the proposed channel selection method on this data set. The same methodology was used to perform channel selection, as described in section 3.2.4.2. We eventually reduced the number of channels from 64 to 12 and even further lower down to 2 based on the computed weights associated with channels. A drop of accuracy in the range of 2 ~ 3% was observed when the channels were reduced from 64 to 12, as seen in Table 3.4. Meanwhile, at 12-channels, the accuracy seems to be reasonably high and stable and begins to drop when the channel numbers were below 12. Note that the proposed CNN model was trained per instance record of EEG record as it was done in [244], the model learned most of the spatial features and less temporal features. We believe this affected the performance when less than 12 channels were utilized compared to the case of amputee's data in which a moving window was used.

**Table 3. 4:** Channel selection based on Physionet data

Participants	Number of Channels and Accuracy (%)						
	64	20	16	12	8	4	2
P1	96.10	93.23	94.18	92.19	70.28	30.78	24.90
P2	95.77	95.25	94.12	93.27	80.45	29.23	22.13
P3	98.19	96.49	93.54	92.76	79.34	49.36	42.23
P4	99.50	96.18	93.78	90.35	79.76	50.34	22.27

### 3.3 Study 2: CNN in interpretation SSVEP EEG signals for BCI

Another common method for triggering EEG variations and mostly applied in BCI applications is by using a flickering source of light, which results in the generation of Steady-State Visual Evoked Potential (SSVEP). This study reports the application of CNN in interpretation SSVEP for BCI application. SSVEP is an electrical signal generated by the brain (EEG), as a resonance phenomenon evoked when a subject looks at a light source flickering at a constant frequency [132]. Conventional EEG systems typically suffer from motion artefacts, extensive preparation time, and bulky equipment. This study applies the findings in the first study for EEG channels selection (reported in section 3.2.4.2 above) and determining scalp location for placement of the portable EEG devices with few channels by using the CNN model (reported in section 3.2.4.1 above). A 2-channels, fully portable, wireless, flexible scalp EEG device (referred to as ‘SKINTRONICS’) was designed, fabricated and applied in this time-domain analysis of EEG signals using CNN allows for an accurate, real-time classification of steady-state visually evoked potentials on the occipital lobe. With the developed 2-channel scalp EEG device, we managed to perform a wireless, real-time, universal EEG classification for an electronic wheelchair, robotic car, and PowerPoint presentation. We demonstrate a new capability to train a CNN model offline and integrate them into wireless mobile devices for a

real-time, universal EEG classification. Our method achieved an average of 94.54% in offline classification and 94.01% in online classification accuracy at a 0.5-second window of the incoming raw EEG data. The off-line term means the model was trained and tested with the recorded EEG data, the online means a model was trained with the recorded EEG data but tested in real-time application.

This is a joint study with the Georgia Institute of technology. Our role in the study was to perform EEG channels selection for SSVEP paradigm, locate the scalp location for placement of a few channels EEG devices and perform classification of the 2-channels EEG signals by using CNN as a computational model without the need of feature extraction. The role of the Georgia Institute of technology was to design and fabricated a few channels portable EEG device. The implementation of the study is divided into two major areas, a) The application of a 32 channels EEG devices to determine the significant channels for SSVEP applications and the scalp location for placement of a portable few channels EEG device, and b) The application of a fabricated 2-channels EEG device in BCI applications.

### **3.3.1 Determining significant Channels and scalp locations with a 32 channels EEG device**

During brain activities, EEG signals can be detected in more than 100 different locations on the scalp [105]. To detect a specific region of the brain which becomes active during a particular activity, an EEG device with many channels enough to cover the whole scalp needs to be used. For this purpose, a 32 channels EEG device was used in this study.

#### **3.3.1.1 Data collection**

##### **Study description**

Six health subjects were recruited. A 32-channel EEG system (ActiveTwo System, Biosemi B.V.), using water-based conductive gels (SignaGel, Parker Laboratories) was used for data

collection, with a sampling frequency of 256Hz. All the experiments, data collection and demonstrations were conducted in a lab environment at Georgia Institute of Technology. Figure 3.8 shows a participant with a 32- channel EEG system (ActiveTwo System, Biosemi B.V.)



**Figure 3. 8:** A study participant wearing 32 channels EEG devices with electrodes covering the whole scalp [245].

### **3.3.1.2 Methodology**

This section provides the methodology conducted in this study, to acquire the data and training of the model,

#### **I. Equipment setup**

Subjects were asked to focus on four different LEDs, one at a time, flickering at different frequencies (11.1, 12.5, 15.2 and 16.7 Hz) and alpha rhythms. We use alpha rhythms as the null class (no target interface action), intended for the user to relax their eyes in case of fatigue. Alpha rhythms are produced when an individual one closes his/her eyes. For the experiments, subjects were seated in front of the LED stimulus set up, about 0.8 m away from their head at eye level where all four stimuli are presented simultaneously.

#### **II. Data acquisition**

The procedure for gathering training data involved a continuous EEG recording of 15 s for

each class, separated by auditory cues. The classes were recorded in order of alpha rhythms, for the first 15 s, followed by 15 s of gazing at four different LED stimuli. Here, each recording where all five classes are performed constitutes a single trial. This trial was performed six times for each subject. The experimental results showed that training from a continuous stimulus was significantly more effective than a fast ( $<1$  s) cue-guided task. This may be because the relevant SSVEP features required for training are most optimal during these longer sessions of looking at the stimulus.

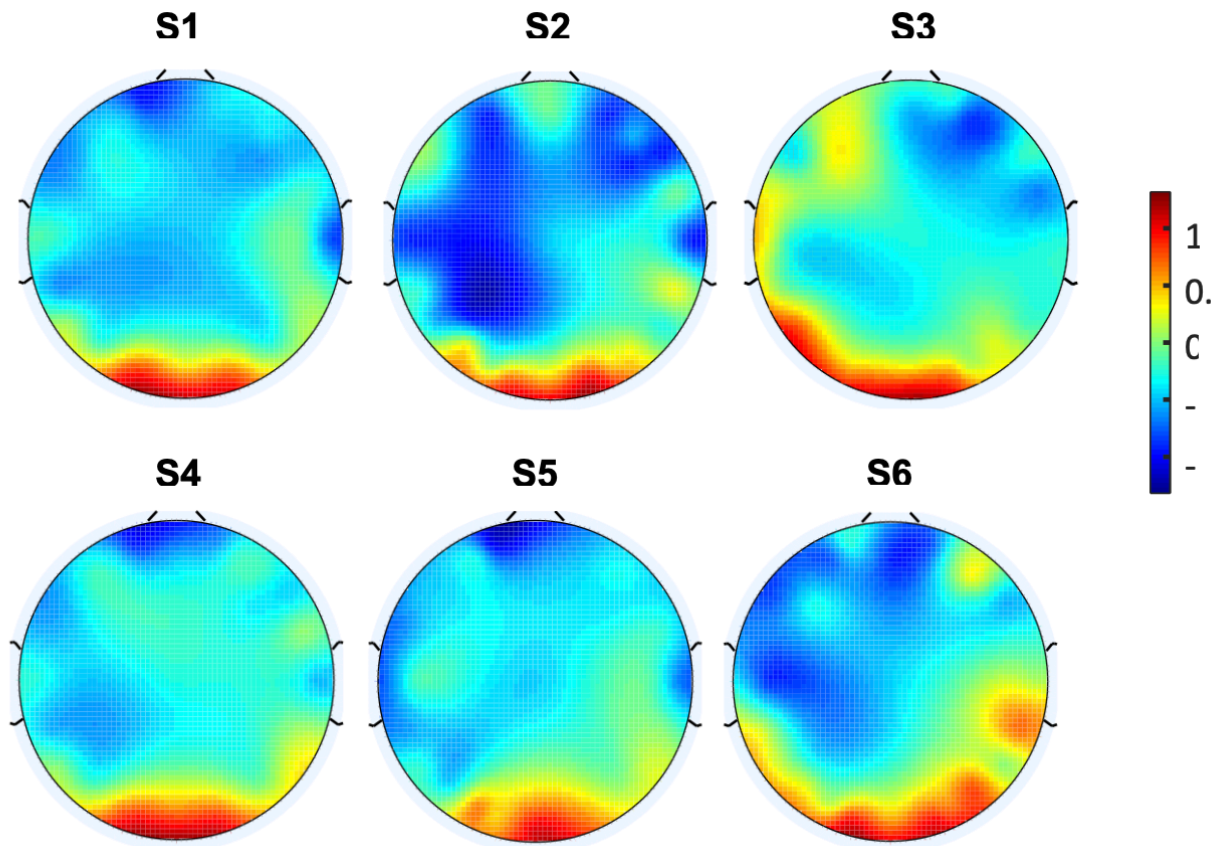
### **III. Model training**

A 2-layer Convolutional Neural Network (CNN) was then trained to classify the data obtained using the 32 channels EEG device. The input was split into epochs of 300 data points (1.172 seconds at 256 Hz). Each epoch was high-pass filtered at 4 Hz and normalized from the range of 0 to 1. Therefore, the inputs were 32 channels x 300 data points, and convolutions were performed along the time domain signal to prevent data mixing across channels.

#### **3.3.1.3 Weights analysis**

##### **Topographic maps**

From the trained network, relevant weights correlated to information presence were extracted and associated with specific channels. Electrode placement was then decided based on the channel locations with the most significant weights. The weights are extracted from the feature maps formed in the first convolution layer, before the pooling layer of the CNN. This is because, in the first convolution layer, the input is still in the same shape as the original input data. The sum of weights of all feature maps was obtained as we have 32 feature maps in the first convolution layer. The weights for each feature map were calculated using the formula as in equation (2) above. A resulting scale showing the activation weights for each channel over the 32 feature maps are shown in Figure 3.9.



**Figure 3. 9:** The topographic maps created from the weights extracted from the trained model show the regions with the highest weights are situated around the virtual cortex region of the human brain.

This exercise is meant to be a proof-of-concept method for mapping biopotential signals from a range of sources with the intention of reducing the number of electrodes and the amount of data to process.

### 3.3.2 2-Chanel EEG device development

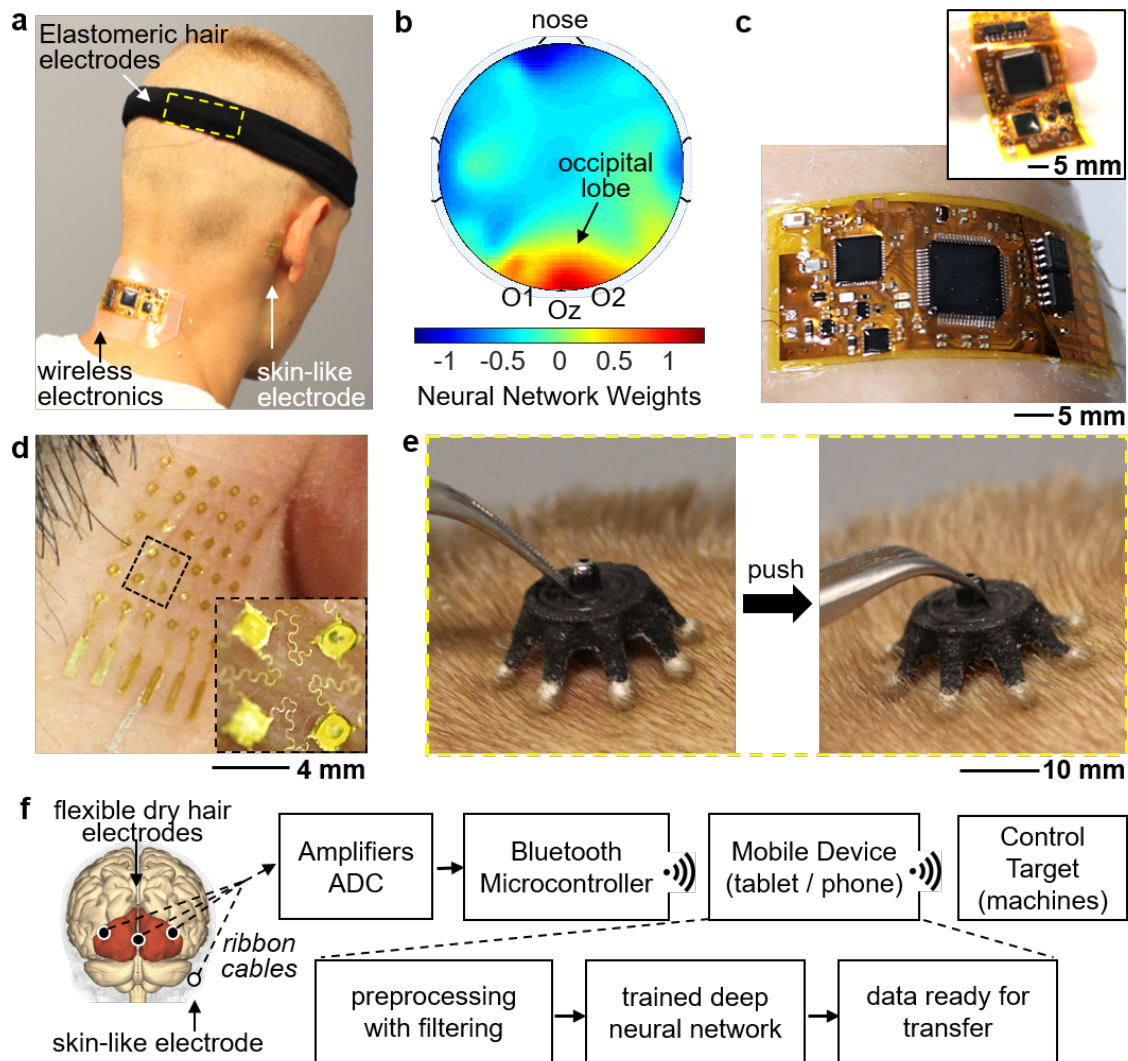
After determining the visual evoked potential were dominant on the visual cortex region of the brain which was covered by just two channels of the 32 channels device, a two channels EEG device was designed and fabricated. The analysis of the weights learned by a CNN model during the use of a 32 channels EEG data is used to isolate the best electrodes from a larger cluster of electrodes, as it does not require prior knowledge of the signal type. This method is useful when the signal's features are difficult to decompose by conventional methods (e.g.

power-spectrum analysis). The photo in Figure. 3.10a captures the design concept of the flexible electronics that minimize the contact area (only 2 channels) on the scalp for a comfortable, dry EEG recording. For precise placement of the 2-electrodes on the scalp for data collection, Figure. 3.10b shows the optimal electrode locations (O1, Oz, and O2) as determined by the analysis of the weights extracted from the topographic maps of a CNN model trained by a 32-channel EEG recording of SSVEP data (detailed in section 3.3.1.3 above). Figure 3.10c shows an overview of a 2 channels wireless, portable scalp electronics for SSVEP-based BCI.

The fully flexible, wearable system enables real-time long-range wireless data acquisition and accurate classification of SSVEP with a high information transfer rate from only two recording channels. Due to the removal of unwanted electrodes for this purpose, SKINTRONICS exhibits a significant reduction of noise and electromagnetic interference, compared to the existing portable EEG systems with rigid electronic components [246]–[248]. Additionally, the use of conformal electronic components allows for easy wearability on the back of the neck or other bare skin locations. This soft system allows for a long-term wear versus other rigid devices with heavy plastic enclosures that hold the electrodes distributed all over the scalp. A set of dry, flexible elastomeric electrodes (Cognionics) that make intimate contact with the hairy region was used during the making of a 2 channels EEG device, resulting in long-term EEG recording [249]. With adequate skin preparation, conformal contact provided by these electrodes allows for superior skin impedance (less than 20 k $\Omega$ ), and therefore lower noise in signal recording and transmission. Due to the use of the flexible contacts of the scalp-mounted dry electrode and the skin-like electrode on the mastoid, we maintained the skin– electrode contact impedance at a relatively low level. Also, there are only three scalp electrodes, effectively secured by using a single headband, which allows the electrode to splay its legs, separating and moving hair, to make effective contact with the scalp as demonstrated in Figure 3.10e. A flow chart in Figure 3.10f summarizes a high-level overview



of data collection, processing, wireless transfer, and machine control.



**Figure 3. 10:** 2-channels EEG system, data acquisition and the BCI system set up [245].

### 3.3.2.1 2 Chanel EEG data acquisition

By using a newly 2-channels fabricated portable EEG device, same experimental setting was used as in 32 channels experiment for data collection by using six participants. Target skin and scalp locations are cleaned with isopropyl alcohol before skin preparation. Abrasive skin preparation gel (NuPrep, Weaver and Co.) was gently applied to each of the electrode locations with a cotton swab to prepare the skin for electrode placement. Excess gel is removed using a gauze pad and the locations are cleaned with alcohol wipes before electrodes

are applied. For the single skin electrode location, adhesive tape is used to remove dead skin cells from the surface and alcohol wipes are used to clean and prepare the skin. Impedances between the hair mount electrodes to ground were maintained below 20 k $\Omega$ . Due to the use of the flexible contacts of the scalp-mounted dry electrode and the skin-like electrode on the mastoid, we maintained the skin– electrode contact impedance at a relatively low level. The recording procedure for training and test data allowed for an additional 0.3 s for gaze-shifting between stimuli. Each subject performed six trials.

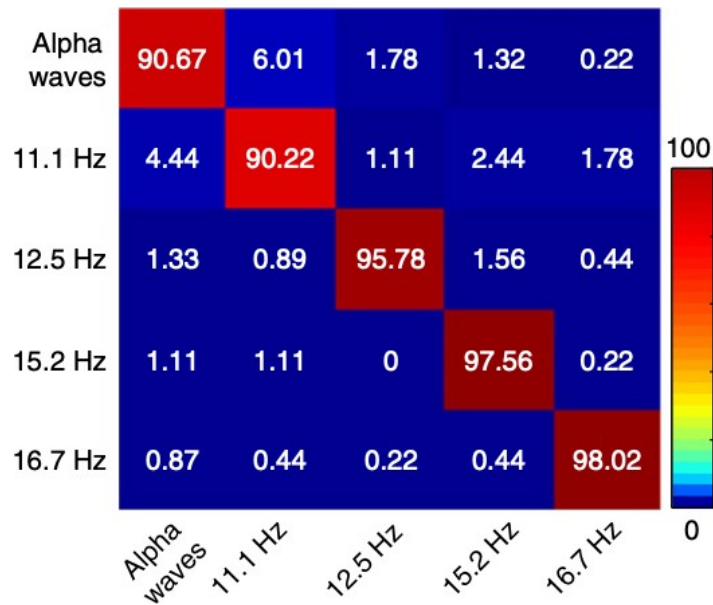
### **3.2.2.2 Data pre-processing and model training**

For all experiments, a 250 Hz sampling rate was used with the SKINTRONICS system. The training data were subdivided into window sizes of 128, 192, 256, 384 and 512 data points (corresponding to 0.512, 0.768, 1.024, 1.536 and 2.048 s data windows) to gauge changes in accuracy over different window sizes. The main aim was to use the smallest window possible to manifest the real-time application. From the training set, there were 2,700 samples from each subject used, 450 samples from each of the 6 recordings. Therefore, for 5 subjects, a single training epoch consisted of training on 13,500 samples.

Data from five subjects was used for training of the model, a sixth subject's data was used to test for universality of our proposed method. Since the universality was tested on only a single subject which is not big enough, our experiment opens the door of further experiment of using our system for universality experiments. For the five subjects, five epochs were used for training and validation and the sixth epochs for testing. After several parameter testing, a training batch size of 64 and learning rate of 1e-3 was selected. An early stopping training criterion was used for training of the model. An average accuracy of  $94.54 \pm 1.10\%$  was achieved across all test subjects for 0.512 sec inputs in the time domain. The table below show the model performance at different window sizes. The confusion matrix is shown in figure 3.11.

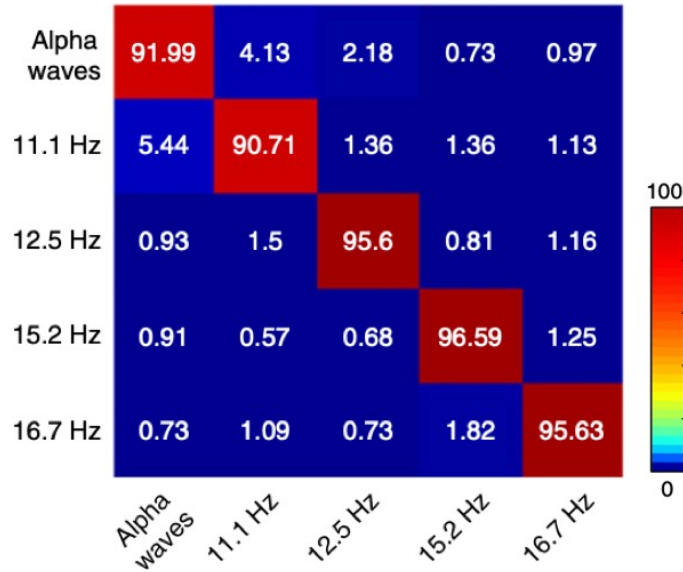
**Table 3. 5:** Testing of different window sizes for each subject, to imitate real-life applications.

Data Length (sec)	Classification Accuracy Per Subject (%)						Average	Standard Error ( $\pm$ )
	S1	S2	S3	S4	S5	S6		
0.512	95.33	95.11	90.09	93.71	96.44	96.53	94.54	0.90
0.768	97.73	95.45	92.73	99.32	96.59	96.59	96.40	0.99
1.024	98.84	96.74	97.00	99.77	96.28	97.49	97.69	0.69
1.536	99.27	96.83	98.05	100.00	97.56	99.32	98.50	0.51
2.048	99.74	99.23	98.46	100.00	99.74	99.69	99.48	0.29



**Figure 3. 11:** Confusion matrix which shows the performance of our system at a 0.5 seconds window of the incoming raw EEG signal.

Our collaborators from Georgia Institute of Technology successfully demonstrated a real-life application of the technology by performing an online test by using our developed CNN model to classify the SSVEP EEG data extracted by the newly developed 2-channels EEG device. Their test involved a control of a wheel chair, a robotic car and a PowerPoint presentation. The online demonstration reached a classification accuracy of 94.01%, and a confusion matrix is shown in Figure 3.12.



**Figure 3. 12:** Online confusing matrix of the CNN model classification, which still maintain higher classification accuracy.

To study the efficiency of CNN models without the need for feature extraction, our collaborators also compared the CNN model performance with SVM, a popular conventional machine learning model that needs feature extraction. Features like Canonical Correlation Analysis (CCA), Power Spectral Density Analysis (PSDA) and Cross-Spectral Density Analysis (CSDA) were used. CNN model against different architectures of SVM and results are reported in Table 3.6. Mean cross-validation (6-folds, one for each subject) test results for all subjects validated for each subject using 2-channels time-domain inputs.

**Table 3. 6:** CNN in comparison to SVM

<b>Data window length (Seconds)</b>	<b>2-CNN-TD (%)</b>	<b>Linear-SVM (%)</b>	<b>Quad-SVM (%)</b>	<b>Cubic-SVM (%)</b>	<b>Gaussian-SVM (%)</b>
0.512	94.54	21.96	82.67	85.93	81.95
0.768	96.40	22.79	84.29	86.69	83.59
1.024	97.69	22.19	86.54	90.41	87.34
1.536	98.50	22.59	87.92	90.96	88.41
2.048	99.48	22.95	84.86	92.51	93.55

### 3.4 Discussions and Conclusion

We demonstrated the use of CNN for the interpretation of multiple classes of motor imagery activities and steady-state visual evoked potentials (SSVEP) from raw EEG signals without the need for handcrafted features. The first study utilised EEG data from real amputees who are the primary target of BCI technologies, while the second study demonstrated the real-life application of highly portable BCI technologies. The proposed CNN possesses a number of characteristics to address the drawbacks associated with conventional machine learning methods. The ability of the proposed CNN model to learn its own features from raw data provides a significant advantage, especially in real-life applications. Handcrafting features from raw data by human experts can sometimes involve the extraction of features that are not related to the classes involved; hence the model's performance can be adversely affected. Using a small moving window of 0.5 second experimented in the first study, allows the possibilities of performing real-time classification on low-cost smart devices. The size of the networks used with fewer parameters to learn, makes it easy for a pre-trained CNN model to be exported and embedded onto smart mobile devices. The use of small CNN models saves computation costs during training as well as running the trained model and an acceptable response time. Furthermore, as our model maintained high classification performance with fewer channels in both studies, the channel reduction technique with CNN demonstrated that it is possible to use portable EEG devices with fewer channels which can reduce the cost of devices while still achieving reasonably high and consistent accuracy. The second study has demonstrated that a small window of data (0.5 seconds) is practical and the response time is acceptable.

Many of the previous methods for EEG channel reduction are performed on the extracted features [250], this will involve manual extraction of features from the obtained data before performing channel selection. This is not the case in our method as all of the drawbacks and limitations for feature extraction are avoided in both of our studies. In addition, our method of

channel selection can be used to locate the important/significant region on a human scalp for a particular activity, providing a systematic guideline for electrode placement for EEG devices, without the need of an expert to locate them. This becomes the case especially when a fewer channel EEG device is to be built and used, it will require precise placement of the device. Although both studies use EEG for BCI application, the successful demonstration of our proposed methods for channel selection, which has shown consistency in two different aspects, the motor imaginary the visual evoked potential. This provides an opportunity for our methods to be utilised in other domains in finding significant features. Machine learning and deep learning models trained on significant features will not only increase the model's performance and robustness, but also reduce computational costs during both training and model deployment.

The next chapter explore the application of CNN in the authentication of cancer cell lines, an area in which much has not been done in the use of cancer cell lines images for authentication purpose.

# **Chapter 4**

## **Image-based Cancer Cell Lines Authentication Using Deep Neural Networks**

This chapter presents a study of this thesis in applying CNN in biomedical images analysis towards the implementation of automatic analysis of biomedical images. By addressing the limitations highlighted in the literature review of manual biomedical images analysis (Section 2.2.3 in the literature review of this thesis), and the limitations of conventional machine learning towards implementation of computerised analysis of biomedical images (section 2.3.4 in literature review of this thesis), this chapter presents a study of the analysis of microscopic images by using CNN, focusing on human cancer cells. The main function of a microscope device is the magnification of substances which cannot be visualised by naked eyes. In medical laboratories, the microscope machine allows for nanometer-scale investigation of cells and molecules [251], hence becomes useful in cancer research which involves studies of cancer cells.

Authentication of cancer cells is an important step in cancer cells biology studies and discovery of anti-cancer drugs. Towards implementation of computerised way for authentication of cancer cell line in the laboratory environment, this study introduces the application of CNN for automatic authentication of cancer cell lines. To implement the study, this chapter is organised as follows. Section 4.1 provide introduction to the study, while section 4.2 describes the study. Through a pilot classification task reported in section 4.2.4, different experiments were conducted to find an optimal training strategy and the optimal model to be used in my analysis. After the optimal model and the optimal training strategy were found, further hyperparameter optimisation was conducted on the optimal model as reported in section 4.2.5. Using the fine-tuned optimal model, the authentication of the cancer cell lines at different stages was conducted and results reported in section 4.3.

## 4.1 Introduction

Cancer is a major cause of death in developed countries and increasingly also in developing countries. Based on the GLOBOCAN 2018 estimates of cancer incidence and mortality by the International Agency for Research on Cancer, an estimated 18.1 million cancer cases (17.0 million excluding nonmelanoma skin cancer) were newly diagnosed in 2018, and 9.6 million individuals (9.5 million excluding nonmelanoma skin cancer) died from cancer in the same year. The cancer incidence increases with age. Hence, the number of cancer cases is anticipated to rise further as a consequence of a rising lifespan. Thus, research is needed to improve cancer therapies.

Cancer cell lines are cancer cells that have been isolated from human or animal cancers and that can be grown continuously as cell cultures in a laboratory. They are important and probably the most commonly used model system for both studying cancer biology and the discovery of novel anti-cancer drugs [175] [252]. However, many cell lines are misidentified, i.e. they have been swapped or contaminated with other cell lines, and as a result, researchers may work with cells that are not what they think they are. This has been a problem since the work with cancer cell lines started and continues to be a problem today [127], [253], [254]. The most applicable test for authentication of cancer cell lines is the Short Tandem Repeat (STR). STRs are short sequences of DNA, normally of length 2-5 base pairs, that are repeated numerous times in a head-tail manner. The STR test primers pick up these repeated segments of DNA to find matches between the cell lines for authentication purpose [255]. Although the STR analysis has been available as a reliable method to identify the genetic origin of a cancer cell line for a while [253], 15-20% of the currently used cell lines have been estimated to be misidentified [256], due to some factors like alteration of cell lines DNA fingerprinting profiles during in vitro manipulations and long-term culture [257]. For example, in a study based on the analysis of 482 different human tumor cell lines, up to 96 cell lines were misidentified [258]. Moreover,



the same study [258] found that STR profiling alone is insufficient to exclude inter-species cross-contamination of human cell lines, and the study argued for the need of additional testing and authentication methods. Hence, additional authentication methods that can be easily applied in the laboratory are highly desirable.

Additionally, methods are needed that reliably discriminate between isogenic cell lines, such as clonal sublines and drug-adapted cancer cell lines, since this is not achievable using STR analysis. An isogenic cell line is a cell line that has been engineered from a parental line through the introduction of a targeted gene mutation [259]. Drug-adapted cancer cell lines are used as models of acquired drug resistance and have been used to identify many relevant drug resistance mechanisms [260]–[264] . Automated image recognition approaches may be an additional way to authenticate cells that may complement STR analysis and enable the differentiation between isogenic cell lines [265]. In this chapter, we are presenting a joint study with the school of biosciences at the University of Kent, which have established the Resistant Cancer Cell Line (RCCL) collection, the worldwide largest collection of drug-adapted cancer cell lines, currently consisting of >1,500 models of acquired resistance [266]. The aim of the study is to develop a computerised way for authentication of cancer cell to complement the current existing method, the STR, which relies on genetic content of a cell line to perform the authentication. Our study also aimed at developing other authentication parameters which can not be achieved by the STR method, such as authentication of isogenic cell lines. The role of scientists from the University of Kent's school of biosciences was to carry out data collection, the contribution of this thesis was to perform an analysis of the data and present the results. To achieve the aim of this aims, this thesis is presenting a method for authentication of cancer cell lines based on cancer cells microscopic image recognition by using CNN.

By applying CNN, this study aims to address the following challenges:

4. How CNN can be used towards the development of fast, easy to use, relatively cheap and computerised ways for authentication of the cancer cell line in the laboratory

environment using microscopic images.

5. What additional authentications of cancer cell lines which can be provided by the application of CNN in the authentication process, to complement the current existing STR authentication method.
6. What techniques can be applied to take advantage of CNN while addressing the unavailability of enough training data, a common problem in medical images.

## **4.2 Study description**

In this study, we investigate the use of two deep learning models, MobileNet and InceptionResNetV2 (details in Section 4.2.1), in the authentication of cancer cell lines through a classification process by using the microscopic images of the cancer cell lines. Both models are based on Convolutional Neural Networks (CNN), which have the capabilities of finding their own related features from the training data without the need for feature engineering.

### **4.2.1 MobileNet and InceptionResnet**

This section describes the models used in our study, Inception ResNet V2 and the MobileNet. Both models are based on Convolutional Neural Network architectures. Inception ResNet V2 has 164 layers with 55 million trainable parameters [267], it consumes more training time and computational resources compared to MobileNet which has only 28 layers and 4 million parameters [268]. As the aim is to demonstrate that it is possible to use deep learning algorithms in the authentication of cancer cell lines in the laboratory settings, two algorithms were chosen by considering the trade-off between model's accuracy and training and deployment expenses.

InceptionResNet combines the two major concepts of Inception and ResNet models as used by C. Szegedy et al. [269]. Inception incorporates each convolution in parallel and concatenates them at the end as proposed by C. Szegedy et al. [230]. The key benefit of the architecture is that, it increases the number of units at each stage without increasing the

computational complexity much [72]. Each inception unit has several non-linear convolution modules with various resolutions which makes it more applicable to tasks that need to process data with multiple resolutions like medical images [270]. Inception which shows good performance at comparatively low computation cost [271]. As Inception networks tend to be very deep, and residual connections adopted from ResNet perform good in training deep networks, the filter concatenation stage of the Inception networks can be replaced by residual connections[27] and give rise to InceptionResNet which takes advantage of both Inception and Residual architectures.

Another model, MobileNet, deploys Depthwise Separable Convolution which applies a single filter to each input channel instead of standard convolution [237]. In MobileNet, a standard convolution is factorised into Depthwise convolution and a  $1 \times 1$  convolution called a pointwise convolution. Depthwise convolution applies a single filter to each input channel. The pointwise convolution then applies a  $1 \times 1$  convolution to combine the outputs of the Depthwise convolution. The MobileNet reduces the computational costs by avoiding standard convolution which filters and combines inputs into a new set of outputs in one step [237].

## **4.2.2 Data sets**

Two datasets were used, the cancer cell lines dataset (see section 4.2.2.1 for details), which is the objective of our authentication task, and the breast cancer cells dataset (see section 4.2.2.2), which is made of publicly available breast cancer cell images [272]. For transfer learning (a concept discussed in section 2.6.5 of Chapter 2 of this thesis), the two models used, the MobileNet and the InceptionResNet V2 were pre-trained on non-medical images, the ImageNet [73]. After determining and fine-tuning the hyper- parameters of the optimum model for our problem, a multi-stage transfer learning approach was therefore conducted on the optimum model configurations by using the breast cancer cells dataset as explained in section 4.2.1.2. During the multi-stage transfer learning, the breast cancer cell images were used as an

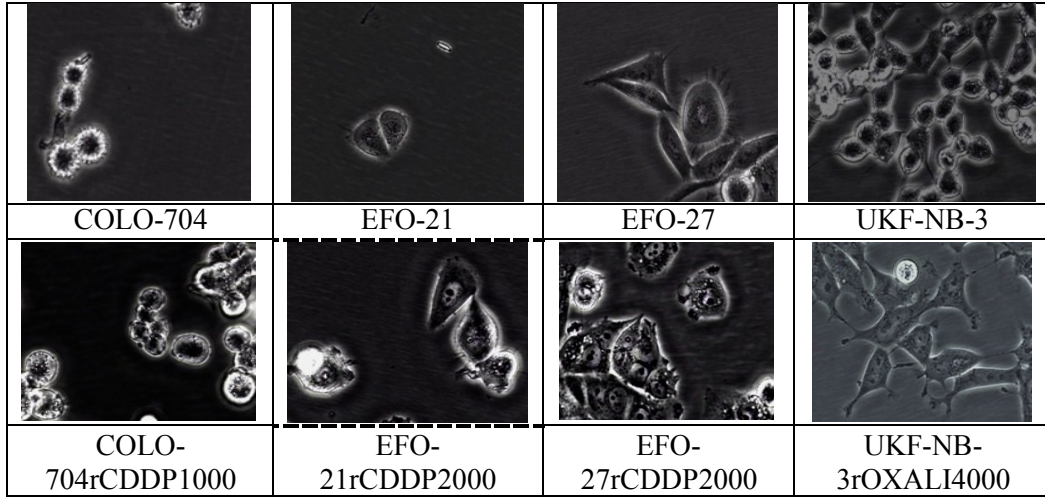
intermediate transfer learning stage before training the model on our target task of cancer cell lines dataset. The two datasets are detailed below,

### 4.2.2.1 Cancer Cell Lines dataset

The dataset consisted of microscopy images of parental cancer cell lines and their sublines, which had been adapted to grow in the presence of anti-cancer drugs. The set of cancer cell lines consisted of three ovarian cancer cell lines (EFO-21, EFO-27, COLO-704) and their sublines adapted to the anti-cancer drug cisplatin, as well as the neuroblastoma cell line UKF-NB-3 and its subline adapted to the anti-cancer drug oxaliplatin. The ovarian cancer cell lines were obtained from DSMZ (Braunschweig, Germany). The neuroblastoma cell line was established from a bone metastasis of a stage IV neuroblastoma patient [273]. All drug-resistant sublines had been established by continuous exposure to stepwise increasing drug concentrations as previously described [273][266] and were derived from the Resistant Cancer Cell Line (RCCL) collection [110]. The cisplatin-resistant ovarian cancer sublines had been adapted to  $1\mu\text{g}/\text{mL}$  (COLO-704<sup>rCDDP1000</sup>) or  $2\mu\text{g}/\text{ml}$  (EFO-27<sup>rCDDP2000</sup> and EFO-21<sup>rCDDP2000</sup>) cisplatin. The oxaliplatin-resistant UKF-NB-3 subline (UKF-NB-3<sup>rOXALI4000</sup>) was adapted to  $4\mu\text{g}/\text{mL}$  oxaliplatin [273]. Table 4.1 shows the number of images per cell line. Image samples for each class are shown in Figure 4.1. The first column shows the parental cell lines and the second column of cell lines shows its corresponding drug treated.

**Table 4. 1:** Number of images per cell line.

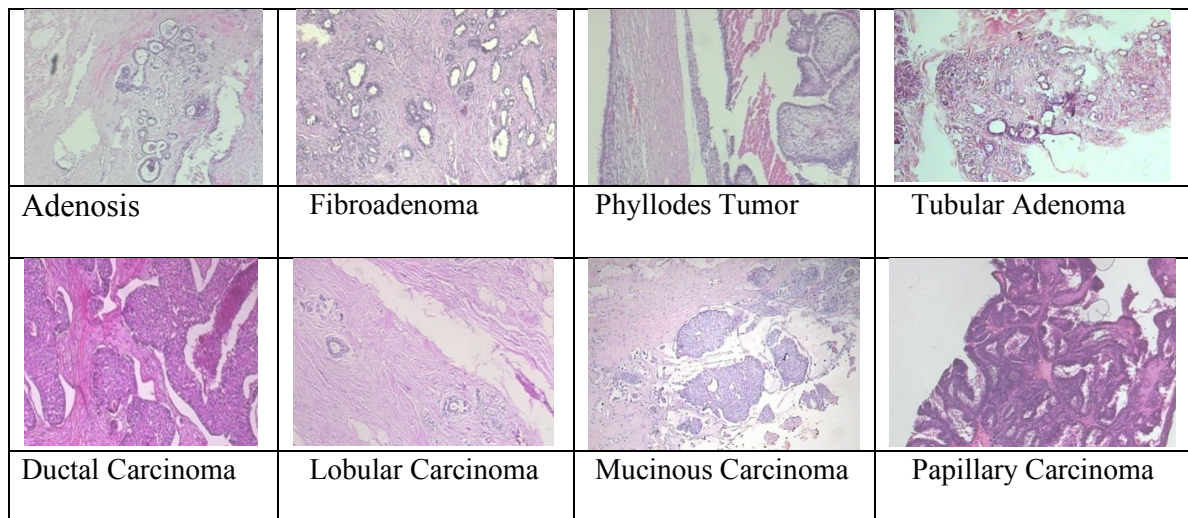
Parental Cell line	Number of images (n)	Drug treated cell line	Number of images (n)
COLO-704	220	COLO-704 <sup>rCDDP1000</sup>	270
EFO-21	220	EFO-21 <sup>rCDDP2000</sup>	220
EFO-27	220	EFO-27 <sup>rCDDP2000</sup>	220
UKF-NB-3	201	UKF-NB-3 <sup>rOXALI4000</sup>	170



**Figure 4. 1:** Cancer Cell Lines sample images. All the images are in RGB, Joint Photographic Experts Group (JPEG) format with a resolution of 2560 x 1922 pixels.

#### 4.2.2.2 Breast Cancer Cells dataset

As most of the available pre-trained models for transfer learning are trained on non-medical images, we proposed a multi-stage transfer learning through an inter-mediate training step using breast cancer cell images, to make sure our selected model had the knowledge of medical images before fine-tuning it for our cancer cell lines cell line authentication task. A multi-stage transfer learning is a technique of transfer learning in which a model can undergo several transfer learning before fine-tuned to its target task. Mostly, the intermediate transfer learning stage involves training a model with a data set from the domain similar to the domain of the target task. For this multi-stage transfer learning, we used publicly available breast cancer cells data set. The data had been produced by the Laboratory of Vision, Robotics, and Imaging (VRI) at the Federal University of Parana, and collected from 82 patients and comprised of 7909 microscopy images of breast tumour tissue [272]. It is divided into two categories with 5429 malignant samples and 2480 benign samples, both with 700x460 pixels resolution and 3 channel RGB. Samples of the dataset are shown in figure 4.2.



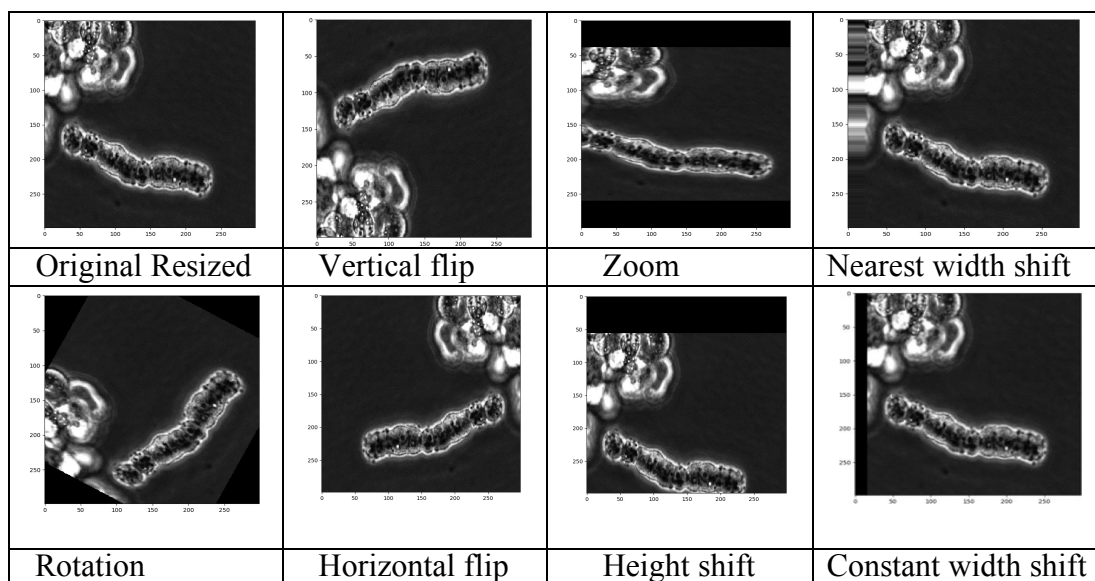
**Figure 4. 2:** Breast cancer sample images as extracted from the publicly available breast cancer dataset, contributed by [272], Laboratory of Vision, Robotics, and Imaging (VRI) at the Federal University of Parana.

### 4.2.3 Data pre-processing

Prior to model training, pre-processing steps like image splitting, cross-validation, conversion to grayscale and different data augmentation methods were performed [234][274]. To ensure each parameter (pixels in our case) has similar data distribution, a method to remove variance in brightness and to enhance the contrast of the images was used. This technique helps with exploding or disappearing gradients. This was done by subtracting each image from its mean pixel-intensity value to make convergence faster while training the network. In the case of RGB images, the mean was calculated channel-wise. All the processing conducted in this paper was done only on the cancer cell line dataset which is our target task.

Our models were trained with both grayscale and RGB formats for comparison to find the optimal method during the pilot authentication task. the conversion to grayscale, data normalisation, and augmentation was done online during the training process. For augmentation, five most popular image augmentation parameters were selected, which are images shifts, flips, rotations, brightness and zoom [275]. Shift and zoom range of 0.2 was selected. A small value was selected for these parameters to make sure each augmented image contains the cancer cells, as large parameters can miss or affect the visualisation of the cancer cells. Other parameters included vertical and horizontal flip, a rotation of 360, and

brightness in the range 0.5,1.0 to 1.5. For brightness, the values less than 1.0 will add darkness to the image and greater than 1.0 will add light to the image. For image splitting, the dataset was separated into two main categories, parental cancer cell lines and their drug-adapted sublines. The four parental cancer cell lines, their four sublines of drug-adapted, and the combination of the two which form a group of eight cell lines were treated as three separate authentication problems. 10-fold cross-validation was performed on all three tasks. Each fold had 10% of the data set selected randomly and without repetition. As we had a relatively small data set, data augmentation approaches were used to artificially increase the sample size as explained in section 2.6.5 of the literature review. As different data augmentation techniques have been proved to contribute different performance on the model [234], two data augmentation approaches were employed and compared, which are the nearest width shift and Constant width shift. Examples of augmented images are shown in Figure 4.3.



**Figure 4. 3:** Image augmentation samples with different twisting parameters with nearest and constant width shifts pixels assignments.

During data collection, the collected dataset may have images taken in a limited set of conditions but we might fall short in a variety of conditions that are not accounted for, like the orientation and zooming effects. Performing operations like flipping the image horizontally, vertically, padding, cropping, rotating and zooming on the training data during

data augmentation will not only increase the amount of the training data, but also the diversity of the data which were not originally present, hence a model gets to generalise well.

The cell line data we used contain a set of four parental cancer cell lines and their sublines adapted to anti-cancer drugs. This dataset enables us to consider the ability to distinguish between cancer cell lines of different genetic origins and also those of the same genetic origin. We further perform separate classifications of the four parental cell lines and their respective drug-adapted sublines in a separate classification task as part of the authentication process. In contrast to *Meng et.al* [276], whose method depends on expensive machinery that is only available in very few specialised laboratories, our approach is based on phase-contrast images from well plates or cell culture flasks that can be obtained with every standard inverted cell culture microscope and will, therefore, be available to every standard cell culture lab. Our authentication method for drug-adapted sublines based on Convolutional Neural Network (CNN) will complement STR and help to ensure that researchers know what cell lines they are working with.

#### **4.2.4 Pilot classification task**

This section introduces a pilot authentication which was conducted to determine the optimal model, pre-processing, and training strategy for the cell lines authentication task. In this process, three authentication tasks were performed by using the two selected models with different method combinations. The three tasks (see inset A of Figure 4.4), were:

- a) The authentication of parental cancer cell lines,
- b) The authentication of drug-treated cancer cell lines and,
- c) the authentication of the combined dataset of parental and drug-treated cancer cell lines.

We adopted a  $k$ -fold cross-validation ( $k=10$  in our case) because when comparing the model's performance using simple train/test split, different results may be obtained each time when different data samples are selected in the train/test sets [277]. During the pilot classification, for each class in all three tasks, eight folds were used for training, one for



validation, and the remaining one fold for testing. For fair comparison across all three authentication tasks, the folds selected for training, validation, and testing were kept constant during the pilot classification task. All results were analysed and compared and the optimal model and the optimal training strategy were selected based on the model’s performance. Optimal model is the highest performing model in this thesis’s case. For validation and testing, only the original (un-augmented) images were used [278]. The data preparation procedures, training, and model selection procedures are illustrated in insets A and B of Figure 4.4, respectively. Two off-the-shelf Deep Neural Network models, InceptionResNet V2 and MobileNet, which have proven to work well in recognising pictures of general objects. Both models are CNN architectures. The two models were trained with different strategies to find the optimal strategy and the optimal model. Figure 4.4 below shows the whole process of pilot classification.

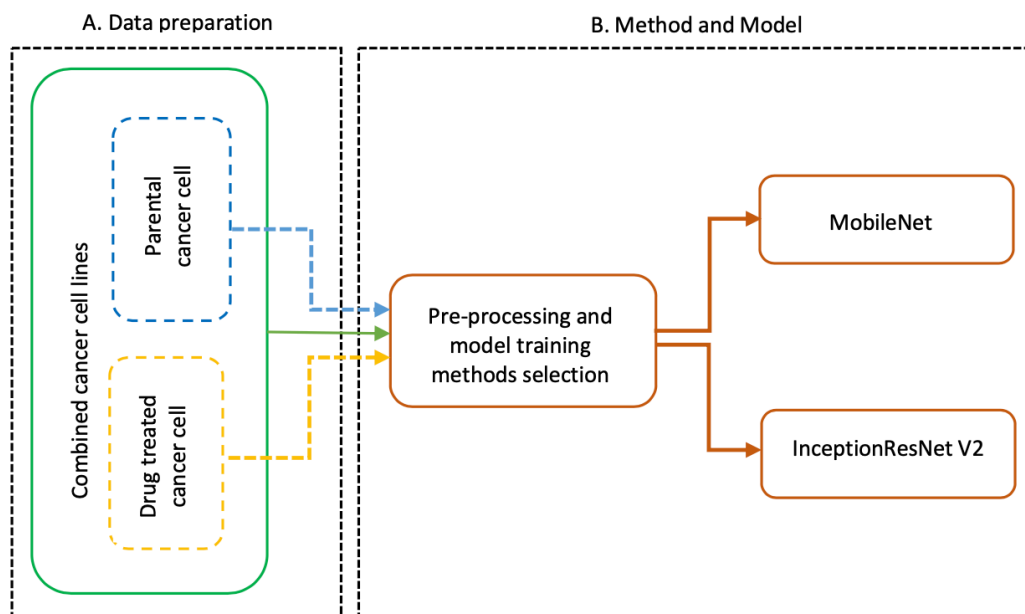


Figure 4. 4: Pilot authentication task which was conducted to determine the optimal model with the optimal training strategy.

#### 4.2.4.1 Training strategies

Different strategies for data pre-processing and model training were tested to determine the optimum performing strategy. Each possible combination of the training strategies was

systematically fed to each of the two models separately and the model performance was used to determine the best performing strategy. Table 4.2 shows the different combinations of training strategies.

**Table 4. 2:** Training strategies selection process to determine the optimal method for our authentication task

Grayscale	RGB	Without Data Augmentation	With Data Augmentation	Without Transfer Learning	With Transfer Learning	Combination number
✓		✓		✓		I
✓			Nearest	✓		II
✓			Constant	✓		III IV
	✓	✓			✓	V
	✓		Nearest		✓	VI
	✓		Constant		✓	VII
	✓	✓		✓		VIII
	✓		Nearest	✓		IX
	✓		Constant	✓		

No tests were done on transfer learning for grayscale image, because the pre-trained models used for transfer learning are trained on the ImageNet dataset which has only RGB images.

#### 4.2.4.2 Performance measure metric

The F1-score was chosen as a measure of performance as it considers both recall and precision. Recall is a true positive rate of a model and precision is the positive predictive value of the model. In machine learning, the true positive rate, also referred to sensitivity or recall, is a percentage of positive samples correctly classified as positive, false negative rate is a percentage of positive samples incorrectly classified as negative, false positive rate is a percentage of negative samples incorrectly classified as positive and true negative rate is a percentage of negative samples correctly classified as negative [279]. Recall and precision can be calculated as in equations 1 and 2 respectively, and F1 score as in equation 3. TP stands for true positive, FN for false- negative and FP for false positive.

$$Recall = \frac{TP}{TP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$F1 \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}} \quad (3)$$

### 4.2.4.3 Pilot authentication results

This section presents pilot authentication results of the experiments conducted during a pilot authentication task as explained in sections 4.2.3.1 and 4.2.3.2, and summarised in figure 4.4 above. The experiments were conducted through a classification process in three authentication tasks as mentioned above.

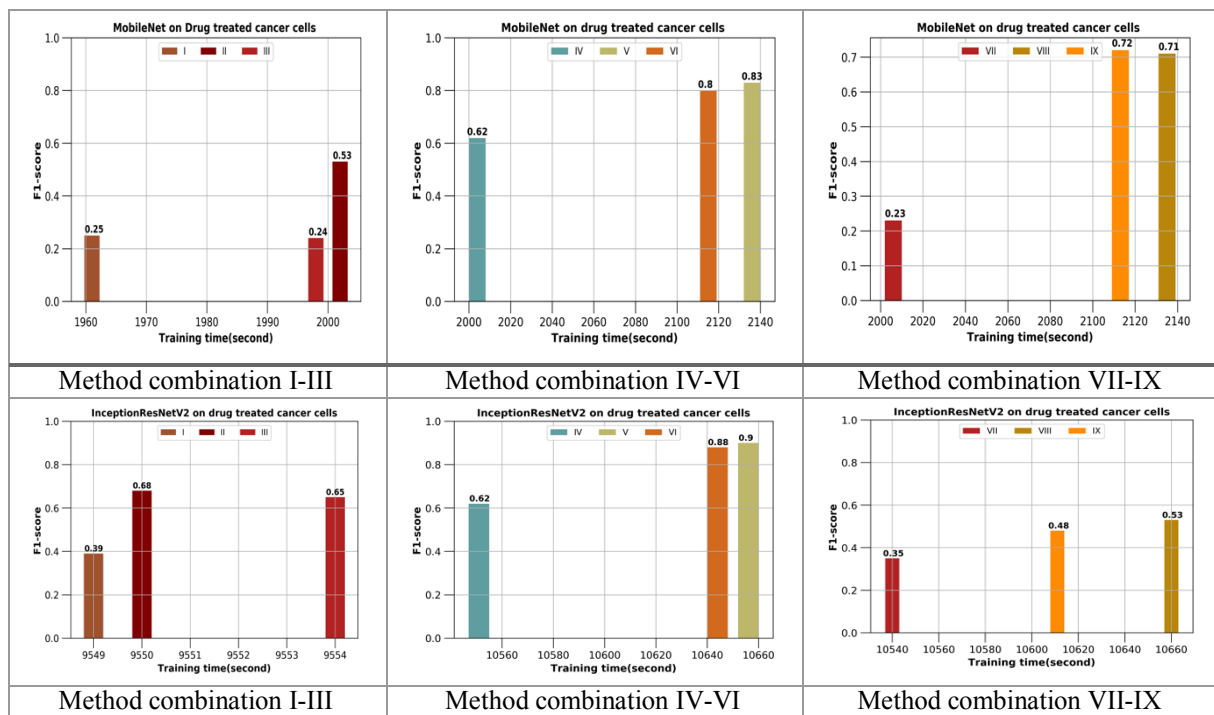
It is more computationally expensive to train a model with RGB than with grayscale images, as it took 50 ms more to train a model with RGB images. This is because RGB images carry more information which can be essential for a model to learn. Hence the two ways, training with grayscale and with RGB images were tested and results compared. Both models were trained with normalised images. For data augmentation, several hyper-parameters were selected. The hyper-parameters selected for augmentation reflected that the cells can be anywhere in an image independently of how the cell culture vessel is placed under the microscope. The selected method and the optimal training strategy were determined during the pilot classification task. For training of the model from scratch we used random initialisation of the weights with Stochastic Gradient Descent (SGD). Random initialisation of weights when training the model neural networks are trained using a stochastic optimization algorithm called stochastic gradient descent. The algorithm uses randomness in order to find a good enough set of weights for the specific mapping function from inputs to outputs in your data that is being learned. It means that your specific network on your specific training data will fit a different network with a different model skill each time the training algorithm is run.

Figures 4.5 - 4.7 below, shows that both models, InceptionResNet V2 and MobileNet, under different training strategy perform reasonably well in our pilot authentication task for cancer

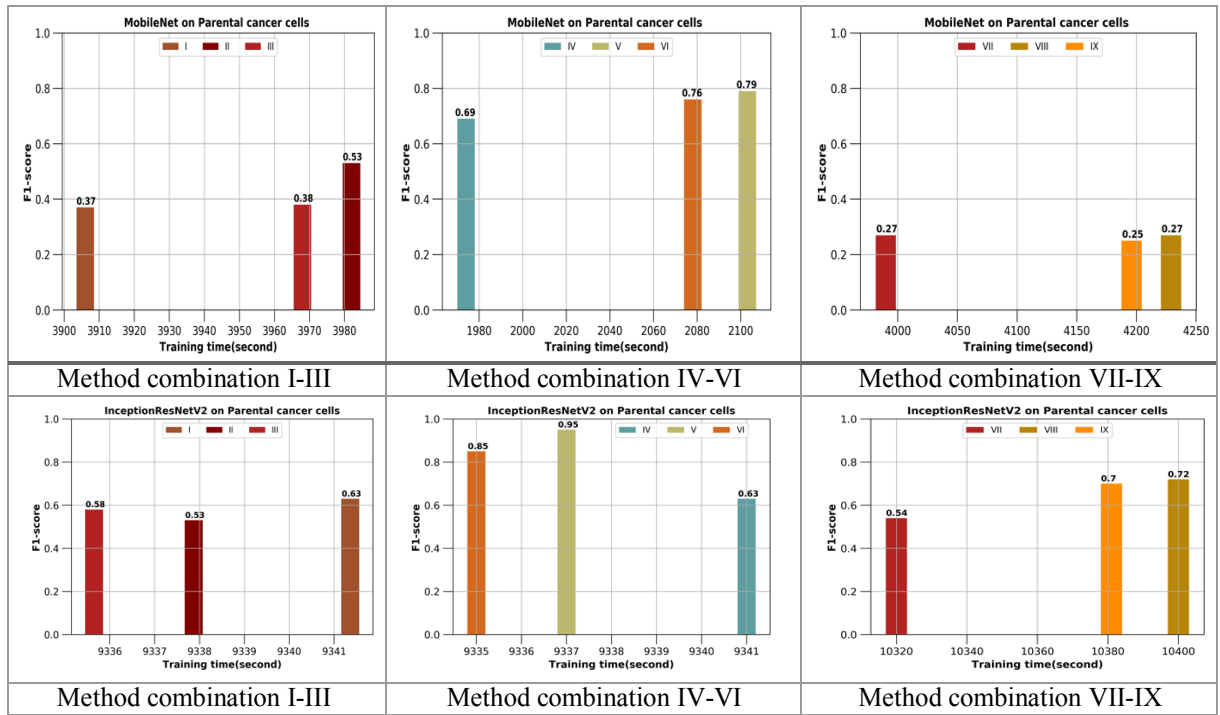
cell line authentication, when trained on RGB images in all three authentication tasks. This is because RGB images contain a large amount of morphological information and thus play a significant role in differentiating cancer cell lines [280]. As the RGB images contain more information, this makes them computational expensive to train compared to grayscale images. Both data augmentation techniques, Nearest width shift and Constant width shift had a positive impact by significantly increasing the F1-score in both models, with the nearest width shift performing better than the constant width shift. With reference to training strategies selections in Table 4.2 above, figures 4.5 - 4.7 show that training strategy I, IV and VII, which did not include augmentation, performed worse than those which included data augmentation, i.e. training strategy II, III, V, VI, VIII and IX. Both models required more computational time for optimisation when applying augmentation since more data were generated by augmenting the training data. Transfer learning (using ImageNet) had a significant positive effect on model performance, as it always performed better (training strategy IV-V) compared to when the models were trained from scratch (training strategies I, II, III, VII, VIII, IX). With transfer learning, both deep neural network models performed better when transfer learning was combined with data augmentation (V, VI) than without data augmentation, method combination IV. A small difference in training time was observed when training the models from scratch or with transfer learning. An early stopping criterion was used. A training time was calculated based on how long the model was trained until its performance flattened. This is because in both cases, the fully connected layers are the ones that get trained and are the ones containing more parameters to be trained, while the deeper layers are responsible for feature extraction. MobileNet optimises faster than InceptionResNet V2 due to its size, (Figure 4.5, 4.6 and 4.7). By observing Figures 4.5 - 4.7, it is clear that in both models, training with RGB and data augmentation (nearest width fill), which is described as training strategy V, performed better than any other combination of approaches.

The performance of the models (F1-scores), with the optimal strategy (method combination

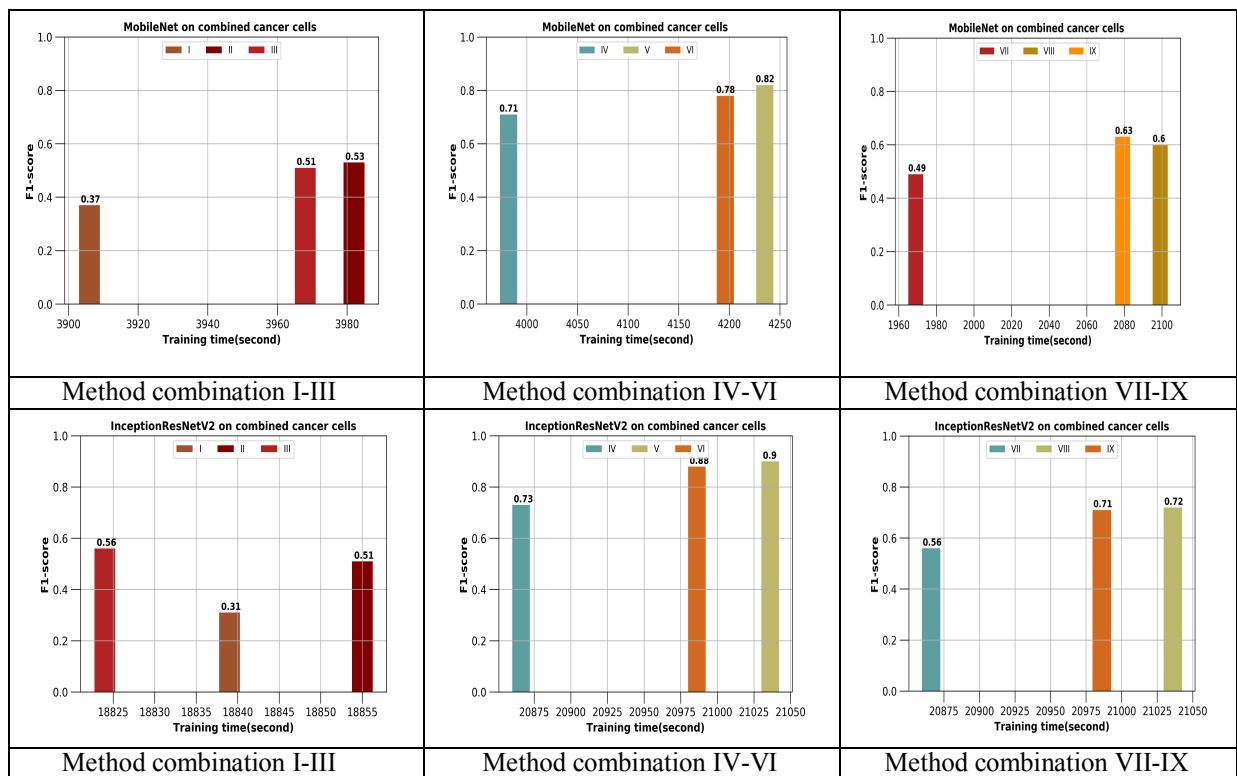
V) observed in our pilot classification on the authentication of an eight-class authentication task, is shown in Table 4.3 below. In training a deep learning model that often takes a long time to train, it is essential to observe and validate the model during training to avoid overfitting. Figure 4.8 below plots the validation curves of our two models during training. From Figure 4.8, it can be seen that both models generalised well on the training set, with InceptionResNet V2 performing better than the MobileNet. InceptionResNet has over 50 million trainable parameters hence become computationally expensive to train and to use the trained model for different applications compared to MobileNet which has only around 4 million trainable parameters which can even be deployed in mobile applications.



**Figure 4. 5:** Drug treated cancer cell lines classification model performance under different methods combinations. Top row for MobileNet and lower row for InceptionResNet V2, each block shows performance of the three tested combinations.



**Figure 4. 6:** Parental cancer cells classification model performance under different combination. Top row for MobileNet, lower row for InceptionResNet V2, each block shows performance on three tested combination.

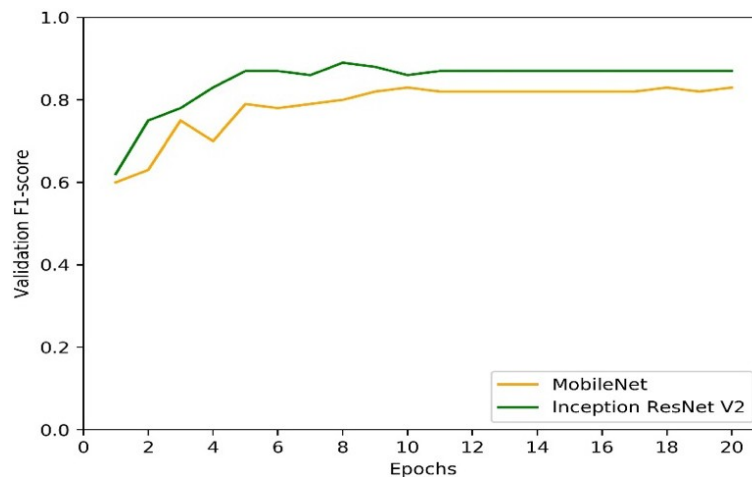


**Figure 4. 7:** Combined cancer cell lines classification model performance. Top row for MobileNet and lower row for InceptionResNet V2, each block shows performance of the three tested combinations

The F1-scores on the test set in Table 4.3 shows that the trained models were able to perform well on the authentication of the unseen data. Table 4.3 and Figure 4.8 show that InceptionResNet V2 performed better on the pilot classification task on the best method combinations determined from Table 4.3. The results are shown in Table 4.3 and Figure 4.8 are based on the eight classes authentication task during the pilot classification task.

**Table 4. 3:** Model comparison based on one-fold, on an eight-class problem

Model	Test F1-Score
InceptionResNet V2	0.88
MobileNet	0.82



**Figure 4. 8:** Learning curves for model comparisons

Based on results from the three authentication tasks, InceptionResNet V2 was selected as the optimum performing model compared to MobileNet. Also, a combination of RGB images, data augmentation, and transfer learning (training strategy V) was observed to be the optimal configuration. Further fine-tuning of hyper-parameters was conducted on the optimal model under optimal training strategy as explained in the next section.

## 4.2.5 Hyper-parameters tuning of the optimal Model

After selecting the optimal model with its optimal training strategies for our task through a pilot classification task, as explained in section 4.2.3 above, further tuning of hyper-parameters was performed to find the optimal hyper-parameters which can increase the performance of our optimal model. For hyper-parameters tuning of our selected optimal model, an eight-class authentication task was used as it is a more complex authentication task and therefore ideal for the fine-tuning task. The fine-tuning was done by using a single fold, as in the pilot task, by keeping constant the training, validation and the test folds throughout the process. The following fine-tuning tests were conducted on our selected model;

### 4.2.5.1 Resizing and cropping

The original images in the dataset have a resolution of 2560x1922 pixels (RGB). In the experiments performed in the pilot classification, the images were resized by squashing the image to 299x299 pixels as a predefined input shape for InceptionResNetV2 and 224x224 pixels for MobileNet. As the predefined input shape of our best performing model was 299x299 pixels, several tests to transform the input image into the required input shape from our sample images were performed and compared. Govindaiah *et.al* [236] resized the images in their data set to a reference image before centre cropping when training their model for screening and assessment of age-related macular degeneration from fundus images. Several resizing and cropping dimensions were tested to find an optimal approach to our problem.

To maintain the aspect ratio, the following methods were applied and then compared with image squashing and reported.

- I. centre crop the image without resizing.
- II. Resize the image to 1280x 961 pixels which is half of the resolution of the original image and then centre crop the image

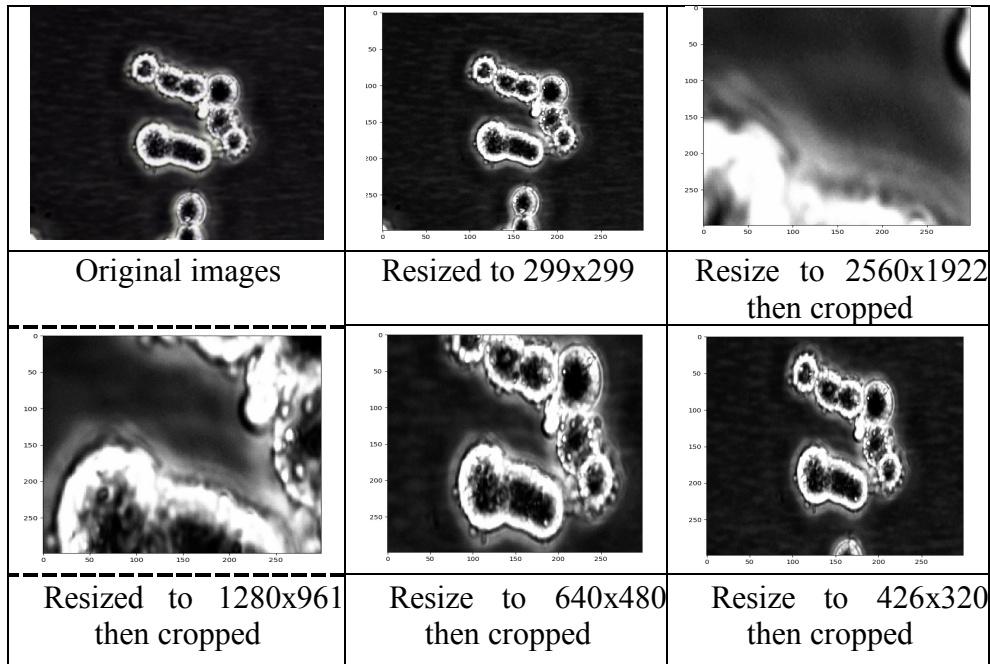


- III. Resize the image to 640x480 pixels which is a quarter of the resolution of the original image and then centre crop the image
- IV. Resize the image to 426x320 pixels which is 1/6 of the resolution of the original image and then centre crop the image.

The model's performance of the model at different resizing is as shown in Table 4.4 below. The aspect ratio of the original image is 01.33:1 whereas the resized images have a 1:1 aspect ratio. Hence, resizing the image by squashing, distorts the aspect ratio of the image. It was observed that resizing only without cropping performed better, which means that the distortion of the aspect ratio did not have an impact on cancer cell authentication. Resized and cropped samples and results of the model performance are shown in Figure 4.9 and Table 4.4 respectively in the supplementary information. By visualising Figure 4.9 it can be seen that resizing and cropping loses some details of the original image while the resized image looks like the original image. Hence, resizing only was taken as the best option.

**Table 4. 4:** Resizing and cropping

<b>Model</b>	<b>Test F1-score</b>
Resize to 2560x1922 then crop	0.44
Resize to 1280x961 then crop	0.62
Resize to 640x480 then crop	0.78
Resize to 426x320 then crop	0.84
Resize to 299x299	0.87



**Figure 4. 9:** Samples of resizing in a different dimension and cropping to find the optimal technique.

#### 4.2.5.2 Fully connected layers

In deep learning models, layers closer to the input learn general features of the input data while deeper layers closer to the output learn more specific features for the target problem [97]. Hence, determining the number of layers to be added on a pre-trained model for transfer learning becomes an important aspect. In fact, the added layers play a crucial role in the learning of specific features for the target problem by the model. When using transfer learning, it is essential to control the depth at which the amount of ‘knowledge’ transfer between the source and target problem is optimal for the target problem. In deep neural networks, most of the parameters are in the fully connected layers [281], as the deeper layers are responsible with feature extractions while the fully connected ones trained with the extracted features. Furthermore, different popular models which have won ImageNet competition has used different number of fully connected layers, AlexNet has 3 fully connected layers, GoogLeNet and ResNet which both contributed to the development of our best performing model as they use Inception and ResNet techniques, respectively, both have one fully connected layer [269][71]. We experimented to find the optimal number of fully connected layers to be

added and the number of neurons in those layers for our problem.

Different numbers of fully connected layers and the number of neurons on the added layers were tested as shown in Table 4.5 below. An architecture with two fully connected layers had low training accuracy and F1-score, but it has a similar F1-score value as a one-fully-connected-layer architecture. Hence, we chose an architecture with one fully connected layer as it is less expensive computationally.

**Table 4. 5:** Model comparison based on one-fold, on an eight-class problem

<b>Fully Connected Layers</b>	<b>Test F1-score</b>	<b>Training accuracy</b>	<b>Training F1-score</b>
1x 256 Neurons	0.83	0.9406	0.96
1x512 Neurons	0.86	0.9432	0.97
1x1024 Neurons	0.87	0.9463	0.97
2x1024 Neurons	0.87	0.9296	0.96

### 4.2.5.3 Batch size

The batch size, which is the number of samples processed before the model is updated during training, has an influence on the model performance in terms of both accuracy and training time. It is reported that the greater the batch size the higher the accuracy of the model [282]. The idea is contradicted in [277] where it is stated that models trained with large batch sizes have a high chance to generalize more poorly than those trained with small batch sizes. Studies [283] and [284] suggested a batch size of no more than 64 samples. In addition, a large batch size has a high computational cost when training a deep learning model; previous studies, [285] and [95], recommended large batch size when training with large datasets, but this choice comes with optimization difficulties. Since there is no clear recommendation of the batch size to use in general, different batch sizes were tested to find the optimum batch size value for our problem.

For batch size selection, the results for the model learning curves are shown in Figure 4.10, and the final results of the model performance at different batch sizes are shown in Table 4.6

below. Based on these results, one can conclude that batch size 16 and Batch size 32 had the same and highest F1 score, in comparison with the other batch sizes. However, the training accuracy and F1-score on the training set was not the same. Batch size 16 had the higher training accuracy and F1-score compared to the batch size 32. Hence, based upon the result, the batch size of 16 was chosen.

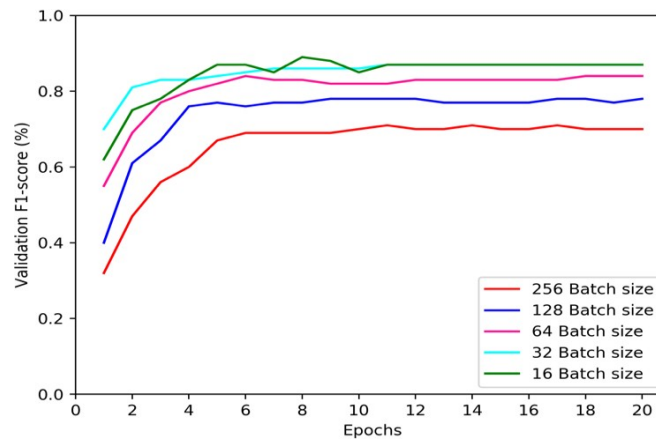


Figure 4. 10: Batch size learning curves

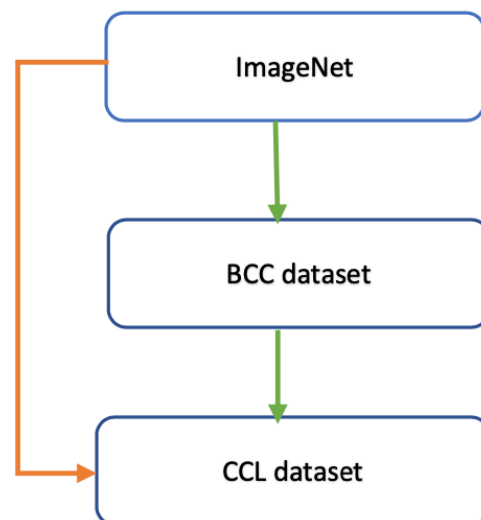
Table 4. 6: Batch size f1-score

Batch Size	Test F1-score	Training accuracy	Training F1-score
16	0.87	0.9463	0.97
32	0.87	0.9355	0.96
64	0.84	0.9185	0.94
128	0.77	0.8937	0.91
256	0.68	0.7885	0.80

#### 4.2.5.4 Multi-stage Transfer learning

A common practice in deep learning is that training and testing data should come from the same domain. In some scenarios, such as training a deep learning model with medical images, that practice becomes more challenging as it is hard to obtain sufficient medical images. Transfer learning could provide a solution by training the models with easily obtained data from a different problem and fine-tune the model with the smaller dataset for the target

problem. Some investigations into medical images have adopted multi-stage transfer learning by fine-tuning their models through other publicly available medical data sets, before fine-tuning the model to their target problem. Ravi *et.al* [97] observed an increment of 4 percent in F1-score when performing multi-stage transfer learning by using publicly available breast cancer data set to fine-tune a deep learning model as an intermediate transfer learning stage, before fine-tuning the model for their target task of classification of breast cancer. As there are common high-level features in cancer cells such as shape and size [286], we used the breast cancer cells data set as an intermediate transfer learning stage, since there is no publicly available data set to our specific problem (cell lines authentication), to the best of our knowledge. Figure 4.11 shows a single-stage transfer learning in the orange line and a multi-stage transfer learning in green. To implement multi stage transfer learning, we load the weights of the ImageNet into model with few new added fully connected layers, and train it with the BCC dataset and save the weight. The saved weights were loaded into the same model, with few new added fully connected layers and trained in the final stage.



**Figure 4. 11:** Single-stage and multi-stage training strategies

A modest increment of 1 percent was observed on the F1-score when a multi-stage transfer learning was conducted compared to the single-stage transfer learning. This may mean that, in

multi-stage transfer learning, our optimal model was able to learn extra cancer-related features from the breast cancer cells data set. This increment of F1-score produced by multi-stage transfer learning might have been small because we used images of breast cancer in the intermediate stage which is little different from our target task, Unlike a study by Ravi *et.al* [97] in which the intermediate stage dataset was of the same domain as the target task. Our target task, in fact, includes images from ovarian and neuroblastoma cancer cells, although they are all digital images of cancer cells. The small increment of F1-score in our case means there might be some little common features among cancer cells from different types of cancer. Hence, a multi stage trained model was used to perform all the authentication tasks reported in the next section.

### **4.3 Authentication using the selected model**

After the fine-tuning of the optimal model, we performed the authentication of the cancer cell line by looking into different authentication angles. Unlike in the pilot classification task and the fine-tuning of the optimal model where the training, validation and test folds were fixed, at this stage, nine folds were used for training and validation and one for testing for all the authentications conducted by using the optimal fine-tuned model using the optimal training strategy. The process was repeated 10 times for each fold to be used for training, validation, and testing, the mean results with their standard deviation of all 10 folds are reported in this thesis. The predictions of any trained deep learning model are subjected to a degree of uncertainty; a trained model can perform well in identifying some classes and perform poorly on other classes. It is necessary to understand how a trained model performs in identifying individual classes trained on instead of just an overall performance, as some classes are hard to authenticate compared to others. We studied by investigating the confidence of our optimal model in authentication of each cancer cell line separately and the findings reported in section 4.3.2. Besides, as it is difficult to get balanced datasets of biomedical images especially when

dealing with rare cases, we investigated the effect of sample size on cancer cell line authentication, reported in section 4.3.3. All the experiments conducted using our fine-tuned optimal model are as explained below;

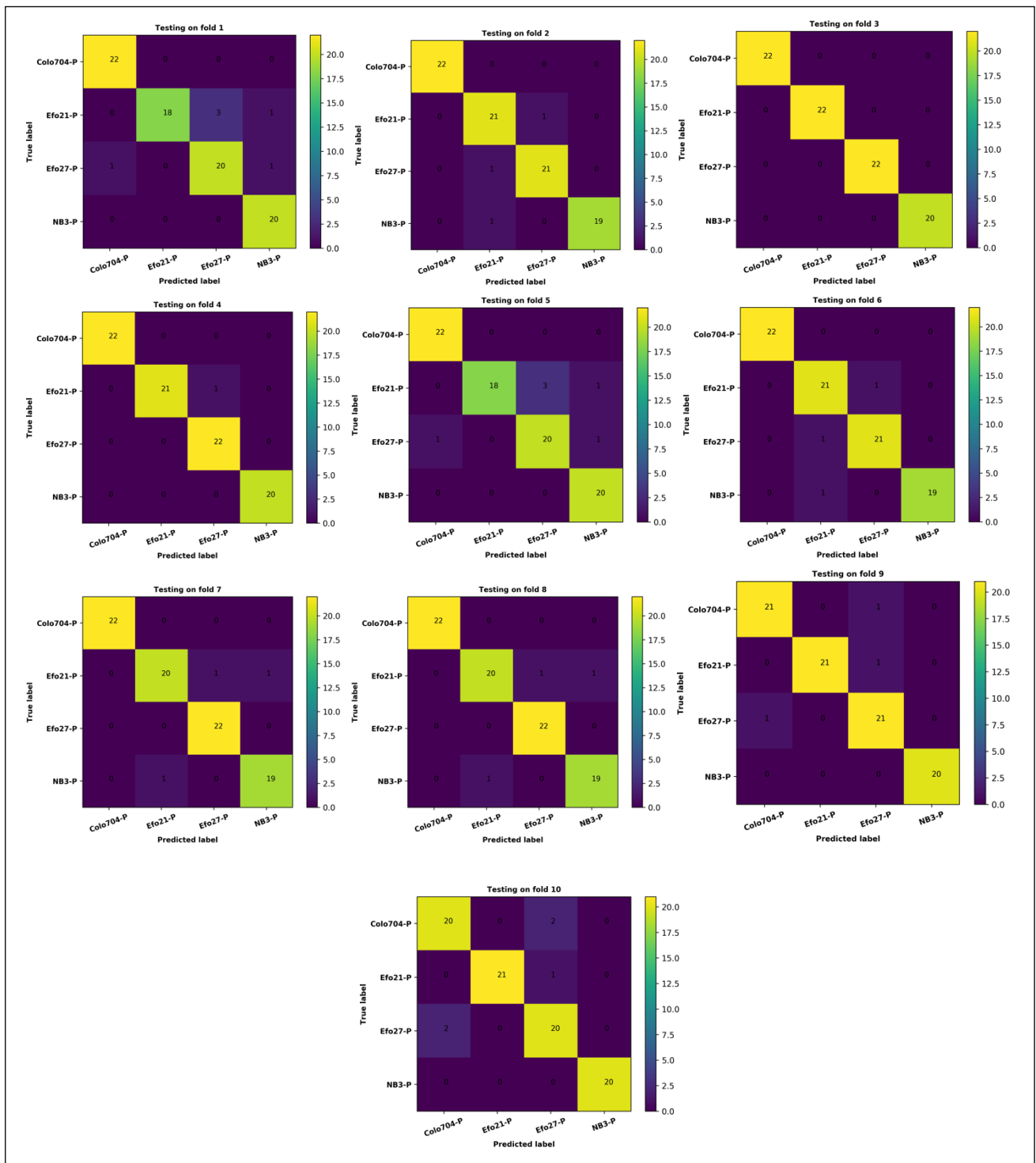
### **4.3.1 Authentication stages**

To complement the current authentication methods used in biology research (i.e. STR), which is not able to differentiate between parental cancer cell lines and their drug-adapted sublines, eight-class authentication of the mixed four (4) parental and the four (4) drug-treated cancer cell lines was conducted. To apply STR profiling for authentication, standardized protocol and a data-driven, quality-controlled and publicly searchable database will be required which is a complex and time-consuming process. To complement STR with a cheaper and quicker computerised method of authenticating cancer cell lines, we further performed authentication of the four (4) parental cancer cell lines and the authentication of the four (4) drug-resistant sublines separately using our optimal fine-tuned model. Furthermore, we perform a two-class task to authenticate the parental cell lines and their drug-adapted sublines in one approach.

#### **4.3.1.1 Four classes authentication**

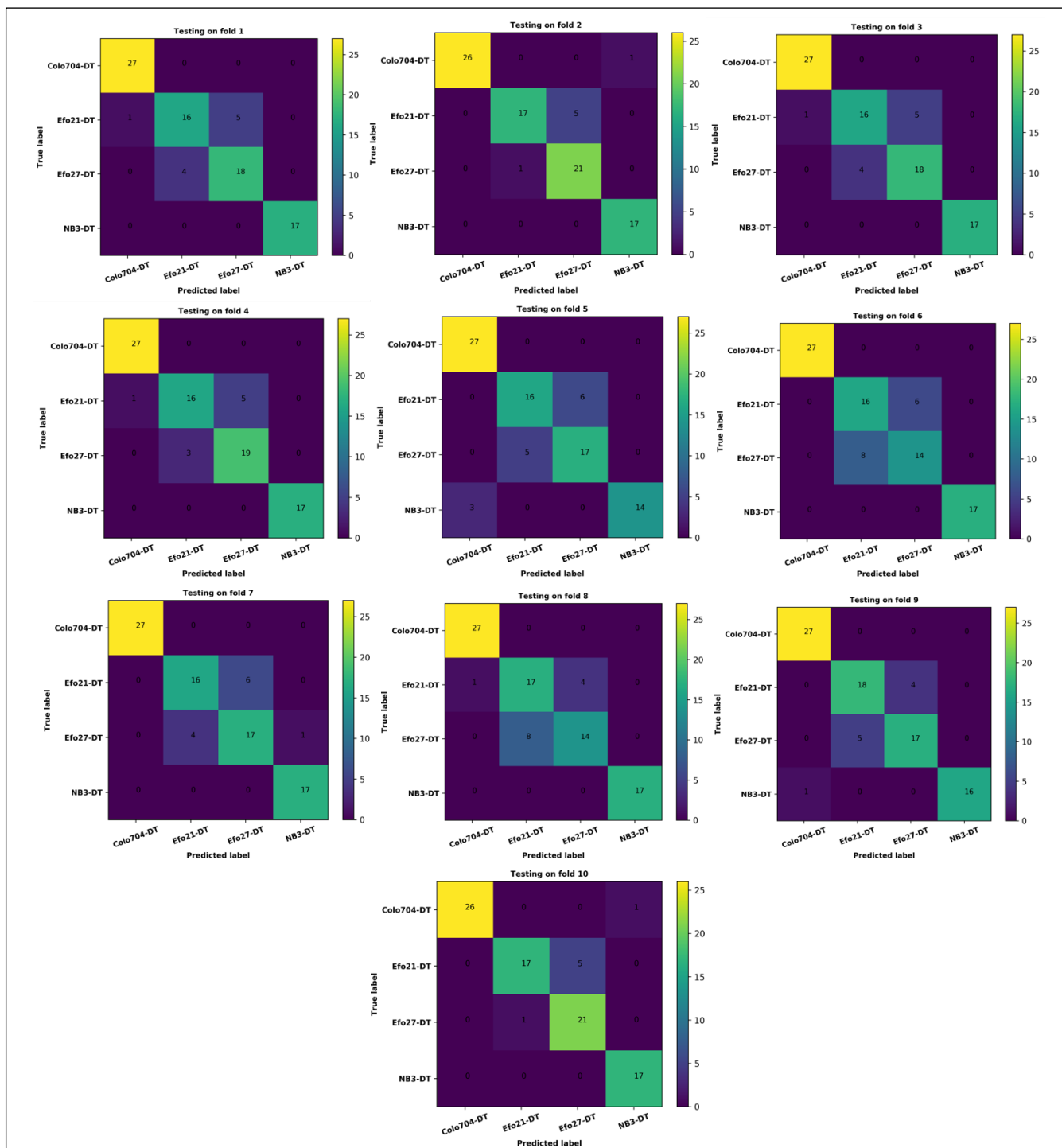
Using the fine-tuned selected optimal model and training strategies, our model performed better on the authentication of parental cancer cell lines than on their drug-adapted sublines (Table 4.7) as two separate authentication tasks. A total of 101 images in the drug-adapted cell lines were misclassified compared to just 30 images in the parental cancer cell lines. The misclassified drug-adapted were classified as other drug-adapted cell lines within the four available drug- adapted cancer cell lines while the misclassified parental cancer cell lines were classified as other parental cancer cell lines among the four available parental cancer lines as can be seen in Figure 4.12 and Figure 4.13 of the confusion matrices of the 10 folds, for parental and drug- adapted cancer cell lines, respectively. This might indicate that drug-

adapted cancer cell lines develop a common resistance mechanism which makes them look more similar and more difficult for the model to differentiate them as they develop some common features as a result of drug treatment and development of resistance mechanism compare to parental cell lines.



**Figure 4.12:** Confusion matrices of parental cancer cell lines. For clear visualisation of the Figure, parental cancer cell lines have been abbreviated as P.





**Figure 4. 13:** Confusion matrices of drug-treated cancer cell lines. For clear visualisation of the Figure, Drug treated cancer cell lines have been abbreviated as DT.

**Table 4. 7:** Four class authentication

Cell type	Mean F1 score	Standard deviation
Parental cancer cell lines	0.96	0.02
Drug treated cancer cell lines	0.91	0.03

### 4.3.1.2 Eight classes authentication

On the combined data set classification problem (eight-class task), which is a more complex authentication task as it combines both four classes of parental and drug-treated cancer cell lines, an average of 0.91 F1-score is obtained with a standard deviation of 0.03, as shown in Table 4.8 below. A mean Area Under the Curve (AUC) of 0.95 was obtained with a standard deviation of 0.01 across the 10-fold. Figure 4.14 and 4.15 shows the ROC curves and the respective AUC for each class and the confusion matrices for the combined cancer cell lines, respectively. In figure 4.13 of the confusion matrices for the combined dataset, it can be noticed that the confusion occurred mostly between the parental class and its respective subline in most of the folds. This is due to the same genetic origin of the cell lines. The differences are only caused by the drug adaptation process.

**Table 4. 8:** Eight classes authentication

Cell type	Mean F1-Score	Standard deviation
Combine cancer cell lines	0.91	0.03

### 4.3.1.3 Per class performance

When training a machine learning model, the model can learn to classify some classes well and fail on other classes of the same problem. This might be due to the quality of the data in a particular class, imbalances between the datasets, or a lack of diversity among the training samples in a particular class that prevents the model from capturing the patterns. It is important, therefore, to see how the trained models perform in classifying each class separately. The model performance in authenticating each class was studied and reported in Table 4.9. These results were extracted from the classification report of the eight-class authentication task. The

average F1-score of each class across the ten folds and its standard deviation was calculated and reported in Table 4.9. The model had low confidence in authenticating the cancer cell lines EFO-21rCDDP<sup>2000</sup> (mean F1-score 0.76) and EFO-27rCDDP<sup>2000</sup> (mean F1-score 0.71), with the highest standard deviations out of all other cell lines. Further analysis of this case is presented later in the paper to find out the reason for this case.

**Table 4. 9:** Model performance per class

Class	Mean F1-score	Standard deviation
COLO-704rCDDP1000	0.95	0.03
COLO-704	0.96	0.03
EFO-21rCDDP2000	0.76	0.12
EFO-21	0.95	0.04
EFO-27rCDDP2000	0.71	0.16
EFO-27	0.95	0.06
UKF-NB-3rOXALI4000	0.98	0.04
UKF-NB-3	0.99	0.03

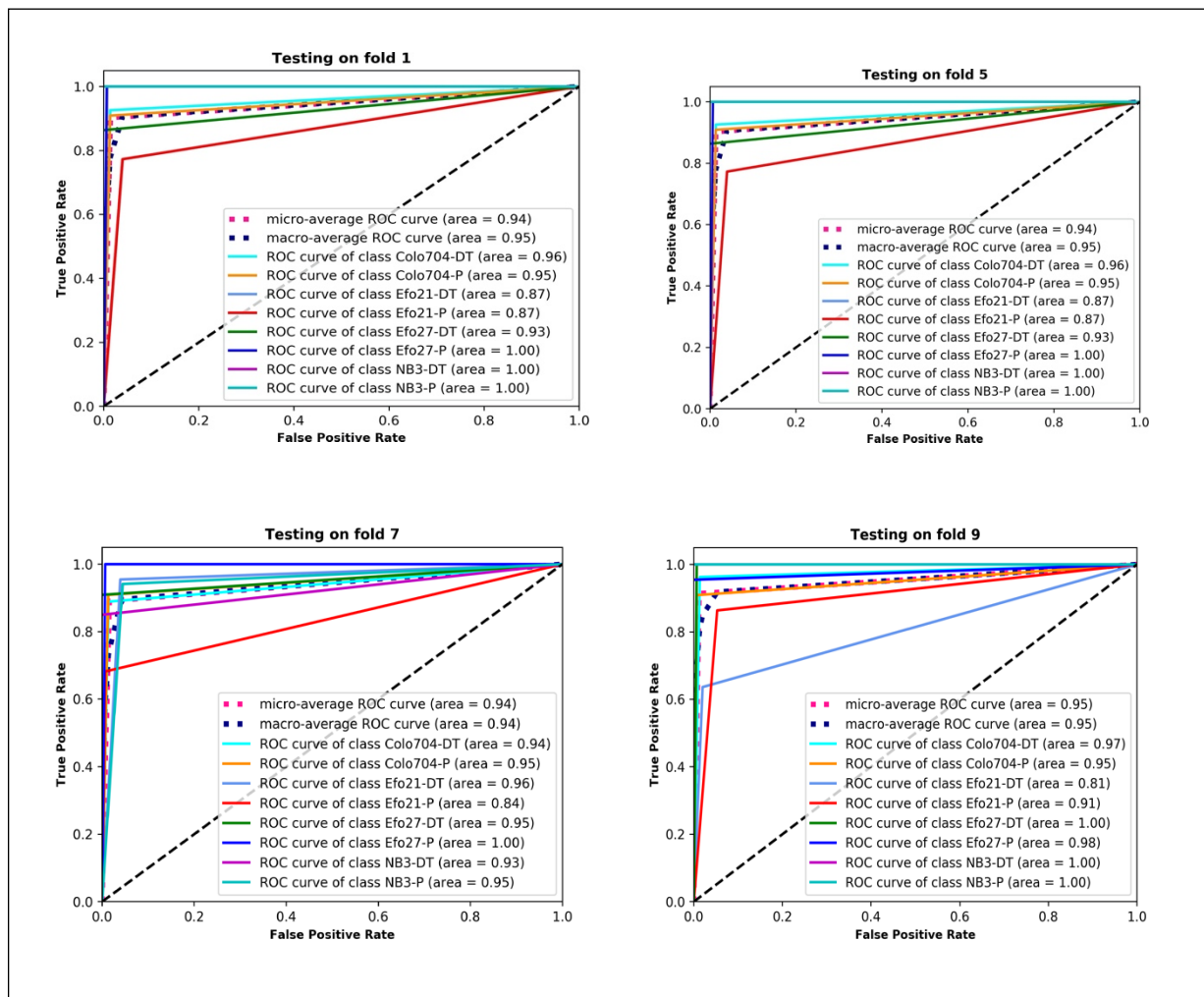
#### 4.3.1.4 Two-classes authentication task

Two class authentications were also performed to see how the model performs in classifying between the parental cell lines and their drug-adapted sublines (Table 4.10). With reference to table 4.1 which shows parental and its subline of drug treated cell line, Table 4.10 shows the classification between the two, for each of the four cell lines. Although there are common features between the parental cell lines and the respective drug-adapted sublines as they come from the same genetic origin, our model managed to authenticate them with an average F1-score of around 0.95.

**Table 4. 10:** Two classes authentication

Cell line	Average F1 score	Standard deviation
COLO-704/ COLO-704rCDDP1000	0.90	0.01
EFO-21/ EFO-21rCDDP2000	0.94	0.02
EFO-27/ EFO-27rCDDP2000	0.98	0.03
UKF-NB-3/ UKF-NB-3rOXALI4000	0.98	0.03
Average	0.95	0.02

The promising results of our method means that our method may be used for authentication of cancer cell lines in conjunction with STR or for tasks where STR cannot be utilised.



**Figure 4. 14:** ROC curves of combined cancer cell lines. For clear visualisation of the Figure, Parental cancer cell lines have been abbreviated as P and DT for Drug treated cancer cell lines, with few folds results been presented.

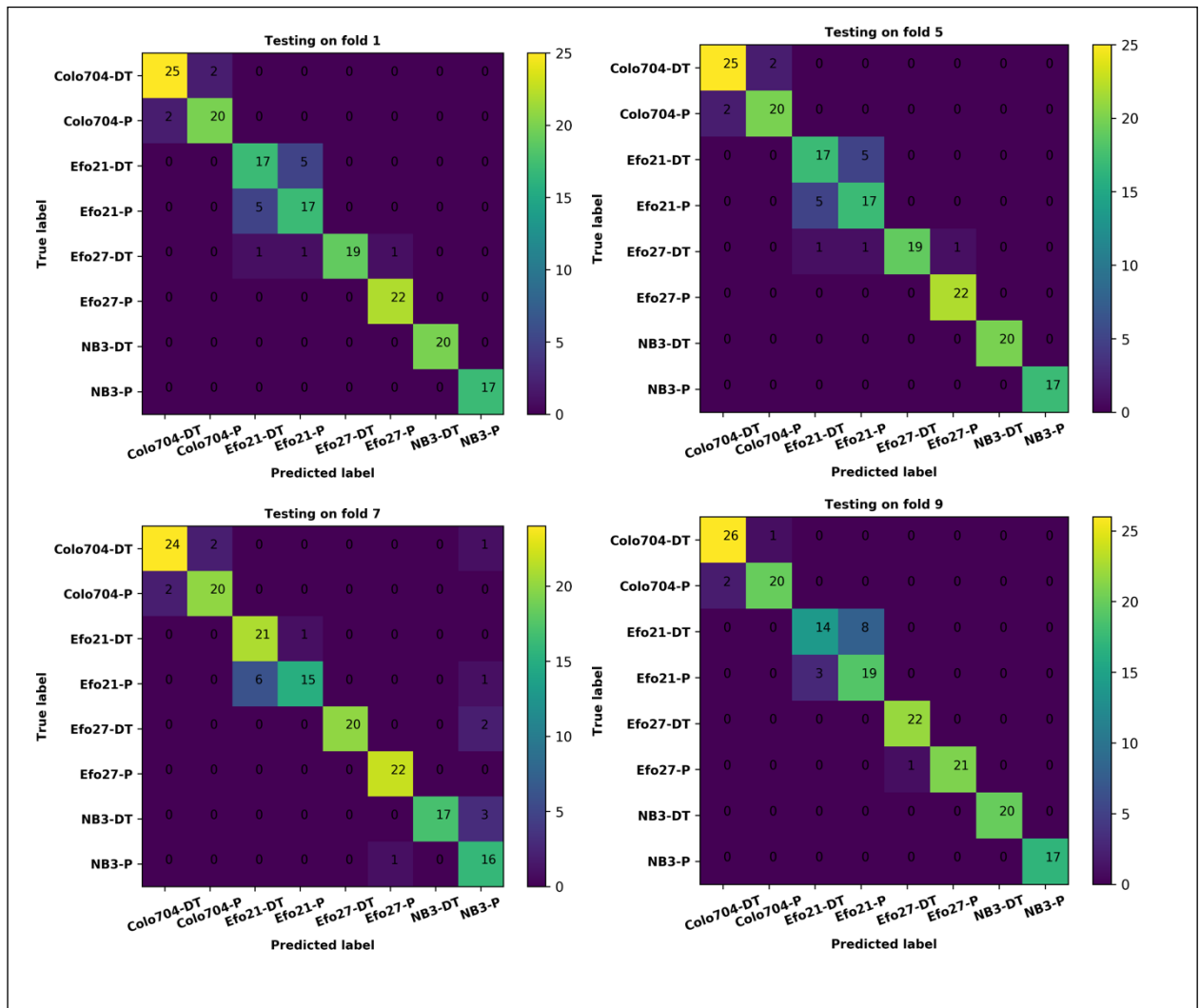


Figure 4.15: Confusion matrices of combined cancer cell lines, with few folds results presented.

### 4.3.2 Effects of sample size

Deep learning can easily overfit when trained on a small sample size, hence it is important to study the effect of sample size when training deep learning models to avoid overfitting [235][265][287]. To study the effect of sample size, we studied our selected optimal model performance at different training sample sizes, on an eight-class classification model based on 10 folds. The number of images in the test fold was kept constant while the number of images in the training folds was reduced in steps of 20%. This experiment was done just to study and give guidance on what data samples amount might be reasonable enough to get acceptable results.

Table 4.11 shows the performance of our optimal model when trained with different training sample sizes. It is crucial to know the optimum number of training samples sufficient for a deep learning model to generalise, especially with biomedical images as it is usually difficult to get large data samples for training. For this investigation, we kept the test sample sizes constant for fair comparison while reducing the training sample size stepwise by 20%, from 100% to 20%. Table 4.11 shows that a drop from 100% to 80% in the number of training images from the original number of training image samples (i.e. from 1566 to 1253 training sample images) had a small impact on the F1-score, with a drop of 3%. Further reduction of the training data set has a significant negative effect on the model, with a significant reduction of the F1-score and an increase in the standard deviation. This suggests that, together with transfer learning and data augmentation techniques, a larger training data sample size will be required for better results.

**Table 4. 11:** 10 folds cross-validation with the training sample size drop.

Percentage (%) of training sample size	Mean F1-score	Standard deviation
100	0.91	0.03
80	0.88	0.04
60	0.80	0.12
40	0.72	0.13
20	0.57	0.16

### 4.3.3 Model’s Classification confidence

Despite the high accuracy attained by a trained CNN, it can have difficulties in predicting some of the classes [288]; this rises the importance of studying the confidence of a trained model in prediction. Some studies advise not to use prediction probabilities as the confidence of a deep learning model [289]. Researchers Y.Gal *et.al* [290] and Z. Ghahramani *et.al* [289] instead, used the probabilities of classes generated by the trained CNN models as the confidence score to reject noises and keep the predictions with high-class probabilities in a face detection task. Also, in an object detection network by Redmon *et.al* [94], probabilities of the bounding

boxes containing objects were used as the confidence score of the model. We, therefore, studied the prediction probabilities of our model in predicting a small set of randomly selected samples from our test set by extracting their probabilities of the assigned class. The predicted probabilities were extracted from the softmax layer of the trained model. Both probabilities incorrect and wrong predictions were studied and reported in this thesis.

The probabilistic confidence of the model when making correct predictions was compared to when incorrect predictions were made (Table 4.12). This was done by randomly picking one image sample from the training sample set of each class. Overall, our model was very confident in predicting the correct class by attaining an average probability confidence of 0.92 for true positive predictions compared to 0.64 for false-positive predictions. This may mean that, the high confidence of the model on correct classification and low confidence in false prediction, this model can be used in real laboratory authentication of cancer cell lines. The lowest confidence of 0.76 and 0.78 were observed when correctly identifying EFO-27<sup>r</sup>CDDP<sup>2000</sup> and EFO- 21<sup>r</sup>CDDP<sup>2000</sup>, respectively. This prompted a further investigation on the case. The model had the highest confidence in classifying COLO-704 and UKF-NB-3 cells with confidence of 0.99 and 1.0 probabilities respectively. The model had a low average confidence of 0.64 when predicting the wrong class, increase of the training samples might increase the confidence of the model presented.

**Table 4. 12:** Model confidence in classifying different classes

Cell line	Correct confidence	Wrong confidence
COLO-704 <sup>r</sup> CDDP1000	0.93	0.76
COLO-704	0.99	0.74
EFO-21 <sup>r</sup> CDDP2000	0.78	0.70
EFO-21	0.96	0.73
EFO-27 <sup>r</sup> CDDP2000	0.76	0.75
EFO-27	0.96	0.79
UKF-NB-3 <sup>r</sup> OXALI4000	0.97	0.68
UKF-NB-3	1.0	0
average	0.92	0.64

#### 4.3.4 Further investigation on EFO-21 and EFO-27

Due to the low performance in authentication of the drug-adapted sublines of EFO-21 and EFO-27 in the per-class performance of the optimal model (Table 4.10) and also in the model confidence investigation (Table 4.12), a further investigation was conducted on these cell lines. This was done by testing the F1-score values by which the model can discriminate between the parental EFO-21 and EFO-27 cell lines and between their drug-adapted sublines (Table 4.13).

**Table 4. 13:** Investigation on Efo-21 and Efo-27

Cell line	Mean F1-score	Standard deviation
Efo-21/Efo-27 Parental	0.94	0.01
Efo-21/Efo-27 Drug treated	0.60	0.05

The results show that our optimal model performed well in the authentication of the parental cancer cell lines, but was less reliable in the authentication of the cisplatin-resistant sublines (Table 4.13). These results are also supported by the performance observed in the confusion matrices of the parental cancer cell lines of EFO-21 and EFO-27 (Figure 4.12), and of their drug-adapted sublines (Figure 4.13). This may suggest that, apart from the drug-adapted cancer cell lines may be developing the same resistance mechanism which becomes harder for a deep learning model to authenticate as seen in Table 4.13 above, the parental cancer cell lines treated with the same anti-cancer drug develops more similar resistance mechanism compared to those treated with different anti-cancer drug.

## 4.4 Summary

The common issue of misidentification of cancer cell lines requires new ways for performing cell line authentication in a laboratory environment. Current authentication methods are expensive, time-consuming, and cannot differentiate between cell lines of the same genetic origin, like parental cancer cell lines and their drug-adapted sublines. Resistance formation is associated with morphological changes that make drug-adapted cell lines distinguishable from the parental cell line. Hence, approaches using computer-aided digital image analysis can be



used to develop effective authentication approaches that can be easily included into the daily laboratory routine and which may complement, or in some cases even replace other authentication methods. Such methods may also enable the discrimination between cell lines of the same genetic origin, a task for which established methods are currently lacking. Such approaches have the potential to improve the reliability of research results due to the reduction of the use of misidentified cell lines. By using deep learning, we have demonstrated that, it is possible to authenticate cancer cell lines, including parental cancer cell lines and their drug-adapted sublines, based on image recognition. Our data also suggest that resistance formation to a certain drug may be associated with specific morphological changes. An improved understanding of such processes may enable the further development of image-based strategies to gain mechanistic insights. Furthermore, our results suggest that, apart from the drug-adapted cancer cell lines developing the same resistance mechanism, cancer cell lines treated with the same drug develop much similar resistance mechanisms compared to those treated with different drugs. Hence, our results are not only promising with regard to the development of novel cell line authentication approaches but also provide initial evidence that image-based methodologies can be developed as tools for the performance of functional and mechanistic studies. Next chapter provides discussion and conclusion of the studies reported in this thesis.

# Chapter 5

## Discussion and Conclusion

Sections 1.1 and 1.2 of Chapter one has introduced the application of conventional machine learning in the development of computerised ways for the analysis of time series biomedical data and biomedical images, respectively. The practice is further discussed in detail in Section 2.3 of the literature Chapter with a focus on the EEG for BCI applications and microscopic images, with its limitations highlighted in Section 2.3.4. Apart from the limitations of conventional machine learning in the analysis, the size of the current existing EEG devices due to the number of channels has affected the development and deployment of real-life BCI applications as discussed in Section 2.4 of the literature Chapter. To address these limitations, this thesis explored the application of deep learning, specifically Convolutional Neural Network (CNN) in the analysis of biomedical data with a focus on EEG signals for BCI application and microscopic images for authentication of cancer cell lines. For the case of EEG based BCI applications towards the design and fabrication of user-friendly EEG devices, this thesis has also explored the application of CNN in the EEG channels selection and in determining precise locations for electrode placement. The present chapter provides a discussion and conclusion of the application of CNN in the healthcare domain for the analysis of biomedical data, with reference to the studies reported in this thesis.

The chapter is broken down as follows: Section 5.1 reports on the discussion of the studies of this thesis, while section 5.2 concludes the findings, looks into the limitations of our suggested method and proposes future research directed towards the development of fully computerised biomedical data analysis based on CNN.

### 5.1 Discussion

The main focus of this thesis was to investigate the use of computerised methods for the

automatic analysis of biomedical data using CNN, with a focus on EEG signals for BCI applications and microscopic images of cancer cells for cancer cell lines authentication. This was done by addressing some of the limitations of conventional machine learning in developing computerised methods for automatic analysis of biomedical data by using CNN. Building around the research questions below, this section discusses the findings of the studies reported in this thesis. By looking into the two common forms of biomedical data, the time series and biomedical images as two separate case studies.

The results of the three studies reported in Chapter 3 and Chapter 4 of this thesis focused on answering the research questions highlighted in section 1.5 of Chapter one. The findings of the studies in answering research questions in case study one is discussed in Section 5.1.1.1 and that of case study 2 are discussed in Section 5.1.1.2. Although CNN can be potential towards implementations of computerised methods for the analysis of biomedical data, its demand for large training datasets has limited its application, especially in biomedical images as highlighted in Section 2.6.4 of literature. Section 5.1.2 discusses the methods applied in this thesis to address this limitation towards the implementations of our suggested methods. For a while, CNN has been termed as a “Black Box” as it’s hard to explain how it makes its prediction, Section 5.1.3 discusses the necessity of visualisation of a trained CNN model for both validation and explanation of the results. The recommendations and implications of this thesis are discussed in Section 5.1.4.

### **5.1.1 CNN in computerised biomedical data analysis**

Studies reported in this thesis has demonstrated the application of CNN in two common forms of biomedical data, time series and biomedical images. By looking into the potential and challenges of applying deep learning in computerised biomedical data analysis, in particular CNN, this section brings together the two case studies and discusses their results as reported in the previous chapters.

### **5.1.1.1 The strength of CNN in automatic feature extraction**

This section addresses the first research questions in both case studies, i.e. application of CNN in biomedical data analysis without the need for manual feature engineering. Unlike conventional machine learning, Deep Learning has a promising potential because of its ability to learn robust feature representation from raw data.

With conventional machine learning, researchers have been applying different manual feature extraction methods with varying number of features selected to train the models. The common approach to manual feature engineering is to build features one at a time using domain knowledge. This is a tedious, time-consuming, involves mathematical computations and error-prone process [291]. This complex process has been a major limitation in applying conventional machine learning in the development of computerised biomedical data analysis, detailed in Section 2.3.4 of the literature Chapter. Furthermore, there is no generic feature extraction scheme which works in all cases, one has to perform several tests to determine a suitable feature extraction method to use for a specific problem [4]. Traditionally, when applying conventional machine learning, one needs to extract many features and perform the selection of the features to identify significant features to use. As it is in the feature extraction process, there is no generic method for feature selection, several methods like filter, wrappers and embedded techniques as explained in Section 2.4 of the literature Chapter has been widely used [292]. The feature selection process is also known to be time-consuming, computationally expensive and experience-dependent [293]. For example, 80% of the extracted EEG features to train a Support vector machine (SVM) model for the diagnosis of Alzheimer's disease turned out to be marginal or irrelevant [294]. In another study, Samuel O.W *et al.* [243] extracted 32 features from EEG data for interpretation motor imaginary activities, 20-time domain and 12 frequency- domain features. To determine the significant features combination, they had to train four different conventional machine learning models like Linear Discriminant Analysis

(LDA), the Multi-Layer Perceptron (MLP), SVM, and Fisher Linear Discriminant classifier (FLD), while monitoring performance metrics such as classification accuracy, F1-score and sensitivity [264]. In the end, they found that only a combination of four (4) time-domain features or ten (10) frequency domain features were significant with accuracies of 90.68% and 99.55%, respectively [138]. On the same dataset as in Samuel's study, our 2 hidden layers CNN model managed to achieve a 99% classification accuracy without the need of manual feature engineering, this comparison can be seen in Table 5.1. Table 5.1 also shows a comparison of the performance of conventional machine learning and that of CNN on interpretation motor imaginary activities. From Table 5.1, it can be seen that CNN outperformed conventional machine learning models in most cases except in literature [295] of ANN trained with frequency-domain features. Originally, EEG data are in the time domain, needs to be transformed into the frequency domain by Fast Fourier Time (FFT) method, a computation expense. Authors in studies [296] and [297] which used CNN, also compared their methods with conventional machine learning, in both cases, CNN outperformed the results they obtained with conventional machine learning. Compared to other studies that used the CNN model, our results outperformed the rest as shown in Table 5.1 in terms of the number of classes involved and the number of hidden layers of the model. The reason why we think our method outperforms other CNN models, is detailed in Section 5.1.1.3. When fewer layers are used, the computational costs of the model also drop, hence make it possible for the trained model to be deployed in portable devices like tablets and mobile phones or even wearable devices like smartwatches [298].

**Table 5. 1:** CNN performance compared to conventional machine learning models which require feature extraction and selection processes in our first study of first case study

Reference	Feature extraction method	Classification method	Number of Layers	Number of Classes	Average classification accuracy (%)
[299]	Sample Entropy	SVM	Not applicable	3	66.11
[300]	Canonical Spatial Pattern (CSP)	LDA	Not applicable	2	80.00%
[301]	Frequency domain features	LDA, MLP), SVM, FLD	Not applicable	5	99.45%
[295]	Frequency domain features	ANN	No details provided	2	90.00%
[302]	Frequency domain features	KNN	Not applicable	2	70.08
Our work	CNN	CNN	2	5	99.00
[296]	CNN	CNN	5	2	90.75
[297]	CNN	CNN	5	2	86.41

In our second study of case study one, our collaborators trained several conventional machine learning models with Canonical Correlation Analysis (CCA), Power Spectral Density Analysis (PSDA) and Cross-Spectral Density Analysis (CSDA) features to interpret SSVEP for BCI application. An average classification accuracy of 65.63% was obtained in their study. As mentioned above and detailed in Section 2.3.4 of Literature, the extraction of these features is time-consuming, experience- dependent and computationally expensive. Without feature extraction, we obtained an average of 94.54% classification accuracy by using CNN. This comparison, together with other literature applied conventional machine learning and CNN on interpretation SSVEP are compared in Table 5.2. It can be seen that CNN outperformed conventional machine learning in most cases.

Although CNN has been widely applied to image data, the convolution and pooling layers are significantly useful in capturing deep features in time series data like the EEG [320], detailed in Section 2.5.1.2 of the literature review. The overall model learns a fixed set of

rules (filters) and applies the rule to the portion of the data that it sees fit best. Although CNN was initially designed for image data, the above-mentioned properties make them efficient in extracting and learning discriminative features useful for the classification task even for time series data, outperforming conventional machine learning as it has been seen in Tables 5.1 and 5.2. The reason why our CNN model performed better than other CNN studies in both cases (also can be seen in Tables 5.1 and 5.2) while we had more classes as explained in Section 5.1.1.3 below.

**Table 5. 2:** CNN performance compared to conventional machine learning models.

Reference	Feature extraction method	Classification method	Number of layers	Number of classes	Average classification accuracy (%)
[245]	CCA, PSDA and CSDA	SVM, LDA, KNN	Not applicable	6	85.93
[303]	CCA, PSDA	SVM	Not applicable	5	80.00
[304]	PSD	ANN	Not provided	6	94.50
[305]	CCA	Decision tree, Naïve Bayes and KNN	Not applicable	4	70.08
Our work	CNN	CNN	2	6	94.54
[68]	CNN	CNN	2	5	94.42

Our findings imply that, CNN can be employed on interpretation raw EEG data more efficiently than conventional machine learning which needs manual feature engineering. This finding is in line with other research work which has shown that CNN without the need of manual feature extraction performs better than conventional machine learning in interpretation raw EEG data [306][307][308]. The architecture size we used in both of our two studies means small and simple architectures can be used and deployed in portable devices for real-time application, as it has also been demonstrated in real-life applications in study 2 of case study one.

The efficiency of CNN over conventional machine learning can also be seen in biomedical images. A study in comparing the performance of conventional machine learning models and

several deep learning models on biomedical image has acknowledged the superiority of deep learning models [69]. Their comparison involved deep learning models like Deep Boltzmann Machine (DBM), Convolutional Encoder Network (CEN) and Convolutional Neural Network (CNN). The superiority of deep learning over conventional machine learning has been mostly on the feature extraction process. Several literature has mentioned the automatic extraction of pixel-level features as a major strength of deep learning when dealing with image data[69][309][310], which is a significant advantage over conventional machine learning models.

Image data are known to have high dimensionality (as each pixel is considered as a feature) which suits the engineering structure of the CNN. CNN uses a convolutional layer to extract features from an input data. During the convolution process, CNN uses filters /patches of a certain dimension  $(A, g)$  which convolves on top of the input data by considering each data point (pixels in case of image data). The ability to extract pixel-level features makes CNN efficient in extracting relevant features which can be missed by human when performing manual feature extraction as it is required in conventional machine learning. As mentioned in Chapter 4, the current authentication methods of cancer cell lines rely on the genetic content of the cell and not the features of the images. Authentication of cancer cell lines based on the cancer cell images is a new research area of computer vision, so no known image features to distinguish the cell lines. Only a few experienced cancer researchers can discriminate cancer cell lines based on their images and they can't explain what features they are looking when making their decision. This may mean that, it will take time and expertise to manually identify and extract relevant and discriminative features if a conventional machine learning model needs to be used instead of a CNN model for authentication purposes. The significant results of CNN in different authentication levels mean CNN can find and extracts discriminative features even in cases when the features are not known by the experts of the field. This implies that, CNN can be reliable in the development of computerised



authentication of cancer cells in the laboratory environment in a much faster and cost-effective way even in the resources limited to cancer research laboratories, compared to the STR.

### **5.1.1.2 Addressing the “big data” problem in specialised domains**

Despite the increase of biomedical data due to advancement in technologies of biomedical devices, factors like low volume, high sparsity [311], and the need of manual annotation of the data [312], has always limited the use of deep learning in specialised domains such as healthcare [313][314]. Further detailed in Section 2.6.4 of Chapter 2, CNN models require a large amount of training data due to the number of parameters that need to be trained. Through answering research question 3 of case study 2, this thesis has addressed data limitation successfully by applying data augmentation and transfer learning techniques, the two most popular and successful methods in dealing with data unavailability. The two techniques are detailed in Section 2.6.5 of Chapter 2. Transfer learning makes sense when low-level features from the problem you are transferring from could be helpful and are relevant to the problem you are transferring to [315]. For example, a model trained with large scale image datasets like ImageNet can be fine-tuned to perform classification on brain MRI images for the detection of brain cancer. Despite the difference between ImageNet and medical images, there are generic representations that might be beneficial to be learned from a large-scale dataset like ImageNet [335]. This is because, all images contain some common low-level features like edges, local shapes, pixel intensity, pixel gradient orientation, colour and texture [316], [317].

Uran *et.al* [338] acknowledged the unavailability of data and the cost of producing sufficient high-quality annotations in EEG based BCI applications and suggested the use of transfer learning techniques. Although some researches have tried to use transfer learning techniques in EEG based BCI applications [318]–[320], privacy concerns associated with EEG signals

have limited the possibility of constructing a large EEG-BCI dataset for this purpose [321]. Privacy has been an issue in this regard as EEG signals reflect brain activities in numerous aspects, the potential abuse of EEG data may lead to severe privacy violations and hence acts like General Data Protection Regulation (GDPR) [322] prohibit organisations from exchanging data without explicit user approval. Another study by M.Kaya *et.al* [323] has mentioned that, development of more effective data processing and analysis methods for EEG BCI has been hindered by a lack of large, uniform and accessible datasets. These limitations have made transfer learning in EEG based BCI applications unpopular, hence the technique was not used in the studies of our first case study of this thesis. The use of transfer learning has achieved significant performance in biomedical images, due to the availability of large-scale image datasets from other domains. Most of the popular used pre-trained models for transfer learning in image data are pre-trained on the ImageNet [324], a large scale image dataset containing over 14 million images of general objects. With a diverse and large dataset like this, the CNN learns how to extract the features during pre-training, the knowledge which will be useful when the model needs to be fine-tuned to a specific problem. When the new dataset of a specific problem is provided, the pre-trained CNN model does not have to learn again on how to extract features as the required parameters are already there with the appropriate values. The model will just extract the features and train the last few fully connected layers with the features of the specific problem. The fully connected layers are simpler and have fewer parameters to learn, hence, only a small dataset can be enough to train them and achieve good results. Transfer learning does not only address the limitation of the dataset, but also lowers computation costs and training time as only the few last layers will be trained instead of the whole CNN model.

As the popular models used in transfer learning are pre-trained on a general object dataset, the ImageNet, which is different from biomedical images, R.K.Samala *et.al* [97] suggested a technique called multi-stage transfer learning, detailed in Section 4.2.5.4 of Chapter 4. In their

study [325], authors reported a boost of 4% in F1-score as they used a common publicly available breast cancer dataset [272] in an intermediate transfer learning stage, before fine-tuning their model to their specific problem, breast cancer also. In our case, only an increase of 1% in F1-score was observed as we used the breast cancer dataset as our intermediate stage for multistage transfer learning, which is different from our specific task, cancer cell lines. According to Yoshua Bengio *et.al* [42], on CNN, the convolution layers near the input layer extract the generic features while the deeper layers extract specific features. This might be a reason for the low performance of multi-stage transfer learning in our case as our model was not trained with data set similar to the target task in the intermediate stage of transfer learning. As stated in Chapter 4, we couldn't use another cancer cell line dataset in our intermediate stage of transfer learning as there is no publicly available cancer cell lines images, to the best of our knowledge. Our findings and that in literature by R.K.Samala *et.al* [97] on multistage transfer learning suggest that ,we may use the dataset from the domain where it is easy to get large training data even when its different from your target task in the first stage of transfer learning as only generic features like pixels [272] are needed, but training the last few layers with another dataset similar to the target task in the intermediate-stage phase to learn the specific domain feature. As stated above, the layers near the input extract the generic features which are common to many image data, while the deeper layers near the output extract the specific features, hence training a model with a dataset which contains features similar to the target task before fine-tuning to the target task will have a positive impact on the model performance.

The other technique applied in this thesis to address the limitation of data availability (Research question 3, case study 2), a common problem in the healthcare domain, is data augmentation, detailed in Section 2.6.5 of Chapter 2. The idea of data augmentation is to generate a synthetic dataset from the available training data by covering the unexplored input space which is not available in the original training dataset [326]. Figure 4.3 in Chapter 4 shows

examples of augmented data. Several methods have been developed to alleviate overfitting in deep learning such as regularisation [327], transfer learning explained above [328], dropout [36] and batch normalisation [329], but only data augmentation has addressed the overfitting problem at the root, at the training data [234]. Data augmentation has been mostly applied in image data, there are very few studies which have explored the method in time series data [99][330], and less so for EEG data [331], [332]. Less attention has been given to explore data augmentation in time series due to factors like temporal dependency property of time series. Unlike image data, the time series data can be transformed in the frequency and time-frequency domain, in this transformed format, data augmentation methods can be designed and implemented. This becomes more complicated when we model multivariate time series where we need to consider the potentially complex dynamics of these variables across time, thus, simply applying those data augmentation methods from image and speech processing may not result in valid synthetic data [333]. Another factor limiting applications of data augmentation in time series is, data augmentation being task-dependent. The data augmentation methods applicable for time series classification may not be valid for time series anomaly detection [328], [333], [334]. In that case, in this thesis, data augmentation technique was applied in image data only, in case study 2.

Data augmentation increases the data diversity and quantity by performing different transformations on the available training data. For example, Google's AutoAugment, a new automated data augmentation technique, introduces 16 geometric and colour-based transformations [335]. As reported in Chapter 4, Section 4.2.3, we have performed operations like resize, flip, zoom and width shift during the data augmentation process. Shown in Section 4.2.4.4 of Chapter 4, several experiments conducted during the pilot classification has proved that, data augmentation contributed to a huge improvement on the model performance by reducing overfitting and improves generalisation as the model gets to learn from more and diverse training data [89]. Through a pilot classification task shown in Section 4.2.4.4 of

Chapter 4, it can be seen in both tables 4.5- 4.7 that, a combination of both transfer learning and data augmentation had a significant impact on the model performance regardless of having a small training dataset.

### **5.1.1.3 Role of CNN in complementing current practice and allowing new applications not possible previously.**

The issue of misidentification of cancer cell lines has always been a problem in cancer research which involves cancer cell lines. According to literature, limited research has been done in the development of computerised ways for the authentication of cancer cell lines. The authentication process mostly relies on the Short Tandem Repeat (STR) test which uses the genetic content of the cell line for authentication. As mentioned in Chapter 4, the STR test is known to be slow, expensive, incapable of authenticating isogenic cell lines, and not available in every cancer research laboratory. In addressing research question 1 of case study 1, we have tried to complement the STR with a computerised way for the authentication of cancer cell lines. Detailed in Chapter 4, we have successfully applied CNN in the authentication of cancer cell lines with an average of 91% F1-score through a classification task by using microscopic images of cancer cell lines.

Unlike STR, microscopes are common devices that can be found in many research laboratories. The development of a CNN based computerised way for authentication of cancer cell lines by using microscopic images will enable even budget limited research laboratories with a cost-effective way for the authentication of cancer cell lines. The study also addressed the second research question of the second case study by providing additional authentication parameters which are important but can't be conducted with the current existing authentication methods. The eight (8) classes authenticated by the CNN with 91% F1-score involved isogenic cancer cells as the four parental cancer cells with their sublines of drug-treated. To further address the isogenic issue, our optimal model attained an average F1-score of 95% when our

optimal model was trained with only two classes of cancer cell lines, the parental and its subline (drug-treated) as both are coming from the same genetic origin hence isogenic. This is an authentication parameter which cannot be performed with STR. This phenomenon of isogenic authentication may be applied in studying the anti-drug resistance behaviour of cancer cells, which can be of importance in the development of effective anti-cancer drugs.

## **5.1.2 Contributions and implications**

This section discusses the contribution of the studies reported in this thesis and its implication to other research conducted in the field of computerised data analysis. It further looks into the implication to practitioners of this domain.

### **5.1.2.1 Contributions**

Building around answering the research questions of the thesis, I identified four major novel contributions, as presented below:

#### **I. CNN towards the implementation of real-world BCI applications.**

Although CNN has been widely applied in interpretation EEG signals for BCI applications, real- world application of the technology has been limited with several factors including the size of the current existing EEG devices. The current existing EEG devices are bulky, containing a lot of channels which makes them prone to noises and difficult to wear. This has limited the application of BCI technologies only in laboratory settings. By restricting the number of channels, can make the design and fabrication of portable and more efficient EEG devices possible. There are several methods for EEG channels selection in literature, but many of them are difficult to use, computationally expensive, knowledge, and experience-dependent, detailed in Section 2.4 of Chapter 2. Towards the implementation of real-world BCI applications, this thesis has contributed a method for EEG channels selection based on

the weights of a trained CNN model. During the training of a CNN model, significant features, EEG channels in our case, are assigned higher weights. Demonstrated in Section 3.2.4 of Chapter 3 in CNN weight analysis, it can be seen that, the method of using the model's learned weights is knowledge and experience independent, computationally affordable, and relatively easier compared to the current existing methods. In using portable EEG devices with fewer channels, precise placement of the electrodes should be highly followed. None of the existing methods for channel selection guides how to precisely place the fewer electrodes on the scalp. By plotting the topographic map with the weights extracted from a trained CNN model, this thesis has further contributed a method for the guidance in knowing the possible number of channels to be used and for precise placement of the fewer selected electrodes. To test the significance of our contributions, the methods were used to implement and demonstrate real-world BCI applications as reported in the second study of the first case study of this thesis in Chapter 3. This was achieved by deploying a trained CNN model in a tablet, receiving EEG data via Bluetooth and performing classification in real time to perform the control mechanism.

## **II. Introducing a computerised and automatic way of authentication of cancer cell lines.**

In both cancer research and the development of anti-cancer drugs, researchers rely on cancer cell lines to conduct their experiments. For this research to be carried out reliably minimising human errors, authentication of cancer cell lines is a crucial step to make the validity of the results obtained. A lot of research involving cancer cell lines is likely not to be reproducible due to the general reluctance of research labs to perform and report results that establish the identity and purity of their cell lines [336]. This has resulted in funding agencies and publishers to put the requirement for authentication of the cell line involved in the studies due to the critical nature of the issue [337]. There are several methods for the authentication of cancer

cell lines, but the STR test is regarded as the gold standard [337]. The STR test is known to be slow, expensive and knowledge dependent [338], hence not available in many research labs working on cancer cell lines. This works by comparing the genetic profiles of the cell lines. “questioned” profile (the sample being tested), while the other is a “reference” profile, ideally from the same donor. Where another sample from that donor is not available, the questioned STR profile should be compared to other samples from that laboratory and databases online [338], a process subject to skills and experience [338]. This resulted in the need for resource-limited research labs to send their cell line samples to other research labs that have this capability of authentication, which adds costs and time expensive process.

This thesis contributes a method for computerised authentication of cancer cell lines by using CNN, to complement the STR. The contributed methods involve the use of microscopic images of cancer cell lines for authentication. Microscopes are one of the common, relatively easier to operate, and affordable laboratory equipment. Microscopic images of cancer cell lines can be easily produced in the laboratory. The use of CNN in the computerised method for the authentication of cancer cell lines based on the microscopic images as reported in Chapter 4 of this thesis, is cost-effective, easy to use, and can be easily deployed in laboratory environment compared to the STR requirements.

### **III. Additional parameters in the authentication of cancer cell lines.**

In the anti-cancer drug discovery and development process, it involves subjecting the cancer cell to different drugs and study how the cancer cell develops the resistance to different drugs. To study drug resistance mechanisms, differentiating the parental and the drug-treated cancer cell lines is of high importance. Since the parental and the drug-treated cell lines are coming from the same genetic origin, they are termed as “Isogenic”. As described above, the STR work by comparing the genetic profiles, hence won’t be able to differentiate the parental and drug-



treated cell lines which are isogenic. Presented in Section 4.3.1.2 of Chapter 4, this thesis has contributed a way for authentication of isogenic cancer cell lines, to complement the STR. This has been demonstrated by authenticating the eight (8) classes which involves the four (4) classes of parental cancer cell lines and their subline of drug-treated shown in Section 4.3.1.2, and in the authentication, the two (2) classes of a parental cancer cell line and its sublines reported Section 4.3.1.3. In this thesis, through cases study 2 reported in Chapter, we have contributed a method that can differentiate the two classes even though they are isogenic. This contributed authentication parameter which is not possible with the current existing authentication methods, including the STR, which might be useful in studying the resistance mechanism of resistant cancer cells to develop effective drugs.

### **5.1.2.2 Implications to other research**

Supported by literature on CNN, the results reported in both two case studies of this thesis, imply that, CNN can extract useful features automatically and perform better than conventional machine learning. By using methods introduced in this thesis, one can use CNN to automatically extract features and perform weight analysis of the trained CNN model to determine the significant features. The ability to determine significant features from a trained CNN model can be used in understanding mechanism, signs and significant features. Understanding of the features can be useful in new areas of research in which significant features are not yet known. In healthcare, this can be useful in knowing and understanding of the features in unexplored areas like cancer cells resistance behaviours as reported in this thesis. This also becomes the case in new diseases in which very little information is known, like in the recent pandemic of COVID-19. There have been variations among radiologists on how to identify the COVID-19 by using imaging technologies like chest X-rays, CT scans, or ultrasound as not much are known about the disease. A study by H.X.Bai *et.al* [339] reported a strong variation among four (4) American and three (3) Chinese radiologists in differentiating

COVID-19 from viral pneumonia on chest CT. Although there are several successful studies in applying CNN on detection of COVID-19 by using chest imaging technologies [340], not much has been done in explaining how the models make their predictions. By using methods contributed to this thesis, it can be possible to identify the significant features, which can be used to understand the disease better.

Much of the existing literature of CNN based BCI applications has been limited to laboratory tests and offline analysis [246][341][308], which do not reflect a realistic real-world environment. Performing EEG based studies like prediction of cognitive load [341], stress [342], or emotion [343] in the lab environment is unrealistic as the lab conditions are known to be calm, while in the real world these situation occurs with a lot of unexpected things happening around such as surrounding noises, and moving people and objects which can distract the user. The limitation of these studies to be conducted only in the laboratory environment has been the case due to the size of the current existing EEG devices mostly, as they compose a large number of channels, which makes the devices prone to capturing noises, difficult to wear and signal deterioration with time as the gel applied between the electrodes and the scalp dries out. Reported in Chapter 3, this thesis has successfully applied CNN in both EEG channels selection and determining scalp locations for fewer electrodes' placement, with high-quality signal recorded. The newly fabricated 2 channels, skin-like wireless EEG sensor, with dry electrodes, recorded better signal quality in terms of Signal to Noise Ratio (SNR) and Information Transmission Rate (ITR). With the use of portable, durable, easy to use, skin likes sensor guided with an accurate way for placement of the electrodes of the sensor on the scalp, while using simple CNN model due to the quality of EEG signals recorded, all of this introduced in this thesis, can help other EEG based researchers to implement their research in a more real-world and realistic environment.

The research and development of a new drug can take up to 12 years and cost at around £1.15 billion [344]. Due to the sensitivity of healthcare, time, and financial costs involved, no

mistake in the drug research can be afforded, hence authentication of biological samples in use in the research is of high importance [345]. Many of the authentication processes are based on comparing the genetic contents of the samples. Genetic based authentication is expensive, needs high knowledge to conduct and requires special machines and facilities. No much work has been done in the development of computerised ways for the authentication process and the misidentification remains a large problem. In a study by Y. Huan *et.al* [346], they uncovered 22 research laboratories out of 28 researched in China were working on misidentified/contaminated cell lines due to the unavailability of the STR in their laboratory environment. Introduced in this thesis, the use of CNN in the authentication process has been faster, relatively easier, and cost-effective, may provide resource-limited drug research laboratories with the ability to authenticate samples they are working on in the laboratory environment. This can also complement the resourceful laboratories with cost-effective and rapid means of the authentication of their biological samples.

Explainable AI has been an important topic in recent times, the visualisation of the trained CNN models may add value to other researchers trying to build explainable AI solutions. The analysis of topographic maps in the case of EEG or class activation maps in the case of biomedical images demonstrated in this thesis can be used to visualise the trained model to validate, make sense of the result and provide an explanation to the users [347]. Furthermore, as the visualisation technique detects the region of the brain that becomes active during a particular situation, the technique can be used in other brain studies like detecting and locating a brain tumour [348] and intention detection, which might be useful in detecting sexual offenders , for early intervention. For biomedical researchers, the visualisation technique can also be useful in detecting and understanding of different medical conditions, like the detection of phantom pain highlighted in the first study of our case study one or the resistance mechanism of cancer cell lines as highlighted in our second case study.

### **5.1.2.3 Implications for practitioners**

The main focus of studies presented in this thesis has been the application of CNN for computerised analysis of biomedical data analysis. Towards improvement and provision of better healthcare services, the contributions made in this thesis may have a significant impact on both practitioners in the healthcare domain and their patients.

Tracking changes in brain activities may be useful in diagnosing of different brain conditions, such as epilepsy [349], anxiety [350], brain tumour [351], and other seizure disorders [352]. Described in Section 3.2.4.2 and 3.3.1.3 of Chapter 3 of this thesis, by using topographic maps generated from a trained CNN model, it is possible to see which parts of the brain become activities under different conditions. For example, during other brain activities, only neurons of a particular part of the brain will fire at about 80 times per second while conditions like seizure are characterised by abnormal firing of neurons [353], as many as 500 times per second. The topographic maps described in the case study one, can be used as a tool by doctors to detect different brain conditions and be used as a second opinion right at their disposal.

For end-users of BCI technologies, the implementation of real-life BCI technologies can improve the quality of life of people with disabilities as they can gain more control over their environment, and provide more fun to those applying BCI in the entertainment industry. Recently, there has been active research interest in smart homes. Smart homes are an extension of home automation for controlling different house parameters like utilities and appliances [354]. Smart homes are not yet a reality for end users as there are still accessibility challenges for the elderly or the disabled, who are the main potential targets for home automation[355] . The BCI control mechanisms demonstrated in this thesis can be expanded to different control mechanisms parameters changing TV channels, switching lights on and off, essential parameters in smart homes [356].

Besides, the ability to deploy CNN solutions in a laboratory environment can also be applied

in early detection and diagnosis of diseases in resource-limited countries. For example, Tanzania, a developing country with over 60 million population has only five hospitals with cancer diagnosis capabilities, with around 20 practicing oncologists [357], and just 15 pathologists [358]. This makes cancer services inaccessible to everyone, only 26% of low-and middle-income countries (LMICs) reported having pathology services available in the public sector, compared to the 90% of high-income countries as reported by the World Health Organisation (WHO) [359]. Around 80% of women with breast cancer in Tanzania are diagnosed in late-stages which is hard to treat, as a result, more than half of women diagnosed with breast cancer in the country die of the disease [359]. The number of breast cancer new cases are approximated to increase by 82% by 2030 in the country [359]. The ability to detect breast cancer as it was seen in the intermediate stage in a multi-stage transfer learning described in this thesis may mean that, CNN is deployed for early detection and diagnosis of different diseases to complement the current existing diagnosis methods. Applying CNN can reduce the burden and save lives as most of the disease can be treated when detected early and lower the treatment costs.

## **5.2 Conclusion**

This section provides the conclusion of this thesis by summarising the results of the studies reported, highlighting the limitation, and point out the future research direction in order to realise real-life applications of CNN in the computerised and automatic analysis of biomedical data analysis.

### **5.2.1 Limitations**

This thesis has been divided into two cases studies with the aim of exploring the two major forms of biomedical data, the time series data and the image data. In the first case study, we explored the application of CNN in interpretation EEG data for BCI application with two studies reported in this thesis. In the second cases study, we explored the application of CNN

in the authentication of cancer cell lines by using microscopic images of cancer cells. This section will highlight the limitations in the two case studies separately.

In the application of CNN in interpretation EEG signals for BCI application in the first case study, both of the two BCI studies reported only a limited number of control classes. The first study demonstrated only five classes, while the second study demonstrated just six classes. To have BCI applications in daily life applications, there should be more control options. For the case of channels selection, the main limitation of this method is the pooling layer on CNN. We extracted learned weights from the first convolution layer as the learned features across the channels are still intact. The first layers in CNN learn only low-level features hence the weights learned for high-level features in the deeper layers were not accounted for in performing channels selection. Besides, both studies applied a pause in between classes, in different aspects. In the first study, the pause between the classes was manually removed while in the second study, the pose was included as a classification task. The pause was introduced as a way to avoid mental fatigue, a common problem when someone imagine a particular action repetitively as it happen in motor imaginary activities [360], and visual fatigue when someone focuses on a light source continuously as it happen in triggering of Steady State Visual Evoked Potential (SSVEP) [361]. In real life, people don't pause during control activities, instead they switch from one activity to another seamlessly. Another major limitation of the real life BCI application is the size of the current existing EEG devices, as previously highlighted. Although this thesis contributed a channel selection method towards the development of portable EEG devices, our method is still limited to the need of a larger EEG device to identify significant channels and also to determine scalp location for the electrode's placement. The first study has applied a 64 channels EEG device and a second study has applied a 32 channels EEG device.

For the second case study in the authentication of cancer cell lines, our study was conducted on just four (4) cancer cell lines, while there are hundreds of cancer cell lines [362]. Our

findings need to be validated on more cancer cell lines. In addition, our method for visualisation is highly limited since it can't be validated as there are no current methods for authentication of cancer cell lines based on the features of cancer cell images. Furthermore, a multi-stage transfer learning has shown a significant increase of 4% in F1-score in another study by Ravi *et.al* [87] because they used a dataset from their domain in the intermediate stage. Only a 1% increase in F1-score was observed in our case, this is because we were limited with the unavailability of publicly available cancer cell lines to be used in the intermediate stage. Finally, although techniques like data augmentation and transfer learning has shown significant improvement in the model performance, the unavailability of enough training data in medical images is still a huge limitation. In our case, we obtained only an average of 217 images per class were used, while the popular ImageNet dataset has more than 500 images for each class [73].

## **5.2.2 Future research directions**

By looking into the need and significance of computerised methods for the analysis of biomedical data, this section points several research directions towards the implementation of computerised analysis of biomedical data by using CNN, to mitigate some of the limitations addressed above, in each case study.

In both of our two studies in case study one, we have conducted experiments on a few classes as in other work of literature, more control classes should be considered to realise the real-life application of BCI technologies. The limitation of the pulling layer in extracting weights for EEG channels selection can be avoided by removing the pooling layers, although this might result in a high computational coast. Thus, further studies should be conducted in this direction to explore our method more. To address the pause issue between classes of EEG events, we propose the use of continuous flow of event EEG data, and train the models to detect and reject the unknown events. The main reason why mental fatigue happens is due to the repetition of

activities in order to get enough training data of the intended events. Recording of continuous flow of data will require longer periods of data recording. This will provide users with more freedom to perform other activities while using a BCI system. In the case of the need of large EEG devices to determine significant channels and their scalp location, although the use of larger EEG devices comes with its limitations, we recommend the method to be used as large devices covers the whole head which can be used to provide individual's brain information before deploying BCI application to them. This can lead into personalised development of BCI applications. Finally, from the visualisation of the trained CNN model by using a topographic map, abnormalities were seen on the third participant in our first study as can be seen in Figure 3.8 of Chapter 3. It was later found out that, the participant complained about the phantom pain. Future research can also apply some of the findings of this research which were not in our research goals. For example, further research can be done on how to use the topographic maps from a trained CNN model to detect other physiological abnormalities of the brain. We also suggest more research to be done in applying our method of channel selection in other domains for feature selection.

In our second case study, our authentication contribution is limited to just four cancer cell lines. For this method of computerised authentication of cancer cell lines to be applied in research labs, it needs to be validated on a greater number of cancer cell lines. As currently, the authentication process looks into the genetic content of the cell line and not on the image, we recommend further studies to be conducted on the visualisation of the trained CNN model to make sense of the model towards implementations of explainable artificial intelligence. With reference to the last two limitations of the second case study highlighted above, this thesis argues the need of availability of public repositories of images of cancer cell lines. There are different cancer cell banks around the world in which researchers used to perform genetic comparisons, but not much has been done on the cancer cell images as many researchers focused on the genetic content of the cell and not the images.



### **5.2.3 Concluding remarks**

By addressing the research questions outlined in chapter 1, this thesis explored the application of the Convolutional Neural Network (CNN) in developing a computerised way for the analysis of biomedical data. To this end, this thesis presented the analysis of biomedical data by using CNN in three studies covering the two common types of biomedical data, the time series, and image data.

For the case of time series biomedical data, CNN was applied in the interpretation of raw EEG signals for BCI applications without the need for feature engineering in the first two studies of this thesis as reported in case study 1. A method for EEG channels selection and determining locations on the scalp for placement of electrodes of a portable EEG device by using CNN was explored. This led to a fabrication of 2 channels, wireless, portable and durable skin like EEG sensor. In biomedical images, we successfully performed the authentication of the cancer cell lines to complement the current existing authentication methods. With our method of using CNN, we also performed authentication of isogenic cancer cell lines, an authentication parameter which cannot be performed by the currently existing authentication methods.

Implications of these techniques and potential applications for which they might be used in both researchers and practitioners' environments have been highlighted in this chapter. It was proposed in the discussion that, the various methods introduced by this thesis have significant implications for the studies of EEG based BCI applications and anti-cancer drug development. However, it should be noted that the techniques reported within the body of this thesis, such applications are conceptual at this time and further research is recommended to fully establish how these techniques might be used to complement and further extended the currently existing methods.

# References

- [1] B. Savoie, F. J. Lexa, and P. Nagy, “Radiologist Technologist Communication,” *J. Am. Coll. Radiol.*, vol. 10, no. 2, pp. 144–145, Feb. 2013.
- [2] A. C. Granero, F. Fuentes-hurtado, and V. N. Ornedo, “A Comparison of Physiological Signal Analysis Techniques and Classifiers for Automatic Emotional Evaluation of Audiovisual Contents,” vol. 10, no. July, pp. 1–14, 2016.
- [3] H. Müller, R. Reihs, K. Zatloukal, and A. Holzinger, “Analysis of biomedical data with multilevel glyphs,” vol. 15, no. Suppl 6, pp. 1–12, 2014.
- [4] R. Ikeda, M. Nishimura, Y. Sun, M. Wada, and K. Nakashima, “Simple HPLC-UV determination of nucleosides and its application to the authentication of Cordyceps and its allies,” *Biomed. Chromatogr.*, vol. 22, no. 6, pp. 630–636, 2008.
- [5] O. W. Samuel *et al.*, “Determining the Optimal Window Parameters for Accurate and Reliable Decoding of Multiple Classes of Upper Limb Motor Imagery Tasks,” *2018 IEEE Int. Conf. Cyborg Bionic Syst.*, pp. 422–425, 2018.
- [6] D. Ganguly, S. Chakraborty, M. Balitanas, and T. Kim, “Medical Imaging : A Review Medical Imaging : A Review,” no. September 2010, 2015.
- [7] A. Casson, M. Abdulaal, M. Dulabh, S. Kohli, S. Krachunov, and E. Trimble, “Electroencephalogram,” in *Seamless Healthcare Monitoring: Advancements in Wearable, Attachable, and Invisible Devices*, 2018, pp. 45–81.
- [8] A. Bullen, “Microscopic imaging techniques for drug discovery,” *Nat. Rev. Drug Discov.*, vol. 7, pp. 54–67, 2008.
- [9] A. Sidhu and T. S. Dillon, *Biomedical Data and Applications*, vol. 224. 2009.
- [10] E. Thomas and W. Albritton, “Office microscopy in the diagnosis of infectious diseases,” *Can. Fam. Physician*, vol. 34, pp. 379–383, Feb. 1988.
- [11] Y.-C. Zhu, C. Dufouil, C. Tzourio, and H. Chabriat, “Silent brain infarcts: a review of MRI diagnostic criteria,” *Stroke*, vol. 42, no. 4, pp. 1140–1145, Apr. 2011.
- [12] A. Thesis and M. O. F. Technology, *Eeg Signal Analysis and Classification Eeg Signal Analysis and Classification*, no. 22306087. 2004.
- [13] B. Estrany, P. Fuster, A. Garcia-Mas, and Y. Luo, “EOG signal processing and analysis for controlling computer by eye movements,” 2009.
- [14] M. Hallett *et al.*, “Central EMG and tests of motor control. Report of an IFCN committee,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 90, no. 6, pp. 404–432, 1994.
- [15] C. Qin, D. Yao, Y. Shi, and Z. Song, “Computer-aided detection in chest radiography based on artificial intelligence: a survey,” *Biomed. Eng. Online*, vol. 17, no. 1, p. 113, 2018.
- [16] C. Roentgen, “Medical Applications of X Rays,” pp. 25–34, 1895.
- [17] A. Carovac, F. Smajlovic, and D. Junuzovic, “Application of Ultrasound in Medicine,” vol. 2011, no. c, pp. 168–171, 2011.
- [18] B. K. C. Hazen and H. System, “Diagnosis of Infectious Disease Samples for Testing Culture of Microorganisms Testing of a Microorganism ’ s Susceptibility and Sensitivity to Antimicrobial Drugs,” pp. 2–5, 2020.
- [19] Yun ZhouLori BoydLori BoydCeleste LawsonCeleste Lawson, “Errors in Medical Imaging and Radiography Practice: A Systematic Review,” *ournal Med. Imaging Radiat. Sci.*, vol. 46, no. 10, 2015.

- [20] A. Pinto, "Spectrum of diagnostic errors in radiology," *World J. Radiol.*, vol. 2, no. 10, p. 377, 2010.
- [21] and L. F. Mittelstadt, Brent Daniel, "The ethics of big data: current and foreseeable issues in biomedical contexts," *Sci. Eng.*, vol. 22, no. 6, pp. 303–341, 2016.
- [22] S. A. Akar, S. Kara, S. Agambayev, and V. Bilgic, "Nonlinear analysis of EEG in major depression with fractal dimensions," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2015-Novem, pp. 7410–7413, 2015.
- [23] S. K. Ameri *et al.*, "Thinnest Transparent Epidermal Sensor System Based on Graphene," pp. 484–487, 2016.
- [24] K. H. Tse, W. H. Luk, and M. C. Lam, "Pocket-sized versus standard ultrasound machines in abdominal imaging," *Singapore Med. J.*, vol. 55, no. 6, pp. 325–333, Jun. 2014.
- [25] W. T. Sobol, "Recent advances in MRI technology: Implications for image quality and patient safety," *Saudi J. Ophthalmol.*, vol. 26, no. 4, pp. 393–399, 2012.
- [26] Y. Peng, Z. Wu, and J. Jiang, "A novel feature selection approach for biomedical data classification," *J. Biomed. Inform.*, vol. 43, no. 1, pp. 15–23, 2010.
- [27] N. Kühl, M. Goutier, R. Hirt, and G. Satzger, "Machine Learning in Artificial Intelligence: Towards a Common Understanding," 2019.
- [28] M. Riesenhuber and T. Poggio, "Models of object recognition," *Nat. Neurosci.*, vol. 3, no. 11, pp. 1199–1204, 2000.
- [29] R. D. Peacocke and D. H. Graf, "An Introduction to Speech and Speaker Recognition," in *Readings in Human–Computer Interaction*, R. M. BAECKER, J. GRUDIN, W. A. S. BUXTON, and S. GREENBERG, Eds. Morgan Kaufmann, 1995, pp. 546–553.
- [30] A. T. D. A and R. K. B, "The potential for artificial intelligence in healthcare," vol. 6, no. 2, pp. 94–98, 2019.
- [31] C. Engineering, "Holalu Seenappa Sheshadri and Arumugam Kandaswamy 1 INTRODUCTION 2 MATERIALS AND METHODS 3 SIMULATED RESULTS," no. 2, pp. 105–107, 2006.
- [32] W. Sittiprapaporn, "The design and interpretation of EEGs, FFTs, visual EPs, auditory EPs and auditory P300 ERPs studies with topographic brain mapping," *Journal of Engineering and Applied Sciences*, vol. 7, no. 6, pp. 399–404, 2012.
- [33] H. Singh, G. D. Schiff, M. L. Graber, I. Onakpoya, and M. J. Thompson, "The global burden of diagnostic errors in primary care," *BMJ Qual. Saf.*, vol. 26, no. 6, pp. 484–494, 2017.
- [34] R. Bose, A. Khasnobish, S. Bhaduri, and D. N. Tibarewala, "Performance analysis of left and right lower limb movement classification from EEG," in *2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN)*, 2016, pp. 174–179.
- [35] V. Varshney, D. Singh, and A. Tiwari, "Deep learning and its application in silent sound technology," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2016, pp. 2716–2721.
- [36] R. Kumar, R. Srivastava, and S. Srivastava, "Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features," *J. Med. Eng.*, vol. 2015, pp. 1–14, 2015.
- [37] L. I. U. Jingwei, C. Yin, and Z. Weidong, "Deep Learning EEG Response Representation for Brain Computer Interface," *2015 34th Chinese Control Conf.*, pp. 3518–3523, 2015.
- [38] S. N. Abdulkader, A. Atia, and M. S. M. Mostafa, "Brain computer interfacing: Applications and challenges," *Egypt. Informatics J.*, vol. 16, no. 2, pp. 213–230, 2015.
- [39] M. Zhu, W. Liu, and P. Wargoeki, "Changes in EEG signals during the cognitive activity at varying air temperature and relative humidity.," *J. Expo. Sci. Environ. Epidemiol.*, vol. 30, no. 2, pp. 285–298, Mar. 2020.

- [40] M. A. Persinger, "Brain electromagnetic activity and lightning: potentially congruent scale-invariant quantitative properties.," *Front. Integr. Neurosci.*, vol. 6, p. 19, 2012.
- [41] T. Kaufmann, A. Herweg, and A. Kübler, "Toward brain-computer interface based wheelchair control utilizing tactually-evoked event-related potentials," *J. Neuroeng. Rehabil.*, vol. 11, no. 1, p. 7, 2014.
- [42] N. Padfield, J. Zabalza, H. Zhao, V. Masero, and J. Ren, "EEG-based brain-computer interfaces using motor-imagery: Techniques and challenges," *Sensors (Switzerland)*, vol. 19, no. 6, pp. 1–34, 2019.
- [43] E. R. Miranda and A. Brouse, "Interfacing the Brain Directly with Musical Systems: On Developing Systems for Making Music with Brain Signals," *Leonardo*, vol. 38, no. 4, pp. 331–336, 2005.
- [44] A. E. Selim, M. A. Wahed, and Y. Kadah, "Machine Learning Methodologies in Brain-Computer Interface Systems Machine Learning Methodologies in P300 Speller Brain-Computer Interface Systems Abstract," no. May 2014, 2009.
- [45] Y. Yang and J. Zhou, "Recognition and analyses of EEG&ERP signals related to emotion: From the perspective of psychology," *2005 First Int. Conf. Neural Interface Control. Proc.*, no. May, pp. 96–99, 2005.
- [46] A. Temko, E. Thomas, W. Marnane, G. Lightbody, and G. Boylan, "EEG-based neonatal seizure detection with Support Vector Machines," *Clin. Neurophysiol.*, vol. 122, no. 3, pp. 464–473, 2011.
- [47] H. Cecotti and A. Gra, "Convolutional Neural Networks for P300 Detection with Application to Brain-Computer Interfaces," vol. 33, no. 3, pp. 433–445, 2011.
- [48] B. Practices *et al.*, "Effectiveness of Laboratory Practices to Reduce Specimen Labeling Errors at the Time of Specimen Collection in Healthcare Settings : Systematic Review," no. September 2017, pp. 244–258, 2015.
- [49] C. C. A. Azikiwe, C. C. Ifezulike, I. M. Siminialayi, L. U. Amazu, J. C. Enye, and O. E. Nwakwunite, "A comparative laboratory diagnosis of malaria: microscopy versus rapid diagnostic test kits," *Asian Pac. J. Trop. Biomed.*, vol. 2, no. 4, pp. 307–310, Apr. 2012.
- [50] C. Raje and J. Rangole, "Detection of Leukemia in microscopic images using image processing," 2014, pp. 255–259.
- [51] N. K. Al-Salihy and T. Ibrikci, "Classifying breast cancer by using decision tree algorithms," *ACM Int. Conf. Proceeding Ser.*, pp. 144–148, 2017.
- [52] P. Care, R. O. Meachern, A. J. Pesch, and S. H. Patel, "Fundamentals of Diagnostic Error in Imaging," no. 5, pp. 6–8, 2018.
- [53] D. Imaging, B. J. Brice, Y. W. Kim, and T. Army, "To err is human ; analysis finds radiologists very human," *Diagn. Imaging*, no. December, pp. 1–2, 2010.
- [54] C. Ohrt, Purnomo, M. A. Sutamihardja, D. Tang, and K. C. Kain, "Impact of Microscopy Error on Estimates of Protective Efficacy in Malaria-Prevention Trials," *J. Infect. Dis.*, vol. 186, no. 4, pp. 540–546, 2002.
- [55] D. L. Rubin, N. H. Shah, and F. Natalya, "Biomedical ontologies : a functional perspective," vol. 9, no. 1, 2007.
- [56] A. S. Parihar, "A study on brain tumor segmentation using convolution neural network," *Proc. Int. Conf. Inven. Comput. Informatics, ICICI 2017*, no. Icici, pp. 198–201, 2018.
- [57] R. B. Parikh *et al.*, "Machine Learning Approaches to Predict 6-Month Mortality Among Patients With Cancer," *JAMA Netw. Open*, vol. 2, no. 10, pp. e1915997–e1915997, 2019.
- [58] E. Bercovich and M. C. Javitt, "Medical Imaging: From Roentgen to the Digital Revolution, and Beyond," *Rambam Maimonides Med. J.*, vol. 9, no. 4, p. e0034, 2018.

- [59] J. C. Fu, S. K. Lee, S. T. C. Wong, J. Y. Yeh, A. H. Wang, and H. K. Wu, "Image segmentation feature selection and pattern classification for mammographic microcalcifications," vol. 29, pp. 419–429, 2005.
- [60] J. J. Nirschl *et al.*, "A deep-learning classifier identifies patients with clinical heart failure using whole-slide images of H&E tissue," *PLoS One*, vol. 13, no. 4, pp. 1–16, 2018.
- [61] G. L. X. Li, O.W. Samuel, X. Zhang, H. Wang, P. Fang, "A motion-classification strategy based on sEMG-EEG signal combination for upper-limb amputees," *J. Neuroeng. Rehabil.*, vol. 14, no. 1, p. 2, 2017.
- [62] A. L. Goldberger *et al.*, "Current Perspective," 2000.
- [63] L. Zhang, L. Lu, I. Nogues, R. M. Summers, S. Liu, and J. Yao, "DeepPap: Deep convolutional networks for cervical cell classification," *IEEE J. Biomed. Heal. Informatics*, vol. 21, no. 6, pp. 1633–1643, 2017.
- [64] R. Rao and K. Makkithaya, "Learning from a Class Imbalanced Public Health Dataset: a Cost-based Comparison of Classifier Performance," *Int. J. Electr. Comput. Eng.*, vol. 7, p. 2215, 2017.
- [65] K. Huang, H. Yang, I. King, and M. R. Lyu, "Imbalanced learning with a biased minimax probability machine," *IEEE Trans. Syst. Man, Cybern. Part B*, vol. 36, no. 4, pp. 913–923, 2006.
- [66] G. Ciccarelli *et al.*, "Comparison of Two-Talker Attention Decoding from EEG with Nonlinear Neural Networks and Linear Methods," no. January, pp. 1–10, 2019.
- [67] J. Seo, W. H. Kim, W. Baek, B. Nam, and S. H. Noh, "Failure-atomic slotted paging for persistent memory," *Int. Conf. Archit. Support Program. Lang. Oper. Syst. - ASPLOS*, vol. Part F1271, no. 1, pp. 91–104, 2017.
- [68] N.-S. Kwak, K.-R. Müller, and S.-W. Lee, "A convolutional neural network for steady state visual evoked potential classification under ambulatory environment," *PLoS One*, vol. 12, no. 2, p. e0172578, Feb. 2017.
- [69] M. F. Rachmadi, M. C. Valdés-hernández, M. Leonora, F. Agan, and T. Komura, "Deep Learning vs . Conventional Machine Learning : Pilot Study of WMH Segmentation in Brain MRI with Absence or Mild Vascular Pathology †," pp. 1–19.
- [70] T. N. Sainath, A. Mohamed, B. Kingsbury, and B. Ramabhadran, "Deep convolutional neural networks for LVCSR," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 8614–8618.
- [71] KOUSTUBH, "ResNet, AlexNet, VGGNet, Inception: Understanding various architectures of Convolutional Networks." .
- [72] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," *31st AAAI Conf. Artif. Intell. AAAI 2017*, pp. 4278–4284, 2017.
- [73] GitHUB, "ImageNet 1000." .
- [74] R. K. Tripathy, S. Mahanta, and S. Paul, "Artificial intelligence-based classification of breast cancer using cellular images," *RSC Adv.*, vol. 4, no. 18, pp. 9349–9355, 2014.
- [75] M. Hosseini, S. Member, D. Pompili, and S. Member, "Optimized Deep Learning for EEG Big Data and Seizure Prediction BCI via Internet of Things," vol. 3, no. 4, pp. 392–404, 2017.
- [76] M. Rashid *et al.*, "Current Status, Challenges, and Possible Solutions of EEG-Based Brain-Computer Interface: A Comprehensive Review.," *Front. Neurobot.*, vol. 14, p. 25, 2020.
- [77] S. Saha, K. A. Mamun, K. Ahmed, R. Mostafa, and R. Ganesh, "Progress in Brain Computer Interfaces : Challenges and Trends Progress in Brain Computer Interfaces : Challenges and Trends," no. January, 2019.
- [78] S.-A. Park, H.-J. Hwang, J.-H. Lim, J.-H. Choi, H.-K. Jung, and C.-H. Im, "Evaluation of feature extraction methods for EEG-based brain-computer interfaces in terms of robustness to slight

- changes in electrode locations,” *Med. Biol. Eng. Comput.*, vol. 51, no. 5, pp. 571–579, 2013.
- [79] G. Baranauskas, “What limits the performance of current invasive brain machine interfaces?,” *Front. Syst. Neurosci.*, vol. 8, p. 68, Apr. 2014.
- [80] B. J. Gluckman, “Optimal-channel Selection Algorithms in Mental Tasks based Brain-computer Interface,” pp. 0–5, 2018.
- [81] T. Alotaiby, F. E. A. El-Samie, S. A. Alshebeili, and I. Ahmad, “A review of channel selection algorithms for EEG signal processing,” *EURASIP J. Adv. Signal Process.*, vol. 2015, no. 1, p. 66, Aug. 2015.
- [82] K.-J. Wang *et al.*, “Brain-computer interface combining eye saccade two-electrode EEG signals and voice cues to improve the maneuverability of wheelchair.,” *IEEE Int. Conf. Rehabil. Robot.*, vol. 2017, pp. 1073–1078, Jul. 2017.
- [83] and S. S. S.J. Schiff, A. Aldroubi, M. Unser, “Fast wavelet transformation of EEG,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 91, no. 6, pp. 442–445, 1994.
- [84] M. I. Razzak, S. Naz, and A. Zaib, “Deep Learning for Medical Image Processing : Overview , Challenges and Future,” pp. 1–30.
- [85] N. Pakenham-Walsh and F. Bukachi, “Information needs of health care workers in developing countries: a literature review with a focus on Africa,” *Hum. Resour. Health*, vol. 7, p. 30, Apr. 2009.
- [86] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” *J. Big Data*, 2019.
- [87] R. K. Samala, H. P. Chan, L. Hadjiiski, M. A. Helvie, C. D. Richter, and K. H. Cha, “Breast cancer diagnosis in digital breast tomosynthesis: Effects of training sample size on multi-stage transfer learning using deep neural nets,” *IEEE Trans. Med. Imaging*, vol. 38, no. 3, pp. 686–696, 2019.
- [88] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” pp. 1–14, 2014.
- [89] N. Dvornik, J. Mairal, S. Member, and C. Schmid, “On the Importance of Visual Context for Data Augmentation in Scene Understanding,” pp. 1–15.
- [90] J. Luo, M. Wu, D. Gopukumar, and Y. Zhao, “Big Data Application in Biomedical Research and Health Care: A Literature Review,” *Biomed. Inform. Insights*, vol. 8, pp. 1–10, Jan. 2016.
- [91] J. N. Kok, E. J. W. Boers, W. A. Kusters, P. Van Der Putten, and M. Poel, “SA NE M SC PL O E – C EO AP LS TE S M SC PL O E –.”
- [92] M. Drahanaky *et al.*, “We are IntechOpen , the world ’ s leading publisher of Open Access books Built by scientists , for scientists TOP 1 %,” *Intech*, vol. i, no. tourism, p. 13, 2016.
- [93] L. Deng, G. Hinton, and B. Kingsbury, “New types of deep neural network learning for speech recognition and related applications: An overview,” *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, pp. 8599–8603, 2013.
- [94] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 779–788, 2016.
- [95] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, “On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima,” pp. 1–16, 2016.
- [96] J. Chang, J. Yu, T. Han, H. J. Chang, and E. Park, “A method for classifying medical images using transfer learning: A pilot study on histopathology of breast cancer,” *2017 IEEE 19th Int. Conf. e-Health Networking, Appl. Serv. Heal. 2017*, vol. 2017-Decem, pp. 1–4, 2017.
- [97] R. K. Samala, H. P. Chan, L. Hadjiiski, M. A. Helvie, C. D. Richter, and K. H. Cha, “Breast cancer diagnosis in digital breast tomosynthesis: Effects of training sample size on multi-stage

- transfer learning using deep neural nets,” *IEEE Trans. Med. Imaging*, vol. 38, no. 3, pp. 686–696, 2019.
- [98] O. D. Anderson, “More effective time-series analysis and forecasting,” *J. Comput. Appl. Math.*, vol. 64, no. 1, pp. 117–147, 1995.
- [99] A. T. Jebb, L. Tay, W. Wang, and Q. Huang, “Time series analysis for psychological research: examining and forecasting change,” *Front. Psychol.*, vol. 6, p. 727, 2015.
- [100] E. Techniques, “Electrocardiography and Introduction to Electrophysiologic Techniques,” pp. 145–181, 1924.
- [101] Y. Chu, X. Zhao, J. Han, and Y. Su, “Physiological Signal-Based Method for Measurement of Pain Intensity,” *Front. Neurosci.*, vol. 11, p. 279, 2017.
- [102] P. Thodoroff, J. Pineau, and A. Lim, “Learning Robust Features using Deep Learning for Automatic Seizure Detection,” pp. 1–12.
- [103] F. Wang, C. Zhou, X. Hao, S. Wang, and G. Yang, “BCI Control System for Humanoid Robot Based on Motor Imaginary,” *2013 25th Chinese Control Decis. Conf.*, pp. 5140–5143, 2013.
- [104] T. M, “Fundamentals of Eeg Measurement,” *Meas. Sci. Rev.*, vol. 2, no. 2, pp. 1–11, 2002.
- [105] “Intro to Brain Computer Interface.” .
- [106] G. Fiscon *et al.*, “Combining EEG signal processing with supervised methods for Alzheimer’s patients classification,” *BMC Med. Inform. Decis. Mak.*, vol. 18, no. 1, pp. 1–10, 2018.
- [107] X.-W. Wang, D. Nie, and B.-L. Lu, “EEG-Based Emotion Recognition Using Frequency Domain Features and Support Vector Machines BT - Neural Information Processing,” 2011, pp. 734–743.
- [108] M. Li and B. Lu, “Emotion classification based on gamma-band EEG,” in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 1223–1226.
- [109] Y. Wang, X. Gao, B. Hong, C. Jia, and S. Gao, “Brain-Computer Interfaces Based on Visual Evoked Potentials,” *IEEE Eng. Med. Biol. Mag.*, vol. 27, no. 5, pp. 64–71, 2008.
- [110] M. H. Diaz *et al.*, “Order and Chaos in the Brain: Fractal Time Series Analysis of the EEG Activity During a Cognitive Problem Solving Task,” *Procedia Comput. Sci.*, vol. 55, pp. 1410–1419, 2015.
- [111] M. Z. Parvez and M. Paul, “EEG signal classification using frequency band analysis towards epileptic seizure prediction,” *16th Int’l Conf. Comput. Inf. Technol. ICCIT 2013*, no. December 2013, pp. 126–130, 2014.
- [112] K. Elf, E. Ronne-Engstrom, R. Semnic, E. Rostami-Berglund, J. Sundblom, and M. Zetterling, “Continuous EEG monitoring after brain tumor surgery,” *Acta Neurochir. (Wien).*, vol. 161, no. 9, pp. 1835–1843, Sep. 2019.
- [113] J. Gwizdka, Y. Moshfeghi, F. E. Pollick, J. Mostafa, and O. Bergman, “Applications of neuroimaging in information science: Challenges and opportunities,” *Proc. ASIST Annu. Meet.*, vol. 50, no. 1, 2013.
- [114] L. Astolfi *et al.*, “Imaging the social brain by simultaneous hyperscanning during subject interaction,” *IEEE Intell. Syst.*, vol. 26, no. 5, pp. 38–45, 2011.
- [115] Q. Wu, Y. Zeng, C. Zhang, L. Tong, and B. Yan, “An EEG-Based Person Authentication System with Open-Set Capability Combining Eye Blinking Signals,” *Sensors (Basel).*, vol. 18, no. 2, Jan. 2018.
- [116] R. Rao and R. Scherer, “Brain-Computer Interfacing [In the Spotlight],” *Signal Process. Mag. IEEE*, vol. 27, pp. 150–152, 2010.
- [117] Y. Zhu *et al.*, “PET Mapping for Brain-Computer Interface Stimulation of the Ventroposterior Medial Nucleus of the Thalamus in Rats with Implanted Electrodes,” *J. Nucl. Med.*, vol. 57, no. 7, pp. 1141–1145, Jul. 2016.

- [118] N. R. Anderson, T. Blakely, G. Schalk, E. C. Leuthardt, and D. W. Moran, “Electrocorticographic (ECoG) correlates of human arm movements,” *Exp. Brain Res.*, vol. 223, no. 1, pp. 1–10, 2012.
- [119] K. M. Szostak, L. Grand, and T. G. Constandinou, “Neural Interfaces for Intracortical Recording: Requirements, Fabrication Methods, and Characteristics,” *Front. Neurosci.*, vol. 11, p. 665, 2017.
- [120] C. J. Stam, “Use of magnetoencephalography (MEG) to study functional brain networks in neurodegenerative disorders,” *J. Neurol. Sci.*, vol. 289, no. 1, pp. 128–134, 2010.
- [121] W. Zhang, C. Tan, F. Sun, H. Wu, and B. Zhang, “A Review of EEG-Based Brain-Computer Interface Systems Design,” *Brain Sci. Adv.*, vol. 4, no. 2, pp. 156–167, Dec. 2018.
- [122] D. Wu, C.-Y. Li, and D.-Z. Yao, “Scale-Free Music of the Brain,” *PLoS One*, vol. 4, no. 6, p. e5915, Jun. 2009.
- [123] N. Kosmyna and A. Lécuyer, “Designing Guiding Systems for Brain-Computer Interfaces,” *Front. Hum. Neurosci.*, vol. 11, p. 396, 2017.
- [124] S.-W. Chen and Y.-C. Lai, “A signal-processing-based technique for P300 evoked potential detection with the applications into automated character recognition,” *EURASIP J. Adv. Signal Process.*, vol. 2014, no. 1, p. 152, 2014.
- [125] C. Neuper, M. Wörtz, and G. Pfurtscheller, “ERD/ERS patterns reflecting sensorimotor activation and deactivation,” *Prog. Brain Res.*, vol. 159, pp. 211–222, 2006.
- [126] G. Pfurtscheller, C. Neuper, and W. Mohl, “Event-related desynchronization (ERD) during visual processing,” *Int. J. Psychophysiol.*, vol. 16, no. 2, pp. 147–153, 1994.
- [127] S. P. J. M. Horbach and W. Halffman, “The ghosts of HeLa: How cell line misidentification contaminates the scientific literature,” *PLoS One*, vol. 12, no. 10, pp. 1–16, 2017.
- [128] J. A. Seoane *et al.*, “Biomedical Data Integration in Computational Drug Design and Bioinformatics,” *Curr. Comput. Aided. Drug Des.*, vol. 9, Nov. 2012.
- [129] U. Strehl, U. Leins, G. Goth, C. Klinger, T. Hinterberger, and N. Birbaumer, “Self-regulation of Slow Cortical Potentials : A New Treatment for Children With,” 2019.
- [130] N. Ute Strehl, Ulrike Leins, Gabriella Goth, Christoph Klinger, Thilo Hinterberger, “Self-regulation of slow cortical potentials: an event-related fMRI study,” *Pediatrics*, vol. 118, no. 5, 2016.
- [131] X. Mao *et al.*, “Progress in EEG-based brain robot interaction systems,” *Comput. Intell. Neurosci.*, vol. 2017, 2017.
- [132] Z. Işcan and V. V. Nikulin, “Steady state visual evoked potential (SSVEP) based brain-computer interface (BCI) performance under different perturbations,” *PLoS One*, vol. 13, no. 1, pp. 1–17, 2018.
- [133] A. B. Schwartz, “Leading Edge Perspective Movement: How the Brain Communicates with the World,” *Cell*, vol. 164, no. 6, pp. 1122–1135, 2016.
- [134] H. Rao, M. Korczykowski, J. Pluta, A. Hoang, and J. A. Detre, “Neural correlates of voluntary and involuntary risk taking in the human brain: an fMRI Study of the Balloon Analog Risk Task (BART),” *Neuroimage*, vol. 42, no. 2, pp. 902–910, 2008.
- [135] J. F. Delgado Saa and M. Çetin, “Hidden conditional random fields for classification of imaginary motor tasks from EEG data,” *Eur. Signal Process. Conf.*, no. Eusipco, pp. 1377–1381, 2011.
- [136] “Nonprimary motor cortex - Wikipedia.”
- [137] G. Samuel, Oluwarotimi & Li, Xiangxin & Zhang, Xu & Wang, Hui & Li, “A Hybrid Non-invasive Method for the Classification of Amputee’s Hand and Wrist Movements,” *Int. Conf. Biomed. Heal. Inform.*, pp. 161–166, 2015.
- [138] O. W. Samuel, Y. Geng, X. Li, and G. Li, “Towards Efficient Decoding of Multiple Classes of



- Motor Imagery Limb Movements Based on EEG Spectral and Time Domain Descriptors,” *J. Med. Syst.*, vol. 41, no. 12, p. 194, Oct. 2017.
- [139] J. Chen, D. Zhang, A. K. Engel, Q. Gong, and A. Maye, “Application of a single-flicker online SSVEP BCI for spatial navigation,” *PLoS One*, vol. 12, no. 5, pp. 1–13, 2017.
- [140] H. Jia, Y. Li, and D. Yu, “Attenuation of long-range temporal correlations of neuronal oscillations in young children with autism spectrum disorder,” *NeuroImage. Clin.*, vol. 20, pp. 424–432, 2018.
- [141] B. Z. Id and P. Cao, “Classification of high dimensional biomedical data based on feature selection using redundant removal,” pp. 1–19, 2019.
- [142] H. K. Huang, “Medical imaging informatics research and development trends—An editorial,” *Comput. Med. Imaging Graph.*, vol. 29, no. 2, p. 91, 2005.
- [143] “Biomedical Imaging - an overview \_ ScienceDirect Topics.” .
- [144] W. Huda and R. B. Abrahams, “X-Ray-Based Medical Imaging and Resolution,” *Am. J. Roentgenol.*, vol. 204, no. 4, pp. W393–W397, Mar. 2015.
- [145] S. P. Power, F. Moloney, M. Twomey, K. James, O. J. O’Connor, and M. M. Maher, “Computed tomography and patient risk: Facts, perceptions and uncertainties,” *World J. Radiol.*, vol. 8, no. 12, pp. 902–915, Dec. 2016.
- [146] A. Seery, H. Tager-Flusberg, and C. A. Nelson, “Event-related potentials to repeated speech in 9-month-old infants at risk for autism spectrum disorder,” *J. Neurodev. Disord.*, vol. 6, no. 1, p. 43, 2014.
- [147] S. Sippel, K. Muruganandan, A. Levine, and S. Shah, “Review article: Use of ultrasound in the developing world,” *Int. J. Emerg. Med.*, vol. 4, p. 72, Dec. 2011.
- [148] A. Berger, “Magnetic resonance imaging,” *BMJ*, vol. 324, no. 7328, p. 35, Jan. 2002.
- [149] J.-F. Xian, M. Chen, and Z.-Y. Jin, “Magnetic resonance imaging in clinical medicine: current status and potential future developments in China,” *Chinese medical journal*, vol. 128, no. 5, pp. 569–570, Mar-2015.
- [150] H. Kurt, G. Ekici, A. Murat, and Ö. Aksu, “On the concept ‘ Microscope ’: Biology student teachers ’ cognitive structure,” vol. 8, no. 19, pp. 1859–1874, 2013.
- [151] B. Ngasala and S. Bushukatale, “Evaluation of malaria microscopy diagnostic performance at private health facilities in Tanzania,” *Malar. J.*, vol. 18, no. 1, p. 375, 2019.
- [152] L. Al-Lamki, “Radiation Exposure from Medical Imaging: A Wake-up Call for Oman!,” *Sultan Qaboos Univ. Med. J.*, vol. 11, no. 1, pp. 1–4, Feb. 2011.
- [153] I. Scholl, T. Aach, T. M. Deserno, and T. Kuhlen, “Challenges of medical image processing,” *Comput. Sci. - Res. Dev.*, vol. 26, no. 1–2, pp. 5–13, 2011.
- [154] M. M. Hasan, M. W. Alam, K. A. Wahid, S. Miah, and K. E. Lukong, “A Low-Cost Digital Microscope with Real-Time Fluorescent Imaging Capability,” *PLoS One*, vol. 11, no. 12, pp. 1–8, 2016.
- [155] “Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013,” *Lancet (London, England)*, vol. 385, no. 9963, pp. 117–171, Jan. 2015.
- [156] R. A. Mahumud, M. Sultana, N. Sheikh, M. N. Ali, D. K. Mitra, and A. R. Sarker, “Diagnostic Errors in Low and Middle-Income Countries: Future Health and Economic Burden for Patient Safety,” *Austin Emerg Med*, 2016.
- [157] “Study Suggests Medical Errors Now Third Leading Cause of Death in the U.” .
- [158] E. P. Balogh, B. T. Miller, J. R. Ball, of Sciences Engineering, Medicine, and others, “Overview of diagnostic error in health care,” in *Improving Diagnosis in Health Care*, National Academies Press (US), 2015.

- [159] E. A. Walker and H. H. Abujudeh, “Understanding and Confronting Our Mistakes : The Epidemiology of Error in Radiology and Strategies for Error Reduction 1,” pp. 1668–1676, 2015.
- [160] J. Roski, G. W. Bo-Linn, and T. A. Andrews, “Creating value in health care through big data: opportunities and policy implications.,” *Health Aff. (Millwood)*, vol. 33, no. 7, pp. 1115–1122, Jul. 2014.
- [161] “healthcare data.” .
- [162] H. B. Barlow, “Unsupervised learning,” *Neural Comput.*, vol. 1, no. 3, pp. 295–311, 1989.
- [163] X. Zhu and A. B. Goldberg, “Introduction to semi-supervised learning,” *Synth. Lect. Artif. Intell. Mach. Learn.*, vol. 3, no. 1, pp. 1–130, 2009.
- [164] A. Soofi and A. Awan, “Classification Techniques in Machine Learning: Applications and Issues,” *J. Basic Appl. Sci.*, vol. 13, pp. 459–465, 2017.
- [165] B. P. Chacko, V. R. V. Krishnan, G. Raju, and P. B. Anto, “Handwritten character recognition using wavelet energy and extreme learning machine,” *Int. J. Mach. Learn. Cybern.*, vol. 3, no. 2, pp. 149–161, 2012.
- [166] E. Kremic and A. Subasi, “Performance of random forest and SVM in face recognition.,” *Int. Arab J. Inf. Technol.*, vol. 13, no. 2, pp. 287–293, 2016.
- [167] M. J. Myers, “Voice recognition software and a hand-held translation machine for second-language learning,” *Comput. Assist. Lang. Learn.*, vol. 13, no. 1, pp. 29–41, 2000.
- [168] N. F. Osman, E. R. McVeigh, and J. L. Prince, “Imaging heart motion using harmonic phase MRI,” *IEEE Trans. Med. Imaging*, vol. 19, no. 3, pp. 186–202, 2000.
- [169] E. Erkan and M. Akbaba, “A study on performance increasing in SSVEP based BCI application,” *Eng. Sci. Technol. an Int. J.*, vol. 21, no. 3, pp. 421–427, 2018.
- [170] A. Dasgupta, Y. V. Sun, I. R. König, J. E. Bailey-Wilson, and J. D. Malley, “Brief review of regression-based and machine learning methods in genetic epidemiology: the Genetic Analysis Workshop 17 experience,” *Genet. Epidemiol.*, vol. 35 Suppl 1, no. Suppl 1, pp. S5–S11, 2011.
- [171] R. Xu and D. Wunsch, “Survey of clustering algorithms,” *IEEE Trans. neural networks*, vol. 16, no. 3, pp. 645–678, 2005.
- [172] A. Alsayat and H. El-Sayed, “Efficient genetic K-Means clustering for health care knowledge discovery,” in *2016 IEEE 14th International Conference on Software Engineering Research, Management and Applications (SERA)*, 2016, pp. 45–52.
- [173] A. B. Gonçalves *et al.*, “Feature Extraction and Machine Learning for the Classification of Brazilian Savannah Pollen Grains,” *PLoS One*, vol. 11, no. 6, pp. 1–20, 2016.
- [174] A. Chandiok and D. K. Chaturvedi, “Machine learning techniques for cognitive decision making,” in *2015 IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions (WCI)*, 2015, pp. 1–6.
- [175] and W. T. Liu, Zhichao, Brian Delavan, Ruth Roberts, “Lessons learned from two decades of anticancer drugs,” *Trends Pharmacol. Sci.* 38, vol. 38, no. 10, pp. 852–872, 2017.
- [176] C. Paper, M. Todisco, D. Casali, G. Saggio, and L. Bianchi, “SVM Classification of EEG Signals for Brain Computer Interface,” no. January, 2009.
- [177] R. Singla and B. A. Haseena, “Comparison of SSVEP Signal Classification Techniques Using SVM and ANN Models for BCI Applications,” vol. 4, no. 1, pp. 6–10, 2014.
- [178] L. Sun, B. Jin, H. Yang, J. Tong, C. Liu, and H. Xiong, “Unsupervised EEG Feature Extraction Based on Echo State Network,” *Inf. Sci. (Ny)*, vol. 475, Sep. 2018.
- [179] R. Kumar, R. Srivastava, and S. Srivastava, “Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features,” *J. Med. Eng.*, vol. 2015, p. 457906, 2015.

- [180] C. Demir and B. Yener, “Automated cancer diagnosis based on histopathological images: a systematic survey. Rensselaer Polytechnic Institute Technical Report TR-05-09,” *Dept. Comput. Sci., Rensselaer Polytech. Inst., Troy, NY, USA, Tech. Rep.*, vol. TR-05-09, no. February, pp. 1–16, 2005.
- [181] X. Wang, S. Wang, Y. Zhu, and X. Meng, “Image segmentation based on Support Vector Machine,” in *Proceedings of 2012 2nd International Conference on Computer Science and Network Technology*, 2012, pp. 202–206.
- [182] S. Tian, J. Yu, and C. Yin, “Anomaly Detection Using Support Vector Machines BT - Advances in Neural Networks – ISNN 2004,” 2004, pp. 592–597.
- [183] M. H. Id *et al.*, “A systematic review of the diagnostic accuracy of artificial intelligence-based computer programs to analyze chest x-rays for pulmonary tuberculosis,” pp. 1–19, 2019.
- [184] S. Lahmiri and M. Boukadoum, “Hybrid Discrete Wavelet Transform and Gabor Filter Banks Processing for Features Extraction from Biomedical Images,” vol. 2013, 2013.
- [185] B. G. Teo and S. K. Dhillon, “An automated 3D modeling pipeline for constructing 3D models of MONOGENEAN HARDPART using machine learning techniques,” *BMC Bioinformatics*, vol. 20, no. Suppl 19, p. 658, Dec. 2019.
- [186] D. Komura and S. Ishikawa, “Machine Learning Methods for Histopathological Image Analysis,” *Comput. Struct. Biotechnol. J.*, vol. 16, Sep. 2017.
- [187] C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, “Key challenges for delivering clinical impact with artificial intelligence,” *BMC Med.*, vol. 17, no. 1, p. 195, 2019.
- [188] R. Miotto, F. Wang, S. Wang, and X. Jiang, “Deep learning for healthcare: review, opportunities and challenges,” *Brief. Bioinform.*, vol. 19, May 2017.
- [189] B. Peter and P. Nicholls, “Human-Machine Interaction and Feedback via Physiological Sensing of Oral Ingestive Activity,” 2019.
- [190] M. East, R. Syndrome, I. Committee, G. Surveillance, and S. Who, “Laboratory testing for coronavirus disease ( COVID-19 ) in suspected human cases,” no. March, pp. 1–7, 2020.
- [191] M.VASANTHA, BHARATHI, DR.V.SUBBIAH, and S.DHAMODHARAN, “Medical Image Feature, Extraction, Selection And Classification,” *Int. J. Eng. Sci. Technol.*, vol. 2, Jun. 2010.
- [192] and F. S. S.M. Zhou, J.Q. Gan, “Classifying mental tasks based on features of higher-order statistics from EEG signals in brain–computer interface,” *Inf. Sci. (Ny).*, vol. 178, no. 6, 2008.
- [193] K. J. Parsons, W. J. Cooper, and R. C. Albertson, “Limits of Principal Components Analysis for Producing a Common Trait Space: Implications for Inferring Selection, Contingency, and Chance in Evolution,” *PLoS One*, vol. 4, no. 11, pp. 1–4, 2009.
- [194] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, “Breast cancer histopathological image classification using Convolutional Neural Networks,” *2016 Int. Jt. Conf. Neural Networks*, pp. 2560–2567, 2016.
- [195] J. GORDON, “No Title,” [https://twitter.com/random\\_forests](https://twitter.com/random_forests). .
- [196] M. Z. Baig, N. Aslam, and H. P. H. Shum, “Filtering techniques for channel selection in motor imagery EEG applications: a survey,” *Artif. Intell. Rev.*, Feb. 2019.
- [197] T. Otaiby, F. Abd El-Samie, S. Alshebeili, and I. Ahmad, “A review of channel selection algorithms for EEG signal processing,” *EURASIP J. Adv. Signal Process.*, vol. 2015, Aug. 2015.
- [198] J. K. Feng *et al.*, “An Optimized Channel Selection Method Based on Multifrequency CSP-Rank for Motor Imagery-Based BCI System,” *Comput. Intell. Neurosci.*, vol. 2019, p. 8068357, 2019.
- [199] H. Liu and L. Yu, “Toward integrating feature selection algorithms for classification and clustering,” *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 4, pp. 491–502, 2005.
- [200] K. H. Lai, Z. Zainuddin, and P. Ong, “Efficient feature selection using a hybrid algorithm for the task of epileptic seizure detection,” *AIP Conf. Proc.*, vol. 1605, no. 1, pp. 209–214, Jul. 2014.

- [201] A. Hernández-Blanco, B. Herrera-Flores, D. Tomás, and B. Navarro-Colorado, “A Systematic Review of Deep Learning Approaches to Educational Data Mining,” *Complexity*, vol. 2019, p. 1306039, 2019.
- [202] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya, “Deep learning for healthcare applications based on physiological signals: A review,” *Comput. Methods Programs Biomed.*, vol. 161, pp. 1–13, 2018.
- [203] O. Access, “The barriers to initiating lung cancer care in low-and middle-income countries,” vol. 8688, pp. 1–10, 2020.
- [204] R. Taneja, “STOCK MARKET PREDICTION USING REGRESSION,” pp. 813–815, 2018.
- [205] A. Shewalkar, D. Nyavanandi, and S. A. Ludwig, “Performance Evaluation of Deep Neural Networks Applied to Speech Recognition: RNN, LSTM and GRU,” *J. Artif. Intell. Soft Comput. Res.*, vol. 9, no. 4, pp. 235–245, 2019.
- [206] D. Chen *et al.*, “Deep learning and alternative learning strategies for retrospective real-world clinical data,” *npj Digit. Med.*, no. May, pp. 1–5, 2019.
- [207] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science (80-. )*, vol. 313, no. 5786, pp. 504–507, 2006.
- [208] K. Yu, “Large-scale deep learning at Baidu,” in *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, 2013, pp. 2211–2212.
- [209] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, “Convolutional neural networks: an overview and application in radiology,” *Insights Imaging*, vol. 9, no. 4, pp. 611–629, 2018.
- [210] S. Albawi, T. Abed Mohammed, and S. ALZAWI, *Understanding of a Convolutional Neural Network*. 2017.
- [211] C. S. Burrus and T. W. Parks, *and Convolution Algorithms*. Citeseer, 1985.
- [212] O. Publications, “That luulutuoote,” vol. 2, 2017.
- [213] B. J. Wythoff, “Backpropagation neural networks: A tutorial,” *Chemom. Intell. Lab. Syst.*, vol. 18, no. 2, pp. 115–155, 1993.
- [214] J. E. Solem and N. C. Overgaard, “A geometric formulation of gradient descent for variational problems with moving surfaces,” in *International Conference on Scale-Space Theories in Computer Vision*, 2005, pp. 419–430.
- [215] A. Baldominos, A. Cervantes, Y. Saez, and P. Isasi, “A Comparison of Machine Learning and Deep Learning Techniques for Activity Recognition using Mobile Devices,” *Sensors (Basel)*, vol. 19, no. 3, p. 521, Jan. 2019.
- [216] T. Young, D. Hazarika, S. Poria, and E. Cambria, “Recent trends in deep learning based natural language processing [Review Article],” *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55–75, 2018.
- [217] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, “Deep supervised learning for hyperspectral data classification through convolutional neural networks,” *Int. Geosci. Remote Sens. Symp.*, vol. 2015-Novem, pp. 4959–4962, 2015.
- [218] N. Kwak and C.-H. Choi, “Input feature selection for classification problems.,” *IEEE Trans. neural networks*, vol. 13, no. 1, pp. 143–159, 2002.
- [219] Z. Mao, W. X. Yao, and Y. Huang, “EEG-based Biometric Identification with Deep Learning,” pp. 609–612, 2017.
- [220] N. Coudray *et al.*, “Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning,” *Nat. Med.*, vol. 24, no. 10, pp. 1559–1567, 2018.
- [221] J. Wang, X. Yang, H. Cai, W. Tan, C. Jin, and L. Li, “Discrimination of Breast Cancer with Microcalcifications on Mammography by Deep Learning,” *Sci. Rep.*, vol. 6, no. June, pp. 1–9, 2016.

- [222] B. Zhao, H. Lu, S. Chen, J. Liu, and D. Wu, “Convolutional neural networks for time series classification,” *J. Syst. Eng. Electron.*, vol. 28, pp. 162–169, Feb. 2017.
- [223] X. Lun, Z. Yu, T. Chen, F. Wang, and Y. Hou, “A Simplified CNN Classification Method for MI-EEG via the Electrode Pairs Signals,” *Front. Hum. Neurosci.*, vol. 14, p. 338, 2020.
- [224] H. Ismail, F. Germain, and J. Weber, “Deep learning for time series classification : a review,” pp. 1–44.
- [225] A. Ravi, N. Heydari, and N. Jiang, “User-Independent SSVEP BCI Using Complex FFT Features and CNN Classification,” in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2019, pp. 4175–4180.
- [226] V. Lawhern, A. Solon, N. Waytowich, S. Gordon, C. Hung, and B. Lance, “EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces,” *J. Neural Eng.*, vol. 15, Nov. 2016.
- [227] N. D. Kurniawan, “MRI in the Study of Animal Models of Neurodegenerative Diseases.,” *Methods Mol. Biol.*, vol. 1718, pp. 347–375, 2018.
- [228] D. Wang and A. Khosla, “Deep Learning for Identifying Metastatic Breast Cancer,” pp. 1–6.
- [229] R. Poplin *et al.*, “Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning,” *Nat. Biomed. Eng.*, vol. 2, no. 3, pp. 158–164, 2018.
- [230] A. G. Howard *et al.*, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” 2017.
- [231] C. M. Bunney, P. E., Zink, A. N., Holm, A. A., Billington, C. J., & Kotz, “乳鼠心肌提取 HHS Public Access,” *Physiol. Behav.*, vol. 176, pp. 139–148, 2017.
- [232] Z. Hussain, F. Gimenez, D. Yi, and D. Rubin, “Differential Data Augmentation Techniques for Medical Imaging Classification Tasks,” *AMIA ... Annu. Symp. proceedings. AMIA Symp.*, vol. 2017, pp. 979–984, Apr. 2018.
- [233] M. Christopher *et al.*, “Performance of Deep Learning Architectures and Transfer Learning for Detecting Glaucomatous Optic Neuropathy in Fundus Photographs,” *Sci. Rep.*, vol. 8, no. 1, pp. 1–13, 2018.
- [234] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” *J. Big Data*, vol. 6, no. 1, p. 60, Jul. 2019.
- [235] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Med. Image Anal.*, vol. 42, no. December 2012, pp. 60–88, 2017.
- [236] A. Govindaiah, M. A. Hussain, R. T. Smith, and A. Bhuiyan, “Deep convolutional neural network based screening and assessment of age-related macular degeneration from fundus images,” *Proc. - Int. Symp. Biomed. Imaging*, vol. 2018-April, no. Isbi, pp. 1525–1528, 2018.
- [237] ILSVRC, “Large Scale Visual Recognition Challenge (ILSVRC).” .
- [238] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, “Residual and plain convolutional neural networks for 3D brain MRI classification,” *Proc. - Int. Symp. Biomed. Imaging*, pp. 835–838, 2017.
- [239] J. Huang *et al.*, “Speed/accuracy trade-offs for modern convolutional object detectors,” *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 3296–3305, 2017.
- [240] “PhysioNet.” .
- [241] F. C. Morabito, C. Ieracitano, L. Bonanno, A. Bramanti, S. De, and I. C. N. Bonino-pulejo, “Deep Convolutional Neural Networks for Classification of Mild Cognitive Impaired and Alzheimer ’ s Disease Patients from Scalp EEG Recordings,” 2016.
- [242] and T. . M. S. Kikkert, M. Melvin, D.H. Slater, h. Johansen-Berg, I. Tracey, “Motor correlates of phantom limb pain,” *Cortex*, vol. 95, pp. 29–36, 2017.

- [243] X. Li, O. W. Samuel, X. Zhang, H. Wang, P. Fang, and G. Li, "A motion-classification strategy based on sEMG-EEG signal combination for upper-limb amputees," *J. Neuroeng. Rehabil.*, vol. 14, no. 1, p. 2, Jan. 2017.
- [244] X. Zhang, L. Yao, Q. Z. Sheng, S. S. Kanhere, T. Gu, and D. Zhang, "Converting Your Thoughts to Texts: Enabling Brain Typing via Deep Feature Learning of EEG Signals," in *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2018, pp. 1–10.
- [245] M. Mahmood *et al.*, "Fully portable and wireless universal brain–machine interfaces enabled by flexible scalp electronics and deep learning algorithm," *Nat. Mach. Intell.*, vol. 1, no. 9, pp. 412–422, 2019.
- [246] C.-T. Lin *et al.*, "A Wireless Multifunctional SSVEP-Based Brain–Computer Interface Assistive System," *IEEE Trans. Cogn. Dev. Syst.*, vol. 11, no. 3, pp. 375–383, 2018.
- [247] J. Xu and B. Zhong, "Review on portable EEG technology in educational research," *Comput. Human Behav.*, vol. 81, pp. 340–349, 2018.
- [248] X. Chen, B. Zhao, Y. Wang, S. Xu, and X. Gao, "Control of a 7-DOF robotic arm system with an SSVEP-based BCI," *Int. J. Neural Syst.*, vol. 28, no. 08, p. 1850018, 2018.
- [249] Y. M. Chi, Y. Wang, Y.-T. Wang, T.-P. Jung, T. Kerth, and Y. Cao, "A practical mobile dry EEG system for human computer interfaces," in *International Conference on Augmented Cognition*, 2013, pp. 649–655.
- [250] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.
- [251] X. Deng *et al.*, "Application of atomic force microscopy in cancer research," *J. Nanobiotechnology*, vol. 16, no. 1, p. 102, 2018.
- [252] S. V. Sharma, D. A. Haber, and J. Settleman, "Cell line-based platforms to evaluate the therapeutic efficacy of candidate anticancer agents," *Nat. Rev. Cancer*, vol. 10, no. 4, pp. 241–253, 2010.
- [253] J. R. W. Masters, "Cell line misidentification: The beginning of the end," *Nat. Rev. Cancer*, vol. 10, no. 6, pp. 441–448, 2010.
- [254] C. Korch *et al.*, "Authentication of M14 melanoma cell line proves misidentification of MDA-MB-435 breast cancer cell line," *Int. J. Cancer*, vol. 142, no. 3, pp. 561–572, 2018.
- [255] "STR Profiling for Cell Line Authentication \_ Biocompare\_ The Buyer's Guide for Life Scientists."
- [256] P. S. Lohar, *Textbook of Biotechnology*. MJP Publisher, 2019.
- [257] W. Parson *et al.*, "Cancer cell line identification by short tandem repeat profiling: Power and limitations," *FASEB J.*, vol. 19, pp. 434–436, Apr. 2005.
- [258] X. Bian, Z. Yang, H. Feng, H. Sun, and Y. Liu, "A Combination of Species Identification and STR Profiling Identifies Cross-contaminated Cells from 482 Human Tumor Cell Lines," *Sci. Rep.*, vol. 7, no. 1, pp. 1–10, 2017.
- [259] "Isogenic Cell Lines for Neuroscience & Cancer Research \_ ASC."
- [260] M. Michaelis, M. N. Wass, and J. Cinatl, "Drug-adapted cancer cell lines as preclinical models of acquired resistance," 2019.
- [261] R. L. Juliano and V. Ling, "A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants," *Biochim. Biophys. Acta - Biomembr.*, vol. 455, no. 1, pp. 152–162, 1976.
- [262] S. P. Cole *et al.*, "Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line," *Science (80-. )*, vol. 258, no. 5088, pp. 1650–1654, 1992.
- [263] J. A. Engelman *et al.*, "MET Amplification Leads to Gefitinib Resistance in Lung Cancer by

Activating ERBB3 Signaling,” *Science (80-. )*, vol. 316, no. 5827, pp. 1039–1043, 2007.

- [264] A. S. Crystal *et al.*, “Patient-derived models of acquired resistance can identify effective drug combinations for cancer,” *Science (80-. )*, vol. 346, no. 6216, pp. 1480–1486, 2014.
- [265] C. M. Bunney, P. E., Zink, A. N., Holm, A. A., Billington, C. J., & Kotz, “乳鼠心肌提取 HHS Public Access,” *Physiol. Behav.*, vol. 176, no. D1, pp. 139–148, 2017.
- [266] M. Michaelis *et al.*, “Adaptation of cancer cells from different entities to the MDM2 inhibitor nutlin-3 results in the emergence of p53-mutated multi-drug-resistant cancer cells,” *Cell Death Dis.*, vol. 2, no. 12, 2011.
- [267] A. Benali Amjoud and M. Amrouch, “Convolutional Neural Networks Backbones for Object Detection BT - Image and Signal Processing,” 2020, pp. 282–289.
- [268] W. Wang, Y. Li, T. Zou, X. Wang, J. You, and Y. Luo, “A Novel Image Classification Approach via Dense-MobileNet Models,” *Mob. Inf. Syst.*, vol. 2020, p. 7602384, 2020.
- [269] C. Szegedy *et al.*, “Going deeper with convolutions,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 1–9, 2015.
- [270] S. J. Kim *et al.*, “Deep transfer learning-based hologram classification for molecular diagnostics,” *Sci. Rep.*, vol. 8, no. 1, pp. 1–12, 2018.
- [271] S. Hira, A. Bai, and S. Hira, “An automatic approach based on CNN architecture to detect Covid-19 disease from chest X-ray images,” *Appl. Intell.*, 2020.
- [272] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, “A Dataset for Breast Cancer Histopathological Image Classification,” vol. 63, no. 7, pp. 1455–1462, 2016.
- [273] E. Saintas *et al.*, “Acquired resistance to oxaliplatin is not directly associated with increased resistance to DNA damage in SK-N-ASrOXALI4000, a newly established oxaliplatin-resistant sub-line of the neuroblastoma cell line SK-N-AS,” *PLoS One*, vol. 12, no. 2, pp. 1–17, 2017.
- [274] L. Perez and J. Wang, “The Effectiveness of Data Augmentation in Image Classification using Deep Learning,” 2017.
- [275] S. O. Gara, K. Mcguinness, S. O. Gara, and K. Mcguinness, “ARROW @ TU Dublin Comparing Data Augmentation Strategies for Deep Image Classification Comparing Data Augmentation Strategies for Deep Image Classification,” 2019.
- [276] T. Gomes, S. A. Teichmann, and C. Talavera-López, “Immunology Driven by Large-Scale Single-Cell Sequencing,” *Trends Immunol.*, vol. 40, no. 11, pp. 1011–1021, Nov. 2019.
- [277] J. D. Rodríguez, A. Pérez, and J. A. Lozano, “Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 569–575, 2010.
- [278] D. Ciregan, U. Meier, and J. Schmidhuber, “Multi-column deep neural networks for image classification,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 3642–3649, 2012.
- [279] H. Wang and H. Zheng, “True Positive Rate BT - Encyclopedia of Systems Biology,” W. Dubitzky, O. Wolkenhauer, K.-H. Cho, and H. Yokota, Eds. New York, NY: Springer New York, 2013, pp. 2302–2303.
- [280] C. Kanan and G. W. Cottrell, “Color-to-grayscale: Does the method matter in image recognition?,” *PLoS One*, vol. 7, no. 1, 2012.
- [281] S. H. S. Basha, S. R. Dubey, V. Pulabaigari, and S. Mukherjee, “Impact of Fully Connected Layers on Performance of Convolutional Neural Networks for Image Classification,” 2019.
- [282] H. Wang, K. Ren, and J. Song, “A closer look at batch size in mini-batch training of deep auto-encoders,” *2017 3rd IEEE Int. Conf. Comput. Commun. ICC3 2017*, vol. 2018-Janua, no. 1, pp. 2756–2761, 2018.
- [283] M. Derezhinski, D. Mahajan, S. S. Keerthi, S. V. N. Vishwanathan, and M. Weimer, “Batch-

expansion training: An efficient optimization framework,” *Int. Conf. Artif. Intell. Stat. AISTATS 2018*, pp. 736–744, 2018.

- [284] M. Takáč, A. Bijral, P. Richtárik, and N. Srebro, “Mini-batch primal and dual methods for SVMs,” *30th Int. Conf. Mach. Learn. ICML 2013*, no. PART 3, pp. 2059–2067, 2013.
- [285] P. Goyal *et al.*, “Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour,” 2017.
- [286] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [287] F. Yu, A. Seff, Y. Zhang, S. Song, T. Funkhouser, and J. Xiao, “LSUN: Construction of a Large-scale Image Dataset using Deep Learning with Humans in the Loop,” 2015.
- [288] S. Wan, T. Y. Wu, W. H. Wong, and C. Y. Lee, “Confnet: Predict with Confidence,” *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2018-April, pp. 2921–2925, 2018.
- [289] Y. Gal and Z. Ghahramani, “Dropout as a Bayesian Approximation: Appendix,” *33rd Int. Conf. Mach. Learn. ICML 2016*, vol. 3, pp. 1661–1680, 2016.
- [290] H. Li, Z. Lin, X. Shen, and J. Brandt, “Cascade Face Detection,” pp. 5325–5334, 2015.
- [291] “Why Automated Feature Engineering Will Change the Way You Do Machine Learning \_ by Will Koehrsen \_ Towards Data Science.” .
- [292] C.-W. Chen, Y.-H. Tsai, F.-R. Chang, and W.-C. Lin, “Ensemble feature selection in medical datasets: Combining filter, wrapper, and embedded feature selection results,” *Expert Syst.*, vol. n/a, no. n/a, p. e12553, Apr. 2020.
- [293] S. Optimization, “SS symmetry An Approach for Streaming Data Feature Extraction Based on Discrete Cosine Transform and Particle,” 2020.
- [294] L. R. Trambaiolli, N. Spolaôr, A. C. Lorena, R. Anghinah, and J. R. Sato, “Feature selection before EEG classification supports the diagnosis of Alzheimer’s disease,” *Clin. Neurophysiol.*, vol. 128, no. 10, pp. 2058–2067, 2017.
- [295] V. A. Maksimenko *et al.*, “Artificial Neural Network Classification of Motor-Related EEG: An Increase in Classification Accuracy by Reducing Signal Complexity,” *Complexity*, vol. 2018, p. 9385947, 2018.
- [296] Z. Tang, K. Zhang, C. Li, S. Sun, Q. Huang, and S. Zhang, “Motor Imagery Classification Based on Deep Convolutional Neural Network and Its Application in Exoskeleton Controlled by EEG,” *Chinese J. Comput.*, vol. 2016, pp. 1–15, Nov. 2016.
- [297] Z. Tang, C. Li, and S. Sun, “Single-trial EEG classification of motor imagery using deep convolutional neural networks,” *Optik (Stuttg.)*, vol. 130, pp. 11–18, 2017.
- [298] C. Chen *et al.*, “Deep Learning on Computational-Resource-Limited Platforms: A Survey,” *Mob. Inf. Syst.*, vol. 2020, p. 8454327, 2020.
- [299] L. Wang, G. Xu, S. Yang, M. Guo, W. Yan, and J. Wang, *Motor imagery BCI research based on sample entropy and SVM*. 2012.
- [300] C. Chen, C. Wu, C. Lin, and S. Chen, “A novel classification method for motor imagery based on Brain-Computer Interface,” in *2014 International Joint Conference on Neural Networks (IJCNN)*, 2014, pp. 4099–4102.
- [301] and G. L. [10] O.W. Samuel, Y. Geng, X. Li, “Towards Efficient Decoding of Multiple Classes of Motor Imagery Limb Movements Based on Eeg Spectral and Time Domain Descriptors,” *J. Med. Syst.*, vol. 41, no. 12, p. 194, 2017.
- [302] N. Isa, A. Amir, M. Ilyas, and M. Razalli, “The Performance Analysis of K-Nearest Neighbors (K-NN) Algorithm for Motor Imagery Classification Based on EEG Signal,” *MATEC Web Conf.*, vol. 140, p. 1024, Jan. 2017.
- [303] H. Jian and K. Tang, “Improving classification accuracy of SSVEP based BCI using RBF SVM



with signal quality evaluation,” in *2014 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, 2014, pp. 302–306.

- [304] “No Title,” 2015.
- [305] Z. İş and V. V Nikulin, “Steady state visual evoked potential ( SSVEP ) based brain-computer interface ( BCI ) performance under different perturbations,” pp. 1–17, 2018.
- [306] L. Automata, “Feature Selection for Motor Imagery EEG Classification Based on Firefly Algorithm and Learning Automata,” 2017.
- [307] W.-L. Mao, H. Fathurrahman, Y. Lee, and T. Chang, “EEG dataset classification using CNN method,” *J. Phys. Conf. Ser.*, vol. 1456, p. 12017, Jan. 2020.
- [308] S. Hwang, K. Hong, G. Son, and H. Byun, “Learning CNN features from DE features for EEG-based emotion recognition,” *Pattern Anal. Appl.*, vol. 23, no. 3, pp. 1323–1335, 2020.
- [309] A. S. Lundervold and A. Lundervold, “An overview of deep learning in medical imaging focusing on MRI,” *Z. Med. Phys.*, vol. 29, no. 2, pp. 102–127, 2019.
- [310] D. Shen, G. Wu, and H.-I. Suk, “Deep Learning in Medical Image Analysis,” *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.
- [311] D. Chen *et al.*, “Deep learning and alternative learning strategies for retrospective real-world clinical data,” *npj Digit. Med.*, vol. 2, no. 1, p. 43, 2019.
- [312] B. Lutnick *et al.*, “Iterative annotation to ease neural network training : Specialized machine learning in medical image analysis,” pp. 1–15.
- [313] M. Ghassemi, T. Naumann, P. Schulam, L. Andrew, I. Y. Chen, and R. Ranganath, “A Review of Challenges and Opportunities in Machine Learning for Health University of Toronto and Vector Institute , Toronto , Canada ; 2 Microsoft Research .,”
- [314] C. S. Perone and J. Cohen-adad, “Promises and limitations of deep learning for medical image segmentation,” pp. 1–2, 2019.
- [315] R. S. Mitchell, J. G. Michalski, and T. M. Carbonell, *An artificial intelligence approach*. Springer, 2013.
- [316] R. Goyenka, “A Study on Low Level Features and High Level Features in CBIR,” pp. 13473–13482, 2017.
- [317] R. Roy, A. Roy, A. Konar, and A. Nagar, *Design of a Computationally Economical Image Classifier using Generic Features*. 2019.
- [318] M. Tangermann *et al.*, “Review of the BCI competition IV,” *Front. Neurosci.*, vol. 6, p. 55, 2012.
- [319] F. Fahimi, Z. Zhang, W. B. Goh, T.-S. Lee, K. K. Ang, and C. Guan, “Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI,” *J. Neural Eng.*, vol. 16, no. 2, p. 26007, 2019.
- [320] H. He and D. Wu, “Transfer learning for Brain--Computer interfaces: A Euclidean space data alignment approach,” *IEEE Trans. Biomed. Eng.*, vol. 67, no. 2, pp. 399–410, 2019.
- [321] E. E. G. Analysis and R. Mill, “Applying Transfer Learning To Deep Learned Models For,” no. DL, pp. 1–16.
- [322] S. Gilang and P. Putra, “SME Credit Scoring Using Social Media Data.”
- [323] M. Kaya, M. K. Binli, E. Ozbay, H. Yanar, and Y. Mishchenko, “A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces,” *Sci. Data*, vol. 5, no. 1, p. 180211, 2018.
- [324] J. Deng, A. C. Berg, K. Li, and L. Fei-fei, “What Does Classifying More Than 10 , 000 Image Categories Tell Us ?,” pp. 71–84, 2010.
- [325] L. Jin and H. Liang, “Deep learning for underwater image recognition in small sample size situations,” *Ocean. 2017 - Aberdeen*, vol. 2017-October, no. 61379007, pp. 1–4, 2017.

- [326] S. C. Wong and M. D. McDonnell, “Understanding data augmentation for classification : when to warp ?,” 2016.
- [327] A. Kendall and Y. Gal, “What uncertainties do we need in bayesian deep learning for computer vision?,” in *Advances in neural information processing systems*, 2017, pp. 5574–5584.
- [328] A. J. Ratner, H. Ehrenberg, Z. Hussain, J. Dunnmon, and C. Ré, “Learning to compose domain-specific transformations for data augmentation,” in *Advances in neural information processing systems*, 2017, pp. 3236–3246.
- [329] S. Ioffe and C. Szegedy, “Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift.”
- [330] A. Le Guennec, S. Malinowski, and R. Tavenard, “Data Augmentation for Time Series Classification using Convolutional Neural Networks,” 2016.
- [331] Z. Zhang *et al.*, “A Novel Deep Learning Approach with Data Augmentation to Classify Motor Imagery Signals,” *IEEE Access*, vol. PP, p. 1, Jan. 2019.
- [332] M. M. Krell and S. K. Kim, “Rotational data augmentation for electroencephalographic data,” in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 471–474.
- [333] Q. Wen, L. Sun, X. Song, J. Gao, X. Wang, and H. Xu, *Time Series Data Augmentation for Deep Learning: A Survey*. 2020.
- [334] K. Chaitanya, N. Karani, C. F. Baumgartner, A. Becker, O. Donati, and E. Konukoglu, “Semi-supervised and task-driven data augmentation,” in *International conference on information processing in medical imaging*, 2019, pp. 29–41.
- [335] E. D. Cubuk, B. Zoph, D. Man, V. Vasudevan, and Q. V Le, “Learning Augmentation Strategies from Data,” no. Section 3, 2012.
- [336] W. G. Dirks *et al.*, “Cell line cross-contamination initiative: an interactive reference database of STR profiles covering common cancer cell lines,” *Int. J. cancer*, vol. 126, no. 1, pp. 303–304, 2010.
- [337] J. L. Almeida, K. D. Cole, and A. L. Plant, “Standards for Cell Line Authentication and Beyond,” pp. 1–9, 2016.
- [338] A. Capes-davis *et al.*, “Match criteria for human cell line authentication : Where do we draw the line ?,” vol. 2519, pp. 2510–2519, 2013.
- [339] H. Bai *et al.*, “Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT,” *Radiology*, vol. 296, p. 200823, Mar. 2020.
- [340] P. R. A. S. Bassi and R. Attux, “A Deep Convolutional Neural Network for COVID-19 Detection Using Chest X-Rays.”
- [341] N. Friedman, T. Fekete, K. Gal, and O. Shriki, “EEG-Based Prediction of Cognitive Load in Intelligence Tests,” *Front. Hum. Neurosci.*, vol. 13, p. 191, 2019.
- [342] S. M. U. Saeed, S. M. Anwar, H. Khalid, M. Majid, and A. U. Bagci, “EEG based Classification of Long-term Stress Using Psychological Labeling,” *Sensors (Basel)*, vol. 20, no. 7, p. 1886, Mar. 2020.
- [343] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, “EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation,” *Sci. World J.*, vol. 2014, p. 627892, 2014.
- [344] “Drug development\_ the journey of a medicine from lab to shelf \_ Career Feature \_ Pharmaceutical Journal.” .
- [345] S. Cardoso, L. Valverde, A. Odriozola, X. Elcoroaristizabal, and M. M. de Pancorbo, “Quality standards in Biobanking: authentication by genetic profiling of blood spots from donor’s original sample,” *Eur. J. Hum. Genet.*, vol. 18, no. 7, pp. 848–851, 2010.

- [346] Y. Huang, Y. Liu, C. Zheng, and C. Shen, “Investigation of Cross-Contamination and Misidentification of 278 Widely Used Tumor Cell Lines,” *PLoS One*, vol. 12, no. 1, p. e0170384, Jan. 2017.
- [347] A. Kaur and Y. Guan, “Phantom limb pain: A literature review,” *Chinese J. Traumatol. = Zhonghua chuang shang za zhi*, vol. 21, no. 6, pp. 366–368, Dec. 2018.
- [348] N. B. Bahadure, A. K. Ray, and H. P. Thethi, “Image Analysis for MRI Based Brain Tumor Detection and Feature Extraction Using Biologically Inspired BWT and SVM,” *Int. J. Biomed. Imaging*, vol. 2017, p. 9749108, 2017.
- [349] S. Noachtar and J. Rémi, “The role of EEG in epilepsy: a critical review,” *Epilepsy Behav.*, vol. 15, no. 1, pp. 22–33, 2009.
- [350] E. I. Martin, K. J. Ressler, E. Binder, and C. B. Nemeroff, “The neurobiology of anxiety disorders: brain imaging, genetics, and psychoneuroendocrinology,” *Psychiatr. Clin. North Am.*, vol. 32, no. 3, pp. 549–575, Sep. 2009.
- [351] M. Thiyagarajan, “BRAIN TUMOUR DETECTION VIA EEG SIGNALS,” *Indian J. Appl. Res.*, vol. 9, pp. 213–215, Jan. 2019.
- [352] F. Buchthal, O. Svensmark, and H. Simonsen, “Relation of EEG and seizures to phenobarbital in serum,” *Arch. Neurol.*, vol. 19, no. 6, pp. 567–572, 1968.
- [353] E. Y. Kimchi and S. S. Cash, “Seizures at the scale of individual neurons,” *Brain*, vol. 138, no. 10, pp. 2807–2808, Sep. 2015.
- [354] H. Yang, W. Lee, and H. Lee, “IoT Smart Home Adoption: The Importance of Proper Level Automation,” *J. Sensors*, vol. 2018, p. 6464036, 2018.
- [355] N. Kosmyrna, F. Tarpin-Bernard, N. Bonnefond, and B. Rivet, “Feasibility of BCI Control in a Realistic Smart Home Environment,” *Front. Hum. Neurosci.*, vol. 10, p. 416, Aug. 2016.
- [356] M. J. Rodrigues, O. Postolache, and F. Cercas, “Physiological and Behavior Monitoring Systems for Smart Healthcare Environments : A Review,” pp. 1–26, 2020.
- [357] “Cancer in My Community\_ The Barriers to Cancer Care in Tanzania \_ Cancer.” .
- [358] C. A. Chao, L. Huang, K. Visvanathan, K. Mwakatobe, N. Masalu, and A. F. Rositch, “Understanding women’s perspectives on breast cancer is essential for cancer control: knowledge, risk awareness, and care-seeking in Mwanza, Tanzania,” *BMC Public Health*, vol. 20, no. 1, p. 930, 2020.
- [359] E. Detection, “TANZANIA BREAST HEALTH CARE ASSESSMENT,” 2017.
- [360] R. Xu *et al.*, “How Physical Activities Affect Mental Fatigue Based on EEG Energy, Connectivity, and Complexity,” *Front. Neurol.*, vol. 9, p. 915, Oct. 2018.
- [361] M.-H. Lee, J. Williamson, Y.-E. Lee, and S.-W. Lee, “Mental fatigue in central-field and peripheral-field steady-state visually evoked potential and its effects on event-related potential responses,” *Neuroreport*, vol. 29, no. 15, pp. 1301–1308, Oct. 2018.
- [362] J.-P. Gillet, S. Varma, and M. M. Gottesman, “The clinical relevance of cancer cell lines.,” *J. Natl. Cancer Inst.*, vol. 105, no. 7, pp. 452–458, Apr. 2013.