PREDICTING EPILEPTIC SEIZURES OF PEDIATRIC PATIENTS USING PHASE-AMPLITUDE COUPLING AND DEEP LEARNING

by

Lina Mahmoud Saleh Alqam

B.S., in Mechatronics Engineering, Birzeit Univerity, 2018

Submitted to the Institute of Biomedical Engineering
in partial fulfillment of the requirements
for the degree of
Master of Science
in
Biomedical Engineering

Boğaziçi University
2022

PREDICTING EPILEPTIC SEIZURES OF PEDIATRIC PATIENTS USING PHASE-AMPLITUDE COUPLING AND DEEP LEARNING

APPROVED BY:

Prof. Dr. Ahmet Ademoğlu .....
(Thesis Advisor)

Assoc. Prof. Dr. Esin Öztürk Işık .....

Assist. Prof. Dr. İpek Şen .....

DATE OF APPROVAL: 5 October 2022

ACKNOWLEDGMENTS

In the very first place, I want to express my gratitude to the Almighty God, Allah. Your blessings shone down upon me at every turn of this journey. All thanks that You chose this path and this place for me.

There is a widespread belief that if you study with Al-Muallim, you will undergo a personal metamorphosis and gain new insights. I would like to thank my advisor, Prof. Dr. Ahmet Ademoglu for being my mentor. What is entropy? I said: messy energy. I am so thankful that I took his tough courses, from Signal Processing and Neuroimaging to Biomedical Instrumentation. Overall, I liked the philosophy lectures, coffee breaks, and hours of writing on the board.

I would like to thank my bestie, my father, Mahmoud. Despite our differences, our minds are remarkably similar, thanks to Granny and Grandpa. Thanks to my mother, Najia, my adorable twin. She is the one who continues to believe in my ability to do strange things. Sister Reem, and the little Maria whom without her talent non of my ideas would come out. Ibrahim and Baibers my super bro gang. I would like to thank my friend Nadeen Zaid for our weekly Friday video calls.

I'd like to express my gratitude to the people of Istanbul, who quickly became like a second family to me. Let there be movies and bedtime tales, Gozde Tokmak. Poems by Janat Ilyas 'O Yaar, my accomplice in all our crimes. Come, I'll show you my latest violin licks, Dilek Yildirm, and her perfect pasta. Manti, coffee, burgers, and homemade lemon blueberry cake are part of our girly nights, Merve Gul Kantarci. As well as everyone who supported me and encouraged me to keep going even when times were tough.

This is where I say thanks to Huden Nese and Ahmed Alramly, two friends from the Neurosignal Laboratory who always had my back. Here at our lab, we joined a

variety of summer courses and workshops. Chocolate, coffee with cardamom, pizza and deep conversations.

Shushi the fish, Felfel the kitten, all the biomedical dogs with ADHD who are still barking at me; thanks for creating stories.

ACADEMIC ETHICS AND INTEGRITY STATEMENT

I, Lina Mahmoud Saleh Alqam, hereby certify that I am aware of the Academic Ethics and Integrity Policy issued by the Council of Higher Education (YÖK) and I fully acknowledge all the consequences due to its violation by plagiarism or any other way.

Name :

Signature:

Date:

ABSTRACT

PREDICTING EPILEPTIC SEIZURES OF PEDIATRIC PATIENTS USING PHASE-AMPLITUDE COUPLING AND DEEP LEARNING

Epilepsy is one of the most prevalent neurological disorders in the world. It is estimated that between four and six children out of every thousand suffer from epilepsy. Visually diagnosing seizure activity can be challenging for medical professionals because doing so correctly requires a great deal of observation and practice. Recent developments in deep learning technology have made it so that erroneous diagnoses of seizure activity are occurring much less frequently and that significantly less time is needed to forecast when a seizure will occur. This study improves upon existing metrics by utilizing cross-frequency coupling gleaned from scalp electroencephalogram (EEG) and transfer learning inside deep neural network architectures. We conducted our study using data from the CHB-MIT dataset, which included 23 children as participants. In the first experiment, we culled all 18 channels from the dataset; in the second, we narrowed it down to just one. Fine-tuning boosted the accuracy of the first experiment to 87.21%, and the accuracy of the second experiment to 96.82%. Images of the phase-amplitude coupling help predict the next seizure when fed into the VGG-19 architecture. Several tests are used to assess the models' efficacy. In-depth analyses are performed and reported on.

Keywords: Scalp EEG, Pediatric Epileptic Seizures, Phase-Amplitude Coupling, Prediction, Transfer Learning.

ÖZET

PEDİATRİK HASTALARIN EPİLEPTİK NÖBETLERİNİ FAZ-GENLİK BAĞLANIM VE DERİN ÖĞRENME KULLANARAK TAHMİN ETME

Epilepsi, dünyadaki en yaygın nörolojik bozukluklardan biridir. Her bin çocuktan dördü ila altısı arasında epilepsi hastası olduğu tahmin edilmektedir. Nöbet aktivitesini görsel olarak teşhis etmek tıp uzmanları için zor olabilir çünkü bunu doğru bir şekilde yapmak çok fazla gözlem ve uygulama gerektirir. Derin öğrenme teknolojisindeki son gelişmeler, nöbet aktivitesine ilişkin hatalı teşhislerin çok daha az sıklıkta ortaya çıkmasını ve bir nöbetin ne zaman gerçekleşeceğini tahmin etmek için önemli ölçüde daha az zamana ihtiyaç duyulmasını sağlamıştır. Bu çalışma, kafa derisi Elektroensefalogramından (EEG) toplanan çapraz frekans bağlanımını ve derin sinir ağı mimarileri içinde transfer öğrenmeyi kullanarak mevcut ölçümleri geliştirir. Çalışmamızı, katılımcı olarak 23 çocuğu içeren CHB-MIT veri setinden elde edilen verileri kullanarak gerçekleştirdik. İlk deneyde, veri setinden 18 kanalın hepsini seçtik; ikincisinde, onu sadece bire indirdik. İnce ayar, ilk deneyin doğruluğunu %87.21'e ve ikinci deneyin doğruluğunu %96.82'ye yükseltti. Faz-genlik bağlanımının görüntüleri, VGG-19 mimarisine beslendiğinde bir sonraki nöbetin tahmin edilmesine yardımcı oldu. Modellerin etkinliğini değerlendirmek için çeşitli testler kullanıldı. Derinlemesine analizler yapıldı ve raporlandı.

Anahtar Sözcükler: EEG, Pediatrik Epileptik Nöbetler, Faz-Genlik Bağlanımı, Tahmin, Transfer Öğrenme.

TABLE OF CONTENTS

ACKNOWLEDGMENTS iii
ACADEMIC ETHICS AND INTEGRITY STATEMENT v
ABSTRACT vi
ÖZET vii
LIST OF FIGURES x
LIST OF TABLES xiii
LIST OF SYMBOLS xiv
LIST OF ABBREVIATIONS xv
1. INTRODUCTION 1
1.1 Research Motivation 1
1.2 Research Goals 2
1.3 Research Outline 3
2. BACKGROUND 4
2.1 Scalp EEG 4
2.2 Epilepsy 5
2.2.1 Classical Definition and History 5
2.2.2 Clinical Definition 8
3. SIGNAL PROCESSING 10
3.1 Bioelectrical Signals 10
3.2 Wavelets 12
3.2.1 Continuous Wavelet Transform 14
3.2.2 Complex Morlet Transform 15
4. METHODOLOGY 17
4.1 Pediatric Dataset 17
4.2 Filtering EEG 18
4.3 Segmentation of EEG 21
4.4 Cross Frequency Coupling 22
4.5 Phase Amplitude Coupling 23
4.6 PAC Steps 25

4.6.1 Illustrated Algorithm 27
4.7 Thresholding 30
5. VGG-19 AND TRANSFER LEARNING 31
5.1 Transfer Learning 31
5.2 CNN Terminology 32
5.3 VGG-19 34
5.3.1 Forward Propagation for VGG-19 35
5.3.2 Backpropagation for VGG-19 38
5.4 Structuring The Model 39
5.5 Parameters for VGG-19 and Transfer Learning 41
5.5.1 Trainable and Frozen Layers: All Channels 42
5.5.2 Fine Tuning: One Channel Variance Based 43
6. EXPERIMENTAL RESULTS 45
6.1 Temporal and Spatial information 45
6.2 Delta-Gamma Coupling Images 53
6.3 VGG-19 Training and Testing 57
6.4 VGG-19 Results 57
7. DISCUSSION OF RESULTS 60
7.1 Phase Amplitude Coupling Images as Epilepsy Biomarker 60
7.2 Comparison between Literature and our Proposed Approach 61
7.3 VGG-19 and Transfer Learning Outcome 62
8. CONCLUSION 64
REFERENCES 65

LIST OF FIGURES

Figure 2.1	A human brain in its normal state, an epileptic brain with focal seizure in the temporal lobe, and an epileptic brain with generalized epilepsy.	8
Figure 3.1	Configuration of an observed EEG signal including biological artifacts.	11
Figure 3.2	Wavelets with different scales.	12
Figure 3.3	Different Heisenberg boxes in time- frequency axis, the top brown box represents low frequency resolution and higher temporal resolution opposite to the green box which boasted spatial resolution and degraded the temporal resolution.	13
Figure 4.1	Configuration of EEG and pediatric patients.	18
Figure 4.2	Bipolar longitudinal montage employs a 10-20 system for electrode placement.	19
Figure 4.3	EEG recording for Channel FP1-F7 without using filter.	20
Figure 4.4	EEG recording for Channel FP1-F7 after using Butterworth filter.	20
Figure 4.5	Schematic figure of slicing out the ictal period while maintaining the other periods.	22
Figure 4.6	EEG signal segmented along the time axis.	22
Figure 4.7	Schematic figure of Comodulogram between low and high frequencies, the grey tile shows the expected region where the activity ought to appear.	26
Figure 4.8	Pediatric patient where amplitude and phase are extracted and created a coupling.	27
Figure 4.9	Steps regrading phase amplitude coupling images.	28
Figure 4.10	Explaining the third step.	28
Figure 4.11	The following steps in computing the comodulation starting from normalization, calculating Shannon entropy to evaluate the phase amplitude indices .	29
Figure 5.1	Transfer learning method	32

Figure 5.2 The VGG-19 network's structural layout. There are a total of 16 convolutions in the network, all linked together using ReLUs, and 5 maxpooling layers. The convolutions have filter maps with a starting value of 64 and a maximum value of 512. After the convolutions, a linear classifier is constructed, with three fully-connected (fc) layers, the first two of which contain 4096 features and the third of which contains 1000. Each value is then mapped by the softmax layer, which is connected to the final fc layer, to the probability of belonging to one of the ImageNet competition's thousand classes. 34

Figure 5.3 VGG-19 convolutional layers order and configurations [72]. 35

Figure 5.4 Flowchart for using VGG-19 and Transfer Learning. 40

Figure 5.5 Database file that contains pixel images file and inside there are two files related to the training and the validation images of preictal and interical. 41

Figure 5.6 The procedure of Transfer Learning with VGG-19, the upper architecture describes the full model of VGG-19 while the lower architecture is the proposed one used in the study with the difference in the softmax layer where we use two classes instead of 1000 classes. The weights were imported from the original VGG-19 model using Transfer Learning. 43

Figure 5.7 The procedure of Transfer Learning with VGG-19, the upper architecture describes the full model of VGG-19 while the lower architecture is the proposed one used in the study with the difference in the softmax layer where we use two classes instead of 1000 classes. The weights were imported from the original VGG-19 model using Transfer Learning. 44

Figure 6.1 Changing in spatial information during three states, interictal, preictal, and ictal respectively. 45

Figure 6.2 Patient *Cb01*'s preictal period is shown in white, and the ictal period is shown in blue on their EEG recording, which spans 30 minutes. 46

Figure 6.3	Patient *Cb01* as explained from the summery, seizure start time: 1467 seconds and seizure end time: 1494 seconds.	47
Figure 6.4	Session number 29 for patient *Cb18* in which there is no seizure activity within a 10-second window.	48
Figure 6.5	Session number 29 for patient *Cb18* in which there is seizure activity within a 10-second window.	49
Figure 6.6	Ictal period for the 18 channels used in this study.	50
Figure 6.7	Interictal period for the 18 channels used in this study.	51
Figure 6.8	Preictal period for the 18 channels used in this study.	52
Figure 6.9	Seizure prediction for one channel, clarifying the false prediction and true prediction.	53
Figure 6.10	Comodulation image of ictal phase.	54
Figure 6.11	Comodulation image of ictal phase.	54
Figure 6.12	Comodulation image of preictal phase.	55
Figure 6.13	Comodulation image of preictal phase	55
Figure 6.14	Comodulation image of peictal phase.	56
Figure 6.15	Comodulation image of interictal phase.	56
Figure 6.16	Comparison of the accuracy between spatial averaging for the channels with/without fine tuning and the accuracy of one channel with/without fine tuning.	57
Figure 6.17	Comparison of the sensitivity between spatial averaging for the channels with/without fine tuning and the sensitivity of one channel with/without fine tuning.	58
Figure 6.18	Comparison of the specificity between spatial averaging for the channels with/without fine tuning and the specificity of one channel with/without fine tuning.	58

LIST OF TABLES

Table 4.1	Information about CHB-MIT pediatric patient, their gender, age, the number of seizures an the number of sessions.	21
Table 4.2	Abnormalities related to coupling with gamma wave and its corresponding condition with the references	27
Table 6.1	Accuracy, sensitivity, specificity, and false positive rate, f1 score values for different patients	59
Table 7.1	The average metrics for using VGG-19 and Transfer Learning.	62
Table 7.2	The average metrics for using VGG-19, Transfer learning and Fine Tuning.	62

LIST OF SYMBOLS

Ψ	Mother Wavelet
ψ	Daughter Wavelet
ψ_0	Complex Morlet Transform
Φ_{fp}	Phase of Low Frequency
ω	Frequency
τ	Translation Parameter
$\mathbf{a}^{[\mathbf{i}]}$	Convolution Layer number i
A_{fA}	Amplitude of High Frequency
$\mathbf{b}^{[\mathbf{i}]}$	Bias for Layer number i
f_L	Low frequency
f_H	High frequency
g	ReLU Function
$\mathbf{m}^{[\mathbf{i}]}$	Pooling Layer number i
$\mathcal{L}$	Loss Function
Q	Entropy
s	scaling factor
W	Continuous Wavelet Transform
$\mathbf{W}^{[\mathbf{i}]}$	Weight for Layer number i
x_b	Probability of the Phase across the number of b bins
$\mathbf{x}$	Tensor Input
$\hat{\mathbf{y}}$	Tensor Output
$\mathbf{z}$	Convolution Function

LIST OF ABBREVIATIONS

EEG	Electroencephalogram
SPS	Seizure Prediction System
PAC	Phase Amplitude Coupling
VGG	Visual Geometry Group
TF	Transfer Learning
FP	False Positive
FPR	False Positive Rate
TN	True Negative
TP	True Positive
FN	False Negative
AED	Anti-epileptic Drug
FIR	Finite Impulse Response
IIR	Infinite Impulse Response
FT	Fourier Transfer
DC	Direct Current
LFO	Low Frequency Oscillation
HFO	High Frequency Oscillation
CFC	Cross Frequency Coupling
CWT	Continuous Wavelet Transform
LF	Low Frequency
HF	High Frequency
KL	Kullback-Leibler
AI	Artificial Intelligence
STFT	Short Time Fourier Transform

1. INTRODUCTION

1.1 Research Motivation

Epilepsy is a chronic brain disease characterized by repeated unprovoked seizures, with various manifestations: changes in consciousness, loss of spatial orientation and/or uncontrolled movements, sensitivity, and mental disorders [1–4]. Epilepsy is detected by using multiple modalities such as Electroencephalogram (EEG). The chaotic and recurrent seizures are one of the most common neurological disorders that cause intellectual deficits, therefore, an accurate automated Seizure Prediction System (SPS) is a requirement for recognizing the seizure onset time before the actual seizure. However, current state-of-art algorithms for neurological disorder prediction do not consider the impact of the high prediction accuracy and detection latency of seizures. Therefore, our new significant study will design a commercial deep-learning prediction system for forecasting seizure onset accurately and validate the use of frequency coupling biomarkers. Therefore, estimating the upcoming seizures is predominantly considered the most challenging predictive problem of neurological disorders with chaotic EEG patterns [5, 6]. From there, predicting seizures was diagnosed via some wavelet transforms for feature extraction [7–10]. How to extract high-level features from chaotic EEG signals, store and predict seizures within a proper time period is not investigated yet. Thus, our contribution here is expected to be a completed state-of-art deep learning architecture for seizure prediction systems. This contribution is significant because it is the helm of epileptic prediction research that is expected to lead to the development of the epilepsy management field. Once these management strategies become available, there is the promise of improving the quality of life for epileptic patients.

1.2 Research Goals

Developing a Seizure Prediction System (SPS) with phase-amplitude coupling (PAC) indices with deep-learning architecture by analyzing EEG signals of pediatric patients with epilepsy.

1. Determine whether PAC will be able to extract the features of epileptic EEG signals. In some previous epileptic studies, the cross-frequency coupling method has been used to extract the features of sub-bands of EEG data. Therefore, we will use another biomarker from the same family: phase-amplitude frequency coupling PAC.

   Expected Results, Interpretation, Possible Pitfalls: We expect that PAC indices will serve as feature extraction for a 10-seconds window for EEG recordings with seizures. Also, we expect that some regions will have a higher delta-gamma frequency which will be stored in a memory for the upcoming learning process entitled seizure approach signature. Interpretation of the result suggests that each extraction will correspond to the input pre-seizure epileptic states. For real-time computation, PAC will be more expensive compared to classical methods.

2. Determine whether a state-of-art deep learning network will predict the future seizure onset precisely. In this work, we will use VGG-19 with an input of pixel images (high/low frequency). For prediction, we will design a unique Transfer Learning algorithm that will install VGG-19 inside its architecture for testing the subjects after the training session. We expect to see an improvement among the results of the confusion matrix that beholds sensitivity, specificity, accuracy, and precision.

   Expected Results, Interpretation, Possible Pitfalls: We expect smaller windows 10- seconds will show an improvement in the seizure prediction system. Interpretation will aid in the classification of some states before predicting the correct state and to avoid a wide/narrow time range for detection latency.

1.3 Research Outline

This study contains 7 chapters with the introduction:

Chapter 2 presents a background overview of the Electroencephalogram (EEG) signal's characteristics and applications. Furthermore, it elucidates the development of epilepsy diagnosis from its earliest explanations to the advances in EEG.

Chapter 3 showcases the fundamental concepts of signal processing, and details about the pre-processing stage's use of a Butterworth filter, for example, are provided here. The significance of wavelets in spatio-temporal resolution is then discussed, along with their application to the analysis of EEG signals.

Chapter 4 explains the steps for working with EEG recordings. The algorithmic procedures for extracting the phases and amplitudes, as well as the idea of cross-frequency coupling and the primary method of phase-amplitude coupling, are also described. In addition, it demonstrates how the continuous EEG recordings of pediatric CHB-MIT patients were segmented.

Chapter 5 explores the deep learning approach for seizure prediction. Both transfer learning and the VGG-19 model were incorporated into this research. It describes the steps that need to be taken in order to structure the model, as well as its significance in the field of biomedical engineering.

Chapter 6 highlights the results obtained by using phase-amplitude coupling images and feeding them to VGG-19 model via transfer learning.

Chapter 7 discusses the experimental results from PAC images as biomarkers for epilepsy to the results of VGG-19 with/without fine tuning.

Chapter 8 is the final part that carries the conclusion which additionally leads to future views using one channel for recording EEG.

2. BACKGROUND

This chapter presents essential information regarding the main topics in the research. Section 2.1 gives a concise summery of the features and applications of the Electroencephalogram signal in neurophysiology. Section 2.2 examines the history of epilepsy diagnosis, from early interpretations such as demonic possession or a divine curse through the limitations of scientific perspectives in different eras and the relevance of EEG in our contemporary knowledge of the illness. The final section gives a brief overview of the medical definition of epilepsy.

2.1 Scalp EEG

> *"The EEG pattern is never simple or singular. We cannot extract one independent variable in the classical manner; we have to deal with the interaction of many unknowns and variables, the time. We have to deal with the rhythms algebraically."*
>
> —Grey Walter, The Living Brain (1953)

The first EEG devices, known as analog EEGs, relied on paper to record patients' continuous brain activity. This was done prior to the advent of quantitative EEG analysis [11]. The advancements in computing technology have made it possible to sample EEG signals and store them as information in numerical form in computers. Numerous computational analysis methods for EEG is made possible by the fact that we are dealing with digital data [11]. For EEG recording, at least two electrodes are needed because the technique records the potential difference between two distinct locations on the scalp [12]. Modern EEG recorders can take readings from a variety of scalp locations simultaneously. Electroencephalograms may be recorded employing either a bipolar or unipolar technique [13]. The bipolar technique involves pairing two coils or

electrodes and monitoring the resulting voltage differentials [13]. By using a reference electrode and one or more working electrodes, the unipolar (or monopolar) technique can measure and record the potential differences between the electrodes. The potential differences between two electrodes are measured and recorded using the unipolar (or monopolar) method, which requires a reference electrode and one or more working electrodes. Unipolar recordings allow the reference electrode to be placed anywhere in theory; however, average reference is typically used because the voltage differential on the scalp surface fluctuates depending on where the reference electrode is placed. In order to quickly and easily compare each electrode's potential to the average, it is sufficient to subtract the average of all electrodes' potential differences from each electrode's potential difference [14]. Topography, a spatial representation of potential distributions across the scalp, is greatly enhanced by normative reference [14].

2.2 Epilepsy

Epilepsy is a chronic brain disease characterized by repeated unprovoked seizures, with various manifestations: changes in consciousness, loss of spatial orientation and/or uncontrolled movements, sensitivity, and mental disorders [15, 16]. Around 70 million people are diagnosed with epilepsy, which is detected by using multiple modalities such as electroencephalogram EEG [17].

2.2.1 Classical Definition and History

The word seizure derives from the Greek verb epilambanein, which means "to take hold" or "attack" [4]. The Mesopotamian moon god (Sumerian: Nanna) is blamed in ancient records for causing a disease known as *the falling disease*, which was also seen as divine retribution [18]. In Sumerian tablets, it was described around 2500 BC that an individual with a twisted neck, wide-open eyes with tense peripheral parts of the body (arms and legs), foaming at the mouth, and a general state of confusion. This description is currently referred to as a focal tonic seizure [19]. Based on the behavioral

manifestations of a seizure, Caelius Aurelianus deduced that epilepsy could come from the brain (idiopathic) or from other parts of the body (sympathetic) and then spread to the brain. He coined the term *aura* from Greek to describe an epileptic patient who felt moving cold breezes around his extremities [20].

There was a substantial regression in the scientific community's understanding of epilepsy during the Middle Ages, and the notion of demonic possession once more predominated the public discourse. This era's literature makes use of phrases like "falling evil" and related terms for evil spirits [20]. Due to the lack of specificity in the description of seizures during this time period, people often confused epilepsy with other mental disorders. Ibnu-Sina, also known as Avicenna, was a Muslim physician who lived during the Golden Age of Islam. During this time period in the Middle East, he made contributions to the field of epilepsy by writing the his Medicine. He theorized that the location from which a seizure emanated (the brain, the stomach, the spleen, or the whole body) could be deduced from its clinical presentation. He described seizures manifesting as changes in temperament ("to a state resembling melancholy" and distraction accompanied by violent reactions) and psycho-neurological dysfunctions, such as intense anxiety and excitement, preceding seizures [19].

As a result of the new questions and evidence that emerged during the Renaissance, the scope of medical understanding was greatly expanded. Vertigo's [21] link to epilepsy was also called into question; doctors often struggled to tell the two conditions apart. Even more so in cases where the epilepsy was thought to have begun in the uterus, it was often difficult to tell the difference between hysterical and epileptic attacks [22]. It used to be thought that in order for a person to have a seizure, their entire body had to tremble, and they even had to fall to the ground. These days, we know that neither of these things is required. William Cullen [23] provided an accurate description of the fact that one need not lose consciousness in order to be affected by convulsions, as they can merely be localized to a specific area of the body. Gerard Swieten [24] provided a detailed explanation of the clinical characteristics of various types of seizures, including screams and salivation, while also attempting to explain why these symptoms occur. It is believed that Samuel Tissot penned the first modern

book on epilepsy. In his book, Tissot [25] introduced new terms such as outburst and eruption.

By the end of the 19th century, public opinion had shifted enough that people with epilepsy were no longer locked up alongside criminals and the insane. Consequently, medical professionals were able to keep a close eye on them and coin a new vocabulary to describe the effects of these seizures. Louis Calmeil [26] is credited with being the first to identify status epilepticus, the medical term for a patient experiencing continuous convulsions. At last, William James West [27] provided an in-depth account of his son's bouts with infantile spasms.

During John Hughlings Jackson's lifetime, neurology and epilepsy emerged as distinct medical specialties [28]. When making his historical mark, Jackson synthesized the insights of many predecessors with his own unique perspective. Jackson backed up and cited statements made by Theodore Herpin in his monograph on incomplete attacks [29]. Hemiplegic patients were the focus of Louis Bravais' research into a previously unknown form of epilepsy. Unlike those with generalized epilepsy, Hemiplegic patients have involuntary movements on the either left or right part of the body, typically the patient's dominant side, following that a paralysis of that limb [29]. However, he stopped short of providing analysis, and it wasn't until Jackson's insights that his work was widely hailed. Epileptic hemiplegia is the term that Robert Todd coined to describe what is now known as Todd's paralysis, a condition that is similar to that described by Prichard and Bravais [30]. Following in the footsteps of his predecessors, Jackson investigated unilateral seizure disorders like epilepsy. His terminology varied depending on his speculation about the epileptic brain region of origin. His description of the progression of seizures from one area to another (Jacksonian epilepsy) anticipated the concept of focal aware motor seizures. Jackson also provided extensive coverage of seizures, describing a surrealistic mode [30].

Clinical descriptions of epilepsy have improved substantially with the invention of electroencephalography (EEG) in the second half of the nineteenth century and the first decade of the twentieth century [31]. The discovery of two separate EEG patterns

lends additional credence to the concept that focal epilepsy is distinct from generalized epilepsy, as shown in Figure 2.1. Using EEG patterns, Frederic Andrews Gibbs and his coworkers identified the three main types of clinical seizures.

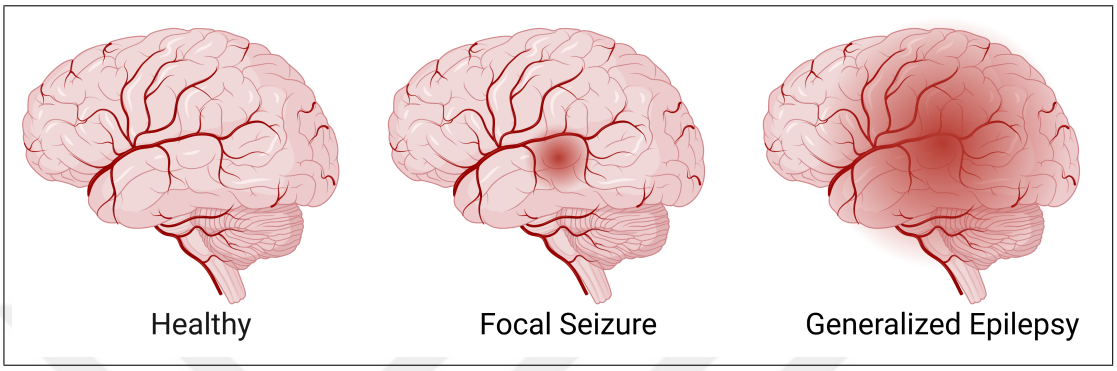

Figure 2.1 A human brain in its normal state, an epileptic brain with focal seizure in the temporal lobe, and an epileptic brain with generalized epilepsy.

2.2.2 Clinical Definition

A change in neurologic function, or seizure, can be defined as an episode that is brought on by the brain's neurons firing too quickly and synchronously [32]. When referring to seizures, the term "epileptic seizure" is used to differentiate those brought on by abnormal neuronal firing from non-epileptic events. Epilepsy can be brought on by a wide variety of factors, each of which points to a fundamental problem in the functioning of the brain. Because it is a secondary condition that is only temporary and not a chronic state, a seizure that is triggered by an insult that can be reversed, such as a fever or hypoglycemia, does not meet the criteria for being classified as epileptic [33]. The term "epilepsy syndrome" is used to describe a collection of symptoms that manifest in patients with epilepsy and that share commonalities with regard to seizure groups, the time when the patients experience their first onset, gentic linkage and responsiveness to therapies like anti-epileptic drugs (AED) or medicines [34].

The medical condition known as epilepsy is a spectrum disorder that encompasses more than just seizures that happen on their own and repeatedly. Additional

stress of caring for a patient or member of a family who is suffering from a chronic illness [35] is required, especially epileptic children. The mental co-occurring disorders, once thought to be a result of uncontrolled seizures or medication side effects, are now understood to be intrinsic to the disorder, appearing prior to the seizures and being attributable to a neuronal network disorder. Even a single seizure can have long-term consequences for brain growth, altering the receptor and causing neural cell loss and so affecting cognition and behavior [36].

3. SIGNAL PROCESSING

The fundamental concepts of signal processing are discussed in this chapter. Section 3.1 includes a brief and concise summary of bioelectrical signals transmitted from within the body of a living being. Section 3.2 clarifies the significance of wavelets in spatiotemporal resolution is then discussed, along with their application to the analysis of EEG signals, which are notoriously difficult to decipher due to their inherent chaos. Furthermore, in order to lay the groundwork for the following chapter, this section analyzes the continuous wavelet transform as well as the complex morlet transform.

3.1 Bioelectrical Signals

Information about a living thing can be transmitted via electrical activity in the body or a bioelectrical signal, which is defined as a change in the magnitude or quantity of a biological variable over temporal and spatial domains [37]. These bioelectrical vital signals are monitored, measured, and analyzed in clinical settings [38].

Bioelectrical signals are produced by nerve and muscle cells, and they are the direct result of electrochemical changes that take place both within and between cells. An action potential is generated in a nerve or muscle cell when it receives a stimulus strong enough to cause the cell to "activate, or cross a threshold. The action potential can be measured with either intracellular or extracellular electrodes and represents a transient influx of ions across the cell membrane. Excited cells can "fire" action potentials that travel along their axons to other nearby cells. As a result of the coordinated activation of numerous cells, an electric field is created that spreads throughout the living tissue. With the aid of surface electrodes, these variations in extracellular potential at the epidermis of the tissue or organism can be recognized and measured. Electrocardiograms, electrogastrograms, electromyograms, and electroencephalograms (ECG, EGG, EEG, and EMG) are all examples of this phenomenon [38].

Biological and physiological processes in the human body, such as changes in brain metabolism, can affect the method of combining EEG source signals. A number of these shifts can be traced back to the manner in which the brain is being utilized. Nonlinear behavior can be inferred from the system as a result of these effects. Analysis of such a system is notoriously challenging, and to date, no one has built a comprehensive model of the system to facilitate the study of brain signals [39].

The signals' nonstationarity can be quantified by analyzing their temporal distributions. Technically speaking, a process is said to be stationary if and only if the signal distribution remains unchanging over a large enough interval of time.

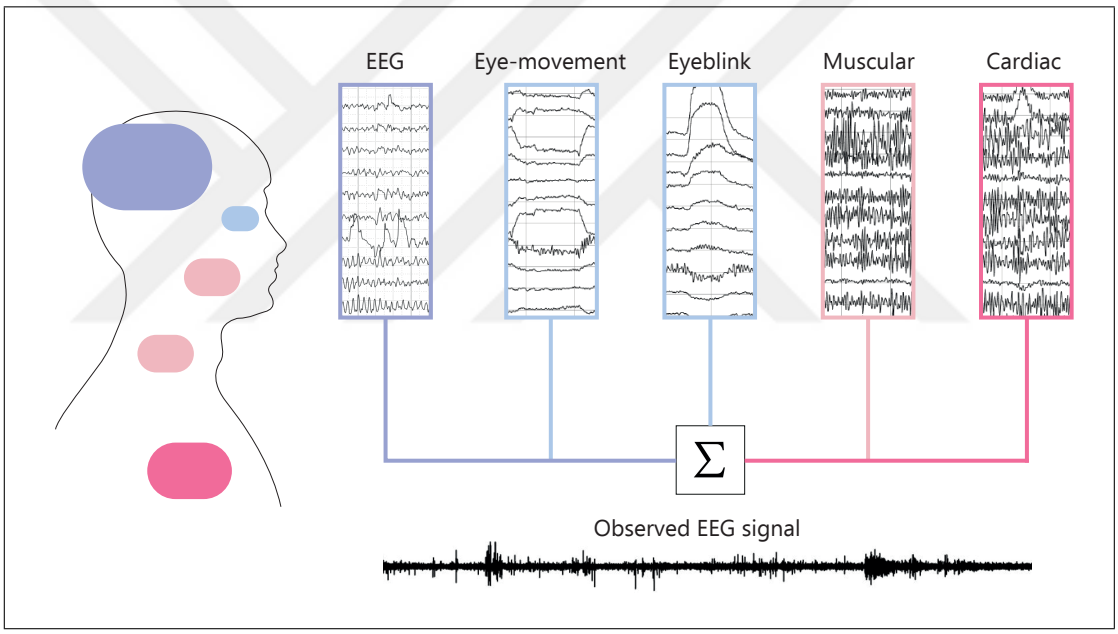

Figure 3.1 Configuration of an observed EEG signal including biological artifacts.

However, stationarity in a broad sense is not always necessary; sometimes it is sufficient for certain statistics to remain stable (or show only minor variations) over time. Though the multichannel EEG distribution is often modeled as a multivariate Gaussian, its mean and covariance often vary across time segments. EEGs are therefore only considered stationary within very brief intervals and are more accurately described as quasi-stationary [39].

A 10-second window of EEG is considered stable, despite the fact that it can

vary due to rapid changes in the brain state, which experiences contamination by some intrinsic artifacts such as eye blinks, muscle activity, and heartbeats [40].

3.2 Wavelets

The term *wavelet* itself was coined by Morlet and Grossmann in the early 1980s, with the two researchers drawing on ideas from both areas of mathematics [41]. They referred to it with the French word *ondelette*, which literally translates to *small wave* [41]. Mathematical analysis, quantum physics, electrical engineering, seismic geology, and even medicine [42] all played a role in the development of wavelets.

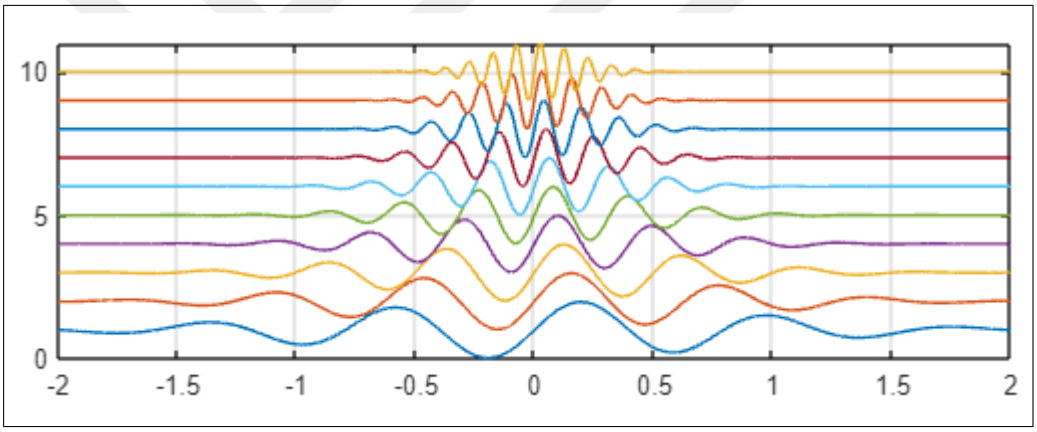

Figure 3.2 Wavelets with different scales.

Mathematical functions called "wavelets" can segment information into its constituent frequency bands, allowing researchers to examine each band separately at a level of detail appropriate to its size. The process of representing a function using wavelets is known as a wavelet transform.

In a more mathematical sense, a wavelet is a function that can be applied to the task of partitioning partitioning a time function into a variety of scaled elements [43]. In most cases, each element of a scale can be associated with a certain frequency range. Understanding something on a larger scale requires being able to examine its minute details. For this to happen, the frequency of interest in the unknown signal must be

relatively close to that of the wavelet. The core of numerous real-world applications of wavelet theory is built around this idea of resonance [44].

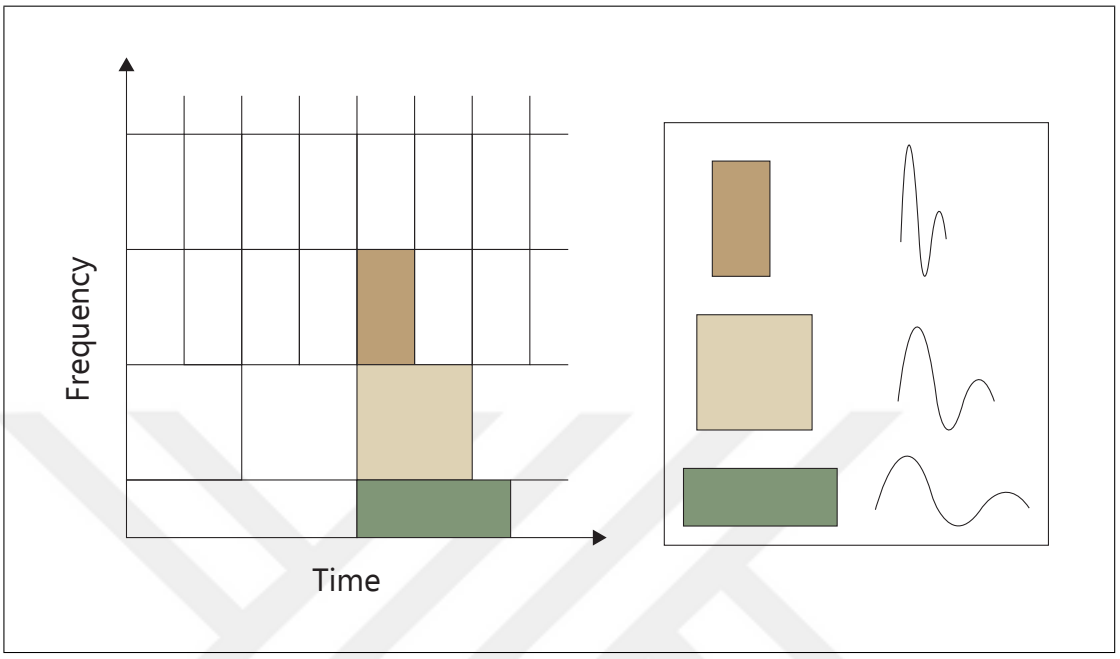

Figure 3.3 Different Heisenberg boxes in time- frequency axis, the top brown box represents low frequency resolution and higher temporal resolution opposite to the green box which boasted spatial resolution and degraded the temporal resolution.

The last twenty years have been crucial to the development of the theory of wavelets [45]. A wavelet is an oscillation that begins with an amplitude of zero, grows, and then shrinks to zero again. It's easiest to think of this phenomenon as a *brief oscillation*, like something a seismograph would pick up on [46]. Typically, wavelets are designed to have a set of characteristics that makes them appropriate for use in signal processing.

The convolutional "shift, multiply, and sum" technique allows for the extraction of information from an unknown signal by combining wavelets with subsets of the signal [47]. Surprisingly, many fields beyond traditional signal processing make use of the Fourier transform. The mathematical basis for wavelets has grown significantly since their introduction as an alternative to the Fourier transform. The first uses of wavelets were in signal processing and filtering. Although initially developed for use in the study of waves, wavelets have since found numerous other applications [48].

3.2.1 Continuous Wavelet Transform

A wavelet is a function that has both finite energy and a zero average. A family of *wavelet daughters*, denoted by the notation $\psi_{\tau,s}$, can be obtained from a mother wavelet $\psi \in L^2(\mathbf{R})$ by simply scaling and translating ψ.

$$\psi_{\tau,s}(t) := \frac{1}{\sqrt{|s|}}\psi\left(\frac{t-\tau}{s}\right) \tag{3.1}$$

The width of a wavelet is controlled by the dilation or scaling factor s, while its location is set by the translation parameter τ. Scaling causes expansion (if $|s| > 1$) or contraction (if $|s| < 1$) of a wavelet, while translation moves it forward or backward in time. The CWT along time is defined as:

$$W_{x;\psi}(\tau, s) = \int_{-\infty}^{\infty} x(t)\frac{1}{\sqrt{|s|}}\psi^*\left(\frac{t-\tau}{s}\right)dt \tag{3.2}$$

While τ describes the position of the wavelet in the time domain, s describes the location of the wavelet in the frequency spectrum. By translating the initial series to a specified function, the wavelet transform makes it possible to collect time and frequency information at the same time.

$$\int \psi(t)dt = 0 \tag{3.3}$$

Eq. 3.3 implies that finite energy function which satisfies the admissibility condition can be used as a mother wavelet. In order for the wavelet to integrate to

zero, it is necessary for there to be some oscillations in the mother wavelet, which is a negative value in one location to cancel out the positive value in another location.

The wavelet and Fourier transforms have very similar mathematical expressions. The primary distinctions are that the Fourier transform lacks a time localization parameter and uses cosine and sine functions rather than a wavelet function.

$$W_x(\tau, s) = \frac{\sqrt{|s|}}{2\pi} \int_{-\infty}^{\infty} \Psi^*(s\omega) X(\omega) e^{i\omega\tau} d\omega \tag{3.4}$$

Here, $\Psi(\omega)$ when it fulfills the admissibility condition, it will be entitled as a mother wavelet. The mother wavelet Ψ is a window function that is used in the construction of the wavelet transform. There is typically an admissibility condition that the mother wavelet must meet, such as:

$$\int \frac{\Psi(\omega)^2}{|\omega|} d\omega < +\infty \tag{3.5}$$

3.2.2 Complex Morlet Transform

Morlet wavelets are defined as sine waves that have been truncated using a Gaussian [49]. Time-frequency analysis makes use of the complex Morlet wavelet, which can be thought of as the real-valued Gaussian being squeezed by a complex-valued sine wave. After that, the time series signal is convolved with the complex Morlet wavelet, resulting in a signal with complex values, which can then be used in phase extraction.

Morlet wavelet which is a modulated Gaussian window is:

$$w(t|\sigma, t_0, f_0) = e^{-\frac{(t-t_0)^2}{2\sigma^2}} e^{j2\pi f_0 t} \tag{3.6}$$

where σ is the window length, in time and in frequency domains, t_0 is the center of the window in time and f_0 is the center of the window in frequency.

The admissibility condition of a wavelet cannot be checked because the Fourier transform of a wavelet does not equal zero at the origin. It is possible to find a solution to the issue by limiting the maximum number of wavelet oscillations to $5 < \omega_c < 6$.

The central frequency of the Morlet wavelet filter in Fourier form is calculated to be $f_c = 0.8125$ Hz. Herein, ω_c defines the central frequency of the band pass filter of the Morlet wavelet.

4. METHODOLOGY

This chapter delves into the pipeline of preprocessing pediatric patients' continuous EEG data. Section 4.1 presents an overview of CHB-MIT pediatric patients dataset. Section 4.2 describes the filtering technique for EEG records. Section 4.3 further describes how the continuous EEG data were divided into segments. Section 4.4 builds on the concept of cross-frequency coupling, and Section 4.5 describes the fundamental approach of phase-amplitude coupling, as well as the steps with an illustrated algorithm to extract the phases and amplitudes. Finally, section 4.6 specifies the determined threshold used in the research.

A significant part of the methodology for treating the EEG signals was adopted from [50], which used a different dataset. The study is the MS thesis of Lucasius where phase-ampltidue coupling technique was utilized. Volterra Neural Network was used for prediction, while we use VVG-19 and Transfer Learning.

4.1 Pediatric Dataset

The CHB-MIT database contains continuous EEG data from 23 pediatric patients diagnosed with epilepsy at Children's Hospital Boston [51]. According to Table 4.1, there are five males aged three to twenty-two and seventeen females aged one and a half to nineteen. Multiple days' worth of EEG signals were captured using the International 10-20 method, as shown in Figure 4.2. They also employed 23 electrodes, however in our study we reduced the number to 18 electrodes that are localized in all 23 patients. The channels are set up as follows: FP1-F3, F3-C3, C3-P3, P3-O1, FP1-F7, F7-T7, T7-P7, P7-O1, FP2-F4, F4-C4, C4-P4, P4-O2 FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ, and CZ-PZ as noted in Figure 4.2. The recordings are sampled at a rate of 256 hertz with 16 bit resolution. The CHB-MIT data was reviewed and annotated by specialists, who identified 182 epileptic epochs across 129 sessions. Ictal state, the

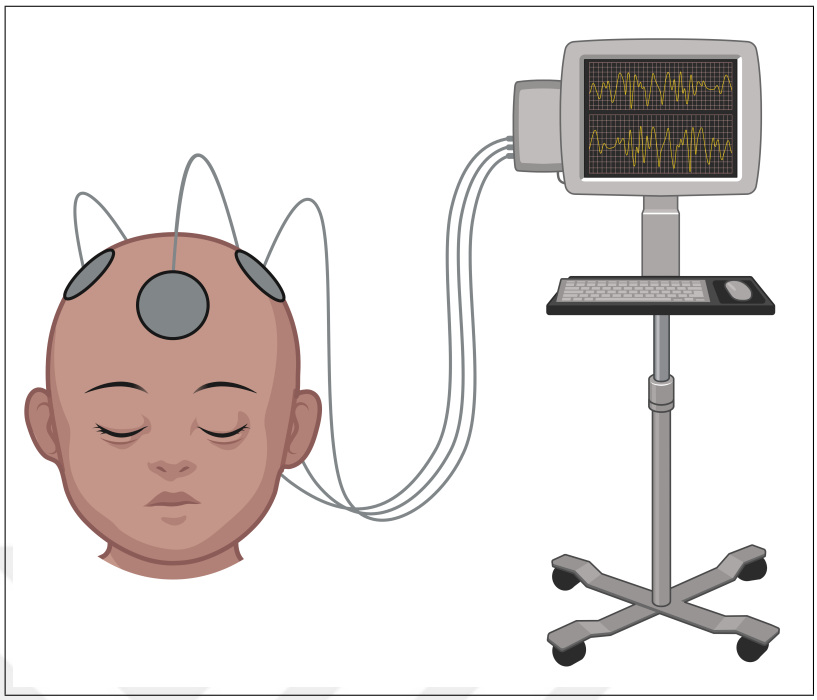

Figure 4.1 Configuration of EEG and pediatric patients.

time period during which a patient experiences the onset of a seizure [52], is easily detectable from raw signals by experts. The time that passes normally in between epileptic attacks is called the interictal period [52]. The preictal stage occurs between the interictal and ictal stages of an epileptic seizure [52].

4.2 Filtering EEG

It is essential that the filtering procedure does not result in any morphological changes that could be misinterpreted as pathological [53]. For this reason, filters employed in biomedical settings must feature a linear phase frequency response at least in the signal frequency range [53]. One frequent technique for dealing with problems caused by a nonlinear phase response is to first apply a forward-time causal filter to the noisy data, then apply the same filter to the output of the time-reversed process, and finally apply the original forward-time causal filter once more. The technique employed is known as "forward-backward filtering", and it is applicable in non-causal filter design

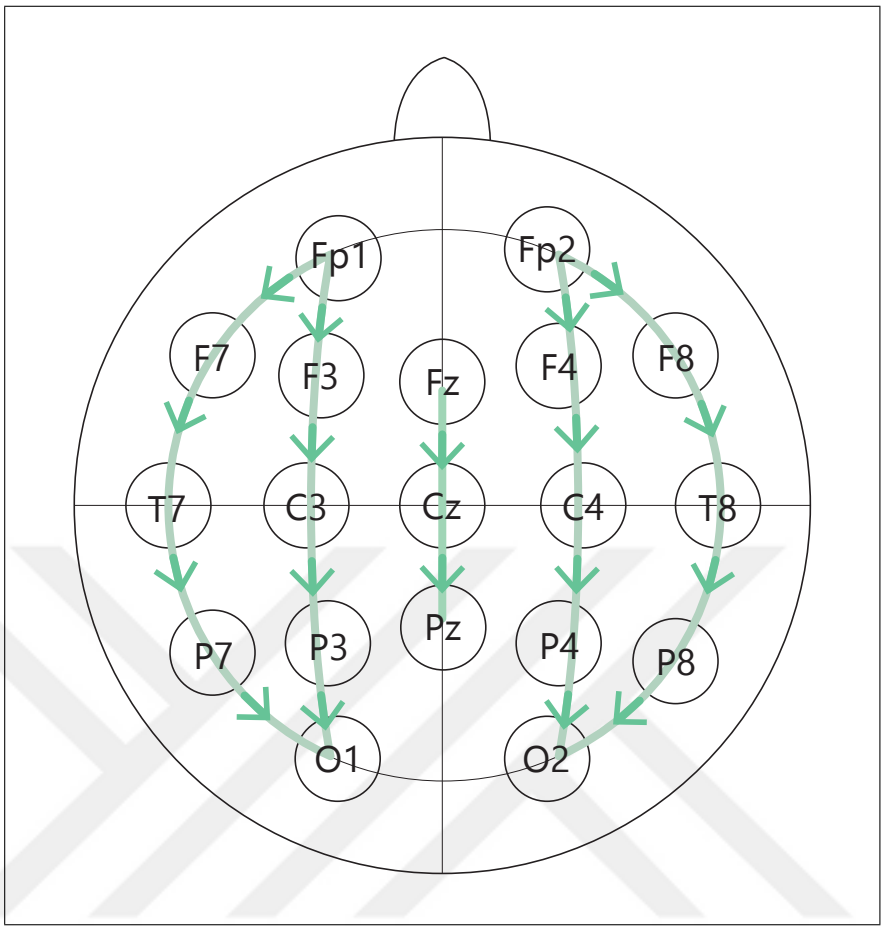

Figure 4.2 Bipolar longitudinal montage employs a 10-20 system for electrode placement.

and zero-phase filtering [54].

According to the wavelets preprocessing, the EEG waveforms between 1 Hz and 150 Hz are significant [55]. Recordings of EEG signals will be flawed due to the presence of the artifacts discussed earlier. One of the most important steps in the preprocessing is to clean up the EEG by eliminating any unwanted artifacts.

Bandpass filters between 0.5 and 130 Hz only let in the relevant frequencies while blocking out the noise frequencies created by unwanted actions. Therefore, to efficiently eliminate and dampen artifacts, it is necessary to create an algorithm tailored to the specific application at hand. The EEG signals are then processed with the following filters:

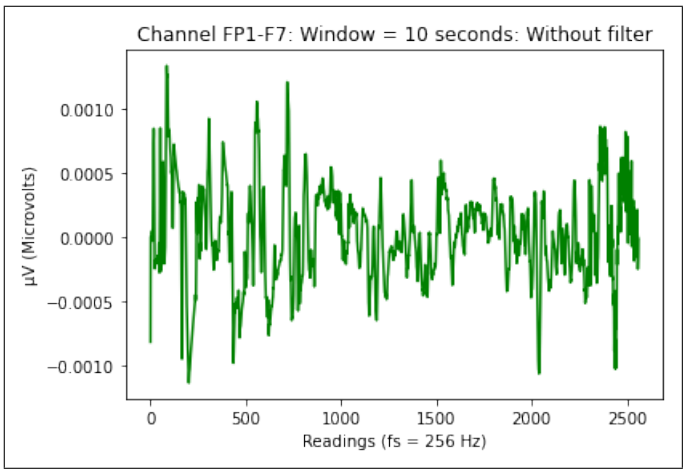

Figure 4.3 EEG recording for Channel FP1-F7 without using filter.

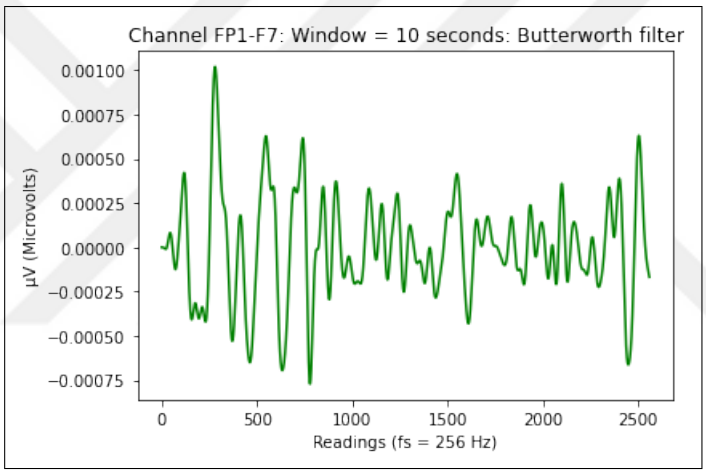

Figure 4.4 EEG recording for Channel FP1-F7 after using Butterworth filter.

1. The high-frequency content is filtered out using a Butterworth low-pass filter with order = 16 and cutoffs (0.5-50) Hz.

2. The DC component is removed and low frequencies are cut off using a Butterworth high-pass filter.

3. A notch filter is used to muffle appliances that cause interference on the power line at exactly 60 hertz.

Table 4.1
Information about CHB-MIT pediatric patient, their gender, age, the number of seizures an the number of sessions.

Patient ID	Gender	Age (Y)	No. of seizures	No. of sessions
chb01	Female	11	7	44
chb02	Male	11	3	36
chb03	Female	14	7	38
chb04	Male	22	4	42
chb05	Female	7	5	39
chb06	Female	1.5	10	18
chb07	Female	14.5	3	19
chb08	Male	3.5	5	21
chb09	Female	10	4	19
chb10	Male	3	7	25
chb11	Female	12	3	35
chb12	Female	2	40	24
chb13	Female	3	12	33
chb14	Female	9	8	26
chb15	Male	16	20	40
chb16	Female	7	10	19
chb17	Female	12	3	21
chb18	Female	22	4	3
chb19	Female	19	3	30
chb20	Female	6	8	29
chb21	Female	13	4	33
chb22	Female	9	3	30
chb23	Female	6	7	9

4.3 Segmentation of EEG

It is necessary to segment the raw CHB-MIT dataset EDF files, which contain a continuous long-term signal lasting around an hour, in order to obtain the trials required for seizure prediction, namely the interictal and preictal intervals.

We collected both the interictal and preictal intervals for each patient for a

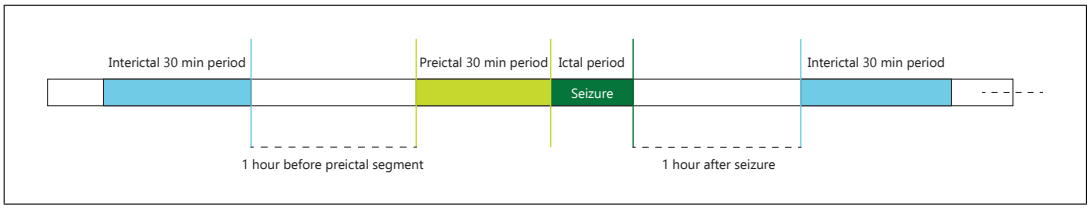

Figure 4.5 Schematic figure of slicing out the ictal period while maintaining the other periods.

period of 30 minutes, with each interval being collected separately, as depicted in Figure 4.5. In order to maintain the reliability of the prediction, ictal intervals were not used. In order to improve the quality of the EEG data, each selected EEG segment has its duration cut down to 10 seconds, and there is a 30% percent overlap between the segments, as shown in Figure 4.6.

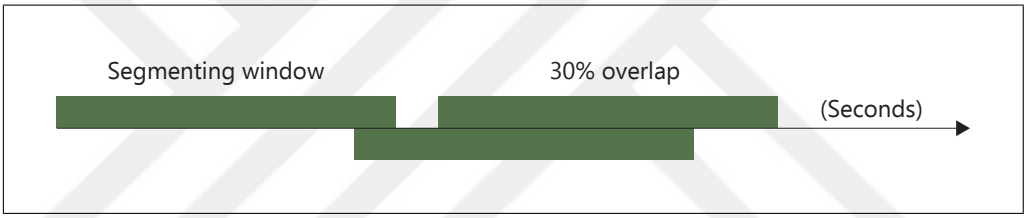

Figure 4.6 EEG signal segmented along the time axis.

4.4 Cross Frequency Coupling

In general, the data shows cross-frequency coupling or frequency-dependent patterns, either consistently or inconsistently at the same frequency within or across recordings [56]. In linear and Gaussian processes, coherence can quantify the degree to which two frequency components are dependent on one another as well as determine whether or not they are statistically independent of one another [56].

In some EEG studies, the seizure onset region showed a significantly greater Cross Frequency Coupling (CFC) between high frequency oscillations implying gamma waves and low frequency oscillations implying theta and alpha waves [57]. The CFC is determined by filtering the information into high and low frequency bands [58]. The next step is to analyze the filtered signals for their amplitude and phase, and then

check if the two are related as given by:

$$(\phi_{f_p}(t), A_{f_A}(t)) = \begin{pmatrix} \phi_{f_p}(1) & A_{f_A}(1) \\ \phi_{f_p}(2) & A_{f_A}(2) \\ \phi_{f_p}(3) & A_{f_A}(3) \\ \phi_{f_p}(4) & A_{f_A}(4) \\ \vdots & \vdots \end{pmatrix} \tag{4.1}$$

where $\phi_{f_p}(i)$ is the phase of low frequency band activity of time index i. If two frequencies have the same index in time, then $A_{f_A}(i)$ is the amplitude of that frequency.

4.5 Phase Amplitude Coupling

The concept of phase-amplitude coupling, or PAC, is gaining traction in the field of neuroscience as a potential clue that could unlock the mysteries of both normal and abnormal information processing in the brain [59]. The phase of a lower-frequency oscillation can modulate the amplitude of a higher-frequency oscillation [55], a phenomenon known as amplitude-phase coupling (PAC).

In signal processing chapter, we reached Eq. **??** were we obtained the real an the imaginary components of CWT:

$$W(f, n) = \omega(f, n) + j\tilde{\omega}(f, n) \tag{4.2}$$

The high frequency amplitude:

$$A(f_H, n) = \sqrt{\omega^2(f_H, n) + \tilde{\omega}^2(f_H, n)} \tag{4.3}$$

The low frequency phase:

$$\phi(f_L, n) = \arctan \frac{\tilde{\omega}(f_L, n)}{\omega(f_L, n)} \tag{4.4}$$

The PAC is calculated from these signals by summing the amplitudes over the phases and binning them according to the given time window in order to convert the phase readings into probability values.

The phases are first partitioned into $N = 18$ bins of 20 degree each, and then the amplitudes are normalized over the sum of all the bins.

$$x_b\left(f_L, f_H, n\right) = \frac{\langle A\left(f_H, n\right)\rangle_b}{\sum_{k=1}^{N}\langle A\left(f_H, n\right)\rangle_k} \tag{4.5}$$

Each x_b represents a different probability value of the bth bin. After that, the probability distribution is utilized in the computation of an entropy measure, which is as follows:

$$Q(f_L, f_H, n) = -\sum_{b=1}^{N} x_b(f_L, f_H, n)\log(x_b(f_L, f_H, n)) \tag{4.6}$$

As a final step, the PACs are calculated by normalizing the entropy measures

by the maximum entropy given by a uniform distribution with an entropy of $Q_{max} = \log N$:

$$\text{Coupling Index}(f_L, f_H, n) = \frac{Q_{max} - Q(f_L, f_H, n)}{Q_{max}} \tag{4.7}$$

The multi-variant time series $y_e(t) = 1, 2, ...E$ channels. Hence, the global measure of Coupling Index:

$$\overline{\text{Coupling Index}}(f_L, f_H) = \langle \text{Coupling Index}(f_L, f_H) \rangle_E \tag{4.8}$$

where $\langle \dots \rangle$ is the spatial average over E channels. The $\overline{\text{Coupling Index}}$ is calculated over f_L and f_H resulting into 131×91 comodulation images, whereas, we adapted the spatial average technique from a research of epileptic participants' using iEEG and EEG recordings by Grigorovsky et al. [60].

4.6 PAC Steps

The CFC in the form of phase-amplitude coupling (PAC) was computed, which assesses how much the phase of LFOs modulates the amplitude of HFOs [61]. We developed an approach based on Tort et al. algorithm [55] to extract PAC indices from continuous EEG multi-channel recordings for use in comodulation images that can handle epileptic data. Our main algorithm as illustrated in Figure 4.9, takes into account the sampling frequency and the low/high frequency to extract (frequency, time, scales) for each block, which are then used to process the phase block (channels, low frequency, samples) and the amplitude block (channels, high frequency, samples).

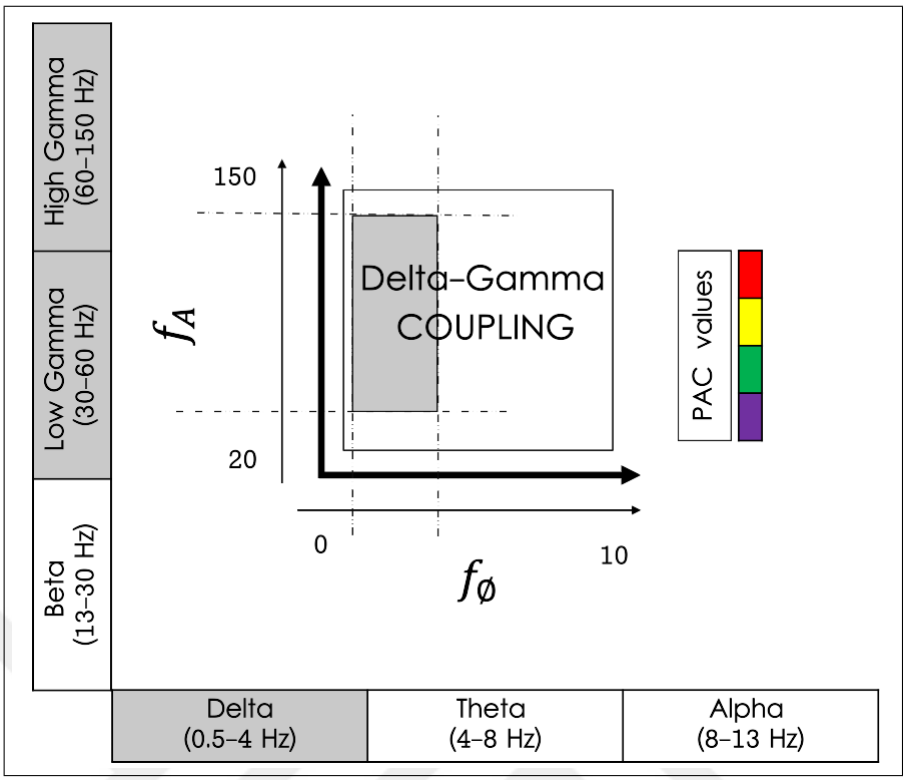

Figure 4.7 Schematic figure of Comodulogram between low and high frequencies, the grey tile shows the expected region where the activity ought to appear.

Extraction of low and high frequency ranges and a complex morlet transform were performed independently on each channel.

We used Fast Fourier Transform (FFT) for this purpose of convolution. Following that, we segmented each phase block into 18 bins depending on its placement along a 360-degree axis that was subsequently subdivided into 20-degree increments. In the algorithm shown in Figure 4.10, the phase block is subjected to a checking system in which it is determined whether or not the phase's current location is larger or smaller than the original phase location, after which a new phase location block consisting of zeros and ones is created and multiplied (tensor multiplication) with the amplitude block. The probability density function shown in Eq. 4.5 was obtained by adding a fourth dimension and then normalizing the original result. The entropy was computed using Eq. 4.6 to reveal the deviation from a uniform distribution. If NAN was detected, it was changed to 0, and the PAC image did not include any tiles representing the activity. Otherwise, Eq. 4.7 was used to determine the PAC index, which quantifies

the variation between phase and amplitude over a specified time interval, as seen in Figure 4.8. These images proposed as a potential preictal biomarker for the evaluation of epilepsy prior to surgery, Table 4.2 presents the coupling and the condition related to coupling and the conducted studies.

Table 4.2
Abnormalities related to coupling with gamma wave and its corresponding condition with the references

Phase-to-Gamma Coupling	Condition	Reference
Alpha-to-Gamma Coupling	Autism	[62]
Theta-to-Gamma Coupling	Alzheimer	[63]
Delta-to-Gamma Coupling	Epilepsy	[?]
Beta-to-Gamma Coupling	Parkinson	[64]

4.6.1 Illustrated Algorithm

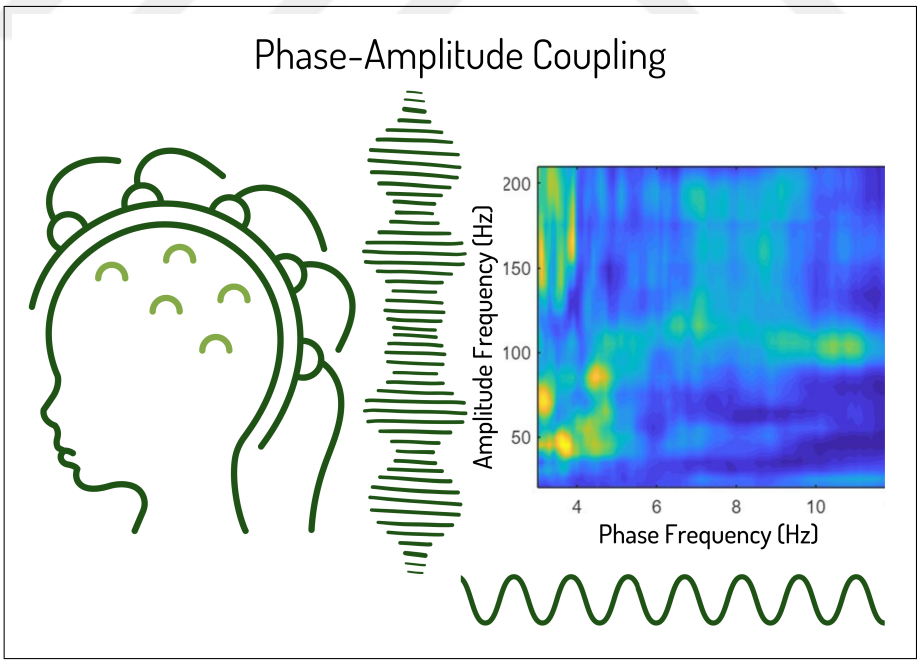

Figure 4.8 Pediatric patient where amplitude and phase are extracted and created a coupling.

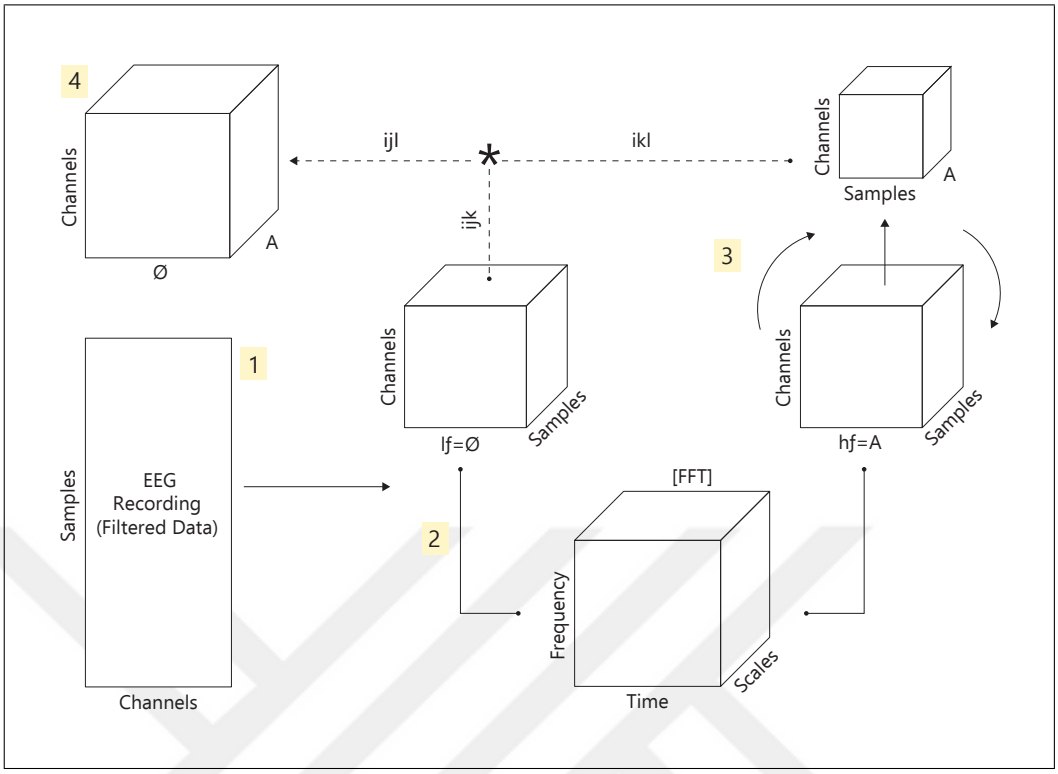

Figure 4.9 Steps regrading phase amplitude coupling images.

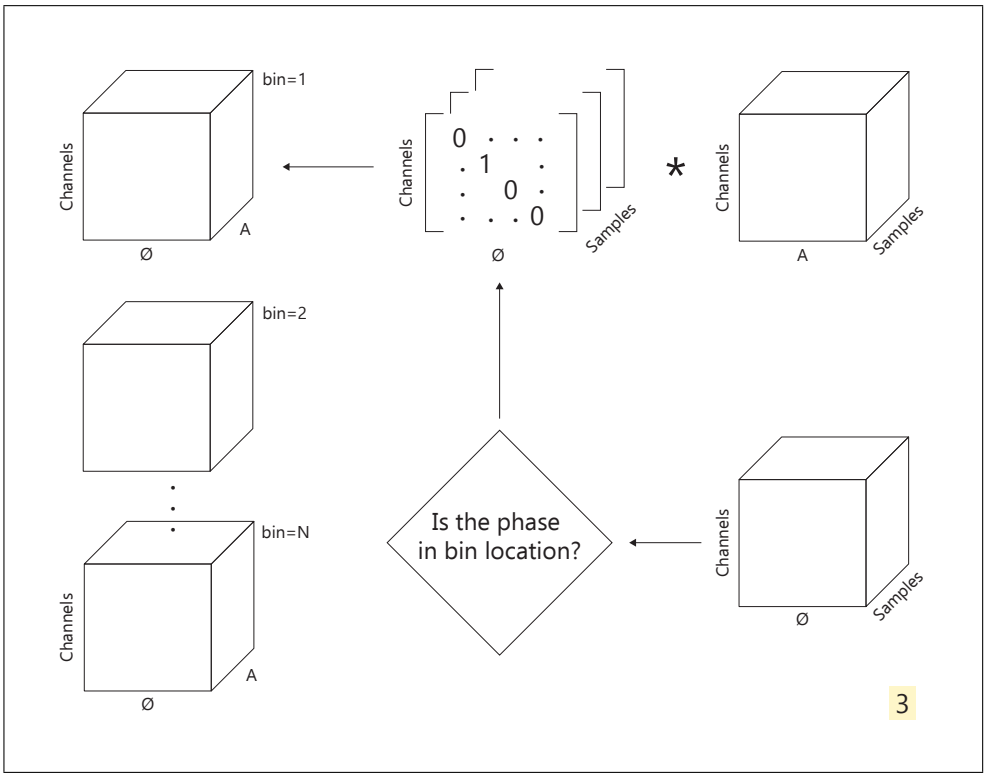

Figure 4.10 Explaining the third step.

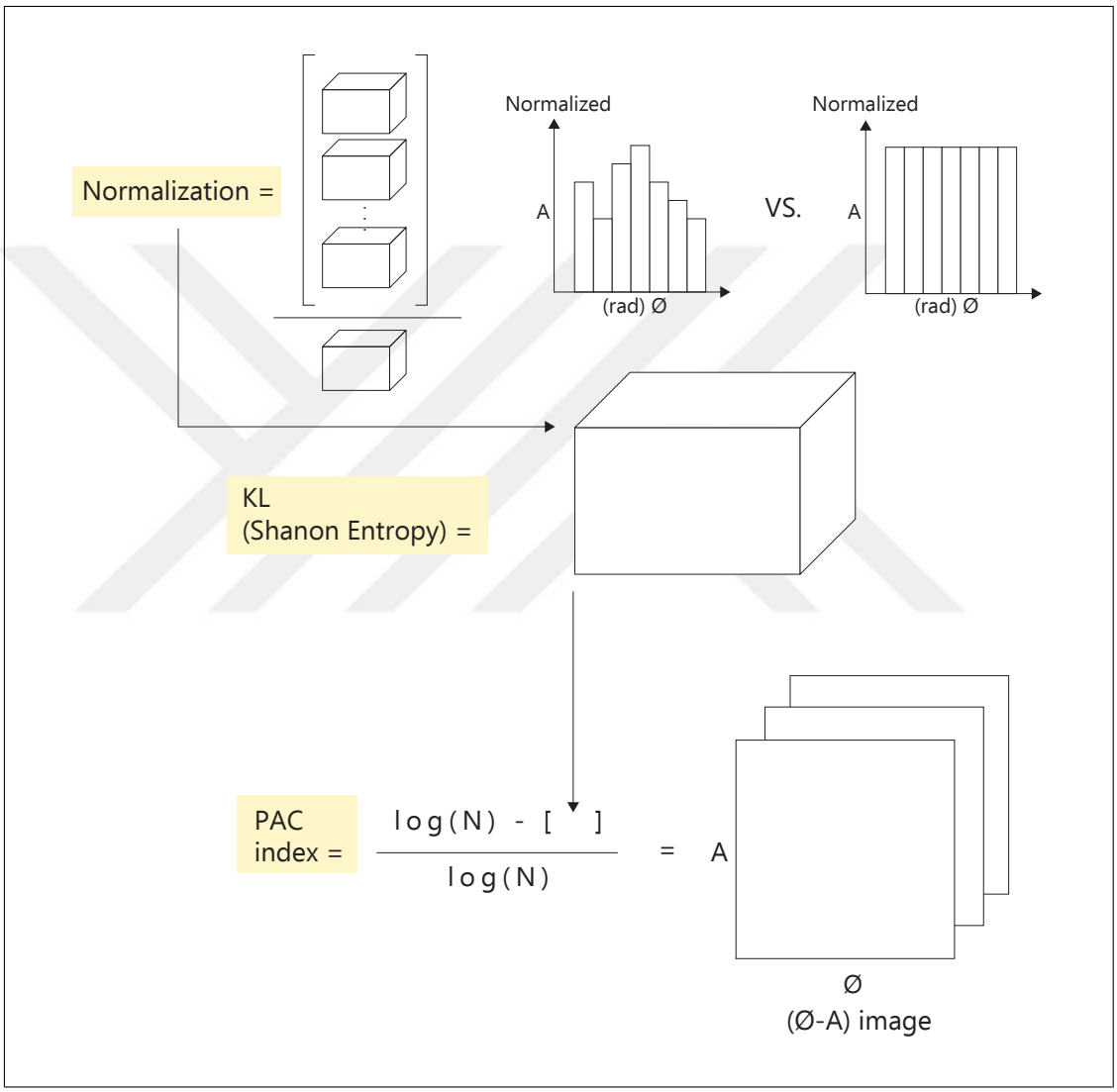

Figure 4.11 The following steps in computing the comodulation starting from normalization, calculating Shannon entropy to evaluate the phase amplitude indices .

4.7 Thresholding

The PAC indices are randomized in order to cut down on the amount of computation that is required during training and testing. A PAC Index's binary representation threshold can be anywhere from 0% to 100% of the index's maximum value. For the purposes of this study, a threshold of 60% was chosen; any pixel values that were higher than 60% of the maximum value were assigned the value 1, while the remaining pixel values were assigned the value 0.

5. VGG-19 AND TRANSFER LEARNING

This chapter investigates the deep learning approach to the problem of seizure prediction. Transfer learning was utilized and the VGG-19 model was implemented as a deep CNN model. Section 5.1 describes transfer learning as a technique and its usefulness in transferring pretrained model knowledge. Section 5.2 defines the fundamental CNN nomenclature. VGG-19 outperforms baselines on several tasks and datasets outside of ImageNet, as shown in Section 5.3. The next subsections include the equations for forward and backprorogation for this model. The processes for structuring the model are covered in Section 5.4. Finally, Section 5.5 describes the parameters utilized in the study to train the model.

5.1 Transfer Learning

The ability of various systems to quickly adapt to new roles, responsibilities, and environments is investigated in the field of transfer learning [65]. In a nutshell, the machine learning paradigm known as "transfer learning" refers to an algorithm's use of knowledge gained from one or more application scenarios to enhance its learning performance in a "target" scenario. Machine learning systems can use external data and models to supplement what is lacking in the target domain, allowing them to better solve problems. This strengthens the system's dependability and robustness by reducing the likelihood that the machine learning model will deviate too greatly from its baseline performance in response to unforeseen perturbations. Transfer learning is a new learning paradigm that can be viewed in contrast to traditional machine learning, which relies on large amounts of well-defined training data as the input.

Transfer learning systems are able to increase their range of applicability by transferring knowledge from one domain to another by incorporating information from different domains to broaden their usefulness. This allows the systems to be used in

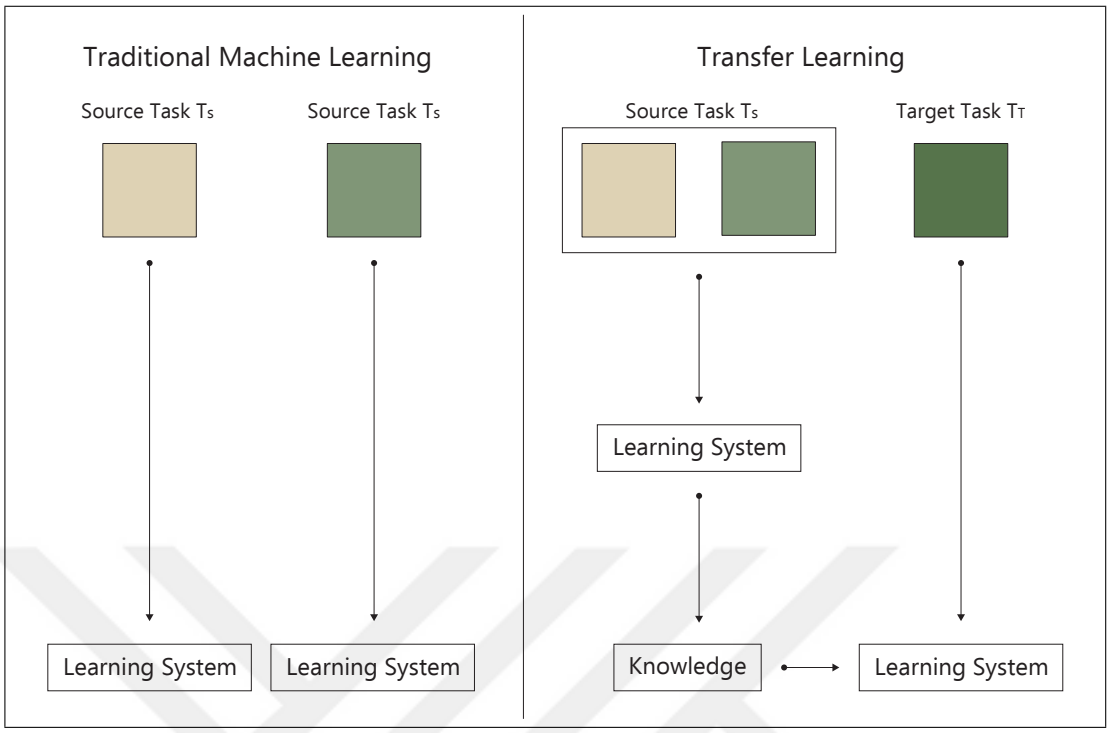

Figure 5.1 Transfer learning method

contexts beyond the ones for which they were originally created. This capacity for generalization contributes to making artificial intelligence more accessible and robust in many domains where specialized AI talent or resources (such as computing power, data, and hardware) may be scarce [66]. This is accomplished by making AI more capable of generalization.

5.2 CNN Terminology

The Convolutional Neural Network (CNN) is an up-and-coming deep artificial neural network that can perform convolutional computation and serves as a model algorithm for deep learning [67]. The CNN can be taught in a supervised, unsupervised, or semi-supervised environment and uses an approach that is similar to how humans perceive visual information [67]. Natural language processing, voice recognition, and image classification are just a few examples of where this network has been put to good use over the years. The CNN model is usually made up of four layers:

the convolutional layer, the pooling layer, the fully connected layer, and the Softmax classification layer. The filtering effect can be improved through iterative application of the convolutional and pooling operations. Many state-of-the-art CNN models have been developed, including ResNet, AlexNet, GoogLeNet, and VGG-19.

Convolutional layers, in particular, are in charge of pinpointing the local combination of features extracted from deeper layers. These neurons conduct a convolution, as opposed to the straightforward matrix multiplication carried out by the fully connected neurons [68]. To organize the weights of a convolutional neural network, **kernels** (also known as "filters") are used. Despite having less width and height than the input, most elementary CNN architectures require them to be proportional to the depth. When the image's input layer is considered, dimensions are typically w (width), h (height), and d (depth) (or, more formally, the three dimensions that reconcile RGB encoding). Therefore, stage one filtration must have a depth of three. Convolution is accomplished by applying a kernel to an input image and then multiplying that kernel by a set of weights that represent intensities within the kernel's application to the input image. The result is stored in the output matrix, also referred to as a feature map or activation map. These filters are available in a wide range of varieties. The network employs the following filter after fully processing the input. The activation map's depth is independent of both the input and the kernel depths and is instead proportional to the number of kernels used on the input image. Overfitting can be avoided with the help of **pooling layers** by combining features that are similar to those in the previous activation map. *The maxpooling layer*, which determines which value is the greatest, is the one used most frequently. The output of the step that involves convolution and pooling is then passed on to the **fully connected layer**, which flattens it and then makes a prediction regarding the label that most accurately describes the image. Like a standard feed-forward neural network, the inputs to the fully connected layer are multiplied by the layer's weights and then added together [69]. In the next step, an **activation function** is used to generate the final result. The results are then passed on to the next fully connected layer. Each class label is associated with a neuron in the final layer, which generates a probability distribution.

5.3 VGG-19

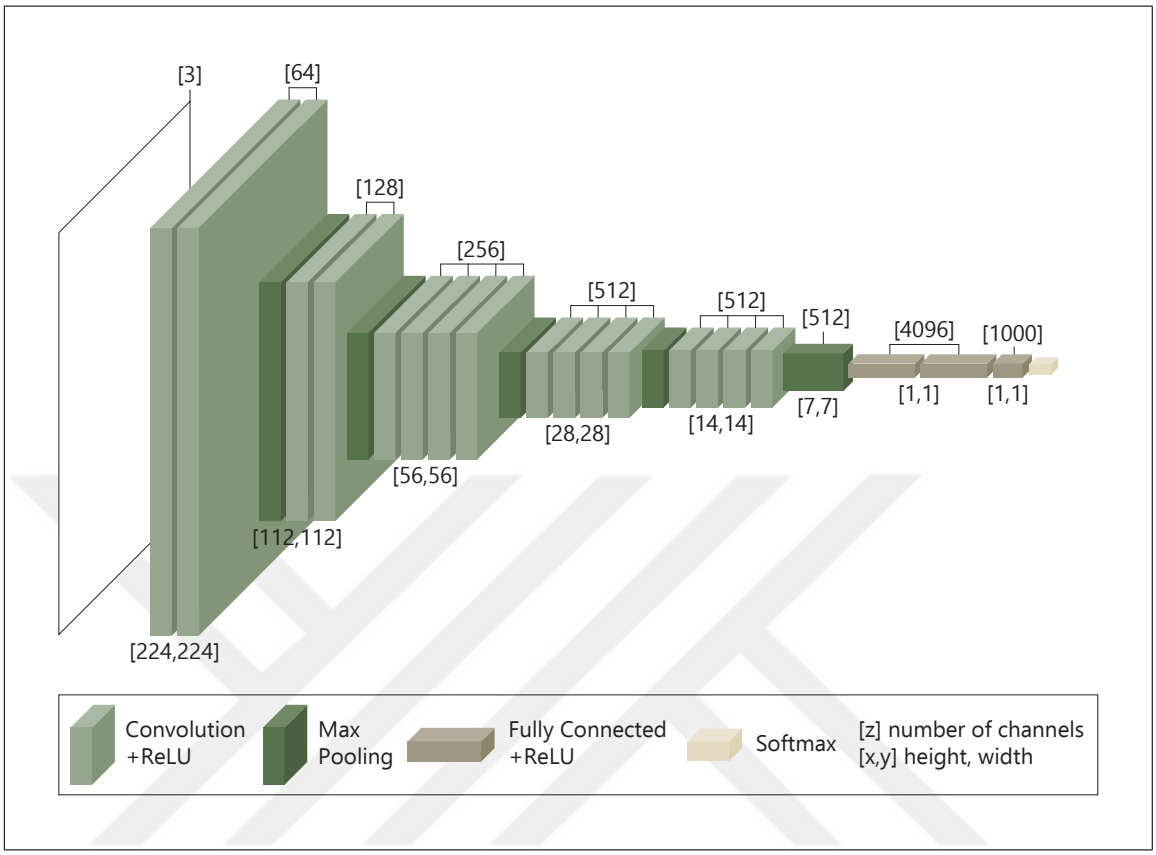

Figure 5.2 The VGG-19 network's structural layout. There are a total of 16 convolutions in the network, all linked together using ReLUs, and 5 maxpooling layers. The convolutions have filter maps with a starting value of 64 and a maximum value of 512. After the convolutions, a linear classifier is constructed, with three fully-connected (fc) layers, the first two of which contain 4096 features and the third of which contains 1000. Each value is then mapped by the softmax layer, which is connected to the final fc layer, to the probability of belonging to one of the ImageNet competition's thousand classes.

This network received an RGB image with a fixed size of (224×224), indicating that the matrix had the shape of $(224, 224, 3)$. Before training, only the training set's average RGB value was removed from each pixel. They did this by using 3-by-3 kernels with 1-pixel strides, thus covering the entire image. The image's spatial resolution was kept intact through the use of spatial padding. Maximum pooling was carried out using a stride 2 window size of 2 by 2. The next step was to create the Rectified linear unit (ReLu), which greatly outperforms earlier models based on *tanh* or *Sigmoid* functions in terms of classification accuracy and computational speed. Used a total of three fully connected layers, the first two being 4096 in size, the second having 1000 channels to facilitate 1000-way classification, and the third being a softmax function.

Since the authors of VGG Net [70] made the models available to the public, they can be used as-is or with minor modifications for other tasks that are similar to the ones for which they were developed. The models were merely used as an effective classification architecture for many other datasets. Facial recognition is another application that benefits from transfer learning [71]. Other frameworks, such as Keras and Pytorch, make weights readily available for customization and use.

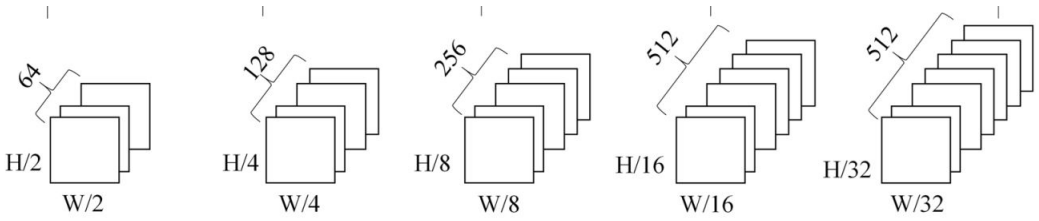

Figure 5.3 VGG-19 convolutional layers order and configurations [72].

For each convolutional layer: (Kernel = 3×3) and stride = 1 pixel.
For each maxpooling layer: (Kernel = 2×2) and stride = 2 pixels.

- The first block depth = 64: Conv1-1, Conv1-2, maxpool.
- The second block depth = 128: Conv2-1, Conv2-2, maxpool.
- The third block depth = 256: Conv3-1, Conv3-2, Conv3-3, Conv3-4, maxpool.
- The fourth block depth = 512: Conv4-1, Conv4-2, Conv4-3, Conv4-4, maxpool.
- The fifth block depth = 512: Conv5-1, Conv5-2, Conv5-3, Conv5-4, maxpool.

5.3.1 Forward Propagation for VGG-19

The input image, denoted by $\mathbf{x}$, is a three-dimensional tensor with the coordinates $(224, 224, 3)$, and the output, denoted by $\hat{\mathbf{y}}$, is a matrix $(1000, 1)$. Consider $\mathbf{a}^{[1]}$ to be the first convolution layer, which has the dimension $(224, 224, 64)$. The transition from the fully connected layer $\mathbf{a}^{[18]}$ to the softmax layer $\mathbf{a}^{[19]}$ starts off in the same manner as it does for other fully connected layers, as is customary for fully connected

layers in general. In order to obtain $\mathbf{z}^{[19]}$, we first carry out a matrix multiplication with the help of $\mathbf{W}^{[19]}$, and then we add a bias with the assistance of $\mathbf{b}^{[19]}$. Since $\mathbf{a}^{[18]}$ is a row vector with the dimensions $(4096, 1)$ and $\mathbf{z}^{[18]}$ is a row vector with the dimensions $(1000, 1)$, the dimensions of $\mathbf{W}^{[19]}$ are $(1000, 4096)$ and the dimensions of $\mathbf{b}^{[19]}$ are $(100, 1)$ respectively.

The convolution function:

$$\mathbf{z}^{[19]} = \mathbf{W}^{[19]}\mathbf{a}^{[18]} + \mathbf{b}^{[19]} \tag{5.1}$$

Softmax function:

$$a_{i,1}^{[19]} = \sigma_{i,1}\mathbf{z}^{[19]} \tag{5.2}$$

The ith element in $\mathbf{a}$:

$$a_{i,1}^{[19]} = \frac{e^{z_{i,1}^{[19]}}}{\sum_{j=1}^{1000} e^{z_{i,1}^{[19]}}} \tag{5.3}$$

The probability that the training instance image is from one of thousand classes:

$$\hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_{1000} \end{bmatrix} \tag{5.4}$$

The probability that y is equal to class i given an input image x:

$$\hat{y}_i = P(y = i|\mathbf{x}) \tag{5.5}$$

To compute this partial derivative, we apply the quotient rule to the derivative:

$$\frac{\partial a_{(1,1)}^{[19]}}{\partial z_{(2,1)}^{[19]}} = \frac{0 * \sum_{j=1}^{1000} e^{z_{(1,1)}^{[19]}} - e^{z_{(1,1)}^{[19]}} e^{z_{(2,1)}^{[19]}}}{\left(\sum_{j=1}^{1000} e^{z_{(1,1)}^{[19]}}\right)^2}$$

$$\frac{\partial a_{(1,1)}^{[19]}}{\partial z_{(2,1)}^{[19]}} = \frac{-e^{z_{(1,1)}^{[19]}} e^{z_{(2,1)}^{[19]}}}{\left(\sum_{j=1}^{1000} e^{z_{(1,1)}^{[19]}}\right)^2}$$

$$\frac{\partial a_{(1,1)}^{[19]}}{\partial z_{(2,1)}^{[19]}} = -a_{(1,1)}^{[19]} a_{(2,1)}^{[19]}$$

To generalize:

$$\frac{\partial a_{(i,1)}^{[19]}}{\partial z_{(j,1)}^{[19]}} = \begin{cases} a_{(i,1)}^{[19]} \left(1 - a_{(i,1)}^{[19]}\right) & i = j \\ -a_{(i,1)}^{[19]} a_{(j,1)}^{[19]} & i \neq j \end{cases}$$

The representation of Jacobian matrix:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{z}^{[19]}} = \begin{bmatrix} -\frac{y_1}{a_{(1,1)}^{[19]}} & -\frac{y_2}{a_{(2,1)}^{[19]}} & \cdots & -\frac{y_{1000}}{a_{(1000,1)}^{[19]}} \end{bmatrix}$$

$$\begin{bmatrix} a_{(1,1)}^{[19]} \left(1 - a_{(1,1)}^{[19]}\right) & -a_{(1,1)}^{[19]} a_{(2,1)}^{[19]} & \cdots & -a_{(1,1)}^{[19]} a_{(999,1)}^{[19]} & \cdots - a_{(1,1)}^{[19]} a_{(1000,1)}^{[19]} \\ \vdots & \vdots & \ddots & \vdots & \\ -a_{(1000,1)}^{[19]} a_{(1,1)}^{[19]} & -a_{(1000,1)}^{[19]} a_{(2,1)}^{[19]} & \cdots & -a_{(1000,1)}^{[19]} a_{(999,1)}^{[19]} & \cdots a_{(1000,1)}^{[19]} \left(1 - a_{(1000,1)}^{[19]}\right) \end{bmatrix}$$

The partial derivative of $\mathcal{L}$ with respect to $z_{(1,1)}^{[19]}$:

$$\frac{\partial \mathcal{L}}{\partial z_{(1,1)}^{[19]}} = -\frac{y_1}{a_{(1,1)}^{[19]}} a_{(1,1)}^{[19]} \left(1 - a_{(1,1)}^{[19]}\right) + \frac{y_2}{a_{(2,1)}^{[19]}} a_{(2,1)}^{[19]} a_{(1,1)}^{[19]} + \frac{y_3}{a_{(3,1)}^{[19]}} a_{(3,1)}^{[19]} a_{(1,1)}^{[19]} + \cdots + \frac{y_{1000}}{a_{(1000,1)}^{[19]}} a_{(1000,1)}^{[19]} a_{(1,1)}^{[19]}$$

$$\frac{\partial \mathcal{L}}{\partial z_{(1,1)}^{[19]}} = -y_1 \left(1 - a_{(1,1)}^{[19]}\right) + y_2 a_{(1,1)}^{[19]} + y_3 a_{(1,1)}^{[19]} + \cdots + y_{1000} a_{(1,1)}^{[19]}$$

$$\frac{\partial \mathcal{L}}{\partial z_{(1,1)}^{[19]}} = -y_1 + y_1 a_{(1,1)}^{[19]} + y_2 a_{(1,1)}^{[19]} + y_3 a_{(1,1)}^{[19]} + \cdots + y_{1000} a_{(1,1)}^{[19]}$$

$$\frac{\partial \mathcal{L}}{\partial z_{(1,1)}^{[19]}} = -y_1 + a_{(1,1)}^{[19]} \sum_{i}^{1000} y_i$$

5.3.2 Backpropagation for VGG-19

For Maxpooling Layers:

The loss with respect to the last convectional layer equals,

$$\frac{\partial \mathcal{L}}{\partial \mathbf{a}^{[16]}} = \frac{\partial \mathcal{L}}{\partial \mathbf{m}^{[16]}} \frac{\partial \mathbf{m}^{[16]}}{\partial \mathbf{a}^{[16]}} \tag{5.6}$$

In the following formula, only the denominator changes. For that, we take the denominator tensor array and break it down properly into a 1D vector.

$$\left[\frac{\partial m_{(1,1,1)}^{[16]}}{\partial a_{(1,1,1)}^{[16]}} \quad \cdots \frac{\partial m_{(1,1,1)}^{[16]}}{\partial a_{(14,14,512)}^{[16]}} \right] \tag{5.7}$$

Jacobin Matrix is similar to the above formula, only the nominator changes. For this, we also decompose it into a 1D vector.

$$\begin{bmatrix} \frac{\partial m_{(1,1,1)}^{[16]}}{\partial a_{(1,1,1)}^{[16]}} \\ \vdots \\ \frac{\partial m_{(7,7,512)}^{[16]}}{\partial a_{(1,1,1)}^{[16]}} \end{bmatrix} \tag{5.8}$$

For Conventional Layers:

ReLU function:

$$g(z) = \begin{cases} z, & z > 0 \\ 0, & z < 0 \end{cases} \tag{5.9}$$

The derivative of ReLU function:

$$g'(z) = \begin{cases} 1, & z > 0 \\ \text{undefined}, & z = 0 \\ 0, & z < 0 \end{cases} \tag{5.10}$$

The diagonal matrix,

$$\frac{\partial \mathbf{a}^{[16]}}{\partial \mathbf{z}^{[16]}} = \begin{bmatrix} g'\left(z^{[16]}_{(1,1,1)}\right) & 0 & \cdots & 0 & 0 \\ 0 & g'\left(z^{[16]}_{(2,1,1)}\right) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & g'\left(z^{[16]}_{(13,14,512)}\right) & 0 \\ 0 & 0 & \cdots & 0 & g'\left(z^{[16]}_{(14,14,512)}\right) \end{bmatrix} \tag{5.11}$$

5.4 Structuring The Model

The VGG-19 network is a deep structure that is a hodgepodge of convolutional, and fully connected layers. In this study, the network accepts RGB images as input. There are 19 levels of complexity in VGG-19. In comparison to other CNNs, VGG-19 performs better for a finite number of training epochs.

The pipeline toward employing VGG-19 and transfer learning begins with downloading the pixel input folder that contains all the PAC images extracted using Complex Morlet continuous wavelet transform and phase-amplitude coupling. Step two, as noted in Figure 5.5 involves reading in different PAC images from a storage database created previously with 80% training directory and 20% validation directory, with the folder's name serving as the input label for each cases interical and preictal. The third step involves the importing of pre-trained network VGG-19 model and by assigning the weights to ImageNet, as seen in Figure 5.4. For VGG-19, the input layer size of $224 \times 224 \times 3$ where 3 indicates the RGB channels. For the fourth step, we swap out the 1000 classes layer from the ImageNet dataset for just two classes taken directly

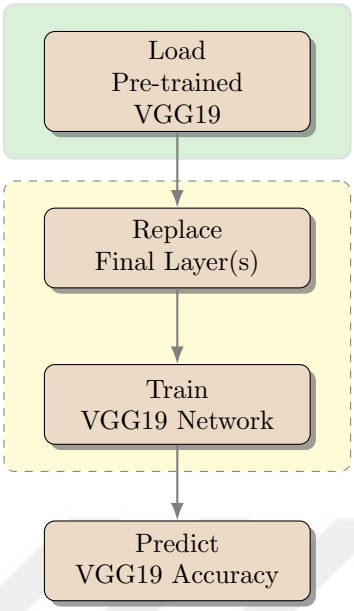

Figure 5.4 Flowchart for using VGG-19 and Transfer Learning.

from our own. The goal of transfer learning is to apply the knowledge in $\mathcal{D}_s$ to help learn the knowledge in $\mathcal{D}_t$.

Herein;

$$\mathcal{D}_s \neq \mathcal{D}_t$$

$$\mathcal{T}_s \neq \mathcal{T}_t$$

The domain of source and target:

$$\mathcal{D}_s = \{\mathcal{X}_s, \overbrace{P(\mathcal{X}_s)}^{\text{marginal}}\}$$

$$\mathcal{D}_t = \{\mathcal{X}_t, P(\mathcal{X}_t)\}$$

The tasks of source and target:

$$\mathcal{T}_s = \{\mathcal{Y}_s, \overbrace{P(\mathcal{Y}_s, \mathcal{X}_s)}^{\text{conditional}}\}$$

$$\mathcal{T}_t = \{\mathcal{Y}_t, P(\mathcal{Y}_t, \mathcal{X}_t)\}$$

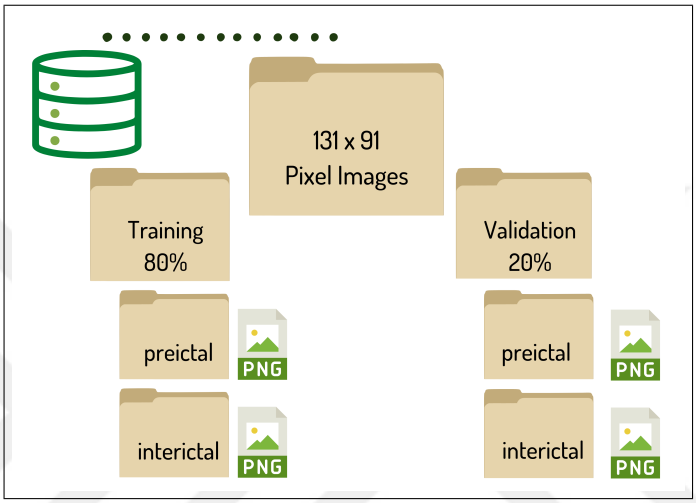

Figure 5.5 Database file that contains pixel images file and inside there are two files related to the training and the validation images of preictal and interical.

Both training and test data are passed to the objects $\mathbb{F}_{train}$ and $\mathbb{F}_{test}$, respectively, along with the folder containing the training data in each case. The information will be organized into the following folders as shown in Figure 5.5.

5.5 Parameters for VGG-19 and Transfer Learning

Epilepsy prediction model developed with VGG-19 and Transfer Learning is implemented in Python. Images of the phase amplitude coupling of preprocessed CHB-MIT data were used as training samples for this study. Google Colab Pro was the platform for the deep learning workstation, with Python 3.7 and the TensorFlow 2.0 framework utilized for deep learning. The gathered 9640 phase amplitude coupling images are utilized as the model's training set after being preprocessed by pixel scaling to mitigate the effect of sample imbalance. Separate folders for the training and testing

sets, with an 8:2 split. A deep learning algorithm needs a threshold value to tell the difference between interictal and preictal samples, and this value is referred to as the threshold of target detection. The model then passes the score for the desired region via the Softmax layer before declaring prediction complete. To illustrate, if the model score is 90%, it means that the preictal has a 90% chance of being chosen as a true label by the VGG-19 algorithm. During threshold setting, the algorithm favors PAC images with high entropy that are closer to the true label, while rejecting PAC images with lower scores. This results in a model with high precision but low recall. The low threshold is biased to accept low score object frame areas, leading to excellent recall but poor precision. The threshold is chosen to 0.6 after a series of studies to guarantee the model's excellent precision and recall. In order to prevent the loss of data diversity due to overly rapid gradient descent, we use ADAM optimizer. Additionally, we train the network model with 20 epochs, 64 batches per iteration, and a learning rate of 0.00001.

5.5.1 Trainable and Frozen Layers: All Channels

A popular CNN model, VGG-19, was released in 2014 with 19 layers and has shown remarkable performance in many image-related tasks since then. The small-sized filters that are a hallmark of VGG-19 make it ideal for our task of identifying the frequency difference between preictal and interictal states. It consists of a transferred VGG-19 and top trainable layers. Pretraining is done on the ImageNet database with the transferred VGG-19, which has had its output layer removed. Two additional trainable full connection layers and a softmax output layer with two neurons representing preictal and normal brain activity have been added. During training, the network would retain the preset values for the parameters of VGG-19 trained on ImageNet. The rest of the tunable parameters are given arbitrary values to start with.

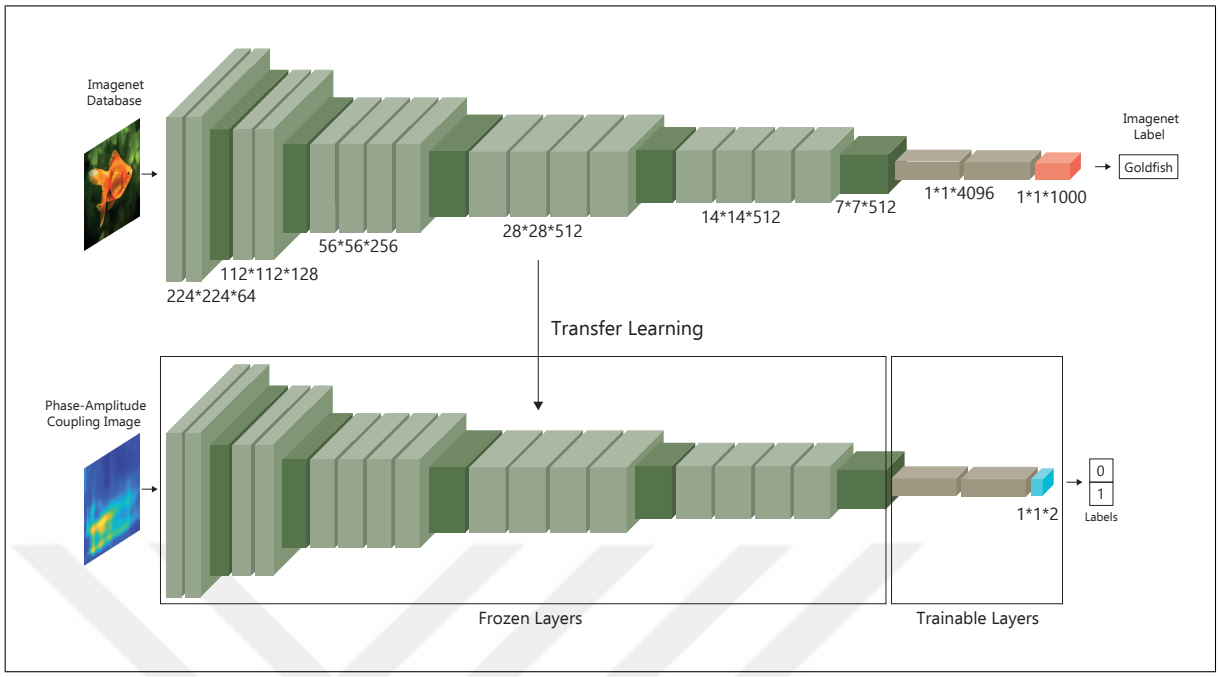

Figure 5.6 The procedure of Transfer Learning with VGG-19, the upper architecture describes the full model of VGG-19 while the lower architecture is the proposed one used in the study with the difference in the softmax layer where we use two classes instead of 1000 classes. The weights were imported from the original VGG-19 model using Transfer Learning.

5.5.2 Fine Tuning: One Channel Variance Based

The selection of channel P7-O1 was based on many CHB studies that build methods for calculating the variance across all channels and subjects. Using their innovative algorithm, JS Ra et al. [73] confirmed that P7-O1 is also the channel with the highest entropy and variance. Our modified VGG-19 model allowed us to ditch the classifier. In addition, we used two additional hidden layers $< 1024, 512 >$, given that the task of epilepsy prediction is bipartite case either preictal or interictal. Each fully-connected layer has had a dropout layer (0.7) appended to it. Moreover, our improved model made use of L2 regularization to lessen the overfitting. In this model, ”softmax" is used as the activation function for the output layer.

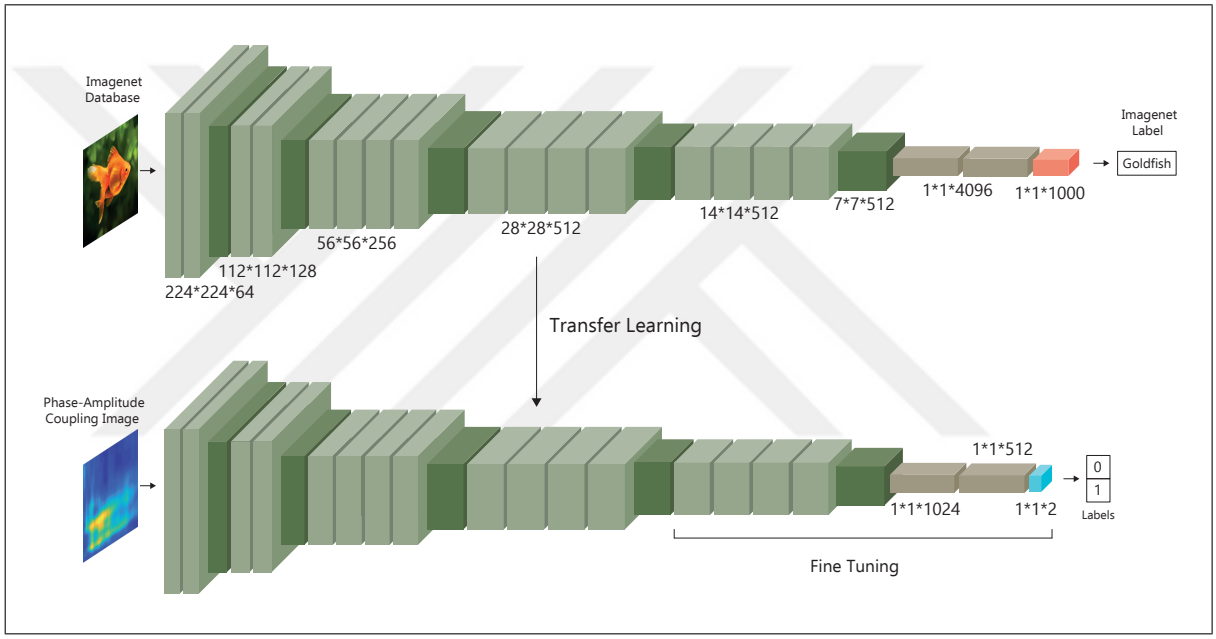

Figure 5.7 The procedure of Transfer Learning with VGG-19, the upper architecture describes the full model of VGG-19 while the lower architecture is the proposed one used in the study with the difference in the softmax layer where we use two classes instead of 1000 classes. The weights were imported from the original VGG-19 model using Transfer Learning.

6. EXPERIMENTAL RESULTS

This chapter clarifies our study's experimental results. Section 6.1 displays the spatiotemporal information of the filtered dataset that will be utilized in subsequent research. In Section 6.2, the results of phase-amplitude coupling images are shown. Section 6.3 demonstrates the two strategies used in transfer learning to improve model metrics. Section 6.4 presents the findings and effects of applying VGG-19 and transfer learning to predict seizures.

6.1 Temporal and Spatial information

In addition to the temporal (changes over time) information depicted in Figure 6.2, EEG also contains spatial information. The signal over the head varies prior to (interictal and preictal) and during an epileptic seizure. When employing the Welch method, we focus on the average power spectral density (PSD) which it divides the data into overlapping segments, generates a modified periodogram for each segment, and averages the periodograms to produce an estimate of the power spectral density, as illustrated in Figure 6.1.

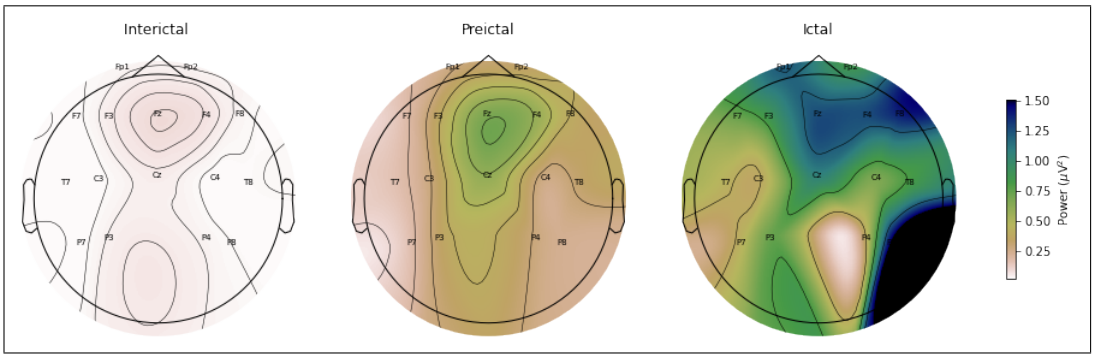

Figure 6.1 Changing in spatial information during three states, interictal, preictal, and ictal respectively.

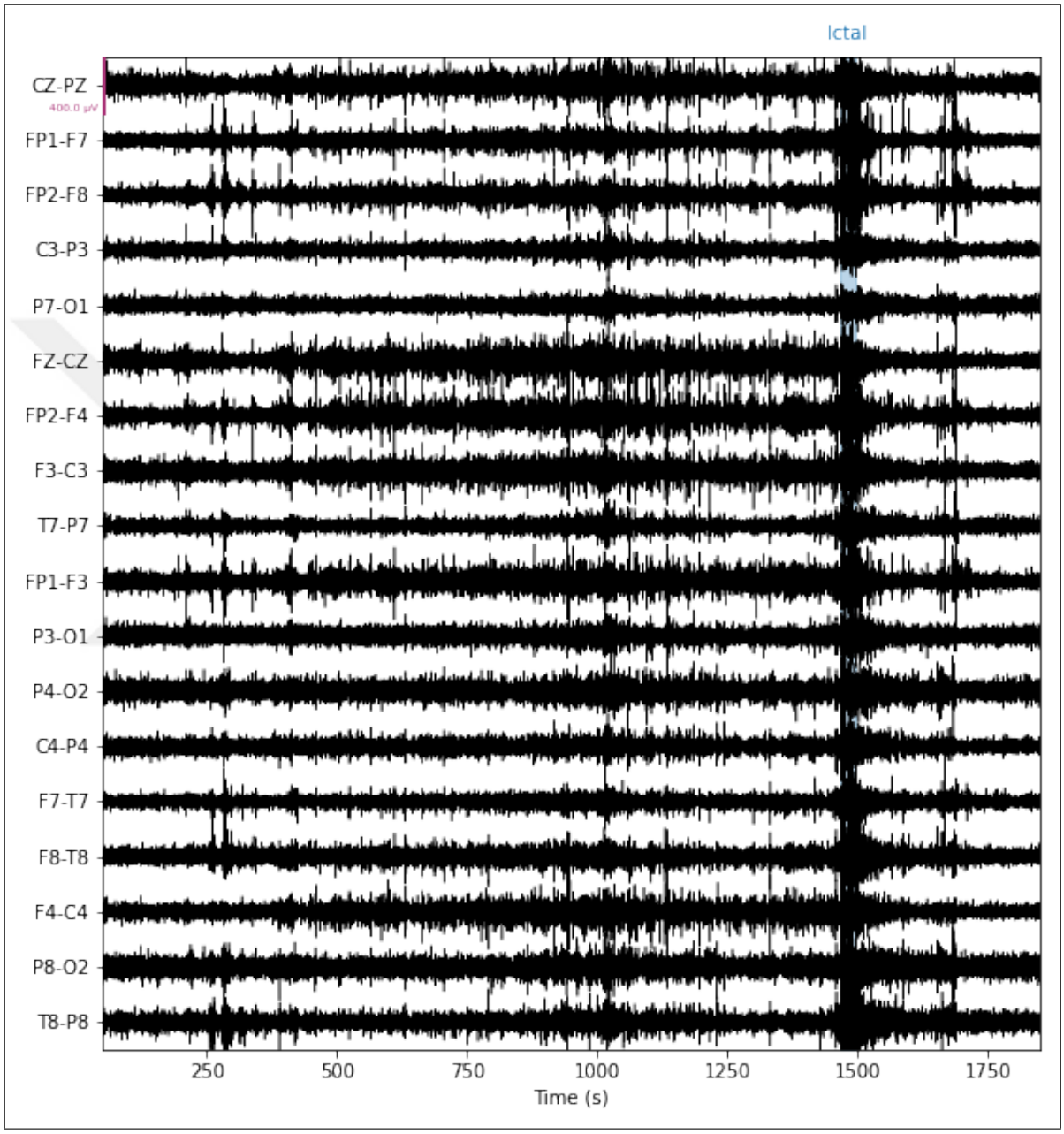

Figure 6.2 Patient *Cb01*'s preictal period is shown in white, and the ictal period is shown in blue on their EEG recording, which spans 30 minutes.

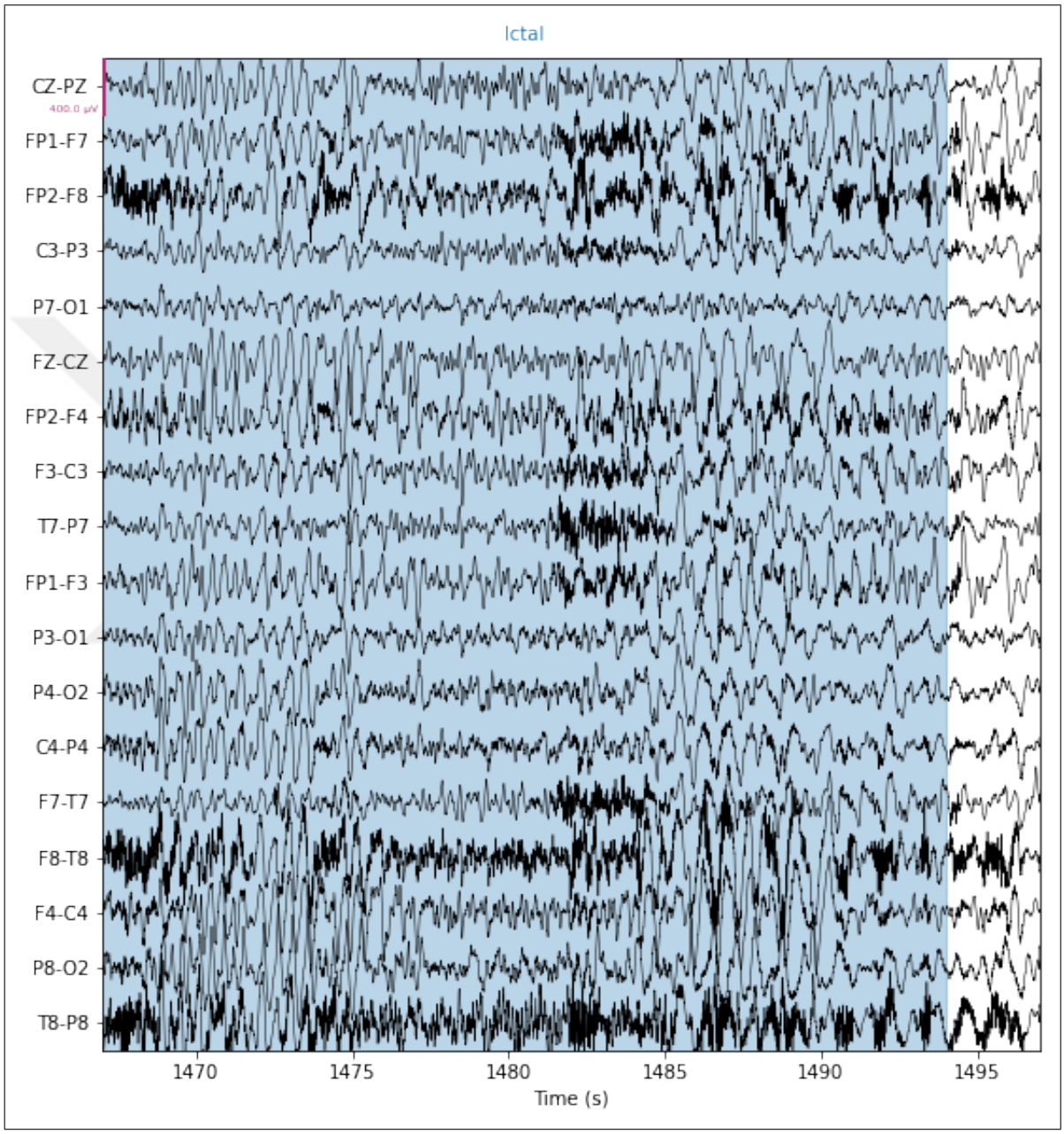

Figure 6.3 Patient *Cb01* as explained from the summery, seizure start time: 1467 seconds and seizure end time: 1494 seconds.

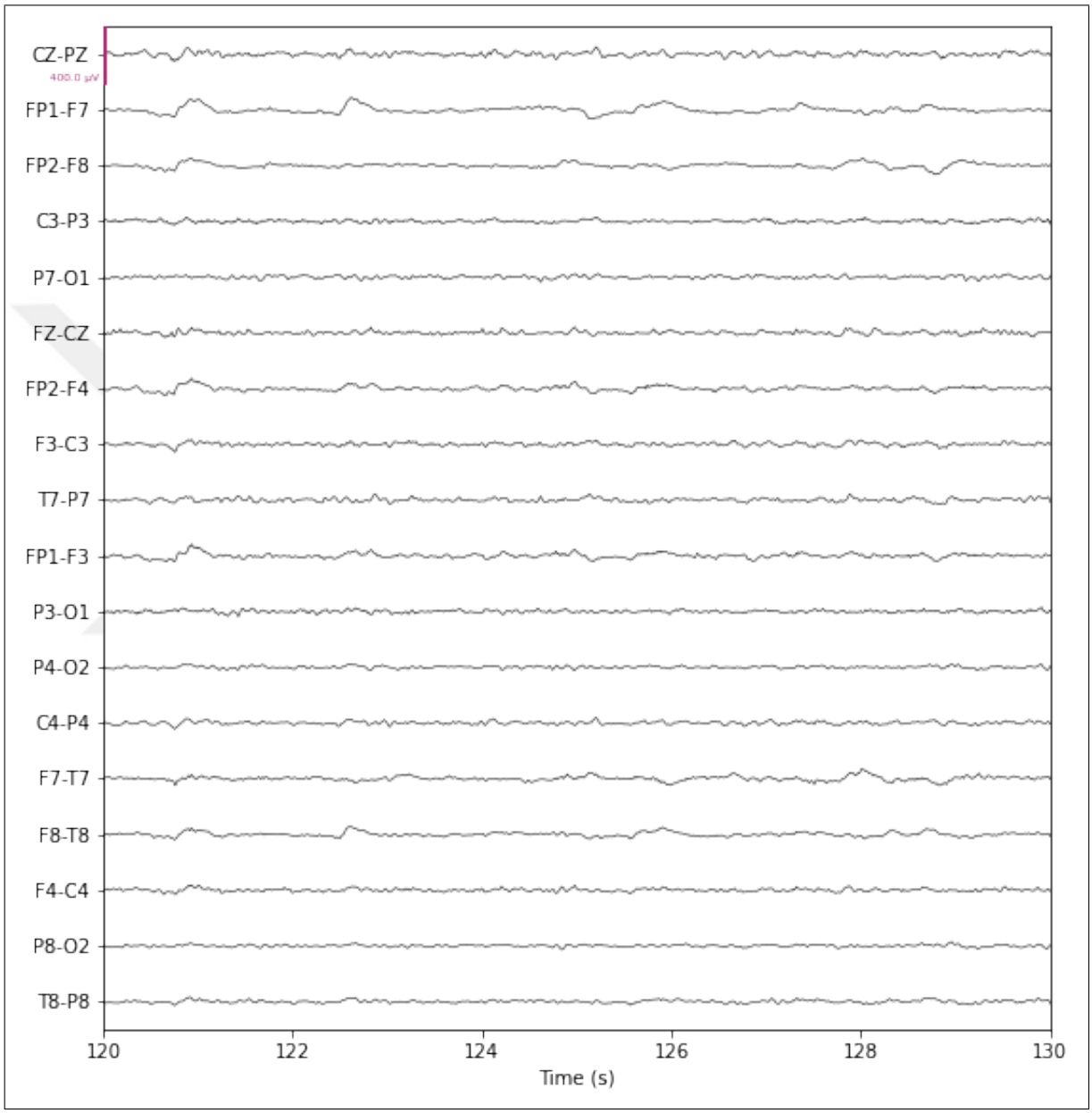

Figure 6.4 Session number 29 for patient *Cb18* in which there is no seizure activity within a 10-second window.

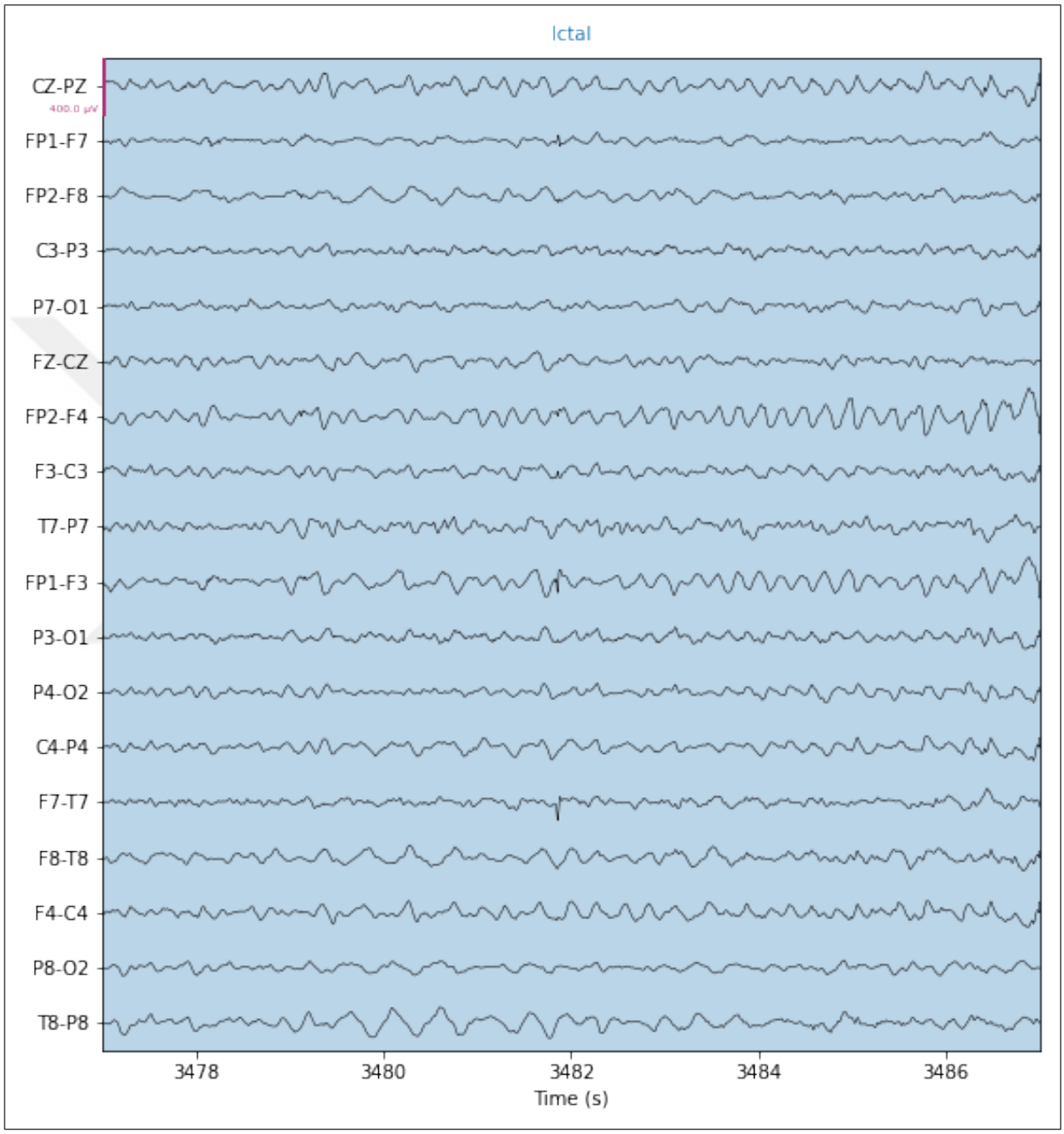

Figure 6.5 Session number 29 for patient *Cb18* in which there is seizure activity within a 10-second window.

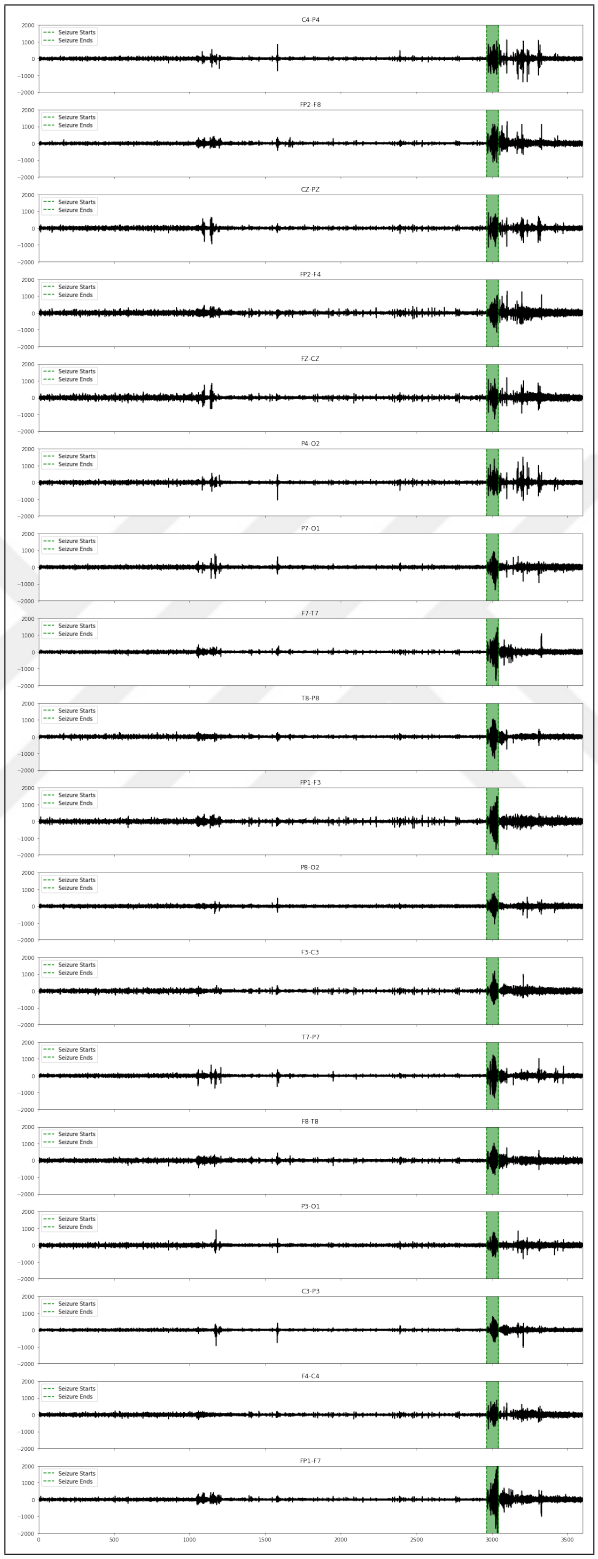

Figure 6.6 Ictal period for the 18 channels used in this study.

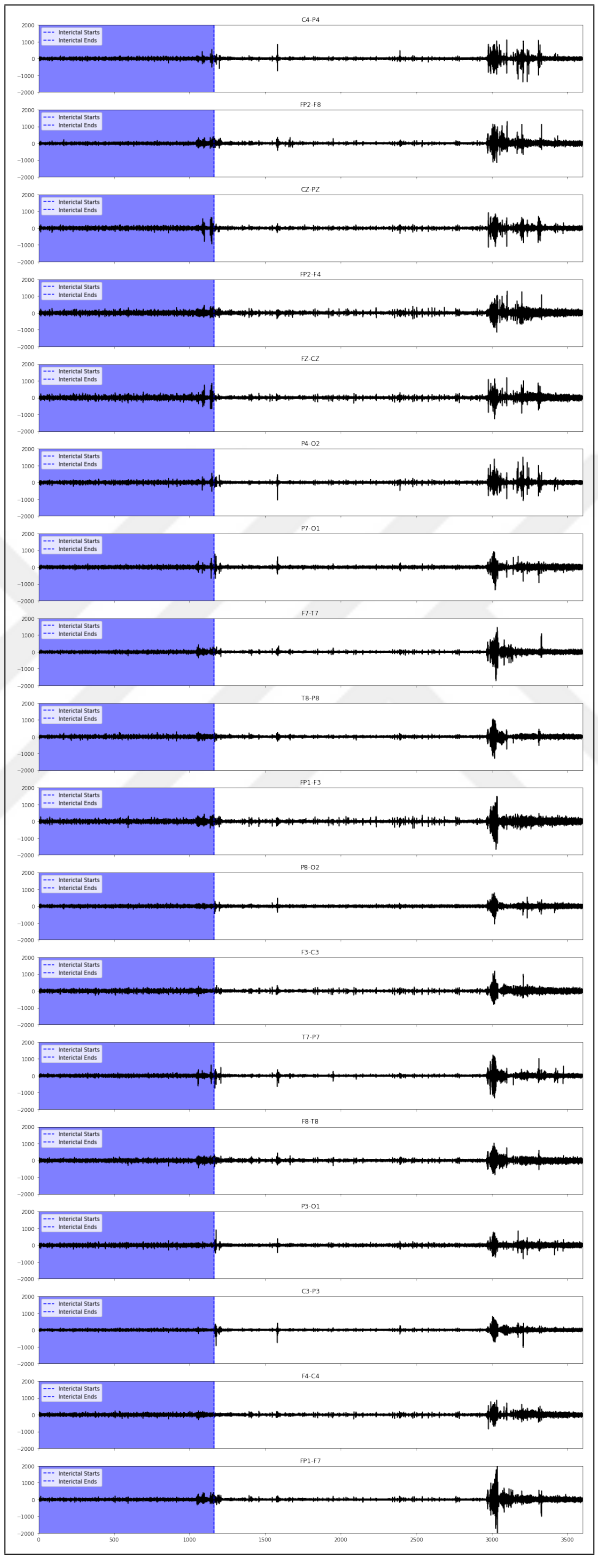

Figure 6.7 Interictal period for the 18 channels used in this study.

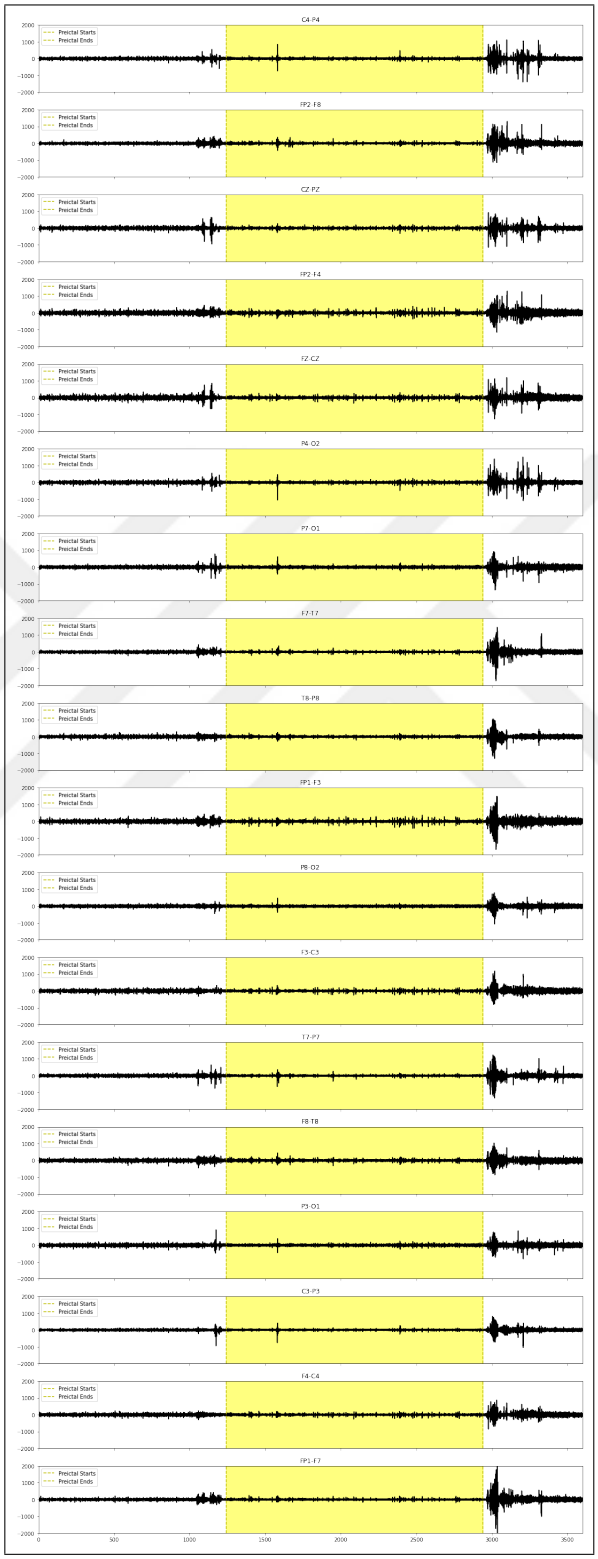

Figure 6.8 Preictal period for the 18 channels used in this study.

6.2 Delta-Gamma Coupling Images

Figure 6.6 shows an enlarged segment of an electroencephalogram (EEG) from the CHBMIT dataset, which was recorded from a patient diagnosed with epilepsy, is shown here with accompanying clinical annotations. The specialist only defines the onset (the start of the ictal state) and the offset (the end of the ictal state). However, in accordance with section 4.3, the preictal and interictal states were tailored in a different manner. The preictal, interictal, and ictal phases are represented by the green, blue, and yellow bars, respectively. All patients in the dataset exhibited delta-gamma profile activity during the preictal state. The sensitivity was 87.6%. On close inspection, the majority of PAC Comodulation images showed a distinct coupling between the delta-gamma frequency, as shown in Figure 6.14. As anticipated, preictal images revealed a substantial amount of activity in the delta-gamma region, as shown in Figure 6.12. Results that have high specificity, such as 89.9%, indicate that there is a good chance that the patient is in a preictal state, which should serve as a warning to the patient about an impending seizure.

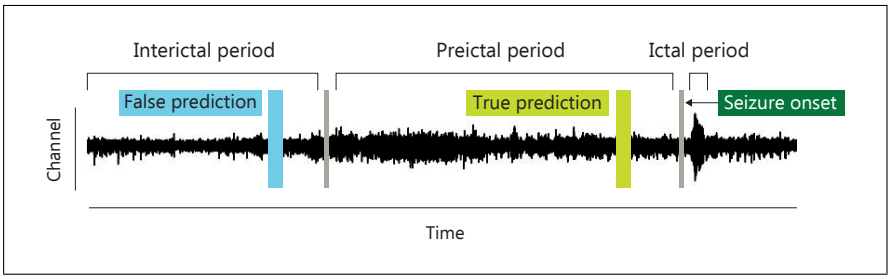

Figure 6.9 Seizure prediction for one channel, clarifying the false prediction and true prediction.

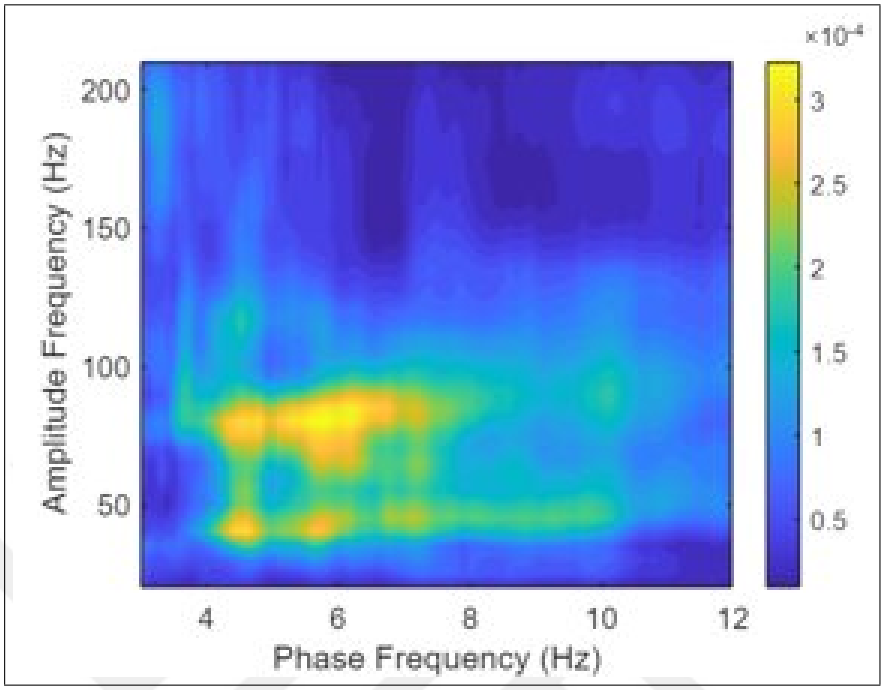

Figure 6.10 Comodulation image of ictal phase.

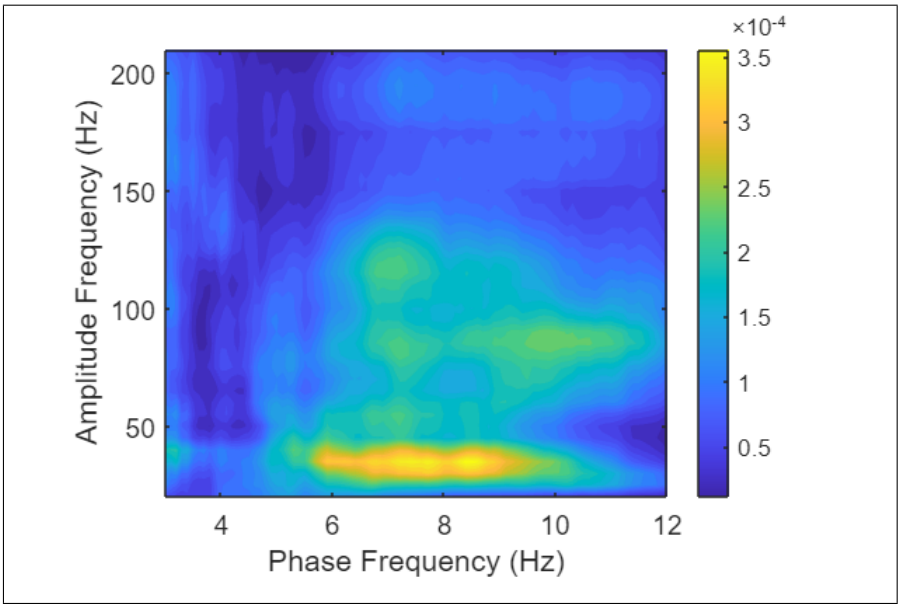

Figure 6.11 Comodulation image of ictal phase.

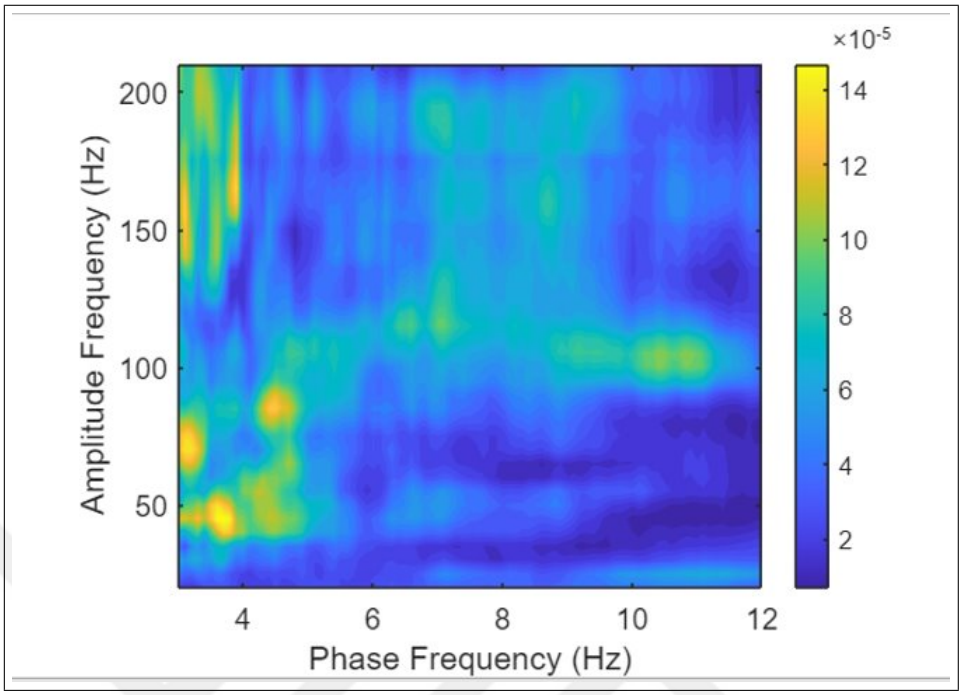

Figure 6.12 Comodulation image of preictal phase.

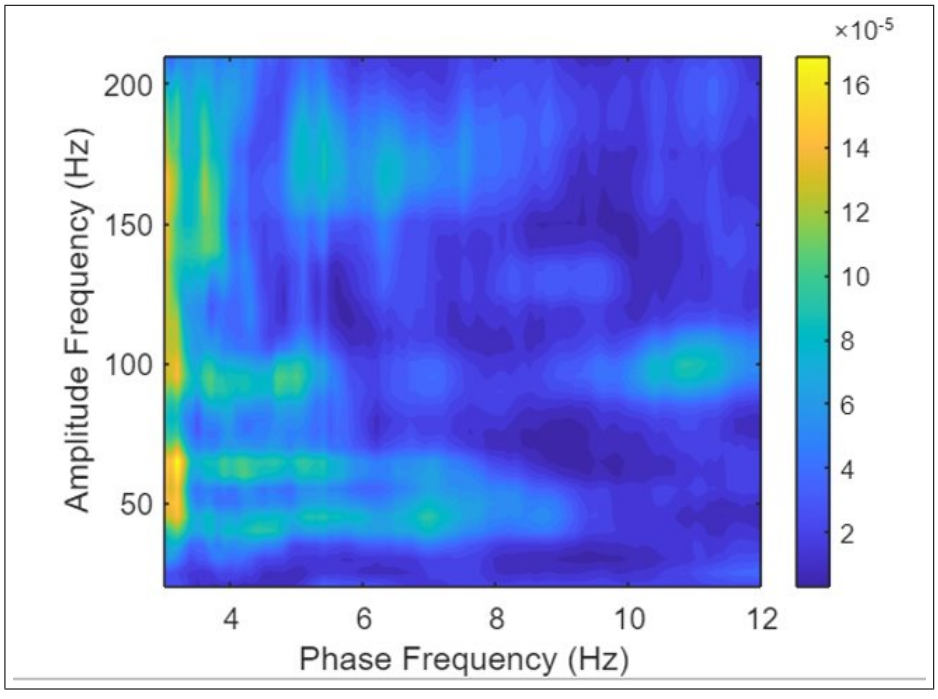

Figure 6.13 Comodulation image of preictal phase

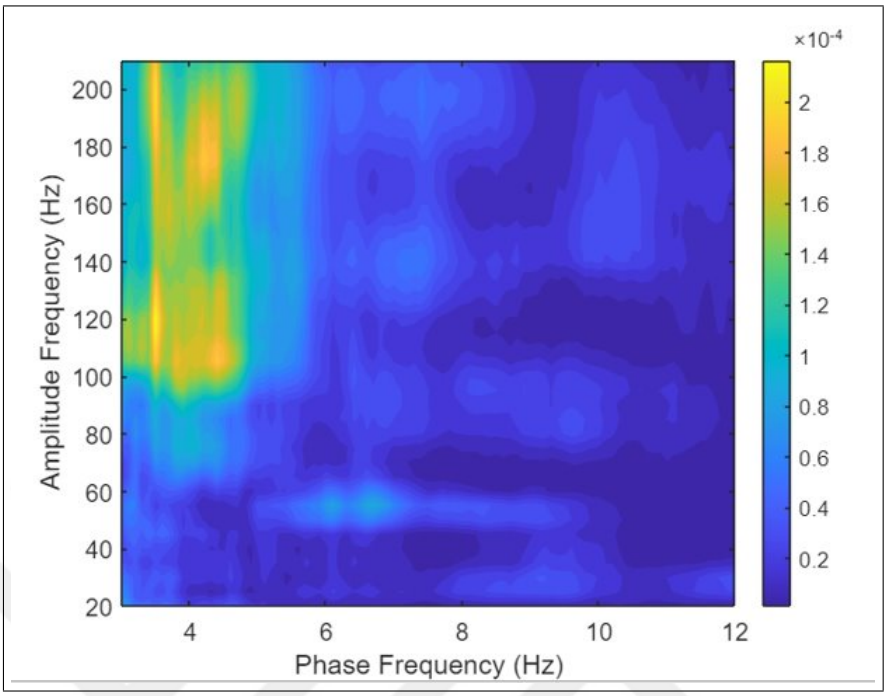

Figure 6.14 Comodulation image of peictal phase.

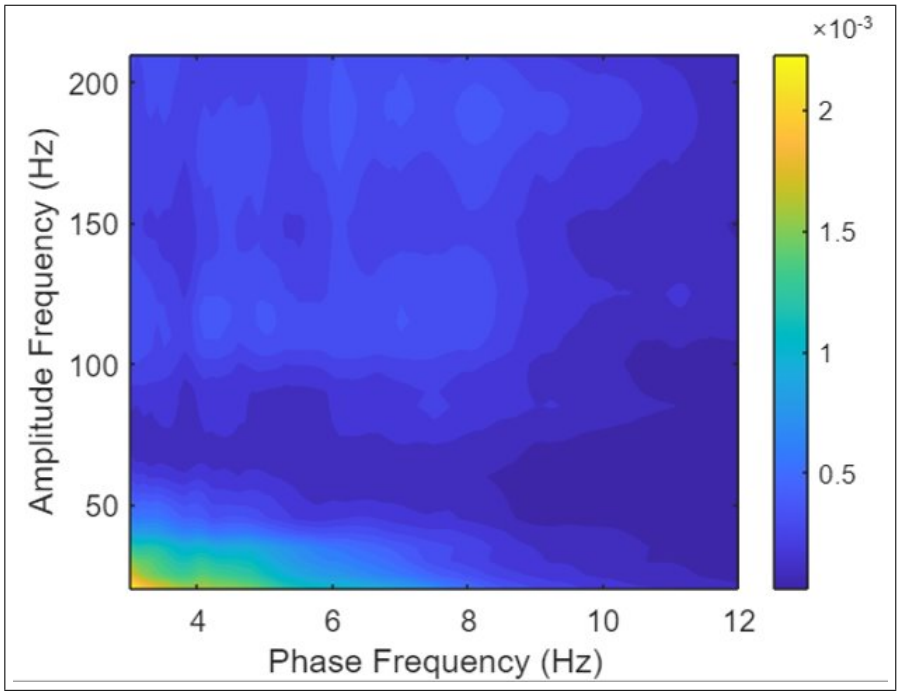

Figure 6.15 Comodulation image of interictal phase.

6.3 VGG-19 Training and Testing

With Google Colab Pro, we use Python to implement transfer learning with the VGG-19. One such library is Keras, which was originally developed for Tensorflow. The number of layers in this model is consistent. Nevertheless, we still set a number of parameters based on the number of datasets used. Initial sample preparation involves splitting the data into a test set and a training set and shuffling them. Secondly, since the network will not be retrained, the trainable parameter has been set to false. Since there are only two classes to distinguish between in this experiment "interictal" and "preictal" the final classification layer is fixed at 2. Initially, the image is scaled to a size of $(224 \times 224 \times 3)$. Both the batch size and the epoch have been set to 64 and 20, respectively.

6.4 VGG-19 Results

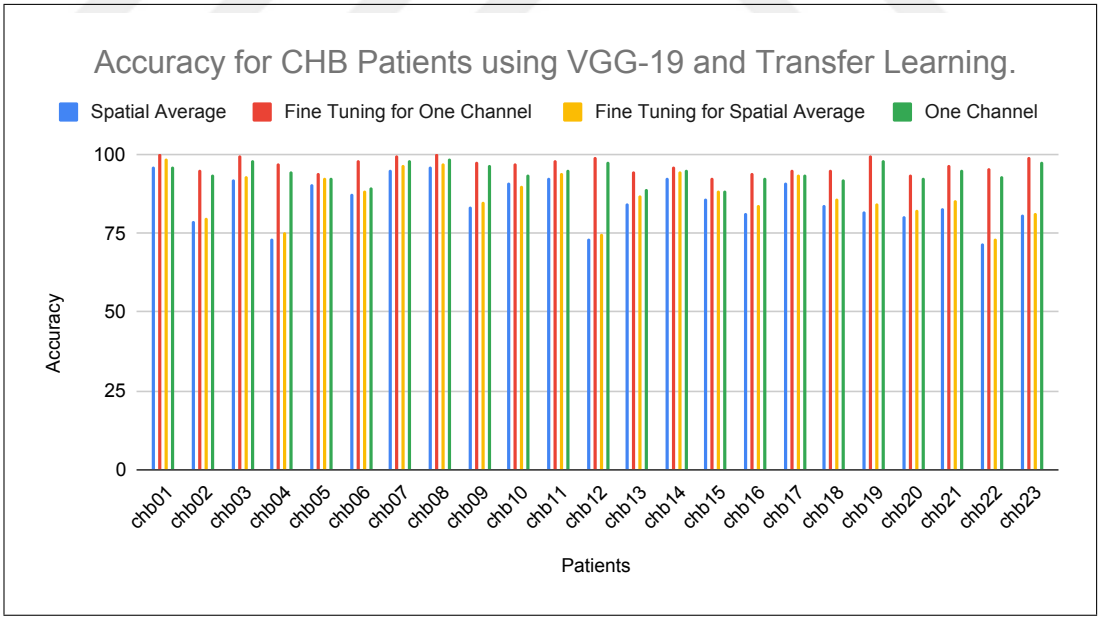

Figure 6.16 Comparison of the accuracy between spatial averaging for the channels with/without fine tuning and the accuracy of one channel with/without fine tuning.

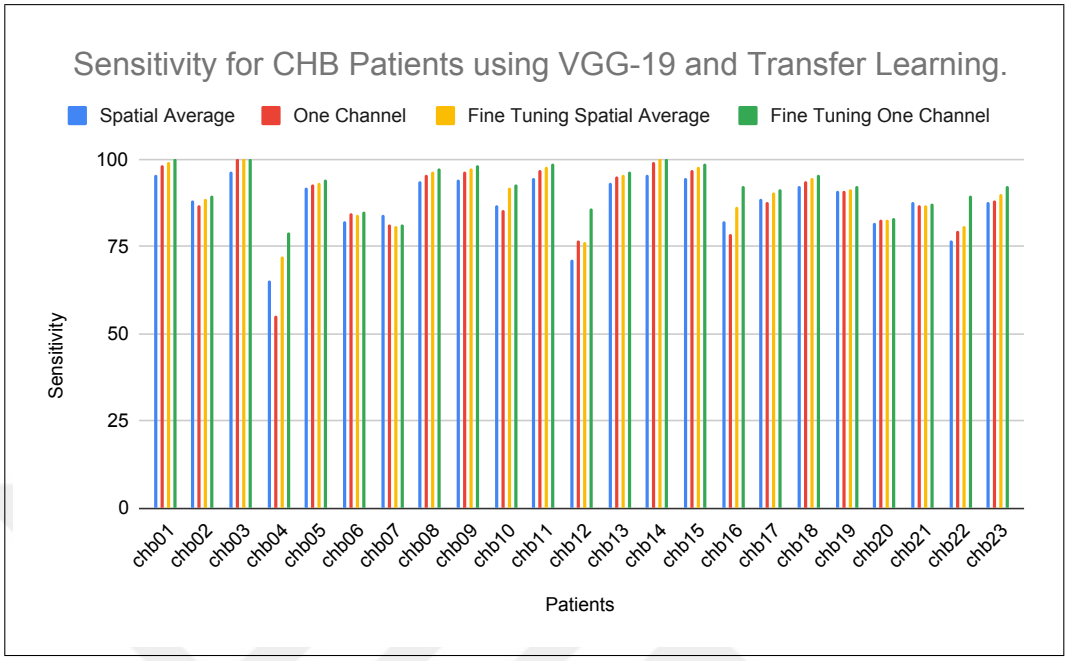

Figure 6.17 Comparison of the sensitivity between spatial averaging for the channels with/without fine tuning and the sensitivity of one channel with/without fine tuning.

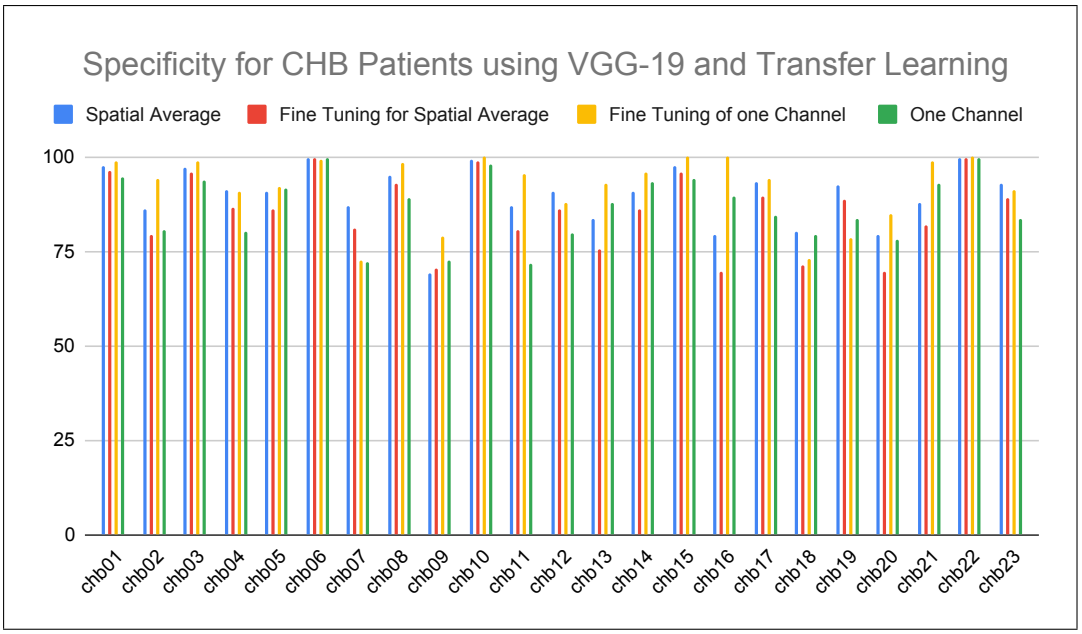

Figure 6.18 Comparison of the specificity between spatial averaging for the channels with/without fine tuning and the specificity of one channel with/without fine tuning.

Table 6.1
Accuracy, sensitivity, specificity, and false positive rate, f1 score values for different patients

Subjects	Accuracy	Sensitivity	Specificity	FPR	F1-score
chb01	96.23	95.3	97.61	0.021	0.94
chb02	78.94	87.97	85.91	0.14	0.869
chb03	91.82	96.3	97.33	0.026	0.967
chb04	73.26	65.46	91.06	0.089	0.687
chb05	90.32	91.82	90.83	0.091	0.946
chb06	87.34	82.3	99.78	0.002	0.902
chb07	95.23	84.21	87.1	0.21	0.89
chb08	96.2	93.6	95.21	0.05	0.835
chb09	83.52	94.11	68.94	0.13	0.916
chb10	91.07	86.99	99.16	0.008	0.879
chb11	92.54	94.41	86.94	0.103	0.922
chb12	73.34	71.2	90.86	0.091	0.778
chb13	84.66	93.27	83.36	0.166	0.901
chb14	92.35	95.72	90.73	0.092	0.936
chb15	86.1	94.39	97.42	0.025	0.953
chb16	81.61	82.4	79.12	0.208	0.817
chb17	91.1	88.45	93.1	0.112	0.876
chb18	83.91	92.5	80.24	0.197	0.877
chb19	82.1	90.8	92.6	0.071	0.906
chb20	80.54	81.97	79.12	0.208	0.817
chb21	83.1	87.87	87.69	0.123	0.876
chb22	71.59	76.8	99.77	0.002	0.756
chb23	80.81	88.8	92.77	0.071	0.906
Average	85.55	87.68	89.85	0.097	0.87

7. DISCUSSION OF RESULTS

In this study, we focused primarily on transfer learning and phase amplitude coupling to create deep learning models for seizure prediction using EEG data. It is fair to say that the predetermined objectives of the research were accomplished.

Despite the computationally intensive nature of the training phase that comes immediately prior to the onset of each seizure, the proposed method is surprisingly uncomplicated: Phase amplitude coupling indices are computed by first performing some simple preprocessing operations, and then performing more advanced operations, such as entropy calculations. The VGG-19 model is being put to use in a practical setting where transfer learning is being employed for predictive purposes. Due to the short duration, real-time applications tailored to the frequency of seizures in pre-operative monitoring may be feasible. Most of the prototypes showcased a minimal number of electrodes as well, which is promising in terms of both power savings and user comfort. However, the present methodology is outperformed in terms of sensitivity when compared to other studies employing data from the same database, which may suggest the requirement for more complex models.

7.1 Phase Amplitude Coupling Images as Epilepsy Biomarker

The hypothesis that phase-amplitude coupling can be viewed as a biomarker for epilepsy will remain untested because the study population consists entirely of subjects with epilepsy, and no healthy controls were included in the analysis.

Alternately, variations in phase-amplitude coupling activity are visible in real-time EEG recordings. According to Figure 6.15, there is no color shift in the comodulation image, indicating that there is no frequency coupling during the interictal stage. Figure 6.14 shows that there is a frequency coupling between the delta (0.5-4 Hz) and gamma (20-150 Hz) bands in the comodulation image during the preictal stage. Tak-

ing into consideration that the colors with the highest degree of warmth signify the strongest coupling. Figure 6.10 shows a comodulation image with a color shift from theta (4-8 Hz) and gamma (20-150 Hz) during the ictal state, indicating an intense frequency coupling.

The amplitude distribution across phases is different in the preictal state compared to the interictal state in terms of Kullback-Leibler divergence. It is the same for ictal situation, which was not included in the computation prior to prepossessing. Color and tone distribution of an image are described by its texture, which is then compared to the image's data via entropy. Our results reveal that entropy increases in preictal compared to interictal images.

7.2 Comparison between Literature and our Proposed Approach

Our research involved 18 EEG scalp channels, 23 pediatric patients with epilepsy, and 182 seizures. The prepossessing involved filtering using the Butterworth filter and then extracting phase-amplitude coupling images using Complex Morlet Continuous Wavelet analysis and Entropy calculations. Transfer learning using the VGG-19 pretrained model was carried out for the prediction portion. The *primary difficulty*, then, was in locating a problem analogous to this one in the same conditions to perform the comparison.

For the sake of comparison, we selected the Ibrahim et al. [74] study in which they extracted spectrogram images from the CHBMIT dataset using Short Time Fourier Transform (STFT) and then used CNN multiple architectures to calculate metrics such as accuracy and specificity. We compared the role of employing complex morlet continuous wavelet and phase amplitude coupling images (comodulogram) with spectogram images under a comparable scenario as the following: For Chb01, Ibrahim et al. study has an accuracy (interictal versus preictal) of 90.35, however our approach has an accuracy of 96.23. Additionally, the specificity of spectogram pictures is 90.44, whereas our approach has a specificity of 97.61, which represents the capacity to accurately identify

negative results. For Chb03, the accuracy of spectogram images is 67.57, whereas the accuracy of our method is 91.82, and the specificity of their method and our method is 67.69 and 97.33, respectively.

7.3 VGG-19 and Transfer Learning Outcome

The VGG-19 deep transfer model, with the output layer swapped out for a new softmax output layer containing just two neurons to represent preictal and interictal states. The three fully connected layers were replaced by two fully connected layers for the two scenarios. While in the first scenario we accounted for the average all channels in our analysis, in the second we focused on the effect of a single channel. The research found that using only one channel which is (P7-O1) resulted in a higher accuracy 94.30 than using the average of all channels did 85.55.

Table 7.1
The average metrics for using VGG-19 and Transfer Learning.

VGG-19 and Transfer Learning	Avg. Accuracy	Avg. Sensitivity	Avg. Specificity
Spatial Averaging	85.55%	87.62%	89.85%
One Channel	94.30%	88.17%	85.69%

Table 7.2
The average metrics for using VGG-19, Transfer learning and Fine Tuning.

VGG-19, TL and Fine Tuning	Avg. Accuracy	Avg. Sensitivity	Avg. Specificity
Spatial Averaging	87.21%	90.17%	86.49%
One Channel	96.82%	92.16%	92.03%

The first strategy was to freeze the model and make the fully connected layers flexible for training as seen in Figure 5.6. The second strategy was utilizing fine-tuning to enhance the model's accuracy. Unfreezing the upper layers of the convolutional base and retraining the network with a low learning rate constitutes fine-tuning as seen in Figure 5.7. When evaluated based on the average of all channels, these strategies increased accuracy from 85.55 to 87.21, and when evaluated based on one-channel

(P7-O1) variance, they increased accuracy from 94.30 to 96.82.

8. CONCLUSION

With the help of a transfer learning model, we tested the VGG-19 network's performance in predicting EEG seizures across a variety of input channels, from a single channel to a full multi-channel data set. When compared to other models, ours had much better test accuracy, sensitivity, specificity, and training time. We found that by using a transfer learning approach with VGG-19, we could greatly reduce the training time required for the model without sacrificing its performance.

After a deep neural network has been trained, we propose exploring post-processing with a matching network architecture inspired by metric learning; preliminary statistics show encouraging results. We updated a strategy for modifying the histogram of network distance metrics in order to make it resemble the distributions that are observed in statistical models. At the moment, we have developed a metric learning architecture to evaluate training results from intermediate layer output. However, in order to analytically compare the histograms, we require a fitting method. We developed an automated method for verifying a compatibility network. The learned model's statistical results showed that the distributions of metric distances were very different during seizure and non-seizure times.

If we want to create a class-discriminating algorithm that can adapt to new data, we need to learn more about how the epochs' distribution behaviors change in response to the labels. We updated the procedure so that it takes maps showing low and high frequencies of images as inputs. Here, we take an image-based approach to dealing with EEG signals. By treating multi-channel signals like images, we were able to achieve significant gain. We are committed to developing more effective strategies for learning the signal images in order to optimize training while minimizing computational expenses. This will provide us with an effective tool for the development of more in-depth seizure prediction systems.

REFERENCES

1. Nowack, W. J., "Book review: Epilepsy: Frequency, causes, and consequences, by w. allen hauser and dale c. hesdorffer. published in 1990 by demos publications, new york, 378 pages," *Journal of Child Neurology*, Vol. 6, no. 4, pp. 374–374, 1991.

2. Fisher, R. S., "What clinicians want to know from epilepsy researchers," *Epilepsia*, Vol. 50, no. 3, pp. 364–367, 2009.

3. Berg, A. T., and J. J. Millichap, "The 2010 Revised Classification of Seizures and Epilepsy," *CONTINUUM: Lifelong Learning in Neurology*, Vol. 19, pp. 571–597, 2013.

4. Stafstrom, C. E., and L. Carmant, "Seizures and epilepsy: an overview for neuroscientists," *Cold Spring Harbor perspectives in medicine*, Vol. 5, no. 6, 2015.

5. Kuhlmann, L., K. Lehnertz, M. P. Richardson, B. Schelter, and H. P. Zaveri, "Seizure prediction — ready for a new era," *Nature Reviews Neurology*, Vol. 14, no. 10, pp. 618–630, 2018.

6. Sharmila, A., and P. Geethanjali, "A review on the pattern detection methods for epilepsy seizure detection from eeg signals," *Biomedical Engineering / Biomedizinische Technik*, Vol. 64, no. 5, pp. 507–517, 2019.

7. Li, J., J. Yan, X. Liu, and G. Ouyang, "Using permutation entropy to measure the changes in eeg signals during absence seizures," *Entropy*, Vol. 16, no. 6, pp. 3049–3061, 2014.

8. Adeli, H., Z. Zhou, and N. Dadmehr, "Analysis of eeg records in an epileptic patient using wavelet transform," *Journal of Neuroscience Methods*, Vol. 123, no. 1, pp. 69–87, 2003.

9. Ocak, H., "Automatic detection of epileptic seizures in eeg using discrete wavelet transform and approximate entropy," *Expert Systems with Applications*, Vol. 36, no. 2, Part 1, pp. 2027–2036, 2009.

10. Nakate, A., and P. Bahirgonde, "Feature Extraction of EEG Signal using Wavelet Transform," *International Journal of Computer Applications*, Vol. 124, no. 2, pp. 21–24, 2015.

11. Caspers, H., *Mechanisms of EEG Generation — Historical and Present Aspects*, pp. 1–11. Boston, MA: Birkhäuser Boston, 1993.

12. Im, C.-H., *Basics of EEG: Generation, Acquisition, and Applications of EEG*, pp. 3–11. Singapore: Springer Singapore, 2018.

13. Kutluay, E., and G. P. Kalamangalam, "Montages for Noninvasive EEG Recording," *Journal of Clinical Neurophysiology*, Vol. 36, pp. 330–336, 9 2019.

14. Jung, K.-Y., *Analysis of EEG in Medically Intractable Epilepsy*, pp. 177–192. Singapore: Springer Singapore, 2018.

15. Werz, M. A., and I. L. Pita, *Epilepsy Syndromes*, Saunders, 2011.

16. Anderson, K., N. , and C. Resource, *Seizures and Epilepsy Syndromes*, 2019.

17. Thijs, R. D., R. Surges, T. J. O'Brien, and J. W. Sander, "Epilepsy in adults," *The Lancet*, Vol. 393, no. 10172, pp. 689–701, 2019.

18. Wilson, J. V. K., and E. H. Reynolds, "Translation and analysis of a cuneiform text forming part of a babylonian treatise on epilepsy," *Medical History*, Vol. 34, no. 2, 1990.

19. Kaculini, C. M., A. J. Tate-Looney, and A. Seifi, "The History of Epilepsy: From Ancient Mystery to Modern Misconception," *Cureus*, 2021.

20. Bladin, P. F., and M. J. Eadie, *Medical Aspects of the History of Epilepsy*, pp. 27–33. London: Springer London, 2010.

21. Jiang, Y., and X. Zhou, "Frontal lobe epilepsy manifesting as vertigo: a case report and literature review," *Journal of International Medical Research*, Vol. 48, no. 9, 2020. PMID: 32877235.

22. Lidz, R. W., "Hysterical seizures," *Psychosomatics*, Vol. 22, pp. 805–821, 9 1981.

23. Eadie, M., "The origin of the concept of partial epilepsy," *Journal of Clinical Neuroscience*, Vol. 6, no. 2, pp. 103–105, 1999.

24. Jenkins, J., "Dr samuel auguste tissot," *Journal of Medical Biography*, Vol. 7, no. 4, pp. 187–191, 1999.

25. Macia, F., and Y. Ardagna, "Historical reflections on neurosyphilis based on the 1826 treatise on general paralysis in demented patients by louis florentin calmeil (1798–1895)," *Revue Neurologique*, Vol. 174, no. 4, pp. 247–254, 2018.

26. Pavone, P., A. Polizzi, S. D. Marino, G. Corsello, R. Falsaperla, S. Marino, and M. Ruggieri, "West syndrome: a comprehensive review," *Neurological Sciences*, Vol. 41, no. 12, pp. 3547–3562, 2020.

27. Hogan, R. E., and K. Kaiboriboon, "The âdreamy stateâ: John hughlings-jacksonâs ideas of epilepsy and consciousness," *American Journal of Psychiatry*, Vol. 160, no. 10, pp. 1740–1747, 2003.

28. Pearce, J., "Théodore herpin: neglected contributions in the understanding of epilepsy," *European Neurology*, Vol. 54, no. 3, pp. 135–139, 2005.

29. Reynolds, E. H., "Robert bentley todd's electrical theory of epilepsy," 2005.

30. Eadie, M., "Louis françois bravais and jacksonian epilepsy," *Epilepsia*, Vol. 51, no. 1, pp. 1–6, 2010.

31. Magiorkinis, E., A. Diamantis, K. Sidiropoulou, and C. Panteliadis, "Highights in the history of epilepsy: the last 200 years," *Epilepsy Research and Treatment*, 2014.

32. Engel, J., "Introduction to temporal lobe epilepsy," *Epilepsy Research*, Vol. 26, no. 1, pp. 141–150, 1996. Mechanisms of Chronic Models of Epilepsy.

33. Scheffer, I. E., S. Berkovic, G. Capovilla, M. B. Connolly, J. French, L. Guilhoto, E. Hirsch, S. Jain, G. W. Mathern, S. L. Moshé, D. R. Nordli, E. Perucca, T. Tomson, S. Wiebe, Y. Zhang, and S. M. Zuberi, "ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology," *Epilepsia*, Vol. 58, no. 4, pp. 512–521, 2017.

34. St. Louis, E., W. Rosenfeld, and T. Bramley, "Antiepileptic Drug Monotherapy: The Initial Approach in Epilepsy Management," *Current Neuropharmacology*, Vol. 7, no. 2, pp. 77–82, 2009.

35. Ellis, N., D. Upton, and P. Thompson, "Epilepsy and the family: a review of current literature," *Seizure*, Vol. 9, no. 1, pp. 22–30, 2000.

36. Engel Jr, J., *Seizures and Epilepsy*, Vol. 83, Oxford University Press, 2013.

37. Semmlow, J., "Chapter 1 - the big picture: Bioengineering signals and systems," in *Circuits, Signals and Systems for Bioengineers (Third Edition)* (Semmlow, J., ed.), Biomedical Engineering, pp. 3–50, Academic Press, third edition ed., 2018.

38. EscabÃ, M., "Chapter 11 - biosignal processing," in *Introduction to Biomedical Engineering (Third Edition)* (Enderle, J. D., and J. D. Bronzino, eds.), Biomedical Engineering, pp. 667–746, Boston: Academic Press, third edition ed., 2012.

39. *Fundamentals of EEG Signal Processing*, ch. 4, pp. 77–113. John Wiley Sons, Ltd, 2021.

40. Kanoga, S., and Y. Mitsukura, "Review of artifact rejection methods for electroencephalographic systems," in *Electroencephalography* (Sittiprapaporn, P., ed.), ch. 6, Rijeka: IntechOpen, 2017.

41. Miner, N. E., "An introduction to wavelet theory and analysis," 1998.

42. Chan, Y., *Wavelet basics*, Springer Science & Business Media, 1994.

43. Akin, M., "Comparison of wavelet transform and fft methods in the analysis of eeg signals," *Journal of medical systems*, Vol. 26, no. 3, pp. 241–247, 2002.

44. Cancino-De-Greiff, H. F., R. Ramos-Garcia, and J. V. Lorenzo-Ginori, "Signal de-noising in magnetic resonance spectroscopy using wavelet transforms," *Concepts in Magnetic Resonance: An Educational Journal*, Vol. 14, no. 6, pp. 388–401, 2002.

45. Avanzo, C. D., V. Tarantino, P. Bisiacchi, and G. Sparacino, "A wavelet methodology for eeg time-frequency analysis in a time discrimination task," *International Journal of Bioelectromagnetism*, Vol. 11, no. 4, pp. 185–188, 2009.

46. Addison, P., J. Watson, and T. FENG, "Low-oscillation complex wavelets," *Journal of Sound and Vibration*, Vol. 254, no. 4, pp. 733–762, 2002.

47. Li, X., X. Yao, J. Jefferys, and J. Fox, "Computational neuronal oscillations using morlet wavelet transform," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp. 2009–2012, IEEE, 2006.

48. De Moortel, I., S. Munday, and A. W. Hood, "Wavelet analysis: the effect of varying basic wavelet parameters," *Solar Physics*, Vol. 222, no. 2, pp. 203–228, 2004.

49. Lee, D. T., and A. Yamamoto, "Wavelet analysis: theory and applications," *Hewlett Packard journal*, Vol. 45, pp. 44–44, 1994.

50. Lucasius, C., *Noninvasive EEG-Based Prediction of Infantile Spasms Using Deep Learning Networks with Phase Amplitude Cross-Frequency Coupling.* PhD thesis, University of Toronto (Canada), 2020.

51. Shoeb, A. H., and J. V. Guttag, "Application of machine learning to epileptic seizure detection," in *ICML*, 2010.

52. Fisher, R. S., H. E. Scharfman, and M. deCurtis, "How Can We Identify Ictal and Interictal Abnormal Activity?," *Issues in Clinical Epileptology: A View from the Bench*, pp. 3–23, 2014.

53. De Cheveigne, A., and I. Nelken, "Filters: When, why, and how (not) to use them," *Neuron*, Vol. 102, no. 2, pp. 280–293, 2019.

54. Kheirati Roonizi, A., and C. Jutten, "Forward-backward filtering and penalized least-Squares optimization: A Unified framework," *Signal Processing*, Vol. 178, p. 107796, 1 2021.

55. Tort, A. B. L., R. Komorowski, H. Eichenbaum, and N. Kopell, "Measuring phase-amplitude coupling between neuronal oscillations of different frequencies," *Journal of Neurophysiology*, Vol. 104, no. 2, pp. 1195–1210, 2010. PMID: 20463205.

56. Jensen, O., and L. L. Colgin, "Cross-frequency coupling between neuronal oscillations," *Trends in Cognitive Sciences*, Vol. 11, no. 7, pp. 267–269, 2007.

57. Jiruska, P., C. Alvarado-Rojas, C. A. Schevon, R. Staba, W. Stacey, F. Wendling, and M. Avoli, "Update on the mechanisms and roles of high-frequency oscillations in seizures and epileptic disorders," *Epilepsia*, Vol. 58, no. 8, pp. 1330–1339, 2017.

58. Levesque, M., P. Salami, J. Gotman, and M. Avoli, "Two Seizure-Onset Types Reveal Specific Patterns of High-Frequency Oscillations in a Model of Temporal Lobe Epilepsy," *Journal of Neuroscience*, Vol. 32, no. 38, pp. 13264–13272, 2012.

59. Onslow, A. C., R. Bogacz, and M. W. Jones, "Quantifying phaseâamplitude coupling in neuronal network oscillations," *Progress in Biophysics and Molecular Biology*, Vol. 105, no. 1, pp. 49–57, 2011. BrainModes: The role of neuronal oscillations in health and disease.

60. Grigorovsky, V., D. Jacobs, and B. L. Bardakjian, "Delta-gamma phase-amplitude coupling as a biomarker of postictal generalized eeg suppression," *Brain Communications*, Vol. 2, no. 2, 2020.

61. Canolty, R. T., and R. T. Knight, "The functional role of cross-frequency coupling," *Trends in Cognitive Sciences*, Vol. 14, pp. 506–515, 11 2010.

62. Berman, J. I., S. Liu, and L. Bloy, "Alpha-to-Gamma Phase-Amplitude Coupling Methods and Application to Autism Spectrum Disorder," *Brain Connectivity*, Vol. 5, no. 2, pp. 80–90, 2015.

63. Goodman, Kumar, S., and R. Zomorrodi, "Theta-Gamma Coupling and Working Memory in Alzheimer's Dementia and Mild Cognitive Impairment," *Frontiers in Aging Neuroscience*, Vol. 10, 2018.

64. Gong, R. Wegscheider, M., and Mühlberg, "Spatiotemporal features of beta-gamma phase-amplitude coupling in Parkinson‘s disease derived from scalp EEG," *Brain*, Vol. 144, no. 2, pp. 487–503, 2020.

65. Pan, S. J., "Transfer learning," *Learning*, Vol. 21, pp. 1–2, 2020.

66. Senator, T. E., "Transfer learning progress and potential," *AI Magazine*, Vol. 32, no. 1, pp. 84–84, 2011.

67. Sony, S., K. Dunphy, A. Sadhu, and M. Capretz, "A systematic review of convolutional neural network-based structural condition assessment techniques," *Engineering Structures*, Vol. 226, 2021.

68. Liu, M., J. Shi, Z. Li, C. Li, J. Zhu, and S. Liu, "Towards better analysis of deep convolutional neural networks," *IEEE transactions on visualization and computer graphics*, Vol. 23, no. 1, pp. 91–100, 2016.

69. Khan, A., A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial intelligence review*, Vol. 53, no. 8, pp. 5455–5516, 2020.

70. Simonyan, K., and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv*, 2014.

71. Liu, Z., J. Li, Z. Shen, G. Huang, S. Yan, and C. Zhang, "Learning efficient convolutional networks through network slimming," in *Proceedings of the IEEE international conference on computer vision*, pp. 2736–2744, 2017.

72. Jeong, S., J. Yun, H.-G. Yeom, H. Lim, J. Lee, and B. C. Kim, "Deep learning based discrimination of soft tissue profiles requiring orthognathic surgery by facial photographs," *Scientific Reports*, Vol. 10, 2020.

73. Ra, J. S., T. Li, and Y. Li, "A novel permutation entropy-based eeg channel selection for improving epileptic seizure prediction," *Sensors*, Vol. 21, no. 23, pp. 79–72, 2021.

74. Ibrahim, F. E., H. M. Emara, and W. e. a. Elshafai, "Deep learning based seizure detection and prediction from electroencephalography signals," *International Journal for Numerical Methods in Biomedical Engineering*, Vol. 38, no. 6, 2022.