

Module No:	<b>COMP6013</b>	Module title:	<b>Bsc Computing Project</b>
Degree Programme :		<b>Bsc (Hons) Computer Science</b>	
Project title :		<b>Real-time preictal detection through the application of machine learning to Electroencephalogram signals.</b>	
Supervisor :		<b>Kashinath Basu</b>	
Due date and time:		<b>19 Apr 2024 - 13:00</b>	
Estimated total time to be spent on assignment:			<b>90 hours per student</b>
<b>Student No:</b>		<b>Student Name:</b>	
<b>19066041</b>		<b>William Riddell</b>	

**Statement of Compliance (please tick to sign)**

☐

I declare that the work submitted is my own and that the work I submit is fully in accordance with the University regulations regarding assessments ([www.brookes.ac.uk/uniregulations/current](http://www.brookes.ac.uk/uniregulations/current))

**Regulations governing the deposit and use of Oxford Brookes University Projects and Dissertations**

Copies of projects/dissertations, submitted in fulfilment of Modular Programme requirements and achieving marks of 60% or above, shall normally be kept by the Library.

**Statement of Permission (please tick to sign)**

☐

	<p><b>Thank you for agreeing to have your work (i.e. dissertation and associated deliverables) deposited in the Repository's Undergraduate eDissertation collection.</b></p> <p>By depositing this work, you agree to the following Terms and Conditions:</p> <p>You confirm that:</p> <ul style="list-style-type: none"> <li>a. you are the copyright owner and/or have the right to grant us licence to hold your work in our institutional repository (currently RADAR).</li> </ul> <p>We agree to:</p> <ul style="list-style-type: none"> <li>b. add the work to the COMP6013 Collection so that it is available to Oxford Brookes University members for the lifetime of the institutional repository.</li> <li>c. convert the work if necessary for long term preservation.</li> </ul> <p>We reserve the right to remove third party copyright material that we may identify in the work. We reserve the right to remove the work for any legal or administrative reason.</p>
--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Use of AI Tools**

You are required to use this [form](#) to declare which AI tools you have used and how you have used them. Please complete the form and attach it to your submission as an Appendix, if you have used such tools.

## LEARNING OUTCOMES

<b>On successful completion of this module, students will be able to achieve the module following learning outcomes (LOs):</b>	
1	Create, design, manage, plan, carry out, and evaluate a project involving the solution of a practical problem set in an appropriate social and economic context, taking into account other relevant factors such as risk
2	Apply practical and analytical skills acquired in the programme to the investigation of a substantial topic
3	Apply the scientific method and report findings using accepted formalisms
4	Identify and utilise trustworthy information sources, such as the ACM Digital Library to develop a coherent understanding of issues in the domain
5	Demonstrate the ability to carry out a substantial piece of work independently and critically evaluate the student's achievements and their own personal development
6	Use appropriate technologies such as online libraries and databases to find, critically evaluate and utilise both non-specialist and technical information pertinent to the project
7	Demonstrate an awareness of and work in a manner guided by the legal, professional, ethical, security and social issues relevant to the IT and telecommunications industry

<b>Engineering Council AHEP4 LOs assessed (from S1 2022-23):</b>	
<b>B3</b>	Select and apply appropriate computational and analytical techniques to model broadly-defined problems, recognising the limitations of the techniques employed
<b>B4</b>	Select and evaluate technical literature and other sources of information to address broadly-defined problems
<b>B5</b>	Design solutions for broadly-defined problems that meet a combination of societal, user, business and customer needs as appropriate. This will involve consideration of applicable health & safety, diversity, inclusion, cultural, societal, environmental and commercial matters, codes of practice and industry standards
<b>B6</b>	Apply an integrated or systems approach to the solution of broadly-defined problems
<b>B7</b>	Evaluate the environmental and societal impact of solutions to broadly-defined problems
<b>B8</b>	Identify and analyse ethical concerns and make reasoned ethical choices informed by professional codes of conduct
<b>B9</b>	Use a risk management process to identify, evaluate and mitigate risks (the effects of uncertainty) associated with a particular project or activity
<b>B10</b>	Adopt a holistic and proportionate approach to the mitigation of security risks
<b>B13</b>	Select and apply appropriate materials, equipment, engineering technologies and processes
<b>B15</b>	Apply knowledge of engineering management principles, commercial context, project management and relevant legal matters
<b>B17</b>	Communicate effectively with technical and non-technical audiences

### **FORMATIVE FEEDBACK OPPORTUNITIES**

**Your supervisor will give you the following formative feedback:**

- Weekly, during project supervision meetings
- Written feedback on Proposal (See Appendix A)
- Written feedback on Progress Report (See Appendix B)
- Feedback on presentation draft

### **SUMMATIVE FEEDBACK DELIVERABLES**

<b>Deliverable description and instructions</b>	<b>Weighting out of 100%</b>
Presentation (see Appendix C) comprising: a) presentation of software, with video URL b) project slides c) summary poster (i.e. the final project slide)	<b>10%</b>
Final Report (see Appendix D) comprising: a) written dissertation b) software artefact URL link to source code	<b>90%</b>

### **ASSIGNMENT IN DETAIL**

See Handbook Appendices A – D for assignment details and marking grid.

Real-time preictal detection through the  
application of machine learning to  
Electroencephalogram signals.

William Riddell

April 17, 2024

Word Count: 10,000

Supervised by Kashinath Basu

Faculty of Technology, Design and Environment  
School of Engineering, Computing and Mathematics

# Contents

<b>1</b>	<b>Abstract</b>	<b>12</b>
<b>2</b>	<b>Acknowledgements</b>	<b>12</b>
<b>3</b>	<b>Introduction</b>	<b>12</b>
3.1	Background . . . . .	13
3.2	Aim and Objectives . . . . .	13
3.3	Project Requirements . . . . .	14
<b>4</b>	<b>Background Review</b>	<b>14</b>
4.1	Datasets . . . . .	14
4.2	Machine Learning (ML) Models . . . . .	19
<b>5</b>	<b>Project Methodology and Development</b>	<b>24</b>
5.1	Methodology . . . . .	24
5.1.1	Approach . . . . .	24
5.1.2	Development System . . . . .	25
5.1.3	Version Control . . . . .	25
5.1.4	Software and Libraries Used . . . . .	26
5.1.5	Schedule . . . . .	26
5.2	Development . . . . .	27
5.2.1	Electroencephalogram (EEG) Data Extraction . . . . .	27
5.2.2	Short-Time Fourier Transform (STFT) . . . . .	29
5.2.3	Notch Filter . . . . .	30
5.2.4	Synthesize Data . . . . .	32
5.2.5	Convolutional Neural Network (CNN) and tuning methodology . . . . .	35
5.2.6	Real Time Simulation . . . . .	40
5.3	Problems Encountered . . . . .	41
<b>6</b>	<b>Review</b>	<b>43</b>
6.1	Objectives and Requirements . . . . .	43
6.2	Results . . . . .	44
6.3	Future Advancements . . . . .	44
6.3.1	Generic Model . . . . .	44
6.3.2	EEG node configuration . . . . .	44
6.3.3	STFT parameters . . . . .	44

6.3.4	Model Architecture . . . . .	45
6.3.5	Prediction System . . . . .	45
6.3.6	Online Learning . . . . .	45
<b>7</b>	<b>Professional Issues and Risks</b>	<b>45</b>
7.1	Professional Issues . . . . .	45
7.2	Risks . . . . .	47
7.2.1	Risk Matrix . . . . .	47
7.2.2	Risk Assessment . . . . .	48
<b>8</b>	<b>Conclusion</b>	<b>49</b>
8.1	Objective Review . . . . .	49
8.2	Final Thoughts . . . . .	51
<b>9</b>	<b>Bibliography</b>	<b>52</b>
<b>A</b>	<b>Links</b>	<b>55</b>
A.1	GitHub . . . . .	55
A.2	Video Demonstration . . . . .	55
<b>B</b>	<b>Results</b>	<b>55</b>
B.1	Tables . . . . .	56
B.2	Plots . . . . .	66
B.2.1	Timings . . . . .	66
B.2.2	Training Loss . . . . .	67
B.2.3	Training Accuracy . . . . .	69
B.2.4	Testing Accuracy . . . . .	71
B.2.5	Testing Recall . . . . .	73
B.2.6	Testing F1 . . . . .	83
B.2.7	Testing Precision . . . . .	93
<b>C</b>	<b>Misc.</b>	<b>102</b>
<b>D</b>	<b>Presentation, Poster, and Ethics Form</b>	<b>112</b>

## List of Figures

1	Third Order Butterworth bandpass filter . . . . .	20
---	---------------------------------------------------	----

2	Fourier transform equation . . . . .	20
3	Cross-Correlation . . . . .	21
4	Maximal Lyapunov ( $\lambda$ ) exponent . . . . .	22
5	Phase-locking Synchrony (SPLV) extraction from the wavelet transformation . . . . .	22
6	Gantt Chart . . . . .	27
7	Extracted raw EEG data. (Science 2024) . . . . .	27
8	Diagram of the 10-20 International System. ( <i>21 electrodes of International 10-20 system for EEG.svg</i> 2010) . . . . .	28
9	Head of extracted EEG data in CSV format. The shown data is approx 1/5 second and each column represents an single EEG node's data. . . . .	29
10	Short-Time Fourier Transform (STFT) Window Extraction . .	29
11	An STFT window with high frequency resolution. . . . .	31
12	Sliding Window technique for synthesizing data spectrogram images . . . . .	34
13	Calculate the desired window size. "desired_size" is the length of the windows in seconds. "sampling_rate" is the sampling rate of the recorded data. . . . .	34
14	Calculate the "step" or "speed" of the sliding window. Can also be described as the offset between window starting points. "target" is the desired number of windows. . . . .	34
15	Produces an set of pairs, where an pair describes the start and ending index of an single window. The set length is the same as "target" . . . . .	34
16	Categorical Cross-Entropy . . . . .	35
17	Optimization function performance comparison. . . . .	36
18	Screenshot of running real-time simulation. . . . .	41
19	Recorded time metrics. . . . .	66
20	Heatmap of loss during training for models with architecture 2.4.64. . . . .	67
21	Heatmap of loss during training for models with architecture 2.4.128 . . . . .	68
22	Heatmap of accuracy during training for models with architecture 2.4.64. . . . .	69
23	Heatmap of accuracy during training for models with architecture 2.4.128 . . . . .	70

24	Heatmap of accuracy during testing for models with architecture 2.4.64. . . . .	71
25	Heatmap of accuracy during testing for models with architecture 2.4.128 . . . . .	72
26	Heatmap of recall for ictal class during testing for models with architecture 2.4.64. . . . .	73
27	Heatmap of recall for preictal class during testing for models with architecture 2.4.64. . . . .	74
28	Heatmap of recall for interictal class during testing for models with architecture 2.4.64. . . . .	75
29	Heatmap of recall (weighted average) during testing for models with architecture 2.4.64. . . . .	76
30	Heatmap of recall (macro average) during testing for models with architecture 2.4.64. . . . .	77
31	Heatmap of recall for ictal class during testing for models with architecture 2.4.128. . . . .	78
32	Heatmap of recall for preictal class during testing for models with architecture 2.4.128. . . . .	79
33	Heatmap of recall for interictal class during testing for models with architecture 2.4.128. . . . .	80
34	Heatmap of recall (weighted average) during testing for models with architecture 2.4.128. . . . .	81
35	Heatmap of recall (macro average) during testing for models with architecture 2.4.128. . . . .	82
36	Heatmap of F1 for ictal class during testing for models with architecture 2.4.64. . . . .	83
37	Heatmap of F1 for preictal class during testing for models with architecture 2.4.64. . . . .	84
38	Heatmap of F1 for interictal class during testing for models with architecture 2.4.64. . . . .	85
39	Heatmap of F1 (weighted average) during testing for models with architecture 2.4.64. . . . .	86
40	Heatmap of F1 (macro average) during testing for models with architecture 2.4.64. . . . .	87
41	Heatmap of F1 for ictal class during testing for models with architecture 2.4.128. . . . .	88
42	Heatmap of F1 for preictal class during testing for models with architecture 2.4.128. . . . .	89



43	Heatmap of F1 for interictal class during testing for models with architecture 2.4.128. . . . .	90
44	Heatmap of F1 (weighted average) during testing for models with architecture 2.4.128. . . . .	91
45	Heatmap of F1 (macro average) during testing for models with architecture 2.4.128. . . . .	92
46	Heatmap of precision for ictal class during testing for models with architecture 2.4.64. . . . .	93
47	Heatmap of precision for preictal class during testing for models with architecture 2.4.64. . . . .	94
48	Heatmap of recall for interictal class during testing for models with architecture 2.4.64. . . . .	95
49	Heatmap of precision (weighted average) during testing for models with architecture 2.4.64. . . . .	96
50	Heatmap of precision (macro average) during testing for models with architecture 2.4.64. . . . .	97
51	Heatmap of precision for ictal class during testing for models with architecture 2.4.128. . . . .	98
52	Heatmap of precision for preictal class during testing for models with architecture 2.4.128. . . . .	99
53	Heatmap of precision for interictal class during testing for models with architecture 2.4.128. . . . .	100
54	Heatmap of precision (weighted average) during testing for models with architecture 2.4.128. . . . .	101
55	Heatmap of precision (macro average) during testing for models with architecture 2.4.128. . . . .	102
56	An STFT window with high time resolution. . . . .	105

## List of Tables

1	The Datasets analysed . . . . .	15
2	Channel Characteristics . . . . .	16
3	Temporal properties . . . . .	17
4	Subject properties . . . . .	18
5	Suggested applications . . . . .	19

6	Results from (Savadkoochi, Oladunni & Thompson 2020) showing results for the Time Domain, Frequency Domain, and Time-Frequency Domain . . . . .	21
7	Cross-Correlation results. . . . .	22
8	Non-Linear interdependence results. . . . .	23
9	Dynamical Entrainment; difference of short-term Lyapunov exponents (DSTL) results. . . . .	23
10	Phase-locking Synchrony (SPLV) results. . . . .	23
11	Entropy of Phase Difference results. . . . .	23
12	Distribution of Wavelet Coherence results. . . . .	23
13	Cross-Correlation results. . . . .	24
14	List of Libraries used . . . . .	26
15	List of Short-Time Fourier Transform (STFT) parameters . . . . .	30
16	List of Short-Time Fourier Transform (STFT) parameters choices . . . . .	30
17	Recording information for each subject. Valid represents if an subject has enough data, and is therefore suitable for training. . . . .	33
18	The possible values for each tuning variable. . . . .	38
19	Objectives taken from 3.2. . . . .	43
20	Requirements taken from 3.3. . . . .	43
21	Risk Matrix . . . . .	48
22	Risk Assessment . . . . .	49
23	Time duration metrics for each model in format %H:%M:%S.%f. . . . .	56
24	Accuracy and Loss metrics (average across epochs) during model training. . . . .	57
25	Accuracy metrics obtained during testing. . . . .	58
26	Generated Confusion Matrix during the model tuning process. . . . .	60
27	Generated Normalized Confusion Matrix during the model tuning process. . . . .	62
28	F1 metrics obtained during testing. . . . .	63
29	Precision metrics obtained during testing. . . . .	64
30	Recall metrics obtained during testing. . . . .	65

## Acronyms

**AI** Artificial Intelligence. 12

**ANN** Artificial Neural Network. 42

**CHB-MIT** Children’s Hospital Boston. 24, 26–28, 32, 41, 46, 49, 102

**CNN** Convolutional Neural Network. 5, 12, 19, 21–24, 29, 35–37, 44, 45

**CSV** Comma-separated values. 7, 28, 29

**DSTL** Dynamical Entrainment; difference of short-term Lyapunov exponents. 10, 20, 22, 23

**EDF** European Data Format. 26, 27

**EEG** Electroencephalogram. 5, 7, 12–14, 24, 26–30, 40, 41, 43–45, 49, 50, 102

**EEGs** Electroencephalograms. 13, 14, 29

**kNN** K Nearest Neighbor. 19–21

**LOOCV** Leaving One Out Cross-Validation. 36

**LR** Logistical Regression. 19, 21–24

**ML** Machine Learning. 5, 12–14, 17, 19, 21, 24–26, 30, 32, 43, 46, 49, 50

**OOM** Out of Memory. 38, 42

**OOP** Object-oriented Programming. 28

**SPLV** Phase-locking Synchrony. 7, 10, 20, 22, 23

**STFT** Short-Time Fourier Transform. 5, 7, 10, 12, 20, 24–26, 29–31, 36, 41, 44, 50

**SVM** Support Vector Machine. 19–23

# 1 Abstract

“Epilepsy is one of the most common serious brain conditions, affecting over 70 million people worldwide.” (Thijs, Surges, O’Brien & Sander 2019) Antiepileptic Drugs have been developed and are currently in widespread use, however, “one third of people with epilepsy will not have adequate seizure control with the current medications [available Antiepileptic Drugs (AEDs)]. For these patients, the situation has improved very little in the last few decades” (Galanopoulou, Buckmaster, Staley, Moshé, Perucca, Engel Jr, Löscher, Noebels, Pitkänen, Stables et al. 2012). Furthering this, 35% of epilepsy patients become resistant to AED medication (Moghim & Corne 2014). This paints a bleak situation for people suffering from epilepsy; “There is an urgent need for more effective and better-tolerated treatments” (Löscher & Schmidt 2011).

This project describes a non-intrusive solution that classifies the current state of a patient in real-time. A preprocessing and classification pipeline has been developed which classifies Electroencephalogram (EEG) signals into one of three categories (interictal, preictal, or ictal) through the use of a Convolutional Neural Network (CNN) and Short-Time Fourier Transform (STFT). This project has also tuned a selection of models, with 15 different models boasting 100% accuracy against a validation dataset.

# 2 Acknowledgements

I would like to thank my project supervisor, Kashinath Basu for his support throughout the project. Our weekly meetings gave invaluable advice and direction which helped me stay on track and encouraged me to develop a project which I am incredibly proud of.

# 3 Introduction

Over the last 20 years, Artificial Intelligence (AI) has seen a large evolution through the use of Machine Learning (ML); the statistical analysis of data which leads to the unveiling of characteristics and connections. (Awad & Khanna 2015). There has been a large uptake of applying ML techniques to biomedical data, increasing the speed and accuracy of prediction, detection, diagnosis, and prognosis.

Electroencephalograms (EEGs) measure the electrical signals in the brain. EEGs have a great use in giving an insight into the inner workings of the brain, for example allowing us to pick up abnormalities preceding and during their occurrence. “A seizure is a burst of uncontrolled electrical activity between brain cells (also called neurons or nerve cells) that causes temporary abnormalities in muscle tone or movements (stiffness, twitching or limpness), behaviors, sensations or states of awareness.” (Medicine n.d.) Due to this, monitoring the brain’s electrical activity through the use of an EEG, and applying analysis through an ML model may allow us to detect the preictal period. “An automated accurate prediction of seizures will significantly improve the quality of life of patients and reduce the burden on caregivers” (Acharya, Hagiwara & Adeli 2018)

### 3.1 Background

“Because of their unpredictable nature, uncontrolled seizures represent a major personal handicap and source of worry for patients. In addition, persistent seizures constitute a considerable burden on healthcare resources.” (Assi, Nguyen, Rihana & Sawan 2017) Due to this both medication and surgery are available to applicable patients, although with 30% patients being refractory to drug therapy, and an equally bleak surgery success rate; 75% in lesional cases, and 50% in nonlesional cases for temporal lobe cases along with 60% in lesional cases and merely 35% in nonlesional for frontal lobe cases (Assi et al. 2017), a large population of patients would therefore greatly benefit from a prediction system in their daily life.

### 3.2 Aim and Objectives

This project will aim to develop a consistent ML model trained to classify either preictal, interictal, and ictal periods. The model will have to achieve a high degree of accuracy ( $\geq 90\%$ ) when being applied to EEG data in real time. Furthermore, a real-time simulation will need to be developed along with an ML pipeline and model parameter tuning.

#### Objectives

1. Research and find a suitable dataset allowing for the classification of the interictal, preictal, and ictal periods.

2. Research and find an suitable ML model approach.
3. Create a data preparation and preprocessing pipeline that extracts the dataset into an ML format suitable for training and testing.
4. Undergo parameter tuning and model architecture tuning of the selected ML model approach.
5. Produce a simulation interface that streams EEG data in real-time through the preprocessing pipeline and into the optimal model. The simulation should show the current classification to the user every second.

### 3.3 Project Requirements

The formal requirements of this project are stated in

1. Develop and preprocessing pipeline which accepts raw EEG data.
2. Tune a number of models to find the optimal variable configuration.
3. Ensure that the optimal model has an accuracy of  $\geq 90\%$  against a validation dataset.
4. From EEG data extraction to prediction, the process must take at most 1 second.
5. Create a real-time simulation mimicking a real-life application.

## 4 Background Review

### 4.1 Datasets

(Wong, Simmons, Rivera-Villicana, Barnett, Sivathamboo, Perucca, Ge, Kwan, Kuhlmann, Vasa et al. 2023) reviews 10 datasets available to download. It evaluates the way the EEGs were physically setup on the subject, the subjects themselves, and the data's properties. Wong et al. also state their opinion on what tasks suit what dataset, with the main two tasks being either detection or prediction.

**Dataset**

University of Bonn

CHB-MIT Scalp EEG

Melbourne-NeuroVista seizure trial (Neurovista Ictal)

Kaggle UPenn and Mayo Clinic’s Seizure Detection Challenge

Neurology and Sleep Centre Hauz Khas

Kaggle American Epilepsy Society Seizure Prediction Challenge

Kaggle Melbourne-University AES-MathWorks-NIH Seizure Prediction Challenge

TUH EEG Seizure Corpus (TUSZ)

Siena Scalp EEG

Helsinki University Hospital EEG

Table 1: The Datasets analysed

Within these datasets, Wong et al. were also able to find the way the EEG nodes were positioned on the subject’s cranium, along with whether the EEG nodes were either placed intracranial or extracranial. Wong et al. also the number of channels that are contained in the raw EEG data for each dataset.

<b>Dataset</b>	<b>Number of channels</b>	<b>Placement method</b>	<b>Type of signal</b>
University of Bonn	1	International 10–20 system, Intracranial	Scalp/Intracranial EEG
CHB-MIT Scalp EEG	18	International 10–20 system/Nomenclature	Scalp EEG
Melbourne-NeuroVista seizure trial (NeuroVista Ictal)	16	Intracranial	Intracranial EEG
Kaggle UPenn and Mayo Clinic’s Seizure Detection Challenge	16–76	Intracranial	Intracranial EEG
Kaggle American Epilepsy Society Seizure Prediction Challenge	16	Intracranial	Intracranial EEG
Neurology and Sleep Centre Hauz Khas	1	International 10–20 System	Scalp EEG
Kaggle Melbourne-University AES-MathWorks-NIH Seizure Prediction Challenge Data	16	Intracranial	Intracranial EEG
TUH EEG Seizure Corpus (TUSZ)	23–31	International 10–20 system / Nomenclature	Scalp EEG
Helsinki University Hospital EEG	19	International 10–20 system	Scalp EEG
Siena Scalp EEG	20/29	International 10–20 system/Nomenclature	Scalp EEG

Table 2: Channel Characteristics

Wong et al. also noted along with this data that the “University of Bonn dataset contains a mixture of both scalp and intracranial EEG data where



scalp EEG from healthy subjects was taken, while intracranial EEG was taken from subjects with a history of seizures.” (Wong et al. 2023). This may present a skew on the ML model during training.

<b>Dataset</b>	<b>Noncontinuous data</b>	<b>Short-term continuous data</b>	<b>Continuous data</b>
University of Bonn	Yes	No	No
CHB-MIT Scalp EEG	No	Yes	Yes
Melbourne-NeuroVista seizure trial (Neurovista Ictal)	N/A	N/A	N/A
Kaggle UPenn and Mayo Clinic’s Seizure Detection Challenge	Yes	No	No
Kaggle American Epilepsy Society Seizure Prediction Challenge	Yes	No	No
Neurology and Sleep Centre Hauz Khas	Yes	No	No
Kaggle Melbourne-University AES-MathWorks-NIH Seizure Prediction Challenge Data	Yes	No	No
TUH EEG Seizure Corpus (TUSZ)	No	Yes	No
Helsinki University Hospital EEG	No	Yes	No
Siena Scalp EEG	No	Yes	No

Table 3: Temporal properties

Wong et al. ordered the datasets into groups, either continuous or non-continuous data. For the continuous data they separated out datasets that record for less than 24 hours in a single go, these were classified as “Short-term continuous” data.

<b>Dataset</b>	<b>Number of subjects</b>	<b>Subject type</b>
University of Bonn	10	Human
CHB-MIT Scalp EEG	23	Human
Melbourne-NeuroVista seizure trial (NeuroVista Ictal)	12	Human
Kaggle UPenn and Mayo Clinic’s Seizure Detection Challenge	12	Human & Canine
Kaggle American Epilepsy Society Seizure Prediction Challenge	7	Human & Canine
Neurology and Sleep Centre Hauz Khas	10	Human
Kaggle Melbourne-University AES-MathWorks-NIH Seizure Prediction Challenge Data	3	Human
TUH EEG Seizure Corpus (TUSZ)	642	Human
Helsinki University Hospital EEG	79	Human
Siena Scalp EEG	14	Human

Table 4: Subject properties

Wong et al. also were able to identify the number of subjects within each dataset. Within the two “Kaggle” datasets there are Canine subjects, making them unsuitable for this project.

Within the review, they also produced tables displaying the segment information for each dataset, breaking down the recording length and frequency, along with the number of events and segments. This information should not weigh into which dataset suits the idea of preictal prediction so shall be left out in this background review. Wong et al. also discussed the idea of the class imbalance problem, where the number and length of each ictal period are unbalanced. Two datasets, “University of Bonn” and the “Neurology and Sleep Centre Hauz Khas” have addressed this issue and have balanced their data between ictal, preictal, interictal, and neonatal periods.

Taking the research into account Wong et al. suggested which dataset

suits either prediction or detection. “Since the aim of seizure prediction is to forecast impending seizures, EEG recordings that include preictal and interictal data should be used for the study, while the aim of seizure detection is to detect ongoing seizure events, hence, EEG recordings that contain ictal and interictal data should be used.” (Wong et al. 2023).

<b>Dataset</b>	<b>Application</b>
University of Bonn	Seizure detection
CHB-MIT Scalp EEG	Seizure detection/Prediction
Melbourne-NeuroVista seizure trial (NeuroVista Ictal)	Seizure detection/Prediction
Kaggle UPenn and Mayo Clinic’s Seizure Detection Challenge	Seizure detection
Kaggle American Epilepsy Society Seizure Prediction Challenge	Seizure prediction
Neurology and Sleep Centre Hauz Khas	Seizure detection/Prediction
Kaggle Melbourne-University AES-MathWorks-NIH Seizure Prediction Challenge Data	Seizure prediction
TUH EEG Seizure Corpus (TUSZ)	Seizure detection/Prediction
Helsinki University Hospital EEG	Seizure detection/Prediction
Siena Scalp EEG	Seizure detection/Predictio

Table 5: Suggested applications

## 4.2 Machine Learning (ML) Models

A series of papers have been reviewed with the following ML model types and feature extraction processes being used:

- Machine Learning (ML) models
  - K Nearest Neighbor (kNN).
  - Support Vector Machine (SVM).
  - Logistical Regression (LR).
  - Convolutional Neural Network (CNN).

- Preprocessing Pipelines
  - Time domain, using the third order Butterworth bandpass filter.
  - Frequency domain using the Fourier transform.
  - Time-Frequency domain using Wavelet Decomposition.
  - Cross-correlation.
  - Non-linear Interdependence.
  - Dynamical Entrainment; difference of short-term Lyapunov exponents (DSTL).
  - Phase-locking Synchrony (SPLV).
  - Entropy of the phase difference.
  - Wavelet Coherence.
  - Short-Time Fourier Transform (STFT)

## K Nearest Neighbor (kNN) and Support Vector Machine (SVM)

(Savadkoobi et al. 2020) built a feature extraction process that extracted the “time domain” using the Butterworth filter (1-70 Hz) 1, the “frequency domain” using a Fourier transform 2, and the “time-frequency domain” using Wavelet decomposition for the entire dataset. The resulting data is then split into its 5 brain wave bands; Delta, Theta, Alpha, Beta, and Gamma. From these 4 variables were extracted, mean, variance, skewness, and kurtosis, leading to 60 total extracted features.

$$y(n) = \sum_{i=0}^N a_i \cdot x(n-i) + \sum_{j=1}^N b_j \cdot y(n-j)$$

Figure 1: Third Order Butterworth bandpass filter

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-i2\pi\xi x} dx.$$

Figure 2: Fourier transform equation

A prediction was calculated from an kNN and SVM model for each separate domain. The results are shown in 6 showing the potential of each model.

Model	Accuracy %		
	TD	FD	T-FD
Support Vector Machine (SVM)	99.5	100	100
K Nearest Neighbor (kNN)	99.5	99	99.5

Table 6: Results from (Savadkoobi et al. 2020) showing results for the Time Domain, Frequency Domain, and Time-Frequency Domain

The classification times however were not included in the report. Extraction in real-time may not be applicable for an approach so extensive in its feature extraction.

### Logistical Regression (LR), Support Vector Machine (SVM), and Convolutional Neural Network (CNN)

Mirowski et al. used the (Freiburg 2024) dataset containing intracranial recordings and measured the performance of an LR, SVM, and CNN ML models. They compiled 6 different preprocessing pipelines and tested the selected models against each. The preprocessing pipelines are as follows:

- The Cross-correlation between pairs of EEG channels was calculated with delays ranging from -0.5 to 0.5. The delays allowed for the propagation and processing time of brainwaves. Cross-correlation describes the amount  $f$  has to be shifted along the  $x$  axis to equal  $g$ . Only the maximum value of the cross-correlation values was retained for training.

$$(f \star g)(\tau) \triangleq \int_{-\infty}^{\infty} \overline{f(t)} g(t + \tau) dt$$

Figure 3: Cross-Correlation

- Non-linear interdependence “which measures the distance, in state-space, between time-delay embedded trajectories of two EEG channels” was also extracted. This is a bivariate feature that “measures the

Euclidian distance, in reconstructed state-space, between trajectories described by two EEG channels”.

- The third method was Dynamical Entrainment; difference of short-term Lyapunov exponents (DSTL), where a Lyapunov exponent 4 describes the rate of separation for two trajectories, which in this report was based off “a common measure of the chaotic nature of a signal” (Mirowski, Madhavan, LeCun & Kuzniecky 2009)

$$\lambda = \lim_{t \rightarrow \infty} \lim_{|\delta \mathbf{Z}_0| \rightarrow 0} \frac{1}{t} \ln \frac{|\delta \mathbf{Z}(t)|}{|\delta \mathbf{Z}_0|}$$

Figure 4: Maximal Lyapunov ( $\lambda$ ) exponent

- The last 3 features were Phase-locking Synchrony (SPLV), Entropy, and Coherence of the phase difference. These were extracted from the wavelet transformation The wavelet transformation extracts the frequency-specific and time-dependent phases.

$$SPLV_{a,b}(f) = \frac{1}{N} \sum_{l=1}^N e^{i(\phi_{a,f}(l) - \phi_{b,f}(l))}$$

Figure 5: Phase-locking Synchrony (SPLV) extraction from the wavelet transformation

The following results show the number of patients with perfect seizure prediction results “(no false positives, all seizures predicted)” from (Mirowski et al. 2009) for each model, for each preprocessing pipeline.

Cross-Correlation		
LR	CNN	SVM
4	9	4
19%	43%	19%

Table 7: Cross-Correlation results.

Non-Linear interdependence		
LR	CNN	SVM
3	10	5
19%	48%	24%

Table 8: Non-Linear interdependence results.

DSTL
SVM
1
5%

Table 9: Dynamical Entrainment; difference of short-term Lyapunov exponents (DSTL) results.

SPLV		
LR	CNN	SVM
10	13	7
48%	62%	33%

Table 10: Phase-locking Synchrony (SPLV) results.

Entropy of Phase Difference		
LR	CNN	SVM
9	11	7
43%	52%	33%

Table 11: Entropy of Phase Difference results.

Distribution of Wavelet Coherence		
LR	CNN	SVM
11	15	8
52%	71%	38%

Table 12: Distribution of Wavelet Coherence results.

These results show the strength of an CNN when compared to other ML models as it's obtained the highest results across all tables. LR however can also achieve strong results when paired with the correct preprocessing pipeline. Both methods would be applicable for this project assuming that the time from recording to final prediction is less than a second.

## Convolutional Neural Network (CNN)

Another CNN approach was implemented by Truong et al. (Truong, Nguyen, Kuhlmann, Bonyadi, Yang, Ippolito & Kavehei 2018) who applied an Short-Time Fourier Transform (STFT) transform 10 on 30 seconds of EEG data. Interference was then removed from the produced matrices by applying a notch filter. The final result was then fed into an CNN producing these results:

No. of seizures	Sensitivity (%)	False Positive Rate (/h)
59	$81.4 \pm 0.0$	$0.06 \pm 0.00$

Table 13: Cross-Correlation results.

# 5 Project Methodology and Development

## 5.1 Methodology

### 5.1.1 Approach

From the datasets discussed in 4.1, the Children's Hospital Boston (CHB-MIT) EEG dataset has been chosen due to the large amount of continuous data, suitable for extracting the preictal period 5. The dataset also has EEG nodes positioned on the subject's scalps, fitting with the use case of this project. Furthermore, the CHB-MIT EEG dataset has produced summary plain-text files for each of the subject's recordings, stating when every seizure began and ended allowing for a straightforward methodology when configuring the EEG tuning and training environment.

Through the analysis of various ML models an CNN will be used for this project. Both (Truong et al. 2018) and (Mirowski et al. 2009) show the potential CNN have when being used in preictal prediction. The approach described in (Truong et al. 2018) was able to achieve similar or better results in



comparison with other preprocessing pipelines and/or ML model configurations while using a less computationally intense preprocessing method. This method was based around an STFT extraction and is described in (Truong et al. 2018). This may allow for the classification to be done in real-time, meeting the project’s objectives 3.2. To ensure this requirement is met a real-time simulation will be developed, showing the raw EEG data and the final prediction every second.

### **5.1.2 Development System**

Due to the linear nature of this project, a Waterfall methodology will be used. Waterfall ensures that previous stages of the software are complete before moving on, this will keep the project on track and moving in a forward direction, crucial for the time-limited aspect of this project. Furthermore, due to the analysis undergone through literature reviews the direction and therefore the goals of the project are clearly defined; this further supports the choice of a Waterfall approach.

### **5.1.3 Version Control**

This project will be under version control through the use of Git, along with GitHub as a remote repository. Both the dataset and the trained models will not be under version control. The use of a version control system and remote repository ensures that if anything goes awry a backup of all commits is stored remotely, securing any progress made in the project. Git also gives the benefit of many features; notably branches and stashing, allowing for developing ideas, features, and bugs without affecting the “main” code, and merging allows for the combination of said branches.

#### 5.1.4 Software and Libraries Used

Name:	Use case:
Google Cloud	Used to download the CHB-MIT EEG dataset
Keras (TensorFlow)	Used to build, train and test the ML models.
Matplotlib	Used in the real-time simulation to plot the Spectrogram and CHB-MIT EEG data.
MNE	Provides classes for loading EEG data from both CSV and EDF files. Also used for running an Short-Time Fourier Transform (STFT) extraction.
NumPy	Utilized functions and classes to store and manipulate loaded data.
Pandas	Utilized functions and classes to store and manipulate loaded data.
scikit-learn	Provides functions used to produce performance metrics for trained ML models.
nicegui.py	An framework which allows for web-application development.

Table 14: List of Libraries used

#### 5.1.5 Schedule

The schedule describes the way the Waterfall methodology has affected the project; it shows how it's been broken down into smaller, distinct stages with progress being blocked until the previous stage has been completed. The overlap in March and April is when the models are being tuned, therefore the only active development is on the real-time simulation.

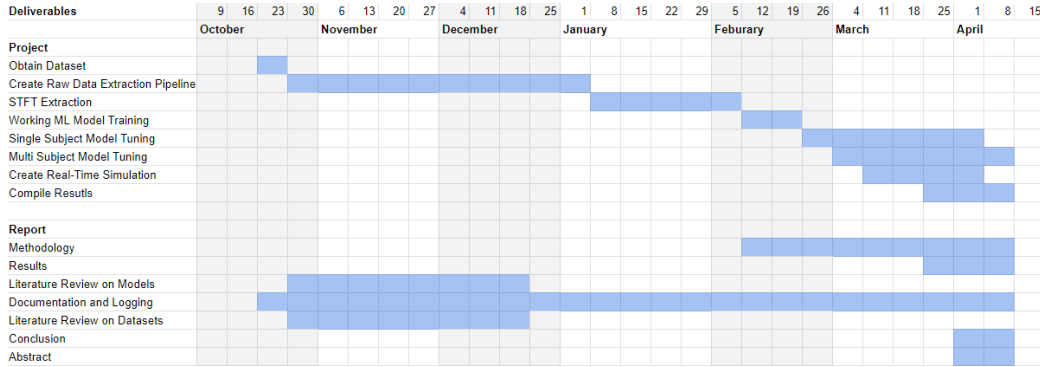


Figure 6: Gantt Chart

## 5.2 Development

### 5.2.1 Electroencephalogram (EEG) Data Extraction

The European Data Format (EDF) (Kemp, Värri, Rosa, Nielsen & Gade 1992) is a file format that was designed for the archival and exchange of EEG recordings (Kemp & Roessen 2013) and was the format of the CHB-MIT EEG dataset. EDF files contain raw EEG signals plotted against a time axis and contain metadata such as the name and recording frequency for each node. 7 shows an plot of raw EEG data.

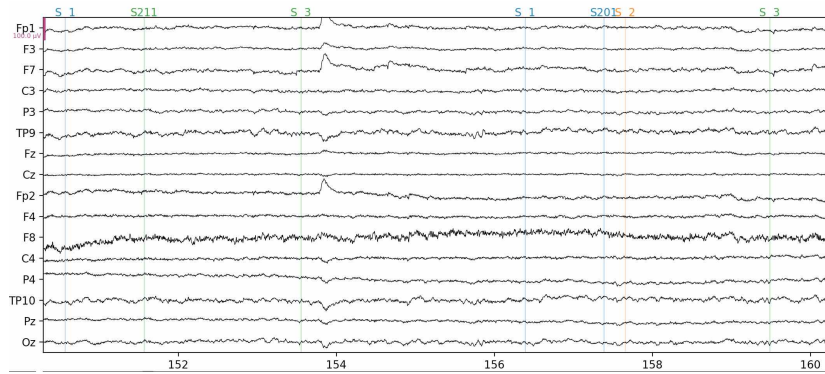


Figure 7: Extracted raw EEG data. (Science 2024)

Each node's name describes its position within the internationally accepted 10-20 system shown in 8. The 10-20 system was popularized due to

its ability to cover all brain regions in a method proportional to the skull size and shape. This ensures that the inter-electrode spacing is equal, allowing for consistent measurements to be taken regardless of subject. (Morley, Hill & Kaditis 2016)

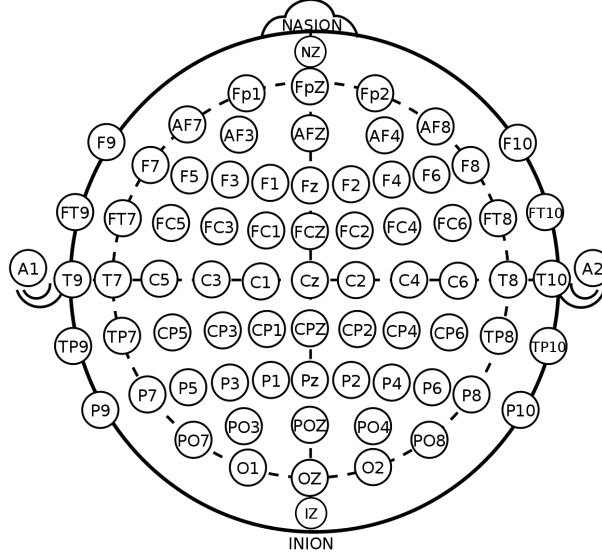


Figure 8: Diagram of the 10-20 International System. (*21 electrodes of International 10-20 system for EEG.svg* 2010)

The CHB-MIT dataset also contained summary files (see 3). From these summary files an Object-oriented Programming (OOP) representation of the dataset was created. This OOP representation was generated for every recording and shows which subject it belonged to, the name of each EEG node within the recording, also representing each node’s position in the 10-20 system, the recording frequency of the EEG, and any seizure’s start and end times contained within that recording. This OOP representation allowed for the labeling of the ictal, preictal, and interictal periods.

For this project, the EEG signals were extracted and stored in an CSV file. This allowed for a more flexible configuration of tuning scenarios such as EEG node combinations or subject-specific tuning, further discussed in ???. Within the CHB-MIT EEG dataset each EEG configuration was not equal as some recordings contained nodes not present in other recordings. Due to this, the common channels were extracted to allow for a fair comparison of

the results. There were 17 channels with the names: FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, P8-O2, FZ-CZ, CZ-PZ. An example of the extracted data can be seen at 9. For each recording 3 CSV files were created, “/ictal/master.csv”, “/preictal/master.csv” and “/interictal/master.csv”. If a period appeared more than once in a single recording it was concatenated to its corresponding “master.csv” file.

FP1	FP2	FP3	FP4	FP5	FP6	FP7	FP8	FP9	FP10	FP11	FP12	FP13	FP14	FP15	FP16	FP17	FP18	FP19	FP20	FP21	FP22	FP23	FP24	FP25	FP26	FP27	FP28	FP29	FP30	FP31	FP32	FP33	FP34	FP35	FP36	FP37	FP38	FP39	FP40	FP41	FP42	FP43	FP44	FP45	FP46	FP47	FP48	FP49	FP50	FP51	FP52	FP53	FP54	FP55	FP56	FP57	FP58	FP59	FP60	FP61	FP62	FP63	FP64	FP65	FP66	FP67	FP68	FP69	FP70	FP71	FP72	FP73	FP74	FP75	FP76	FP77	FP78	FP79	FP80	FP81	FP82	FP83	FP84	FP85	FP86	FP87	FP88	FP89	FP90	FP91	FP92	FP93	FP94	FP95	FP96	FP97	FP98	FP99	FP100	FP101	FP102	FP103	FP104	FP105	FP106	FP107	FP108	FP109	FP110	FP111	FP112	FP113	FP114	FP115	FP116	FP117	FP118	FP119	FP120	FP121	FP122	FP123	FP124	FP125	FP126	FP127	FP128	FP129	FP130	FP131	FP132	FP133	FP134	FP135	FP136	FP137	FP138	FP139	FP140	FP141	FP142	FP143	FP144	FP145	FP146	FP147	FP148	FP149	FP150	FP151	FP152	FP153	FP154	FP155	FP156	FP157	FP158	FP159	FP160	FP161	FP162	FP163	FP164	FP165	FP166	FP167	FP168	FP169	FP170	FP171	FP172	FP173	FP174	FP175	FP176	FP177	FP178	FP179	FP180	FP181	FP182	FP183	FP184	FP185	FP186	FP187	FP188	FP189	FP190	FP191	FP192	FP193	FP194	FP195	FP196	FP197	FP198	FP199	FP200	FP201	FP202	FP203	FP204	FP205	FP206	FP207	FP208	FP209	FP210	FP211	FP212	FP213	FP214	FP215	FP216	FP217	FP218	FP219	FP220	FP221	FP222	FP223	FP224	FP225	FP226	FP227	FP228	FP229	FP230	FP231	FP232	FP233	FP234	FP235	FP236	FP237	FP238	FP239	FP240	FP241	FP242	FP243	FP244	FP245	FP246	FP247	FP248	FP249	FP250	FP251	FP252	FP253	FP254	FP255	FP256	FP257	FP258	FP259	FP260	FP261	FP262	FP263	FP264	FP265	FP266	FP267	FP268	FP269	FP270	FP271	FP272	FP273	FP274	FP275	FP276	FP277	FP278	FP279	FP280	FP281	FP282	FP283	FP284	FP285	FP286	FP287	FP288	FP289	FP290	FP291	FP292	FP293	FP294	FP295	FP296	FP297	FP298	FP299	FP300	FP301	FP302	FP303	FP304	FP305	FP306	FP307	FP308	FP309	FP310	FP311	FP312	FP313	FP314	FP315	FP316	FP317	FP318	FP319	FP320	FP321	FP322	FP323	FP324	FP325	FP326	FP327	FP328	FP329	FP330	FP331	FP332	FP333	FP334	FP335	FP336	FP337	FP338	FP339	FP340	FP341	FP342	FP343	FP344	FP345	FP346	FP347	FP348	FP349	FP350	FP351	FP352	FP353	FP354	FP355	FP356	FP357	FP358	FP359	FP360	FP361	FP362	FP363	FP364	FP365	FP366	FP367	FP368	FP369	FP370	FP371	FP372	FP373	FP374	FP375	FP376	FP377	FP378	FP379	FP380	FP381	FP382	FP383	FP384	FP385	FP386	FP387	FP388	FP389	FP390	FP391	FP392	FP393	FP394	FP395	FP396	FP397	FP398	FP399	FP400	FP401	FP402	FP403	FP404	FP405	FP406	FP407	FP408	FP409	FP410	FP411	FP412	FP413	FP414	FP415	FP416	FP417	FP418	FP419	FP420	FP421	FP422	FP423	FP424	FP425	FP426	FP427	FP428	FP429	FP430	FP431	FP432	FP433	FP434	FP435	FP436	FP437	FP438	FP439	FP440	FP441	FP442	FP443	FP444	FP445	FP446	FP447	FP448	FP449	FP450	FP451	FP452	FP453	FP454	FP455	FP456	FP457	FP458	FP459	FP460	FP461	FP462	FP463	FP464	FP465	FP466	FP467	FP468	FP469	FP470	FP471	FP472	FP473	FP474	FP475	FP476	FP477	FP478	FP479	FP480	FP481	FP482	FP483	FP484	FP485	FP486	FP487	FP488	FP489	FP490	FP491	FP492	FP493	FP494	FP495	FP496	FP497	FP498	FP499	FP500	FP501	FP502	FP503	FP504	FP505	FP506	FP507	FP508	FP509	FP510	FP511	FP512	FP513	FP514	FP515	FP516	FP517	FP518	FP519	FP520	FP521	FP522	FP523	FP524	FP525	FP526	FP527	FP528	FP529	FP530	FP531	FP532	FP533	FP534	FP535	FP536	FP537	FP538	FP539	FP540	FP541	FP542	FP543	FP544	FP545	FP546	FP547	FP548	FP549	FP550	FP551	FP552	FP553	FP554	FP555	FP556	FP557	FP558	FP559	FP560	FP561	FP562	FP563	FP564	FP565	FP566	FP567	FP568	FP569	FP570	FP571	FP572	FP573	FP574	FP575	FP576	FP577	FP578	FP579	FP580	FP581	FP582	FP583	FP584	FP585	FP586	FP587	FP588	FP589	FP590	FP591	FP592	FP593	FP594	FP595	FP596	FP597	FP598	FP599	FP600	FP601	FP602	FP603	FP604	FP605	FP606	FP607	FP608	FP609	FP610	FP611	FP612	FP613	FP614	FP615	FP616	FP617	FP618	FP619	FP620	FP621	FP622	FP623	FP624	FP625	FP626	FP627	FP628	FP629	FP630	FP631	FP632	FP633	FP634	FP635	FP636	FP637	FP638	FP639	FP640	FP641	FP642	FP643	FP644	FP645	FP646	FP647	FP648	FP649	FP650	FP651	FP652	FP653	FP654	FP655	FP656	FP657	FP658	FP659	FP660	FP661	FP662	FP663	FP664	FP665	FP666	FP667	FP668	FP669	FP670	FP671	FP672	FP673	FP674	FP675	FP676	FP677	FP678	FP679	FP680	FP681	FP682	FP683	FP684	FP685	FP686	FP687	FP688	FP689	FP690	FP691	FP692	FP693	FP694	FP695	FP696	FP697	FP698	FP699	FP700	FP701	FP702	FP703	FP704	FP705	FP706	FP707	FP708	FP709	FP710	FP711	FP712	FP713	FP714	FP715	FP716	FP717	FP718	FP719	FP720	FP721	FP722	FP723	FP724	FP725	FP726	FP727	FP728	FP729	FP730	FP731	FP732	FP733	FP734	FP735	FP736	FP737	FP738	FP739	FP740	FP741	FP742	FP743	FP744	FP745	FP746	FP747	FP748	FP749	FP750	FP751	FP752	FP753	FP754	FP755	FP756	FP757	FP758	FP759	FP760	FP761	FP762	FP763	FP764	FP765	FP766	FP767	FP768	FP769	FP770	FP771	FP772	FP773	FP774	FP775	FP776	FP777	FP778	FP779	FP780	FP781	FP782	FP783	FP784	FP785	FP786	FP787	FP788	FP789	FP790	FP791	FP792	FP793	FP794	FP795	FP796	FP797	FP798	FP799	FP800	FP801	FP802	FP803	FP804	FP805	FP806	FP807	FP808	FP809	FP810	FP811	FP812	FP813	FP814	FP815	FP816	FP817	FP818	FP819	FP820	FP821	FP822	FP823	FP824	FP825	FP826	FP827	FP828	FP829	FP830	FP831	FP832	FP833	FP834	FP835	FP836	FP837	FP838	FP839	FP840	FP841	FP842	FP843	FP844	FP845	FP846	FP847	FP848	FP849	FP850	FP851	FP852	FP853	FP854	FP855	FP856	FP857	FP858	FP859	FP860	FP861	FP862	FP863	FP864	FP865	FP866	FP867	FP868	FP869	FP870	FP871	FP872	FP873	FP874	FP875	FP876	FP877	FP878	FP879	FP880	FP881	FP882	FP883	FP884	FP885	FP886	FP887	FP888	FP889	FP890	FP891	FP892	FP893	FP894	FP895	FP896	FP897	FP898	FP899	FP900	FP901	FP902	FP903	FP904	FP905	FP906	FP907	FP908	FP909	FP910	FP911	FP912	FP913	FP914	FP915	FP916	FP917	FP918	FP919	FP920	FP921	FP922	FP923	FP924	FP925	FP926	FP927	FP928	FP929	FP930	FP931	FP932	FP933	FP934	FP935	FP936	FP937	FP938	FP939	FP940	FP941	FP942	FP943	FP944	FP945	FP946	FP947	FP948	FP949	FP950	FP951	FP952	FP953	FP954	FP955	FP956	FP957	FP958	FP959	FP960	FP961	FP962	FP963	FP964	FP965	FP966	FP967	FP968	FP969	FP970	FP971	FP972	FP973	FP974	FP975	FP976	FP977	FP978	FP979	FP980	FP981	FP982	FP983	FP984	FP985	FP986	FP987	FP988	FP989	FP990	FP991	FP992	FP993	FP994	FP995	FP996	FP997	FP998	FP999	FP1000	FP1001	FP1002	FP1003	FP1004	FP1005	FP1006	FP1007	FP1008	FP1009	FP1010	FP1011	FP1012	FP1013	FP1014	FP1015	FP1016	FP1017	FP1018	FP1019	FP1020	FP1021	FP1022	FP1023	FP1024	FP1025	FP1026	FP1027	FP1028	FP1029	FP1030	FP1031	FP1032	FP1033	FP1034	FP1035	FP1036	FP1037	FP1038	FP1039	FP1040	FP1041	FP1042	FP1043	FP1044	FP1045	FP1046	FP1047	FP1048	FP1049	FP1050	FP1051	FP1052	FP1053	FP1054	FP1055	FP1056	FP1057	FP1058	FP1059	FP1060	FP1061	FP1062	FP1063	FP1064	FP1065	FP1066	FP1067	FP1068	FP1069	FP1070	FP1071	FP1072	FP1073	FP1074	FP1075	FP1076	FP1077	FP1078	FP1079	FP1080	FP1081	FP1082	FP1083	FP1084	FP1085	FP1086	FP1087	FP1088	FP1089	FP1090	FP1091	FP1092	FP1093	FP1094	FP1095	FP1096	FP1097	FP1098	FP1099	FP1100	FP1101	FP1102	FP1103	FP1104	FP1105	FP1106	FP1107	FP1108	FP1109	FP1110	FP1111	FP1112	FP1113	FP1114	FP1115	FP1116	FP1117	FP1118	FP1119	FP1120	FP1121	FP1122	FP1123	FP1124	FP1125	FP1126	FP1127	FP1128	FP1129	FP1130	FP1131	FP1132	FP1133	FP1134	FP1135	FP1136	FP1137	FP1138	FP1139	FP1140	FP1141	FP1142	FP1143	FP1144	FP1145	FP1146	FP1147	FP1148	FP1149	FP1150	FP1151	FP1152	FP1153	FP1154	FP1155	FP1156	FP1157	FP1158	FP1159	FP1160	FP1161	FP1162	FP1163	FP1164	FP1165	FP1166	FP1167	FP1168	FP1169	FP1170	FP1171	FP1172	FP1173	FP1174	FP1175	FP1176	FP1177	FP1178	FP1179	FP1180	FP1181	FP1182	FP1183	FP1184	FP1185	FP1186	FP1187	FP1188	FP1189	FP1190	FP1191	FP1192	FP1193	FP1194	FP1195	FP1196	FP1197	FP1198	FP1199	FP1200	FP1201	FP1202	FP1203	FP1204	FP1205	FP1206	FP1207	FP1208	FP1209	FP1210	FP1211	FP1212	FP1213	FP1214	FP1215	FP1216	FP1217	FP1218	FP1219	FP1220	FP1221	FP1222	FP1223	FP1224	FP1225	FP1226	FP1227	FP1228	FP1229	FP1230	FP1231	FP1232	FP1233	FP1234	FP1235	FP1236	FP1237	FP1238	FP1239	FP1240	FP1241	FP1242	FP1243	FP1244	FP1245	FP1246	FP1247	FP1248	FP1249	FP1250	FP1251	FP1252	FP1253	FP1254	FP1255	FP1256	FP1257	FP1258	FP1259	FP1260	FP1261	FP1262	FP1263	FP1264	FP1265	FP1266	FP1267	FP1268	FP1269	FP1270	FP1271	FP1272	FP1273	FP1274	FP1275	FP1276	FP1277	FP1278	FP1279	FP1280	FP1281	FP1282	FP1283	FP1284	FP1285	FP1286	FP1287	FP1288	FP1289	FP1290	FP1291	FP1292	FP1293	FP1294	FP1295	FP1296	FP1297	FP1298	FP1299	FP1300	FP1301	FP1302	FP1303	FP1304	FP1305	FP1306	FP1307	FP1308	FP1309	FP1310	FP1311	FP1312	FP1313	FP1314	FP1315	FP1316	FP1317	FP1318	FP1319	FP1320	FP1321	FP1322	FP1323	FP1324	FP1325	FP13
-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	------

This can also be explained as taking a Fourier transform of the EEG signals after a window function has been applied, and then sliding a window across the result. The sliding window transforms the one-dimensional output from the Fourier transform into two-dimensional data allowing for visual analysis.

There are various parameters for an STFT transformation:

Window Function	Used to isolate signal currently undergoing analysis. Optimal functions have low to no artefacts left in the signal and creates no discontinuities at section boundaries.
Window Size	Changes the size of the window function. This affects the resolution of both time and frequency, leading to the uncertainty principal; either variable will be in high resolution. See 11 and 56
Time Step (step size or hop size)	This is the distance between windows. Influences window over or underlap, as well as directly affecting computational load.

Table 15: List of Short-Time Fourier Transform (STFT) parameters

For this project these variables were selected:

Window Function	Sine Window: $w[n] = \sin\left(\frac{\pi n}{N}\right) = \cos\left(\frac{\pi n}{N} - \frac{\pi}{2}\right)$ , $0 \leq n \leq N$ .
Window Size	7680. The length of the input data.
Time Step (step size or hop size)	3840. This leads to an window overlap.

Table 16: List of Short-Time Fourier Transform (STFT) parameters choices

See 56 for an Spectrogram produced with these parameters.

### 5.2.3 Notch Filter

“Power line interference may severely corrupt neural recordings at 50/60 Hz and harmonic frequencies. The interference is usually non-stationary and can vary in frequency, amplitude, and phase.” (Keshtkaran & Yang 2014) This poses a large issue when training an ML model against EEG recordings as

the interference may affect the model's ability to pick up on characteristics, or may mislead the model.

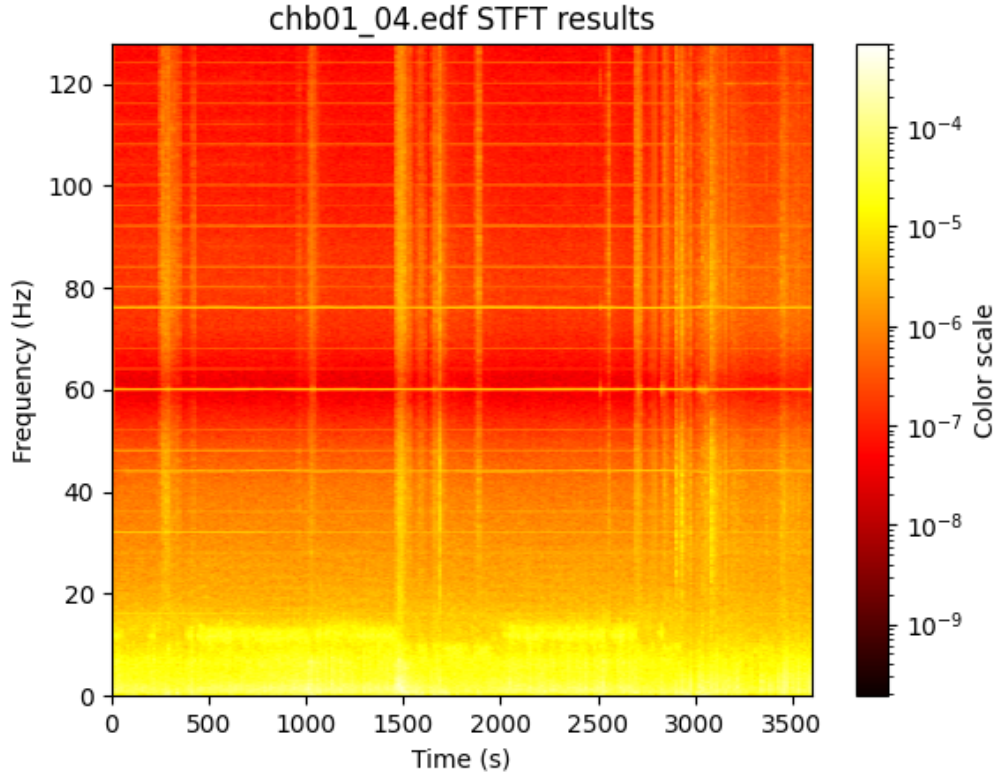


Figure 11: An STFT window with high frequency resolution.

11 clearly shows power line interference with a horizontal band around 60 Hz. (Keshtkaran & Yang 2014) discusses the different methods for removing power line interference, although due to the objectives stated in 3.2, a compute-intensive method needed to be avoided to allow the project to meet the real-time requirement. Through analysis done by Keshtkaran et al., a notch filter is fairly computationally light, although it does however have drawbacks such as not entirely removing the interference (Keshtkaran & Yang 2014). A notch filter was therefore applied to the raw data during the application of the STFT function.

#### 5.2.4 Synthesize Data

Within the CHB-MIT dataset there was a class imbalance with the interictal period dwarfing all other classes for every subject. A visualization of the data can be seen below in 17 which was generated from the extracted data. All subjects have very little data for the preictal and ictal classes, with most of them having under 30 seconds of recorded data. If a model was trained off synthesized data for these subjects, regardless of the synthesization method, the results would be greatly misleading and the models would not work in a real-life application. Therefore it was decided that only subjects “chb06” and “chb12” will be utilized for the training and tuning of ML models within this project.



	Recordings			Line Count			Length (Seconds)		
	Ictal	Pre	Inter	Ictal	Pre	Inter	Ictal	Pre	Inter
chb01	7	7	42	456	9334	37352484	1.78	36.46	145908.14
chb02	3	3	36	15	3118	32489237	0.059	12.18	126911.08
chb03	7	7	38	416	8201	35004081	1.625	32.04	136734.69
chb04	3	3	42	2385	15164	143801293	9.32	59.23	561723.8
chb05	5	5	39	568	7344	35927152	2.22	28.69	140340.44
chb06	7	7	18	17803	41291	61439324	69.54	161.3	239997.36
chb07	3	3	19	331	21128	61769049	1.29	82.53	241285.35
chb08	5	5	20	929	11739	18420150	3.63	45.86	71953.71
chb09	3	3	19	6648	19716	62519544	25.19	77.02	244216.97
chb10	7	7	25	461	26621	46068086	1.8	103.99	179953.46
chb11	3	3	35	812	3682	32052414	3.17	14.38	125204.74
chb12	10	10	21	8058	10643	18532198	31.48	41.57	72391.40
chb13	8	8	33	4435	8988	30391012	17.32	35.11	118714.90
chb14	7	7	26	1608	12433	23940969	6.28	48.57	93519.41
chb15	14	14	40	7160	12350	36843574	27.97	48.24	143920.21
chb16	6	6	19	4381	5832	17495378	17.11	22.78	68341.3
chb17	3	3	21	299	7678	1934639	1.17	29.99	75572.03
chb18	6	6	36	329	9142	32822357	1.29	35.71	128212.33
chb19	3	3	30	242	5657	27569463	0.95	22.1	107693.21
chb20	6	6	29	2569	6147	25421628	10.04	24.01	99303.23
chb21	4	4	33	207	7451	30240352	0.81	29.11	118126.34
chb22	3	3	31	210	7004	28557334	0.82	27.36	111552.09
chb23	3	3	9	12020	6111	24455749	46.95	23.87	95530.27

Table 17: Recording information for each subject. Valid represents if an subject has enough data, and is therefore suitable for training.

A sliding window method was used to synthesize the data for subjects “chb06” and “chb12”. The sliding window method moves a window of 30s across the data at speed  $x$ . Variable speed allows for the generation of a specific number of windows, see 12 for a diagram of this process. 13, 14, and 15 were taken from the project’s implementation and describe the algorithm used to synthesize the data.

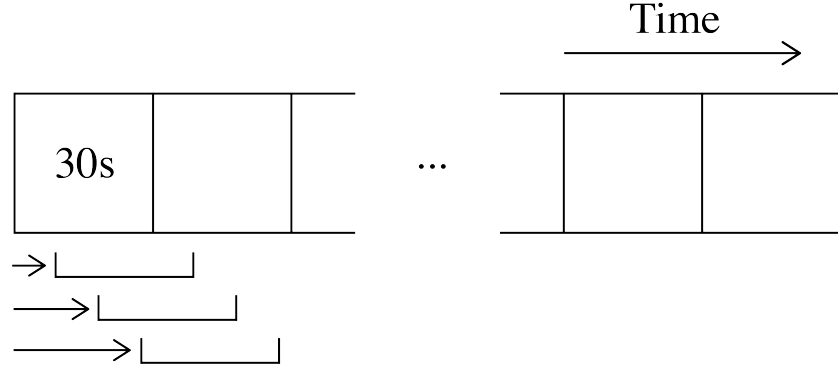


Figure 12: Sliding Window technique for synthesizing data spectrogram images

$$\text{window\_size} = \text{desired\_size} \times \text{sampling\_rate} = 30 \times 256$$

Figure 13: Calculate the desired window size. “desired\_size” is the length of the windows in seconds. “sampling\_rate” is the sampling rate of the recorded data.

$$\text{step} = \frac{\text{total\_rows} - \text{window\_size}}{\text{target} - 1}$$

Figure 14: Calculate the “step” or “speed” of the sliding window. Can also be described as the offset between window starting points. “target” is the desired number of windows.

$$\text{windows} = \{(\lfloor i \times \text{step} \rfloor, \lfloor i \times \text{step} + \text{window\_size} \rfloor) \mid i \in \{0, 1, 2, \dots, \text{target} - 1\}\}$$

Figure 15: Produces an set of pairs, where an pair describes the start and ending index of an single window. The set length is the same as “target”

### 5.2.5 Convolutional Neural Network (CNN) and tuning methodology

#### Overview

Due to the lack of suitable subjects discussed in 5.2.4 this project focused on tuning an CNN and its architecture. The number of convolutional blocks and dense layers, as well as the dense layer's complexity, were setup as the model's architecture tuning variables, this is further disused in 5.2.5 along with the other variables used for tuning. The loss function and optimizer however were not changed between model iterations and were set to Categorical Cross-Entropy and the Adam Optimizer respectively. Categorical Cross Entropy 16 is a common loss function for multi-class classification applications such as this project, therefore all models within this project utilized the same Categorical Cross Entropy function allowing for the comparison of the loss metrics.

$$H(y, \hat{y}) = - \sum_i y_i \log(\hat{y}_i)$$

Figure 16: Categorical Cross-Entropy

The Adam optimizer combines two different versions of stochastic gradient descent; Adaptive Gradient Algorithm and Root Mean Square Propagation. Both of these optimizer functions handle the learning rate for each node within a neural network allowing for more efficient training. The two functions manage the learning rates differently with the Adaptive Gradient Algorithm being more applicable for natural language and computer vision problems, Root Mean Square Propagation however suites online learning and non-stationary problems (Brownlee 2021). Adam combines these two approaches, realizing the benefit of each of them, and configures the learning rates accordingly. The performance of Adam when compared to other optimizer functions can be seen in 17 (Brownlee 2021).

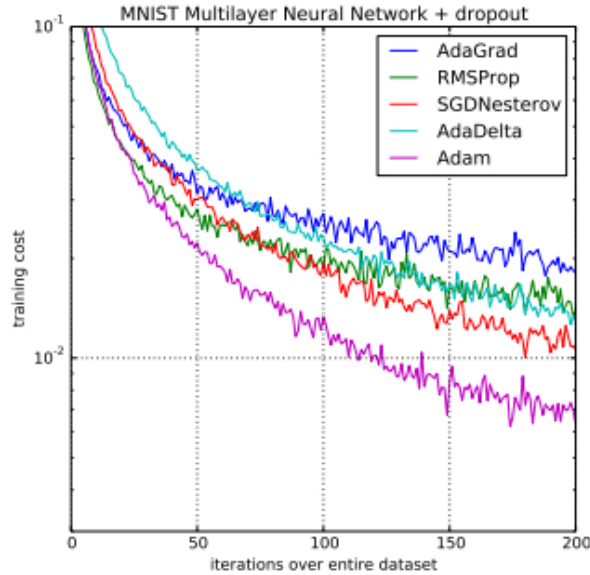


Figure 17: Optimization function performance comparison.

With only a single machine and the time available specified in ??, applying Leaving One Out Cross-Validation (LOOCV) for all architecture and variable configurations would have been impossible. Therefore a testing training split ratio needed to be decided on. The time costs of running testing, especially on complex CNN models were high, consequently, a large split for testing would be impractical as it would further drive up the training testing times. Due to the nature of the way the data was synthesized as well, a smaller testing split would be representative of all the available data for each class, this therefore led the project to use an 80% training and 20% testing split.

## Architecture

The CNN was dynamically built in accordance with the current variable configuration. By default, a single 2-dimensional convolutional block was included with an input shape of  $(17 \times 3841 \times 2)$  such that the model would accept a 30-second STFT window. This convolutional block had 32 filters with a kernel size of  $(3 \times 3 \times 3)$ . The dynamic convolutional blocks were then built, increasing the number of filters following the formula  $32 \times (2^i)$  where

$i$  is the convolutional block index. Each convolutional block is followed by a batch normalization, and after all convolutional blocks, a 2 dimensional Max Pooling layer with a pool size of  $(2 \times 2)$  is applied, along with a Flatten layer preparing the data to be passed to the dense layers.

After the convolutional blocks and related layers have been added the dense layers are built. Again, a single dense layer is added by default which has an output shape of  $(1, 3)$ , preceding this dense layer is the dynamically generated dense layers. These layers have a set output shape specified in the architecture tuning variables.

Listing 1: Function which dynamically builds an CNN model from the specified parameters. Logging has been removed.

```

1      def compile_model(self, num_conv_layers=4,
2                          num_dense_layers=4, dense_layer_size=64):
3          self.num_conv_layers = num_conv_layers
4          self.num_dense_layers = num_dense_layers
5          self.dense_layer_size = dense_layer_size
6
7          if num_dense_layers < 1 or num_conv_layers < 0:
8              raise ValueError
9
10         model = models.Sequential()
11
12         # add convolutional layers
13         model.add(layers.Conv2D(32, kernel_size=(3, 3),
14                                 activation='relu', input_shape=(17, 3841,
15                                                         2)))
16
17         for i in range(1, num_conv_layers):
18             model.add(layers.Conv2D(32 * (2**i),
19                                     kernel_size=(3, 3), activation='relu'))
20             model.add(layers.BatchNormalization())
21
22         model.add(layers.MaxPooling2D(pool_size=(2, 2)))
23         model.add(layers.Flatten())
24
25         # add dense layers
26         for i in range(num_dense_layers-1):
27             model.add(layers.Dense(dense_layer_size,
28                                     activation='relu'))
29

```

```

24         model.add(layers.Dense(3, activation='softmax'))
25         model.compile(loss='categorical_crossentropy',
                        optimizer='adam', metrics=['accuracy'])

```

**Tuning Variables** 18 shows the different variables and their possible values used for model tuning. A model was trained for possible configurations of these variables.

Model Parameters			
Num. Conv. Layers	2	3	
Num. Dense Layers	4	8	12
Dense Layer Size	64	128	256
Architecture Parameters			
Epochs	1	8	32
Batch Size	16	32	48

Table 18: The possible values for each tuning variable.

The number of convolutional layers had to be limited; as the filter size increased exponentially with each layer an Out of Memory (OOM) error would occur when the filter size was too large. Therefore only 2 and 3 convolutional layers were tested. The number of epochs was also limited due to training time. With this configuration of tuning variables, 162 models were required to be trained.

**Metrics** For each model, a plain text file was created during its tuning process. This plain text file recorded the start and end times of each process, including its model compiling, model training, and model testing. The file also contained metrics such as the training and testing accuracy and loss, the final confusion matrix, and the normalized confusion matrix. A classification report was also generated which calculated the precision, recall, F1 score, and support for each class. The classification report also calculated the accuracy of the model. An example of this file can be seen in 2. The results are available along with a discussion in 6.2.

Listing 2: Logging file programmatically generated during the tuning process. Training accuracy and loss lines have been trimmed for readability.

-----

```

current configuration

conv layers count      :    2
dense layer count      :    4
dense layer size       :   128

epochs      :   32
batch_size  :   32

dataset generation start : 2024-04-06 21:00:09.473945
one hot matrix:
['interictal', 'preictal', 'ictal']
{'interictal': 0, 'preictal': 1, 'ictal': 2}
[[1.  0.  0.]
 [0.  1.  0.]
 [0.  0.  1.]]
target subjects : ['chb06']
dataset generation end : 2024-04-06 21:00:09.666774
compiling model start : 2024-04-06 21:00:09.669026
compiling model end : 2024-04-06 21:00:09.847257
training start : 2024-04-06 21:00:09.848724
training end : 2024-04-07 01:59:22.883511
training accuracy: [0.974943995475769,
                    0.997187077999115, 0.9978642463684082,
                    0.9973433613777161, 1.0...]
training loss: [1.4391484260559082,
                0.01381457969546318, 0.019184233620762825,
                0.043409526348114014...]
saving model start : 2024-04-07 01:59:22.893082
saving model end : 2024-04-07 01:59:34.687428
cmatrix prediction start : 2024-04-07 01:59:34.691563
cmatrix prediction end : 2024-04-07 02:02:52.020467
Confusion Matrix:
[[1626    0    0]
 [    0 1543    0]
 [    0    0 1631]]
Normalized Confusion Matrix:
[[1.  0.  0.]

```

```

[0. 1. 0.]
[0. 0. 1.]]
Classification Report:

```

	precision	recall	f1-score	support
interictal	1.00	1.00	1.00	1626
preictal	1.00	1.00	1.00	1543
ictal	1.00	1.00	1.00	1631
accuracy			1.00	4800
macro avg	1.00	1.00	1.00	4800
weighted avg	1.00	1.00	1.00	4800

### 5.2.6 Real Time Simulation

A real-time simulation was built to show the potential of the preprocessing and prediction pipeline, proving its ability to output a prediction from raw EEG data each second. The real-time solution was built with “NiceGui.py” and the solution to setup model tuning was adapted to work in this new environment, implementing each step as defined above. The NiceGui solution has a hard-coded path to a saved Keras model. This model is loaded before the simulation begins and is used to run the predictions. 18 contains an image of the real-time simulation.



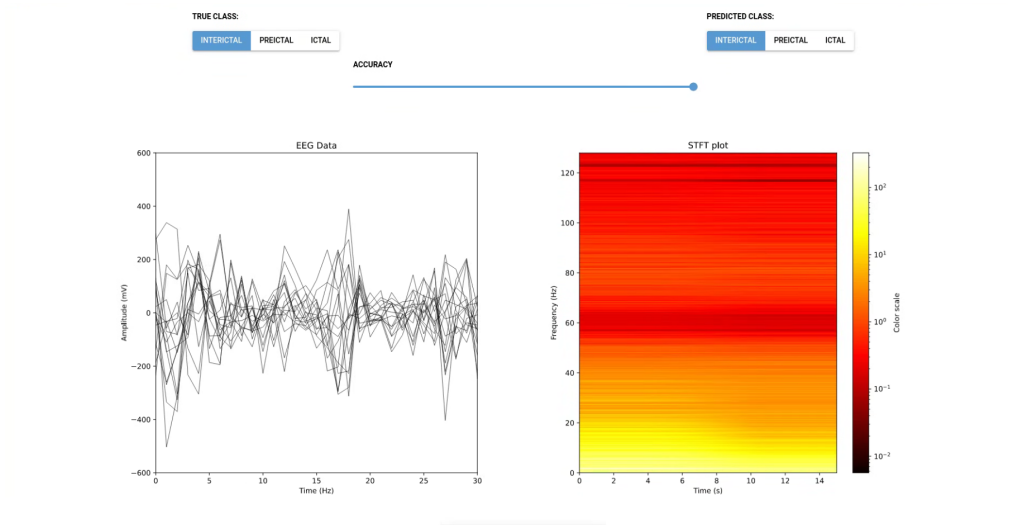


Figure 18: Screenshot of running real-time simulation.

Within the simulation you can toggle which class the EEG data is being pulled from, this then replaces the last 30 seconds of streamed data with data extracted from the newly selected class. This, in turn, refreshed the two plots on the webpage, with the left plot representing the signals from each EEG node, and the right plot being the produced STFT window used for prediction. The predicted class then appears in the top right, and the accuracy bar, representing the accuracy as a percentage is updated.

### 5.3 Problems Encountered

There have been many issues faced throughout this project. From the very start managing bodies of datasets would not respond, this impacted my decision-making when selecting one of the datasets discussed in 4.1. Furthermore, once the CHB-MIT dataset was selected there were discrepancies within the summary files, and the extracted data with channel names not matching. The CHB-MIT dataset also did not record the EEG values in the same manner for all subjects, this led to some data points from EEG nodes to be discarded as those specific EEG channels were not used for all subjects. On average this meant over 25% of recorded data had to be removed. Of what was recorded, some subjects had very little data for certain classes with an overabundance of recorded interictal data. Having such little data

for certain classes decreases the confidence of the synthesized data, as even though enough data points are available, the synthesized data may not be representative of a real-life situation.

This project’s approach changed substantially across its development. Originally development attempted to extract numeric and statistical features for each recorded point, more akin to traditional ANN methodologies, however, this approach was abandoned as it would not have been able to meet the real-time objective of this project.

Tuning the models raised more problems than anything else in this project, with OOM issues being a common problem occurring throughout the model tuning process and the development of the real-time simulation. All code affected with OOM problems had to be carefully re-written to remove any memory leaks as well as implementing generator functions which for example would open the dataset paths used for training only when required. Some OOM errors were not avoidable, however. Certain model architecture configurations would lead to memory errors with convolutional block complexity growing exponentially only a maximum of 3 convolutional blocks were able to be used. Dense layer count and complexity on top also added to the problem.

There were also many failed attempts at tuning the models. A re-occurring error was missing data points or even classes when tuning, this led to many trained models being invalid as they would only predict 2, or even 1 class due to its training data. Even when tuning was seemingly working this issue would appear with no exceptions being raised or problems being apparent. Over 250 trained models 4 have been discarded due to this problem alone, this has drastically reduced the results available for discussion.

## 6 Review

### 6.1 Objectives and Requirements

Index	Objective	Achieved
1	Research and find an suitable dataset allowing for the classification of the interictal, preictal, and ictal periods.	True
2	Research and find an suitable ML model approach.	True
3	Create a data preperation and preprocessing pipeline which extracts the dataset into an ML format suitable for training and testing.	True
4	Undergo parameter tuning and model architecture tuning of the selected ML model approach.	True
5	Produce a simulation interface which streams EEG data in real time through the preprocessing pipeline and into the optimal model. The simulation should show the current classification to the user every second.	True

Table 19: Objectives taken from 3.2.

Index	Requirements	Achieved
1	Develop and preprocessing pipeline which accepts raw EEG data.	True
2	Tune an number of models to find the optimal variable configuration.	True
3	Ensure that the optimal model has an accuracy of $\geq 90\%$ against an validation dataset.	True
4	From EEG data extraction to prediction, the process must take at most 1 second.	True
5	Create an real-time simulation mimicking an real life application.	True

Table 20: Requirements taken from 3.3.

The above tables correspond to the items in 3.3 and 3.2. Upon review, even with the tuning issues discussed in 7.1, this project was able to achieve all of

its objectives and requirements.

## **6.2 Results**

## **6.3 Future Advancements**

### **6.3.1 Generic Model**

All models have been trained against a single subject's data, this has allowed the CNN to pick up on their own characteristics for each class. Moving forward, a generic model for all subjects should be trained and evaluated against a single-subject model. Through this evaluation, a definitive answer can be found, pointing towards either a generic model having the same capabilities as a single-subject model or if a model for this application would have to be developed on a per-subject basis.

### **6.3.2 EEG node configuration**

The current model utilizes 17 EEG nodes which are all positioned on the subject's cranium. In a real-life application reducing the number of EEG nodes will drastically increase the convenience of having an EEG based prediction device. Therefore a tuning process, where models are developed with different configurations for all possible EEG node combinations will need to be developed. Reducing the number of EEG nodes will also decrease the price of a prediction system along with the chance of a node, and therefore the device failing.

### **6.3.3 STFT parameters**

STFT has 2 variables, being the window size and the time step. The window size affects the time and frequency resolution of the final image see 56. The time step defines the distance between two successive windows. Both of these variables should be tuned as a balance of the time and frequency resolution needs to be found to produce optimal results. This will be a lengthy process as different CNN convolutional block configurations will need to be tested for each STFT parameter combination.

### 6.3.4 Model Architecture

This project attempts to find the optimal CNN architecture, although due to issues discussed in 7.1 not all model architectures were tested. Therefore an approach that tunes the architecture would be beneficial as it may show that a less complex model will have the ability to produce just as successful metrics. Simplifying the model will allow it to run on a small or embedded device, perfect for a prediction system.

### 6.3.5 Prediction System

A prediction system prototype should be created to allow for real-life trials. OpenBCI is a company that produces EEG devices that can stream data in real-time to a machine. OpenBCI advertises EEG nodes, microcontrollers, and 3D printed models which hold the components on their website. OpenBCI has also made the STL files for their 3D models accessible through their GitHub [https://github.com/OpenBCI/Ultracortex/tree/master/Mark\\_IV](https://github.com/OpenBCI/Ultracortex/tree/master/Mark_IV). Using these materials along with either their microcontroller or a custom one running off open source software such as <https://github.com/Pi-EEG/EEGwithRaspberryPI> should allow for an easy platform to create such a prediction device.

### 6.3.6 Online Learning

Due to a lack of datasets for this application, converting the model's training strategy to online learning, along with collecting data in real-time through the prediction system discussed above, will allow for a model to continuously refine itself, picking up on characteristics for that specific subject. This approach, if successful should allow for the highest confidence, with minimal false positives alleviating unneeded stress for the user.

## 7 Professional Issues and Risks

### 7.1 Professional Issues

1. Legal Issues:

- If the project was to be implemented into a seizure prediction system then GDPR legislation would have to be upheld; if the model

is utilizing online learning then the data being stored will not be anonymous, meaning the data will have to be deleted when required and available to send to the user on their request. If the data however was truly anonymous, akin to the CHB-MIT dataset used in this project, then the data does not fall under GDPR. Another legal issue could arise if a seizure was not predicted correctly leading to a user suing the company producing the prediction system. This could be solved through an agreement that the end user has to accept, stating the producers of the product, including the developers are not liable.

## 2. Social Issues:

- People from all walks of life have epilepsy, and due to financial costs, some may not be able to afford the prediction system. This will be an issue that may not be solvable by the producers of the software and instead may need to rely on funding such as a governmental body, healthcare service, or charity. The producers can make the software open source allowing users to fit it onto their own hardware, although that will take time, money, and knowledge for the individual to set up an equivalent system.

## 3. Ethical Issues:

- The final product should attempt to achieve the same accuracy regardless of age or gender, although due to the differences in the way each person's brain functions and its characteristics, this may be a difficult task to achieve. In an ideal world either a generic model would exist, or models would be trained off datasets specific to age or gender allowing for common characteristics between individuals in those groups to be picked up on, however, there is a significant lack of data currently available making this unfeasible.

## 4. Environmental Issues:

- EEG nodes are comparatively inexpensive to their possible benefits, due to the low price of EEG nodes the creation of a prediction device won't have a large material cost. The ML model tuning process however will have a large energy cost. This will need to be taken into account when setting up a system that tunes models

at an industry scale. Steps will have to be taken to minimize the energy consumption of the system.

5. Intellectual Property Issues:

- There will be an issue when deciding who owns the recorded data. If the company that produces the final product owns it then they can utilize the large amount of data to further refine and train more advanced models, allowing them to tackle other issues such as the ones discussed in “Social Issues”, although this should be a choice for the end user. Another issue may be patenting the product and which components should or can be patented or copyrighted.

6. Accessibility Issues:

- As discussed epileptic patients come from all walks of life which means the final device needs to have an accessible interface, allowing everyone to have a clear indication of the state of the device, including the preictal period alert or even if the device is on. Different people will need different alert methods, and the final device should be extensible, allowing the final patient to fit it to their disability. Some variants of the device should have any combination of light, sound, and vibration alerts, as well as notifications to any of their other devices. This will allow anyone, regardless of disability to be alerted if a preictal period has been detected.

## **7.2 Risks**

### **7.2.1 Risk Matrix**

The numbers in each cell of the risk matrix corresponds to the items in 7.1.

Probability	Harm Severity			
	Minor	Marginal	Critical	Catastrophic
Certain				
Likely				
Possible				
Unlikely			4, 6	3, 5, 7
Rare				1, 2

Table 21: Risk Matrix

### 7.2.2 Risk Assessment

No.	Risk	Impact	Mitigation Strategy
1	Personal data could be leaked through an attack	Fail to adhere to GDPR. Criminal Offence	Implement encryption for data as well as increasing security measures
2	Personal data is leaked by unauthorized employee	Fail to adhere to GDPR. Criminal Offence	Tighten access controls. Educate employees about password and general security
3	Personal data is leaked by an authorized employee	Fail to adhere to GDPR. Criminal Offence	Employ education techniques stating the importance of adhering to GDPR as well as minimizing access to sensitive data.
4	False Negatives for the individual	End users will be unprepared for their seizures and may be caught in an unfavourable situation depending on reliance on the final product.	Implement cross validation techniques. Request all missed seizures to be logged or automatically detected for further training and inspection.



5	False Positives for the individual	A patient may experience un-needed stress or stop important activities due to false positives. Could have measurable knock on effects for an individual.	False positives should be recorded which will allow for further development of the model.
6	False Negatives in a healthcare environment	Staff may not be prepared for an seizure, increasing reaction times	See “False Negatives for the individual”
7	False Positives in a healthcare environment	Staff may waste time preparing for a seizure which never occurs, where their help may be needed elsewhere	See “False Positives for the individual”

Table 22: Risk Assessment

## 8 Conclusion

### 8.1 Objective Review

1. *Research and find a suitable dataset allowing for the classification of the interictal, preictal, and ictal periods.* - Two literature reviews were undergone, evaluating 10 possible datasets; outlining where the EEG nodes were positioned, their placement method, the recording style (continuous or non-continuous), their application, and even their subject type with some datasets evaluated containing canine subjects. This literature review allowed for the CHB-MIT dataset to be selected due to its appropriate node placement, recording style, and human-only subjects.
2. *Research and find an suitable ML model approach.* - The second literature review successfully researched different preprocessing pipelines along with ML models, this literature review was crucial in giving an insight into different approaches and which one would fit into a real-time

application. Without these two literature reviews the project would have met serious issues when attempting to meet the accuracy and real-time requirements of this project.

3. *Create a data preparation and preprocessing pipeline that extracts the dataset into an ML format suitable for training and testing.* - The pipeline was successfully implemented by using an STFT function along with a notch filter to remove interference. This approach was influenced by the ML model literature review. This pipeline was chosen as the number of extracted features, steps, and compute-heavy functions was low, giving it the best chance to meet the real-time requirement. The STFT window exceeded expectations and fully meets this requirement, allowing for effective, but fast preprocessing.
4. *Undergo parameter tuning and model architecture tuning of the selected ML model approach.* - For tuning ML models the classes had to be balanced. The chosen approach was to synthesize data by using a sliding window technique, this was used to generate windows until all 3 classes had an equal number of data points. Once the data has been synthesized the pre-processing pipeline was then applied, generating thousands of spectrogram images. Through the use of “sklearn”, “Tensorflow”, and “Keras” these spectrograms could then be loaded and used for tuning, testing or running a real-time simulation. The tuning process however was slightly buggy, and with more time available this hopefully would of been resolved allowing for a greater number of models to be trained and evaluated. Nevertheless, the results obtained through the tuning process also exceeded my expectations with testing results being near perfect for all tuned models.
5. *Produce a simulation interface which streams EEG data in real-time through the preprocessing pipeline and into the optimal model. The simulation should show the current classification to the user every second.* - The developed simulation proves the pipeline’s ability to run in real-time which confirms that the project has met the real-time requirement. The simulation streams 17x256 data points each second for the currently selected class and produces a prediction using a trained model. In the future, this simulation can be easily modified to accept real recorded data from a physical EEG setup. This is discussed in ???. At the beginning of this project, this integration was going to be

incorporated into the requirements, however after evaluating the time given it was dropped.

## **8.2 Final Thoughts**

This project has successfully achieved all requirements and objectives, producing the groundwork for a successful seizure prediction system. It's proved that the technology has the ability to accurately classify seizure periods, and predict seizures up to 20 minutes before they begin. If a detection system was to be developed it would alleviate large amounts of stress in the day-to-day lives of epileptic patients, something which would be invaluable for them. Even with this groundwork, there are still a lot of developments that can be done, however even with the state of the current project, if a prediction system was to be based on this paper it would be able to exceed expectations and even change the way people go about their day to day lives.

## 9 Bibliography

### References

- 21 electrodes of International 10-20 system for EEG.svg* (2010). Accessed: 2024-04-10.  
**URL:** [https://commons.wikimedia.org/wiki/File:21\\_electrodes\\_of\\_International\\_10-20\\_system\\_for\\_EEG.svg](https://commons.wikimedia.org/wiki/File:21_electrodes_of_International_10-20_system_for_EEG.svg)
- Acharya, U. R., Hagiwara, Y. & Adeli, H. (2018), ‘Automated seizure prediction’, *Epilepsy & Behavior* **88**, 251–261.
- Assi, E. B., Nguyen, D. K., Rihana, S. & Sawan, M. (2017), ‘Towards accurate prediction of epileptic seizures: A review’, *Biomedical Signal Processing and Control* **34**, 144–157.
- Awad, M. & Khanna, R. (2015), *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, Apress.  
**URL:** <https://books.google.co.uk/books?id=qPGQnAEACAAJ>
- Brownlee, J. (2021), ‘Gentle introduction to the adam optimization algorithm for deep learning’, ^1^. Accessed: 2024-04-14.
- Freiburg, S. P. P. (2024), ‘Eeg database’, <https://epilepsy.uni-freiburg.de/freiburg-seizure-prediction-project/eeg-database>. Accessed: 2024-04-07.
- Galanopoulou, A. S., Buckmaster, P. S., Staley, K. J., Moshé, S. L., Perucca, E., Engel Jr, J., Löscher, W., Noebels, J. L., Pitkänen, A., Stables, J. et al. (2012), ‘Identification of new epilepsy treatments: issues in preclinical methodology’, *Epilepsia* **53**(3), 571–582.
- Kemp, B. & Roessen, M. (2013), ‘European data format now supports video’, *Sleep* **36**(7), 1111–1111.
- Kemp, B., Värri, A., Rosa, A. C., Nielsen, K. D. & Gade, J. (1992), ‘A simple format for exchange of digitized polygraphic recordings’, *Electroencephalography and clinical neurophysiology* **82**(5), 391–393.

- Keshtkaran, M. R. & Yang, Z. (2014), ‘A fast, robust algorithm for power line interference cancellation in neural recording’, *Journal of neural engineering* **11**(2), 026017.
- Löscher, W. & Schmidt, D. (2011), ‘Modern antiepileptic drug development has failed to deliver: ways out of the current dilemma’, *Epilepsia* **52**(4), 657–678.
- Medicine, J. H. (n.d.), ‘Types of seizures, what is a seizure?’.  
**URL:** <https://www.hopkinsmedicine.org/health/conditions-and-diseases/epilepsy/types-of-seizures>
- Mirowski, P., Madhavan, D., LeCun, Y. & Kuzniecky, R. (2009), ‘Classification of patterns of eeg synchronization for seizure prediction’, *Clinical neurophysiology* **120**(11), 1927–1940.
- Moghimi, N. & Corne, D. W. (2014), ‘Predicting epileptic seizures in advance’, *PloS one* **9**(6), e99334.
- Morley, A., Hill, L. & Kaditis, A. (2016), ‘10-20 system eeg placement’, *European Respiratory Society, European Respiratory Society*.
- Savadkoobi, M., Oladunni, T. & Thompson, L. (2020), ‘A machine learning approach to epileptic seizure prediction using electroencephalogram (eeg) signal’, *Biocybernetics and Biomedical Engineering* **40**(3), 1328–1341.
- Science, N. D. (2024), ‘What is eeg? - neural data science in python’. Accessed: 2024-04-10.  
**URL:** <https://neuraldatascience.io/7-eeg/about-eeg.html>
- Thijs, R. D., Surges, R., O’Brien, T. J. & Sander, J. W. (2019), ‘Epilepsy in adults’, *The lancet* **393**(10172), 689–701.
- Truong, N. D., Nguyen, A. D., Kuhlmann, L., Bonyadi, M. R., Yang, J., Ippolito, S. & Kavehei, O. (2018), ‘Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram’, *Neural Networks* **105**, 104–111.
- Wong, S., Simmons, A., Rivera-Villicana, J., Barnett, S., Sivathamboo, S., Perucca, P., Ge, Z., Kwan, P., Kuhlmann, L., Vasa, R. et al. (2023),

‘Eeg datasets for seizure detection and prediction—a review’, *Epilepsia*  
*Open* .

## **A Links**

### **A.1 GitHub**

[https://github.com/N3utra1/COMP6013\\_Dissertation](https://github.com/N3utra1/COMP6013_Dissertation)

### **A.2 Video Demonstration**

<https://drive.google.com/file/d/1VGSPdRV1AAZp02pIs3qw4FANiYk2peBG/view?usp=sharing>

## **B Results**

Model key: [convolution block count].[dense layer count].[dense layer size]/[batch size].[epochs]

## B.1 Tables

Model	Training	Testing	Total
2.4.128/16.1	0:09:34.252395	0:02:29.318160	0:12:16.936242
2.4.128/16.32	4:55:14.648736	0:02:36.181157	4:58:03.207583
2.4.128/16.8	1:13:53.250455	0:02:28.278624	1:16:34.098996
2.4.128/32.1	0:10:16.908274	0:02:58.134548	0:13:27.358200
2.4.128/32.32	4:59:13.034787	0:03:17.328904	5:02:42.546522
2.4.128/32.8	1:21:12.806743	0:02:50.145996	1:24:15.803703
2.4.128/48.1	0:10:53.791005	0:02:57.700320	0:14:06.449949
2.4.128/48.32	5:04:12.389344	0:02:40.036783	5:06:58.710978
2.4.128/48.8	1:21:59.435414	0:02:31.724462	1:24:43.405594
2.4.64/16.1	0:09:16.671517	0:02:20.833553	0:11:43.794613
2.4.64/16.32	4:49:23.103991	0:02:40.019918	4:52:09.480089
2.4.64/16.8	1:18:51.356595	0:02:55.648672	1:21:53.885149
2.4.64/32.1	0:09:53.286427	0:02:30.918573	0:12:30.448678
2.4.64/32.32	5:01:06.399645	0:03:07.540450	5:04:20.224694
2.4.64/32.8	1:21:48.011924	0:02:32.670039	1:24:27.044614
2.4.64/48.1	0:09:58.299346	0:02:59.123882	0:13:03.698828
2.4.64/48.32	5:04:12.389344	0:02:40.036783	5:06:58.710978
2.4.64/48.8	1:20:23.212631	0:02:46.982105	1:23:16.378890

Table 23: Time duration metrics for each model in format %H:%M:%S.%f.



Model	Training Accuracy	Training Loss
2.4.128/16.1	0.9832265377044678	0.8224268555641174
2.4.128/16.32	0.9993602503091097	0.02814496521023102
2.4.128/16.8	0.9976428672671318	0.11281578045975693
2.4.128/32.1	0.975464940071106	1.4378010034561157
2.4.128/32.32	0.9989679381251335	0.04745509265126133
2.4.128/32.8	0.9962754622101784	0.18717686411091886
2.4.128/48.1	0.9669219255447388	2.318654775619507
2.4.128/48.8	0.9958652406930923	0.2898520281258925
2.4.64/16.1	0.9799447655677795	0.6853911876678467
2.4.64/16.32	0.9992593247443438	0.024666143968849415
2.4.64/16.8	0.9969331100583076	0.08956567873635081
2.4.64/32.1	0.9619211554527283	1.725386619567871
2.4.64/32.32	0.9987660832703114	0.05414362490514908
2.4.64/32.8	0.9953247904777527	0.21542102738881397
2.4.64/48.1	0.9655154347419739	2.379462718963623
2.4.64/48.32	0.9989125896245241	0.07438817310655671
2.4.64/48.8	0.994374118745327	0.31704871635884047

Table 24: Accuracy and Loss metrics (average across epochs) during model training.

Model	Testing Accuracy
2.4.128/16.1	0.99
2.4.128/16.32	1.00
2.4.128/16.8	1.00
2.4.128/32.1	1.00
2.4.128/32.32	1.00
2.4.128/32.8	1.00
2.4.128/48.1	1.00
2.4.128/48.8	1.00
2.4.64/16.1	1.00
2.4.64/16.32	1.00
2.4.64/16.8	1.00
2.4.64/32.1	1.00
2.4.64/32.32	1.00
2.4.64/32.8	1.00
2.4.64/48.1	1.00
2.4.64/48.32	1.00
2.4.64/48.8	1.00

Table 25: Accuracy metrics obtained during testing.

Model	Confusion Matrix		
	1593	33	0
2.4.128/16.1	0	1543	0
	0	0	1631
	1626	0	0
2.4.128/16.32	0	1543	0
	0	0	1631
	1625	1	0
2.4.128/16.8	0	1543	0
	0	0	1631
	1626	0	0
2.4.128/32.1	0	1543	0
	0	0	1631
	1626	0	0
2.4.128/32.32	0	1543	0
	0	0	1631
	1626	0	0
2.4.128/32.8	0	1543	0
	0	0	1631
	1626	0	0
2.4.128/48.1	0	1543	0
	0	0	1631
	1626	0	0
2.4.128/48.8	0	1543	0
	0	0	1631
	1626	0	0
2.4.64/16.1	0	1543	0
	0	0	1631
	1626	0	0
2.4.64/16.32	0	1543	0
	0	0	1631
	1626	0	0
2.4.64/16.8	0	1543	0
	0	0	1631
	1626	0	0
2.4.64/32.1	5	1538	0
	0	0	1631
	1626	0	0

	1626	0	0
2.4.64/32.32	0	1543	0
	0	0	1631
	1626	0	0
2.4.64/32.8	0	1543	0
	0	0	1631
	1626	0	0
2.4.64/48.1	0	1543	0
	0	0	1631
	1626	0	0
2.4.64/48.32	0	1543	0
	0	0	1631
	1626	0	0
2.4.64/48.8	0	1543	0
	0	0	1631

Table 26: Generated Confusion Matrix during the model tuning process.

Model	Normalized Confusion Matrix		
2.4.128/16.1	1.0	0.02093909	0.0
	0.0	0.97906091	0.0
	0.0	0.0	1.0
2.4.128/16.32	1.0	0.0	0.0
	0.0	1.0	0.0
	0.0	0.0	1.0
2.4.128/16.8	1.0	0.000647668394	0.0
	0.0	0.999352332	0.0
	0.0	0.0	1.0
2.4.128/32.1	1.0	0.0	0.0
	0.0	1.0	0.0
	0.0	0.0	1.0
2.4.128/32.32	1.0	0.0	0.0
	0.0	1.0	0.0
	0.0	0.0	1.0
2.4.128/32.8	1.0	0.0	0.0
	0.0	1.0	0.0
	0.0	0.0	1.0
2.4.128/48.1	1.0	0.0	0.0
	0.0	1.0	0.0
	0.0	0.0	1.0
2.4.128/48.8	1.0	0.0	0.0
	0.0	1.0	0.0
	0.0	0.0	1.0
2.4.64/16.1	1.0	0.0	0.0
	0.0	1.0	0.0
	0.0	0.0	1.0
2.4.64/16.32	1.0	0.0	0.0
	0.0	1.0	0.0
	0.0	0.0	1.0
2.4.64/16.8	1.0	0.0	0.0
	0.0	1.0	0.0
	0.0	0.0	1.0
2.4.64/32.1	0.9969344	0.0	0.0
	0.0030656	1.0	0.0
	0.0	0.0	1.0

	1.0	0.0	0.0
2.4.64/32.32	0.0	1.0	0.0
	0.0	0.0	1.0
	1.0	0.0	0.0
2.4.64/32.8	0.0	1.0	0.0
	0.0	0.0	1.0
	1.0	0.0	0.0
2.4.64/48.1	0.0	1.0	0.0
	0.0	0.0	1.0
	1.0	0.0	0.0
2.4.64/48.32	0.0	1.0	0.0
	0.0	0.0	1.0
	1.0	0.0	0.0
2.4.64/48.8	0.0	1.0	0.0
	0.0	0.0	1.0

Table 27: Generated Normalized Confusion Matrix during the model tuning process.

Model	Interictal	Preictal	Ictal	Macro Average	Weighted Average
2.4.128/16.1	0.99	0.99	1.00	0.99	0.99
2.4.128/16.32	1.00	1.00	1.00	1.00	1.00
2.4.128/16.8	1.00	1.00	1.00	1.00	1.00
2.4.128/32.1	1.00	1.00	1.00	1.00	1.00
2.4.128/32.32	1.00	1.00	1.00	1.00	1.00
2.4.128/32.8	1.00	1.00	1.00	1.00	1.00
2.4.128/48.1	1.00	1.00	1.00	1.00	1.00
2.4.128/48.8	1.00	1.00	1.00	1.00	1.00
2.4.64/16.1	1.00	1.00	1.00	1.00	1.00
2.4.64/16.32	1.00	1.00	1.00	1.00	1.00
2.4.64/16.8	1.00	1.00	1.00	1.00	1.00
2.4.64/32.1	1.00	1.00	1.00	1.00	1.00
2.4.64/32.32	1.00	1.00	1.00	1.00	1.00
2.4.64/32.8	1.00	1.00	1.00	1.00	1.00
2.4.64/48.1	1.00	1.00	1.00	1.00	1.00
2.4.64/48.32	1.00	1.00	1.00	1.00	1.00
2.4.64/48.8	1.00	1.00	1.00	1.00	1.00

Table 28: F1 metrics obtained during testing.

Model	Interictal	Preictal	Ictal	Macro Average	Weighted Average
2.4.128/16.1	1.00	0.98	1.00	0.99	0.99
2.4.128/16.32	1.00	1.00	1.00	1.00	1.00
2.4.128/16.8	1.00	1.00	1.00	1.00	1.00
2.4.128/32.1	1.00	1.00	1.00	1.00	1.00
2.4.128/32.32	1.00	1.00	1.00	1.00	1.00
2.4.128/32.8	1.00	1.00	1.00	1.00	1.00
2.4.128/48.1	1.00	1.00	1.00	1.00	1.00
2.4.128/48.8	1.00	1.00	1.00	1.00	1.00
2.4.64/16.1	1.00	1.00	1.00	1.00	1.00
2.4.64/16.32	1.00	1.00	1.00	1.00	1.00
2.4.64/16.8	1.00	1.00	1.00	1.00	1.00
2.4.64/32.1	1.00	1.00	1.00	1.00	1.00
2.4.64/32.32	1.00	1.00	1.00	1.00	1.00
2.4.64/32.8	1.00	1.00	1.00	1.00	1.00
2.4.64/48.1	1.00	1.00	1.00	1.00	1.00
2.4.64/48.32	1.00	1.00	1.00	1.00	1.00
2.4.64/48.8	1.00	1.00	1.00	1.00	1.00

Table 29: Precision metrics obtained during testing.



Model	Interictal	Preictal	Ictal	Macro Average	Weighted Average
2.4.128/16.1	0.98	1.00	1.00	0.99	0.99
2.4.128/16.32	1.00	1.00	1.00	1.00	1.00
2.4.128/16.8	1.00	1.00	1.00	1.00	1.00
2.4.128/32.1	1.00	1.00	1.00	1.00	1.00
2.4.128/32.32	1.00	1.00	1.00	1.00	1.00
2.4.128/32.8	1.00	1.00	1.00	1.00	1.00
2.4.128/48.1	1.00	1.00	1.00	1.00	1.00
2.4.128/48.8	1.00	1.00	1.00	1.00	1.00
2.4.64/16.1	1.00	1.00	1.00	1.00	1.00
2.4.64/16.32	1.00	1.00	1.00	1.00	1.00
2.4.64/16.8	1.00	1.00	1.00	1.00	1.00
2.4.64/32.1	1.00	1.00	1.00	1.00	1.00
2.4.64/32.32	1.00	1.00	1.00	1.00	1.00
2.4.64/32.8	1.00	1.00	1.00	1.00	1.00
2.4.64/48.1	1.00	1.00	1.00	1.00	1.00
2.4.64/48.32	1.00	1.00	1.00	1.00	1.00
2.4.64/48.8	1.00	1.00	1.00	1.00	1.00

Table 30: Recall metrics obtained during testing.

## B.2 Plots

### B.2.1 Timings

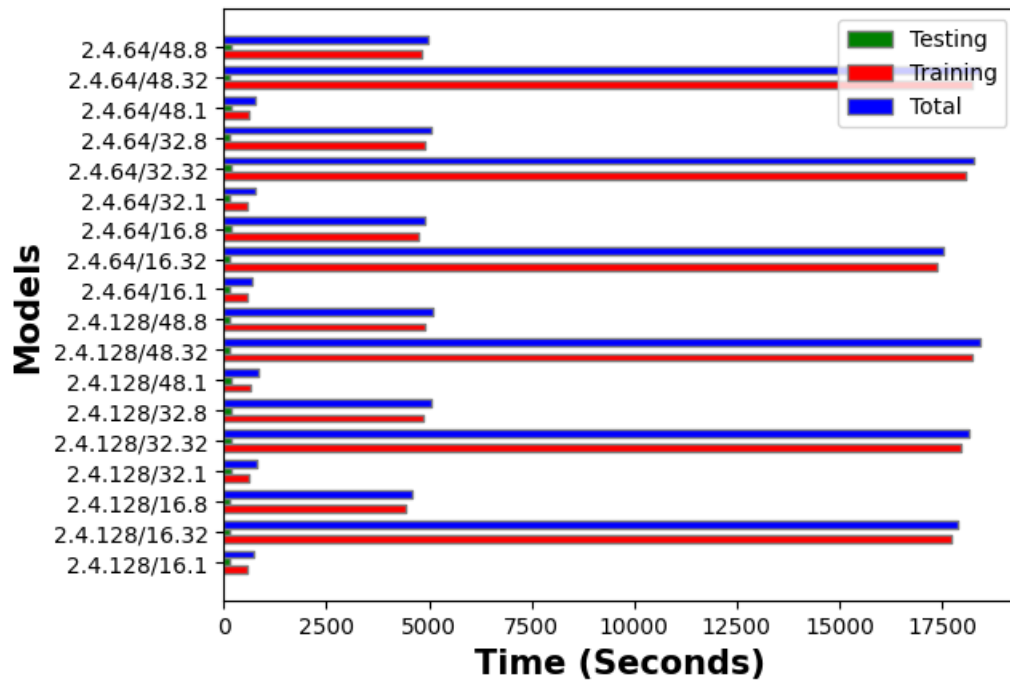


Figure 19: Recorded time metrics.

### B.2.2 Training Loss

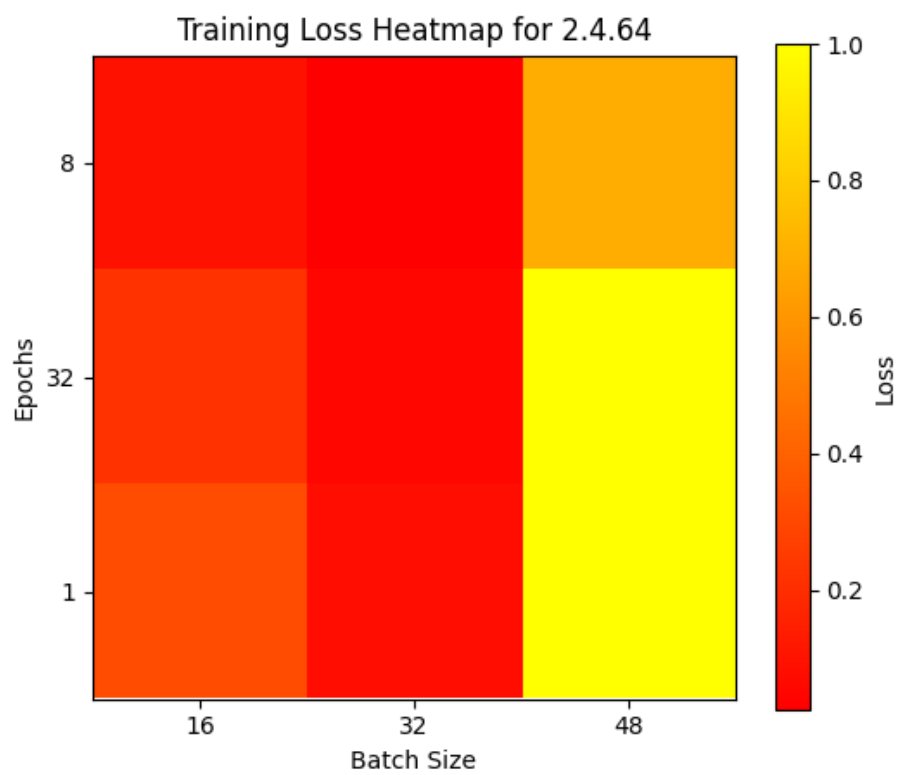


Figure 20: Heatmap of loss during training for models with architecture 2.4.64.

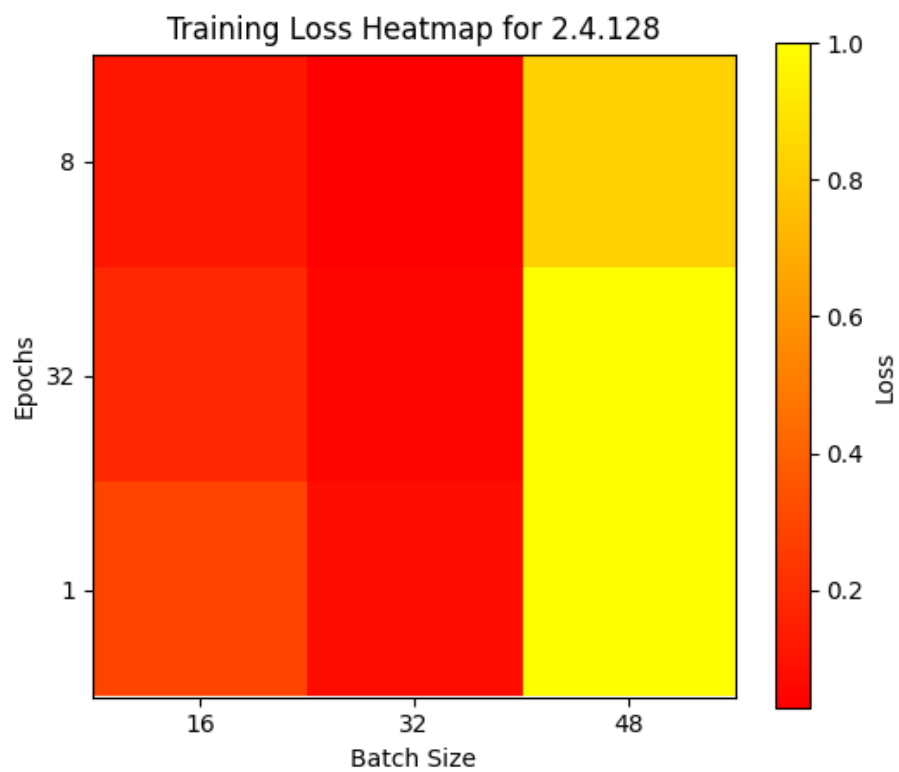


Figure 21: Heatmap of loss during training for models with architecture 2.4.128

### B.2.3 Training Accuracy

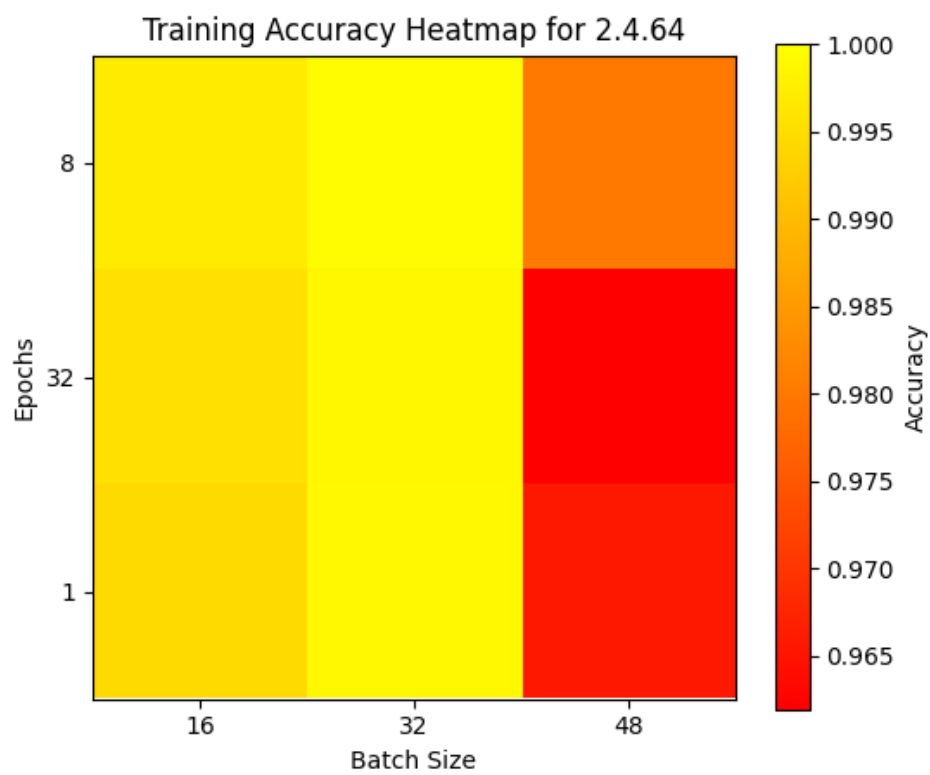


Figure 22: Heatmap of accuracy during training for models with architecture 2.4.64.

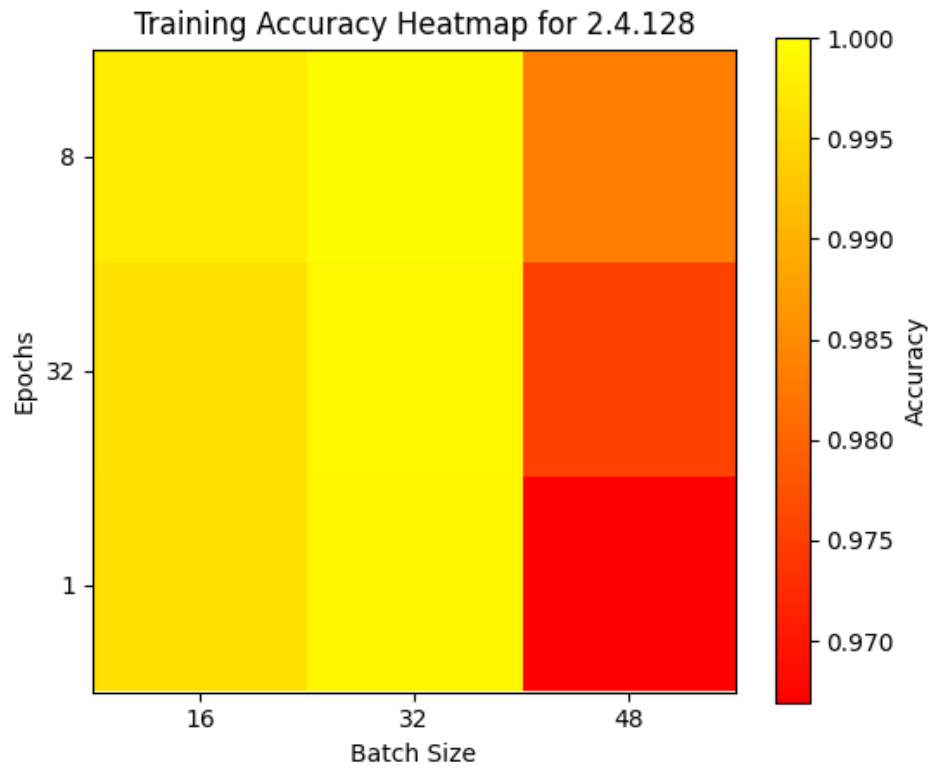


Figure 23: Heatmap of accuracy during training for models with architecture 2.4.128

### B.2.4 Testing Accuracy

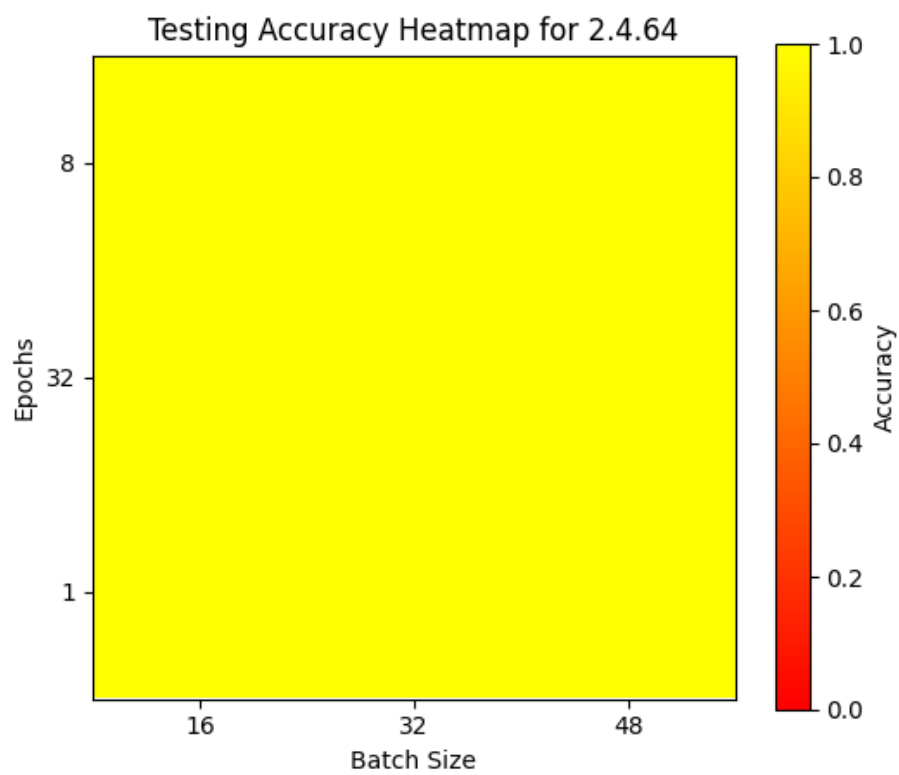


Figure 24: Heatmap of accuracy during testing for models with architecture 2.4.64.

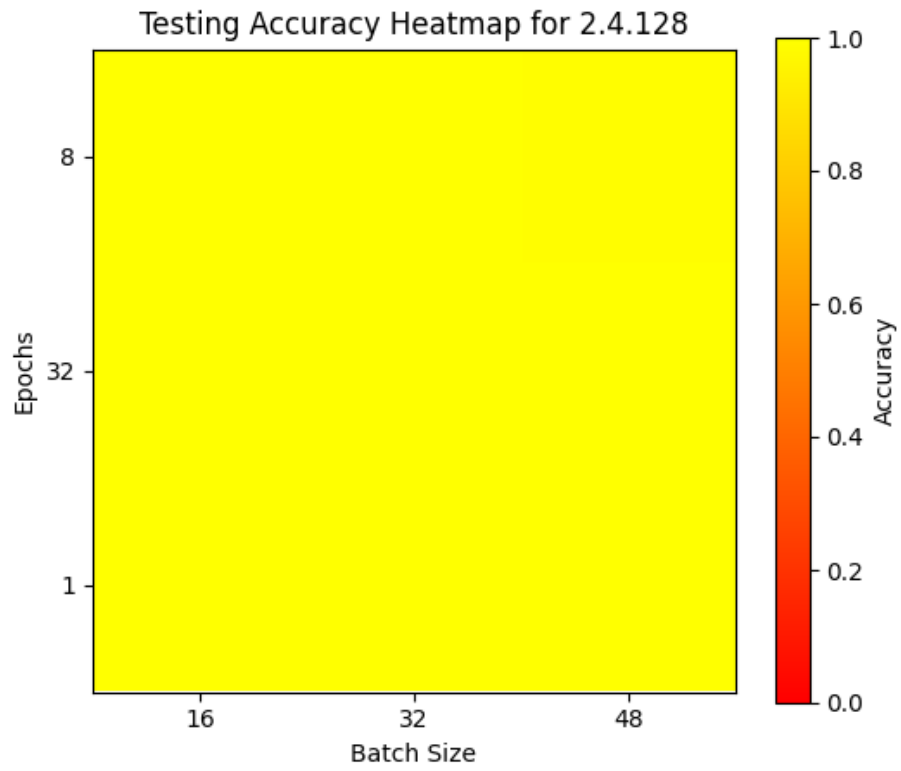


Figure 25: Heatmap of accuracy during testing for models with architecture 2.4.128



### B.2.5 Testing Recall

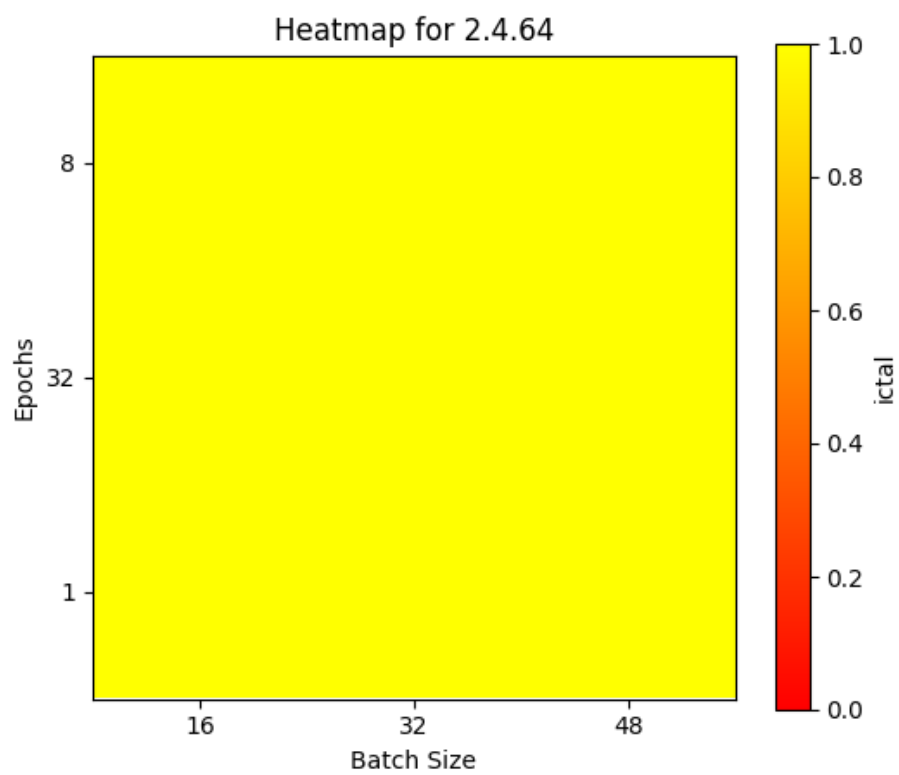


Figure 26: Heatmap of recall for ictal class during testing for models with architecture 2.4.64.

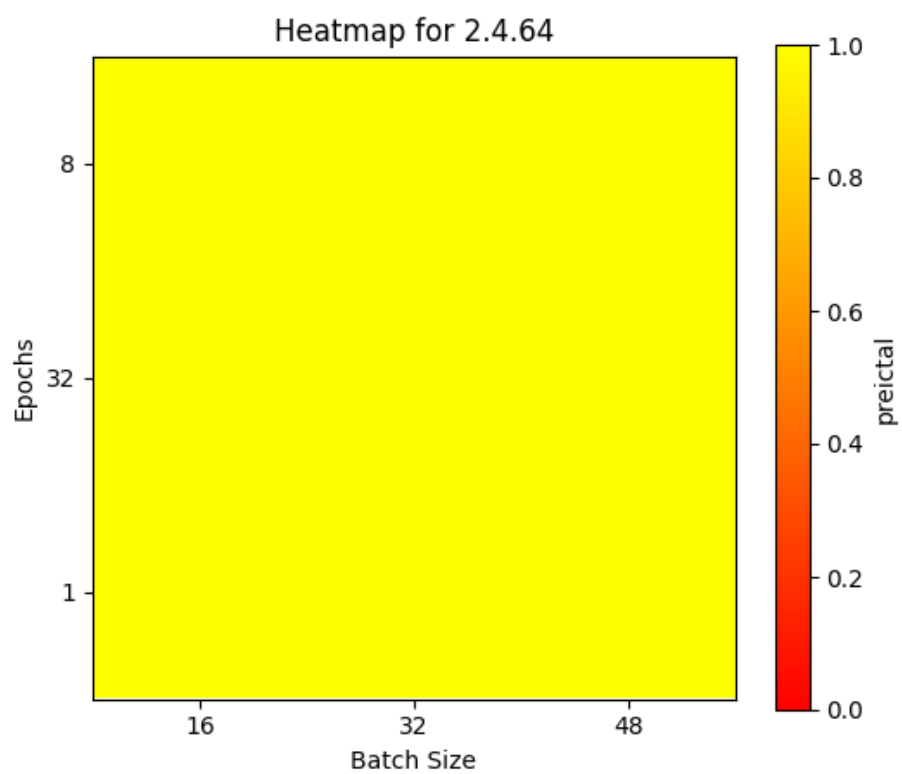


Figure 27: Heatmap of recall for preictal class during testing for models with architecture 2.4.64.

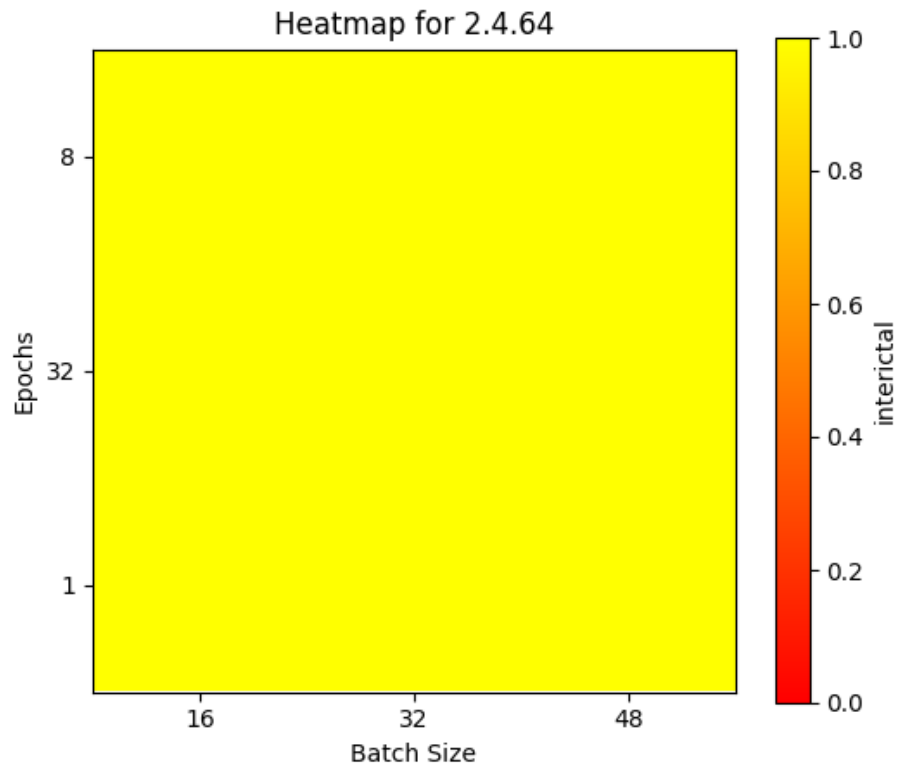


Figure 28: Heatmap of recall for interictal class during testing for models with architecture 2.4.64.

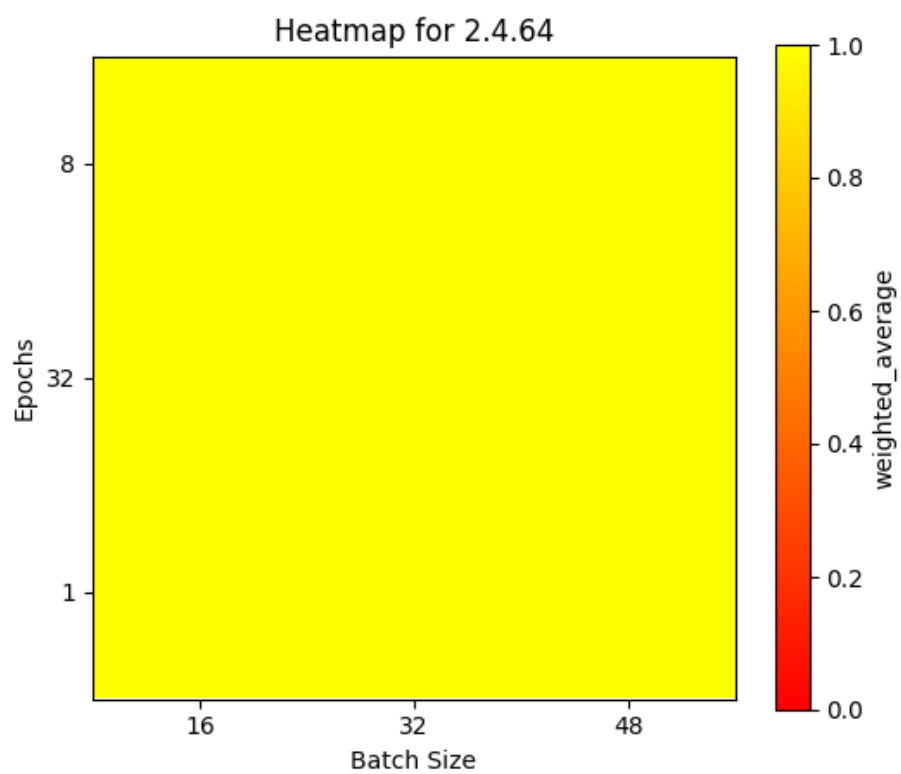


Figure 29: Heatmap of recall (weighted average) during testing for models with architecture 2.4.64.

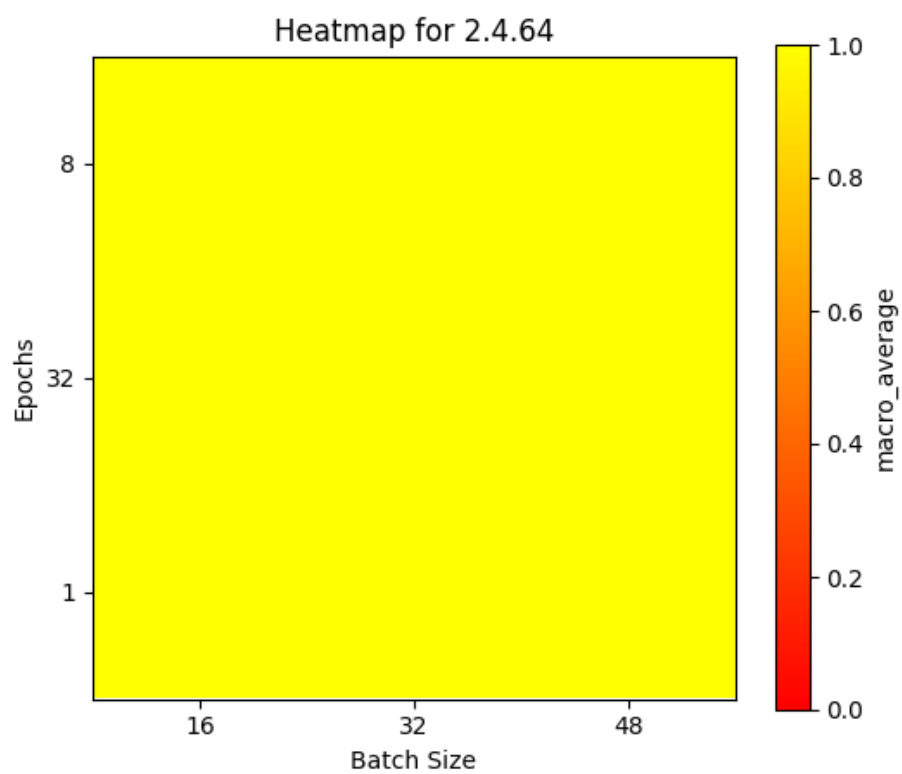


Figure 30: Heatmap of recall (macro average) during testing for models with architecture 2.4.64.

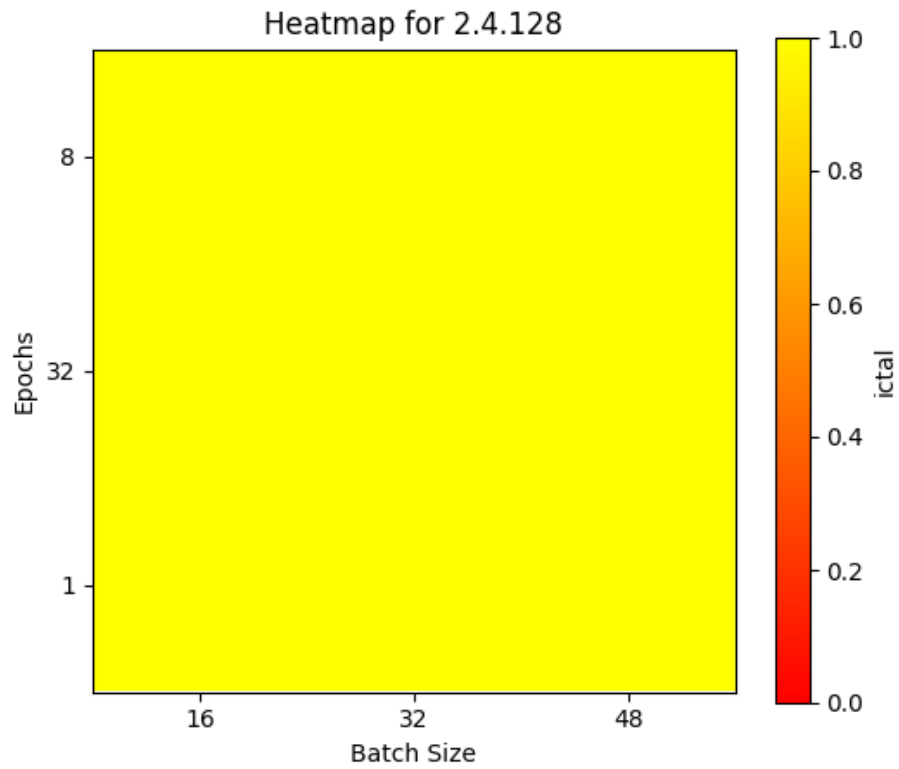


Figure 31: Heatmap of recall for ictal class during testing for models with architecture 2.4.128.

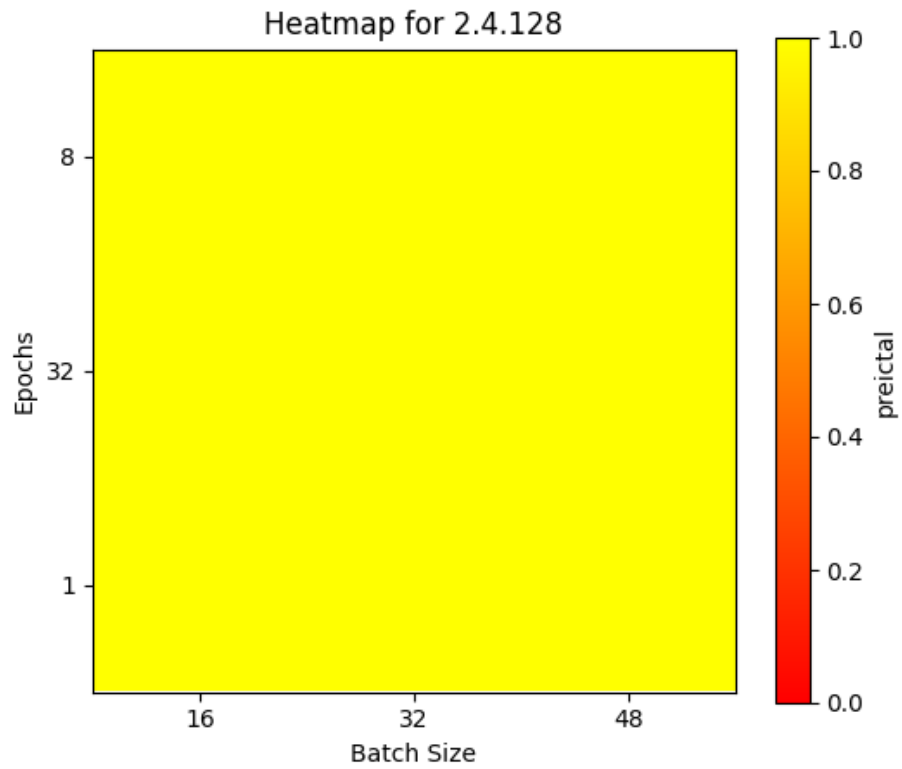


Figure 32: Heatmap of recall for preictal class during testing for models with architecture 2.4.128.

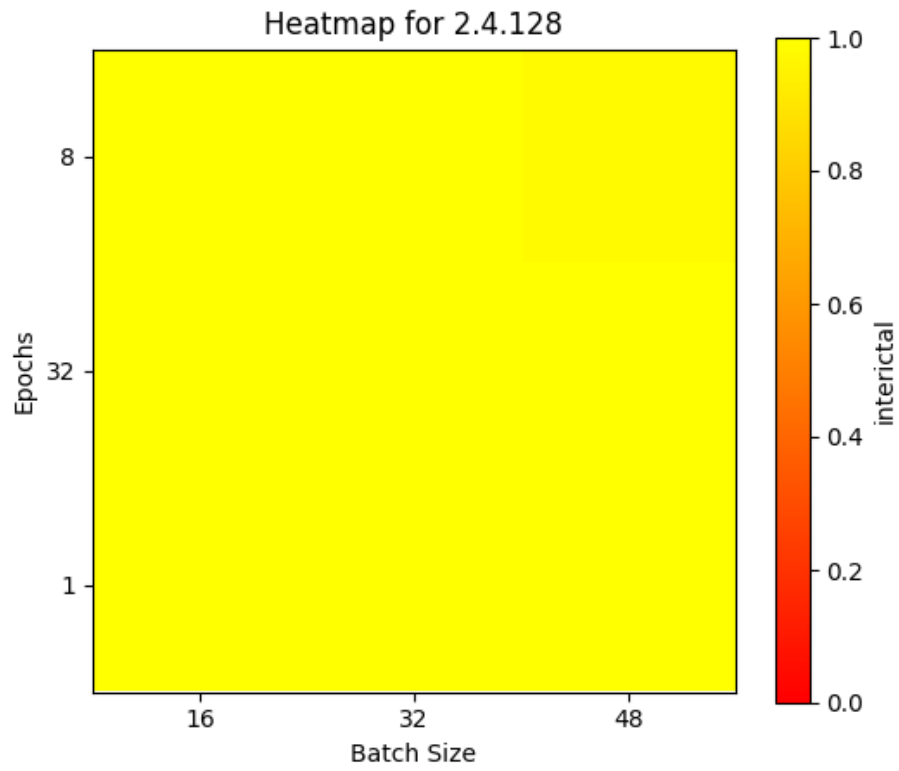


Figure 33: Heatmap of recall for interictal class during testing for models with architecture 2.4.128.



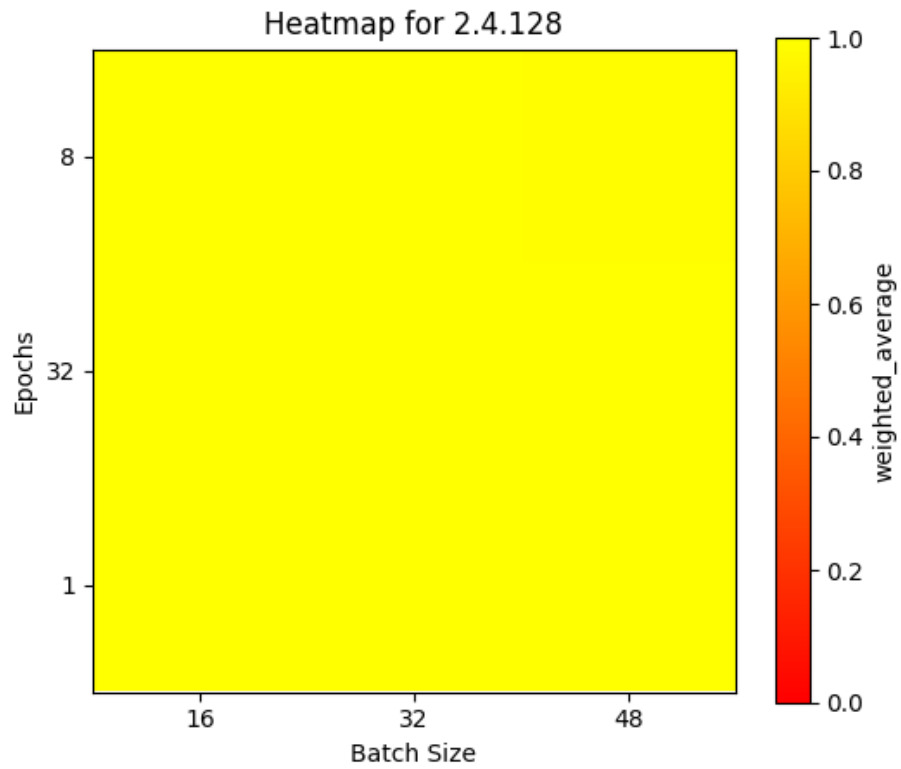


Figure 34: Heatmap of recall (weighted average) during testing for models with architecture 2.4.128.

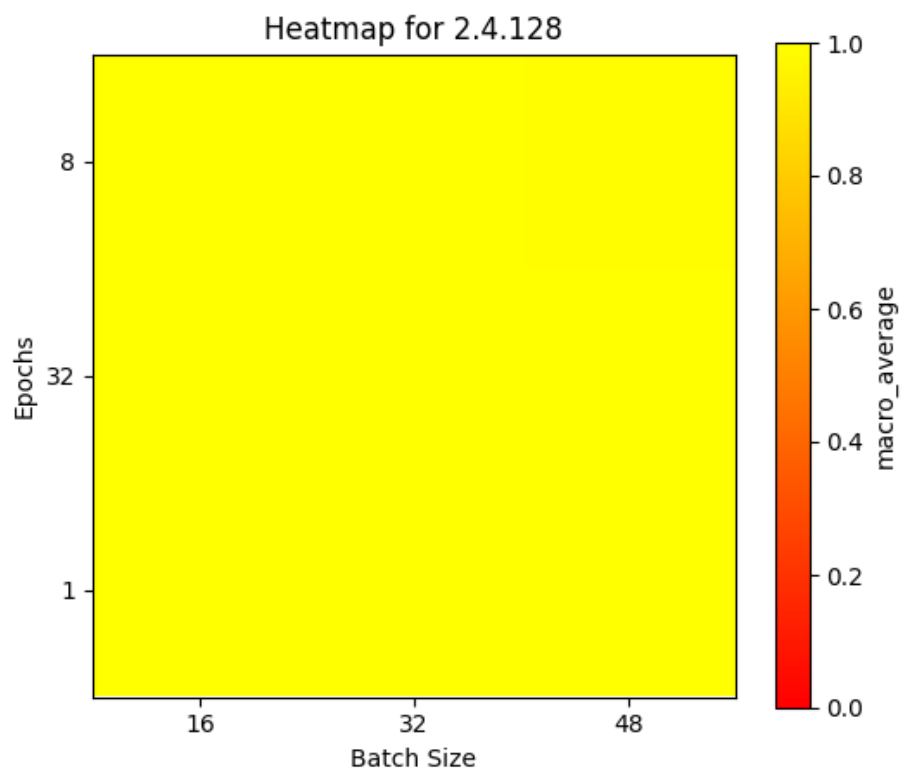


Figure 35: Heatmap of recall (macro average) during testing for models with architecture 2.4.128.

### B.2.6 Testing F1

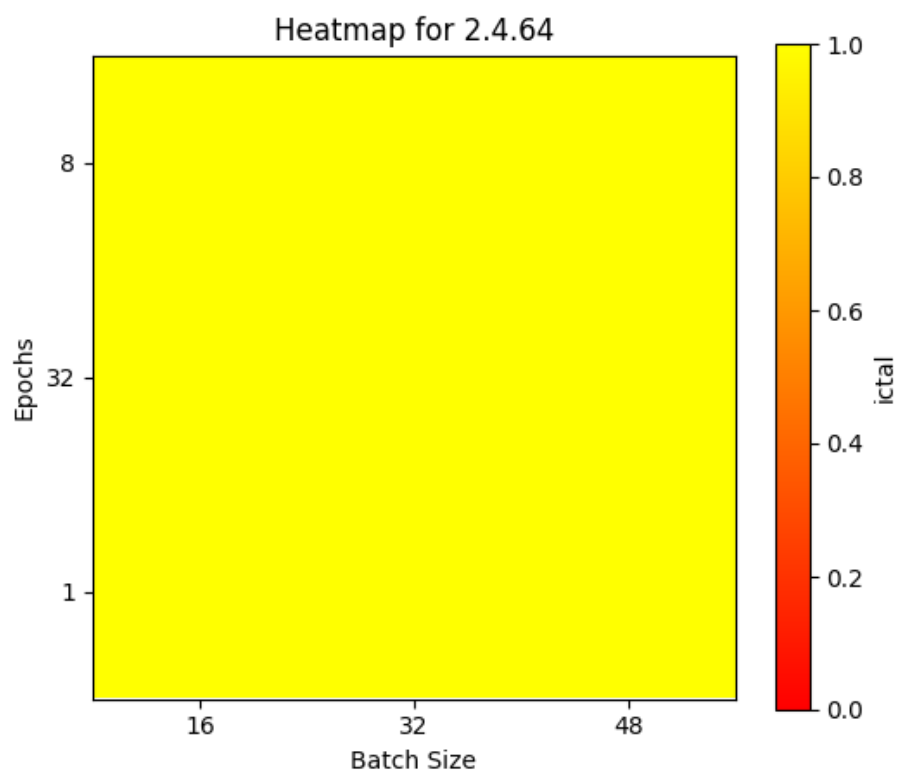


Figure 36: Heatmap of F1 for ictal class during testing for models with architecture 2.4.64.

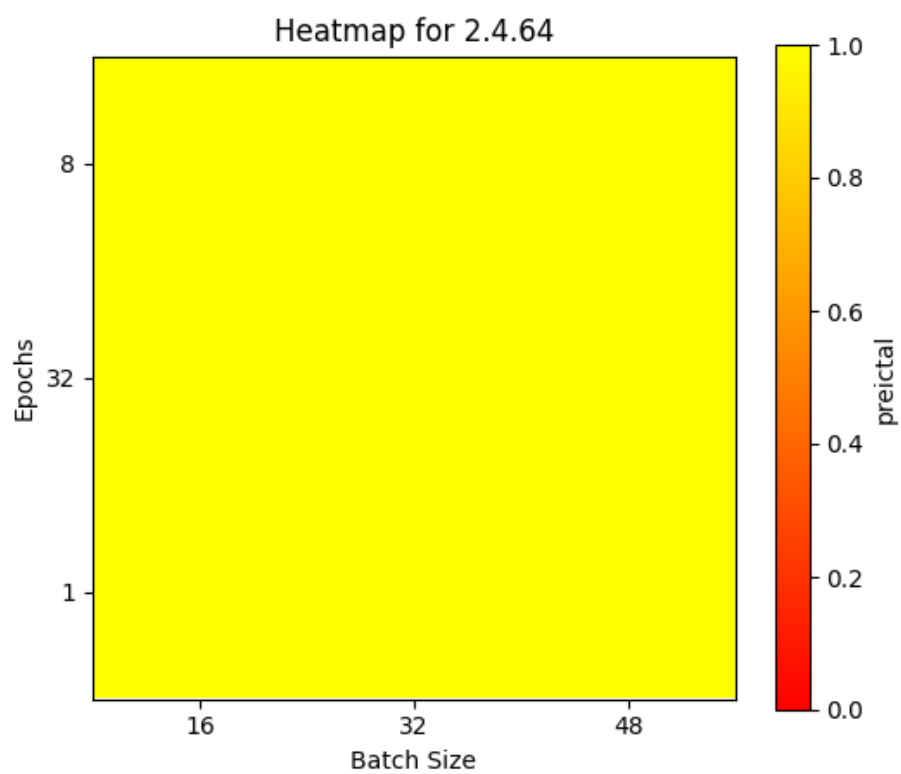


Figure 37: Heatmap of F1 for preictal class during testing for models with architecture 2.4.64.

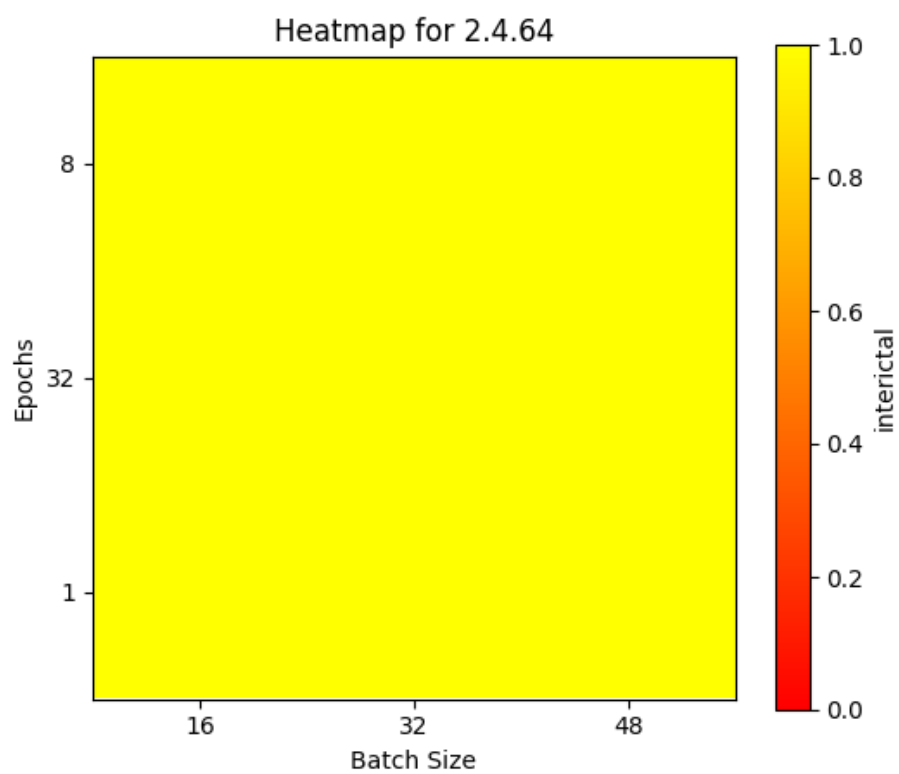


Figure 38: Heatmap of F1 for interictal class during testing for models with architecture 2.4.64.

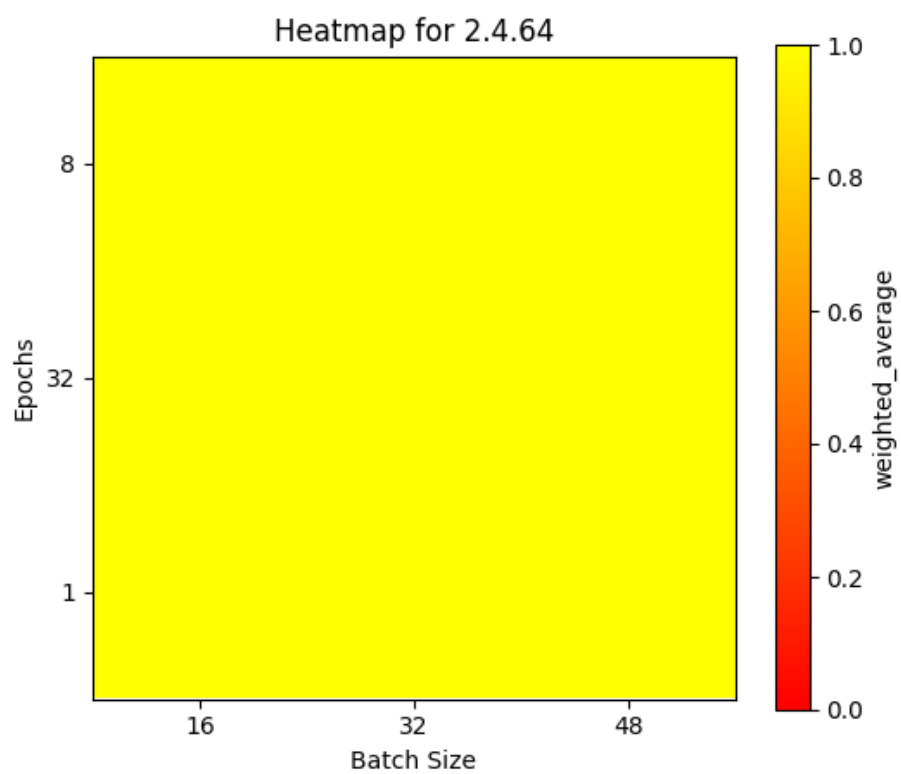


Figure 39: Heatmap of F1 (weighted average) during testing for models with architecture 2.4.64.

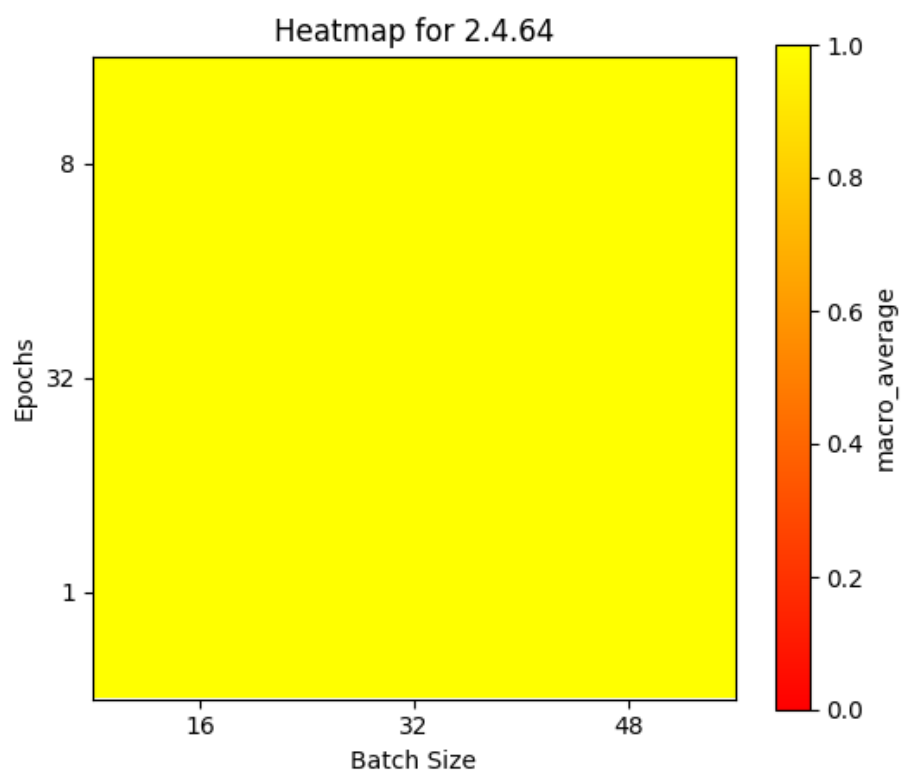


Figure 40: Heatmap of F1 (macro average) during testing for models with architecture 2.4.64.

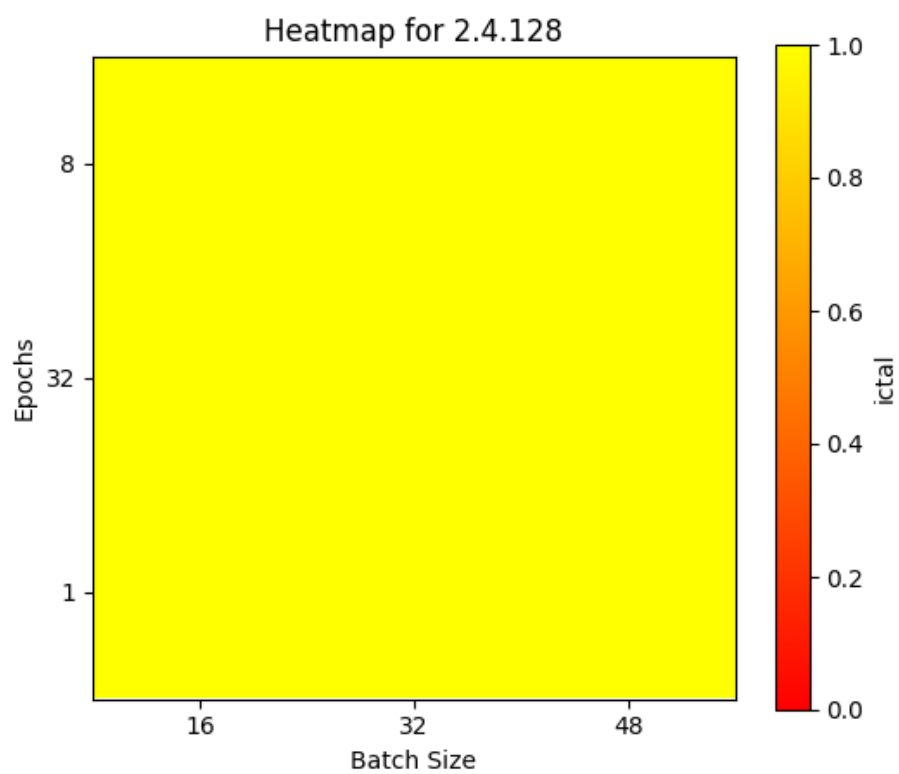


Figure 41: Heatmap of F1 for ictal class during testing for models with architecture 2.4.128.



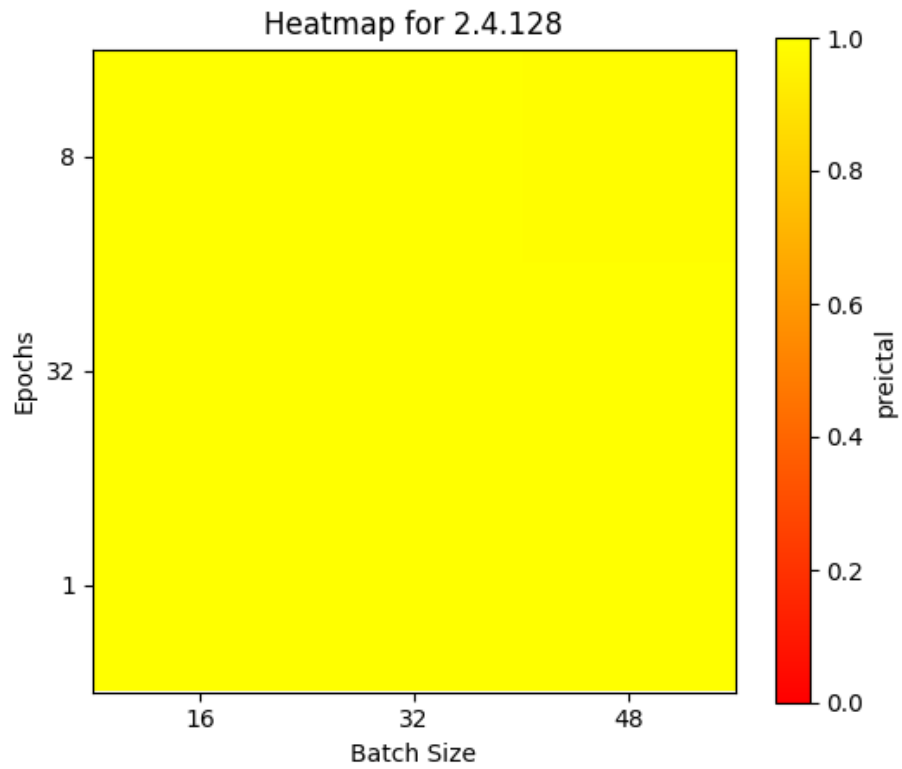


Figure 42: Heatmap of F1 for preictal class during testing for models with architecture 2.4.128.

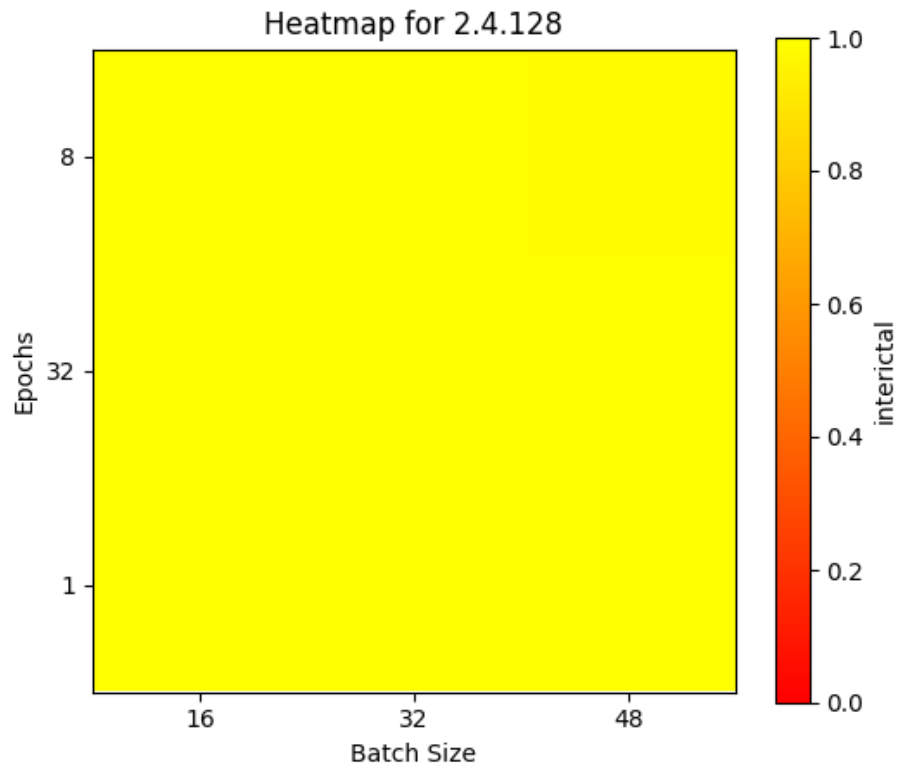


Figure 43: Heatmap of F1 for interictal class during testing for models with architecture 2.4.128.

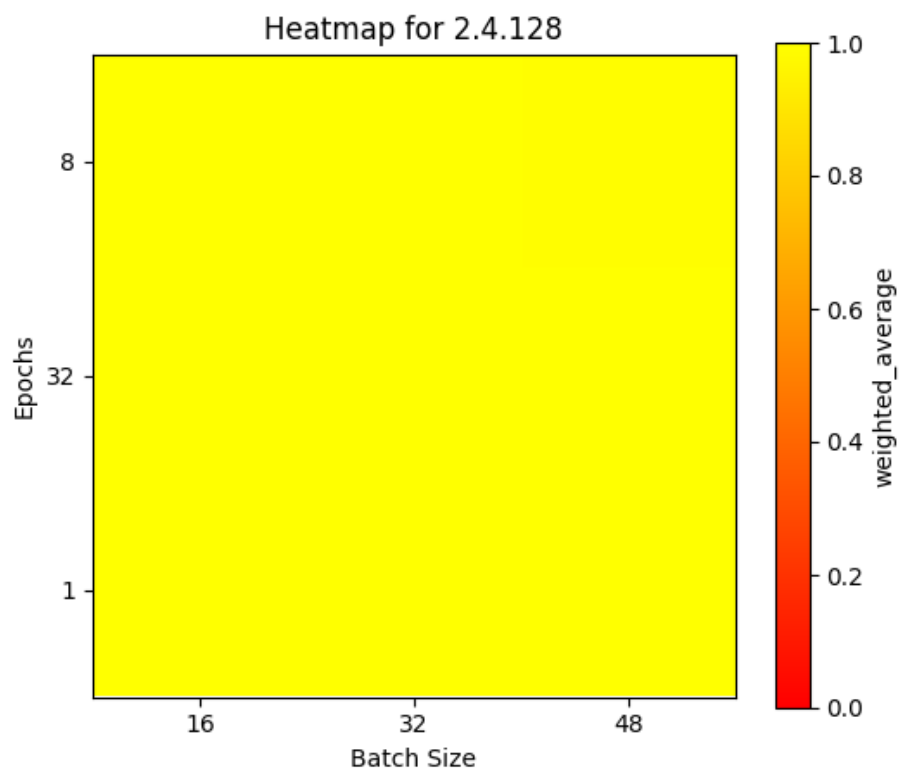


Figure 44: Heatmap of F1 (weighted average) during testing for models with architecture 2.4.128.

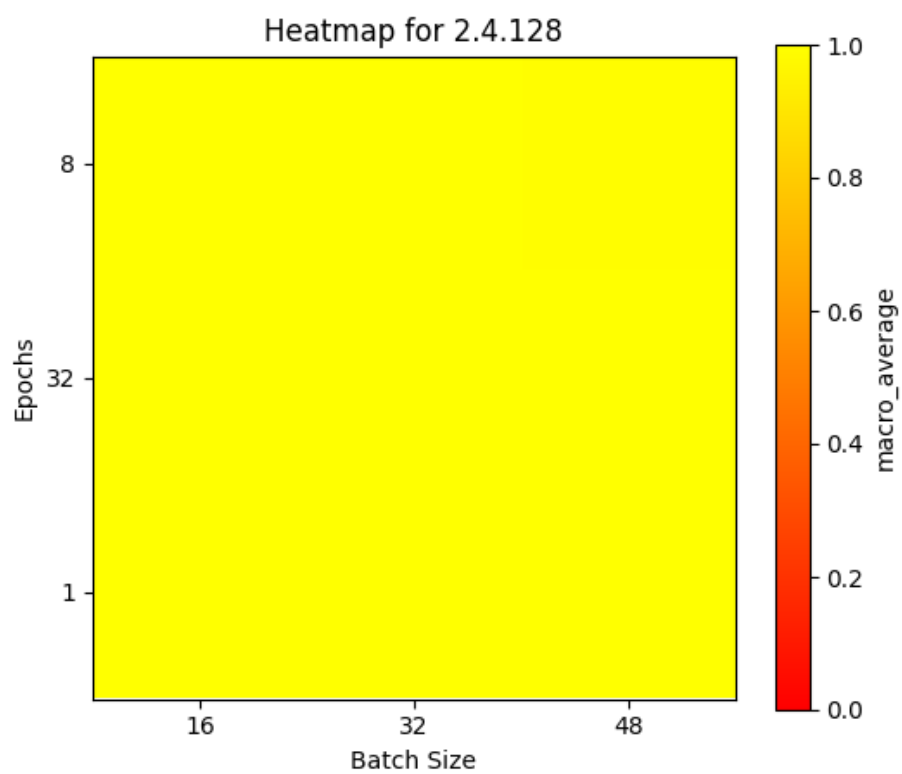


Figure 45: Heatmap of F1 (macro average) during testing for models with architecture 2.4.128.

### B.2.7 Testing Precision

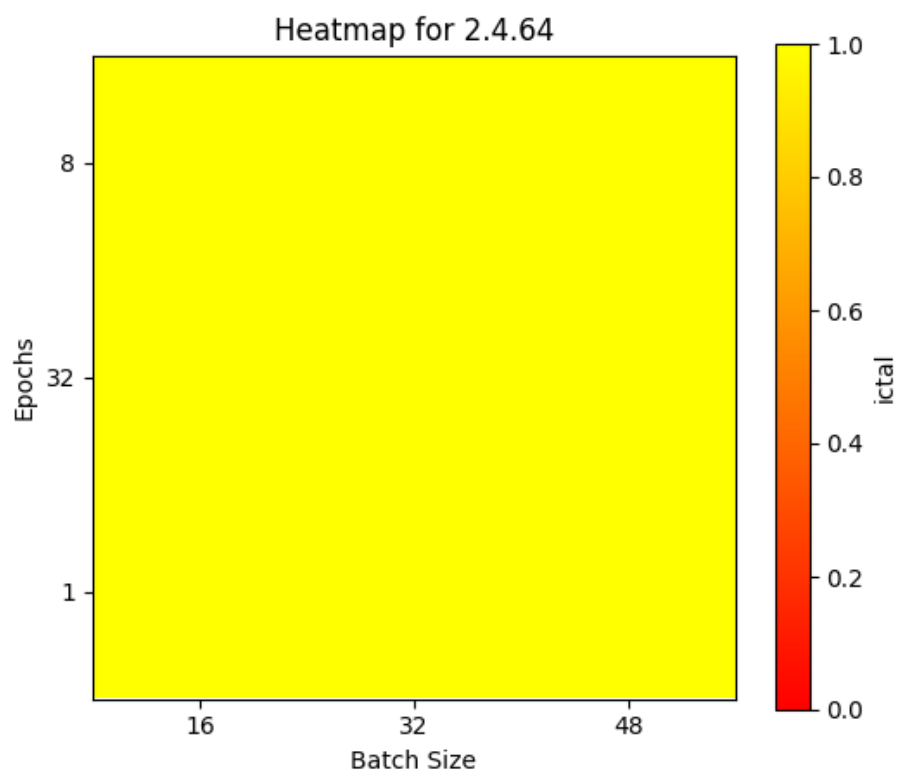


Figure 46: Heatmap of precision for ictal class during testing for models with architecture 2.4.64.

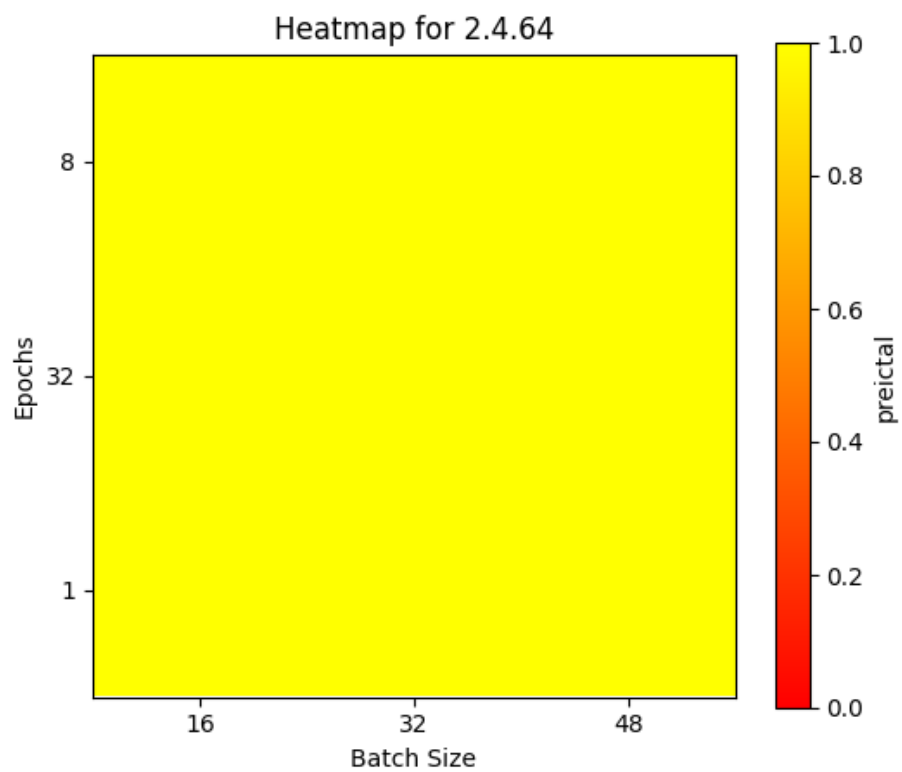


Figure 47: Heatmap of precision for preictal class during testing for models with architecture 2.4.64.

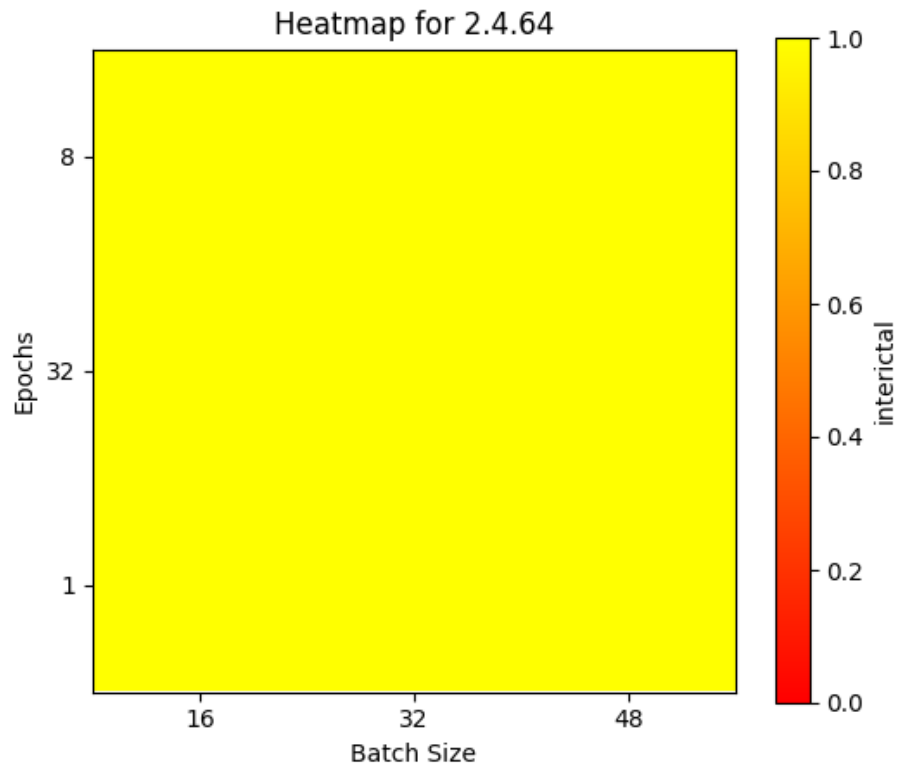


Figure 48: Heatmap of recall for interictal class during testing for models with architecture 2.4.64.

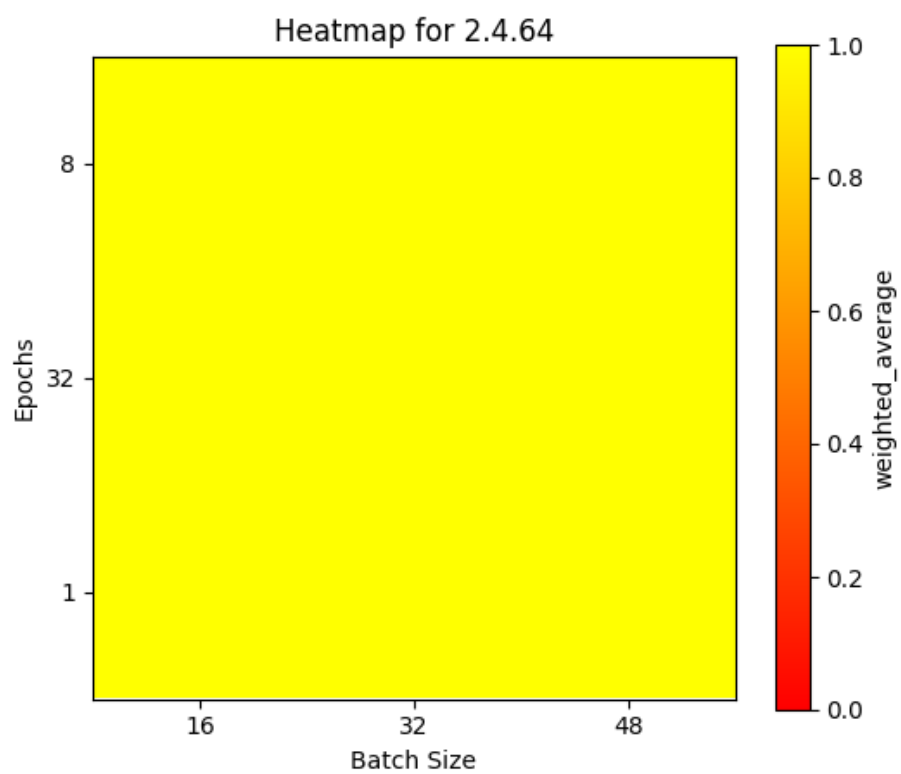


Figure 49: Heatmap of precision (weighted average) during testing for models with architecture 2.4.64.



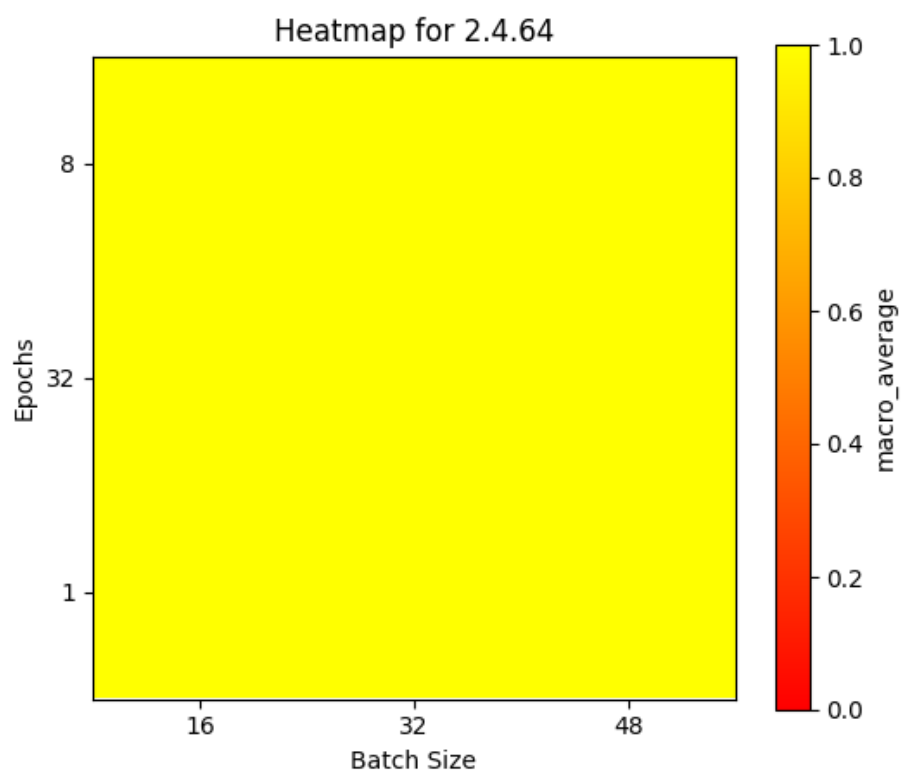


Figure 50: Heatmap of precision (macro average) during testing for models with architecture 2.4.64.

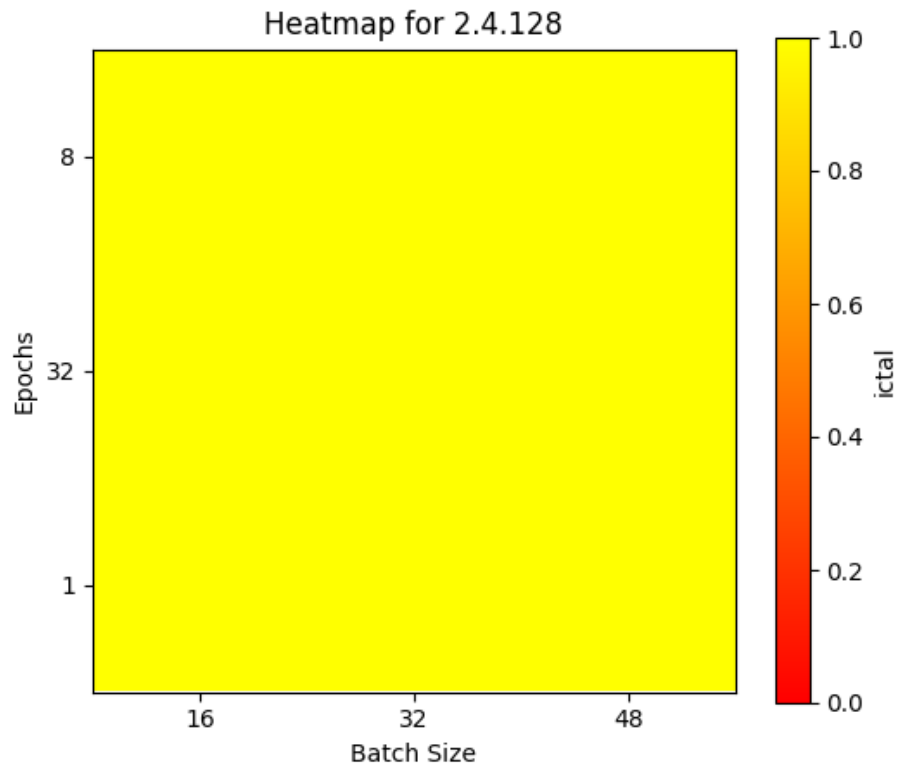


Figure 51: Heatmap of precision for ictal class during testing for models with architecture 2.4.128.

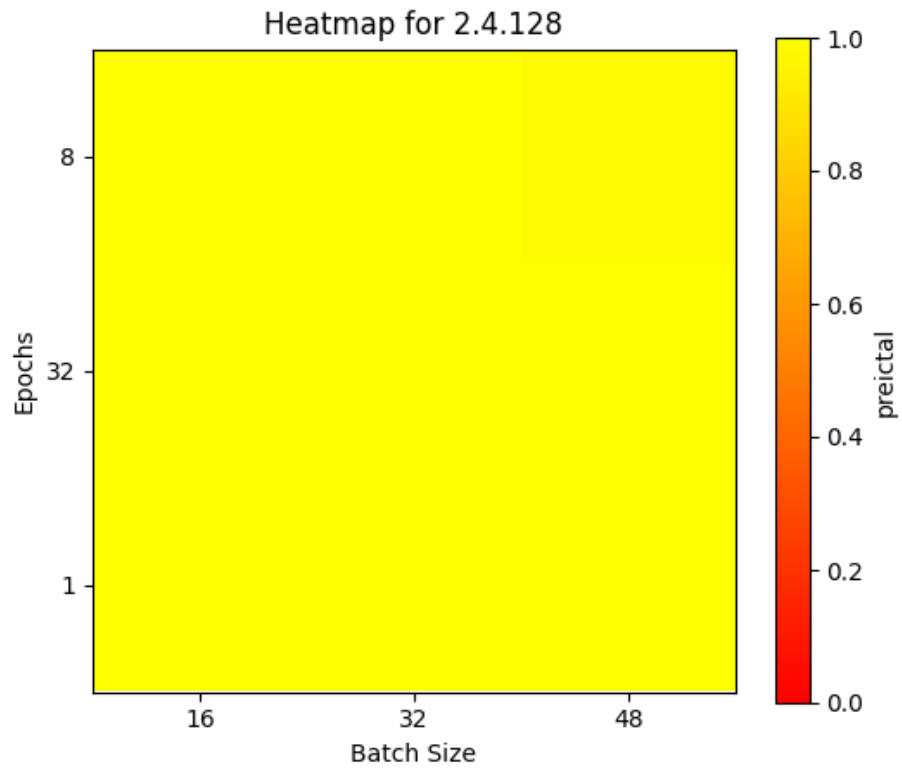


Figure 52: Heatmap of precision for preictal class during testing for models with architecture 2.4.128.

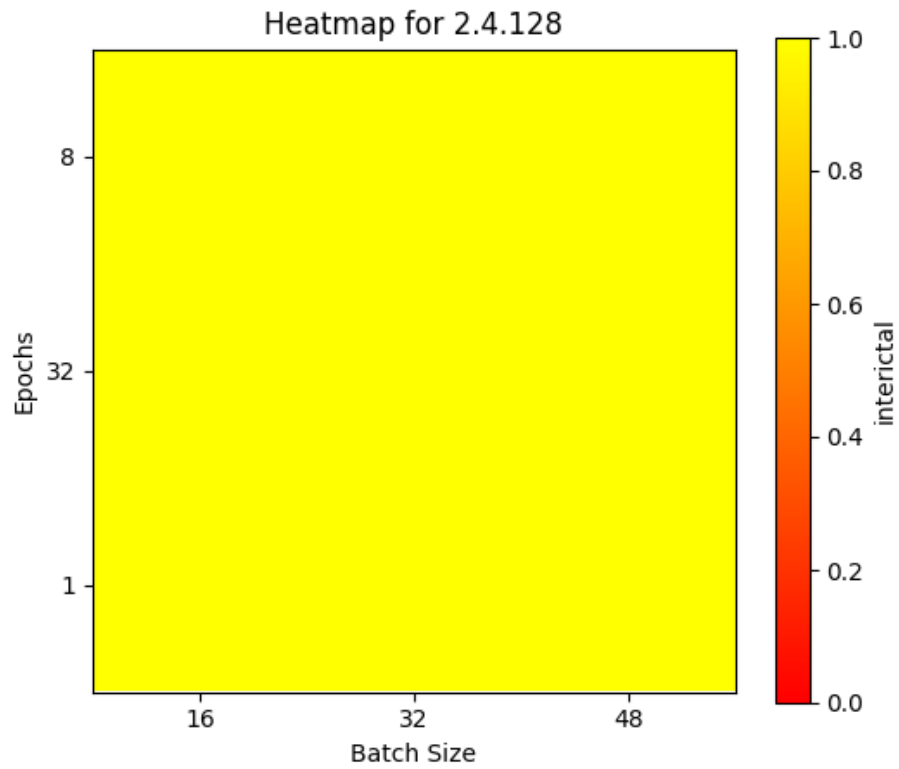


Figure 53: Heatmap of precision for interictal class during testing for models with architecture 2.4.128.

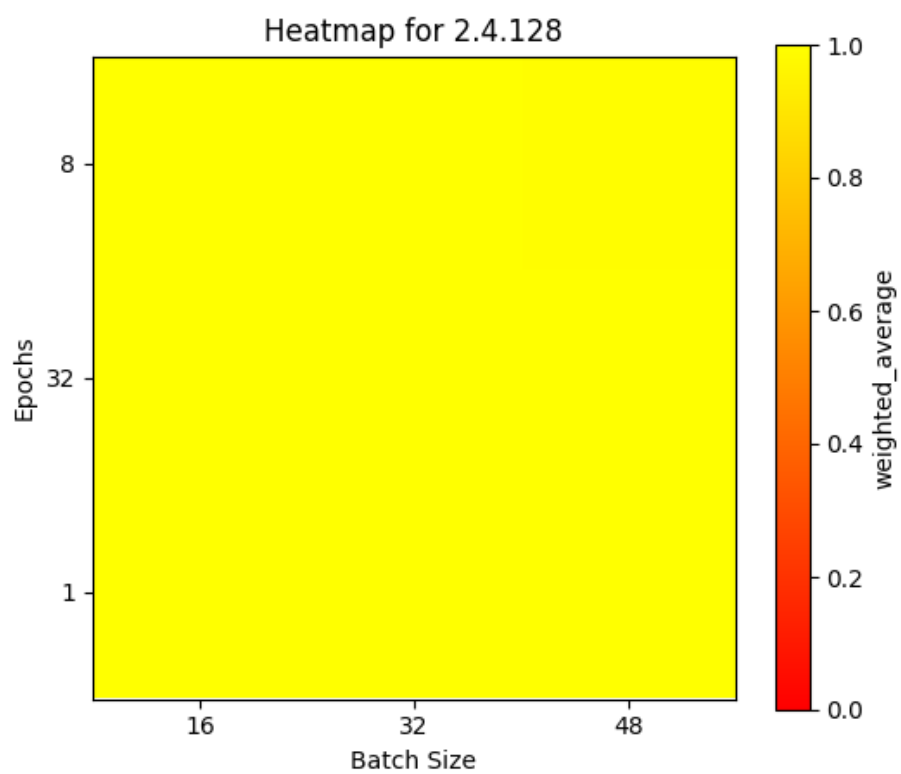


Figure 54: Heatmap of precision (weighted average) during testing for models with architecture 2.4.128.

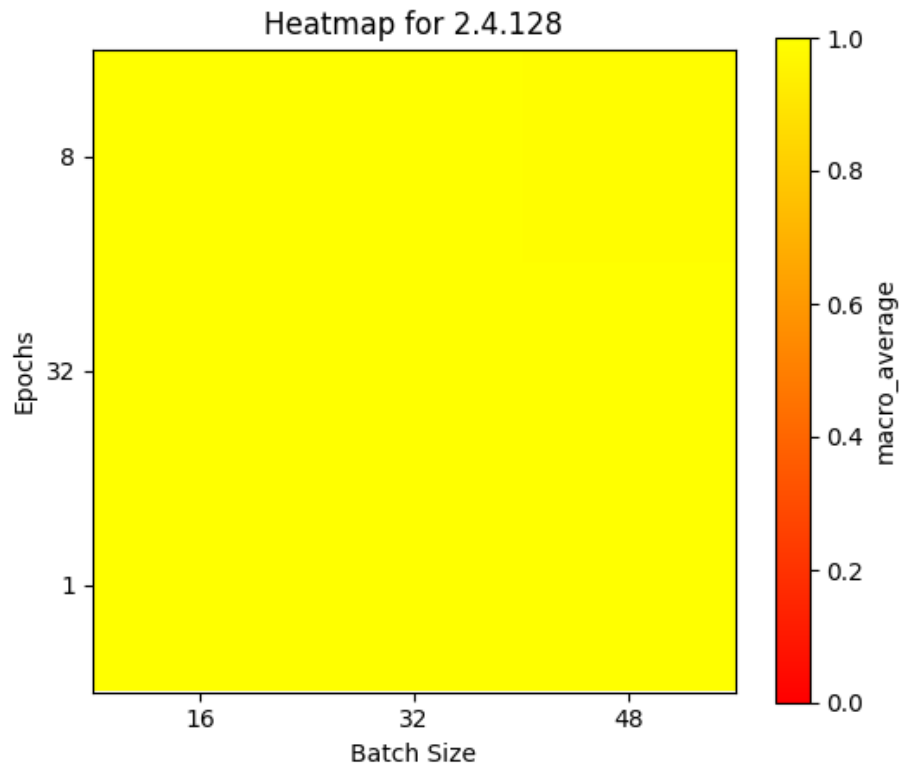


Figure 55: Heatmap of precision (macro average) during testing for models with architecture 2.4.128.

## C Misc.

Listing 3: Example Summary file from the CHB-MIT EEG dataset.

```
File Name: chb06_03.edf
File Start Time: 03:09:42
File End Time: 7:09:42
Number of Seizures in File: 0

File Name: chb06_04.edf
File Start Time: 07:09:51
File End Time: 10:50:52
```

Number of Seizures in File: 2  
Seizure 1 Start Time: 327 seconds  
Seizure 1 End Time: 347 seconds  
Seizure 2 Start Time: 6211 seconds  
Seizure 2 End Time: 6231 seconds

File Name: chb06\_05.edf  
File Start Time: 10:51:20  
File End Time: 14:51:20  
Number of Seizures in File: 0

File Name: chb06\_06.edf  
File Start Time: 14:51:23  
File End Time: 18:51:23  
Number of Seizures in File: 0

File Name: chb06\_07.edf  
File Start Time: 18:51:31  
File End Time: 22:51:31  
Number of Seizures in File: 0

File Name: chb06\_08.edf  
File Start Time: 22:51:39  
File End Time: 26:51:39  
Number of Seizures in File: 0

File Name: chb06\_09.edf  
File Start Time: 02:51:47  
File End Time: 6:51:47  
Number of Seizures in File: 1  
Seizure 1 Start Time: 12500 seconds  
Seizure 1 End Time: 12516 seconds

File Name: chb06\_10.edf  
File Start Time: 06:51:54  
File End Time: 10:51:54  
Number of Seizures in File: 1  
Seizure 1 Start Time: 10833 seconds  
Seizure 1 End Time: 10845 seconds

File Name: chb06\_12.edf  
File Start Time: 14:52:10  
File End Time: 18:52:10  
Number of Seizures in File: 0

File Name: chb06\_13.edf  
File Start Time: 18:52:20  
File End Time: 22:52:20  
Number of Seizures in File: 1  
Seizure 1 Start Time: 506 seconds  
Seizure 1 End Time: 519 seconds

File Name: chb06\_14.edf  
File Start Time: 22:52:35  
File End Time: 26:52:35  
Number of Seizures in File: 0

File Name: chb06\_15.edf  
File Start Time: 02:52:43  
File End Time: 6:52:43  
Number of Seizures in File: 0

File Name: chb06\_16.edf  
File Start Time: 06:52:51  
File End Time: 7:43:21  
Number of Seizures in File: 0

File Name: chb06\_17.edf  
File Start Time: 07:45:51  
File End Time: 11:45:51  
Number of Seizures in File: 0

File Name: chb06\_18.edf  
File Start Time: 11:45:55  
File End Time: 13:58:03  
Number of Seizures in File: 1  
Seizure 1 Start Time: 7799 seconds  
Seizure 1 End Time: 7811 seconds

File Name: chb06\_24.edf



File Start Time: 08:23:24  
File End Time: 12:23:24  
Number of Seizures in File: 1  
Seizure 1 Start Time: 9387 seconds  
Seizure 1 End Time: 9403 seconds

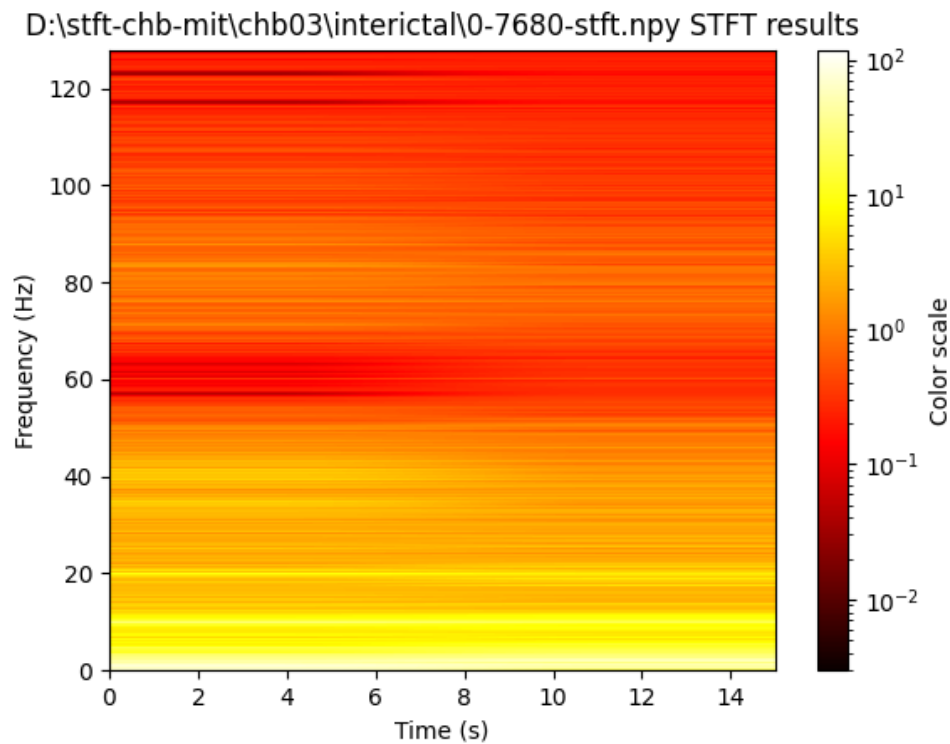


Figure 56: An STFT window with high time resolution.

Listing 4: List of failed trained models. See 7.1

```
/data/results-archive/2.4.128/16.1.keras  
/data/results-archive/2.4.128/16.32.keras  
/data/results-archive/2.4.128/16.8.keras  
/data/results-archive/2.4.128/32.1.keras  
/data/results-archive/2.4.128/32.32.keras  
/data/results-archive/2.4.128/32.8.keras  
/data/results-archive/2.4.128/48.1.keras
```

/data/results-archive/2.4.128/48.8.keras  
/data/results-archive/2.4.256/16.1.keras  
/data/results-archive/2.4.256/16.32.keras  
/data/results-archive/2.4.256/16.8.keras  
/data/results-archive/2.4.256/32.1.keras  
/data/results-archive/2.4.256/32.32.keras  
/data/results-archive/2.4.256/32.8.keras  
/data/results-archive/2.4.256/48.1.keras  
/data/results-archive/2.4.256/48.32.keras  
/data/results-archive/2.4.256/48.8.keras  
/data/results-archive/2.4.64/16.1.keras  
/data/results-archive/2.4.64/16.32.keras  
/data/results-archive/2.4.64/16.8.keras  
/data/results-archive/2.4.64/32.1.keras  
/data/results-archive/2.4.64/32.32.keras  
/data/results-archive/2.4.64/32.8.keras  
/data/results-archive/2.4.64/48.1.keras  
/data/results-archive/2.4.64/48.32.keras  
/data/results-archive/2.4.64/48.8.keras  
/data/results-archive/2.8.64/16.1.keras  
/data/results-archive/2.8.64/16.32.keras  
/data/results-archive/2.8.64/16.8.keras  
/data/results-archive/2.8.64/32.1.keras  
/data/results-archive/2.8.64/32.32.keras  
/data/results-archive/2.8.64/32.8.keras  
/data/results-archive/2.8.64/48.1.keras  
/data/results-archive/2.8.64/48.8.keras  
/data/results-archive/48.32.keras  
/data/results-archive/ictal-only-prediction/2.12.128/16.1.keras  
/data/results-archive/ictal-only-prediction/2.12.128/16.8.keras  
/data/results-archive/ictal-only-prediction/2.12.128/32.1.keras  
/data/results-archive/ictal-only-prediction/2.12.128/32.8.keras  
/data/results-archive/ictal-only-prediction/2.12.256/16.1.keras  
/data/results-archive/ictal-only-prediction/2.12.256/16.8.keras  
/data/results-archive/ictal-only-prediction/2.12.256/32.1.keras  
/data/results-archive/ictal-only-prediction/2.12.256/32.8.keras  
/data/results-archive/ictal-only-prediction/2.12.64/16.1.keras  
/data/results-archive/ictal-only-prediction/2.12.64/16.8.keras  
/data/results-archive/ictal-only-prediction/2.12.64/32.1.keras  
/data/results-archive/ictal-only-prediction/2.12.64/32.8.keras

/data/results-archive/ictal-only-prediction/2.4.128/16.1.keras  
/data/results-archive/ictal-only-prediction/2.4.128/16.8.keras  
/data/results-archive/ictal-only-prediction/2.4.128/32.1.keras  
/data/results-archive/ictal-only-prediction/2.4.128/32.8.keras  
/data/results-archive/ictal-only-prediction/2.4.256/16.1.keras  
/data/results-archive/ictal-only-prediction/2.4.256/16.8.keras  
/data/results-archive/ictal-only-prediction/2.4.256/32.1.keras  
/data/results-archive/ictal-only-prediction/2.4.256/32.8.keras  
/data/results-archive/ictal-only-prediction/2.4.64/16.1.keras  
/data/results-archive/ictal-only-prediction/2.4.64/16.8.keras  
/data/results-archive/ictal-only-prediction/2.4.64/32.1.keras  
/data/results-archive/ictal-only-prediction/2.4.64/32.8.keras  
/data/results-archive/ictal-only-prediction/2.8.128/16.1.keras  
/data/results-archive/ictal-only-prediction/2.8.128/16.8.keras  
/data/results-archive/ictal-only-prediction/2.8.128/32.1.keras  
/data/results-archive/ictal-only-prediction/2.8.128/32.8.keras  
/data/results-archive/ictal-only-prediction/2.8.256/16.1.keras  
/data/results-archive/ictal-only-prediction/2.8.256/16.8.keras  
/data/results-archive/ictal-only-prediction/2.8.256/32.1.keras  
/data/results-archive/ictal-only-prediction/2.8.256/32.8.keras  
/data/results-archive/ictal-only-prediction/2.8.64/16.1.keras  
/data/results-archive/ictal-only-prediction/2.8.64/16.8.keras  
/data/results-archive/ictal-only-prediction/2.8.64/32.1.keras  
/data/results-archive/ictal-only-prediction/2.8.64/32.8.keras  
/data/results-archive/imbalanced\_training/2.4.64/16.1.keras  
/data/results-archive/imbalanced\_training/2.4.64/16.8.keras  
/data/results-archive/imbalanced\_training/2.4.64/32.1.keras  
/data/results-archive/imbalanced\_training/2.4.64/32.8.keras  
/data/results-archive/imbalanced\_training\_2/2.12.128/16.1.keras  
/data/results-archive/imbalanced\_training\_2/2.12.128/16.8.keras  
/data/results-archive/imbalanced\_training\_2/2.12.128/32.1.keras  
/data/results-archive/imbalanced\_training\_2/2.12.128/32.8.keras  
/data/results-archive/imbalanced\_training\_2/2.12.256/16.1.keras  
/data/results-archive/imbalanced\_training\_2/2.12.256/16.8.keras  
/data/results-archive/imbalanced\_training\_2/2.12.256/32.1.keras  
/data/results-archive/imbalanced\_training\_2/2.12.256/32.8.keras  
/data/results-archive/imbalanced\_training\_2/2.12.64/16.1.keras  
/data/results-archive/imbalanced\_training\_2/2.12.64/16.8.keras  
/data/results-archive/imbalanced\_training\_2/2.12.64/32.1.keras  
/data/results-archive/imbalanced\_training\_2/2.12.64/32.8.keras

/data/results-archive/imbalanced\_training\_2/2.4.128/16.1.keras  
/data/results-archive/imbalanced\_training\_2/2.4.128/16.8.keras  
/data/results-archive/imbalanced\_training\_2/2.4.128/32.1.keras  
/data/results-archive/imbalanced\_training\_2/2.4.128/32.8.keras  
/data/results-archive/imbalanced\_training\_2/2.4.256/16.1.keras  
/data/results-archive/imbalanced\_training\_2/2.4.256/16.8.keras  
/data/results-archive/imbalanced\_training\_2/2.4.256/32.1.keras  
/data/results-archive/imbalanced\_training\_2/2.4.256/32.8.keras  
/data/results-archive/imbalanced\_training\_2/2.4.64/16.1.keras  
/data/results-archive/imbalanced\_training\_2/2.4.64/16.8.keras  
/data/results-archive/imbalanced\_training\_2/2.4.64/32.1.keras  
/data/results-archive/imbalanced\_training\_2/2.4.64/32.8.keras  
/data/results-archive/imbalanced\_training\_2/2.8.128/16.1.keras  
/data/results-archive/imbalanced\_training\_2/2.8.128/16.8.keras  
/data/results-archive/imbalanced\_training\_2/2.8.128/32.1.keras  
/data/results-archive/imbalanced\_training\_2/2.8.128/32.8.keras  
/data/results-archive/imbalanced\_training\_2/2.8.256/16.1.keras  
/data/results-archive/imbalanced\_training\_2/2.8.256/16.8.keras  
/data/results-archive/imbalanced\_training\_2/2.8.256/32.1.keras  
/data/results-archive/imbalanced\_training\_2/2.8.256/32.8.keras  
/data/results-archive/imbalanced\_training\_2/2.8.64/16.1.keras  
/data/results-archive/imbalanced\_training\_2/2.8.64/16.8.keras  
/data/results-archive/imbalanced\_training\_2/2.8.64/32.1.keras  
/data/results-archive/imbalanced\_training\_2/2.8.64/32.8.keras  
/data/results-archive/results/2.4.64/16.1.keras  
/data/results-archive/results/2.4.64/16.8.keras  
/data/results-archive/results/2.4.64/32.1.keras  
/data/results-archive/results - Copy/2.4.64/32.1.keras  
/data/results-archive/results - Copy/4.4.64/32.4.keras  
/data/results-archive/training\_data\_broken/2.12.128/16.1.keras  
/data/results-archive/training\_data\_broken/2.12.128/16.32.keras  
/data/results-archive/training\_data\_broken/2.12.128/16.8.keras  
/data/results-archive/training\_data\_broken/2.12.128/32.1.keras  
/data/results-archive/training\_data\_broken/2.12.128/32.32.keras  
/data/results-archive/training\_data\_broken/2.12.128/32.8.keras  
/data/results-archive/training\_data\_broken/2.12.128/48.1.keras  
/data/results-archive/training\_data\_broken/2.12.128/48.32.keras  
/data/results-archive/training\_data\_broken/2.12.128/48.8.keras  
/data/results-archive/training\_data\_broken/2.12.256/16.1.keras  
/data/results-archive/training\_data\_broken/2.12.256/16.32.keras

/data/results-archive/training\_data\_broken/2.12.256/16.8.keras  
/data/results-archive/training\_data\_broken/2.12.256/32.1.keras  
/data/results-archive/training\_data\_broken/2.12.256/32.32.keras  
/data/results-archive/training\_data\_broken/2.12.256/32.8.keras  
/data/results-archive/training\_data\_broken/2.12.256/48.1.keras  
/data/results-archive/training\_data\_broken/2.12.256/48.32.keras  
/data/results-archive/training\_data\_broken/2.12.256/48.8.keras  
/data/results-archive/training\_data\_broken/2.12.64/16.1.keras  
/data/results-archive/training\_data\_broken/2.12.64/16.32.keras  
/data/results-archive/training\_data\_broken/2.12.64/16.8.keras  
/data/results-archive/training\_data\_broken/2.12.64/32.1.keras  
/data/results-archive/training\_data\_broken/2.12.64/32.32.keras  
/data/results-archive/training\_data\_broken/2.12.64/32.8.keras  
/data/results-archive/training\_data\_broken/2.12.64/48.1.keras  
/data/results-archive/training\_data\_broken/2.12.64/48.32.keras  
/data/results-archive/training\_data\_broken/2.12.64/48.8.keras  
/data/results-archive/training\_data\_broken/2.4.256/16.1.keras  
/data/results-archive/training\_data\_broken/2.4.256/16.32.keras  
/data/results-archive/training\_data\_broken/2.4.256/16.8.keras  
/data/results-archive/training\_data\_broken/2.4.256/32.1.keras  
/data/results-archive/training\_data\_broken/2.4.256/32.32.keras  
/data/results-archive/training\_data\_broken/2.4.256/32.8.keras  
/data/results-archive/training\_data\_broken/2.4.256/48.1.keras  
/data/results-archive/training\_data\_broken/2.4.256/48.32.keras  
/data/results-archive/training\_data\_broken/2.4.256/48.8.keras  
/data/results-archive/training\_data\_broken/2.8.128/16.1.keras  
/data/results-archive/training\_data\_broken/2.8.128/16.32.keras  
/data/results-archive/training\_data\_broken/2.8.128/16.8.keras  
/data/results-archive/training\_data\_broken/2.8.128/32.1.keras  
/data/results-archive/training\_data\_broken/2.8.128/32.32.keras  
/data/results-archive/training\_data\_broken/2.8.128/32.8.keras  
/data/results-archive/training\_data\_broken/2.8.128/48.1.keras  
/data/results-archive/training\_data\_broken/2.8.128/48.32.keras  
/data/results-archive/training\_data\_broken/2.8.128/48.8.keras  
/data/results-archive/training\_data\_broken/2.8.256/16.1.keras  
/data/results-archive/training\_data\_broken/2.8.256/16.32.keras  
/data/results-archive/training\_data\_broken/2.8.256/16.8.keras  
/data/results-archive/training\_data\_broken/2.8.256/32.1.keras  
/data/results-archive/training\_data\_broken/2.8.256/32.32.keras  
/data/results-archive/training\_data\_broken/2.8.256/32.8.keras

/data/results-archive/training\_data\_broken/2.8.256/48.1.keras  
/data/results-archive/training\_data\_broken/2.8.256/48.32.keras  
/data/results-archive/training\_data\_broken/2.8.256/48.8.keras  
/data/results-archive/training\_data\_broken/2.8.64/16.1.keras  
/data/results-archive/training\_data\_broken/2.8.64/16.32.keras  
/data/results-archive/training\_data\_broken/2.8.64/16.8.keras  
/data/results-archive/training\_data\_broken/2.8.64/32.1.keras  
/data/results-archive/training\_data\_broken/2.8.64/32.32.keras  
/data/results-archive/training\_data\_broken/2.8.64/32.8.keras  
/data/results-archive/training\_data\_broken/2.8.64/48.1.keras  
/data/results-archive/training\_data\_broken/2.8.64/48.32.keras  
/data/results-archive/training\_data\_broken/2.8.64/48.8.keras  
/data/results-archive/training\_data\_broken/3.12.128/16.1.keras  
/data/results-archive/training\_data\_broken/3.12.128/16.32.keras  
/data/results-archive/training\_data\_broken/3.12.128/16.8.keras  
/data/results-archive/training\_data\_broken/3.12.128/32.1.keras  
/data/results-archive/training\_data\_broken/3.12.128/32.32.keras  
/data/results-archive/training\_data\_broken/3.12.128/32.8.keras  
/data/results-archive/training\_data\_broken/3.12.128/48.1.keras  
/data/results-archive/training\_data\_broken/3.12.128/48.32.keras  
/data/results-archive/training\_data\_broken/3.12.128/48.8.keras  
/data/results-archive/training\_data\_broken/3.12.256/16.1.keras  
/data/results-archive/training\_data\_broken/3.12.256/16.32.keras  
/data/results-archive/training\_data\_broken/3.12.256/16.8.keras  
/data/results-archive/training\_data\_broken/3.12.256/32.1.keras  
/data/results-archive/training\_data\_broken/3.12.256/32.32.keras  
/data/results-archive/training\_data\_broken/3.12.256/32.8.keras  
/data/results-archive/training\_data\_broken/3.12.256/48.1.keras  
/data/results-archive/training\_data\_broken/3.12.256/48.32.keras  
/data/results-archive/training\_data\_broken/3.12.256/48.8.keras  
/data/results-archive/training\_data\_broken/3.12.64/16.1.keras  
/data/results-archive/training\_data\_broken/3.12.64/16.32.keras  
/data/results-archive/training\_data\_broken/3.12.64/16.8.keras  
/data/results-archive/training\_data\_broken/3.12.64/32.1.keras  
/data/results-archive/training\_data\_broken/3.12.64/32.32.keras  
/data/results-archive/training\_data\_broken/3.12.64/32.8.keras  
/data/results-archive/training\_data\_broken/3.12.64/48.1.keras  
/data/results-archive/training\_data\_broken/3.12.64/48.32.keras  
/data/results-archive/training\_data\_broken/3.12.64/48.8.keras  
/data/results-archive/training\_data\_broken/3.4.128/16.1.keras

/data/results-archive/training\_data\_broken/3.4.128/16.32.keras  
/data/results-archive/training\_data\_broken/3.4.128/16.8.keras  
/data/results-archive/training\_data\_broken/3.4.128/32.1.keras  
/data/results-archive/training\_data\_broken/3.4.128/32.32.keras  
/data/results-archive/training\_data\_broken/3.4.128/32.8.keras  
/data/results-archive/training\_data\_broken/3.4.128/48.1.keras  
/data/results-archive/training\_data\_broken/3.4.128/48.32.keras  
/data/results-archive/training\_data\_broken/3.4.128/48.8.keras  
/data/results-archive/training\_data\_broken/3.4.256/16.1.keras  
/data/results-archive/training\_data\_broken/3.4.256/16.32.keras  
/data/results-archive/training\_data\_broken/3.4.256/16.8.keras  
/data/results-archive/training\_data\_broken/3.4.256/32.1.keras  
/data/results-archive/training\_data\_broken/3.4.256/32.32.keras  
/data/results-archive/training\_data\_broken/3.4.256/32.8.keras  
/data/results-archive/training\_data\_broken/3.4.256/48.1.keras  
/data/results-archive/training\_data\_broken/3.4.256/48.32.keras  
/data/results-archive/training\_data\_broken/3.4.256/48.8.keras  
/data/results-archive/training\_data\_broken/3.4.64/16.1.keras  
/data/results-archive/training\_data\_broken/3.4.64/16.32.keras  
/data/results-archive/training\_data\_broken/3.4.64/16.8.keras  
/data/results-archive/training\_data\_broken/3.4.64/32.1.keras  
/data/results-archive/training\_data\_broken/3.4.64/32.32.keras  
/data/results-archive/training\_data\_broken/3.4.64/32.8.keras  
/data/results-archive/training\_data\_broken/3.4.64/48.1.keras  
/data/results-archive/training\_data\_broken/3.4.64/48.32.keras  
/data/results-archive/training\_data\_broken/3.4.64/48.8.keras  
/data/results-archive/training\_data\_broken/3.8.128/16.1.keras  
/data/results-archive/training\_data\_broken/3.8.128/16.32.keras  
/data/results-archive/training\_data\_broken/3.8.128/16.8.keras  
/data/results-archive/training\_data\_broken/3.8.128/32.1.keras  
/data/results-archive/training\_data\_broken/3.8.128/32.32.keras  
/data/results-archive/training\_data\_broken/3.8.128/32.8.keras  
/data/results-archive/training\_data\_broken/3.8.128/48.1.keras  
/data/results-archive/training\_data\_broken/3.8.128/48.32.keras  
/data/results-archive/training\_data\_broken/3.8.128/48.8.keras  
/data/results-archive/training\_data\_broken/3.8.256/16.1.keras  
/data/results-archive/training\_data\_broken/3.8.256/16.32.keras  
/data/results-archive/training\_data\_broken/3.8.256/16.8.keras  
/data/results-archive/training\_data\_broken/3.8.256/32.1.keras  
/data/results-archive/training\_data\_broken/3.8.256/32.32.keras

/data/results-archive/training\_data\_broken/3.8.256/32.8.keras  
/data/results-archive/training\_data\_broken/3.8.256/48.1.keras  
/data/results-archive/training\_data\_broken/3.8.256/48.32.keras  
/data/results-archive/training\_data\_broken/3.8.256/48.8.keras  
/data/results-archive/training\_data\_broken/3.8.64/16.1.keras  
/data/results-archive/training\_data\_broken/3.8.64/16.32.keras  
/data/results-archive/training\_data\_broken/3.8.64/16.8.keras  
/data/results-archive/training\_data\_broken/3.8.64/32.1.keras  
/data/results-archive/training\_data\_broken/3.8.64/32.32.keras  
/data/results-archive/training\_data\_broken/3.8.64/32.8.keras  
/data/results-archive/training\_data\_broken/3.8.64/48.1.keras  
/data/results-archive/training\_data\_broken/3.8.64/48.32.keras  
/data/results-archive/training\_data\_broken/3.8.64/48.8.keras  
/data/results-combined/2.4.16/16.1.keras  
/data/results-combined/2.4.16/16.32.keras  
/data/results-combined/2.4.16/16.8.keras  
/data/results-test/2.4.128/16.1.keras  
/data/results-test/2.4.128/16.32.keras  
/data/results-test/2.4.128/16.8.keras  
/data/results-test/2.4.128/32.1.keras  
/data/results-test/2.4.128/32.32.keras  
/data/results-test/2.4.128/32.8.keras  
/data/results-test/2.4.128/48.1.keras  
/data/results-test/2.4.128/48.8.keras  
/data/results-test/2.4.64/16.1.keras  
/data/results-test/2.4.64/16.32.keras  
/data/results-test/2.4.64/16.8.keras  
/data/results-test/2.4.64/32.1.keras  
/data/results-test/2.4.64/32.32.keras  
/data/results-test/2.4.64/32.8.keras  
/data/results-test/2.4.64/48.1.keras  
/data/results-test/2.4.64/48.32.keras  
/data/results-test/2.4.64/48.8.keras

## D Presentation, Poster, and Ethics Form



# Real-time preictal detection through the application of machine learning to Electroencephalogram signals.

## Final Major Project

William Riddell - 19066041

# Contents

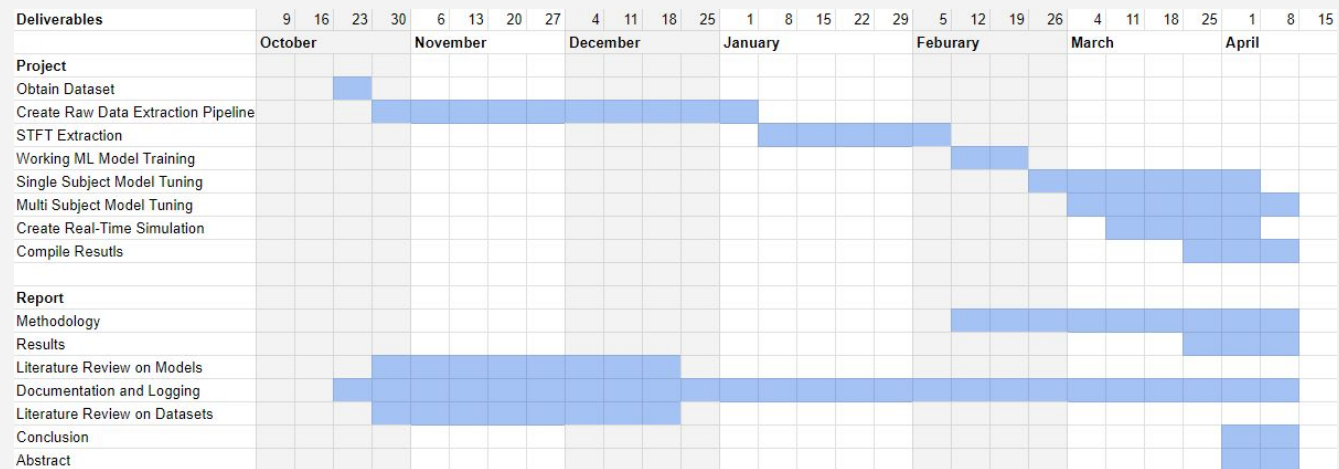
- Epilepsy
- Order of work
- Solution
  - Dataset
  - Preprocessing
  - Convolutional  
Neural Network
  - Post Processing
  - System Results
- Conclusion

# Epilepsy

- “Epilepsy is one of the most common serious brain conditions, affecting over 70 million people worldwide.” - Thijs, R.D., Surges, R., O'Brien, T.J. and Sander, J.W., 2019. Epilepsy in adults. *The Lancet*, 393(10172), pp.689-701.
- “One third of people with epilepsy will not have adequate seizure control with the current medications [available Antiepileptic Drugs (AEDs)]. For these patients the situation has improved very little in the last few decades” - Galanopoulou, A.S., Buckmaster, P.S., Staley, K.J., Moshé, S.L., Perucca, E., Engel Jr, J., Löscher, W., Noebels, J.L., Pitkänen, A., Stables, J. and White, H.S., 2012. Identification of new epilepsy treatments: issues in preclinical methodology. *Epilepsia*, 53(3), pp.571-582.
- “Of those treated with AEDs, 35% become resistant to medication” - Moghim, N. and Corne, D.W., 2014. Predicting epileptic seizures in advance. *PloS one*, 9(6), p.e99334.
- “There is an urgent need for more effective and better-tolerated treatments to control drug-resistant seizures, as well as for innovative therapies to prevent, stop, or reverse the development of epilepsy and epilepsy-related comorbidities” - Löscher, W. and Schmidt, D., 2011. Modern antiepileptic drug development has failed to deliver: ways out of the current dilemma. *Epilepsia*, 52(4), pp.657-678.

# Order of Work

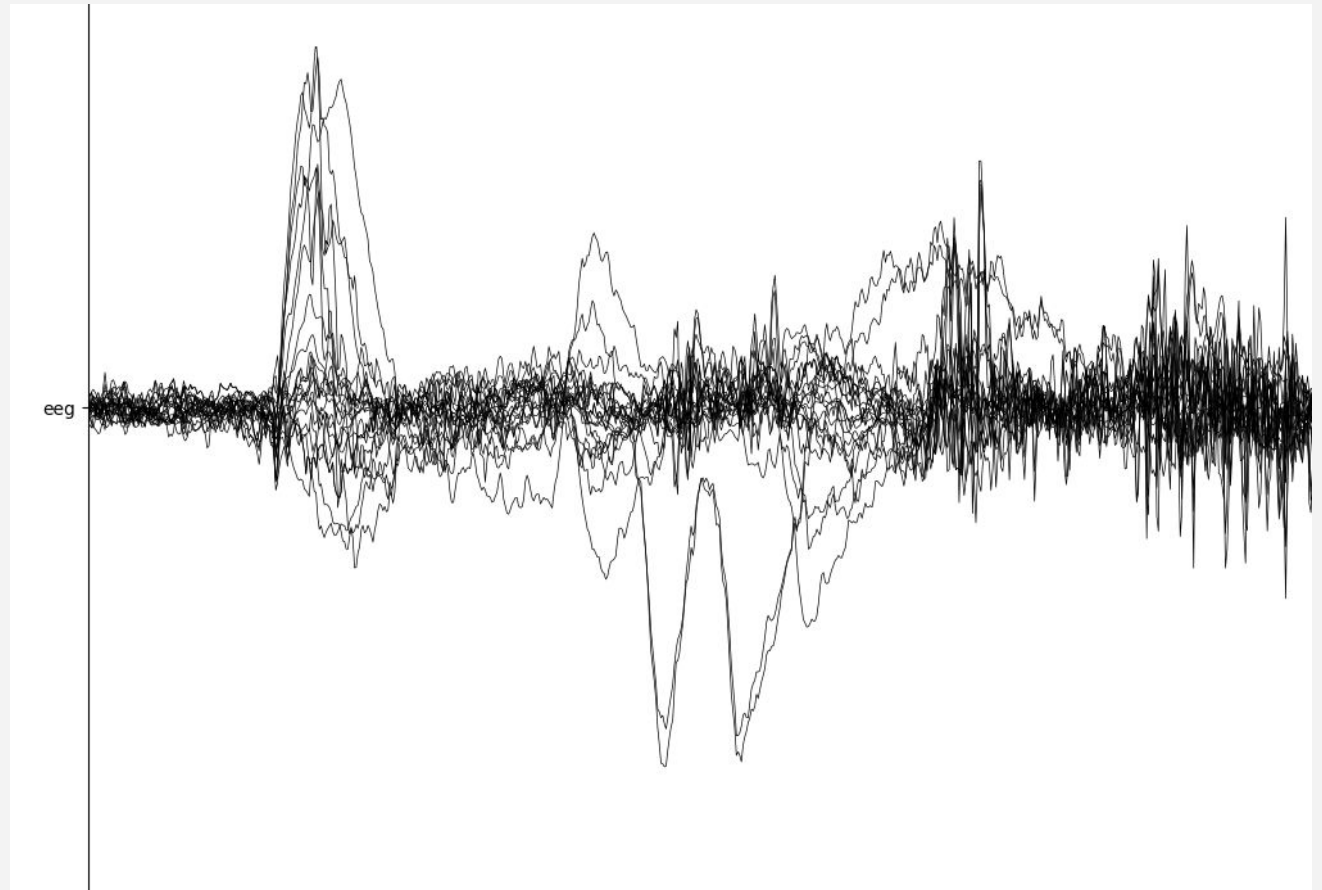
- Emphasis on literature review findings instead of personal experiments.
- Increased the time allocated for hyper-parameter tuning due to the training times for a single model.
- Added time to allow for further developments, such as subject agnostic model training or node configuration tuning.



# The Solution

# CHB-MIT EEG

- 23 unique subjects.
- 10-20 International System for Scalp EEG.
- 17 common channels between all subjects.
- Continuous recordings with labeled seizure times.
- Open source; hosted through Google Cloud.

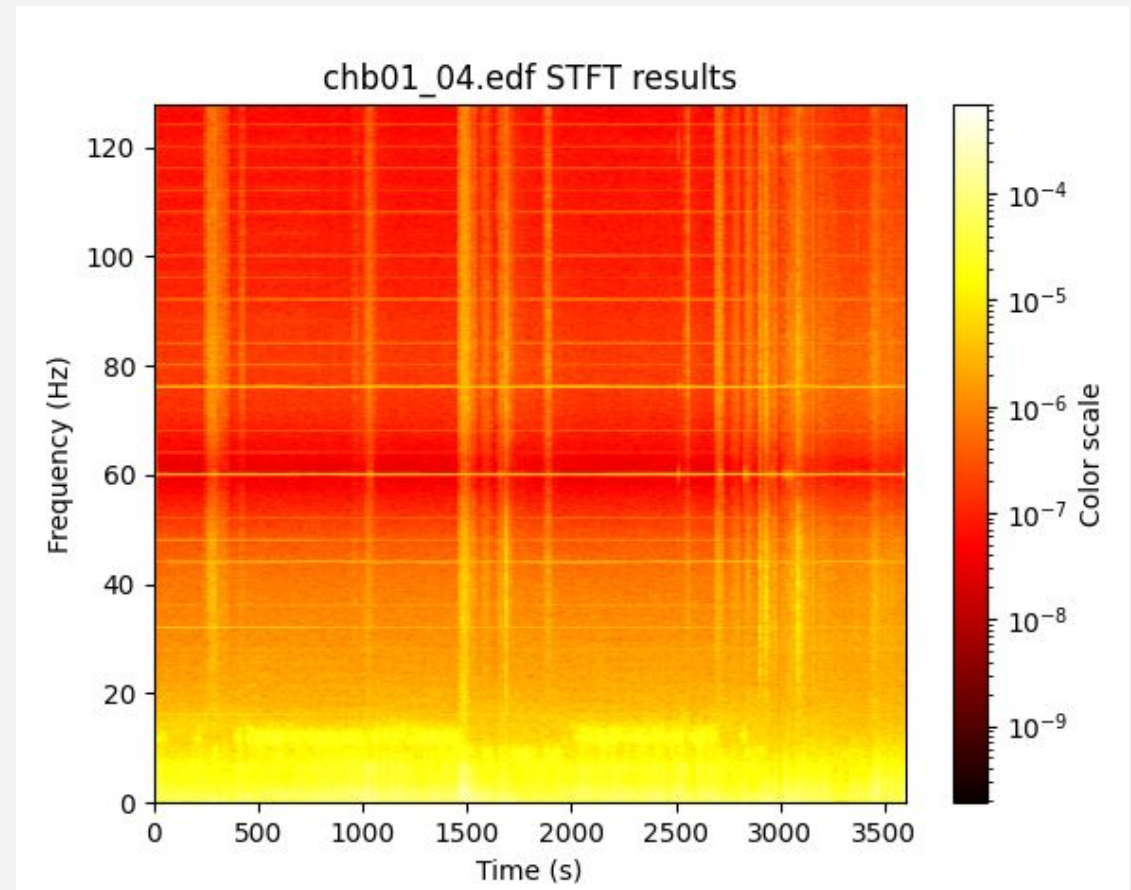


A Butterfly representation of subject 1's EEG recording.  
[https://github.com/N3utra1/COMP6013\\_Dissertation.git](https://github.com/N3utra1/COMP6013_Dissertation.git)

# Preprocessing

- Short-time Fourier transform.
- Band Pass Filter.
- Class balancing.
- Approach influenced by “Convolutional neural networks for seizure prediction and scalp electroencephalogram”.

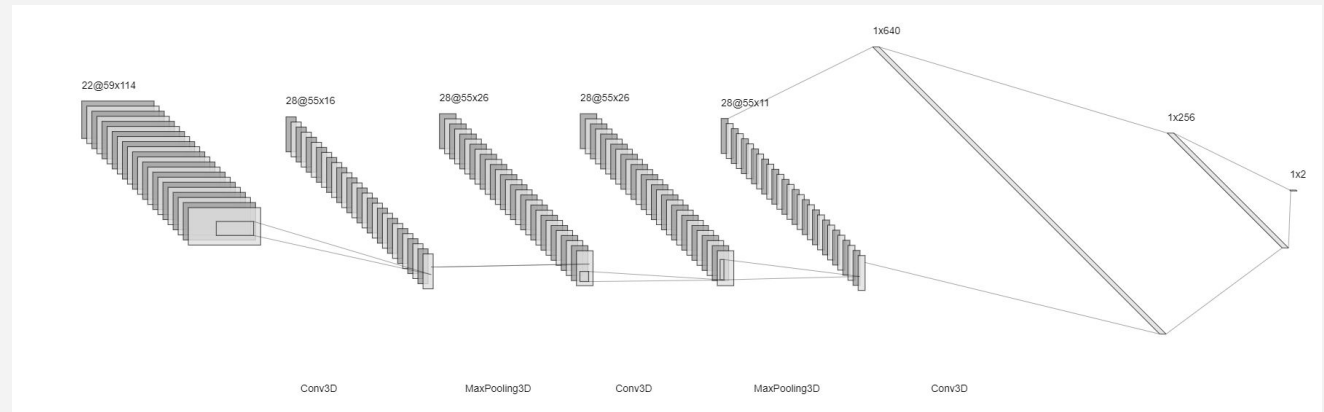
Truong, N.D., Nguyen, A.D., Kuhlmann, L., Bonyadi, M.R., Yang, J., Ippolito, S. and Kavehei, O., 2018. Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram. *Neural Networks*, 105, pp.104-111.



Heat map of a Short-time Fourier transform result. Seizure between 1467-1494 seconds..  
[https://github.com/N3utra1/COMP6013\\_Dissertation.git](https://github.com/N3utra1/COMP6013_Dissertation.git)

# The Convolution Neural Network

- The CNN is created using Python 3 along with the TensorFlow and Keras.
- The solution is an image recognition approach to the problem.
- CNN model architecture has been tuned, along with the hyper-parameters.

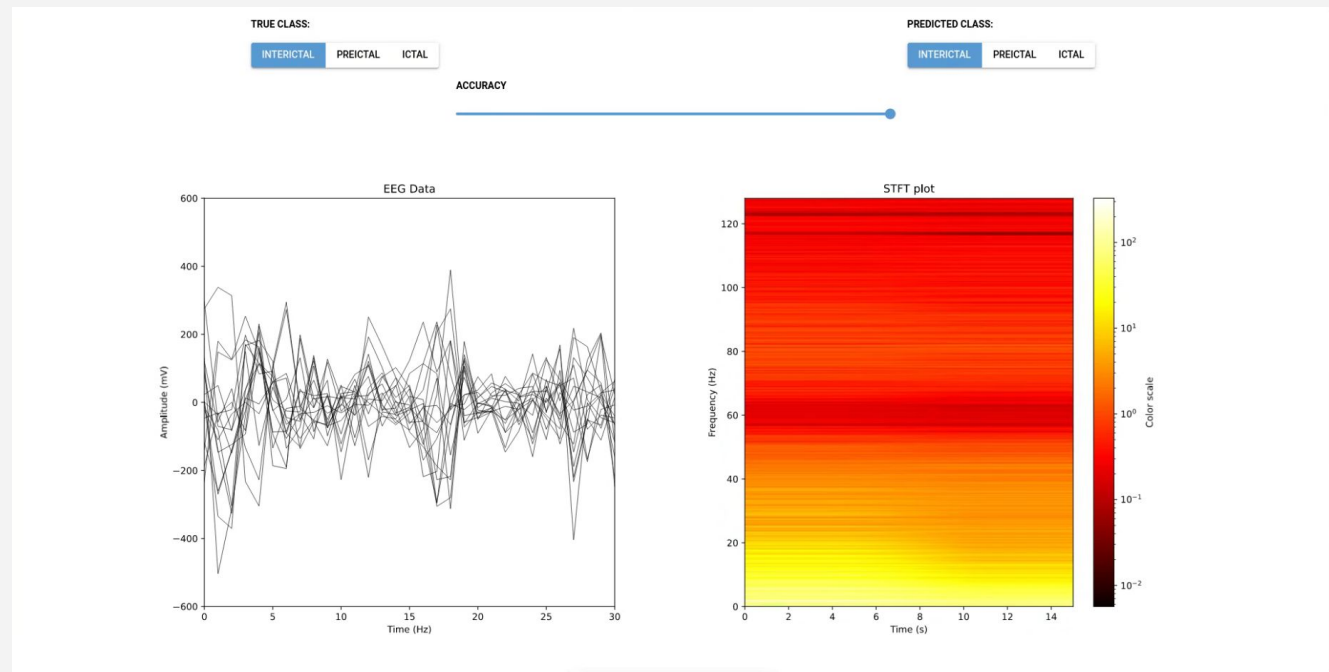


Representation of an model architecture. Created using <https://alexlenail.me/NN-SVG/LeNet.html>



# Real Time Simulation

- Shows the preprocessing pipeline's ability to run in real-time.
- User selects an class in the top left and the model shows it's prediction in the top right.
- The plot on the left shows the last 30 seconds of raw EEG data. 256 new data points for each EEG node arrive each second.
- The right plot is the STFT spectrogram image being inputted into the ML model.



# Model Results

- Models trained against a single subject's data, allowing the CNN to pick up on their characteristics.
- The models were tested against 20% of the dataset (unseen).
- 14 models were perfect models with 100% accuracy against the test dataset.
- This shows the potential this approach has for real-time preictal prediction.

Model	Training Accuracy	Training Loss
2.4.128/16.1	0.9832265377044678	0.8224268555641174
2.4.128/16.32	0.9993602503091097	0.02814496521023102
2.4.128/16.8	0.9976428672671318	0.11281578045975693
2.4.128/32.1	0.975464940071106	1.4378010034561157
2.4.128/32.32	0.9989679381251335	0.04745509265126133
2.4.128/32.8	0.9962754622101784	0.18717686411091886
2.4.128/48.1	0.9669219255447388	2.318654775619507
2.4.128/48.8	0.9958652406930923	0.2898520281258925
2.4.64/16.1	0.9799447655677795	0.6853911876678467
2.4.64/16.32	0.9992593247443438	0.024666143968849415
2.4.64/16.8	0.9969331100583076	0.08956567873635081
2.4.64/32.1	0.9619211554527283	1.725386619567871
2.4.64/32.32	0.9987660832703114	0.05414362490514908
2.4.64/32.8	0.9953247904777527	0.21542102738881397
2.4.64/48.1	0.9655154347419739	2.379462718963623
2.4.64/48.32	0.9989125896245241	0.07438817310655671
2.4.64/48.8	0.994374118745327	0.31704871635884047

Table 24: Accuracy and Loss metrics (average across epochs) during model training.

Results table taken from report.

Model key: [convolution block count].[dense layer count].[dense layer size]/[batch size].[epochs]

# Wrapping up

# Future Advancements

- Tuning STFT parameters.
- Further tuning of model architecture.
- Testing different EEG node configurations.
- “From a large number of recent studies, it is clear that BCI is an up-and-coming research area.” - Zabcikova, M., Koudelkova, Z., Jasek, R. and Lorenzo Navarro, J.J., 2022. Recent advances and current trends in brain-computer interface research and their applications. International Journal of Developmental Neuroscience, 82(2), pp.107-123.

# Real-time preictal detection through the application of machine learning to Electroencephalogram signals.

William Riddell Final year project.  
Supervisor: Kashinath Basu

## Introduction

"A seizure is a burst of uncontrolled electrical activity... that causes temporary abnormalities in muscle tone or movements, behaviors, sensations or states of awareness." (Medicine n.d.). "Persistent seizures constitute a considerable burden on healthcare resources." (Assi, Nguyen, Rihana & Sawan 2017). Therefore "An automated accurate prediction of seizures will significantly improve the quality of life of patients and reduce the burden on caregivers" (Acharya, Hagiwara & Adeli 2018). As uncontrolled electrical activity causes seizures, a method has been devised that applies image recognition through the use of machine learning to images generated from EEG data. This method can be run in real-time, outputting a prediction each second, with the ability to predict seizures up to 20 minutes before onset.

## Literature

10 datasets, 6 models, and 10 preprocessing pipelines were evaluated. Results varied with a preprocessing pipeline using "Distribution of Wavelet Coherence" along with a CNN (Mirowski et al. 2009) obtaining some of the highest results.

Another approach was using a pre-processing pipeline with Short Time Fourier Transforms (Truong, Nguyen, Kuhlmann, Bonyadi, Yang, Ippolito & Kavehei 2018). This achieved respectable results of 81.4% sensitivity and 0.06 false positives/hour.

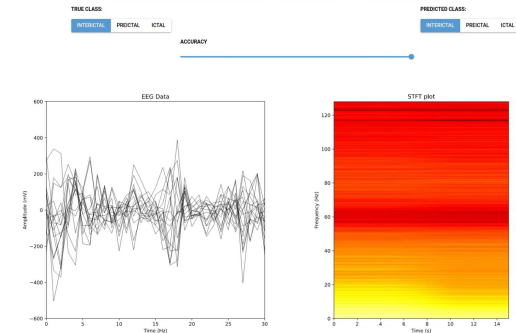
## Ethical Issues

As this project is only training ML models ethical issues are very small, however, if this is embedded in a seizure prediction system the false positive rate will have to be very low; if the system alerts the user and is incorrect, it will cause a large amount of unneeded stress for the end user as they prepare for a seizure which has been incorrectly predicted.

## Methodology

This project used a Short-Time Fourier-Transform which creates an image from the electrical signals in the brain, recorded with an EEG. Interference is then removed by using a notch filter. For training the machine learning models, data was synthesized by using a sliding window method to balance the dataset, however, for prediction, the images can be directly fed into a model which classifies the results.

The project was programmed in Python 3 and used TensorFlow and MNE as the main libraries.



Real Time Simulation. EEG data left, Spectrogram (STFT) image right. Model's predicted class top right.

## Results

This project has trained over 15 models which obtained 100% accuracy during testing. These models are trained against an individual, allowing the neural network to pick up on their specific characteristics. Any of these models can then be used in conjunction with the real-time simulation which has also been developed.

The real-time simulation displays the model's ability to classify the current state of the subject each second, alerting the subject if a seizure is imminent. This shows the potential this project has to greatly alleviate stress for people suffering from seizures.

## Conclusions

This project was able to achieve its objectives, developing a model with an accuracy >90%, along with a preprocessing pipeline that can process the raw data in under a second. The research and development done allows for many future advancements;

- Tuning STFT window parameters directly changes the properties of the model's input images.
- Implementing online learning, such that the model is trained against a subject at all times, further reduces false positives and inaccurate predictions.
- EEG Node configurations should also be tested, possibly showing that the same results can be obtained by using fewer EEG nodes.

With hindsight and more time, other preprocessing pipelines and ML model types would have been integrated into the solution. As the code is open source, structuring the code such that anyone could integrate the future advancements discussed above would allow for many different variable configurations to be tested against many different model types and preprocessing approaches.

## References

<https://drive.google.com/file/d/1VGSPdRVIAAZp02pls3qw4FANiYk2peBG/view?usp=sharing>

[https://github.com/N3utra1/COMP6013\\_Dissertation](https://github.com/N3utra1/COMP6013_Dissertation)

## Faculty of Technology, Design and Environment - Ethics Review Form E1

- This form should be completed jointly by the **Supervisor and Student** who is undertaking a research/major project which involves human participants.
- It is the **Supervisor** who is responsible for exercising appropriate professional judgement in this review.
- Before completing this form, please refer to the University Code of Practice for the Ethical Standards for Research involving Human Participants, available at <http://www.brookes.ac.uk/Research/Research-ethics/> and to any guidelines provided by relevant academic or professional associations.
- Note that the ethics review process needs to fully completed and signed **before fieldwork commences**.

(i) **Project Title:**

**Real-time preictal detection through the application of machine learning to Electroencephalogram signals.**

(ii) **Name of Supervisor and School in which located:**

Kashinath Basu - Faculty of Technology, Design and Environment School of Engineering, Computing and Mathematics

(iii) **Name of Student and Student Number:**

William Riddell - 19066041

(iv) **Brief description of project outlining where human participants will be involved (30-50 words):**

No Human participants will be involved in the process of the project. An pre-recorded dataset will be downloaded containing EEG data against unnamed subjects. An AI will be trained off extracted features from this dataset.

		Yes	No
1.	Does the study involve participants who are unable to give informed consent (e.g. children, people with learning disabilities)?		X
2.	If the study will involve participants who are unable to give informed consent (e.g. children under the age of 18, people with learning disabilities), will you be unable to obtain permission from their parents or guardians (as appropriate)?		X
3.	Will the study require the cooperation of a gatekeeper for initial access to groups or individuals to be recruited (e.g. students, members of a self-help group, employees of a company)?		X
4.	Are there any problems with the participants' right to remain anonymous, or to have the information they give not identifiable as theirs?		X

5.	Will it be necessary for the participants to take part in the study without their knowledge/consent at the time? (e.g. covert observation of people in non-public places?)		x
6.	Will the study involve discussion of or responses to questions the participants might find sensitive? (e.g. own traumatic experiences)		x
7.	Are drugs, placebos or other substances (e.g. food substances, vitamins) to be administered to the study participants?		x
8.	Will blood or tissue samples be obtained from participants?		x
9.	Is pain or more than mild discomfort likely to result from the study?		x
10.	Could the study induce psychological stress or anxiety?		x
11.	Will the study involve prolonged or repetitive testing of participants?		x
12.	Will financial inducements (other than reasonable expenses and compensation for time) be offered to participants?		x
13.	Will deception of participants be necessary during the study?		x
14.	Will the study involve NHS patients, staff, carers or premises?		x

<b>Signed :</b>	K. Basu	<b>Supervisor</b>
<b>Signed :</b>	William Riddell	<b>Student</b>
<b>Date:</b>	17/04/24	

#### What to do now:

1. If you have answered '**no**' to all the above questions:

- The student must **submit** the completed and fully signed E1 form to their **Dissertation Module Leader via Moodle**.
- The student must keep a copy of the E1 form which must be bound into their dissertation as an appendix.
- The supervisor must keep a copy of the E1 form as they are responsible for monitoring compliance during the fieldwork.

1. If you have answered '**yes**' to **any** of the above questions:

- The supervisor and student must complete the TDE E2 form available at <http://www.brookes.ac.uk/Research/Research-ethics/Ethics-review-forms/>
- Note that the information in the E2 must be in **sufficient detail** for the ethical implications to be clearly identified.
- The signed E2 and signed E1 Form must be emailed to Bridget Durning ([bdurning@brookes.ac.uk](mailto:bdurning@brookes.ac.uk)) who is the Faculty Research Ethics Officer (FREO) for review. Please allow **at least two weeks** for this review process.
- If/when approved the FREO will issue an E3 Ethics Approval Notice.
- The student must send the E1, E2 and E3 Notice **to the Dissertation Module Leader**.
- The student must also keep copies which must be bound into their dissertation as an appendix.
- The supervisor must keep a copy of documentation to monitor compliance during field work.

1. If you answered 'yes' to any of questions 1-13 and 'yes' to question 14, an application must be submitted to the appropriate NHS research ethics committee. This is an onerous and time consuming process so the supervisor should liaise early with the FREO if the student is considering this.