SELF SUPERVISED LEARNING OF EEG (ELECTROENCEPHALOGRAM) RAW DATA TO LEARN THE HIDDEN PATTERNS OF HUMAN BRAIN ACTIVITIES.

Thambawita Maddumage Tharindu Akalanka Gunarathna

209327H

Master of Science in Computer Science Specialising in Data Science Engineering and Analytics

Department of Computer Science & Engineering

Faculty of Engineering

University of Moratuwa

Sri Lanka

October 2022

SELF SUPERVISED LEARNING OF EEG (ELECTROENCEPHALOGRAM) RAW DATA TO LEARN THE HIDDEN PATTERNS OF HUMAN BRAIN ACTIVITIES.

Thambawita Maddumage Tharindu Akalanka Gunarathna
209327H

Thesis/Dissertation submitted in partial fulfillment of the requirements for the degree
Master of Science in Computer Science Specialising in Data Science Engineering and Analytics

Department of Computer Science & Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

October 2022

DECLARATION

I declare that this is my own work and this thesis/dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Candidate Name: Gunarathna T. M. T. A.

……………………………………

Signature of Candidate Date:

The above candidate has carried out research for the PhD/MPhil/Masters thesis/dissertation under my supervision. I confirm that the declaration made above by the student is true and correct.

Supervisor Name: Dr. Thanuja D. Ambegoda

……………………………………

Signature of Supervisor Date:

ACKNOWLEDGMENTS

First and foremost, I am extremely grateful to my supervisor, Dr. Thanuja D. Ambegoda for his invaluable advice, continuous support, and patience during my MSc academic research study. His immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life. I would also like to thank Dr. Sapumal Ahangama for introducing me to my supervisor. I would like to thank all the members in the Computer Science and Engineering Department. It is their kind help and support that have made my study and life in the University of Moratuwa a wonderful time. Finally, I would like to express my gratitude to my parents, my sisters and their families and my friends. Without their tremendous understanding and encouragement in the past few years, it would be impossible for me to complete my study.

Abstract

EEG is a non-invasive neuroimaging modality that operates by measuring changes in electrical voltage on the scalp that are induced by cortical activity. In this research, we propose a method for self-supervised learning of EEG raw data to learn the hidden patterns of human brain activities. This work was performed through a pipeline consisting of five phases. Each of the phase's output will be the input for the next phase. Phase 1 is for pre-processing raw EEG sequences into EEG representations that catch the spacial and temporal properties in the original raw EEG sequences. We have followed a relatively less complex method to pre-process raw EEG sequences. In phase 2, pre-processed raw EEG sequences will be learnt by self-supervised representation learning. For that self-supervised vision transformers with DINO will be used. These vision transformers models are computationally more demanding and require more training data therefore more computational resources and training data will be needed. So that at the presence of more training data and computational processing power, self-supervised vision transformer architectures will be expected to produce the best results while outperforming supervised learning architectures. Then at the phase 3, sequences of prototypes for each raw EEG data sequence of the dataset will be generated. To evaluate the prototypes that are generated from raw EEG data, phase 4 and 5 have been used as the downstream task for the self-supervised learning task. For phase 4 and 5, we again used a transformer architecture, that is a BERT based model called RoBERTa to learn the synthetic language generated by phase 3 or to learn the context and the language of generated prototype sequences and by performing a multi class prototype sequence classification, prototype generation for each representation at specific time stamp of raw EEG data sequence can be evaluated. We believe that since the models are computationally demanding and require more training data, the latter explained pipeline of five phases should be improved with more training and performing hyperparameter tuning at a high computational resources and data rich environment.

Keywords: Electroencephalogram, Self-Supervised Learning, Vision Transformers, Natural Language Processing

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
EEG	Electroencephalogram
ECG	Electrocardiogram
EMG	Electromyogram
MRI	Magnetic Resonance Imaging
BCI	Brain Computer Interface
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short Term Memory
SSL	Self Supervised Learning
DINO	Self-distillation with no labels
BERT	Bidirectional Encoder Representations from Transformers