MULTIVARIATE TIME-SERIES ANALYSIS TECHNIQUES FOR DISCOVERING COMPUTATIONAL BIOMARKERS FOR EPILEPSY IN EEG

SIDDHARTH PANWAR

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY DELHI

SEPTEMBER 2021

MULTIVARIATE TIME-SERIES ANALYSIS TECHNIQUES FOR DISCOVERING COMPUTATIONAL BIOMARKERS FOR EPILEPSY IN EEG

by

SIDDHARTH PANWAR

Department of Electrical Engineering

Submitted

in fulfilment of the requirements of the degree of Doctor of Philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

SEPTEMBER 2021

Dedicated to

People suffering from epilepsy

Certificate

This is to certify that the thesis entitled "**Multivariate time-series analysis techniques for discovering computational biomarkers for epilepsy in EEG**" being submitted by **Mr. Siddharth Panwar** to the Department of Electrical Engineering, Indian Institute of Technology Delhi, for the award of the degree of Doctor of Philosophy is the record of the bona-fide research work carried out by him under my supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree. The results contained in this thesis have not been submitted either in part or in full to any other university or institute for the award of any degree or diploma.

(Dr. Shiv Dutt Joshi)
Professor
Department of Electrical Engineering
Indian Institute of Technology Delhi

(Dr. Anubha Gupta)
Professor
Department of Electronics and Communication Engineering
Indraprastha Institute of Information Technology Delhi

(Dr. Puneet Agarwal)
Director and Head
Unit Neurology
Max Super Speciality Hospital, Saket

Acknowledgements

I was raised on the campus of a technical university and was naturally surrounded by researchers throughout my life. I had always noticed that for most doctoral students there was at least one point in their study where there was a real struggle. Some questioned their choice of pursuing the degree and some others were just simply desperate to be done with it. I consider myself fortunate enough to have enjoyed every single moment of my PhD journey. I have often told my friends that if I could afford it, I would do another PhD. I believe such sentiments are not easily found in graduating doctoral students and the credit for this goes directly to my supervisors. Prof. S.D. Joshi, Prof. Anubha Gupta and Dr. Puneet Agarwal together created an environment where I was free to explore without any inhibitions. My creativity was allowed to mature and flourish so I could become a confident independent researcher.

Being my primary advisor, Prof. Joshi deserves my special gratitude. I have not come across a truer academic in my life and probably never would. Before joining the PhD program at IIT Delhi I was a circuit designer and changing my area of research to signal processing was a substantial switch. To build a strong foundation for my research Prof. Joshi conducted personal lecture series for me where no question was too stupid and no idea was too simple. In fact, he told me it is in the simple ideas where the real kernel of knowledge resides. He gave me the courage to look at my research problem with a completely fresh set of eyes and not get influenced by what was done in the past. For someone who was a novice in the area of signal processing, this can be a daunting step but it was made supremely easy with his guidance. As I move ahead in life I take with myself the honor and selflessness with which he has taught countless students. I can only hope that I will be able to manifest a fraction of these traits in my professional life.

Siddharth Panwar

Abstract

Electroencephalography (EEG) is one of the oldest techniques used for studying the brain that continues to be used today because of the several advantages it offers. The EEG device is inexpensive, compact and specializes in capturing the functional activity of the brain. Although it was extremely prominent in the last century, its clinical use has largely been restricted to investigating epilepsy. As opposed to imaging modalities like CT scan and MRI, EEG more naturally lends itself to quantitative techniques because it dynamically captures brain function. However, it is in this aspect that EEG remains largely underutilized.

Techniques aimed at quantitative analysis of EEG have primarily explored three distinct specialization areas within the study of epilepsy, namely, epilepsy diagnosis, seizure prediction and seizure/abnormality detection. This thesis focuses on the first two of the three application areas by trying to identify novel computational biomarkers that help identify pathological features in EEG that are not visible to the naked eye. Although decades of research activity has preceded the work presented in this thesis, we aim to fill some of the gaps that have prevented the studies carried out in the past from translating into tools in the clinic. The most pressing issues were found in the case of epilepsy diagnosis where a single benchmark data set with several shortcomings has been in use for more than a decade for building predictive models.

In this thesis, we begin with the creation of a new data set that overcomes the shortcomings of the previous benchmark data set for epilepsy diagnosis and correcting the flaws in the model evaluation methodology of past studies. Characteristic response vector (CRV) is proposed that extends the notion of an impulse response of a linear system to positive definite matrices. This vector is used to reduce sample covariance matrices of EEG obtained through sliding window technique to single vectors. A novel computational biomarker is discovered in these vectors that helps in distinguishing epilepsy patients from healthy individuals. The predictive model is

evaluated on an external test data set, which, to the best of our knowledge, has not been done in previous studies.

Extending the sliding window technique employed in the predictive model described above, we propose recursive dynamic functional connectivity (rdFC) that builds a multi-scale hierarchical network for neurophysiological data. The rdFC technique is used to build 5-point connectivity patterns in a 3-dimensional Cartesian coordinate system. Using rdFC, a specific correlation structure is discovered that is present universally in a diverse range of human EEG records. This correlation structure is represented using three equivalent rdFC patterns. The dynamics of these patterns is found to be linked with that of seizures and used to build a seizure forecasting system for scalp EEG. The system is highly computationally efficient and satisfies all the guidelines widely accepted for evaluation of seizure prediction systems.

सार

इलेक्ट्रोएन्सेफलोग्राफी (ईईजी) मस्तिष्क का अध्ययन करने के लिए उपयोग की जाने वाली सबसे पुरानी तकनीकों में से एक है जिसका उपयोग आज भी जारी है क्योंकि इसके कई फायदे हैं। ईईजी डिवाइस सस्ता, कॉम्पैक्ट है और मस्तिष्क की कार्यात्मक गतिविधि को कैप्चर करने में माहिर है। यद्यपि यह पिछली शताब्दी में अत्यंत प्रमुख था, इसका नैदानिक उपयोग काफी हद तक मिर्गी की जांच तक ही सीमित रहा है। सीटी स्कैन और एमआरआई जैसे इमेजिंग तौर-तरीकों के विपरीत, ईईजी अधिक स्वाभाविक रूप से खुद को मात्रात्मक तकनीकों के लिए उधार देता है क्योंकि यह गतिशील रूप से मस्तिष्क के कार्य को पकड़ लेता है। हालाँकि, यह इस पहलू में है कि ईईजी का काफी हद तक कम उपयोग किया जाता है।

ईईजी के मात्रात्मक विश्लेषण के उद्देश्य से तकनीकों ने मुख्य रूप से मिर्गी के अध्ययन के भीतर तीन अलग-अलग विशेषज्ञता क्षेत्रों का पता लगाया है, अर्थात् मिर्गी निदान, जब्ती भविष्यवाणी और जब्ती / असामान्यता का पता लगाना। यह थीसिस उपन्यास कम्प्यूटेशनल बायोमार्कर की पहचान करने की कोशिश करके तीन अनुप्रयोग क्षेत्रों में से पहले दो पर केंद्रित है जो ईईजी में रोग संबंधी विशेषताओं की पहचान करने में मदद करते हैं जो नग्न आंखों को दिखाई नहीं देते हैं। यद्यपि इस थीसिस में प्रस्तुत किए गए कार्य से पहले दशकों की शोध गतिविधि हुई है, लेकिन हमारा लक्ष्य कुछ ऐसे अंतरालों को भरना है जिन्होंने अतीत में किए गए अध्ययनों को क्लिनिक में उपकरणों में अनुवाद करने से रोका है। मिर्गी के निदान के मामले में सबसे अधिक दबाव वाले मुद्दे पाए गए जहां कई कमियों के साथ एक एकल बेंचमार्क डेटा सेट एक दशक से अधिक समय से भविष्य कहनेवाला मॉडल बनाने के लिए उपयोग किया जा रहा है।

इस थीसिस में, हम एक नए डेटा सेट के निर्माण के साथ शुरू करते हैं जो मिर्गी के निदान के लिए पिछले बेंचमार्क डेटा सेट की कमियों को दूर करता है और पिछले अध्ययनों के मॉडल मूल्यांकन पद्धति में खामियों को ठीक करता है। विशेषता प्रतिक्रिया वेक्टर (सीआरवी) प्रस्तावित है जो एक रैखिक प्रणाली की आवेग प्रतिक्रिया की धारणा को सकारात्मक निश्चित मैट्रिक्स तक बढ़ाता है। इस वेक्टर का उपयोग स्लाइडिंग विंडो तकनीक के माध्यम से प्राप्त ईईजी के नमूना सहप्रसरण मैट्रिक्स को एकल वैक्टर में कम करने के लिए किया जाता है। इन वैक्टर में एक नया कम्प्यूटेशनल बायोमार्कर खोजा गया है जो मिर्गी के रोगियों को स्वस्थ व्यक्तियों से अलग करने में मदद करता है। भविष्य कहनेवाला मॉडल का मूल्यांकन बाहरी परीक्षण डेटा सेट पर किया जाता है, जो कि हमारे सर्वोत्तम ज्ञान के लिए, पिछले अध्ययनों में नहीं किया गया है।

ऊपर वर्णित भविष्य कहनेवाला मॉडल में नियोजित स्लाइडिंग विंडो तकनीक का विस्तार करते हुए, हम पुनरावर्ती गतिशील कार्यात्मक कनेक्टिविटी (rdFC) का प्रस्ताव करते हैं जो न्यूरोफिज़ियोलॉजिकल डेटा के लिए एक बहु-स्तरीय पदानुक्रमित नेटवर्क बनाता है। rdFC तकनीक का उपयोग 3-आयामी कार्टेशियन समन्वय प्रणाली में 5-बिंदु कनेक्टिविटी पैटर्न बनाने के लिए किया जाता है। rdFC का उपयोग करते हुए, एक विशिष्ट सहसंबंध संरचना की खोज की जाती है जो मानव ईईजी रिकॉर्ड की एक विविध श्रेणी में सार्वभौमिक रूप से मौजूद है। इस सहसंबंध संरचना को तीन समान rdFC पैटर्न का उपयोग करके दर्शाया गया है। इन पैटर्न की गतिशीलता को बरामदगी के साथ जोड़ा जाता है और खोपड़ी ईईजी के लिए एक जब्ती पूर्वानुमान प्रणाली का निर्माण करने के लिए उपयोग किया जाता है। प्रणाली अत्यधिक कम्प्यूटेशनल रूप से कुशल है और जब्ती भविष्यवाणी प्रणाली के मूल्यांकन के लिए व्यापक रूप से स्वीकृत सभी दिशानिर्देशों को पूरा करती है।

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CRA	Characteristic Response Analysis
CRV	Characteristic Response Vector
rdFC	Recursive Dynamic Functional Connectivity
SVM	Support Vector Machine
POCR	Positive Orthant Characteristic Response
BD	Bhattacharyya Distance
LBP	Local Binary Pattern