Certificate

It is certified that the work contained in the thesis titled **Computation of Some Aspects of Language Processing in the Brain Using Machine Learning** by **Ashish Ranjan** (Roll No.: 15191002) has been carried out under my/our supervision and that this work has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all the requirements of Comprehensive Examination, Candidacy, and SOTA for the award of the degree of **Doctor of Philosophy** to the Indian Institute of Technology (Banaras Hindu University), Varanasi

Signature of Supervisor
Anil Kumar Singh
Associate Professor
Department of Computer Science & Engineering
Indian Institute of Technology (BHU)
Varanasi - 221005, India

Signature of Co-Supervisor
Anil Kumar Thakur
Associate Professor
Department of Humanistic Studies
Indian Institute of Technology (BHU)
Varanasi - 221005, India

Declaration

I, **Ashish Ranjan**, certify that the work embodied in this thesis is my bonafide work and carried out by me under the supervision of **Anil Kumar Singh (Department of Computer Science and Engineering)** from **July 2015** to **January 2022**, at the **Department of Humanistic**, Indian Institute of Technology (BHU), Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully copied any other's work, paragraphs, text, data, results, *etc.*, reported in journals, books, magazines, reports dissertations, theses, *etc.*, or available at websites and have not included them in this thesis and have not cited as my work.

Date: 11/01/2022
Place: Varanasi

Ashish Ranjan
Signature of Student
(Ashish Ranjan)

Certificate by the Supervisor

It is certified that the above statement made by the student is correct to the best of my/our knowledge.

Signature of Supervisor
(Anil Kumar Singh)

Signature of Co-Supervisor
(Anil Kumar Thakur)

Forwarded
11/01/2022

Signature of Head of Department
(Dr. Ajit Kumar Mishra)

विभागाध्यक्ष/Head
मानवतावादी अध्ययन विभाग/Department of Humanistic Studies
भारतीय प्रौद्योगिकी संस्थान/Indian Institute of Technology
(काशी हिन्दू विश्वविद्यालय)/(Banaras Hindu University)
वाराणसी-२२१००५ (उ०प्र०)/Varanasi-221005 (U.P.)

Copyright Transfer Certificate

Title of the Thesis: **Computation of Some Aspects of Language Processing in the Brain Using Machine Learning**
Name of Student: **Ashish Ranjan**

Copyright Transfer

Ashish Ranjan

Date: 11/01/2022 **Signature of Student**
Place: Varanasi **(Ashish Ranjan)**

Acknowledgments

I want to take this opportunity to express my deep sense of gratitude to all who helped me directly or indirectly during this thesis work. Firstly, I would like to thank my supervisor, Dr. Anil Kumar Singh, Co-supervisor Dr. Anil Kumar Thakur, and Ex-supervisor Prof. Ravi Bhushan Mishra for being great mentors and the best advisers I could ever have. Their advice, encouragement, and critics are the source of innovative ideas; their inspiration is the cause behind the successful completion of this Thesis work. The confidence shown in me by them was the most significant source of inspiration. It has been a privilege working with them for several years. I am highly obliged to thefaculty members of Computer Science and Engineering Department and the Department of Humanistic Studies for their support and encouragement. I sincerely thank Prof. R.K Mishra, Department of Electrical engineering, and Ex. Head, Department of Humanistic Studies, Prof. P.K. Roy of Department of Biomedical Engineering, Prof. K. K. Shukla, Prof. A. K. Tripathi, Prof. S.K. Singh, Dr. Sukomal Pal of the department of Computer Science and Engineering and. Prof. P.K Panda, Dr. Sanjukta Ghosh, Dr. A.K Mishra, Dr. Vinita Chandra, Dr. Sukhada, Dr. Swasti Mishra, Dr. Amrita Dwivedi, Dr. Vishwanath Dhital, and Dr. Shail Shankar of the Department of Humanistic Studies IIT (BHU), for providing continuous support, encouragement, and advice. I sincerely thank all the Professors, Deans, office staff, support staff, and Ph.D. Research scholars of the India Institute of Technology (BHU) Varanasi, India. I express my gratitude to the Director, Registrars, Deans, Heads, and Student Alumni of the Indian Institute of Technology (BHU) Varanasi.

My memory of my study period at IIT (BHU) can never be complete without mentioning my fellow research scholars. Special thanks to Dr. Vibhav Prakash Singh, Dr. Shailendra Tiwari, Dr. Nagendra Pratap Singh, Dr. Rajesh Kumar, Dr. Tarun Maini, Dr. Sushant Kumar Pandey, Mr. Veeru Rajbhar, Mr. Manoj Bhandari, Mr. Sooraj and his wife Mrs. Greeshma, and Ms. Vandana, Mrs. Shreya for their great help and cooperation.

I extend special thanks to the non-teaching staff in the Department, particularly Mr. Vinay, Mr. Amit, Mr. Rajendra Kumar, Mr. Ravi Bharati, and Mr. Bharat Pandey, for their consistent support.

My parents, Smt. Sita Devi and Shree Ram Dular Singh, gave me the power and brain to work out on this research and their help at every level, made me see this success. Words are insufficient to express my profound gratitude to my loving wife, Dr. Subhadra Kumari, and my son Shreedhar Saraswat for their continuous support. I extend my thanks to my brother Mr. P.K. Prabhakar and my sisters (Smt. Usha Kumari, Smt. Kusum Kumari, Smt. Kumud Kumari, Smt. Vimal Kumari, and Smt. Suman Kumari) who inspired me constantly. The list couldn't be complete without mentioning the names of my grand parents Late. Shri Ram Snehi Singh and Late. Smt. Karpurni Devi. The blessings of my family Guru Shri Swami Jagannathdas mahaprabhu and Shri Lalu baba always kept me on right track. The memory of my nani who loved me most, inspired always.Special thanks to sister-in-law Anjubala her son Batuk and husband Rajeevji for their kindness and support. I am highly obliged for the support and blessing from my father-in-law Shri Kamta Bhakta, mother-in-law Smt. Vidyawati Devi and brother-in-laws Shri Krishna Kumar and Shri Vidyashankar. Last but most, I convey my gratitude to my Guru "Vaidya-Ji" Shri Dhananjay Prasad Gupta for all that I have today. Finally, I would like to end by paying my heartfelt thanks and prayers to the Almighty for his unbound love and grace.

- Ashish Ranjan

Preface

The brain is the most complex organ of the human body. Millions of neurons are connected and pass information to one another in processing thoughts, emotions, motor activities, and linguistic phenomena. Scientists have been investigating the brain for decades to answer the questions related to language functioning in the brain. The analysis of neural activation for the linguistic phenomenon is studied based on neuroimaging data for the last two decades. Cognitive state analysis or reading the brain was always exciting for researchers. Analysis of the human brain while a person is engaged in a particular task, is an essential topic in the recent development of neuro-imaging studies. The advent of non-invasive neuro-imaging has made it possible to analyze the structural and functional paradigm of the brain associated with different cognitive tasks. This opened a new window for research and innovation in neuroscience, medicine, psychology, linguistics, and biomedical engineering. Findings from fMRI, EEG, MEG, and PET contributed a lot from the perspective of neurolinguistics. It is always of great interest for the research community to discover how the brain processes linguistic items. Several approaches exist in literature in which the localization of activated brain areas corresponding to language stimuli is studied. Some research focuses on finding the brain network of different brain areas in communication and language understanding.

This thesis aims to determine the specific activation location for a particular language task and present computational models that can predict specified language entities using fMRI activation patterns as input. We have analyzed the fMRI data from three perspectives- 1. Noun level analysis, 2. Sentence level analysis, and 3. Discourse level analysis. We analyzed nouns from different categories in noun level analysis and formulated a computational model to classify the nouns. We proposed a computational model to classify affirmative and negative sentences in the sentence-level study. We figure out the neural representation of different word categories and emotional words at discourse-level analysis. Finally, we proposed a computational model to identify the mood of the readers.

Noun level analysis- Similar words activate similar neurons in the brain. Hence, the processing of the nouns is attributed to highly activated identical brain regions. Neural representation of nouns can be used as input to the computational model to identify the noun class to which a particular noun belongs. Sixty concrete nouns (with their line drawing) belonging to twelve different categories were shown on the screen, and fMRI data were recorded from nine participants. Taking fMRI data of a particular noun entity as input, our task is to determine the class to which this specific noun belongs. We employed cascaded feature selection and appropriate classification methodology to classify the nouns. We adopted two classification approaches – first classified in binary and then in twelve classes. For binary class classification, we employed the Variance threshold PCA+ LDA in a cascaded fashion for feature selection. We choose MLP and random forest for classification by selecting the prominent features from this approach.

We chose variance threshold + PCA for twelve category classification as feature selection and then used the random forest for classification strategies. We classified the concrete nouns with acceptable accuracy. We have used an only fMRI recording as input to our model. This is the first approach to classifying nouns on this data to the best of our knowledge.

At sentence level analysis, we analyzed the brain's processing of affirmative and negative sentences. This study aimed to devise a computational model that can identify the polarity of the sentence by taking fMRI data as input. Analysis of the cognitive state of the brain in processing negative and affirmative sentences has given rise to diverse results. Some parts of the brain show greater activation in processing negative sentences, while others show enhanced activation for affirmative sentences. For this, we analyzed the data set named Star Plus, data available online. Six participants view a picture on the screen, and next, the sentence explains the image. From here, we extract the fMRI data set corresponding to the sentence. The sentences are of two types- affirmative and negative. The model takes fMRI data and identifies whether a person views affirmative or negative sentences. Protuberant voxels are selected using appropriate feature selection and good classification techniques; our model provides exemplary accuracy.

First, we analyzed the affirmative and negative sentences in the brain based on fMRI data employing the k-NN classification algorithm. The fMRI data for the sentence is extracted, and the info-gain feature selection technique selects prominent feature vectors. Using these feature vectors, classification of brain state was made possible in the sentential negation identification task.

Further, we analyzed the brain's neuronal activity in sentence polarity detection tasks using the multilayer perceptron classification methodology. The whole brain is divided into almost 5000 three-dimensional volumes called voxels, from which prominent voxels are selected using symmetrical uncertainty based on entropy to classify brain state. The proposed method achieved significantly higher accuracy in classifying brain states in processing affirmative and negative sentences. The result also shows that certain brain regions like the left dorsolateral prefrontal cortex (LDLPFC) and calcarine sulcus (CALC) are prominent deterministic areas in classifying affirmative and negative sentences in the brain. In contrast, the right posterior pre-central sulcus (RPPREC) and right supramarginal gyrus (RSGA) are less contributing.

Next, we investigated the processing of affirmative and negative sentences in the brain using a greedy stepwise correlation-based feature selection technique and random forest classification approach; our model can classify the cognitive state in sentence polarity detection task with, on average, 95.41% accuracy. Our result shows that CALC, RDLPFC, and LDLPFC positively contribute to feature selection. In contrast, RPPREC, RSGA, and RFEF add very little to the polarity check. Finally, we employed the SVM-RFE feature selection and Rotation Forest classification technique. We obtained optimal solutions for the polarity detection task with 100% accuracy for some specific sets of attributes and seed values.

We analyzed the fMRI data for chapter 9 from the novel Harry Potter and the Sorcerer's stone at discourse level analysis. The reader was reading the chapter so that one word at a time appeared on the screen. The data was collected in a fashion that one fMRI image was captured for four consecutive words. Using Gumbel distribution, we extracted a token-level fMRI vector and calculated the corresponding activated areas for each POS tag set. We tagged the whole chapter using the spacy tag set.

Further, we analyzed the emotional categories following Paul Ekman's classification. There six kinds of emotions are proposed in literature- anger, fear, disgust, happiness, sadness, and surprise. We analysed all these basic emotions and their corresponding activation patterns in detail. Further, we classified the reader's mood based on the emotion category using MLP and Random Forest classification methodology. Our model proposed identifying the reader's mood with more than 73% accuracy.

Contents

List of Figures

List of Tables

Abbreviations

AI	**A**rtificial **I**ntelligence
BERT	**B**idirectional **E**ncoder **R**epresentations from **T**ransformers
MLP	**M**ulti **L**ayer **P**erceptron
RF	**R**andom **F**orest
RF	**R**otation **F**orest
fMRI	**F**unctional **M**agnetic **R**esonance **I**maging
MRI	**M**agnetic **R**esonance **I**maging
MEG	**M**agnetoencephlography
EEG	**E**lectroencephlography
SVM	**S**upport **V**ector **M**achine
CFS	**C**orrelation based **F**eature **S**election
k-NN	**K** **N**earest **N**eighbour
PET	**P**ositron **E**mission **T**omography
ML	**M**achine **L**earning
ANN	**A**rtificial **N**eural **N**etworks
SOM	**S**elf **O**rganizing **M**ap
TP	**T**rue **P**ositive
FP	**F**alse **N**egative
ROC	**R**eciever **O**perating **C**haracteristic
SMA	**S**upplementary **M**otor **A**rea
ROI	**R**egion of **I**nterest

Symbols

P	Probablity
M	Mean Vector
S	Scatter Matrix
$T2$	Transverse Relaxation Time
$H(X)$	Entropy
δ	Threshold
M_s	Heuristic Merit
r_{cf}	Mean feature class correlation
r_{ff}	Average feature-feature inter-correlation
ϕ	Null
K	Number of neighbours
$h(x)$	Hidden Layer
$W^{(1)}$	Weight Matrix
$O(x)$	Output Vector
G, s	Activation Function
$b^{(1)}$	Bias Vector
μ	Mean
σ	Standard Deviation
Σ	Summation
k	Kohen's Kappa