Declaration

I hereby declare that the work presented in this thesis titled Explainable and Generalized Deep Learning Framework: Study on Atypical Brain Network Development submitted to the Indian Institute of Technology Jodhpur in partial fulfillment of the requirements for the award of the degree of Doctor of Philosophy is a bonafide record of the research work carried out under the supervision of Dr. Romi Banerjee and Dr. Dipanjan Roy. The contents of this thesis, in full or in parts, have not been submitted to, and will not be submitted by me to, any other Institute or University in India or abroad for the award of any degree or diploma.

Certificate

This is to certify that the thesis titled Explainable and Generalized Deep Learning Framework: Study on Atypical Brain Network Development, submitted by Km Bhavna (P19CSE002) to the Indian Institute of Technology Jodhpur for the award of the degree of Doctor of Philosophy, is a bonafide record of the research work done by her under our supervision. To the best of our knowledge, the contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Romi Banerjee
Ph.D. Thesis Supervisor

Dr. Dipanjan Roy
Ph.D. Thesis Supervisor

Abstract

Social cognition refers to the ability to understand, process, and respond to social interactions and the behaviors of others. The Theory-of-Mind (ToM) brain network, associated with social cognition, is responsible for understanding another person's intentions and ideas, comprising regions such as the medial prefrontal cortex, temporoparietal junction, and superior temporal sulcus. These areas are crucial for recognizing and interpreting others' thoughts, intentions, and emotions. Recent studies highlight significant interactions between the Fronto-Parietal Network (FPN) and the Temporo-Parietal Junction (TPJ), which are critical for processing social information and predicting social evaluations. Using functional magnetic resonance imaging (fMRI) data, recent research has focused on the early development of ToM in children, extending into middle childhood and adolescence. By age five, children typically develop the ability to understand others' aims and predict actions based on false-belief paradigms, marking a critical stage in ToM development. Deficits in ToM are evident in neurodevelopmental disorders like autism spectrum disorder (ASD), where individuals show significant impairments in social cognition and communication. Despite extensive research, the heterogeneity within the ASD population and incomplete understanding of its neurobiology pose challenges.

Our research focuses on quantifying the temporal stability of ToM and Pain brain networks from early childhood (3 years) to adulthood, using fMRI-based dynamic functional connectivity analyses. We investigate whether temporal stability patterns are associated with performance in false-belief reasoning tasks, particularly in children aged 3–12 years. To decode cognitive states during a naturalistic movie-watching task, we developed an explainable spatiotemporal connectivity-based graph convolutional neural network (Ex-stGCNN). We further employ an explainable convolutional variational autoencoder (Ex-Convolutional VAE) to predict individual false-belief task performance, identifying key brain regions contributing to these predictions as subject-specific neural fingerprints.

In the context of ASD, which affects social behavior and communication ability, we utilize large-scale fMRI datasets to develop an explainable deep learning framework aimed at identifying both shared and ASD-specific variability in ToM brain networks. Our models classify typically developing (TD) and ASD individuals using functional connectivity and meta-connectivity features derived from ToM, Default-Mode Network (DMN), Central Executive Network (CEN), and Salience Network (SN). Furthermore, we use these features to predict clinical scores related to ASD symptoms, including ADOS-Total, ADOS-Social, DSM-IV, and Full-Scale IQ (FIQ), through connectome-based predictive modeling (CPMM). The model evaluation demonstrated robust classification performance (AUC 0.86) and accurate symptom severity prediction with minimal error across test sets. Finally, we also identify ASD sub-groups through edge-centric meta-connectivity features using a Rough Fuzzy C-means approach. Overall, this work demonstrates the power of explainable deep learning models in bridging brain network dynamics and symptom heterogeneity in ASD, contributing to more individualized insights into atypical brain development.

Acknowledgments

I would like to express my gratitude to my Ph.D. Thesis Supervisors, Dr. Romi Banerjee and Dr. Dipanjan Roy, for being mentors, guides, and friends throughout my Ph.D. journey. Their invaluable guidance, insightful conversations, and constant support made my research enjoyable and rewarding. Their qualities of patience, optimism, and pursuit of perfection are truly inspiring and have left a lasting impact on me. I would also like to extend my gratitude to the members of the Student Research Committee, Prof. Neeraj Jain, Dr. Anand Mishra, and Dr. Angshuman Paul, for their valuable suggestions and guidance during my research work. I am grateful to all the staff of the Computer Science and Engineering Departments at IIT Jodhpur for their continuous and valuable support.

I am deeply appreciative of my batchmates, classmates, and labmates for the regular discussions on research work. A heartfelt thank you to Shubham Sharma, Sneha Prasad, and Alok Shukla for their support, their constant words of encouragement, and for making my stay at the institute a wonderful experience.

I am very grateful to my parents for their unconditional love, endless patience, and unwavering belief in me. They have given me the freedom to choose my path and have always been my pillars of strength. I also thank my brother, Dr. Manish Kumar Bajpai, for his constant support, encouragement, unconditional trust, and for giving me all the opportunities that have shaped me into the person I am today.

Contents

page

Abstract v
Acknowledgements vii
Contents ix
List of Figures xi
List of Tables xv
List of Symbols xvii
List of Abbreviations xix

Chapter 1: Introduction 1
1.1 Brain Network Analysis . . . 1
1.2 Deep Learning Frameworks to Study Brain Functional Connectivity Patterns . . . 2
1.2.1 Functional and Structural Connectivity . . . 2
1.2.2 Resting-State Functional Connectivity . . . 2
1.2.3 Brain Development and Connectivity . . . 4
1.2.4 Clinical Conditions and Functional Connectivity . . . 4
1.2.5 Image Analysis in the Context of Brain-Activated Regions . . . 4
1.2.6 Deep Learning for Analyzing Functional Connectivity . . . 5
1.3 Motivation for the Project . . . 6
1.3.1 Characterizing Development of Brain Networks From Early Childhood . . . 6
1.3.2 Identifying Brain Connectivity-Based Features as Individual Fingerprints . . . 6
1.3.3 Design of Explainable and Generalized AI Framework to Study Atypical Brain Network Development . . . 7
1.4 Research Gap . . . 7
1.4.1 Identification of Brain Features . . . 7
1.4.2 Deep Learning in Predicting Brain Disorder . . . 8
1.4.3 Interpretation of Deep Learning Results . . . 8
1.5 Contributions . . . 8
1.5.1 Cognitive Development of Brain Networks . . . 8
1.5.2 Decoding of Brain Task using Graph Convolutional Neural Networks . . . 10
1.5.3 Identification of Source of Variability Between Autistic and Neuro-typicals using Contrastive Varaitional Auto-Encoder Model . . . 10
1.5.4 Explainable and Generalized AI Framework for Classification of Autistics from Neuro-typicals . . . 11
1.6 Organization of Thesis . . . 11

Chapter 2: Background and Related Work 13
2.1 Neuroimaging Techniques: An Overview . . . 13
2.1.1 Structural MRI and Neuroanatomy . . . 13
2.1.2 Functional MRI and Brain Activity . . . 14
2.1.3 Near-Infrared Spectroscopy (NIRS) and Alternative Methods . . . 14
2.1.4 Artificial Neural Networks in Neuroimaging . . . 14
2.2 Brain Network Modelling . . . 14
2.2.1 Human Brain Mapping . . . 15
2.2.2 Current Frontiers in Brain Network Mapping . . . 16
2.3 Deep Learning . . . 19
2.4 Deep Learning on Brain Connectivity . . . 21
2.4.1 Challenges using Deep Learning on Brain Inspired Model . . . 22

Chapter 3: Characterization of the Temporal Stability of ToM and Pain Functional Brain Networks Carry Distinct Developmental Signatures during Naturalistic Viewing 27
3.1 Introduction . . . 27
3.2 Proposed Approach . . . 29
3.2.1 fMRI Movie Watching Dataset and MRI Preprocessing . . . 29
3.2.2 Explicit ToM Task and False-Belief Composite Score and fMRI Analysis . . . 31
3.2.3 BOLD Phase Coherence and Estimation of Dynamic Functional Connectivity . . . 33
3.2.4 Computation of Dominant Dynamic Functional Connectivity Subspaces . . . 35
3.2.5 Computation of Network Temporal Stability using Dominant dFC Subspaces . . . 36
3.2.6 Validation of Results . . . 37

3.2.7 Stochastic Characterization of dFC . . . 37
3.3 Results . . . 38
3.3.1 Distinct Temporal Stability of ToM and Pain Network in 3-12 years age . . . 38
3.3.2 Association between Performance in false belief task and Temporal Stability of DFC 41
3.4 Discussion . . . 57
3.4.1 Stochastic properties of dynamic functional connectivity and age effects in children . 58
3.4.2 Aging introduces temporal variability in the evolution of dynamic functional connectivity ToM and Pain functional brain networks during movie watching . . . 59
3.4.3 Association between performance on false belief tasks in children and Temporal Stability of DFC in ToM functional brain networks . . . 59
3.5 Summary . . . 60

Chapter 4: Explainable Deep-Learning Framework: Decoding Brain States and Predicting Performance in False-Belief Tasks Across Developmental Age-Group 61
4.1 Introduction . . . 61
4.2 Material and methods . . . 63
4.2.1 Participants and fMRI Preprocessing . . . 63
4.2.2 fMRI Data Analysis and Extraction of Feature Sets . . . 65
4.2.3 Decoding of States using Explainable Spatiotemporal Connectivity based Graph Convolutional Neural Network . . . 66
4.2.4 Prediction of Individual Performance in False-belief Task using Explainable Convolutional Variational Autoencoder Model . . . 70
4.3 Results . . . 76
4.3.1 Computation of FC and ISFC Matrices . . . 76
4.3.2 Decoding of cognitive states using Ex-stGCNN model . . . 76
4.3.3 Prediction of Individual Performance on False-belief tasks . . . 77
4.4 Discussion . . . 84
4.4.1 Decoding of Cognitive states using stGCNN Model . . . 85
4.4.2 Feature identification and brain fingerprinting using Ex-stGCNN . . . 86
4.4.3 Prediction of Individual Performance in False-Belief Taks . . . 86
4.4.4 Limitations and future Scope . . . 87
4.5 Summary . . . 88

Chapter 5: Identifying Functional Variations in Theory of Mind Brain Networks between Autism and Typically Developing Individuals using Contrastive Machine Learning and Explainable DNN Model 89
5.1 Introduction . . . 89
5.2 Material and methods . . . 91
5.2.1 Participants and MRI Preprocessing . . . 91
5.2.2 Resting-State Functional Connectivity . . . 92
5.2.3 Identification of Shared and Specific Variability . . . 93
5.2.4 Classification Using Explainable Artificial Neural Networks (Ex-ANN) Model . . . 95
5.2.5 Identification of neurobiological features underlying classification . . . 98
5.3 Results . . . 99
5.3.1 Calculation of resting state-Functional Connectivity Matrices . . . 99
5.3.2 Identification of Shared And ASD-Specific Features or Source of Variability . . . 100
5.3.3 Classification of ASD and TD Individuals using Ex-ANN Model . . . 100
5.4 Discussion . . . 106
5.5 Summary . . . 108

Chapter 6: A Lightweight, End-to-End Explainable, and Generalized DNN Framework to Classify Autistics from Neuro-Typicals 109
6.1 Introduction . . . 109
6.2 Materials and Method . . . 111
6.2.1 fMRI Dataset and Preprocessing . . . 111
6.2.2 Proposed Method for Calculation of Meta-Connectivity . . . 113
6.2.3 Node-Edge Connectivity Based Graph Attention Network for Classification . . . 114
6.2.4 Symptom-Severity score prediction for ASD Samples . . . 117
6.3 Results . . . 118
6.3.1 Computation of Meta-Connectivity Matrices . . . 118
6.3.2 Classification of ASD and TD Individuals using Proposed Ex-NEGAT Model . . . 118
6.3.3 Prediction of Symptom Severity Score . . . 121
6.4 Discussion . . . 123
6.4.1 Advantages of using meta-connectivity feature to discover new ASD biomarkers . . . 123
6.4.2 Advantages of using MC matrix and attention-based GNN . . . 125
6.4.3 Predicting symptom severity in ASD using MC matrices and identifying relevant ROIs 126

Chapter 7: Identification of Sub-Groups within ASD samples using Rough Fuzzy C-Means Clustering Approach 127

7.1 Introduction . . . 127
7.2 Methods and Materials . . . 129
7.2.1 fMRI Dataset and Preprocessing . . . 129
7.2.2 Calculation of Meta-Connectivity Feature-set . . . 133
7.2.3 Identification of ASD Sub-groups using Rough Fuzzy C-Means Algorithm . . . 134
7.3 Results . . . 137
7.3.1 Computation of Meta-Connectivity Matrices . . . 137
7.3.2 Identification of ASD Sub-Groups using RFCM . . . 138
7.3.3 Association between Identified Clusters and Behavioral Scores . . . 142
7.4 Discussion . . . 142
7.4.1 Analysis of Each Cluster . . . 143
7.4.2 Association between Identified Clusters and Behavioral Scores . . . 143
7.5 Summary . . . 144

Chapter 8: Conclusions 145
8.1 Conclusion . . . 145
8.2 Future Work . . . 147

Bibliography 149

List of Figures

Figure	*Title*	*Page*
1.1	Different types of neuroimaging approaches, such as gene imaging, structural MRI, functional MRI, PET scan, and DTI scan, used to study brain dynamics.	3
1.2	Figure shows the overall pipeline of the thesis, in which, 2 illustrates characterization of temporal stability, Whereas, 3 is representing decoding of brain brain states and prediction of individual performance in false-belief tasks. Finally 4 and 5 represent deep learning model to classify autistics and neurotypicals, as well as identifying sub-groups within ASD-samples and their symptom-severity scores.	9
2.1	Pipeline to study brain dynamics using structural and functional images.	13
2.2	Figure illustrates pipeline to perform classification using machine learning approach.	20
2.3	Figure illustrates pipeline to perform prediction using deep learning approach.	20
2.4	Figure illustrates pipeline to perform prediction using convolutional neural network model.	21
3.1	Illustrative overview of the proposed unsupervised characterization of dFC of Tom and Pain functional brain networks for identifying critical age where reasonable temporal stability is achieved and, subsequently, how temporal stability patterns predict performance in the false-belief task. The pipeline of the analysis proposed in this work is organized as follows: **1)** Identifying brain regions anchored in ToM and Pain networks and extraction of BOLD fMRI time-series signals from ToM and Pain functional brain regions, **2)** Calculation of Angular distance and Mahalanobis distance using after characterization of dominant dFC subspaces using an unsupervised approach, leading to temporal stability computations, and **3)** Applied auto-regression model to check temporal fluctuations and identifying Markovian process that fully explains the dFC evolution.	30
3.2	Movie demonstration: **A)** response magnitude that evoked maximum activation in ToM and pain networks. **B)** depicts the movie scenes where higher activation occurred. T1, T2, T3, T4, and T5 represent the ToM scenes with higher activation, whereas P1, P2, P3, P4, and P5 show the pain scenes with higher activation.	32
3.3	Angular distance matrices depicting temporal patterns of all ROIs of the ToM and Pain networks for six age groups (3-yrs, 4-yrs, 5-yrs, 7-yrs, 8-12 yrs children and Adults). Each entry in the matrix represents principal angle $\phi(t_x, t_y)$ between dominant DFC subspaces t_x and t_y that varies from 0 (lower angular distance) to $\pi/2$ (higher angular distance). 3-yrs and 4-yrs age groups showed shorter-lived patterns of temporal stability, whereas 5-yrs, 7-yrs, 8-12 yrs, and adults showed longer-lived patterns of temporal stability for defined time-courses.	43
3.4	Mahalanobis distance matrices depicting temporal patterns of all ROIs of the ToM and Pain networks for six age groups (3, 4, 5, 7,8-12-year children and Adults) that show low dimension pattern high dimensionality functional connectivity.	45
3.5	Angular distance matrices depicting temporal patterns of ToM network ROIs for each of the six age groups (3, 4, 5, 7, 8-12 yrs and Adults age groups). Earlier, we defined Time-points with maximum activation in ToM network = 18-27, 82-90, 92-102, 140-149, 150-159. We observed longer-lived local spread temporal stability patterns for defined time courses in the ToM network for 5-yrs to adulthood. Whereas we found shorter-lived repeated patterns of temporal stability for 3-yrs and 4-yrs age groups.	46

3.6 Angular distance matrices depicting temporal patterns of pain network ROIs separately for each of the six age groups (3, 4, 5, 7, 8-12 yrs and Adults age groups) (maximum activation in pain network =35-49, 78-85, 88-92, 104-115, 159-168). We observed longer-lived local spread temporal stability patterns for defined time courses in the pain network for 3-yrs to adulthood. 47

3.7 Mahalanobis distance matrices depicting temporal patterns of ToM and pain network ROIs for each of the six age groups (3, 4, 5, 7, 8-12 yrs and Adults age groups). We observe a trend of decreasing high distance values to lower overall distance values from children to adults. 48

3.8 Entropy and Frobenius distances matrices: **A and C** are showing entropy for Angular and Mahalanobid distances including All ROIs, whereas **B and D** are depicting Frobenius distance between ToM and Pain ROIs for Angular and Mahalanobis distance. **E,F,G, and H** are showing entopy for Angular and Mahalanobis distance for ToM and Pain ROIs separately. Statistically significant differences are represented using $*(P \leq 0.05)$, $**(P \leq 0.01)$, $***(P \leq 0.001)$, $****(P \leq 0.0001)$, ns (not significant). 49

3.9 Stochastic characteristics of dFC. The figure shows the principal angle, $\phi(t)$, modeled as an autoregressive $AR(\rho)$ process in stochastic modeling. The value of the model order ρ was systematically changed from 0 to 100. The graphic displays the AIC values associated with the model order. The inset shows the initial minimum value of the AIC (Akaike Information Criterion) and the matching model order. . We observe the best fit model for $\rho \geq 2$. 50

3.10 AIC matrices for ToM Network in six-groups analysis, for Angular Distance. We observed minimum model order of $\rho \geq 4$ for Angular distance. 51

3.11 AIC matrices for ToM Network in six-groups analysis, for Mahalanobis Distance. We observed minimum model order of $\rho \geq 4$ for Mahalanobis distance. 52

3.12 AIC matrices for Pain Network in six-groups analysis, for Angular Distance. We observed minimum model order of $\rho \geq 4$ for Angular distance. 53

3.13 AIC matrices for Pain Network in six-groups analysis, for Mahalanobis Distance. We observed minimum model order of $\rho \geq 4$ for Mahalanobis distance. 54

3.14 Angular distance matrices depicting temporal patterns for false-belief task-based pass, fail, and inconsistent groups. It showed false-belief task is dependent on temporal stability. The figure shows higher temporal stability in ToM network for the pass group, moderate stability for the inconsistent group, and lowest temporal stability for the fail group. 55

3.15 **A, C, E** depicts entropy plots for angular distance for the false-belief task-based pass, fail, and inconsistent groups. **B, D, F** are depicting entropy plots for Mahalanobis distance. In the case of entropy plots, Mahalanobis distance is better at capturing the differences in temporal stability in different groups. 56

4.1 Illustrative overview of end-to-end explainable deep-learning framework for decoding of cognitive states and prediction on performance of false-belief task-based pass, fail, and inconsistent groups. 63

4.2 Movie demonstration: **A)** response magnitude that evoked maximum activation in ToM and pain networks. **B)** depicts the movie scenes with higher activation. $T_i \in [T_1, T_2, T_3, T_4, T_5]$ is representing ToM scenes, and $P_i \in [P_1, P_2, P_3, P_4, P_5]$ is representing pain scenes with higher activation. 64

4.3 The figure illustrates the calculation of FC and ISFC matrices from movie-watching fMRI data that has been further used as input for decoding of cognitive state and prediction of individual performance in false-belief task purpose. Here, S represents task-evoked brain activity; I represents intrinsic brain activity; N represents noise 66

4.4 The architecture of proposed Ex-stGCNN model. Firstly, FC and ISFC matrices are converted into adjacency matrices and then into graph Laplacian. Finally, the graph representation of these matrices is provided as input to the model for training. We trained this proposed model using FC and ISFC matrices separately. The SHAP approach is used to extract neurological brain fingerprints to check the contribution of each brain region in the prediction. 71

4.5 Architecture of proposed Ex-Convolutional VAE model for predicting false-belief task-based pass, fail, and inconsistent groups. The proposed architecture first used the 2D convolutional approach to create a 32-dimensional latent space and trained the prediction model using the ADAM optimization approach (a 32-D latent space was used to train the model). The SHAP approach is used for the explainability of the proposed model. 73

4.6 Performance of Ex-stGCNN model: **A and B** show ROC and precision-recall curves of Ex-stGCNN model using ISFC matrices. 78

4.7 Feature Importance. **A and B** show brain regions that contributed most to the prediction using ISFC and FC matrices. We extracted three dominating ToM regions and three dominating pain regions. **C and D** show a relationship between feature values and the impact they have on the model's predictions. 79

4.8 Performance of GCNN model on Age-wise sub-groups: **A)** shows accuracy of proposed Ex-stGCNN model on sub-groups using ISFC matrices. **B)** shows accuracy using FC matrices. Our results suggest that we can achieve considerable performance using ISFC matrices at early childhood stage. In sub-group analysis, both Walklets and Node2Vec node embedding methods performed approximately the same. 80

4.9 Performance using leave-one-out approach: **A)** shows performance in terms of accuracy for each individual. It shows average accuracy for all 10 time-course for each participant. The results suggested that as age increases, the performance of the model also increases. **B)** shows the weight of each time-course in the prediction. 81

4.10 Functional connectivity between ToM and pain networks for false-belief task-based pass, fail, and inconsistent groups. Pass group showed stronger connectivity between ToM and pain network, whereas inconsistent group showed moderate connectivity, and fail group showed weaker connectivity. 82

4.11 Inter-Subject Functional connectivity between ToM and pain networks for the false-belief task-based pass, fail, and inconsistent groups. The FC and ISFC matrices show similar patterns, i.e., stronger connectivity for the pass group, moderate connectivity for the inconsistent group, and weaker connectivity for the fail group. 82

4.12 Performance for prediction of false-belief task: **A)** Shows importance of features and their SHAP values using ISFC matrices for the proposed Ex-Convolutional VAE model. **B)** shows feature importance using ISFC matrices for proposed Ex-stGCNN model **C)** Shows overall performance. 83

5.1 Illustrative overview of End-to-End Explainable Artificial Intelligence framework for identifying the common source of variability that helps classify ASD samples from neurotypicals using Theory-of-Mind brain network. The contribution of the paper is as follows: Steps 1 and 2 represent data extraction; Step 3 is the calculation of the Functional Connectivity matrix; Step 4 is the identification of the source of variability; Step 5 is a classification of the ASD and TD samples. 91

5.2 Illustrates pipeline for identifying common-source of variability between ASD and TD samples. A Contrastive Variational Autoencoder (CVAE) was trained to separate ASD-specific variations from those shared with TD samples. Representational Similarity Analysis (RSA) compared subject similarities in shared and ASD-specific features to behavioral, demographic, and rs-FC matrices data. 96

5.3 Comparison of Variational and Contrastive-variational Autoencoder model's matrices. CVAE features best capture the common variation between ASD and TD groups, and parameters associated with ASD are best obtained by clinically ASD-specific measures and resting-state functional connectivity. VAE model enable to capture shared and ASD-specific features. 101

5.4 The figure illustrates the model performance curve and precision-recall curve. Here, the results are reported for proposed Ex-ANN, LSTM-RNN, and CapsNet models, in which proposed model performance is the best. 103

5.5 (A) Importance of Features that underlie classification for ASD and TD samples using SHAP approach. (B) Importance of Features that underlie classification for ASD and TD samples using Integrated gradient. Both approaches identified approximately the same brain regions. 103

5.6 Illustration of sub-groups division for the TD population to identify the effect of developmental age on the model's performance. The rs-FC matrices have been shown on average level. rs-FC matrices are showing an increase in correlation between brain regions from 6-25 age range. 104

5.7 Illustration of average rs-FC matrices for sub-groups division in ASD population. the matrices are showing reverse pattern as compare to TD samples, in this there is decrease in correlation between brain region from 6-25 age-ranfe. 105

6.1 Illustrative overview of light-weight, generalized proposed Ex-NEGAT framework for classifying ASD from TD using meta-connectivity as an edge-centric feature set. Step 1 involves BOLD time-series signal extraction and calculation of meta-connectivity matrices; Step 2 shows the details of the proposed architecture NEGAT framework to perform classification. 110

6.2 Calculation of Static, Dynamic functional connectivity, and Meta-connectivity: In the traditional approach **(A)**, neural activity is averaged over time to create a functional connectivity (FC) matrix. We calculated a dynamic FC stream using shorter sliding windows **(B)**, leading to a dynamic FC matrix. In **(C)**, we considered each FC link as a dynamic variable, giving rise to a Meta-Connectivity (MC) matrix by capturing covariances between these dynamic links. 115

6.3 **A) and B)** are depicting classification results using meta-connectivity with window size =30 sec. **C) and D)** are depicting classification results using meta-connectivity with window size = 20 sec. Whereas **E) and F)** are depicting classification results using functional connectivity. **G) and H)** are showing classification results using Dynamic Functional Connectivity. 121

6.4 Performance of proposed model using Leave-one-out approach. The results suggested that the proposed model outperforms using the proposed meta-connectivity as a feature set. . . 123

6.5 Prediction of Symptom-severity scores (ADOS-Total, ADOS-Social, ADOS-Module, ADOS-Communication, ADOS-Stereo, and FIQ scores) using Connectome-based prediction modeling in which Meta-connectivity was used as feature set. **A)** shows results for brain regions whose edge connectivity contributed positively to the prediction. Whereas **B)** depicts results for brain regions whose edge connectivity contributed negatively to the prediction. 124

7.1 Figure illustrates the workflow of proposed Rough Fuzzy C-Means algorithm, in which meta-connectivity between ToM, DMN, CEN, SN brain network for each individual subjected as input. 130

7.2 Calculation of Static, Dynamic functional connectivity, and Meta-connectivity: In the traditional approach **(A)**, neural activity is averaged over time to create a static Functional Connectivity (FC) matrix. We calculated a dynamic FC stream using shorter sliding windows **(B)**, leading to a dynamic FC matrix. In **(C)**, we considered each FC link as a dynamic variable, giving rise to a Meta-Connectivity (MC) matrix by capturing covariances between these dynamic links. 135
7.3 **A)** is showing identification of no. of clusters using Elbow approach. **B), C), and D)** are showing visualization of clusters using different approaches. 138
7.4 Illustrating the association between all three clusters,i.e., how one sample in a cluster is also associated with other clusters with a membership confidence value. The result suggested that sub-group 2 was highly associated with sub-groups 1 and sub-groups 3, whereas sub-group 1 and sub-group 3 were slightly associated with each other. 139
7.5 Average Meta-connectivity matrices for Neuro-typicals, sub-group 1, sub-group 2, and sub-groups 3 that show that sub-group 3 is more close to neuro-typicals and less associated with ASD symptoms. Whereas sub-group 1 is more associated with ASD symptoms. 139
7.6 Figure illustrates the correlation matrix among behavioral scores for each subgroup. We found lower correlation for 3rd sub-group and higher correlation for 1st and 3rd sub-groups. 140
7.7 Figure illustrates gender and age distribution across clusters. In part B, we found that most of the samples belonged to the early developmental age range, and after 25 years of age, there were very few samples. In part A, we found that males were more included in the case of autism. 140
7.8 Figure illustrates regression coefficient with confidence interval and ADOS severity in each subgroup. We found more severity in 1st sub-group. 141

List of Tables

Table	Title	Page
2.1	Survey on Autism Detection using DNN	23
2.2	Survey on Autism Detection using DNN	24
3.1	Table shows selected ToM and pain brain regions and corresponding MNI coordinated for extracting time-series signal.	34
3.2	Table showing the selected time-course where higher activation occurred for ToM and pain networks.	34
3.3	Table with difference and p-values calculated using the post-hoc pair-wise Dunn-Sidak test for Mahalanobis Distance-derived Z-scores for the six age subgroups (3, 4, 5, 7, 8-12 yrs and Adult), and for the False-Belief Performance results	40
3.4	Table with p-values calculated for Kolmogorov-Smirnov test between three (3-5, 7-12 year olds, and Adult) age groups of subjects, for Angular and Mahalanobis Distance entropy values.	40
3.5	Table with p-values calculated for Kolmogorov-Smirnov test between six (3, 4, 5, 7, 8-12 year olds, and Adult) age groups of subjects, for Angular and Mahalanobis Distance entropy values.	41
3.6	Table with p-values calculated for Kolmogorov-Smirnov test for three (3-5, 7-12 year olds, and Adult) and six (3, 4, 5, 7, 8-12, and Adult) age group subjects, Frobenius Distance between TOM and Pain Angular and Mahalanobis Distances.	42
3.7	Table with (the first) minimal AIC score and the corresponding Model Order number, obtained for the stochastic characterization of the Angular and Mahalanobis Distance temporal matrices for 6 age groups (3, 4, 5, 7, 8-12, and Adult) of subjects	44
3.8	Table with (the first) minimal AIC score and the corresponding Model Order number, obtained for the stochastic characterization of the Angular and Mahalanobis Distance temporal matrices for three age group (3-5, 7-12 year olds and Adult) subjects.	45
4.1	ToM and pain brain regions and corresponding MNI-coordinated for extracting time-series signal.	65
4.2	Description of movie-clip events with higher activation for ToM and pain networks.	67
4.3	Table shows an implementation of the proposed GCNN model architecture, where O = No. of task, N is = input size, $F_i \in [F_1, F_2, F_3]$ = no. of filters at i_{th} graph convolutional layer, K = polynomial order of filters.	68
4.4	Table shows proposed architecture of Ex-Convolutional VAE model and the parameters with values that have been used on it.	74
4.5	Table shows the performance of proposed Ex-stGCNN model using ISFC and FC matrices. We found more accurate results using ISFC matrices as compared to FC matrices.	78
4.6	Table shows the comparison between performance of traditional models and proposed model on complete dataset. The proposed model outperforms the existing models.	78
4.7	Table shows the comparison between the performance of multiple models and the proposed model for predicting individual performance on false-belief tasks. The proposed Ex-Convolutional VAE model outperforms as compared to Ex-stGCNN and other existing models.	81

5.1 Demographic Information of ASD and TD groups From ABIDE dataset 93
5.2 Brain region with MNI coordinates that are associated with Theory-of-Mind concept and Default-mode-Network 93
5.3 Proposed architecture of VAE model and the parameters with values used in it. 97
5.4 Proposed architecture of CVAE model and the parameters with values used in it. 98
5.5 Table is showing results identified by CVAE and VAE model. 101
5.6 ASD vs. TD classification accuracies on ABIDE Cohort using different models. The training was performed on the ABIDE I dataset, while testing was performed on the ABIDE II dataset to check the model's generalizability. 102
5.7 Performance of proposed model using the leave-one-out approach, in which testing is performed on single site data as an independent dataset. 106

6.1 Demographic information of ASD and TD groups from ABIDE dataset 111
6.2 Brain region with MNI coordinates that are associated with Theory-of-Mind, Default-mode-Network, Central Executive, and Salience Networks 112
6.3 Performance of proposed Ex-NEGAT model using FC,dFC, and MC matrices as feature set with $d = 3$, where W.s represents window size. 119
6.4 Performance of proposed model using different values of depth d. The table is highlighting Accuracy, F1-Score, Precision, Recall, and Execution time in seconds. 119
6.5 Performance comparison of multiple models including proposed model in the classification of ASD from neuro-typical samples using MC matrices as feature set. 120
6.6 Validation of performance of the proposed EX-NAGET model using the Leave-one-out approach. 122
6.7 Table shows the top 10 pairs of nodes whose edge connectivity contributed most to the prediction. The table shows edge connectivity strength for ASD and TD groups. 122

7.1 Demographic information of ASD and TD groups from ABIDE dataset 131
7.2 Demographic information of participants after pre-processing including ABIDE I and ABIDE II datasets 131
7.3 Brain region with MNI coordinates that are associated with Theory-of-Mind, Default-mode-Network, Central Executive, and Salience Networks 132

List of Abbreviations

AAL	Automated Anatomical Labeling
ABIDE	Autism Brain Imaging Data Exchange
Adam	Adaptive Moment Estimation
ADOS	Autism Diagnostic Observation Schedule
AFNI	Analysis of Functional NeuroImages
AI	Artificial Intelligence
ANN	Artificial Neural Network
ANTs	Advanced Normalization Tools
AR	Auto-Regressive
ASD	Autism Spectrum Disorder
ART	Artifact Detection Tools
BOLD	Blood-Oxygen Level Dependent
CEN	Central Executive Network
CNN	Convolutional Neural Network
CPMM	Connectome-based Prediction Modelling
C-PAC	Configurable Pipeline for the Analysis of Connectomes
CVAE	Convolutional Variational Autoencoder
DCNN	Deep Convolutional Neural Network
DFS	Depth-First Search
DL	Deep Learning
DNN	Deep Neural Network
dFC	Dynamic Functional Connectivity
dFCwalk	Dynamic Functional Connectivity Walk
DMN	Default Mode Network
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition
DTI	Diffusion Tensor Imaging
EEG	Electroencephalography
Ex-Convolutional VAE	Explainable Convolutional Variational Autoencoder
Ex-NEGAT	Explainable and Generalized Node-Edge Connectivity-based Graph Attention Neural Network
Ex-stGCNN	Explainable Spatiotemporal Graph Convolutional Neural Network
FC	Functional Connectivity
FDR	False Discovery Rate
FIQ	Full-Scale Intelligence Quotient
fMRI	Functional Magnetic Resonance Imaging
GCN	Graph Convolutional Network
GNN	Graph Neural Network
GSO	Grid Search Optimization
HCP	Human Connectome Project
HIPAA	Health Insurance Portability and Accountability Act
ICA	Independent Component Analysis
IRB	Institutional Review Board
ISFC	Inter-Subject Functional Connectivity

LSTM-RNN	Long Short-Term Memory Recurrent Neural Network
MC	Meta-Connectivity
ML	Machine Learning
MLP	Multiple Layer Perceptron
MRI	Magnetic Resonance Imaging
MVPA	Multi-Voxel Pattern Analysis
NIRS	Near-Infrared Spectroscopy
PCA	Principal Components Analysis
PET	Positron Emission Tomography
RFCM	Rough-Fuzzy C-Means
RNN	Recurrent Neural Network
ROI	Region of Interest
ROIs	Regions of Interest
rs-FC	Resting-State Functional Connectivity
SAE	Sparse Autoencoder
SGD	Stochastic Gradient Descent
SHAP	SHapley Additive exPlanations
SFC	Static Functional Connectivity
SN	Salience Network
SPM	Statistical Parametric Mapping
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TD	Typically Developing
TE	Echo Time
T1	Longitudinal Relaxation Time
T2	Transverse Relaxation Time
ToM	Theory of Mind
TR	Repetition Time
tsfresh	Time Series Feature Extraction on the Basis of Scalable Hypothesis Tests
VAE	Variational Autoencoder