C3 submission | MIVP Project | Second Semester | Group No : 10

Autism Spectrum Detection using FMRI Data

ABIDE

Autism Brain Imaging
Data Exchange

Submitted by:
Arzoo Khan (MIT2020028)
Ashutosh Gupta (MIT2020029)

Under Guidance of:
Prof. Anupam Agrawal

Introduction

ASD (Autism Spectrum Disorder), is a neurological disorder, that can observed in children between age six to seventeen. ASD affects the social skills, behavioural skills as well as communication skills in individuals. So , it becomes very crucial to detect ASD at an early age, so that the quality of life of children may be improved. FMRI (Functional Magnetic Resonance Imaging), are used in the training of ASD models, as it can be used to study structures of brain and show the correlated fluctuations in BOLD (Blood oxygen level-dependent) signals from different brain regions.

Project Objective

- It has been observed that ASD affects the functional connectivity between different regions of the brain.
- Keeping in mind the popularity of ANN ,and FMRI data we have designed a model that is capable of predicting Autism Spectrum Disorder using ABIDE dataset.

Challenges

1.Data is very complex so the optimization scope is less.
2.Co-relation calculation is not generalized for all parcellations.

Keywords

FMRI data, ABIDE dataset, ANN(Artificial Neural Network),Autism Spectrum Disorder(ASD)

Literature Review

15 min

Previous Work-

1. Automated Detection of Autism Spectrum Disorder Using a Convolutional Neural Network

https://www.frontiersin.org/articles/10.3389/fnins.2019.01325/full#:~:text=We%20detected%20ASD%20patients%20using,the%20patterns%20of%20functional%20connectivity.

Accuracy- 70.22% using the ABIDE I dataset and the CC400 functional parcellation atlas of the brain.

2. **ASD-DiagNet: A Hybrid Learning Approach for Detection of Autism Spectrum Disorder Using fMRI Data**

https://www.frontiersin.org/articles/10.3389/fninf.2019.00070/full

Accuracy- 73.0% using the ABIDE I dataset and the CC200 functional parcellation atlas of the brain.

3. **Diagnosing Autism Spectrum Disorder from Brain Resting-State Functional Connectivity Patterns Using a Deep Neural Network with a Novel Feature Selection Method**

Most recently, Heinsfeld et al. (2018) used a deep learning based approach and achieved 70% accuracy for classifying 1, 035 subjects (505 ASD and 530 controls). They claimed this approach improved the state of the art technique. In their technique, distinct pairwise Pearson's correlation coefficients were considered as features. Two stacked denoising autoencoders were first pre-trained in order to extract lower dimensional data. After training autoencoders, their weights were applied to a multi-layer perceptron classifier (fine-tuning process) which was used for the final classification. However, they also performed classification for each of the 17 sites included in ABIDE dataset separately, and the average accuracy is reported as 52%. The low performance on individual sites was justified to be due to the lack of enough training samples for intra-site training.

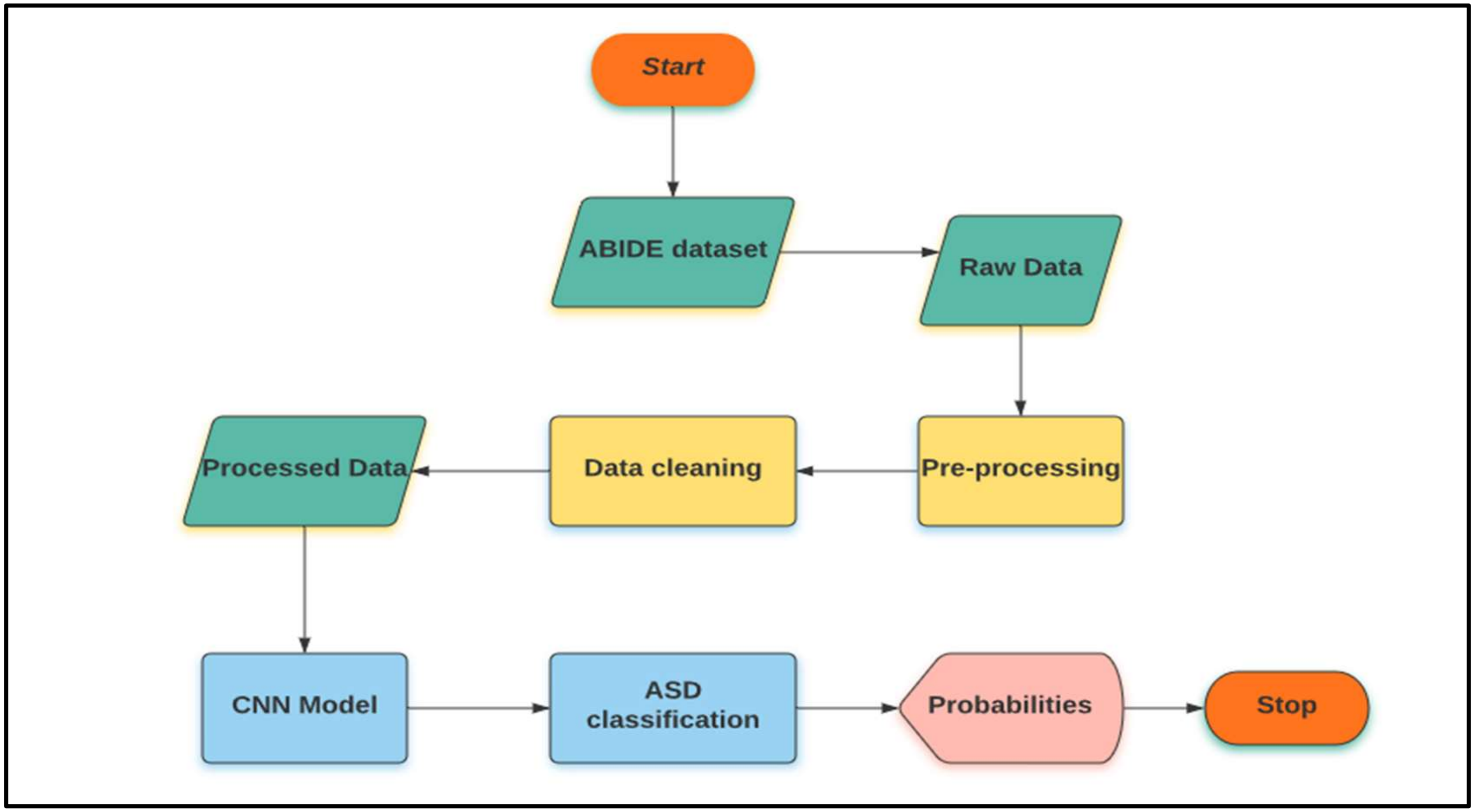
https://www.frontiersin.org/articles/10.3389/fnins.2017.00460/full

Problem Statement and Scope

15 min

Problem Statement : We are required to detect autism in patients based on FMRI data.

Scope : Recently, the use of ANN, has gained a lot of popularity, in dealing with classification and learning problems. We extracted detect brain biomarkers (Which used for of early detection of ASD) in patients suffering from ASD , using the FMRI data. The proposed method is based on detection of ASD using ANN , based on FMRI data. the detection of this disease at an early age is very crucial for human kind. With the advancements in various machine learning and deep learning algorithms, it has now been possible to detect ASD in patients based on the earlier records of various patients. We have introduced a ANN model that is capable of predicting ASD based on FMRI data, since FMRI data is widely used for studying different brain activities. We have used ABIDE dataset that are primary source of FMRI data. We have also compared our proposed model with the different machine learning algorithms like KNN, Random forest, etc. that had previously been used for ASD detection.



Methodology : Keeping in mind the popularity of ANN , we have designed a model that is capable of predicting Autism Spectrum Disorder using ABIDE dataset. During training the batch size was 25, dropout rate is 0.1 and number of epochs is 2000.

Experimental dataset

ABIDE Dataset : Autism Brain Imaging Data Exchange data is a collection of structural and functional brain image data, contributed by 24 different International Brain Imaging Laboratories. It is the most common dataset used in autism detection. ABIDE dataset can be divided into two phases, ABIDE Ⅰ and ABIDE Ⅱ . ABIDE Ⅰ consist of data of 1112 subjects, where 539 subjects were having ASD and 573 were normal taken from 17 sites, whereas ABIDE Ⅱ 593 normal individuals and 521 with ASD, taken from 19 sites.

Pre-Processing of dataset : The different tools available for pre-processing of FMRI data are CCS, CPAC, DPARSF, NIAK , etc. We have used CPAC (Configuring Pipeline of Analysis of Connectome) with nuisance signal removal, motion alignment , low frequency drift as parameters. FILT_Global i.e. Band Pass Filtering and Global Signal Regression were used as processing strategy. The data that is obtained after processing strategy and registration technique, are further treated by calculating the derivatives based on the region of interest. The different region of interests are namely aal(automated anatomical labeling) , ez(eickhof-zilles), ho, TT, dosenach 160 as well as cc200.
After all the above mentioned pre-processing techniques, the desirable dataset is obtained.

Pre-processing

15 min

ABIDE

Autism Brain Imaging Data Exchange

PREPROCESSED

The Preprocessed Connectomes Project (PCP) is pleased to announce the public release and open sharing of preprocessed neuroimaging data from the Autism Brain Imaging Data Exchange (ABIDE). A consortium of the International Neuroimaging Datasharing Initiative (INDI), ABIDE is a collaboration of 16 international imaging sites that have aggregated and are openly sharing neuroimaging data from **539 individuals suffering from ASD** and **573 typical controls**. These 1112 datasets are composed of structural and resting state functional MRI data along with an extensive array of phenotypic information.

Data from ABIDE was preprocessed by five different teams using their preferred tools. Functional preprocessing was performed using: the Connectome Computation System (CCS), the Configurable Pipeline for the Analysis of Connectomes (CPAC), the Data Processing Assistant for Resting-State fMRI (DPARSF) and the NeuroImaging Analysis Kit. Due to the controversies surrounding bandpass filtering and global signal regression, four different preprocessing strategies were performed with each pipeline: all combinations of with and without filtering and with and without global signal correction. To limit the variance between outputs to just preprocessing, statistical derivatives for each pipeline and strategy were calculated by the CPAC software. Structural preprocessing and calculation of cortical measures was performed using three different pipelines: ANTS, CIVET, and FreeSurfer. Refer to the links at the top for more information about the different pipelines and derivatives.

http://preprocessed-connectomes-project.org/abide/

Pre-processing Technique

15 min

Basic Processing

Step	CCS	C-PAC	DPARSF	NIAK
Drop first "N" volumes	4	0	4	0
Slice timing correction	Yes	Yes	Yes	No
Motion realignment	Yes	Yes	Yes	Yes
Intensity normalization	4D Global mean = 1000	4D Global mean = 1000	No	Non-uniformity correction using median volume

Nuisance Signal Removal

Each pipeline implemented some form of nuisance variable regression[1,2] to clean confounding variation due to physiological processes (heart beat and respiration), head motion, and low frequency scanner drifts, from the fMRI signal.

Regressor	CCS	C-PAC	DPARSF	NIAK
Motion	24-param	24-param	24-param	scrubbing and 1st principal component of 6 motion parameters & their squares
Tissue signals	mean WM and CSF signals	CompCor (5 PCs)	mean WM and CSF signals	mean WM and CSF signals
Motion realignment	Yes	Yes	Yes	Yes
Low-frequency drifts	linear and quadratic trends	linear and quadratic trends	linear and quadratic trends	discrete cosine basis with a 0.01 Hz high-pass cut-off

Processing strategies

Each pipeline was used to calculate four different preprocessing strategies:

Strategy	Band-Pass Filtering	Global Signal Regression
filt_global	Yes	Yes
filt_noglobal	Yes	No
nofilt_global	No	Yes
nofilt_noglobal	No	No

http://preprocessed-connectomes-project.org/abide/

Registration

15 min

Registration

A transform from original to template (MNI152) space was calculated for each dataset from a combination of functional-to-anatomical and anatomical-to-template transforms. The anatomical-to-template transforms were calculated using a two step procedure that involves (one or more) linear transform that is later refined with a very high dimensional non-linear transform. When data are written into template space (typically after the calculation of derivatives, except for NIAK) all transforms are used simultaneously to avoid multiple interpolations.

Registration	CCS	C-PAC	DPARSF	NIAK
Functional to Anatomical	boundary-based rigid body (BBR)	boundary-based rigid body (BBR)	rigid body	rigid body
Anatomical to Standard	FLIRT + FNIRT	ANTs	DARTEL	CIVET

Derivatives

Statistical derivatives (e.g., regional homogeneity) were generated from preprocessed functional data for each of the four processing strategies generated from each of the four processing pipelines. As mentioned earlier, these derivatives were all generated using CPAC. Although the calculation of the derivatives were the same for every pipeline, there were differences in each pipeline as to when each derivative was registered to standard space and when smoothing was applied. In every case the final resolution of the calculated derivatives is 3x3x3 mm^3.

http://preprocessed-connectomes-project.org/abide/

Description of modules of flowchart

30 min

ABIDE Dataset : Autism Brain Imaging Data Exchange data is a collection of structural and functional brain image data, contributed by 24 different International Brain Imaging Laboratories.

Raw Data : The FMRI data is raw and pre-processed. Please find the link for raw data
http://preprocessed-connectomes-project.org/abide/download.html

Pre-Processing : The data is preprocessed and converted into labelled data. Please find the link for pre-processed data.
https://drive.google.com/drive/folders/1AxErI3Tib5ALu0lqGLKRFH2cf3UTEkVX?usp=sharing

Data cleaning : After pre-processing still some field contains NAN values, in-order to remove those values, data cleaning is done.

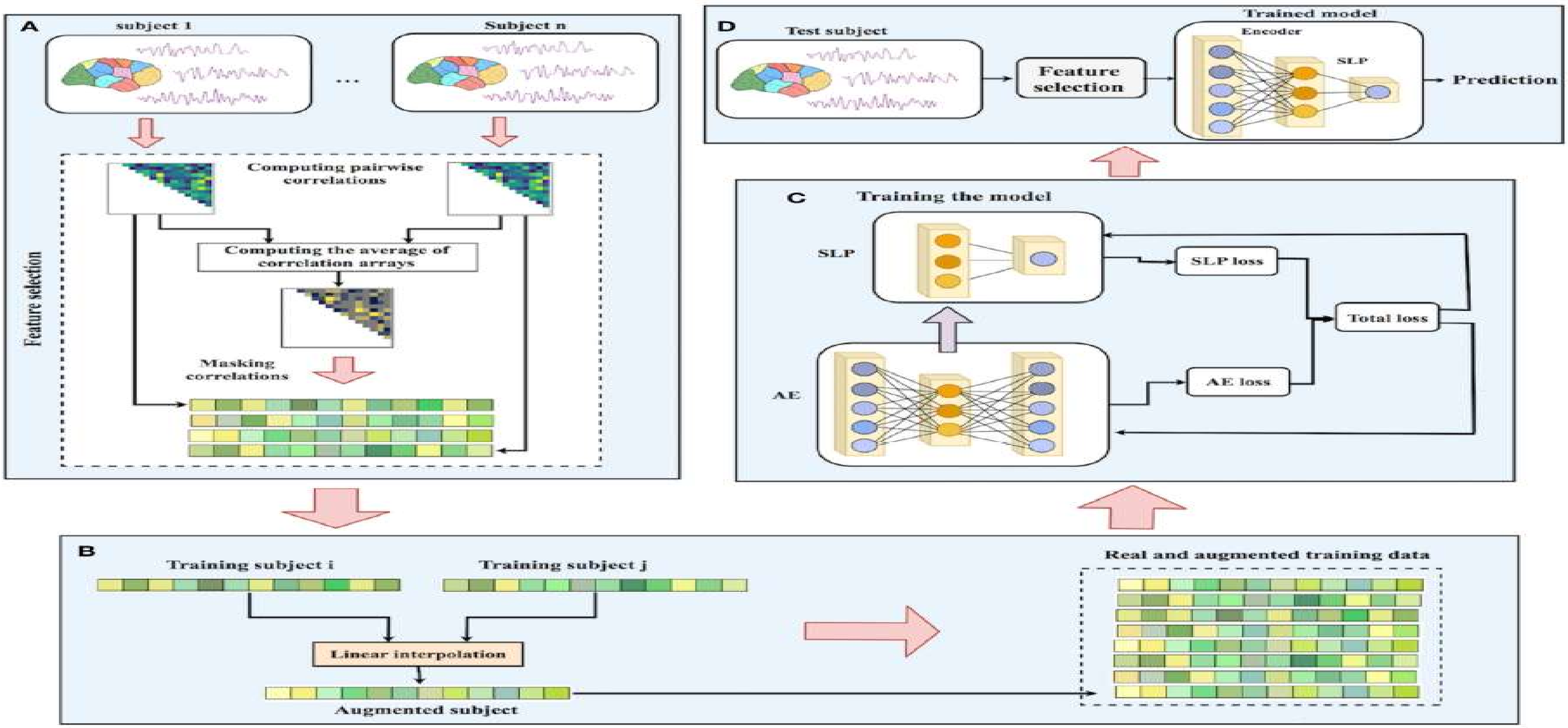
Processed data : After pre-processing and data cleaning , data is obtained that can be used to train the model.

ANN Model : There are six layers in the network including the input and output layer. The input layer has 17 neurons whereas the output layer has 4 neurons. The four hidden layers have 15, 7, 9, 7 neurons respectively. After every level in the network , there is a dropout layer with the dropout rate equal to 0.1.The dropout layer is used to drop the unnecessary features with are irrelevant in the prediction. Relu activation function is used in all the layers except the output layer in which SoftMax is used.

ASD Classification : After training of the model, the subjects are classified whether they have autism or not. The output is obtained in the form of probabilities.

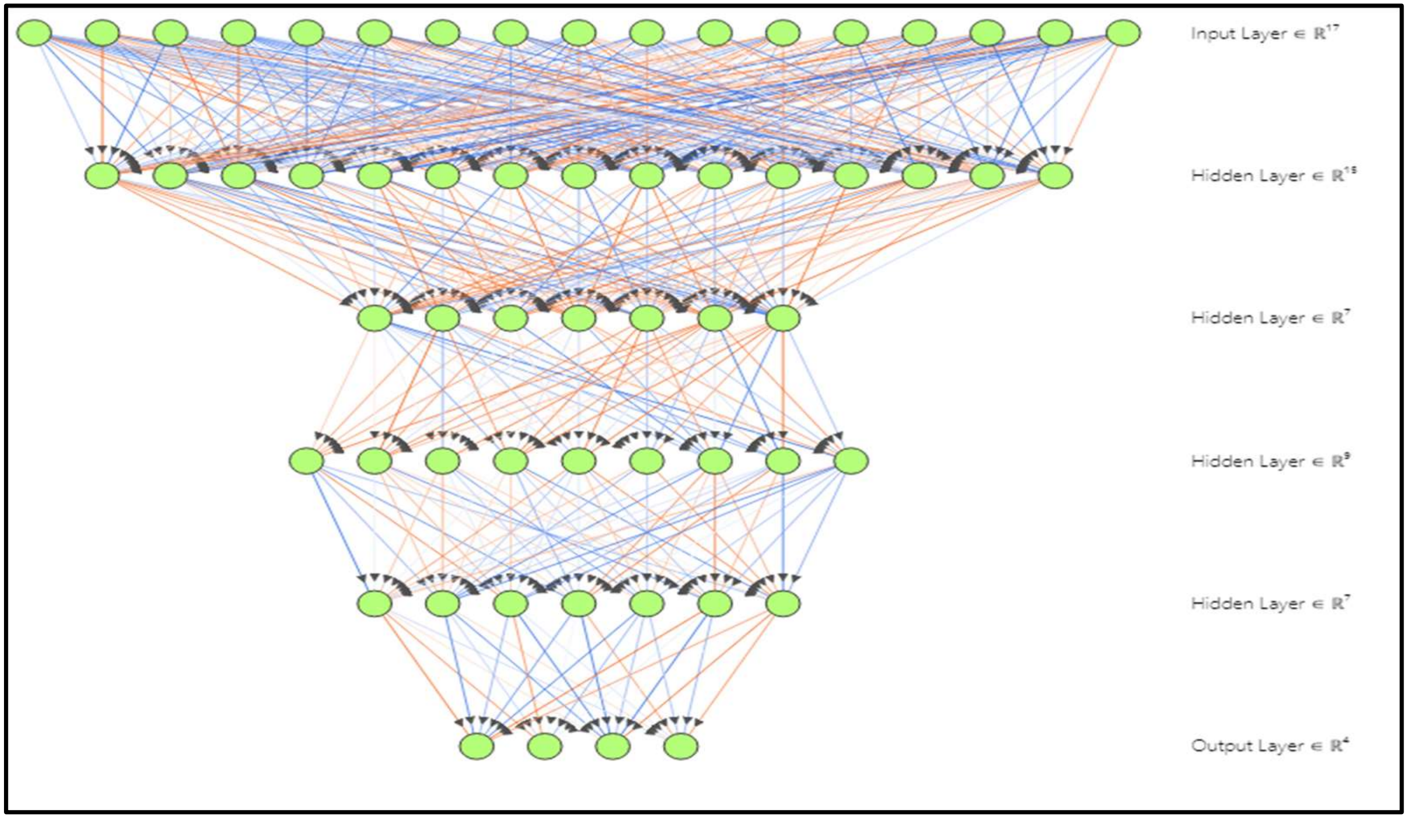
Pre-processing in Depth-

30 min



Taken From – BASE Paper (ASD-DiagNet: A Hybrid Learning Approach for Detection of Autism Spectrum Disorder Using fMRI Data Taban Eslami 1,2, Vahid Mirjalili 3, Alvis Fong1, Angela R. Laird4 and Fahad Saeed2*)

Model Architecture



Language and libraries used for implementation

1. Pandas
2. matplotlib.pyplot
3. NumPy
4. Os
5. functools
6. sklearn.impute
7. Time
8. Pytorch-gpu
9. Keras-gpu
10. Tensorflow-gpu
11. Sys
12. Pickle
13. torch.nn
14. torch.nn.functional
15. sklearn.model_selection
16. torch.optim
17. sklearn.metrics
18. scipy

Features-

Co-Relation Matrix based on Features-

IC_01 - Cerebellum
IC_02 – ACC (Agenesis of corpus callosum) + mpfc (Medial prefrontal cortex)
IC_03 - Caudate
IC_04 - Cerebellum
IC_05 - Calcarine
IC_06 - Calcarine
IC_07 – Precuneus +PCC (The posterior cingulate cortex)
IC_08 - Frontal
IC_09 – IPL (Inferior Parietal Lobule) +AG (angular gyrus)
IC_10 – MTG (Human Middle Temporal Gyrus)
IC_11 - Frontal
IC_12 – SMA (supplementary motor area)
IC_13 - Temporal Pole
IC_14 - Temporal Pole + Fusiform
IC_15 – STG (Superior temporal gyrus)
IC_16 - Middle Occipital
IC_17 - Cerebellum
IC_18 - Cerebellum
IC_20 – MCC (anterior midcingulate cortex in cognitive motor control)
IC_21 - Temporal Pole + Cerebellum
IC_22 - Insula + Caudate
IC_24 – IPL (Inferior Parietal Lobule +Postcentral
IC_26 – Inf +Mid Frontal
IC_28 - Calcarine
IC_29 – MTG (Human Middle Temporal Gyrus)
IC_30 - Inf Frontal

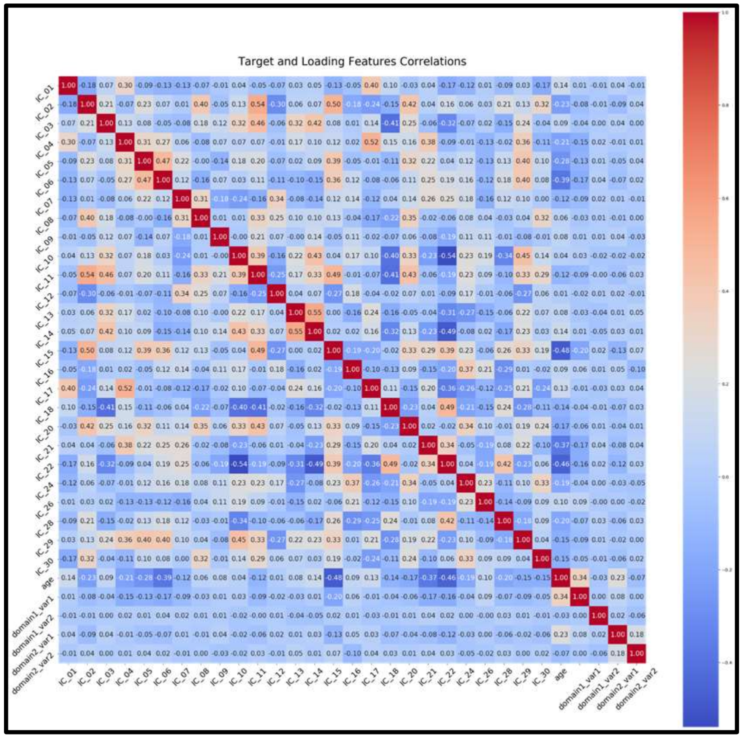
Static Functional Network Connectivity Features-

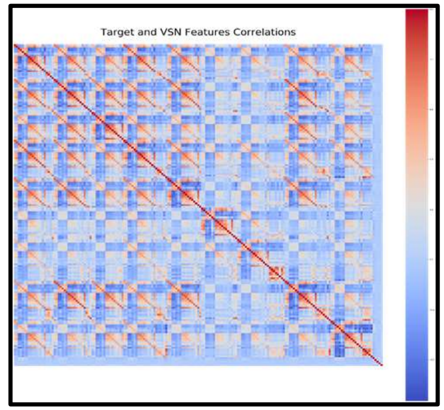
SCN - Sub-cortical Network
ADN - Auditory Network
SMN - Sensorimotor Network
VSN - Visual Network
CON - Cognitive-control Network
DMN - Default-mode Network
CBN - Cerebellar Network

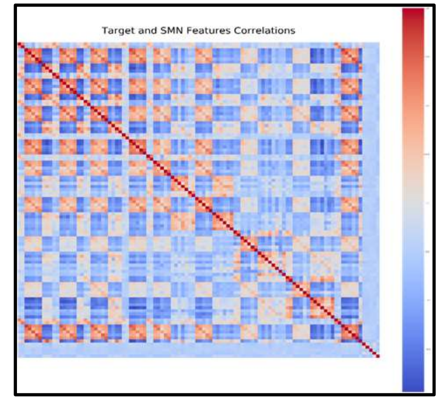
10 min

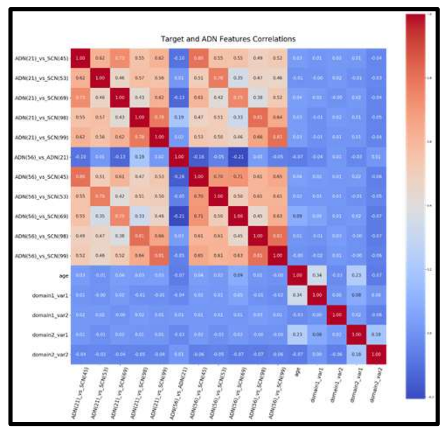
Snapshots (Visualization of Data)-

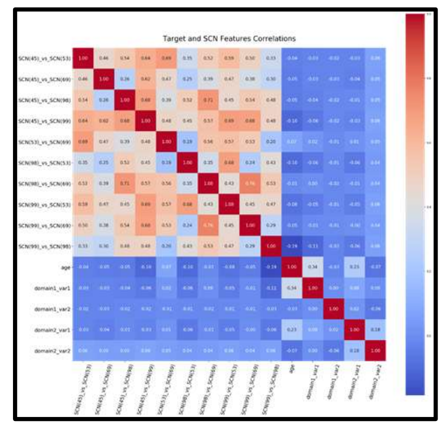
Co-Relation Matrix-





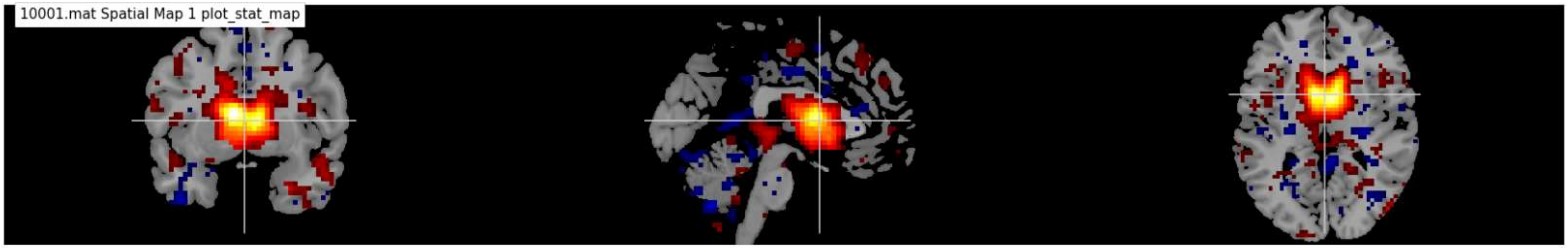


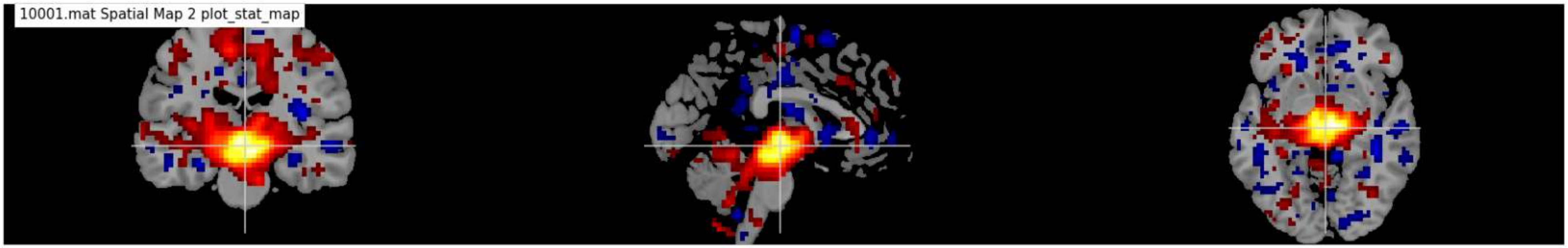




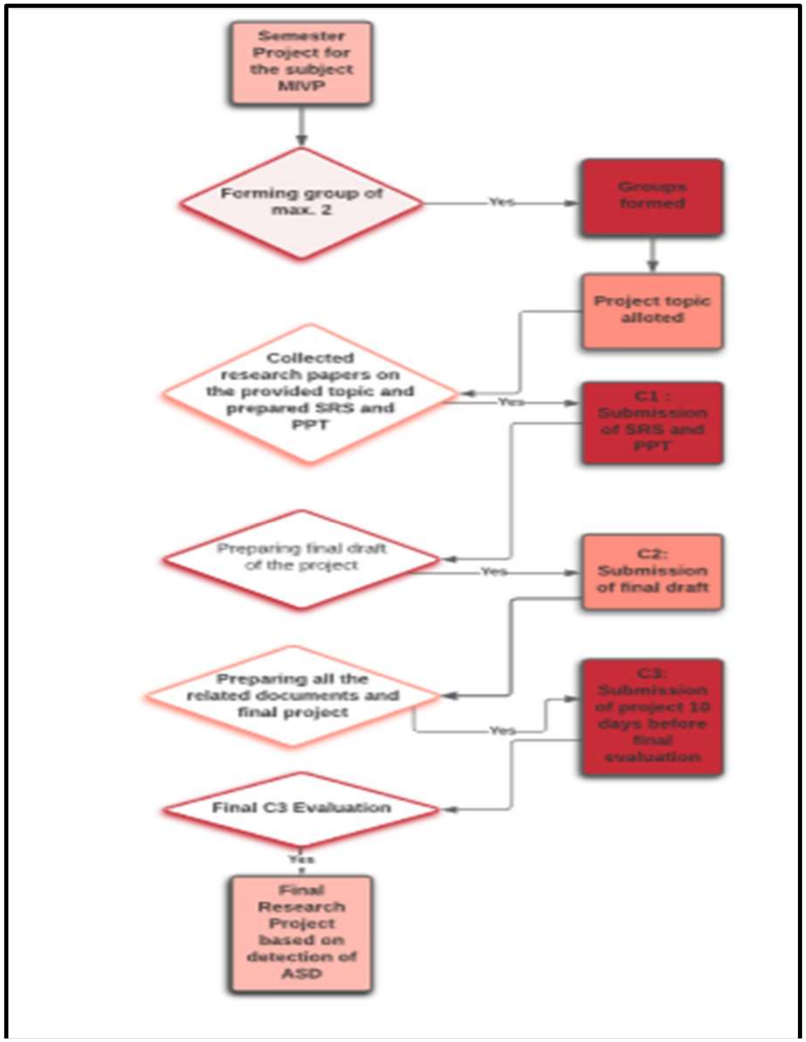
Snapshots-

Visualization Spatial Maps -





Activity Chart



ROC Curve for all the model we have used -

10 min

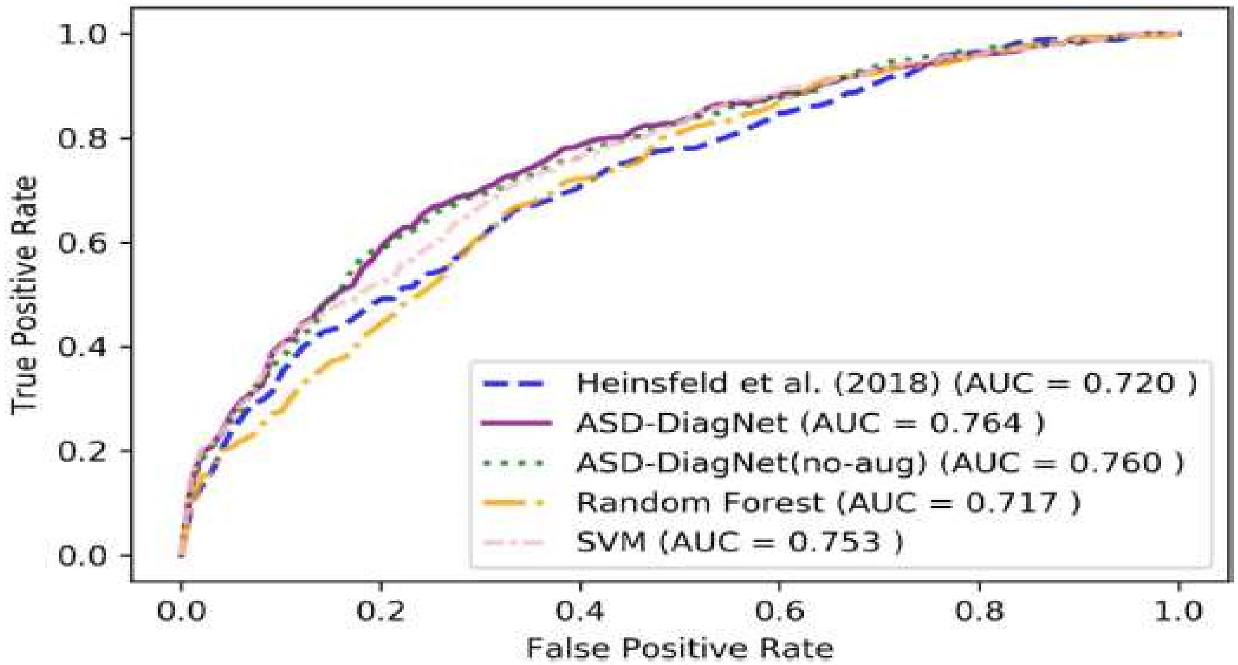


FIGURE 4 | ROC curves of different methods for classification of whole dataset using CC-200 parcellation.

Comparisons with & without Augmentation -

Site	ASD-DiagNet	ASD-DiagNet (no aug.)	Heinsfeld et al., 2018	SVM	Random-Forest
Caltech	52.8	49.9	52.3	46.9	**54.2**
CMU	**68.5**	67.4	45.3	66.6	62.4
KKI	**69.5**	68.6	58.2	66.4	66.6
Leuven	**61.3**	57	51.8	59.8	59.8
MaxMun	48.6	51.4	**54.3**	53.8	49.2
NYU	68	65.1	64.5	**71.4**	61.8
OHSU	**82**	71.9	74	79.4	54.3
Olin	**65.1**	58.8	44	59.5	52.2
Pitt	**67.8**	65.9	59.8	66.3	59.9
SBL	51.6	47.5	46.6	**60**	48.3
SDSU	63	61.3	**63.6**	58.7	62.7
Stanford	**64.2**	53	48.5	51.4	62.1
Trinity	54.1	51.2	**61**	53.1	54.5
UCLA	**73.2**	70.3	57.7	72.1	69.3
USM	68.2	65.1	62	**73.2**	58
UM	63.8	**65.7**	57.6	64.2	64.8
Yale	**63.6**	61.7	53	61.6	55.3
Average	**63.8**	60.7	56.1	62.6	58.6

Bold and color values corresponds to highest accuracy achieved among all datasets.

Method	Accuracy	Sensitivity	Specificity
ASD-DiagNet	**67.5**	63.4	71.5
ASD-DiagNet (no aug.)	64.5	60.9	68
SVM	**67.5**	63.9	70.9
Random forest	65	56.8	72.7
Heinsfeld et al., 2018	63.3	58.6	67.8

Bold values show the highest accuracy among all methods.

Method	Accuracy	Sensitivity	Specificity
ASD-DiagNet	65.3	63.4	66.9
ASD-DiagNet (no aug.)	65.2	61.1	69
SVM	**66.4**	61.6	71
Random forest	65.1	60.3	69.7
Heinsfeld et al., 2018	63.2	59.8	66.4

Bold values show the highest accuracy among all methods.

Analysis -

Results-

	Classifiers	Accuracy score	AUC-ROC Score
0	Random Forest	0.840764	0.735244
1	Logistic Regression	0.668790	0.599839
2	kNN	0.687898	0.619831
3	SVM_rbf	0.796178	0.694598
4	Our Deep Neural Network	0.906212	
5	XGBoost	0.847134	0.747647

Random Forest, KNN, SVM, XGBoost - A comparison of machine learning algorithms for the surveillance of autism spectrum disorder ***Scott H. Lee***
Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques Author - ***Suman Raja, Sarfaraz Masood***.

10 min

References-

[1] ASD-DiagNet: A Hybrid Learning Approach for Detection of Autism Spectrum Disorder Using fMRI Data Taban Eslami 1,2, Vahid Mirjalili 3, Alvis Fong1, Angela R. Laird4 and Fahad Saeed2*

[2] Turner, Jt& Page, Adam & Mohsenin, Tinoosh & Oates, Tim. (2014). Deep Belief Networks used on High Resolution Multichannel Electroencephalography Data for Seizure Detection. 75-81.

[3] Luo Tian-jian, Fan Yachao, Chen Lifei, Guo Gongde, Zhou Changle. (2020). EEG Signal Reconstruction Using a Generative Adversarial Network With Wasserstein Distance and Temporal-Spatial-Frequency Loss. Frontiers in Neuroinformatics, doi: 10.3389/fninf.2020.00015

[4] S. Hwang, K. Hong, G. Son and H. Byun, "EZSL-GAN: EEG-based Zero-Shot Learning approach using a Generative Adversarial Network," 2019 7th International Winter Conference on Brain-Computer Interface (BCI), Gangwon, Korea (South), 2019, pp. 1-4, doi: 10.1109/IWW-BCI.2019.8737322.

[5] R. A. Sree and A. Kavitha, "Vowel classification from imagined speech using sub-band EEG frequencies and deep belief networks," 2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN), Chennai, 2017, pp. 1-4, doi: 10.1109/ICSCN.2017.8085710.

10 min

References-

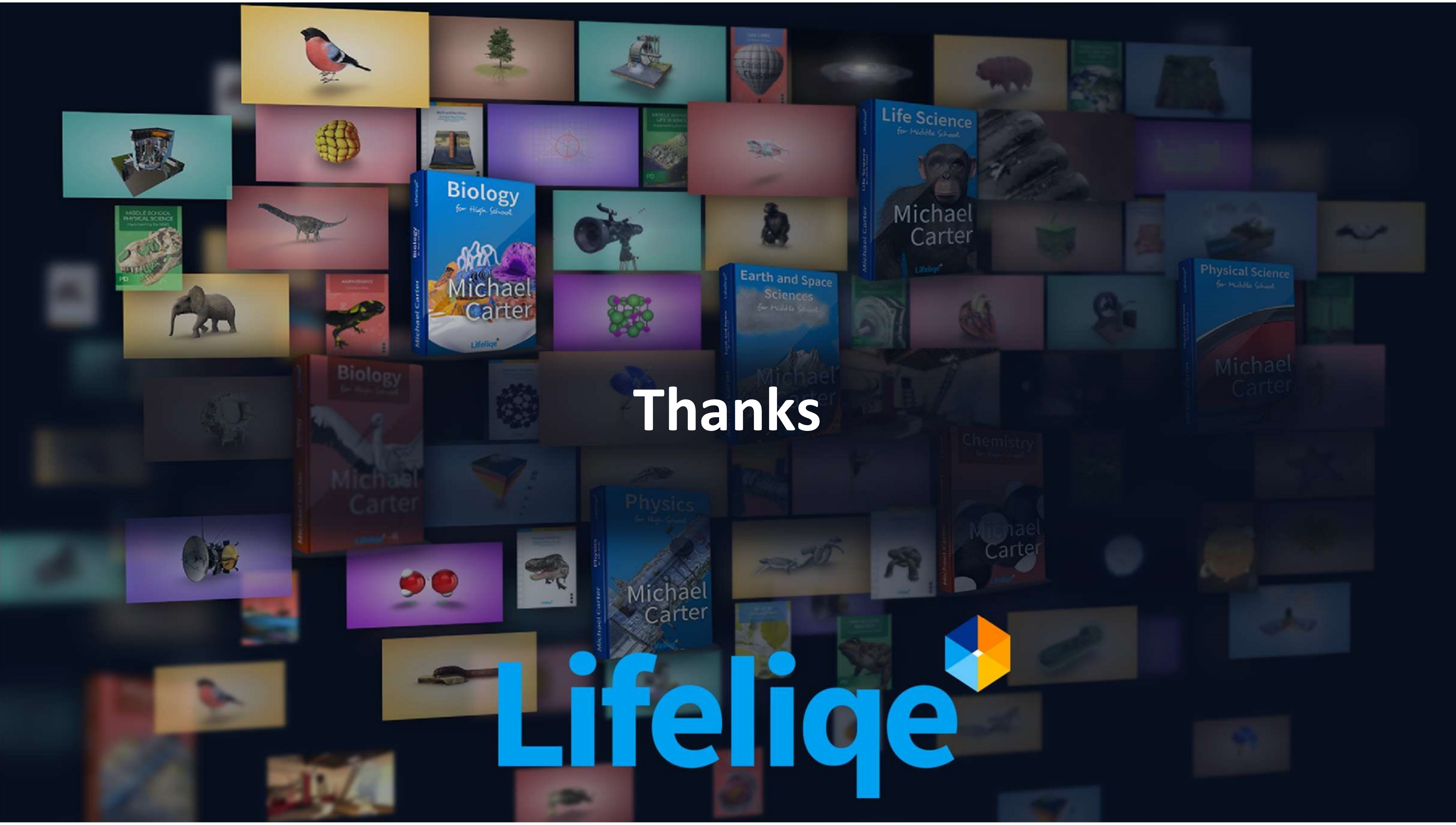
[6] W. Zheng, J. Zhu, Y. Peng and B. Lu, "EEG-based emotion classification using deep belief networks," 2014 IEEE International Conference on Multimedia and Expo (ICME), Chengdu, 2014, pp. 1-6, doi: 10.1109/ICME.2014.6890166.

[7]Al-kaysi A.M., Al-Ani A., Boonstra T.W. (2015) A Multichannel Deep Belief Network for the Classification of EEG Data. In: Arik S., Huang T., Lai W., Liu Q. (eds) Neural Information Processing. ICONIP 2015. Lecture Notes in Computer Science, vol 9492. Springer, Cham. https://doi.org/10.1007/978-3-319-26561-2_5

[8] K. Li, X. Li, Y. Zhang and A. Zhang, "Affective state recognition from EEG with deep belief networks," 2013 IEEE International Conference on Bioinformatics and Biomedicine, Shanghai, 2013, pp. 305-310, doi: 10.1109/BIBM.2013.6732507.

[9] Y. Ren and Y. Wu, "Convolutional deep belief networks for feature extraction of EEG signal," 2014 International Joint Conference on Neural Networks (IJCNN), Beijing, 2014, pp. 2850-2853, doi: 10.1109/IJCNN.2014.6889383.

[10] Hartmann, Kay Gregor et al. "EEG-GAN: Generative adversarial networks for electroencephalographic (EEG) brain signals." ArXiv abs/1806.01875 (2018): n. pag.[11] N. D. Truong, L. Kuhlmann, M. R. Bonyadi, D. Querlioz, L. Zhou and O. Kavehei, "Epileptic Seizure Forecasting With Generative Adversarial Networks," in IEEE Access, vol. 7, pp. 143999-144009, 2019, doi: 10.1109/ACCESS.2019.2944691.


Thanks
Lifeliqe
Biology
for High School
Michael Carter
Life Science
for Middle School
Michael Carter
Earth and Space Sciences
for Middle School
Physical Science
for Middle School
Michael Carter
Physics
for High School
Michael Carter
Chemistry