This work has been accepted in INTERNATIONAL CONFERENCE ON ADVANCED NETWORK TECHNOLOGIES AND INTELLIGENT COMPUTING (ANTIC-2021) for the publication. This copy of paper is for Institute (NIT ROURKELA) documentation.

Autism detection using surface and volumetric morphometric feature of sMRI with Machine learning approach

Mayank Mishra[1[0000-0003-2047-3457]] and Umesh C. Pati[2]

National Institute of Technology Rourkela,
Department of Electronics and Communication Engineering, Rourkela, 769008, India
mmishra1208@gmail.com, ucpati@nitrkl.ac.in

Abstract. Brain imaging has played a very crucial role in the detection of various brain disorders. Among many brain imaging modalities, Magnetic Resonance Imaging (MRI) has proven its importance due to its detailed information regarding the insight of the brain. Autism Spectrum Disorder (ASD) has emerged as a very serious brain disorder due to its late detection among people. It comprises symptoms that are generally ignored, and this creates the urgency for its early detection. This work puts forward the method for the detection of ASD utilizing Machine Learning (ML) with the features extracted from sMRI (Structural Magnetic Resonance Imaging). Surface morphometric and volumetric morphometric features have been utilized for training the machine learning models. The cross-validation approach has been used to avoid overfitting problem occurred during training and testing steps. Machine learning models such as Random Forest (RF), Extra Trees (ET), Linear Support Vector Machine (SVM), Non - Linear SVM, and K- Nearest Neighbors (KNN) have been used for classification between ASD and controls. To evaluate the performance of classification, accuracy, precision, recall, and ROC- AUC score values have been considered.

Keywords: Autism, Brain Imaging, Random Forest, Support Vector Machine, sMRI.

1 Introduction

Brain imaging eases the process for the insight information of the brain with a noninvasive approach. The information can be further utilized for the detection of various brain disorders [1]. Among various brain imaging modalities, MRI holds its uniqueness for its detailed insight and soft tissue information of the brain [2]. Structural brain imaging such as sMRI (Structural Magnetic Resonance Imaging) provides the anatomical information of the brain which can be further utilized for the detection of brain disorders. Among many disorders, ASD has also affected many lives. It is recognized by its symptoms such as weak social communication, repetition in behavior, etc. [3]. Anatomical information of the brain has played a very important role in past

This work has been accepted in INTERNATIONAL CONFERENCE ON ADVANCED NETWORK TECHNOLOGIES AND INTELLIGENT COMPUTING (ANTIC-2021) for the publication. This copy of paper is for Institute (NIT ROURKELA) documentation.

years towards the detection of ASD [4]. The increase in the utilization of machine learning in the field of medical diagnosis has simplified the purpose. Many research works have been done towards the detection of various brain disorders utilizing anatomical information of the brain with machine learning [5].

2 Related Works

Lauren E. Libero et al. have presented a comparative study of cortical surface area, volume, thickness, and gyrification index of the brain of ASD and controls. The study concludes with the alteration observed in the anatomy of the social brain region [6]. Gajendra J. Katuwal et al. have presented the machine learning approach for classification between ASD and controls. The work concludes with high accuracy of the individual site compared to a large heterogeneous dataset [7]. Gajendra J. Katuwal et al. have sub-divided the heterogeneous dataset based on autism severity, VIQ (Verbal IQ), and age. Classification performance has been improved for the subdivision process compared to the whole dataset [8].

Osman Altay et al. presented the work for the prediction of ASD using K- Nearest Neighbors (KNN) and Linear Discriminant Analysis (LDA) classifiers. The dataset utilized for training the model includes a variety of questions encountered by ASD and controls during the diagnosis process [9]. Milan N. Parikh et al. have utilized personal characteristic data for the classification between ASD and controls with machine learning models [10]. Kayleigh K. Hyde et al. presented the survey on the supervised ML approach for the detection of ASD. The survey comprises various aspects of data such as behavioral, brain imaging, developmental, genetic data, etc. [11]. Tania Akter et al. presented the machine learning approach for the early detection of ASD in which dataset has been further sub-categorized based on the age factor such as toddlers, children, adolescents, and adults. The result shows that different machine learning models perform differently when dealing with different sub-categories [12].

Shirajul Islam et al. proposed a machine learning approach for the detection of ASD at an early stage. The dataset has been sub-divided as per the medical, health, and social science criteria. The limitation of this work has been found to be model overfitting [13]. F. Catherine Tamilarasi et al. have utilized the gray level co-occurrence matrix for the feature selection from the thermal face images. Among various machine learning models, SVM performs better in their experiment [14]. Our previous work [19] has been focused on the detection of ASD using surface morphometric features with Decision Trees and Random Forest machine learning models. It also presented the comparative analysis between left and right hemispheric surface morphometric features for these models.

In this work, the machine learning approach has been presented for the detection of ASD utilizing surface morphometric features and volumetric morphometric features.

This work has been accepted in INTERNATIONAL CONFERENCE ON ADVANCED NETWORK TECHNOLOGIES AND INTELLIGENT COMPUTING (ANTIC-2021) for the publication. This copy of document is for Institute (NIT ROURKELA) documentation.

Comparative analysis of Random Forest (RF), Extra Trees (ET), Linear Support Vector Machine (SVM), Non - Linear SVM, and K- Nearest Neighbors (KNN) on the final dataset which includes surface morphometric, as well as volumetric morphometric features towards the ASD detection, has been presented in the further section of this work.

3 The proposed work

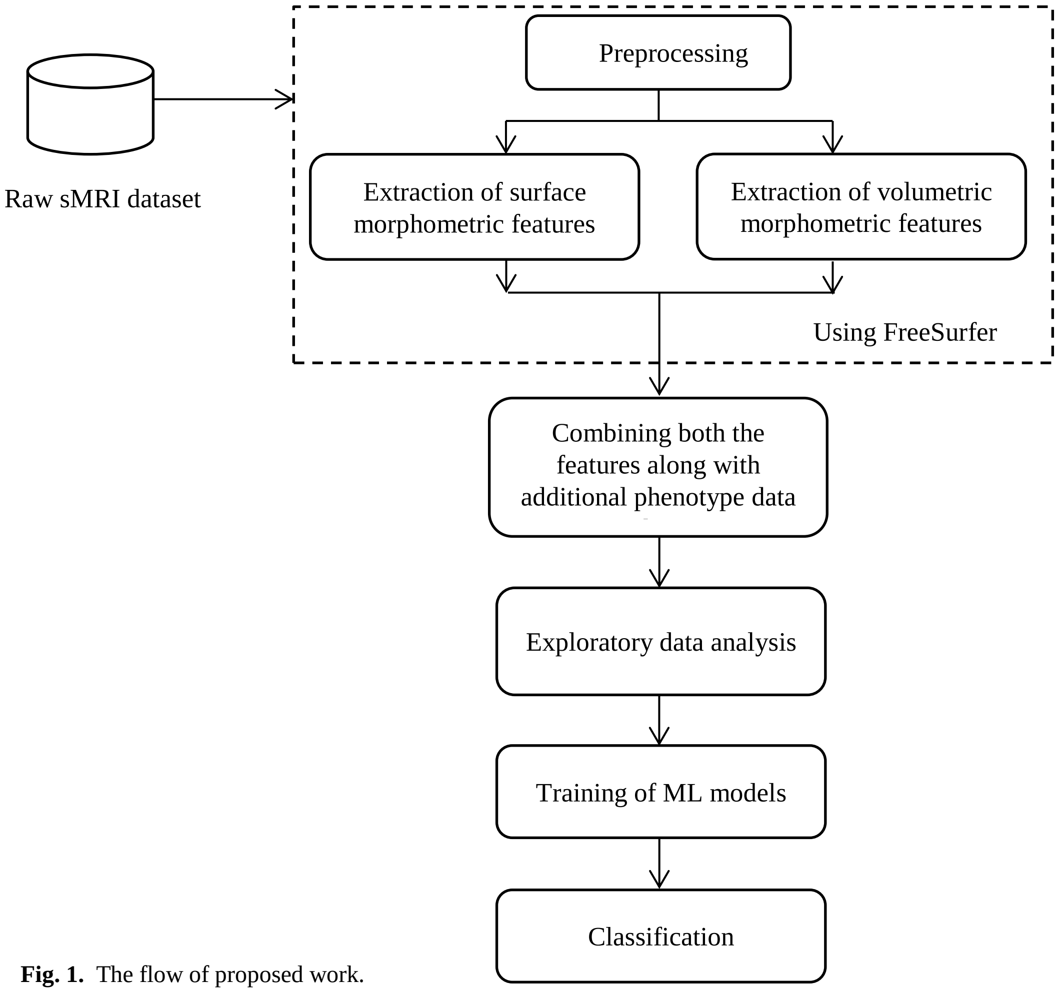

Fig. 1. The flow of proposed work.

The flow of the proposed work follows the pipelines which include the collection of raw MRI datasets, preprocessing, extraction of surface and volumetric morphometric

This work has been accepted in INTERNATIONAL CONFERENCE ON ADVANCED NETWORK TECHNOLOGIES AND INTELLIGENT COMPUTING (ANTIC-2021) for the publication. This copy of paper is for Institute (NIT ROURKELA) documentation.

features, combining of the features, exploratory data analysis, ML model training, and classification.

3.1 Dataset Collection

The ABIDE-1[15] dataset collected from COINS [16] has been utilized for the experimentation. Total 100 T1-weighted sMRI data has been taken for this work, which includes 68 ASD and 32 Controls.

3.2 Preprocessing

Once the dataset is collected it goes through the 'recon-all' pipeline of FreeSurfer [17]. It includes stripping off the skull, normalization of intensity, volumetric labeling, surface parcellation, and volumetric segmentation. The thickness of various surface regions measured using FreeSurfer [17] of Desikan-Killiany Atlas [18] has been considered for surface morphometric features as shown in Fig.2. The volume measured from the segmented regions (such as Left-Lateral-Ventricle, Left-Caudate, Left-Putamen, Left-Hippocampus, etc.) of the sMRI of the brain using FreeSurfer [17] has been considered as volumetric morphometric features. Total 103 features which is the combination of surface as well as volumetric morphometric features have been utilized for the execution of this work. As presented in [7] the performance of the ML model has been improved by adding additional phenotype information to the dataset. The presented approach also utilizes the same approach as mentioned in [7] by combining the additional phenotype data such as age at the time of sMRI scan, VIQ, and performance IQ (PIQ) of the person in the dataset of 103 extracted features.

3.3 Exploratory Data Analysis (EDA)

Once the entire extracted feature along with additional phenotype information is combined, it goes through EDA. It includes finding the missing values in the dataset, removing the incomplete information present in the dataset and lastly finding the Pearson Correlation coefficients [20] between all the features of the dataset. While executing the experiment the threshold of 0.9 (obtained after trial and error approach) has been kept for removing highly correlated features from the dataset.

3.4 Machine Learning Models

For training purposes, Random Forest (total number of decision trees = 150), Extra Trees (total number of decision trees =150), Linear SVM, Non-Linear SVM (degree =2), and KNN (number of neighbors = 6) machine models have been used. All the mentioned parameters have been taken into consideration after the 'trial and error approach'.

This work has been accepted in INTERNATIONAL CONFERENCE ON ADVANCED NETWORK TECHNOLOGIES AND INTELLIGENT COMPUTING (ANTIC-2021) for the publication. This copy of document is for Institute (NIT ROURKELA) documentation.

3.5 Classification

Once the model is trained it is ready for testing purposes. Hence, the classification has been performed for the detection of ASD. To evaluate the classification performance of the models various classification evaluation parameters such as accuracy, precision, recall, and ROC-AUC score have been calculated using the following expressions:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{Recall or True Positive Rate (TPR)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{False Positive Rate (FPR)} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (4)$$

By putting the values of FPR on the x-axis and the values of TPR on the y-axis, ROC is plotted and then the AUC value has been calculated.

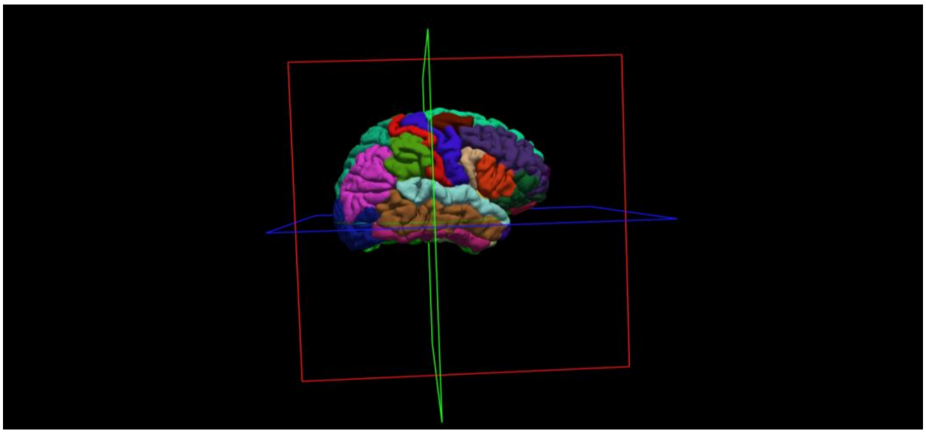

Fig. 2. Surface parcellation using Desikan-Killiany Atlas in FreeSurfer.

This work has been accepted in INTERNATIONAL CONFERENCE ON ADVANCED NETWORK TECHNOLOGIES AND INTELLIGENT COMPUTING (ANTIC-2021) for the publication. This copy of paper is for Institute (NIT ROURKELA) documentation.

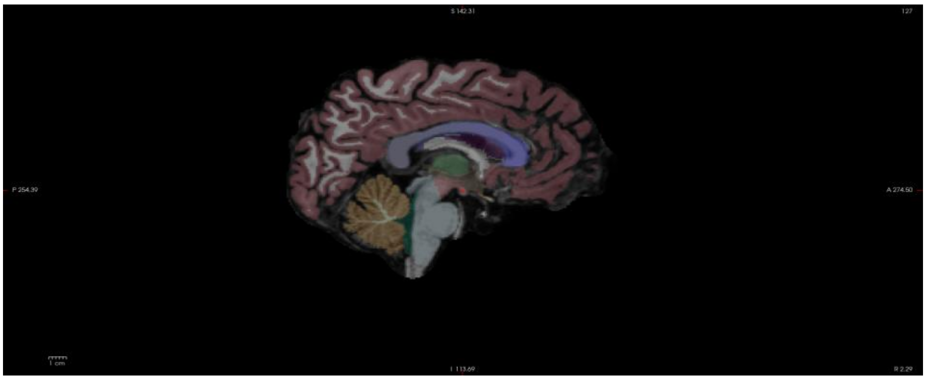

Fig. 3. Sagittal view of volumetric segmentation in FreeSurfer.

4 Results and discussion

Instead of dividing the complete dataset into training and test data set, this approach utilizes the cross-validation method to overcome the "overfitting" problem. Dataset has been split into 'k' folds. For every iteration, 'k-1' folds are treated as a training dataset and the remaining fold has been taken as a test dataset. The number of iteration depends on the number of 'k-fold' used in cross-validation.

For each iteration, the mentioned classification evaluation parameter has been calculated. After the 'k^{th}' iteration, the mean of these parameters has been calculated as shown in Table 1 to Table 5 for every ML model.

Table 1. Classification performance evaluation of Random Forest

k	Accuracy	Precision	Recall	ROC-AUC
5	0.7890	0.8397	0.8800	0.8685
10	0.8000	0.8500	0.8761	0.8700
20	0.8100	0.8733	0.8791	0.8708

Table 2. Classification performance evaluation of Extra Trees

k	Accuracy	Precision	Recall	ROC-AUC
5	0.8000	0.8599	0.8956	0.8740
10	0.8099	0.8721	0.8761	0.8761
20	0.8100	0.8775	0.8791	0.8833

This work has been accepted in INTERNATIONAL CONFERENCE ON ADVANCED NETWORK TECHNOLOGIES AND INTELLIGENT COMPUTING (ANTIC-2021) for the publication. This copy of document is for Institute (NIT ROURKELA) documentation.

Table 3. Classification performance evaluation of Linear- SVM

k	Accuracy	Precision	Recall	ROC-AUC
5	0.7000	0.7399	0.8505	0.7329
10	0.7400	0.7918	0.8666	0.7577
20	0.7300	0.8091	0.8166	0.7416

Table 4. Classification performance evaluation of Non-Linear SVM

k	Accuracy	Precision	Recall	ROC-AUC
5	0.7100	0.8189	0.7318	0.7677
10	0.7100	0.8488	0.7095	0.7803
20	0.7900	0.9174	0.7750	0.8041

Table 5. Classification performance evaluation of KNN

k	Accuracy	Precision	Recall	ROC-AUC
5	0.61	0.7217	0.7098	0.5647
10	0.60	0.7063	0.6952	0.5636
20	0.61	0.7308	0.7125	0.6020

As per the results mentioned from Table 1 to Table 5, the performance of the ML models has been improved as the value of 'k' for the cross-validation increases. Experimental results show the better performance of models at k = 20.

On comparing the performance of ML models at k = 20 as shown in Fig.4, the Extra Trees model performs superior to all the models used for the execution of the experiment. After the Extra Trees model, the Random Forest model also performs well on the various classification evaluation parameters. It has been also observed that Non-linear SVM performs superior to other ML models on the scale of 'precision'.

This work has been accepted in INTERNATIONAL CONFERENCE ON ADVANCED NETWORK TECHNOLOGIES AND INTELLIGENT COMPUTING (ANTIC-2021) for the publication. This copy of paper is for Institute (NIT ROURKELA) documentation.

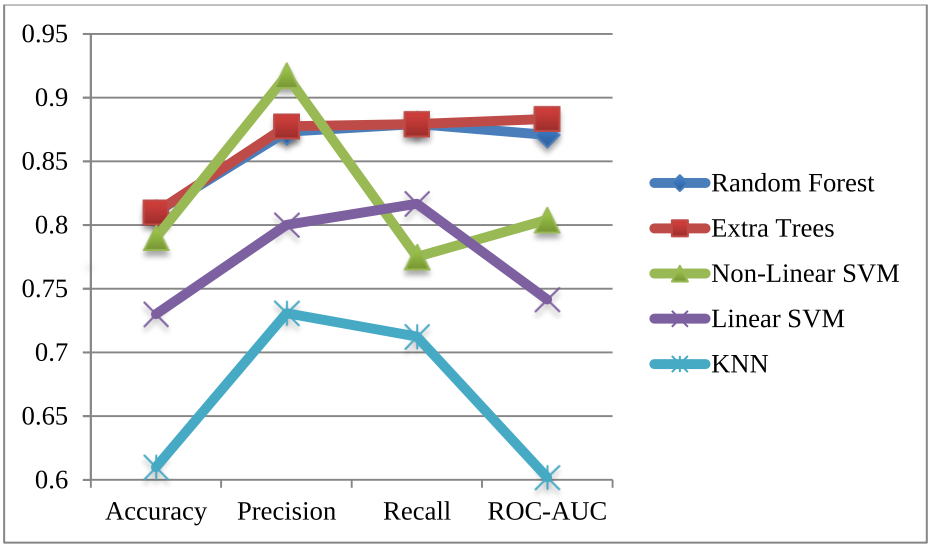

Fig. 4. Graphical comparison between different ML models at k=20 in cross-validation approach.

5 Conclusion

The presented machine learning approach utilizes the volumetric, surface morphometric features along with additional phenotype information for the detection of ASD. To remove the overfitting problem that arises in the smaller dataset, the cross-validation approach has been utilized in the execution of the experiment. It has been observed that the performance of ML models has improved with the increase in the value of 'k' for the cross-validation. Experimental results show the better performance of Extra Tree and Random Forest. Non-Linear SVM also performs well in terms of precision. The limitation of this work is limited dataset utilization towards the execution of the approach. Hence, the results obtained cannot be taken as generalized findings. Future work will be dedicated to overcoming this limitation.

References

1. Bhatele K, Bhadauria.: Brain structural disorders detection and classification approaches: a review. Artificial Intelligence Review 53, 3349- 3401(2020).
2. Möllenhoff K., Oros-Peusquens AM., Shah N.J. Introduction to the Basics of Magnetic Resonance Imaging. In: Gründer G. (eds) Molecular Imaging in the Clinical Neurosciences. Neuromethods, vol 71, pp. 75-98, Humana Press, Totowa, NJ (2012).

This work has been accepted in INTERNATIONAL CONFERENCE ON ADVANCED NETWORK TECHNOLOGIES AND INTELLIGENT COMPUTING (ANTIC-2021) for the publication. This copy of document is for Institute (NIT ROURKELA) documentation.

3. Ecker C, Ginestet C, Feng Y, Johnston P, Lombardo MV, Lai MC, Suckling J, Palaniyappan L, Daly E, Murphy CM, Williams SC, Bullmore ET, Baron-Cohen S, Brammer M, Murphy DG; MRC AIMS Consortium. Brain surface anatomy in adults with autism: the relationship between surface area, cortical thickness, and autistic symptoms. JAMA Psychiatry 70(1), 59-70 (2013).
4. Brašić J. R., Mohamed, M. Human brain imaging of autism spectrum disorders. In Imaging of the human brain in health and disease. Academic Press (2014).
5. Mateos-Pérez J M, Dadar M, Lacalle-Aurioles M, Iturria-Medina Y, Zeighami Y, Evans A C.: Structural neuroimaging as clinical predictor: A review of machine learning applications. NeuroImage: Clinical 20, 506-522 (2018).
6. Libero LE, DeRamus TP, Deshpande HD, Kana RK.: Surface-based morphometry of the cortical architecture of autism spectrum disorders: volume, thickness, area, and gyrification. Neuropsychologia 62, 1-10 (2014).
7. Katuwal G J, Cahill N D, Baum S A, Michael A M.: The predictive power of structural MRI in Autism diagnosis. In: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 4270-4273. Milan, Italy (2015).
8. Katuwal G J., Baum S A, Cahill N D, Michael A M.: Divide and Conquer: Sub-Grouping of ASD Improves ASD Detection Based on Brain Morphometry. PLOS ONE 11(4), 1-24 (2016).
9. Osman A, Mustafa U.: Prediction of the autism spectrum disorder diagnosis with linear discriminant analysis classifier and K-nearest neighbor in children. In: 6th International Symposium on Digital Forensic and Security (ISDFS), pp. 1-4 (2018).
10. Parikh MN, Li H, He L.: Enhancing Diagnosis of Autism with Optimized Machine Learning Models and Personal Characteristic Data. Front Computational Neuroscience 13(9), 1-5 (2019).
11. Hyde, K., Novack, M.N., LaHaye, N. et al.: Applications of Supervised Machine Learning in Autism Spectrum Disorder Research: a Review. Rev J Autism Dev Disord 6, 128–146 (2019).
12. Akter T, Satu S. et al.: Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders. IEEE Access 7, 166509-166527 (2019).
13. Islam S, Akter T. et al.: Autism Spectrum Disorder Detection in Toddlers for Early Diagnosis Using Machine Learning. In: 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), pp. 1-6. Gold Coast, Australia (2020).
14. Tamilarasi F C, Shanmugarn J.: Evaluation of Autism Classification Using Machine Learning Techniques. In 3rd International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 757-761 (2020).
15. Di Martino A, Yan C-G, Li Q, Denio E, Castellanos FX, Alaerts K. et al.: The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular Psychiatry 19, 659–667 (2014).
16. Scott A, Courtney W et al.: COINS: An innovative informatics and neuroimaging tool suite built for large heterogeneous datasets. Frontiers in Neuroinformatics 5 (2011).
17. Fischl B.: FreeSurfer. Neuroimage 62(2), 774–781 (2012).
18. Winkler A M, Kochunov Peter. et al.: Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. NeuroImage 53(3), 1135-1146 (2010).

This work has been accepted in INTERNATIONAL CONFERENCE ON ADVANCED NETWORK TECHNOLOGIES AND INTELLIGENT COMPUTING (ANTIC-2021) for the publication. This copy of paper is for Institute (NIT ROURKELA) documentation.

19. Mishra M, Pati U C.: Autism Spectrum Disorder Detection using Surface Morphometric Feature of sMRI in Machine Learning. In: 8th International Conference on Smart Computing and Communications (ICSCC), pp. 17- 20 (2021).
20. Zhou H, Deng Z. et al.: A new sampling method in particle filter based on Pearson correlation coefficient. Neuro Computing 216, 208-215 (2016).