Diagnose Schizophrenia through Ginni-Index along With Random Forest Classification on Multi-Modal Brain Magnetic Resonance Imaging

Dr. Balaji Rajendra Bombade, Avinash Hanwate, Shivaji Jadhav

Department of Computer Science & Engineering, SGGS IE&T, Nanded, MS, India

brbombade@gmail.com, avinash.hanwate@gmail.com

shivaji.d.jadhav@gmail.com

Abstract— Schizophrenia, as one of serious mental issue, hugely affects patients' lives by influencing how they think and abilities to control-command-execute the routine activities. Basic indications incorporate false convictions, hazy or lost reasoning, hearing voices lessened social engagement, and crazy articulation, and an absence of motivation. Schizophrenia is an extreme and unhealthful psychological illness that has no entrenched, non-intrusive finding biomarker. As of now, because of its side effect cover with other psychological sicknesses (like bipolar issue) it must be analysed subjectively, by procedure of elimination. Understanding Schizophrenia is most critical through non-intrusive process. As the Image processing field progressed, application of its on Schizophrenia was the quest. Therefore, our aim was to accept the challenge and hence we used Kaggle 2014 challenge database (data set is generated using MRI scanning) of Schizophrenia patients and applied Ginni-Index, Random Forest algorithm. Two modalities of magnetic resonance imaging examines are used to accumulate these highlights: functional and structural magnetic resonance imaging. Classification is done using Support Vector Machine (SVM) and Gaussian Kernel. The resultant output is with higher accuracy rate of AUC 0.923.

Keywords—Ginni-Index, Random Forest Classification, Support Vector Machine, MRI, FNC, SBM, Schizophrenia.

I. INTRODUCTION

Schizophrenia, one of serious mental issue, hugely affects patients' lives by influencing how them think and behave [1]. The diagnosis of schizophrenia is still to a great extent reliant on subjects and concentrated on taking out the side effects, which are typically inconspicuous and undefined from other psychological instabilities, for example, bipolar disorder. A few causes are considered by the researchers for schizophrenia. One noteworthy reason for this illness is anomalous mind structure. With the advancement of current electromagnetic methods, brain magnetic resonance imaging information are broadly utilized as a part of clinical and subjective neuroscience to help specialists to analyse different cerebrum issue. Information obtained from MRI procedures (Basic and Functional) turn out to be greatly alluring to scientists in the course of recent years [2].

We used MLSP 2014 Schizophrenia Classification Kaggle Challenge data set that includes data like functional network connectivity (FNC) [3-4] and source-based morphometry (SBM). Both, databases can be further refined from mind maps that are gotten from MRI utilizing independent component analysis (ICA) [5]. FNC are time astute relationship esteems that measure the general association between independent brain mappings. SBM are weights of brain maps obtained from the application of ICA on the dim issue part of the mind, where flag transduction happens and shapes the computational capacity of a cerebrum. Both informational indexes are viewed as significant to schizophrenia discovery, which could improve the expectation exactness of programmed diagnostics if legitimately utilized.

The manifestation of schizophrenia can keep going for a considerable length of time or be deep rooted with different inabilities. Around 1.1 percent of US grown-up populace is accounted for to have a year commonness of this issue malady. Up until this point, this mind issue sickness can't be cured in light of the fact that the reasons for it are still to a great extent obscure. Two noteworthy angles are considered as supporters of the danger of creating schizophrenia. One is qualities and condition and the other one is mind science and structure.

It is realized that schizophrenia occurs inside families yet no hereditary data is found yet. Researchers are as yet attempting to make sense of key qualities that reason the turmoil. Analysts likewise believe that some intricate responses of mind chemicals may assume a part in schizophrenia advancement. All the more imperatively, the mind structure and movement might be somewhat/altogether changed amid the confusion advance that can be identified by neuroimaging and attractive reverberation imaging system [7].

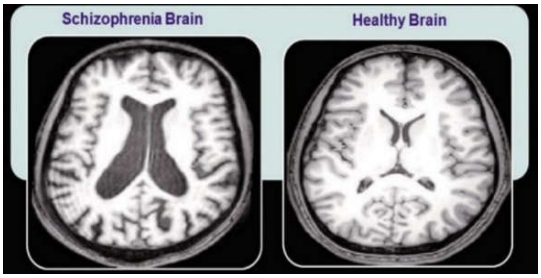

Fig1. Difference between schizophrenia patient's brain and healthy patient's brain.

Past work simply consolidated the structural MRI and functional MRI together as potential features in this arrangement challenge. Notwithstanding, these are two unique estimations on a similar mind neuron network. Basically join these two informational collections may bring about a multi linearity portrayal of our cerebrum picture network. Not quite the same as past work on this issue, we will likely research all the more painstakingly about the sMRI and fMRI informational collections with a specific end goal to comprehend the connection between them by unsupervised machine learning approaches for a superior comprehension and speaking to for features of brains. Based on the information of the chose features, administered machine learning strategies including logistics regression and SVM are connected for the order undertaking.

We aim at classifying healthy people from the people suffering from schizophrenia and schizoaffective disorder. Received information from the magnetic resonance imaging (MRI) was used and analysed properly, which is widely applied in understanding human brains in neuroscience [8]. Supervised machine learning analysis- Classification contains following methods:

A. Support Vector Machine:

Small data sets classification and regression analysis is done using most powerful supervised learning models like Support Vector Machine (SVM). It looks for a projection of features into a certain hyper plane. Within this hyper plane, we can successfully linearly classify all training vectors into classes. For a given training data set, the SVM algorithm is able to generate a model that assigns new data points into one of the classes. As and when applied, hyper plane creates maximum margin from both classes. The points lying on the boundaries of the hyper plane are called support vectors. SVM is able to perform non-linear classification using the kernel techniques, which maps the input into high-dimensional spaces.

B. Neural Network:

Neural Networks (NNs) [08] is a model inspired by biological neural networks and it is a very significant method for the supervised process since NNs can learn from the previous experience/examples and get better output. Suppose we have several input neurons (features), the input neurons will connect to other interconnected neurons (hidden layers) so that they can exchange messages between each other. In order to assure that the model learns from previous experience, we add numeric regression weights on the connection which can be tuned based on experiences and allow the system to learn from the data. Then, by keeping updating the weights, we are able to obtain the optimal output. The network function can be described as,

$$f(x) = \mathrm{h}\left(\sum_i \left(w_i g_i\ (x)\right)\right),$$

Where the h (.) is the sigmoid function for this two-class classification problem, and g (.) can be different activation functions for NNs. In other words, the NNs initially give us a random guess about what it might be, then calculate the equation above and see how far the answer was from the actual one so that the model is able to adjust the connected weight appropriately and ultimately generates an optimal outcome. It is said that NNs can approximate the data well given wisely chosen weights and enough hidden layers.

NNs is very efficient in our case since the advantages of NNs are to deal with large numbers of features and diversity of the data. However, the way that NNs handle the relationship between variables is still not easy for us to understand. Here our input neurons are the features and our output neuron is whether the person has Schizophrenia or not. By repeated learning and cross-validation, the results are improved and the accuracy of our prediction increases as well.

II. PROPOSED METHOD AND ITS IMPLEMENTATION DETAILS

The proposed approach is specially introduced for handwritten Kannada characters. The Kannada script has most of the characters curvy in structure. So the feature extraction method focuses on curvy features of the character images. The complete approach is shown in the Figure 2. The original images are pre-processed in its first step to make it ready for feature extraction.

In order to experiment the proposed method, we used Kaggle 2014 MLSP Challenge data set (data set is generated using MRI scanning's of patents aged between 18 up to 45). Total test set includes 119748 unknown subjects. Initially, we had taken 46 records of terminally ill patients and crossed matched it with 40 peoples with sound brain condition.

The proposed method includes application of random vector into the features set and further calculating mean of Ginni-Index. The variation/increase/decrease in the Ginni-Index helps to know the patients exact condition. But, this is not the end. Further, we apply, Random Forest Classification and Gaussian Kernel with Support Vector Machine to classify the dataset. Based on the provide dataset, we decided to use below feature set:

FNC224, SBM map67, FNC 295, SBM map61,
FNC226, FNC183, FNC33, FNC302,
FNC220, FNC243, FNC37, FNC289,
SBM map36, and others.

This is shown in Fig.2 with random forest plot.

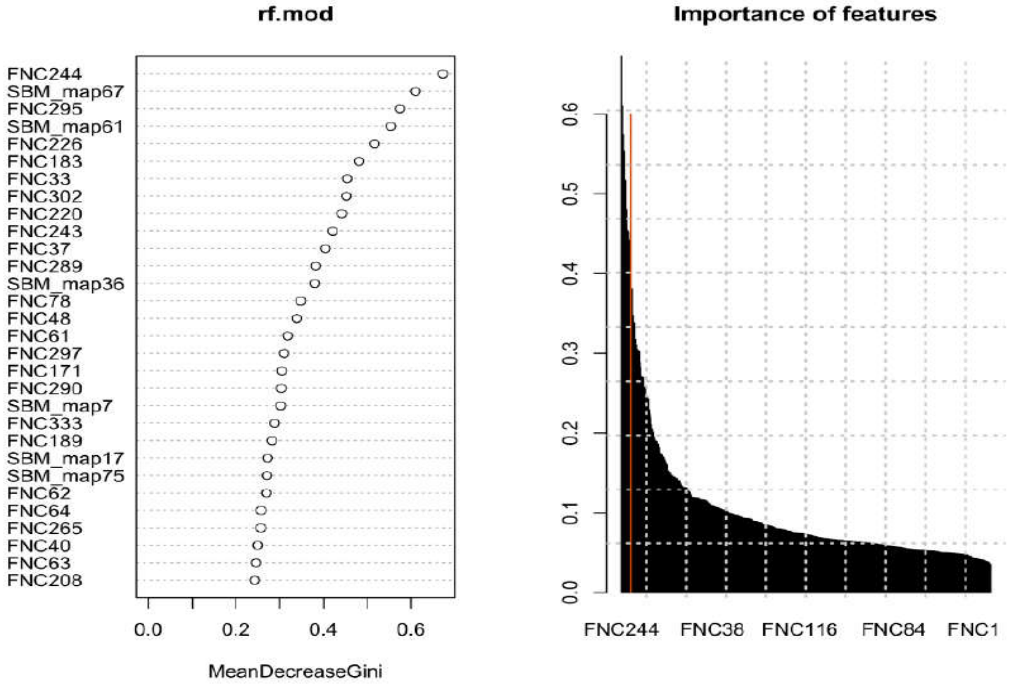

Fig 2. Random Forest Plot of Ginni-Index Important Features.

The resultant value is shown in Figure 3.

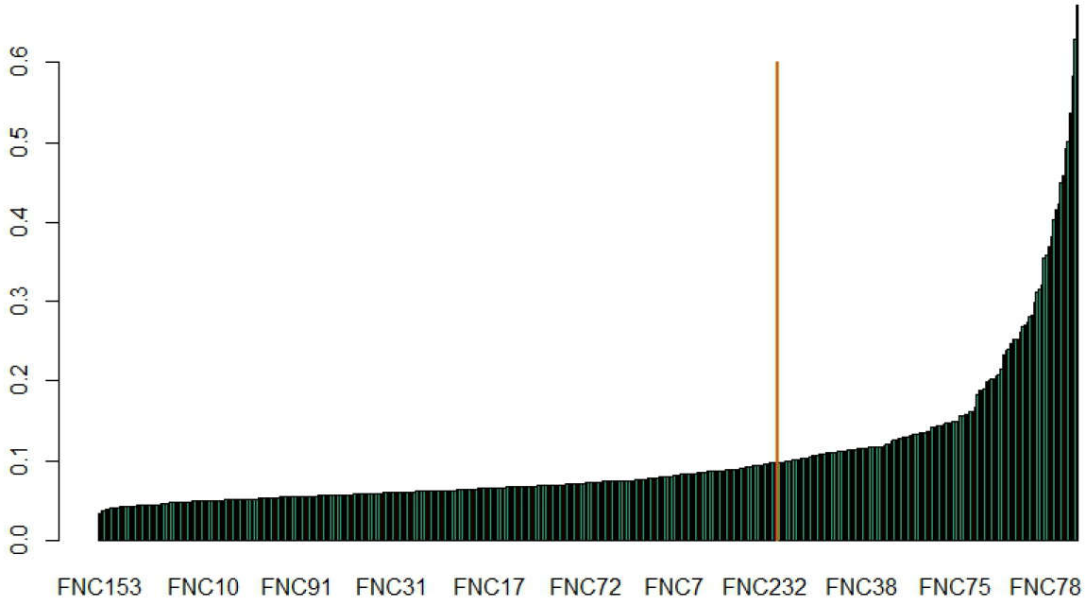

Fig 3: Feature Plot of Important selected features as mentioned above.

Post classification, it is clear that, finding the Schizophrenia becomes easier through scanned MRI images and using above novel technique. The resultant output is reduced dataset with 0.923 AUC.

III. RESULT AND DISCUSSION

The original dataset contained 410 features (32 for SBM and 378 for FNC). After the feature trimming, we ended up with 122 variables. Next, we estimated hyper parameter s (sigma, width parameter) for the Gaussian Radial Basis kernel and tuned C parameter (the penalty factor, controlling trade-off between model complexity and proportion of non-separable instances) using leave-one-out cross validation for the final SVM classifier. The resulted test set area under the receiver operating characteristic curve (AUC) was 0.923. Of note, cross-validated performance of various models that had been tested (RF, boosted trees, neural network, SVM) varied around 0.8 and 0.85 (for overall accuracy) and the public scores that we were receiving after the submissions were unstable. Therefore, we decided not to implement more complex solutions and stopped on a relatively simple model. In general, it was somewhat difficult to evaluate performance of the models, since the mismatch between cross-validated accuracies and the feedback that we were receiving during submissions was very large (with private AUC scores varying around 0.65). Meanwhile, additional feature selection (e.g., recursive feature elimination approaches) and classifier resembling could potentially result in a superior performance. This is kept for feature work.

REFERENCES

[1] *M. R. Arbabshirani, K. Kiehl, G. Pearlson et al., "Classification of schizophrenia patients based on resting-state functional network connectivity," Frontiers in Neuroscience, vol. 7, 2013.*

[2] *Van de Ven, V.G., et al., Functional connectivity as revealed by spatial independent component analysis of fMRI measurements during rest. Hum Brain Map, 2004. 22(3): p. 165-78.*

[3] *Richard, Jonas, et al., Machine learning with Brain Graphics: Predictive Modelling Approaches for Functional Imaging in Systems Neuroscience. IEEE Signal Processing Magazine, 2013, vol.30, no.3, p. 58-70.*

[4] *Kaggle, MLSP 2014 Schizophrenia Classification Challenge, 2014.*

[5] National Institute of Mental Health, Schizophrenia, 2016.
[6] V. Kecman, Learning and Soft Computing, Support Vector machines, Neural Networks and Fuzzy Logic Models, MA: The MIT Press, Cambridge, 2001.
[7] A. V. Lebedev, "The 10th Annual MLSP Competition: Second Place," in IEEE MLSP, Reims, France, 2014, In Press.
[8] Y. J. Li, F. Y. Fan," Classification of schizophrenia and depression by EEG with ANNs", IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, China, September 1-4, 2005.