2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Multi-modal deep learning of functional and structural neuroimaging and genomic data to predict mental illness

Md Abdur Rahaman[1, 2], Jiayu Chen[2], Zening Fu[2], Noah Lewis[1, 2], Armin Iraji[2], Vince D. Calhoun[1, 2]

[1]Georgia Institute of Technology, Atlanta, GA, USA
[2]Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State University, Georgia Institute of Technology, Emory University, Atlanta, GA, USA

Abstract— **Neuropsychiatric disorders such as schizophrenia are very heterogeneous in nature and typically diagnosed using self-reported symptoms. This makes it difficult to pose a confident prediction on the cases and does not provide insight into the underlying neural and biological mechanisms of these disorders. Combining neuroimaging and genomic data with a multi-modal 'predictome' paves the way for biologically informed markers and may improve prediction reliability. With that, we develop a multi-modal deep learning framework by fusing data from different modalities to capture the interaction between the latent features and evaluate their complementary information in characterizing schizophrenia. Our deep model uses structural MRI, functional MRI, and genome-wide polymorphism data to perform the classification task. It includes a multi-layer feed-forward network, an encoder, and a long short-term memory (LSTM) unit with attention to learn the latent features and adopt a joint training scheme capturing synergies between the modalities. The hybrid network also uses different regularizers for addressing the inherent overfitting and modality-specific bias in the multi-modal setup. Next, we run the network through a saliency model to analyze the learned features. Integrating modalities enhances the performance of the classifier, and our framework acquired 88% ($P < 0.0001$) accuracy on a dataset of 437 subjects. The trimodal accuracy is comparable to the state-of-the-art performance on a data collection of this size and outperforms the unimodal and bimodal baselines we compared. Model introspection was used to expose the salient neural features and genes/biological pathways associated with schizophrenia. To our best knowledge, this is the first approach that fuses genomic information with structural and functional MRI biomarkers for predicting schizophrenia. We believe this type of modality blending can better explain the disorder's dynamics by adding cross-modal prospects.**

Clinical Relevance— **This study combinedly learns imaging and genomic features for the classification of schizophrenia. The data fusion scheme extracts modality interactions, and the saliency experiments report multiple functional and structural networks closely connected to the disorder.**

I. Introduction

Schizophrenia (SZ) is a severe neuropsychiatric disorder that manifests with various biological and neural alterations [1]. In general, SZ identification is symptom-based and primarily dependent on the self-acknowledged behavioral syndromes of a subject. As such, the diagnosis gets intricated by the heterogeneity in the patient's psychological state, lack of reliable information from the subject, and dichotomy in clinical trials [2-4]. The complication extends further because the outcomes and effects of this health trait overlap with other psychiatric and neurodegenerative disorders [5]. Neuroimaging can aid schizophrenia research by providing the opportunities of building biomarkers from mechanistic investigations of the structural and functional profile of the healthy and diseased brain. However, generating reliable biomarkers is an extensive research avenue in neuroscience. Currently, genomic data has been showing great potential in predicting schizophrenia[6, 7]. Besides, studies reported a strong association between the genetic risk factor of schizophrenia and brain functional and structural changes [8, 9]. Also, Chen et al. suggest shared genetic risk in schizophrenia and gray matter reduction [10]. Intuitively, compounding genetic and neuroimaging information is promising and might complement each other to characterize schizophrenia.

Deep learning (DL) based studies have reported significant improvement in schizophrenia classification accuracy over the years [11-16]. However, besides the high accuracy, the models provide minimal insights into the associated neural components, effects, and reasoning. Multi-modal deep learning potentially enhances the robustness of inferences made from a DL model for a learning task that can explain distinct aspects of a system under investigation [17, 18]. Qureshi et al. used multi-modal integration of structural magnetic resonance image (sMRI) and functional resonance image (fMRI) for enhancing the accuracy of schizophrenia and healthy control (HC) classification [19]. The study used a relatively small sample (SZ: 72, HC: 72) and applied the ML model on global connectivity features from specific brain regions. More studies performed multi-modal SZ classification, which is mostly confined to imaging modalities only [20-22]. Here, we use sMRI, resting fMRI and, genome-wide single nucleotide polymorphism (SNP) data in a deep learning framework to characterize the disorder. We use independent component analysis (ICA) [23, 24] to decompose fMRI and sMRI scans into independent component features, which are further analyzed by a DL model with SNPs for learning the shared feature subspace. Modality fusion improves the classification accuracy, and the saliency model extracts highly discriminative features for characterizing schizophrenia. We observe a set of structural subdomains, including caudate, anterior cingulate & medial prefrontal, inferior and mid frontal gyrus, and calcarine, that have contributed strongly to the classification. Also, the saliency experiment reports several static functional network connectivity (sFNC) pairs that show substantial association in identifying the disorder. Moreover, the SNPs extracted from genomics data highlight well-documented SZ-related genes and pathways.

II. Data Preprocessing

A. Structural MRI

We preprocessed the sMRI data using statistical parametric mapping (SPM12, http://www.fil.ion.ucl.ac.uk/spm/) under MATLAB 2019 environment for all three datasets. We segmented the structural images into gray matter, white matter, and CSF using modulated normalization algorithm, resulting in outputs as gray matter volume (GMV). Next, the GMV was smoothed using a Gaussian kernel with a full width at half maximum (FWHM) = 6 mm. More details about the preprocessing steps of sMRI and fMRI images are described in prior work [25].

B. Functional MRI Specifications

The fMRI data were preprocessed using the SPM12 toolbox too. The first five time points of the fMRI scans were discarded to guarantee that the magnetization achieves a steady state. A rigid body motion correction was performed to correct subject head motion, followed by the slice-timing correction to account for timing difference in slice acquisition. The fMRI data were subsequently warped into the standard Montreal Neurological Institute (MNI) space using an echo-planar imaging (EPI) template and were slightly resampled to 3 × 3 ×3 mm^3 isotropic voxels. The resampled fMRI images were finally smoothed using a Gaussian kernel with a full width at half maximum (FWHM) = 6 mm.

C. Genomics

The datasets were quality controlled and imputed separately following the pipeline described in our previous work [24][26]. Then three datasets were merged and went through standard Plink [27] preprocessing. After preprocessing, the data were pruned (linkage disequilibrium (LD)) at $r^2 < 0.9$. Finally, we conducted feature selection based on the PGC SZ GWAS [28], such that the analysis used 1280 SNPs residing in the 108 risk loci and showing SZ associations with p < 1e-4. SNPs were retained with population structure corrected using principal component analysis. The details about genetic preprocessing are provided in this study [10].

III. Multi-modal Architecture

Our architecture consists of two major submodules i) decomposition of MRI images and genotyping of genomics sequence ii) deep neural networks (DNN) to learn multi-modal features for SZ classification. Figure 1 presents different parts of our framework. The DNN incorporates three subnetworks to learn the modality features, followed by multiple fully connected layers for the prediction.

A. Group ICA (using Neuromark pipeline) and genotyping

In this study, we applied a fully automated ICA-based pipeline called Neuromark to extract functional and structural networks [29]. Neuromark is a robust analysis pipeline that can capture image features while retaining more individual-level variability. For fMRI, it uses two large healthy controls datasets, the human connectome project (HCP) (http://www.humanconnectomeproject.org/) and the genomics superstruct project (https://www.nitrc.org/projects/gspdata) to construct reproducible spatial priors. Then based on the network priors, we applied fully automated spatially constrained ICA using the NeuroMark approach to each subject from FBIRN [30], COBRE [31], and MPRC [32] dataset to estimate individual-level functional networks and their time courses. Fifty-three intrinsic connectivity networks (ICNs) covering the whole brain were arranged into 7 functional domains [25, 33]: sub-cortical (SC), auditory (AD), sensorimotor (SM), visual (VS), cognitive-control (CC), default-mode (DM), and cerebellar domain (CB). More information about the spatial maps and the domains are included in this study [25]. We calculated the Pearson correlation between the time course (TCs) of ICNs, resulting in a 53 x 53 sFNC matrix for each subject. We vectorize the sFNC matrix using the upper diagonal entries as the functional features in our fMRI subnetwork (encoder). Similarly, for sMRI, the Neuromark pipeline first extracts reliable ICs that are replicated across independent datasets. 6500 subjects from this study [34] were randomly divided into two groups for estimating the spatial priors. Then the priors were used as a reference to extract individual level structural networks and corresponding loadings. Here, 30 structural networks were estimated by the infomax algorithm, and this procedure was

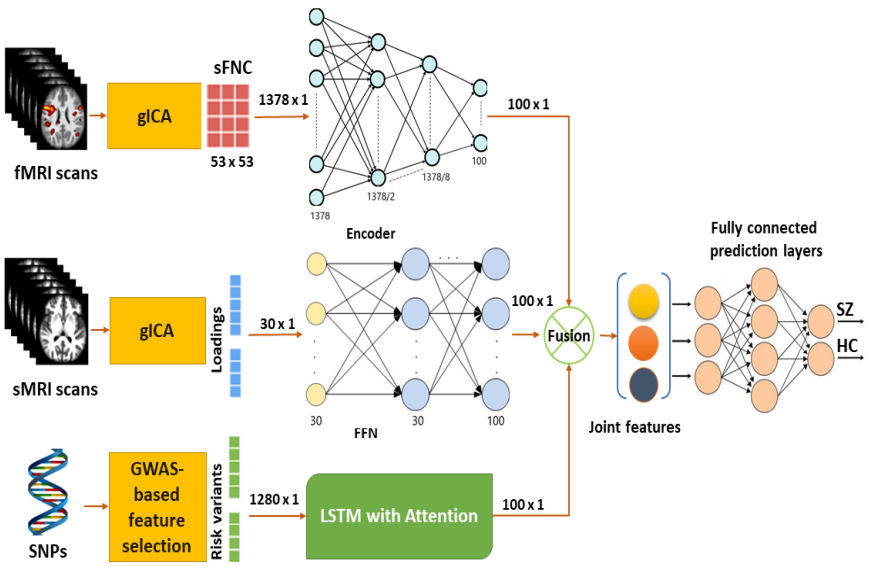

Figure 1. Our multi-modal architecture. It has two submodules, i) group independent component analysis (gICA) and genotyping ii) Deep neural networks (DNN) for learning the modality features. The submodule i) runs two separate gICA for sMRI and fMRI with distinct settings e.g., number of ICs expected. Then it computes the static functional network connectivity (sFNC) matrix for each subject from fMRI gICA and collects the ICA loading matrix from sMRI. Then, the genotyping on the genomics sequence generates SNPs. The DNN submodule consists of four subnetworks. sFNC features are extracted and learned by a compression network (encoder). The next subnetwork is a multilayer feed forward network (FFN) for learning the ICA loading features. For genomics, we use a bi-directional long short-term memory (LSTM) unit with attention mechanism. After that, a fusion mechanism is applied to combine the latent features from three modalities. Finally, the joint features are sent through a series of fully connected layers followed by SoftMax prediction layer. The multimodal losses are back propagated across the DNN submodule using standard optimization technique .

repeated 20 times in ICASSO, in which the best run was selected to ensure the estimation stability. We used the loadings of the structural networks as the sMRI features in the classification. The third modality is based on a set of SNPs generated from genotyping data.

B. Deep learning neural networks

- We use an encoder network consisting of five fully connected (FC) layers to learn low-level features from the sFNC matrix. The weights are initialized uniformly using the Xavier initialization function under the hood of Pytorch [35]. Each layer is followed by a dropout layer and sets the probability to 0.2. The subnetwork used ReLU [36] activation function.
- The sMRI subnetwork is a feed-forward network (FFN) with five FC layers. The FC layers mostly have 30 hidden nodes followed by a dropout layer with a probability of 0.2 except the last layer. The last layer has100 nodes to match up with the latent feature size from other modalities.
- The genetic subnetwork is a bi-directional long short-term memory (bi-LSTM) unit incorporated with the attention mechanism [37]. Although the SNPs data are not temporally dependent, the neighboring SNPs are more spatially allied than the distant ones. We apply light pruning on the SNP data and use bi-LSTM to capture these localized semantics which might help differentiate the disorder.
- Prediction model: The final subnetwork stack of three fully connected (FC) layers is followed by a SoftMax layer. The FC layers have 100 hidden nodes, and the dropout is set to 0.2. We determined these parameters by running random search hyperparameter optimization and nested cross-validation.

C. Fusion

We implemented multiple data fusion techniques to integrate the latent features from the three subnetworks we described earlier, e.g., early, joint, or late fusion of modalities for multi-modal learning [38]. Early fusion refers to integrating multiple input modalities into a standard feature vector [39], late fusion refers to the aggregation of predictions from various models [40, 41]. The late fusion technique is recommended when signals from different modalities do not complement each other, imposing the modality-specific subnetwork to contribute separately to the prediction. We run late fusion for comparison only. We incorporate joint fusion or intermediate fusion that combines the learned feature representations from intermediate layers of modality-specific subnetworks. We aggregate the latent features from sMRI, fMRI, and SNPs as input to the final subnetwork. This helps us observing the relation/interaction between features from multiple modalities in the post hoc analysis. We use weighted concatenation for combining the latent features. The aggregation is performed following this equation,

$$u = [w_1 h_1(t_1), w_2 h_2(t_2), \ldots, w_m h_m(t_m)] \quad (1)$$

where h is a modality-specific neural network. We can describe it as $h_m : \mathbb{R}^{dm} \rightarrow \mathbb{R}^d$ (m represents the modalities, $m = 1, ..., M$). The joint feature u is defined as $u \in \mathbb{R}^d$. The w's are the learnable scaler for each modality. If the final prediction value is p,

$$p = y(u) \quad (2)$$

where y is a neural network and is described as y: $\mathbb{R}$d → $\mathbb{R}$c and c is the number of classes.

D. Multi-modal training

Our implementation uses a cross-entropy loss and *Adam* [42] optimizer with a learning rate of $1\times e^{-4}$ and a batch size of 36 for training from Pytorch. The Adam optimizer updates the weights in the proposed DNN by back-propagating the cross-entropy loss computed from the final layer's prediction and the subject's label. The fully connected layers are followed by the dropout [42] layer with p = 0.2 to reduce overfitting. We also added early stopping for balancing between training and validation loss, which eventually regularizes the model. The data set is divided into training, testing, and validation set. For the validation dataset, we held-out 50 (HC: 25, SZ: 25) subject's data. Next, we run 300 training epochs and use the 10-fold cross-validation [43] method to evaluate the performance of the classifier on the remaining data. The cross-validation method randomly divides the datasets into 10 equal partitions and uses 9 of them to train the model and the remaining one for testing the performance. The technique interchanges the training and testing set and repeats the process ten times. We implement one multi-modal regularization technique for the experimental sanity focused on removing bias by maximizing functional entropies [44]. However, it does not significantly improve overall performance but showed a lower accuracy (Table I).

IV. Results

A. Experiments and Evaluation

We selected 437 subjects from 3 different datasets, COBRE [31], fBIRN [30], and MRPC [32]. The combined dataset has 275 healthy control (HC) and 162 schizophrenia (SZ) subjects. Due to heterogeneity in data preprocessing and lack of common subjects used in available studies, we generate several baselines to compare the performance of our proposed multi-modal classification

Table I. Our experiments and comparisons with baselines

#	Fusion Type	Experiment	Accuracy(%)
1	Unimodal: Functional MRI	Multi-layer feed-forward network	81
2	Unimodal: Structural MRI	Multi-layer feed-forward network	78
3	Unimodal: Genetic data (SNPs)	bi-LSTM with Attention	68
4	Bimodal (mid)	sMRI + fMRI	83
5	Bimodal (mid)	SNPs + sMRI	78
6	Bimodal (mid)	SNPs + fMRI	81
7	Late fusion	fMRI + sMRI + SNPs	78
8	Functional entropies [44]	fMRI + sMRI + SNPs	72
9	**Mid fusion**	**fMRI + sMRI + SNPs**	**88***

framework. Also, compared the performance for several data fusion techniques, including joint fusion, late fusion, additive, multiplicative, and a weighted combination of

latent features. For the saliency experiment on the fully trained model, we use the *Captum* library [45]. Saliency is defined as $S^c(x) = \left|\frac{\partial Y^c}{\partial x}\right|$, or the gradients of the prediction of the correct class with respect to the input. It is a widely used method for interpreting neural networks and has been shown to be an effective measure under scrutiny [46]. Saliency maps are essentially heat maps of each input sample, indicating the relevancy of each feature to the sample's true class.

B. Evaluation

The performance of the model is evaluated on the True positive rate (TPR), false-positive rate (FPR), and accuracy (ACC). These terms are defined as,

- *TPR = TP/(TP+FN)*
- *FPR = FP/(TN+FP),* and
- *ACC = (TP+TN)/(TP+FP+FN+TN).*

Here, *TP* stands for true positives – no. of SZ subjects classified as SZ; *FN* stands for false negatives: no. of SZ classified as HC, *TN* (true negatives): no. of HC subjects classified as HC, *FP* (false positives): no. of HC classified as SZ. Table I shows the accuracy from different models we tested on our dataset. Figure 2 present the Receiver operating characteristic (ROC) curves for the models we studied. We observe our multi-modal neural network achieved an AUC of 0.85 and an average accuracy of 88% across all the testing folds and outperforms other methods.

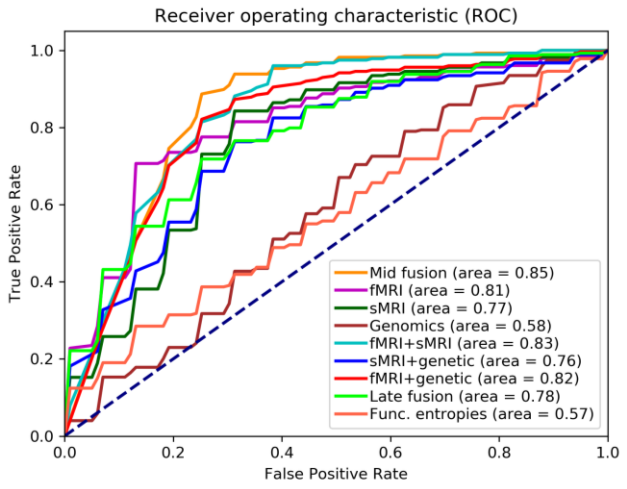

Figure 2. Receiver operating characteristic (ROC) curves and AUC (Area under the ROC curve) for all the models we mentioned in Table I. Our multi-modal joint fusion model outperforms all other baselines.

C. Feature interpretation using saliency model.

For consistency, we calculated the mean of the saliency values across the subjects and the folds. From the average saliency scores for sMRI features, we observe some of the components are showing significantly higher saliency than others. We selected the best five out of thirty components for further investigation. The components are i). Caudate, ii). Anterior cingulate (ACC) and medial frontal cortices (mpFC), iii). Inferior and mid frontal gyrus, iv). Precuneus and posterior cingulate cortex (PCC), and v). Calcarine. Figure 3 shows the mean loading value of these components across HC and SZ subgroups. The differences are tested using a two-sample t-test for statistical significance and controlled for age, sex, and site. Component *i, ii, iii,* and *v* show statistically significant HC-SZ difference at a 1% significance level and denoted by asterisk

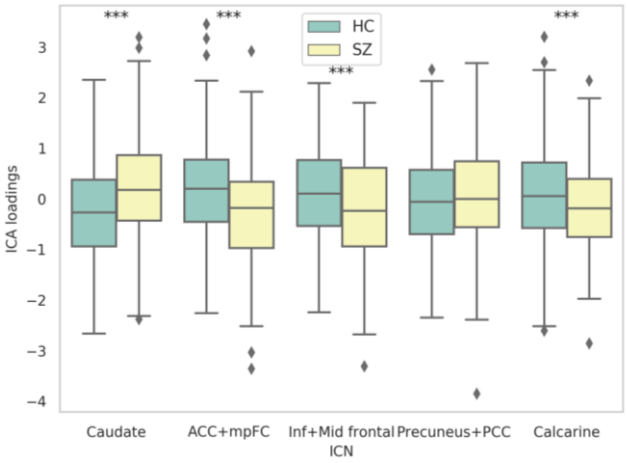

Figure 3. The best five salient sMRI components. The boxplot shows the max, min, and median loading values in the SZ and HC groups. We run a two-sample t-test to test the statistical significance of HC-SZ group differences. Four components (asterisks) show significant differences at a level of 1%. The components are Caudate, Anterior Cingulate (ACC) and medial prefrontal (mpFC) Cortics, Inferior and mid frontal gyrus, and Calcarine

sign in figure 3. HC subjects have a stronger representation (loading) of components *ii, iii,* and *v* than SZ but the weaker representation of caudate than SZ. This might have resulted from the absence or inadequate presence of these structural components (*ii, iii,* and *v*) in SZ subjects. Similarly, the caudate seems to be having weaker activation in the structural MRI of HC subjects. Figure 4 illustrates the mean saliency score of each sFNC pair. We print the mean saliency score for each pair of components in an sFNC matrix for the fMRI features. Figure 4 demonstrates the sFNC matrix annotated with saliency. Value at each cell represents how saliently the pair contributed to the prediction. We can observe from the left subfigure that the most salient pairs are links to the

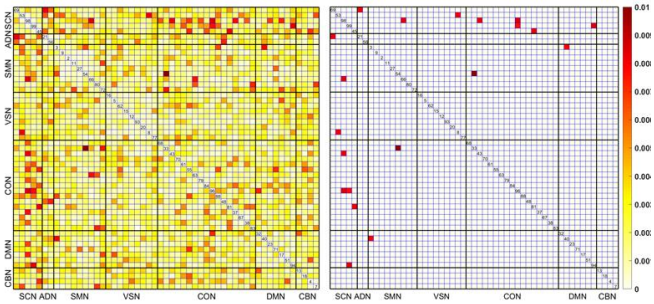

Figure 4. The saliency of fMRI features – a pair of static FNC. We plot the mean saliency value (across the subjects) in the sFNC matrix. The left one illustrates saliency for all the sFNC pairs (1378), and the right one represents the 10 best salient pairs. We observe the salient pairs mainly include components from the subcortical region; thus, the predominant connectivity's are between subcortical and other domains, e.g., visual (VSN), cognitive control (CON), default mode (DMN), etc.

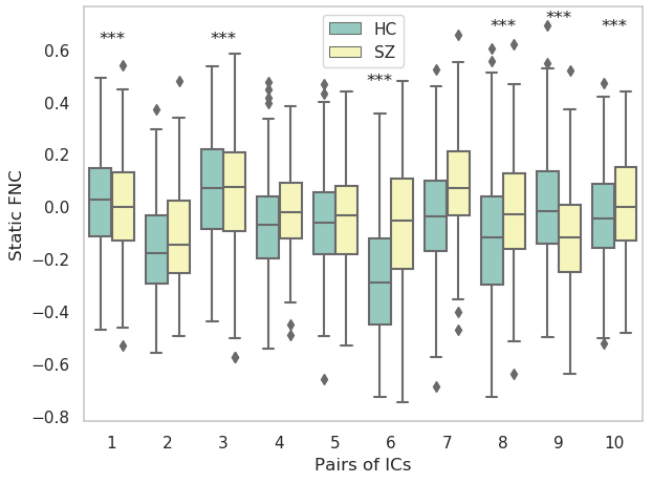

Figure 5. The boxplot demonstrates the max, min, and median of static functional connectivity strength in the 10 most salient sFNC pairs. The pairs are sorted in terms of their saliency score. We run a two-sample t-test to test the statistical significance of HC-SZ group differences. Pair 1, 3, 6, 8, 9, 10 shown to be statistically significant at a level of 1% (asterisks).

subcortical region as shown in the right subfigure. These are the 10 most salient pairs (the connection between two components). We observe from Table II that most of the pair includes at least one component from subcortical networks. So, the highly contributing connections are between subcortical and other regions of the brain. Moreover, the static functional network connectivity strength in those pairs is consistently higher in the SZ subjects than HC. In figure 5, we are showing the mean connectivity strength in HC and SZ

Table II: Salient pairs of components and neuro domains

Pair ID	Comp I	Dom I	Comp II	Dom II	Direction of HC - SZ difference
1	Paracentral lobule	SMN	Insula	CON	HC > SZ
2	Putamen	SCN	Superior frontal gyrus	CON	HC < SZ
3	Putamen	SCN	Precentral gyrus	SMN	HC < SZ
4	Caudate	SCN	Posterior cingulate cortex	DMN	HC < SZ
5	Putamen	SCN	Superior medial frontal gyrus	CON	HC < SZ
6	Caudate	SCN	Superior temporal gyrus	ADN	HC < SZ
7	Thalamus	SCN	Left inferior parietal lobule	CON	HC < SZ
8	Hypothalamus	SCN	Lingual gyrus	VSN	HC < SZ
9	Post central gyrus	SMN	Precuneus	DMN	HC > SZ
10	Caudate	SCN	Superior frontal gyrus	DMN	HC < SZ

subgroups. Six pairs appear to be statistically substantial at a significance level of 1%. From the saliency map computed on SNPs we select SNPs showing higher importance scores across at least 6 folds of cross-validation (total folds = 10). The higher importance score indicates a saliency value greater than the mean saliency of a particular fold. The salient SNPs are annotated to 92 unique genes, and we conducted pathway analysis on these genes using David functional analysis tool [47]. We noted that there were two clusters of genes involved in the dopaminergic synapse ($p = 0.03$) and postsynaptic density ($p = 0.008$) respectively, both have been well documented for SZ [48, 49]. The mentionable involved genes in the two significant pathways are the GRIN2A - glutamate receptor contributing to the slow phase of excitatory postsynaptic current, long-term synaptic potentiation, and learning by similarity. CACNA1C is related to calcium signaling. GRM3 is also a glutamate receptor; CHRNA3 is the cholinergic receptor associated with schizophrenia [50-52].

V. Conclusion

We present a multi-modal classification framework for sMRI, fMRI, and SNPs data. Our model achieved 0.85 ROC with an accuracy of 88% in classifying schizophrenia which outperforms contemporary DNN models. Another primary goal of our study is to investigate the learned features while characterizing the disorder. The saliency maps on the subject-level features help understand the dynamics, effects, and unique aspects of the illness, mainly, how the diseased neural system differs from a healthy system. Experimental results reported a subset of sMRI, fMRI features exhibiting substantial HC/SZ group differences across distinct regions of the brain. Moreover, integrating genetic information compliment the learning from another modality. We observe the unimodal performance of SNPs is lower while blending with other modalities enhances the overall performance. In addition, two identified pathways are associated with SZ. The genes involved seem to regulate the learning process. These analytics could be useful in providing additional clues regarding the neural mechanisms underlying schizophrenia.

VI. Compliance with Ethical Standards

Informed consent was obtained from each participant prior to scanning in accordance with the Internal Review Boards of corresponding institutions.

References

[1] Barry, J., *A Deep Learning Approach to Diagnosing Schizophrenia.* 2019.

[2] Aboraya, A., et al., *The reliability of psychiatric diagnosis revisited: The clinician's guide to improve the reliability of psychiatric diagnosis.* Psychiatry (Edgmont), 2006. **3**(1): p. 41.

[3] Ward, C., et al., *The psychiatric nomenclature: Reasons for diagnostic disagreement.* Archives of General Psychiatry, 1962. **7**(3): p. 198-205.

[4] Yassin, W., et al., *Machine-learning classification using neuroimaging data in schizophrenia, autism, ultra-high risk and first-episode psychosis.* Translational psychiatry, 2020. **10**(1): p. 1-11.

[5] Frisoni, G.B., et al., *The clinical use of structural MRI in Alzheimer disease.* Nature Reviews Neurology, 2010. **6**(2): p. 67-77.

[6] KLASİFİYE, Ş.H.G.V.K., *CLASSIFICATION OF SCHIZOPHRENIA PATIENTS BY USING GENOMIC DATA: A DATA MINING APPROACH.* 2015.

[7] Vivian-Griffiths, T., et al., *Predictive modeling of schizophrenia from genomic data: Comparison of polygenic risk score with kernel support vector machines approach.* American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 2019. **180**(1): p. 80-85.

[8] Passchier, R.V., et al., *Schizophrenia Polygenic Risk and Brain Structural Changes in Methamphetamine-Associated Psychosis in a South African Population.* Frontiers in Genetics, 2020. **11**.

[9] Richards, A.L., et al., *The relationship between polygenic risk scores and cognition in schizophrenia.* Schizophrenia bulletin, 2020. **46**(2): p. 336-344.

[10] Chen, J., et al., *Shared genetic risk of schizophrenia and gray matter reduction in 6p22. 1.* Schizophrenia bulletin, 2019. **45**(1): p. 222-232.

[11] Oh, J., et al., *Identifying Schizophrenia Using Structural MRI With a Deep Learning Algorithm.* Frontiers in Psychiatry, 2020. **11**(16).

[12] Patel, P., P. Aggarwal, and A. Gupta. *Classification of schizophrenia versus normal subjects using deep learning*. in *Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing*. 2016.

[13] Qureshi, M.N.I., J. Oh, and B. Lee, *3D-CNN based discrimination of schizophrenia using resting-state fMRI.* Artificial intelligence in medicine, 2019. **98**: p. 10-17.

[14] Lei, D., et al., *Detecting schizophrenia at the level of the individual: relative diagnostic value of whole-brain images, connectome-wide functional connectivity and graph-based metrics.* Psychological medicine, 2020. **50**(11): p. 1852-1861.

[15] Yan, W., et al., *Discriminating schizophrenia using recurrent neural network applied on time courses of multi-site FMRI data.* EBioMedicine, 2019. **47**: p. 543-552.

[16] Yan, W., et al. *Discriminating schizophrenia from normal controls using resting state functional network connectivity: A deep neural network and layer-wise relevance propagation method*. in *2017 IEEE 27th international workshop on machine learning for signal processing (MLSP)*. 2017. IEEE.

[17] Liu, K., et al., *Learn to combine modalities in multimodal deep learning.* arXiv preprint arXiv:1805.11730, 2018.

[18] Rashid, B. and V. Calhoun, *Towards a brain-based predictome of mental illness.* Human brain mapping, 2020. **41**(12): p. 3468-3535.

[19] Qureshi, M.N.I., et al., *Multimodal discrimination of schizophrenia using hybrid weighted feature concatenation of brain functional connectivity and anatomical features with an extreme learning machine.* Frontiers in neuroinformatics, 2017. **11**: p. 59.

[20] Cetin, M.S., et al. *Multimodal based classification of schizophrenia patients*. in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2015. IEEE.

[21] Cetin, M.S., et al., *Multimodal classification of schizophrenia patients with MEG and fMRI data using static and dynamic connectivity measures.* Frontiers in neuroscience, 2016. **10**: p. 466.

[22] Salvador, R., et al., *Multimodal Integration of Brain Images for MRI-Based Diagnosis in Schizophrenia.* Frontiers in neuroscience, 2019. **13**: p. 1203.

[23] Calhoun, V.D., J. Liu, and T. Adalı, *A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data.* Neuroimage, 2009. **45**(1): p. S163-S172.

[24] Du, Y., et al., *Neuromark: a fully automated ica method to identify effective fmri markers of brain disorders.* medRxiv, 2019: p. 19008631.

[25] Fu, Z., et al., *Dynamic functional network reconfiguration underlying the pathophysiology of schizophrenia and autism spectrum disorder.* Human brain mapping, 2021. **42**(1): p. 80-94.

[26] Chen, J., et al., *Guided exploration of genomic risk for gray matter abnormalities in schizophrenia using parallel independent component analysis with reference.* Neuroimage, 2013. **83**: p. 384-396.

[27] Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* The American journal of human genetics, 2007. **81**(3): p. 559-575.

[28] Ripke, S., et al., *Biological insights from 108 schizophrenia-associated genetic loci.* Nature, 2014. **511**(7510): p. 421.

[29] Du, Y., et al., *NeuroMark: an adaptive independent component analysis framework for estimating reproducible and comparable fMRI biomarkers among brain disorders.* MedRxiv, 2019: p. 19008631.

[30] Keator, D.B., et al., *The function biomedical informatics research network data repository.* Neuroimage, 2016. **124**: p. 1074-1079.

[31] Aine, C.J., et al., *Multimodal Neuroimaging in Schizophrenia: Description and Dissemination.* Neuroinformatics, 2017. **15**(4): p. 343-364.

[32] Adhikari, B.M., et al., *Functional network connectivity impairments and core cognitive deficits in schizophrenia.* Human brain mapping, 2019. **40**(16): p. 4593-4605.

[33] Fu, Z., et al., *Dynamic state with covarying brain activity-connectivity: on the pathophysiology of schizophrenia.* NeuroImage, 2021. **224**: p. 117385.

[34] Abrol, A., et al., *Replicability of time-varying connectivity patterns in large resting state fMRI samples.* Neuroimage, 2017. **163**: p. 160-176.

[35] Paszke, A., et al., *Pytorch: An imperative style, high-performance deep learning library.* arXiv preprint arXiv:1912.01703, 2019.

[36] Xu, B., et al., *Empirical evaluation of rectified activations in convolutional network.* arXiv preprint arXiv:1505.00853, 2015.

[37] Bahdanau, D., K. Cho, and Y. Bengio, *Neural machine translation by jointly learning to align and translate.* arXiv preprint arXiv:1409.0473, 2014.

[38] Huang, S.-C., et al., *Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines.* NPJ digital medicine, 2020. **3**(1): p. 1-9.

[39] Ramachandram, D. and G.W. Taylor, *Deep multimodal learning: A survey on recent advances and trends.* IEEE Signal Processing Magazine, 2017. **34**(6): p. 96-108.

[40] Reda, I., et al., *Deep learning role in early diagnosis of prostate cancer.* Technology in cancer research & treatment, 2018. **17**: p. 1533034618775530.

[41] Yoo, Y., et al., *Deep learning of brain lesion patterns and user-defined clinical and MRI features for predicting conversion to multiple sclerosis from clinically isolated syndrome.* Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2019. **7**(3): p. 250-259.

[42] Kingma, D.P. and J. Ba, *Adam: A method for stochastic optimization.* arXiv preprint arXiv:1412.6980, 2014.

[43] Stork, D., P. Hart, and R. Duda, *Pattern Classification. Hoboken*. 2012, NJ, USA: Wiley.

[44] Gat, I., et al., *Removing Bias in Multi-modal Classifiers: Regularization by Maximizing Functional Entropies.* arXiv preprint arXiv:2010.10802, 2020.

[45] Kokhlikyan, N., et al., *Captum: A unified and generic model interpretability library for pytorch.* arXiv preprint arXiv:2009.07896, 2020.

[46] Adebayo, J., et al., *Sanity Checks for Saliency Maps*. CoRR, 2018.

[47] Huang, D.W., et al., *DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists.* Nucleic acids research, 2007. **35**(suppl_2): p. W169-W175.

[48] Föcking, M., et al., *Proteomic and genomic evidence implicates the postsynaptic density in schizophrenia.* Mol Psychiatry, 2015. **20**(4): p. 424-32.

[49] Roberts, R.C., et al., *Dopaminergic synapses in the caudate of subjects with schizophrenia: relationship to treatment response.* Synapse (New York, N.Y.), 2009. **63**(6): p. 520-530.

[50] Fujii, Y., et al., *Positive associations of polymorphisms in the metabotropic glutamate receptor type 3 gene (GRM3) with schizophrenia.* Psychiatric genetics, 2003. **13**(2): p. 71-76.

[51] Mössner, R., et al., *Further evidence for a functional role of the glutamate receptor gene GRM3 in schizophrenia.* European Neuropsychopharmacology, 2008. **18**(10): p. 768-772.

[52] Petrovsky, N., et al., *Sensorimotor gating is associated with CHRNA3 polymorphisms in schizophrenia and healthy volunteers.* Neuropsychopharmacology, 2010. **35**(7): p. 1429-1439.