Multimodal MRI Study Using Convolutional Neural Networks for Schizophrenia Classification

By

Brendan A. Mitchell

A capstone project submitted for
Graduation with University Honors

June 1, 2020

University Honors
University of California, Riverside

APPROVED

_____
Dr. Xiaoping Hu
Department of Bioengineering

_____
Dr. Richard Cardullo, Howard H Hays Jr. Chair, University Honors

Abstract

Schizophrenia is a complex neurological disorder with no well-established, non-invasive diagnosis biomarker. Currently, diagnosis is based on clinical scores and questionnaires that are subjective and rely on the cooperation of patients - these factors lead to a significant misdiagnosis of schizophrenia. Thus, there is a need for an unbiased diagnostic tool based on objective markers, such as those derived from MRI images. In this paper, I employ a deep learning-based schizophrenia classification method via convolutional neural networks applied to multimodal MRI scans. Compared to classification results of T1-weighted structural MRI and resting-state functional MRI data individually, my multimodal classification performance was more sensitive and specific. This paper posits that multimodal classification may be a promising method for creating an accurate and reliable diagnostic tool for schizophrenia.

INTRODUCTION

Schizophrenia (SZ) is a multifaceted, enigmatic disorder of altered brain connectivity that affects about one percent of the population and is characterized by cognitive, perceptual, and affective deficits.[i] Such symptoms include hallucinations, delusions, disorganized thought, or reduced emotions; however, symptoms and psychopathological course vary from patient to patient and often overlaps with other mental disorders such as bipolar disorder (see Fig. 1) and attention-deficit hyperactivity disorder (ADHD). [ii,iii,iv,v]

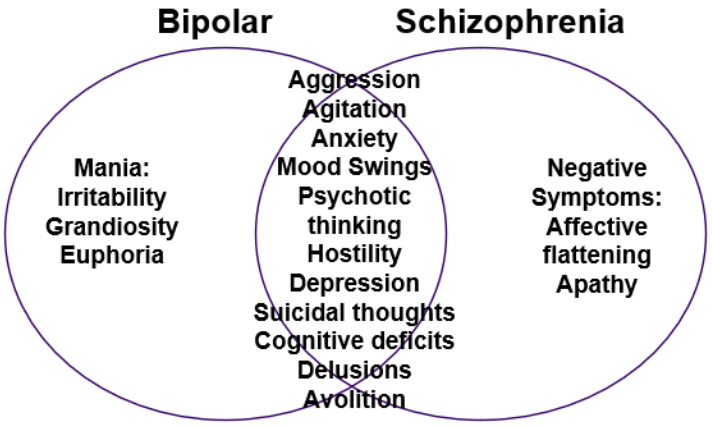

Fig. 1. Overlapping symptoms of bipolar disorder and schizophrenia. The Venn diagram shows the great degree of similarity in symptoms between the two neurological disorders (adapted from Keck PE Jr. et al.).

For these reasons, the current method of classifying schizophrenia, which is mainly based on process of elimination via clinical assessment and the patient's self-reported experiences, can result in significant misdiagnosis and overdiagnosis of schizophrenia. To further elucidate this point, a recent Johns Hopkins study revisited the diagnoses of 43 patients for schizophrenia spectrum disorder by a specialty early psychosis consultation clinic and concluded that 40 (93%) of them were misdiagnosed: 22 (51%) had a different primary psychotic disorder and 18 (42%) had no psychotic disorder at all.[vi] Thus, it is apparent there is a need for a more reliable, objective diagnostic tool.

Considerable effort has been made in this regard: advanced neuroimaging techniques such as magnetic resonance imaging (MRI), in particular, can provide powerful imaging information to help better clarify the structural and functional abnormalities associated with schizophrenia. MRI is a non-invasive, sensitive imaging modality that can measure both structure and function. Structural MRI (sMRI) scans can capture information about brain anatomy and morphology of white matter, gray matter, and cerebrospinal fluid. For example, McDonald et al. found reduced cortical volume and enlarged lateral ventricles in patients relative to controls (see Fig. 2).[vii]

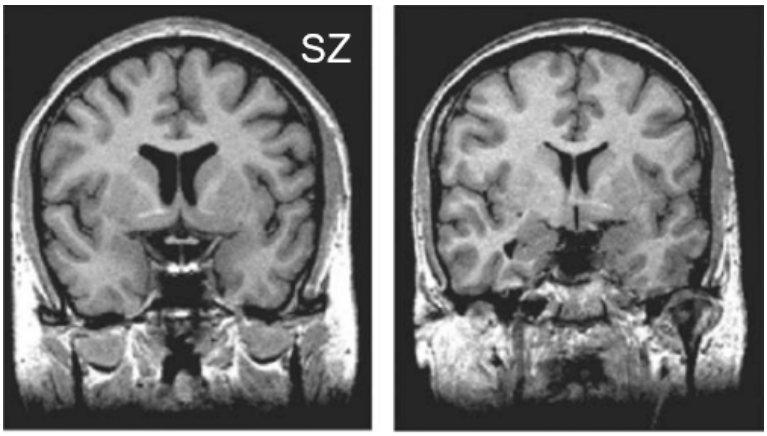

Fig. 2. T1-weighted MRI scans of schizophrenia patients and neurotypical control subjects. Right, schizophrenia (SZ) patient showing enlarged lateral ventricles and reduced cortex volume. Left, control subject for structural comparison (adapted from McDonald et al.).

Functional MRI (fMRI) scans measure the hemodynamic response via the blood oxygenation level dependent (BOLD) signal as an indirect measure of neural activity, capturing information about brain function and connectivity. For example, Whitfield-Gabrieli et al. and Garrity et al. found abnormal functional connectivity in the default mode network in patients relative to controls, in resting-state and task-based fMRI respectively.[viii,ix]

Despite MRI being capable of capturing multimodal information, existing studies are often focused on single-mode study, which only detects single dimensional information by each

modality, providing biomarkers that may either be distinct from or shared by other modalities, thereby missing potentially crucial differences between subject groups. Due in part by single-mode study and schizophrenia heterogeneity, the quantitative alterations between patients and control groups from these studies are often inconsistent and have not yet provided a diagnostic measure that is both sensitive and specific.[x,xi,xii,xiii,xiv] Consideration of multimodal data may unify disparate findings and provide a more comprehensive illustration of altered brain connectivity in schizophrenia – taking advantage of the strengths of each modality and their complementary relationships.

Several studies have shown success using multimodal data for schizophrenia classification. For example, Guo et al. found that three-way classification of schizophrenia from features extracted from fMRI, sMRI, and diffusion tensor imaging (DTI) data performed better than one or two modality classification with an accuracy of 86.52%.[xv] Similarly, Zhuang et al. showed that combining features from resting-state fMRI, sMRI, and DTI achieved the best performance for schizophrenia classification with an accuracy of 84.29%.[xvi] And Cetin et al. showed that the integration of features extracted from fMRI and magnetoencephalography (MEG) data improved schizophrenia classification accuracy over single-mode study.[xvii] As in these three studies, simple machine learning classifiers, like a support-vector machine (SVM), are typically used for multimodal classification tasks.

More recently, however, more complex, deep machine learning methods have been gaining a good reputation for extracting informative features from imaging data.[xviii] Among the various deep learning methods, a convolutional neural network (CNN) is a hierarchical model that outperforms many other algorithms on image classification tasks.[xix] Comprised of alternating convolutional and pooling layers (see below, Deep Machine Learning with CNNs),

CNNs can extract hierarchical features and learn complex patterns from the imaging data for improved classification. For example, Liu et al. utilized cascaded CNNs for multimodal classification of Alzheimer's disease (AD) and achieved an accuracy of 93.26% with MRI and positron emission tomography (PET) data.[xx] Although not a schizophrenia classification study, their success with AD demonstrates the promise of CNNs for multimodal classification.

Motivated by the success of the aforementioned studies and the present dominance of CNNs in various computer imaging tasks, my Honors Capstone project is focused on improving the current, subjective method of diagnosing schizophrenia by employing an objective deep learning-based multimodal schizophrenia classification method by CNNs applied to T1-weighted structural MRI and resting-state functional MRI scans. Here, I hypothesize that multimodal classification will perform better than single-modality classification.

MATERIALS AND METHODS

Data Sets

All data sets used in this work were obtained from the UCLA Consortium for Neuropsychiatric Phenomics LA5c Study (CNP), which are publicly available in the website (www.openneuro.org/datasets/ds000030). The primary goal of CNP was to better understand the functional and anatomical changes associated with neuropsychiatric disorders by considering neuroimaging data, phenotypic information, and clinical and behavioral assessments. The participant population consists of 130 control subjects and individuals diagnosed (following DSM-IV) with schizophrenia (50 subjects), attention deficit/hyperactivity disorder (43 subjects), and bipolar disorder (49 subjects), containing 155 men and 117 women ages between 21 and 50 years. More information regarding the participants and the study can be found in the Poldrack et

al. paper.[xxi] Participants of this study gave written informed consent and the procedures were approved by the Institutional Review Boards at UCLA and the Los Angeles County Department of Mental Health.

In this work, T1-weighted sMRI and resting-state fMRI data acquired from a 3T Siemens Trio scanner was used. The T1-weighted high-resolution anatomical scans (MPRAGE) were collected with the following parameters: TR=1.9s, TE=2.26ms, FOV=250mm, matrix=256x256, sagittal plane, slice thickness=1mm, and 176 slices. The resting-state fMRI scans lasted 304s and the participants were asked to remain relaxed and keep their eyes open—they were not presented any stimuli or asked to respond during the scan—and the scans were collected using a T2*-weighted echoplanar imaging (EPI) sequence with the following parameters: TR=2s, TE=30ms, FOV=192mm, matrix=64x64, oblique slice orientation, slice thickness=4mm, 34 slices, and flip angle=90°. These neuroimaging data were then preprocessed according to the pipeline outlined in the Gorgolewski et al. paper.[xxii] From these preprocessed images, the T1-weighted sMRI and resting-state fMRI images of 50 control subjects and 50 schizophrenia patients were downloaded for the purpose of this project.

These images were then further preprocessed in Python with NumPy, NiBabel, scikit-learn, and Nilearn packages to prepare the input for schizophrenia classification via CNNs. T1-weighted structural images are 3D brain volumes comprised of 2D brain slices, having spatial dimensions x, y, and z. Preprocessing steps for T1-weighted sMRI data included: accessing the image as a NumPy array (shape:182, 218, 182), performing column-wise standardization, assigning class label (0 for control group, 1 for schizophrenia group), concatenating the control and patient data (shape:100, 182, 218, 182; where 100 refers to the total number of subjects), and stacking, column-wise, the labels from both classes (shape: 100,). Functional MR images, like

structural images, have three spatial dimensions x, y, and z, but the 3D volume is acquired over a period of time; time is the fourth dimension. Preprocessing steps for the 4D resting-state fMRI data included: extracting a brain mask for each image, applying the mask to extract the corresponding time series (see Fig. 3), performing column-wise standardization, reshaping the 2D data arrays to have same shape (152, 74362), assigning class label, concatenating the control and patient data (shape:100, 152, 74362), and stacking the labels (shape:100,). After the additional preprocessing, the combined data and label arrays for both modalities were randomly split into training (n=64), validation (n=16), and testing (n=20) sets for classification.

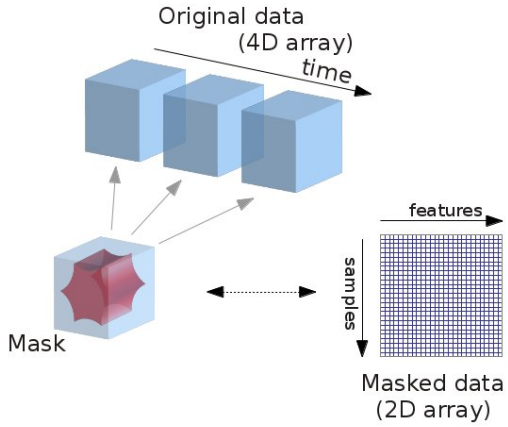

Fig. 3. Transforming 4D fMRI images to 2D data arrays. Applying a brain mask will convert the 4D data into a restructured 2D data representation, time x voxels (adapted from Nilearn documentation).

Deep Machine Learning with CNNs

Several of the aforementioned studies relied on regions of interests (ROIs) and/or extraction of handcrafted features that are determined by prior hypotheses, which may neglect the brain-wide changes associated with schizophrenia, thereby limiting the representation power of extracted features. Distinct from these methods, deep CNNs can learn hierarchical features and patterns from the voxel intensities of the whole-brain images for classification. CNNs are

typically built by stacking alternatively convolutional and pooling layers, a fully connected layer, and a softmax classification layer. Here, the first convolutional layer of a CNN extracts basic, low-level features from the images, the second convolutional layer accepts the features extracted by the first as its inputs to learn more complex features, etc.

A convolutional layer works by convolving the input image with the learned kernel filters, adding a bias term, applying a non-linear activation function, and producing a feature map by each filter. Typically, the number of filters in convolutional layers starts low and increases with each layer. After each convolutional layer, there is usually a pooling layer. In this work, max pooling is used. It works by reducing the feature map along the spatial dimensions by replacing each element with their maximum—ideally, keeping the most influential features. After stacking several convolutional and max pooling layers, the high-level reasoning in the deep CNN is achieved through the fully connected layers. All feature maps are flattened into a 1D vector as the inputs of the fully connected layer. The fully connected layer generates the learned linear combination of all neurons from the previous layer and passes them through a nonlinear activation function. Lastly, a softmax classification layer is fine-tuned by backpropagation to predict class probability (ranging from 0 to 1).

Model Architecture and Implementation

For this work, 2D CNNs are used for the 3D structural MR images and 1D CNNs are used for the 2D resting-state images. The CNN architecture is comprised of repeated blocks of (convolution, convolution, max pooling, dropout), depicted in Fig. 4. Each CNN consists of 4 blocks, followed by a fully connected layer and a softmax layer. Before each convolution, the inputs are padded such that the dimensions of the outputs are equal to that of the inputs. Inspired by the success of Sturmfels et al.,[xxiii] a reverse filter scheme was employed in which the first

convolutional layer has the most filters with the purpose of focusing on feature extraction earlier in the network. The number of convolution filters were set to 64, 32, 16, 8 for the 4 blocks, respectively. For the structural-based 2D CNN, the sizes of convolutional filters were set to 3 × 3 and max pooling sizes of 2 × 2. And for the functional-based 1D CNN, the sizes of filters were set to 3 and max pooling sizes of 2. These layers use the *ReLu* activation function.[xxiv] In addition, dropout layers are utilized to limit overfitting and improve generalization.

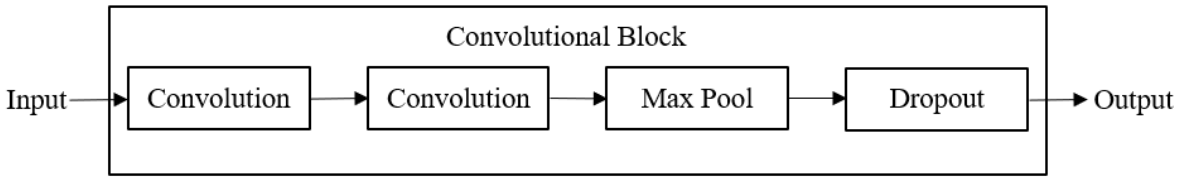

Fig. 4. Convolutional block for CNN architecture. The CNN architecture consists of 4 repeated blocks, the layers within a block have the same number of filters and same sizes.

The classification algorithm in this work was implemented with the Keras library in Python with TensorFlow backend and experiments were conducted on the Hu Lab Cluster. Since the structural and functional images have different dimensions, two CNN models were trained separately on their respective training sets to learn features. The outputs from each CNN were flattened to 1D vectors and passed through separate fully connected and softmax layers for single-mode classification. For multimodal classification, however, the flattened outputs from each CNN were concatenated to form a fully connected layer that contains information from both modalities, and then passed through a softmax layer (see Fig. 5).

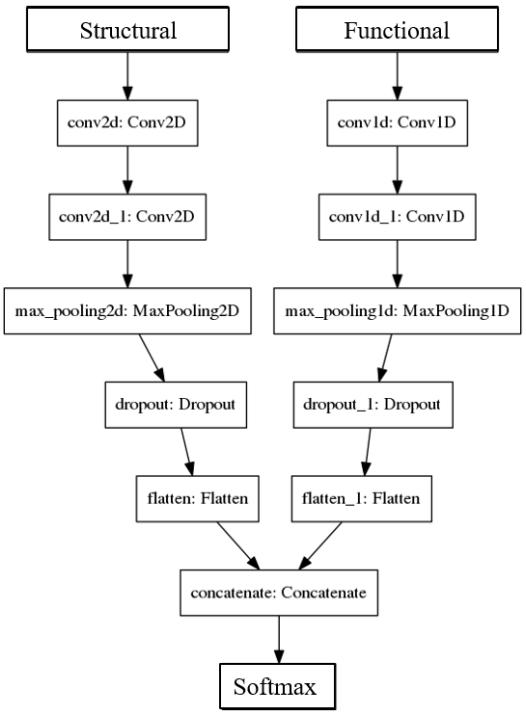

Fig. 5. Simplified multimodal CNN architecture.

Here, the two models train on the labeled images from the control group and patient group to learn discriminative features. The validation set was also used during model training to evaluate the models. The model architectures and parameters were then modified according to the results on the validation set to maximize validation performance (see Fig. 6). After determining the best performing model, the model makes predictions on novel, unlabeled images. For all experiments, the models are trained ten times using the Adam optimizer[xxv] and sparse categorical cross entropy function (due to label format) for loss computations with different random initializations (via Xavier initialization)[xxvi] over 100 epochs – i.e., the training set passes through the models 100 times. For each run, the model weights that achieved the best performance on the validation set during training were saved. The best weights were then used

when evaluating the model performance on the novel test set. All reported performance metrics pertain to the test set unless otherwise specified.

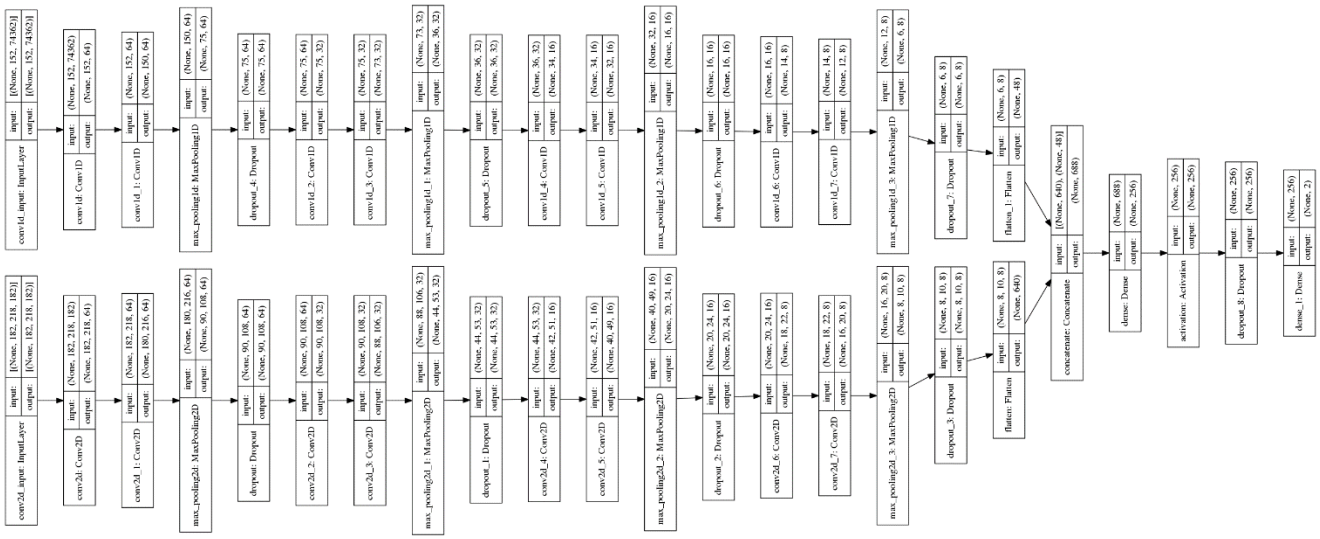

Fig. 6. Full, deep multimodal CNN model with 4 convolutional blocks. Architecture determined after comparing the performances of different configurations.

There are several metrics that can be used to evaluate classification performance. In this work, the following metrics were used for comparison: accuracy (ACC), log-loss, precision, sensitivity (SEN), specificity (SPE), and area under the curve (AUC) score. Accuracy corresponds to the percentage of total images correctly classified. Log-loss measures the uncertainty of the prediction and its variation from correct label. Precision measures the number of images that are correctly classified as positive (belonging to schizophrenia group) out of the total number of positive images. Sensitivity, also referred to as recall, or the true positive rate (TPR), measures the number of images correctly classified as positive out of the number of true positive (correctly predicts schizophrenia) and false negative images (incorrectly predicts control). Specificity, also referred to as the true negative rate (TNR), measures the number of images correctly identified as negative (belonging to control group) out of the total number of negative images. The AUC score for the receiver operating characteristic (ROC) curve, a plot of

TPR vs FPR, for example, indicates how well the probabilities of positive classes are separated from the negative classes. In general, the values of these metrics range from 0 to 1 (or scaled to 0 to 100%), where values close to 1 represent good performance; the opposite is true when referring to loss. It should also be noted that diagnostic tests typically exhibit trade-offs between metrics like specificity and sensitivity.

EXPERIMENTAL RESULTS

Structural, Functional, and Multimodal Classification results

The following experiments were performed on the hold-out testing set of 20 subjects to compare the classification performances of T1-weighted structural MRI, resting-state functional MRI, and multimodal-based models for discriminating schizophrenia (SZ) patients from the normal control subjects (NC). The ROC curves were used to visualize the trade-offs between specificity and sensitivity of the models by plotting the false positive rate (1–specificity) against the true positive rate or sensitivity (see Fig. 7), and to compute the AUC score (see Table 1). In addition to plotting the performances from the three MRI-based models, a "no skill" model (AUC score of 50%) that makes random predictions or only predicts one class (e.g., only predicting schizophrenia) was included for visual comparison. From the graph, all three MRI-based models performed better than the no skill model (red line), suggesting that these models can discriminate between images from each class unlike the no skill model. It is also clear from the graph that the multimodal-based classification model (green curve) performs the best and most consistently.

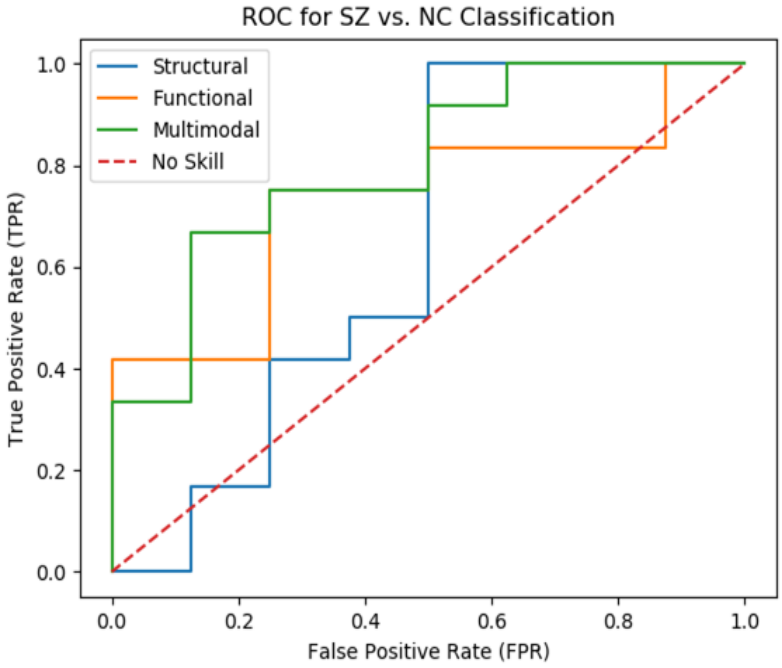

Fig. 7. ROC curves of structural, functional, and multimodal classifications of SZ vs. NC. Red dashed line, performance of "no skill" model that cannot discriminate between classes. Blue curve, structural-based model performance. Orange curve, functional-based model performance. Green curve, multimodal-based model performance. All three MRI models perform better than the no skill model with multimodal being the most consistent.

Table 1 shows the quantitative results of the classification performances for the models. From the results shown in the table, structural-based classification achieves the highest test accuracy of 75% and highest sensitivity, but at the cost of low specificity. In other words, the structural model is great at detecting actual cases of schizophrenia but also has a high rate of false positives. Functional-based classification achieves the lowest test accuracy of 55% and lowest sensitivity but has the highest specificity. Both the structural and functional-based models have reciprocal strengths, sacrificing specificity for better sensitivity, or vice-versa. Multimodal-based classification achieves a high test accuracy of 70% without the sensitivity-specificity trade-off seen with the structural and functional-based models. Furthermore, the multimodal-based classification achieves the highest AUC score of 80.21% and the lowest loss. Thus, it is

clear from these results that multimodal classification performs better than each individual modality.

Table 1 Comparison of Classification Performances on SZ vs. NC

SZ vs. NC	ACC%	SEN%	SPE%	AUC%	Loss
Structural	75	100	37.5	63.54	0.6906
Functional	55	25	100	72.92	0.6888
Multimodal	70	66.67	75	80.21	0.6596

In addition to graphing ROC curves to compare model performances, Precision-Recall (P-R) curves were used to visualize how well the models predict the positive class, schizophrenia (see Fig. 8) and to compute AUC scores (see Table 2). As was done for the ROC curve, the no skill model was included in the P-R curve for comparison. From the graph, all three MRI-based models performed better than the no skill model (red line) and the functional (orange curve) and multimodal-based (green curve) models performed better than the structural-based model, with multimodal being the most consistent. These observations are confirmed by the calculated AUC scores in which the multimodal-based model had the highest AUC score of 86.19% and the structural-based model had the lowest score of 62.26%, as in Table 2.

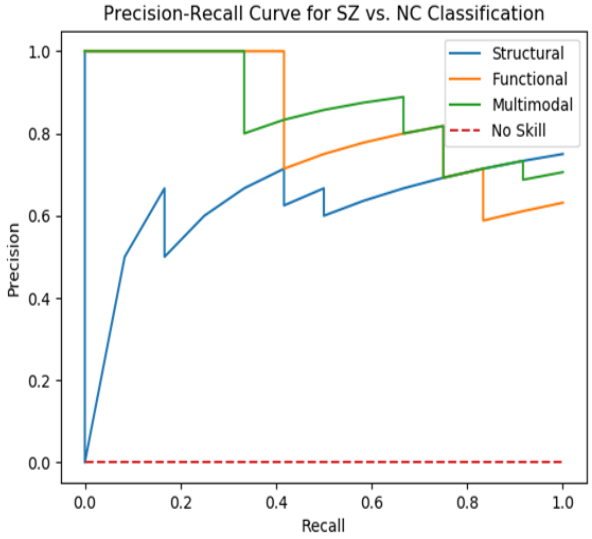

Table 2 Comparing P-R AUC

SZ vs. NC	AUC%
Structural	62.26
Functional	83.49
Multimodal	86.19

Fig. 8. Precision-recall curves of structural, functional, and multimodal classifications of SZ vs. NC. Red dashed line, performance of no skill model. Blue curve, structural-based model performance. Orange, functional-based model performance. Green, multimodal-based performance. All three MRI-based models perform better than the no skill model with multimodal being the most consistent.

DISCUSSION

For my Honors Capstone project, I employed deep CNNs based on multimodal MRI scans to classify, or diagnose, schizophrenia. Although my preliminary findings demonstrate that multimodal classification performs better and more consistently than single-mode study, there are several limitations to my work that should be considered. First, my CNNs were trained on a dataset of images from 100 total subjects, which is exceedingly small for a CNN-based classification task; typically, CNNs are trained on thousands of images for better representation. For this reason, my models were unstable and showed signs of overfitting. Second, using whole-brain images as the input requires training a large number of parameters and is susceptible to noise, increasing computation time and may cause overfitting. Third, the architecture and parameters of the CNN model, such as number of layers, types of layers, and size and number of filters, may not be optimal for these MR images. Fourth, 2D CNNs were used for 3D sMRI data and 1D CNNs for 2D fMRI data, potentially losing rich spatial and temporal information. Fifth, only T1-weighted sMRI and resting-state fMRI were used in this work; other modalities such as DTI may be included to further improve classification performance. Sixth, randomized splitting of data into training, validation, and testing sets resulted in class imbalances (e.g., training set had 28 images belonging to schizophrenia group and 36 belonging to control group). Seventh, it is difficult to visualize the potentially clinically relevant features that were learned by the CNNs.

However, there are some suggestions to address these limitations to improve classification performance. Since the dataset used in this study was small for effective CNN training, it may be beneficial to incorporate other similar datasets to increase sample size and augment (e.g. shift and down-sample) the training data to generate additional images. Rather than directly using whole-brain images as the input, the images can be divided into smaller,

equal-sized chunks to reduce computation costs which may allow for CNNs to extract more relevant features. Several architecture and parameter configurations can be compared to determine the best performing model setup. And, instead of using 2D and 1D CNNs, 3D and 2D CNNs should be employed to better capture information from the structural and functional images, respectively. Other neuroimaging modalities that have been used to characterize schizophrenia abnormalities, such as DTI, should also be considered to provide another clinically relevant perspective of the brain. Additionally, it would be especially interesting to incorporate data from other patient groups such as bipolar disorder and ADHD since they share features with schizophrenia. Finally, instead of randomly splitting the whole dataset into training, validation, and testing sets, cross-validation techniques can be used to preserve class balance.

CONCLUSION

Today, the collection of multimodal brain images from each subject is becoming common practice, paving the way for improved computer vision tasks and medical image analyses based on large multimodal datasets. In this paper, a multimodal classification algorithm based on deep CNNs is used to predict schizophrenia from structural and functional MRI images. The experimental preliminary results demonstrate that multimodal classification achieves better performance over each individual modality, having both higher specificity and sensitivity, the highest AUC scores, and the lowest log-loss. Thus, the results support the main hypothesis driving my project; that is, that multimodal classification does perform better than single-mode study. However, as noted above, future work can address the limitations of this project to improve schizophrenia classification performance. Additionally, it should be noted, that although CNN training and inference is 100% algorithmic, the data used to train the model was labeled by humans which can be subjective and prone to bias.

[i] Picchioni, M. M., & Murray, R. M. (2007, July 14). Schizophrenia. *British Medical Journal*, Vol. 335, pp. 91–95. https://doi.org/10.1136/bmj.39227.616447.BE

[ii] Lang, F. U., Kösters, M., Lang, S., Becker, T., & Jäger, M. (2013, March). Psychopathological long-term outcome of schizophrenia -- a review. *Acta Psychiatrica Scandinavica*, Vol. 127, pp. 173–182. https://doi.org/10.1111/acps.12030

[iii] Bambole, V., Johnston, M., Shah, N., Sonavane, S., Desouza, A., & Shrivastava, A. (2013). Symptom overlap between schizophrenia and bipolar mood disorder: Diagnostic issues. *Open Journal of Psychiatry*, *03*(04), 8–15. https://doi.org/10.4236/ojpsych.2013.34a002

[iv] Buckley, P. F., Gowans, A., Sebastian, C. S., Pathiraja, A., Brimeyer, A., & Stirewalt, E. (2004). The boundaries of schizophrenia: Overlap with bipolar disorders. *Current Psychosis and Therapeutics Reports*, *2*(2), 49–56. https://doi.org/10.1007/s11922-004-0031-8

[v] Corbisiero, S., Riecher-Rössler, A., Buchli-Kammermann, J., & Stieglitz, R. D. (2017). Symptom overlap and screening for symptoms of attention-deficit/hyperactivity disorder and psychosis risk in help-seeking psychiatric patients. *Frontiers in Psychiatry*, *8*(OCT). https://doi.org/10.3389/fpsyt.2017.00206

[vi] Coulter, C., Baker, K. K., & Margolis, R. L. (2019). Specialized consultation for suspected recent-onset schizophrenia: Diagnostic clarity and the distorting impact of anxiety and reported auditory hallucinations. *Journal of Psychiatric Practice*, *25*(2), 76–81. https://doi.org/10.1097/PRA.0000000000000363

[vii] McDonald, C., Grech, A., Toulopoulou, T., Schulze, K., Chapple, B., Sham, P., … Murray, R. M. (2002). Brain volumes in familial and non-familial schizophrenic probands and their unaffected relatives. *American Journal of Medical Genetics - Neuropsychiatric Genetics*, *114*(6), 616–625. https://doi.org/10.1002/ajmg.10604

[viii] Whitfield-Gabrieli, S., Thermenos, H. W., Milanovic, S., Tsuang, M. T., Faraone, S. V., McCarley, R. W., … Seidman, L. J. (2009). Hyperactivity and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(4), 1279–1284. https://doi.org/10.1073/pnas.0809141106

[ix] Garrity, A. G., Pearlson, G. D., McKiernan, K., Lloyd, D., Kiehl, K. A., & Calhoun, V. D. (2007). Aberrant "default mode" functional connectivity in schizophrenia. *American Journal of Psychiatry*, *164*(3), 450–457. https://doi.org/10.1176/ajp.2007.164.3.450

[x] Gandal, M. J., Haney, J. R., Parikshak, N. N., Leppa, V., Ramaswami, G., Hartl, C., … Geschwind, D. H. (2018). Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science*, *359*(6376), 693–697. https://doi.org/10.1126/science.aad6469

[xi] Buchanan, R. W., & Carpenter, W. T. (1994). Domains of psychopathology: an approach to the reduction of heterogeneity in schizophrenia. *The Journal of Nervous and Mental Disease*, *182*(4), 193–204. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10678315

[xii] Joyce, E. M., & Roiser, J. P. (2007, May). Cognitive heterogeneity in schizophrenia. *Current Opinion in Psychiatry*, Vol. 20, pp. 268–272. https://doi.org/10.1097/YCO.0b013e3280ba4975

[xiii] Arnedo, J., Svrakic, D. M., Del Val, C., Romero-Zaliz, R., Hernández-Cuervo, H., Fanous, A. H., … Zwir, I. (2015). Uncovering the hidden risk architecture of the schizophrenias: Confirmation in three independent genome-wide association studies. *American Journal of Psychiatry*, *172*(2), 139–153. https://doi.org/10.1176/appi.ajp.2014.14040435

[xiv] Schennach, R., Meyer, S., Seemüller, F., Jäger, M., Schmauss, M., Laux, G., … Riedel, M. (2012). Response trajectories in "real-world" naturalistically treated schizophrenia patients. *Schizophrenia Research*, *139*(1–3), 218–224. https://doi.org/10.1016/j.schres.2012.05.004

[xv] Guo, S., Huang, C. C., Zhao, W., Yang, A. C., Lin, C. P., Nichols, T., & Tsai, S. J. (2018). Combining multi-modality data for searching biomarkers in schizophrenia. *PLoS ONE*, *13*(2). https://doi.org/10.1371/journal.pone.0191202

[xvi] Zhuang, H., Liu, R., Wu, C., Meng, Z., Wang, D., Liu, D., … Li, Y. (2019). Multimodal classification of drug-naïve first-episode schizophrenia combining anatomical, diffusion and resting state functional resonance imaging. *Neuroscience Letters*, *705*, 87–93. https://doi.org/10.1016/j.neulet.2019.04.039

[xvii] Cetin, M. S., Houck, J. M., Rashid, B., Agacoglu, O., Stephen, J. M., Sui, J., … Calhoun, V. D. (2016). Multimodal classification of schizophrenia patients with MEG and fMRI data using static and dynamic connectivity measures. *Frontiers in Neuroscience*, *10*(OCT), 466. https://doi.org/10.3389/fnins.2016.00466

[xviii] Shen, D., Wu, G., & Suk, H.-I. (2017). Deep Learning in Medical Image Analysis. *Annual Review of Biomedical Engineering*, *19*(1), 221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442

[xix] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

[xx] Liu, M., Cheng, D., Wang, K., & Wang, Y. (2018). Multi-Modality Cascaded Convolutional Neural Networks for Alzheimer's Disease Diagnosis. *Neuroinformatics*, *16*(3–4), 295–308. https://doi.org/10.1007/s12021-018-9370-4

[xxi] Poldrack, R. A., Congdon, E., Triplett, W., Gorgolewski, K. J., Karlsgodt, K. H., Mumford, J. A., … Bilder, R. M. (2016). A phenome-wide examination of neural and cognitive function. *Scientific Data*, *3*, 1–12. https://doi.org/10.1038/sdata.2016.110

[xxii] Gorgolewski, K. J., Durnez, J., & Poldrack, R. A. (2017). Preprocessed Consortium for Neuropsychiatric Phenomics dataset. *F1000Research*, *6*, 1262. https://doi.org/10.12688/f1000research.11964.1

[xxiii] Sturmfels, P., Rutherford, S., Angstadt, M., Peterson, M., Sripada, C., Wiens, J., … Wiens, J. (2018). A Domain Guided CNN Architecture for Predicting Age from Structural Brain Images. In *Proceedings of Machine Learning Research* (Vol. 85). Retrieved from https://gitlab.eecs.umich.edu/mld3/brain_age_prediction

[xxiv] Nair, V., & Hinton, G. E. (2010). Rectified linear units improve Restricted Boltzmann machines. *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, 807–814.

[xxv] Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.

[xxvi] Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Journal of Machine Learning Research*, *9*, 249–256. Retrieved from http://www.iro.umontreal.