Predicting Psychosis using Convolutional Neural Networks on Brain MRIs: A Comparison of Model Architectures

4035323, 4035678,4033882, 4037605

***Abstract*— Magnetic Resonance Imaging (MRI) is a powerful tool for studying the brain and its structure. It is particularly useful for identifying abnormalities in patients with psychiatric disorders such as psychosis.**

In this study, a binary classification task was performed on brain MRIs to predict the condition of psychosis by using convolution based neural networks. Three different models, a deep plain neural network, Residual Network (ResNet), and a wide and deep model, were compared and the results showed that the wide and deep model performed better over the other models with 88 percent train accuracy and 60.4% test accuracy with an error rate of 63%. Additionally, the wide and deep model was able to achieve this performance with fewer training examples, making it more computationally efficient. These findings indicate that this approach could be a valuable addition to the arsenal of diagnostic tools for psychosis and other psychiatric disorders.

***Index Terms*— Convolutional Neural Network (CNN), MRI, Psychosis, ResNet, psychiatric disorder/psychosis.**

I. INTRODUCTION

Psychosis is characterized by hallucinations and delusions, and the exact cause of these symptoms is still unclear. It is influenced by various factors like genetic makeup, environment, and mental health [1]. Most conditions frequently share symptoms with other mental health disorders, making a precise diagnosis of them challenging. When compared to healthy people, patients with psychosis occasionally experience variations in brain volume. A potent tool for the non-invasive imaging of the human brain is magnetic resonance imaging (MRI). It has been extensively employed in the field of neuroscience to investigate several neurological and psychiatric diseases since it enables researchers to analyze brain anatomy and function. By identifying specific neuroanatomical characteristics of a particular disorder, researchers can develop diagnostic tools that are more accurate and specific, leading to improved patient care.

An accurate representation of the brain can be obtained with MRI by recording images of the brain from various angles, in various planes [2]. The intensity of an image at a particular voxel, the three-dimensional pixels within an MRI provide more information on the tissue at that location [3]. An MRI of brain shows greater intensity values in regions with more water or fat, including the white matter. On the other hand, in spaces with more air, intensity values are lower. Tissues such as grey matter, white matter, and cerebrospinal fluid, can be identified and distinguished by examining signal Intensities. The analysis of massive datasets of brain MRI images, which can be time and resource-intensive when done manually, can be automated with the aid of machine learning.

Machine learning applications in psychiatry are still in the early stages of development, and these techniques are not yet widely used in clinical settings. In recent years, medical imaging field has seen an increase in the usage of image processing and Deep learning methods. In this paper, we have analyzed an MRI imaging dataset to understand the differences and classify Healthy Control (HC) and first episode patients (FEP) with various types of psychosis, such as schizophrenia and bipolar disorder. FEP patients are particularly valuable for the research as they are in the early stages of mental disorders and have not yet been exposed to medications or other complicating factors. In this paper, convolution based sequential, Non-sequential and ResNet classification models are developed to classify normal and abnormal brains from MRI dataset. The model is trained on a dataset of MRI images that included both HC and FEP subjects, to identify differences in brain structure and distinguish between the two groups.

The sequential neural networks model such as deep neural networks capture temporal patterns in the data by using feedback connections that allow information to flow from one step in the sequence to the next. Non-sequential models extract spatial features of the images, and the architectures can be designed to increase the depth of the network to improve the feature representation for the final classification. Residual connections are a key feature of ResNet neural network architecture. These connections are additional pathways that bypass one or more layers in the network and allow input data to be transmitted directly to later layers. By allowing the input data to flow directly to later layers, residual connections can help to improve model learning and avoid problems with vanishing gradients, which occur when training deep neural networks. Residual connections play a critical role in the ResNet architecture and contribute to its effectiveness.

The objective of this study is to investigate the effectiveness of using convolutional neural networks (CNNs) to automatically detect psychosis by analyzing brain MRIs. We aim to compare the performance of various CNN model architectures and identify the best performing architecture for this task. By improving the ability to automatically detect psychosis using brain MRIs, this study has the potential to enhance the diagnostic process and provide a more efficient and accurate means of identifying patients with psychosis. The subsequent sections of the paper are organized as follows: Section 2 covers the literature review and Section 3 describes the dataset and the pre-processing and data augmentation methods, the training procedure, and the proposed network architectures. The experimental results and comparative analysis are presented in Section 4. Finally, in Section 5, we present our conclusions and discuss directions for further research.

II. LITERATURE REVIEW

A 2022 study utilized deep learning approaches, specifically CNN models with transfer learning, to classify brain tumor MRI images and achieved an accuracy of 97% by incorporating data augmentation and using the ResNet-50 architecture [5]. Incorporation of data augmentation and transfer learning techniques had enhanced the model's classification performance. In 2021, Z. Li et al. used a ResNet-based CNN to classify brain MRI images of patients with bipolar disorder, first episode psychosis and healthy controls and achieved an accuracy of 98% [1]. There are studies from 2020 and 2019 which have used CNNs, specifically the ResNet model, to classify brain MRI images for disease detection and obtained accuracies of around 95% [6][7]. With the brain MRI images, multi class classification has been implemented with Convolutional networks using models ResNet, VGG, AlexNet. ResNet architecture, had different number of residual layers added creating ResNet-18, ResNet-34, and ResNet-50 architectures. Among these ResNet-50, with 50 layers had the highest accuracy of 95% [7]. These studies demonstrate the potential of CNN models and ResNet-based CNNs for the classification of brain MRI images in the context of psychosis and other mental health conditions in the brain.

There is a limited amount of research available on using non-sequential wide and deep architecture specifically for brain MRI classification. However, studies have been conducted on the implementation of deep and wide models in image classification and how it affects and improves performance [8]. The combination of wide and deep architectures is used in non-sequential model to improve the classification of brain MRI images for psychosis. Wide architectures can effectively model shallow relationships while deep architectures can effectively model deep relationships. The combination of both architectures can be used to discover complex features that can be used to improve the classification of brain MRI images. However, further research is needed to evaluate the performance of ResNet and other CNN architectures on larger and more diverse datasets and pre-processing techniques for improving the accuracy and generalizability of the classifiers. Network architectures with more layers have the capability to extract more intricate patterns from the input data, but they also necessitate larger amounts of data and computational power for both training and operation. The goal of this research is to develop automated diagnostic tools that can aid in the early detection and treatment of psychosis, thus improving patient outcomes. We have developed Sequential, Non-Sequential and a ResNet model with one residual block architecture for our analysis considering the limitation in dataset size and computational power.

III. METHODOLOGY

A. Dataset Sources

The dataset for this study consists of brain MRI images belonging to male and female respondents of various ages from seven different sites. It includes both healthy control participants and individuals with first-episode psychosis who were tested for psychosis, and the results of the test and their diagnosis recorded. The dataset includes a total of 674 records, with 223 individuals exhibiting psychotic symptoms and 451 subjects with normal brain.

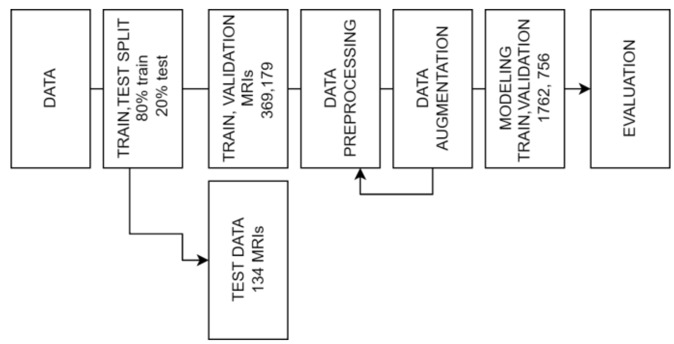

Fig 1. Model pipeline

The dimension of the brain images in the dataset is (79, 85, 80), indicating that each image consists of 79 rows of voxels, 85 columns of voxels and 80 slices of the image. Slices are parts of 2D images taken along the z-axis of these 3D MRI images, each of the slices showing a different cross-sectional view of brain. The MRI Images are classified based on the diagnostic value assigned to everyone. Individuals without symptoms are assigned a diagnostic value of 0 and are classified as normal brains. Those individuals who have been diagnosed with a psychosis disorder, as indicated by a diagnostic value ranging from 1 to 7, are classified as abnormal brains in subsequent analysis. We have created a binary classification task using the pipeline as in Figure 1.

1) Limitations of the dataset

The size of the data was limited to 674 records and the distribution of classes was highly imbalanced the ratio of negative to positive class was twice.

The quality of scans is another challenge, though the size is relatively large we are not sure what postprocessing has been done.

The MRI scans are done with various protocols and different energy pulses so we do not have much clear understanding of the granularity of the properties of data as this will affect the choices that we have to take in implementation of algorithms.

B. Data Pre-processing

To prepare MRI scan data for further analysis, it is necessary to adjust the dimensions of the data during the pre-processing stage so that we can match the input shape the algorithms are expecting. The dimensions of 3D MRI scans are represented by the height, the number of rows in the data; the width, the number of columns in the data; and the depth, the number of slices in the data. Data pre-processing involved cropping, resizing, and normalizing the data to specific dimensions. The images are cropped to narrow down the region of interest.

To trade off image quality and size we have resized the images maintaining the aspect ratio during the resizing process. The intensity values are normalized using the z-score normalization technique to ensure that they all have the same range.

To increase the size of dataset and improve the robustness of the algorithm's we performed data augmentation with rotation, translation, and scaling to the training images. By increasing the size of the dataset through data augmentation, we addressed one way to reduce overfitting is by increasing the size of data. The combined dataset of original and augmented images is then split into two subsets: a training set and a validation set. The training set contains 70% of the images and is used to train the model, while the validation set contains the remaining 30% of the images and is used to evaluate the performance of the model during training.

D. Network Architecture

The architecture is designed to take advantage of the 3-dimensional nature of the MRI data by using a series of 3D convolutional layers. Sequential, Non-Sequential, and ResNet models are implemented to evaluate the performance. The input to the network 3D MRI scans with batch size of 20 and the dimensions of the scans being 30 height, 24 width, 79 slices. All the model architectures described below were designed to be simple to prevent overfitting due to less data. The filters size was restricted to maximum of 20 due to limitation of hardware resource.

The reason behind choosing these three models is to experiment and see which model would be able to generalize and suit the dataset well. There is no specific reason for reshaping the dimensions of the scans, the dimensions were chosen based on the hardware resources available. And there is no specific reason for the values of batch size, epochs, penalizes, optimizer chosen, drop out layer values these are the hyperparameters and throughout the training process were adjusted and tuned accordingly to get the best possible fit. ReLU and sigmoid activation functions were used throughout. The choice for using ReLU in hidden layers is being that it is very computationally less expensive and faster and sigmoid is traditional and standard activation that is used in combination with binary cross entropy loss function other functions like SoftMax could have been chosen as well. The default learning rate is 0.01 unless specified.

Penalization was done to weights and loss functions of the dense layers only this is to allow more freedom to learn complex patterns from the convolution layers. We have experimented with different values and with penalization at convolution layers, but we found that penalization at dense layers gave a better fit compared to penalizing the convolution layers. This suggests that the dense layers were the primary source of overfitting in the model and regularizing them improved the overall performance of the network

One common pattern that can be seen throughout the models is the increase in filters as the network gets deeper. This is done so that deeper layers can more about the abstract features. More filters mean that each convolutional layer can learn more different features, which in turn allows the network to learn more complex representations of the input data. Increasing the number of filters in the deeper layers also helps to combat the vanishing gradient problem, which is a common issue in deep neural networks. As the gradients are passed through the layers during backpropagation, they can become smaller and smaller, making it difficult for the network to learn. By using more filters in the deeper layers, the network can learn multiple features simultaneously which will help to overcome the vanishing gradient problem. Another reason for increasing the number of filters is that, as the depth of the network increases, the size of the feature maps decreases due to the pooling operation. As the feature maps get smaller and smaller, the amount of information that can be captured also decreases. So, increasing the number of filters can compensate for this loss of information.

1) Plain Deep Neural Network

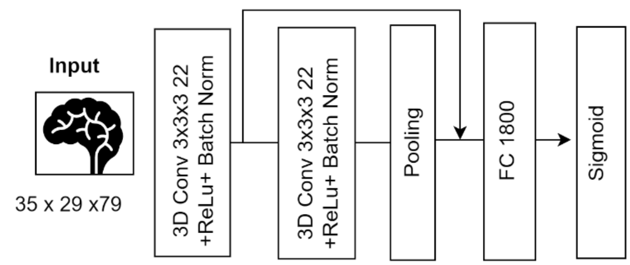

Fig 2. Sequential architecture

The sequential model seen in Figure 2 was designed with tradeoff between the layers and model complexity. Four layers consisting of two convolutions and two fully connected dense layers with drop out layer. The total epochs used to train the model are 200 but using early stopping the training was stopped after 26 epochs, with batch size 20. The batch size has been kept small because of limitations in hardware resources. And training was stopped after 65 epochs to prevent overfitting.

2) Non-sequential Model

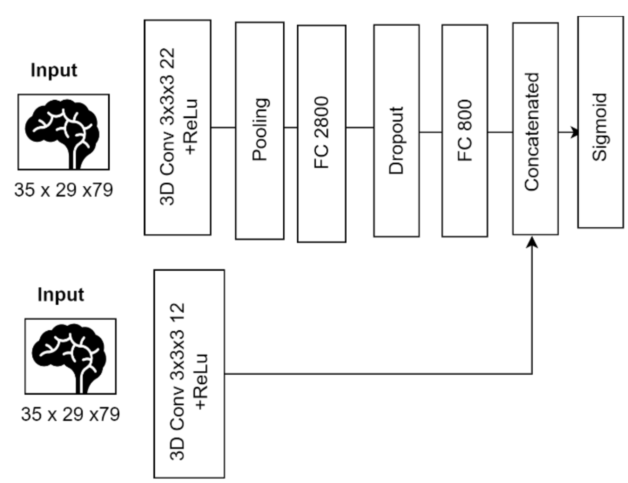

Fig 3. Non-sequential architecture

The Non-Sequential model used was from the idea of Wide and Deep Network developed for recommender system [8]. The architecture uses two inputs as in Figure 3 with just one convolution layer with more filters than the previous model the reason behind being this architecture is using two models a wide to learn general patterns and deep network to learn more complex patterns. The wide and deep part are trained separately. The convolution layer at the wider part was combined with outputs of dense layers resulting in a new feature vector that has combined information from both models and passed through sigmoid activation. The model was trained with a batch size of 20 for 5 epochs with early stopping.

3) Residual Network

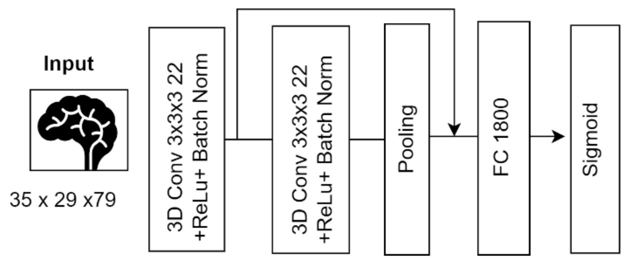

Fig 4. Residual network architecture

The model design (Figure 4) was taken inspiration from ResNet architecture developed for 3D MRI brain classification [4]. The model had just two convolution layers with skip connection added from input of first convolution layer to the output of residual block. This residual block has only one convolution layer as the general residual blocks tend to have two blocks this is to make the network less expensive to train. In this architecture, the addition of the residual block allows the model to extract more complex features from the input, while the residual connection allows the network to still propagate information from the original input through the network, which could help the network learn more efficiently.

IV. MODEL EVALUATION

The training and testing simulation were conducted on Intel i7 workstation with GPU Nvidia GTX 1080 Ti using TensorFlow Framework.

The impact of hyperparameters helped control the model from overfitting at its best. Typically, Res-Net and Plain Deep Neural Network architecture were at high risk of overfitting when compared to the non-sequential models this was observed during the training.

As we kept increasing the layers in sequential plain network, we were able to achieve 100% accuracy but had to trade off with very low validation accuracy. And the case was similar with the rest of the models. But the Res-Net never surfaced an accuracy of higher than 75 percent regardless of long hyper tuning with other parameters.

TABLE I
PERFORMANCE ACCURACY ESIMATED IN PERCENTAGE

METHOD	TRAIN	VALIDATION	TEST
Sequential	93.36	82.28	58.9
Non-Sequential	88.88	83.20	60.44
Res-Net	55.56	25.53	50

TABLE II
CONFUSION MATRIX COMPARISON ACROSS MODELS IN PERCENTAGE

METHOD	TN	FP	FN	TP
Sequential	34	16	25	25
Non-Sequential	43	7	32	18
Res-Net	0	50	0	50

TP=Correctly identified Psychosis
TN= Correctly identified Normal

Experiments were done with variation in batch sizes with lowest being 20 and highest being 180 with steps of 80 and interestingly we found that lower batch size provided us with decent results. This signals that making small gradient corrections might help in this case for this dataset.

For the evaluation purposes, considering risk of data leakage test data was separated and pre-processing was done on all the train, validation, and test sets. The test set pre-processing did not involve using any statistical measures to be used from train and validation set and is completely independent and this pre-processing was done to match the input size the networks are expecting.

TABLE III
COMPARISON OF RECOGNITION RATES ACROSS THE MODELS

METHOD	CLASSES	PRECISION	RECALL	F1-SCORE
Sequential	Class 0	0.58	0.67	0.62
	Class 1	0.61	0.51	0.55
Non-Sequential	Class 0	0.57	0.85	0.68
	Class 1	0.71	0.36	0.48
Res-Net	Class 0	0	0	0
	Class 1	0.5	0.1	0.67

Note:
Class 0 = Normal brain condition
Class 1 = Abnormal brain condition (Psychosis)

Table 1 shows that Sequential model, has high training accuracy but low validation and test accuracy, indicating overfitting. The non-Sequential model, has better training, validation, and test accuracy, suggesting less overfitting.

Custom ResNet has low training, validation, and test accuracy, indicating underfitting. Among the three models, non-sequential seems to have the best balance of training, validation, and test accuracy.

Further analysis of each model in Table 2, All the models are not performing well and are not able to identify underlying patterns. we can see that the Sequential model has a low precision and high false positive rate, while non-Sequential model has a higher precision, but a low recall. The Res-Net model has a perfect recall but a low precision.

Table 3 indicates that sequential model is not performing well for either class and has a low precision and recall for both classes. The non-Sequential model is performing well for predicting normal conditions of the brain (class 0), but not as well for psychosis condition (Class 1), with a low recall but high precision. And lastly, ResNet model is performing well for Class 0, but not as well for Class 1, with a low recall but high precision. Overall, the Non-Sequential model seems to be the best performing model among the three, as it has the highest F1-Score for Class 0, but it's not performing well for Class 1. This conclusion is based on standard convention given to class 0 but may differ based on the concern and context of interest.

This performance of this models was poor when compared to the other variants of Res-Net [1][6][7]. Though there were large differences in the way both simulations were conducted, quality, quantity of data being used these models validate the potential of being capable to be used for the classification tasks.

V. DISCUSSION AND CONCLUSIONS

The use of magnetic resonance imaging (MRI) to study brain disorders has been a promising research area in recent years. Structural alterations, such as variations in the shape or size of specific brain areas, can be spotted through MRI, and this can be useful in identifying potential indicators of psychosis. One approach that has been used to interpret MRI data is the use of machine learning models, such as 3D convolutional neural networks (CNNs), to classify brain images as healthy or diseased.

In this study, different neural network architectures were explored for the classification of brain MRI data, including Sequential, Non-Sequential, and ResNet. Before developing models, pre-processing steps such as cropping, normalization, and resizing were carried out on the dataset to improve efficiency along with Data augmentation to increase the size of the data.

Some of the significant findings in this exploration and comparison were, TensorFlow framework provided an easy implementation, the use of multiple GPUs for distributed training could have been easier as well. This experiment demonstrates the potential of wide and deep architectures and their generalisation coupled with state of art architectures could give better performance. The use of data augmentation played crucial role as the images of Psychosis conditions were quite less and the dataset was imbalanced.

Further studies need to be done to perform segmentations of the brain to improve the performance of the models and there is requirement for large dataset to get better generalisation. Experimenting the dataset with pretrained models could develop new pathways in research. It is worth noting that while structural brain imaging can be used to identify potential indicators of psychosis, a diagnosis of psychosis cannot be made solely based on these tests. A comprehensive assessment that includes examination of symptoms, medical history, and other relevant information is necessary to diagnose psychosis. Additionally, pre-processing and regularization techniques are essential to improve the performance of these models and ensure that the outcomes are applicable to fresh data. Robust models like Random Forest Classifiers can be used after getting the segmented parts of images and using them as features can give new directions in research as this model can be interpreted in much better way than neural networks, providing more insight into the underlying structure of the data and the feature importance.

The efficiency of 3D convolutional neural networks used on brain MRI data for early psychosis identification is examined in the paper. The results indicate that the performance of the adopted models, including sequential, non-sequential, and ResNet, was poor in general but was descent based on complexity of dataset, with wide and deep network maintaining good balance of trade off and test accuracy of 60%. The study also highlights potential for improvement in terms of overall performance and generalisability. Further research is needed to develop these models, test them on a broader range of subjects, and find ways to increase their performance. Such research could help in early detection of psychosis, which is critical to improve patients' health and treatment.

REFERENCES

[1] Z. Li et al., "Deep learning based automatic diagnosis of first-episode psychosis, bipolar disorder and healthy controls," Computerized Medical Imaging and Graphics, vol. 89, p. 101882, Apr. 2021

[2] K. Dang, T. Vo, L. Ngo, and H. Ha, "A deep learning framework integrating MRI image preprocessing methods for brain tumor segmentation and classification," *IBRO Neuroscience Reports*, vol. 13, pp. 523–532, Dec. 2022

[3] N. Maikusa, Y. Shigemoto, E. Chiba, Y. Kimura, H. Matsuda, and N. Sato, "Harmonized Z-Scores Calculated from a Large-Scale Normal MRI Database to Evaluate Brain Atrophy in Neurodegenerative Disorders," *Journal of Personalized Medicine*, vol. 12, no. 10, p. 1555, Sep. 2022

[4] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3D brain MRI classification," IEEE Xplore, Apr. 01, 2017

[5] A. ul Haq, J. P. Li, S. Khan, M. A. Alshara, R. M. Alotaibi, and C. Mawuli, "DACBT: deep learning approach for classification of brain tumors using MRI data in IoT healthcare environment," *Scientific Reports*, vol. 12, p. 15331, Sep. 2022

[6] R. Singh, A. Goel, and D. K. Raghuvanshi, "Computer-aided diagnostic network for brain tumor classification employing modulated Gabor filter banks," The Visual Computer, vol. 37, no. 8, pp. 2157–2171, Sep. 2020

[7] M. Talo, O. Yildirim, U. B. Baloglu, G. Aydin, and U. R. Acharya, "Convolutional neural networks for multi-class brain disease detection using MRI images," *Computerized Medical Imaging and Graphics*, p. 101673, Oct. 2019, doi: 10.1016/j.compmedimag.2019

[8] T. Nguyen, M. Raghu, and S. Kornblith, "Do Wide and Deep Networks Learn the Same Things? Uncovering How Neural Network Representations Vary with Width and Depth," *arXiv:2010.15327 [cs]*, Apr. 2021