Schizophrenia Detection from MR images

Ilia Golub[1] and Prosperity Oguama[1]

[1]*Gonda Interdisciplinary Brain Research Center, Bar-Ilan University, Ramat Gan, Israel*

1 Abstract

Background

Schizophrenia (SZ) is a chronic mental disorder that presents challenges in accurate diagnosis due to its complex nature. Advances in machine learning (ML) and deep learning (DL) have shown promise in automating SZ detection using structural MRI, but variability in preprocessing methods remains a key challenge. **Aim**: This study aims to evaluate the impact of various preprocessing and data augmentation techniques on ML-based classification of schizophrenia from MRI data.

Methods

We utilized the COBRE dataset, which includes structural MRI data from 62 individuals, divided into 30 schizophrenia patients and 32 healthy controls. A support vector classifier (SVC) was employed to classify the data. Various pre-processing techniques—normalization, brain extraction, and cropping—and data augmentation methods—translation, rotation, and Gaussian noise addition—were applied to the MRI images. Model performance was assessed using accuracy, recall, precision, and F1-score.

Results

Normalization and brain extraction improved classification accuracy, with cropping achieving the highest accuracy (0.57) and F1-score (0.69) among the preprocessing techniques. Combining preprocessing methods did not always result in performance improvements and sometimes led to a decrease in accuracy. Data augmentation significantly enhanced model performance, with translation and rotation leading to the highest accuracy (0.79) and F1-score (0.80). However, Gaussian noise had a detrimental effect, reducing accuracy and F1-score.

Conclusion

Our findings indicate that careful selection of preprocessing and augmentation techniques is critical for optimizing model performance in schizophrenia detection. While certain preprocessing steps like normalization and cropping are beneficial, combining them may not always yield additive improvements. Augmentation techniques like translation and rotation proved effective in boosting model performance, while Gaussian noise should be avoided. These results contribute to the development of more standardized preprocessing pipelines for medical imaging, which could enhance the reliability and reproducibility of machine learning models in clinical applications.

Keywords: schizophrenia, sMRI, CNN, SVC

2 Introduction

Schizophrenia (SZ) is a chronic mental disorder affecting 1 in 300 people worldwide, linked to structural brain abnormalities observable via MRI. Machine learning (ML) and deep learning (DL) have advanced SZ diagnosis, yet variability in preprocessing methods—such as noise removal and intensity normalization—remains a major challenge [1,5].

Despite progress in CNN-based SZ detection, studies often lack standardized preprocessing pipelines, impacting model reliability and reproducibility. Benchmark datasets [3,4] report high classification accuracies, but inconsistent

Table 1: The subjects were selected to give a balanced representation of age and gender distributions in the dataset, prioritizing age distribution, which has been shown to affect model performance (ref).

	COBRE		
	Individuals with SZ (n=30)	**Healthy controls (n=32)**	**Total (62)**
Minimum age	19	18	18
Maximum age	66	65	66
Average age	39.6	41	40.3
Gender (Male/ Female)	19/11	17/15	36/26

preprocessing limits comparability.

This project systematically evaluates preprocessing and augmentation techniques (e.g., smoothing, translation) to establish a robust pipeline for SZ detection from MRI. Standardizing these practices could improve the reproducibility and clinical applicability of ML-based diagnostics.

3 Methods

3.1 Dataset and Exploratory Data Analysis

The raw data were obtained from the SchizConnect database (accessed January 1, 2025), which houses structural and functional MRI data. The database provides filtering options to enable users to select data that meets specific criteria including MRI field strength and clinical diagnosis (e.g. schizophrenia, bipolar disorder).

The data obtained originates from the Center of Biomedical Research Excellence (COBRE) dataset in NifTI format. It includes data from 62 individuals, 30 with schizophrenia and 32 healthy controls.

The data from SchizConnect was filtered based on MRI field strength and clinical diagnosis. We retrieved only images captured with 3T MRI scans. This was done to ensure uniformity of the data obtained. Figure 1 presents the example of raw data.
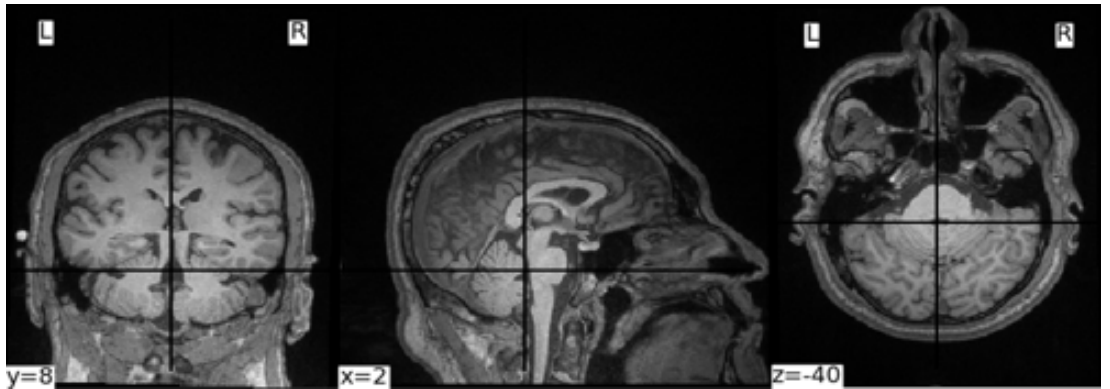

Figure 1: Three image slices representing the average of a 3D column in three views: dorsal, lateral, and superior respectively

Clinical data was made available in form of CSV files. The demographic features of the dataset used in this project are presented in Table 1.

3.2 Image Preprocessing

The major idea behind preprocessing was to establish a standardized pipeline that ensures consistency in the quality and content of the MRI data. This pipeline addressed common issues such as noise, intensity variations, and artifacts, which can affect the performance of machine learning models.

We experimented with different combinations of preprocessing and data augmentation techniques to determine which pipeline yields the best performance of the ML/DL models regarding recall rate, precision, and F1 score.

In order to find the best combination of hyperparameters for each preprocessing method, we performed a grid search and measured the quality metrics of the processed image using a custom test function.

Table 2 presents the preprocessing techniques used and their respective hyperparameters and tools used. Figure 2 gives a comparison of a raw image and a preprocessed image, also indicating the image metrics.
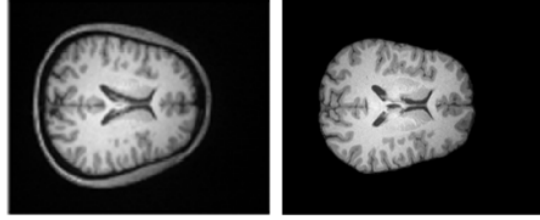

Figure 2: Panel (A) shows a slice of a raw scan. Panel (B) shows a slice of a preprocessed scan. For preprocessed scans, the standard deviation of SNR values is relatively low ($SNR = 25 \pm 1$), which ensures consistency in the image quality and prevents the model from making predictions based on possible artifacts.

3.2.1 Resampling

The first step of the pipeline is resampling the image to a standard resolution to ensure uniform input for the model and consistent noise and signal profiles.

3.2.2 intensity normalization

The second step of the pipeline is normalization of voxel intensities to a consistent range to maintain the relative scale of the data and produce values in a fixed range to satisfy models expectations. The adopted method scales the data to a range of [0, 1].

Normalization also improves the results of brain extraction by helping the algorithm better distinguish between brain and non-brain regions.

Intensity distributions were inspected to ensure normalization has not distorted the data.

3.2.3 Brain extraction

The third step of the pipeline is a brain extraction. The purpose of this stage is to

- remove non-brain regions in order to minimizes noise, making it easier for the network to focus on relevant features;

- subsequently enhance the network's performance, especially in a schizophrenia detection task where small differences in brain anatomy are crucial;

- ensure that the input images are more consistent in terms of content, which can help the model generalize better.

The generated brain mask was refined by closing gaps and removing small regions. The mask was also used as a leverage for the cropping step.

Table 2: Preprocessing techniques used in the study.

Preprocessing Step	Hyperparameters	Tools
Resampling	Voxel size: 2x2x2	Nibabel library
Intensity Normalization	Min-max scaling: [0, 1]	NumPy
Brain Extraction	Modality: T1, U-net	ANTsPyNet
Cropping	Size: based on the largest bounding box among all images	scikit-image, NumPy
Smoothing	Gaussian filter: sigma=0.5	SciPy

3.2.4 Cropping

The fourth step of the pipeline is cropping the brain region to a consistent size across scans.

3.2.5 Smoothing

The fifth step of the pipeline is smoothing the brain region to reduce noise and variability introduced by acquisition artifacts, enhancing the overall quality of the images. This was done to ensure that the downstream tasks are more robust and the model training is more effective. We used a Gaussian filter to reduce high-frequency noise while preserving edges to some extent. Since the neural network performed a classification task, moderate smoothing was carried out to focus on broader patterns.

The sixth step of the pipeline is normalization of the smoothed image to address potential intensity changes and ensure that all images are on a comparable scale.

3.2.6 Validation

The impact of preprocessing on the data's integrity and quality was assessed using a combination of a visual inspection and quantitative metrics. The visual inspection involved plotting a few slices of the MRI before and after each preprocessing step. Validation was performed to ensure that preprocessing steps did not introduce artifacts or distort anatomical structures, and to ensure consistency of the processed images.

Quantitative metrics included Signal-to-Noise Ratio (SNR), Contrast-to-Noise Ratio (CNR), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM).

Relative PSNR for one image. Using a hypothetical "ideal" signal (the maximum intensity value as a baseline), a relative PSNR for one image was calculated. This relative PSNR was used as a way to assess the "quality" of a single image compared to a theoretical perfect image of uniform intensity.

A binary mask was computed using Otsu's thresholding to explicitly select a background region for noise estimation.

The metrics were computed for the entire 3D volume to assess the overall quality or similarity of the entire MRI dataset, capturing global differences.

3.3 Data Augmentation

According to typical variations in brain MRI scans, the following transformations with corresponding parameters were applied to the images:

Option 1 Translation. The operation is applied on the x and y-axis and repeated 10 times by using different values sampled randomly within the range $[-10, 10]$. By this method, 10 images are generated from an original image.

Option 2 Rotation. Rotation angles are selected randomly within the range $[-10°, 10°]$ and clockwise rotations are applied using 10 different angle values. 10 images are generated from an original image by this method.
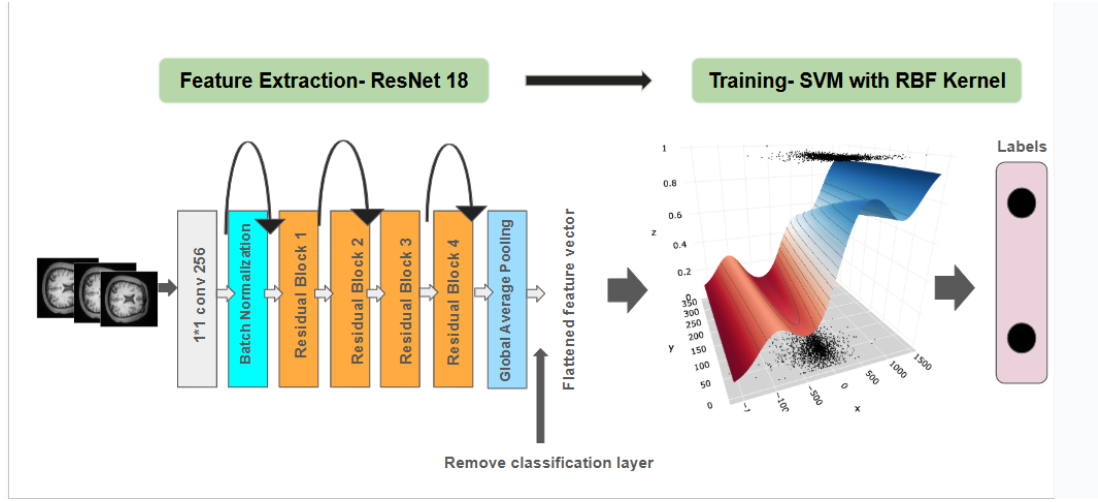
Figure 3: Training architecture and flow.

Option 3 Gaussian noise addition. Noise addition is applied by $I = I + R * I * std + mean$, where I refers to the image, R refers to the Noise. The noise to be added is generated using a Gaussian distribution. The variance of the distribution is computed from the input images while the mean value is set to zero. In this study, three different fixed values (0.1, 0.2, and 0.3) are used and 3 images are generated from an input image.

The augmented samples were plotted to ensure the transformations don't distort anatomical features.

3.4 Model architectures

Convolutional Neural Networks (CNNs) are commonly used in the literature for this task. However, a few papers explored the use of support vector machines for image classification. To gain a full understanding of both approaches, we decided on the combination of a pretrained deep neural network for feature extraction and a support vector classifier for binary classification. The ReNet-18 architecture was selected for its balance between computational load and accuracy. Cross validation was performed with the support vector classifier (SVC) to choose the best hyperparameters, and the hyperparameters that yielded the best results were the radial basis function (RBF) kernel, a gamma value of 0.0001 and a C parameter value of 100. Each 2D slice of the images were fed into the ResNet-18 model as NumPy arrays of dimension `224x224`. This input size corresponds to the size of the images in the ImageNet dataset that was used to train ResNet-18 and was selected for compatibility with the pretrained weights of the ResNet-18 model. Figure 3 gives a representation of the training architecture and flow.

3.5 Training and validation

The dataset was split into a training set (80%) and a validation set (20%).

Accuracy, recall, precision, and F1-score were calculated to asses the performance of the SVC.

4 Results

4.1 Image Preprocessing

The results presented in Table 3 demonstrate the impact of various preprocessing methods on classification performance. The baseline model using raw images (0) achieved an accuracy of 0.51, with an F1-score of 0.63. Applying resampling (1) alone did not alter these results, suggesting that resampling alone does not introduce substantial benefits.

Table 3: Desct

Preprocessing method	Accuracy	Recall	Precision	F1-score
Raw (0)	0.51	0.67	0.60	0.63
Resampling (1)	0.51	0.67	0.60	0.63
Normalization (2)	0.56	0.69	0.64	0.67
Brain extraction (3)	0.54	0.78	0.60	0.68
Cropping (4)	0.57	0.76	0.63	0.69
Smoothing (5)	0.55	0.71	0.62	0.66
1+2	0.57	0.69	0.64	0.67
1+2+3	0.55	0.80	0.61	0.69
1+2+3+4	0.44	0.53	0.56	0.54
1+2+3+4+5	0.48	0.60	0.58	0.59

Table 4: Desct

Augmentation method	Accuracy	Recall	Precision	F1-score
Raw (0)	0.51	0.67	0.60	0.63
Translation (1)	0.75	0.84	0.74	0.79
Rotation (2)	0.79	0.81	0.78	0.80
Gaussian Noise (3)	0.33	0.32	0.53	0.40

Normalization (2) improved accuracy to 0.56 and F1-score to 0.67, indicating that standardizing intensity values enhances model performance. Brain extraction (3) yielded a slightly lower accuracy (0.54) but a notable increase in recall (0.78), which suggests that removing non-brain tissues helps capture more positive instances but may introduce a trade-off in precision.

Cropping (4) led to the highest accuracy (0.57) and an F1-score of 0.69, highlighting its effectiveness in improving signal-to-noise ratio. Smoothing (5) showed a moderate impact, improving results compared to the raw data but not outperforming other methods.

When combining preprocessing techniques, the pattern is less straightforward. The combination of resampling, normalization, and brain extraction (1+2+3) resulted in the highest recall (0.80) while maintaining a strong F1-score (0.69). However, adding cropping (1+2+3+4) caused a drastic performance drop, with accuracy declining to 0.44 and F1-score to 0.54. The full pipeline (1+2+3+4+5) slightly improved over this but still underperformed compared to individual preprocessing steps.

4.2 Data Augmentation

The results presented in Table 4 highlight the impact of various data augmentation methods on model performance. The baseline model using raw data (0) achieved an accuracy of 0.51 and an F1-score of 0.63. In contrast, applying translation (1) significantly improved both accuracy (0.75) and F1-score (0.79), with a notable increase in recall (0.84), indicating that translation is effective at boosting model performance and sensitivity.

Rotation (2) further improved results, with the highest accuracy (0.79) and F1-score (0.80), alongside strong recall (0.81) and precision (0.78), demonstrating its ability to enhance both model robustness and overall performance.

Gaussian noise (3), however, led to a substantial decrease in performance, with an accuracy of only 0.33 and a low F1-score of 0.40. This suggests that adding noise may negatively affect the model, potentially obscuring relevant patterns in the data.

5   Discussion

This study investigated the effects of various preprocessing and data augmentation techniques on the classification performance of schizophrenia from structural MRI data. The results demonstrate that certain preprocessing steps, such as normalization, brain extraction, and cropping, can improve model performance when applied individually. However, combining these steps did not yield consistent improvements, suggesting that over-processing may distort the data and hinder the model's ability to generalize effectively.

Among the individual preprocessing techniques, normalization resulted in the most substantial performance improvement, enhancing accuracy by 9.8%. This can be attributed to the equalization of feature contributions, which allowed the model to make more accurate predictions. Interestingly, this outcome contradicts the conventional understanding that normalizing medical images may flatten clinically significant differences in intensity. The effectiveness of this technique in this study highlights the importance of context when applying preprocessing methods to medical imaging.

Brain extraction, or skull stripping, while beneficial in other studies, had a moderate effect here, improving recall but slightly lowering accuracy. This suggests that removing non-brain tissues enhances the model's sensitivity to schizophrenia cases but may reduce precision, potentially due to the loss of relevant contextual information in the brain image. Cropping, on the other hand, demonstrated a clear advantage by improving accuracy and F1-score. By focusing on the brain region and reducing irrelevant background data, cropping effectively improved the signal-to-noise ratio, allowing the model to focus on clinically relevant features.

However, combining preprocessing steps often led to diminished performance, especially when all techniques were applied together. The full preprocessing pipeline, including resampling, normalization, brain extraction, cropping, and smoothing, resulted in a performance drop, particularly in terms of accuracy and F1-score. This suggests that certain preprocessing steps, when combined, may interfere with each other, creating distortions that prevent the model from learning the most pertinent features of the data.

Regarding data augmentation, both translation and rotation significantly enhanced model performance, with translation improving recall and rotation achieving the highest accuracy and F1-score. These results align with the literature, where such transformations are often used to increase the diversity of the training data, helping the model generalize better to unseen data. Gaussian noise, however, severely degraded performance, suggesting that it may obscure important features in the MRI images, leading to poorer model predictions.

While preprocessing techniques like skull stripping, cropping, and smoothing have been shown to improve performance in various medical imaging tasks, the results in this study suggest that their effectiveness depends on the context and the combination of other preprocessing steps. Furthermore, the adverse effect of Gaussian noise indicates the need for careful consideration of the types of augmentation applied, as certain methods may introduce detrimental effects.

In contrast to some previous studies that have used more complex augmentation techniques, such as elastic deformations, this study focused on simpler methods, which proved to be more effective. Elastic deformations can sometimes introduce unrealistic and unnatural variations, leading to damage and noise that harm model performance [2]. The methods investigated in this study, particularly translation and rotation, were found to be effective in maintaining the integrity of the original images while enhancing model robustness.

Overall, this study highlights the importance of standardizing preprocessing and augmentation techniques in medical imaging, particularly for tasks like schizophrenia detection. While certain methods demonstrated clear benefits, the combination of multiple preprocessing steps should be approached with caution, as it may lead to diminishing returns or reduced model performance. Further research is needed to explore the optimal combination of preprocessing techniques and to investigate the impact of more advanced augmentation methods on model accuracy and generalizability.

6   Conclusion

In this study, we investigated the impact of various preprocessing and data augmentation techniques on the classification performance of schizophrenia from structural MRI images. Our results demonstrate that individual preprocessing methods, particularly normalization, brain extraction, and cropping, can positively influence model performance, improving accuracy, recall, and F1-score. However, combining these techniques did not always lead to enhanced results, suggesting that over-processing may introduce distortions that hinder model generalization.

Data augmentation proved to be a valuable tool, with translation and rotation significantly improving model robustness and performance, particularly in recall and accuracy. Conversely, adding Gaussian noise resulted in a substantial performance decrease, indicating that it may obscure critical features in the data.

The findings emphasize the importance of careful selection and application of preprocessing and augmentation methods in medical imaging tasks. While some techniques show clear benefits, the combination of multiple preprocessing steps must be handled cautiously to avoid diminishing returns. These results underscore the need for a standardized preprocessing pipeline to enhance the reliability, reproducibility, and clinical applicability of machine learning models for schizophrenia detection.

Further work is needed to refine preprocessing pipelines and explore more advanced data augmentation methods to improve the performance of ML-based schizophrenia diagnosis. Ultimately, this study contributes to the ongoing efforts to make ML-driven medical diagnostics more robust and applicable in real-world clinical settings.

Author contributions

Conception: Prosperity Oguama, Ilia Golub; **Data Acquisition**: Ilia Golub; **Preprocessing & Data Augmentation**: Ilia Golub; **Model Building & Testing**: Prosperity Oguama; **Drafting & Editing**: Ilia Golub, Prosperity Oguama. All authors contributed to the study, read, revised and approved the final manuscript.

Data availability statement

The data underlying this article is publicly available and accessible through http://schizconnect.org/.

Financial disclosure

The authors report no financial disclosure.

Conflict of interest

The authors declare no potential conflict of interests.

References

1. S. Gengec Benli and M. Andac. Constructing the schizophrenia recognition method employing glcm features from multiple brain regions and machine learning techniques. *Diagnostics*, 13(13), 2023.

2. T. C. W. Mok and A. C. S. Chung. Learning data augmentation for brain tumor segmentation with coarse-to-fine generative adversarial networks. page 70–80, 2019.

3. J. Oh, B.-L. Oh, K.-U. Lee, J.-H. Chae, and K. Yun. Identifying schizophrenia using structural mri with a deep learning algorithm. *Frontiers in Psychiatry*, 11, 2020.

4. S. Verma, T. Goel, M. Tanveer, W. Ding, R. Sharma, and R. Murugan. Machine learning techniques for the schizophrenia diagnosis: A comprehensive review and future research directions, 2023.

5. J. Zhang, V. Rao, Y. Tian, Y. Yang, N. Acosta, Z. Wan, P.-Y. Lee, C. Zhang, L. Kegeles, S. Small, and J. Guo. Detecting schizophrenia with 3d structural brain mri using deep learning. 06 2022.