



A deep learning approach for mental health quality prediction using functional network connectivity and assessment data

Meenu Ajith¹ · Dawn M. Aycock² · Erin B. Tone³ · Jingyu Liu^{4,7} · Maria B. Misiura^{3,5} · Rebecca Ellis⁶ · Sergey M. Plis¹ · Tricia Z. King^{3,7} · Vonetta M. Dotson^{8,9} · Vince Calhoun¹

Accepted: 16 January 2024 / Published online: 10 February 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

While one can characterize mental health using questionnaires, such tools do not provide direct insight into the underlying biology. By linking approaches that visualize brain activity to questionnaires in the context of individualized prediction, we can gain new insights into the biology and behavioral aspects of brain health. Resting-state fMRI (rs-fMRI) can be used to identify biomarkers of these conditions and study patterns of abnormal connectivity. In this work, we estimate mental health quality for individual participants using static functional network connectivity (sFNC) data from rs-fMRI. The deep learning model uses the sFNC data as input to predict four categories of mental health quality and visualize the neural patterns indicative of each group. We used guided gradient class activation maps (guided Grad-CAM) to identify the most discriminative sFNC patterns. The effectiveness of this model was validated using the UK Biobank dataset, in which we showed that our approach outperformed four alternative models by 4–18% accuracy. The proposed model's performance evaluation yielded a classification accuracy of 76%, 78%, 88%, and 98% for the excellent, good, fair, and poor mental health categories, with poor mental health accuracy being the highest. The findings show distinct sFNC patterns across each group. The patterns associated with excellent mental health consist of the cerebellar-subcortical regions, whereas the most prominent areas in the poor mental health category are in the sensorimotor and visual domains. Thus the combination of rs-fMRI and deep learning opens a promising path for developing a comprehensive framework to evaluate and measure mental health. Moreover, this approach had the potential to guide the development of personalized interventions and enable the monitoring of treatment response. Overall this highlights the crucial role of advanced imaging modalities and deep learning algorithms in advancing our understanding and management of mental health.

Keywords Mental health · rs-fMRI · Deep learning · Assessment data

Introduction

Resting-state functional MRI (rs-fMRI) has become one of the most widely used modalities for analyzing functional links to mental health in the human brain. By analyzing differences in brain connectivity patterns using rs-fMRI, researchers can gain insight into the neural substrates of mental health and potentially identify biomarkers for healthy brain function (Goulas & Margulies, 2021). Studies have demonstrated that the efficacy of rs-fMRI network-based classification can be significantly improved using deep learning techniques (Li et al., 2021). These advancements

have opened the possibility of applying this classification approach to the fast and objective diagnosis of mental conditions such as major depressive disorder (Uyulan et al., 2021), schizophrenia (Liu et al., 2022), anxiety disorder (Al-Ezzi et al., 2021), bipolar disorder (Cheng et al., 2022) and post-traumatic stress disorder (Saba et al., 2022). Recent studies use rs-fMRI and deep learning to predict cognitive decline in healthy aging individuals (Chen et al., 2021), showing potential for individualized interventions aimed at promoting and maintaining mental health.

The majority of current clinical criteria for determining the severity of mental health symptoms rely on the subjective assessment of the patient's symptoms and self-reported medical history. More recently, predictive modeling based on machine learning (ML) has been used to interpret neuroimaging data to determine symptom severity or cognitive

✉ Meenu Ajith
majith@gsu.edu

Extended author information available on the last page of the article

impairment. The neuroimaging field has shown a rising interest in ML technologies due to challenges in integrating an enormous amount of information in neuroimaging scans. Supervised ML algorithms are mathematical models created to identify patterns in known data and use that information to predict patterns in new data. This approach can be applied to clinical populations, including individuals with mental health conditions. For example, functional connectivity (FC) between brain regions measured by rs-fMRI in depressed individuals demonstrate distributed variations across the entire brain (Craddock et al., 2009). Another study (Zeng et al., 2012) utilizing FCs and linear support vector machines (SVM) achieved an accuracy of 94% while classifying between patients with depression and healthy controls. Similarly, a statistical machine learning method such as partial least squares (PLS) regression was used to predict different clinical measures, such as the Positive and Negative Affect Schedule (PANAS), Beck Depression, Inventory-II (BDI-II), Snaith-Hamilton Pleasure Scale (SHAPS), and age from functional connectivity data (Yoshida et al., 2017). These predicted clinical scores were further used to classify the depressed patients from healthy controls with 80% accuracy.

Nevertheless, recent advances in deep learning approaches show that, particularly for complex high-dimensional datasets such as fMRI data, the deep models show a significant improvement in performance over standard ML models (Su et al., 2020). Deep learning algorithms may be trained to recognize abnormalities in fMRI data that are linked with certain mental health problems, and these features can then be used to identify the existence or intensity of a mental health condition. It can also be used to produce tailored treatment options for people with mental health concerns in addition to identifying and predicting them. Furthermore, deep learning models have been successfully utilized on raw fMRI data to perform classification tasks such as detecting distinct brain states or conditions (Riaz et al., 2018). Convolutional neural networks (CNN) are amongst the most widely used deep learning models for connectome-based classification and this is particularly significant given how well CNN performs in image classification as well as object recognition (Kawahara et al., 2017). Yet, the accessibility of a significant number of training samples is a crucial need for deep learning approaches. Hence, very basic CNN models should be constructed for fMRI-based applications, in accordance with the quantity of data that is accessible.

This paper focuses on classifying participants into different mental health categories based on sFNC data from rs-fMRI. The self-reported behavioral measures of mental health from the UK Biobank were aggregated to obtain a mental health score for each subject. These mental health scores were subjected to Gaussian mixture model (GMM) clustering to obtain the optimum number of categories

for classification. Finally, the labels for classification were obtained after placing the participants into four different classes such as excellent, good, fair, and poor mental health. Following this, the sFNC features were input into a one-dimensional convolutional neural network (1D-CNN) to extract useful connectivity parameters for categorizing mental health. The key contributions of this study are as follows: (1) the novel method used a combination of neuroimaging data and a set of self-reported assessment data on questions related to mental health to provide a flexible prediction of mental health quality; (2) the automatic computation of sub-categories of mental health quality in any population; (3) interpreting the deep learning model by identifying salient regions in the sFNC associated with each mental health quality category; (4) improved generalization and robustness by training and optimizing the deep learning model on a large dataset; and (5) the model also demonstrated superior performance when compared to other state-of-the-art machine learning algorithms.

Methods

Participants

The data for this analysis were acquired from the UK Biobank database (Miller et al., 2016). The sample comprised 34606 participants, whose ages ranged from 53 to 87 (69.75 ± 7.43) years as shown in Table 1. Participants included 19120 females (53.1%) and 16880 males (46.8%).

fMRI data acquisition and preprocessing

A 32-channel head coil 3-Tesla (3T) Siemens Skyra scanner was used to scan all the participants. Next, resting-state

Table 1 Demographic information from the UKBioank database

Characteristics	Number
Participants, n	34,606
Age (years), mean (SD)	69.75 (7.43)
Male, n (%)	16,880 (46.8)
Female, n (%)	19,120 (53.1)
Fed up feeling	
Yes	12,134 (33.7)
No	22,472 (62.4)
Frequency of unenthusiasm and disinterest in the last 2 weeks	
Not at all	28,643 (79.5)
Several days	4896 (13.6)
More than half the days	617 (1.7)
Nearly everyday	450 (1.2)

fMRI images were obtained using a gradient-echo echo planar imaging (GE-EPI) technique. The acquisition parameters consist of no iPAT, fat saturation, flip angle (FA) = 52°, spatial resolution = $2.4 \times 2.4 \times 2.4$ mm, field-of-view (FOV) = (88 × 88 × 64 matrix), repeat time (TR) = 0.735 s, echo time (TE) = 39 ms and 490 volumes. Also, eight slices were acquired concurrently and hence the multiband acceleration factor was set to eight. During the 6-minute and 10 s resting-state scanning phases, participants were asked to passively look at a crosshair and stay relaxed.

We conducted several preprocessing procedures on the UK Biobank database. To reduce the effects of subject-specific motion, we used MCFLIRT (Jenkinson et al., 2002), an intra-modal motion correction tool. To evaluate brain scans among participants, we employed grand-mean intensity normalization to scale the full 4D dataset by a single multiplicative factor. Further, to eliminate residual temporal drifts, we filtered the data with a high-pass temporal filter and rectified geometric aberrations using FSL's Topup tool (Andersson et al., 2003). After EPI unwarping, we employed a gradient distortion correction (GDC) unwarping stage. Next, structural artifacts were eliminated using Independent Component Analysis (ICA) along with FMRIB's ICA-based X-noiseifier (Salimi-Khorshidi et al., 2014). Furthermore, the data were standardized to an MNI EPI template with FLIRT, succeeded by SPM12. Finally, the data were smoothed with a Gaussian filter with a full width at half maximum (FWHM) of 6 mm.

Following preprocessing, we applied a completely automated spatially constrained independent component analysis (ICA) using the NeuroMark (Du et al., 2020) technique on the resting state-fMRI data. This utilizes an adaptive-ICA technique, such as group information guided ICA (GIG-ICA) (Du & Fan, 2013) or spatially constrained ICA (Lin et al., 2010), for automating the estimation and labeling of connectivity features specific to individual subjects. First, independent components (ICs) were calculated using two large-sample healthy control datasets (HCs). Second, replicable intrinsic connection networks (ICNs) were obtained by comparing and evaluating the spatial maps of ICs from various datasets. The highly replicated ICNs were then used as network templates in an adaptive ICA technique to automatically estimate subject-specific functional networks and related time courses (TCs). Through the application of NeuroMark, various network features are accessible, encompassing spatial functional networks, inter-network functional connectivity, graph measures of functional organization, and frequency information concerning network fluctuations, viewed from both static and dynamic standpoints. For instance, in the case of functional network connectivity (FNC), the sFNC can be derived by computing Pearson correlations between time courses (TCs) of intrinsic connectivity networks (ICNs), resulting in an sFNC matrix that reflects the interactions

between any two networks. According to their functional and anatomical characteristics (Allen et al., 2014), the top 53 ICNs were classified into seven functional domains: subcortical (SC: 5 ICNs), auditory (AUD: 2 ICNs), sensorimotor (SM: 9 ICNs), visual (VIS: 9 ICNs), cognitive control (CC: 17 ICNs), default mode (DM: 7 ICNs), and cerebellar (CB: 4 ICNs). We used the sFNC as input to our model and as the basis for all subsequent analyses in this study. The ICNs that were used as network templates are represented by an IC, accompanied by its corresponding functional domain, primary brain region, and peak coordinate (Du et al., 2020). We used the sFNC as input to our model and as the basis for all subsequent analyses in this study.

Mental health category identification

Self-reported questionnaires, while valuable for capturing subjective experiences, are susceptible to various biases that can compromise data reliability and validity. Social desirability bias (Bispo Júnior, 2022) prompts respondents to portray themselves favorably with societal norms by underreporting socially undesirable behaviors. Simultaneously, it leads to overstating positive behaviors, potentially skewing the true prevalence of certain attitudes or actions. Response set bias (McGee Ng et al., 2016) introduces consistent responses irrespective of the context, like consistently opting for extreme or neutral choices, impacting the accuracy and reliability of the answers provided. Language and cultural differences can lead to misunderstandings or misinterpretations, particularly in diverse populations, affecting response consistency and comparability across groups. Additionally, the restricted ability to capture contextual details and the challenge of measuring gradual changes over time also limit the effectiveness of these questionnaires, especially in assessing complex constructs like personality traits. These biases underscore the importance of rigorous questionnaire design, pretesting across diverse demographics, and considering cultural details to enhance the reliability and validity of data derived from self-reported questionnaires.

However, despite their potential for introducing subjective biases and inaccuracies, self-reported questionnaires play a vital role in capturing individuals' subjective experiences and perceptions. Objective clinical assessments alone often fail to capture these subjective aspects. While their correlation with neuroimaging findings can pose challenges, the UKB's substantial sample size and diverse array of phenotypic, imaging, and biological measurements provide a unique opportunity to address these concerns. By utilizing the extensive sample size of the UK Biobank, researchers can account for and mitigate potential biases arising from individual variations. The inclusion of a large and diverse population allows for a more comprehensive analysis that can help identify and control confounding factors, thereby enhancing the

robustness and reliability of the findings. In the case of mental health, the recent depressive symptoms-based questions from the UK Biobank are recommended for imaging-based research and represent a more robust metric for evaluating depressive symptoms. Additionally, despite the small effect sizes for individual imaging-derived phenotypes, the multivariate associations between brain imaging-derived phenotypes and mental health suggest a meaningful relationship between brain biomarkers and mental health outcomes in the UK Biobank database. Also, including self-reported measures facilitates longitudinal studies, enabling the tracking of changes in mental health over time within the same individual.

The self-reported questionnaires from 34606 participants were collected from the UKBiobank database for creating the labels. While the UK Biobank contains mental health data from different sources, we focus on using the assessment center questions completed using the touch screen on the day of the scan. Table 2 shows the 20 questions and the corresponding responses for each question. Here we normalized the responses to the questions so that the range is from 0 to 1. For instance, in the case of mood swings, '0' corresponds to no mood swings, and '1' corresponds to having mood swings. Whereas in the case of frequency of depressed mood in the last 2 weeks, '0' denotes not at all, '0.33' denotes several days, '0.67' denotes more than half the days, and '1' denotes nearly every day of depressed mood.

The first 12 questions included in the table enable the calculation of the Eysenck Neuroticism (N-12) score. Individuals with high neuroticism scores are more prone to negative moods and to experience sensations such as anxiety, worry, fear, wrath, frustration, and loneliness. As a result, people with high neuroticism scores are regarded to be at risk of developing mood disorders, anxiety disorders, and substance use disorders (Barlow et al., 2021). On the other hand, questions 16–19 reflect recent depressive symptoms (RDS-4), a continuous measure of depression symptom severity acquired at the time of scanning. The RDS-4 uses four self-report questions to measure low mood, indifference, restlessness, and weariness. Each question inquires about current symptom occurrences, especially within the past 2 weeks. The four response alternatives are: not at all, several days, more than half the days, and practically every day. In comparison to N-12, RDS-4 assesses the current state of depressed symptoms, whereas N-12 assesses personality traits. Later Smith and colleagues created a categorical (case-control) measure of the lifetime incidence of depression using questions from the evaluation data (Smith et al., 2013). This was represented using questions 14 and 15 and they served as an indication of the subject's probable depressive status. Nevertheless, these questions did not distinguish between isolated and recurring depressive episodes. For instance, if the participants indicated they had seen a doctor or a psychiatrist for nerves, worry, stress, or depression, their depression status was set to 1.

Table 2 The questions and responses related to mental health that are used in this study from the UKBioank database

No.	Mental health questionnaire	Responses
1.	Mood swings	0, 1
2.	Miserableness	0, 1
3.	Irritability	0, 1
4.	Sensitivity/hurt feeling	0, 1
5.	Fedup feeling	0, 1
6.	Nervous feeling	0, 1
7.	Worrier anxious feeling	0, 1
8.	Tense/highly strung	0, 1
9.	Worry too long after embarrassment	0, 1
10.	Suffer from nerves	0, 1
11.	Loneliness/isolation	0, 1
12.	Guilty feeling	0, 1
13.	Risk taking	0, 1
14.	Seen a doctor/gp for nerves, anxiety, tension or depression	0, 1
15.	Seen a psychiatrist for nerves, anxiety, tension or depression	0, 1
16.	Frequency of depressed mood in last 2 weeks	0, 0.33, 0.67, 1
17.	Frequency of unenthusiasm disinterest in last 2 weeks	0, 0.33, 0.67, 1
18.	Frequency of tenseness restlessness in last 2 weeks	0, 0.33, 0.67, 1
19.	Frequency of tiredness lethargy in last 2 weeks	0, 0.33, 0.67, 1
20.	Illness, injury, bereavement, stress in last 2 years	0, 0.17, 0.33, 0.5, 0.67, 1

In this study, we selected 20 questions for each subject based on previous studies conducted in UK Biobank associated with mental health (Dutt et al., 2022). A mental health score was calculated for each subject by summing their normalized responses to the questions. Here the maximum possible score for any subject is 20 and the minimum score is 0. A histogram of the mental scores over the 34606 participants is shown in Fig. 1. According to the histogram, the primary conclusion is that participants with low mental health scores have excellent mental health while those who score closer to 20 have poorer mental health. In this problem, the number of categories of mental health quality is not predefined. Hence we use the Gaussian Mixture Model clustering (Fraley & Raftery, 2002) method to automatically find the different groups present in the data.

GMMs are unsupervised probabilistic models that follow the assumption that all data points are generated from a fixed set of Gaussian distributions. This approach distributes data points into distinct groups using the soft clustering technique. Multiple Gaussian distributions are fitted to the data and the distribution parameters such as mean, variance, and weight are calculated for each cluster. The probability of each data point belonging to a cluster is determined after learning these parameters. The univariate mental health score data are from a normal distribution with mean μ and variance σ^2 . Expectation maximization (Dempster et al., 1977) is used to estimate the mixture model's parameters when the number of clusters is known. This is an iterative strategy with the property that the maximum likelihood of the data strictly rises with each additional iteration. There are two phases in the expectation-maximization process. Initially, the mean and variances are assigned randomly. Next, the posterior probability that each data point belongs to a cluster is determined in the expectation phase using the current mean and variances. The cluster means and variances are recalculated in the maximization

stage using the probability obtained in the expectation step. The steps are repeated to get a maximum likelihood estimate until the algorithm's convergence.

We used the Bayesian Information Criterion (BIC) (Schwarz, 1978) to determine the ideal number of clusters. The BIC compares the maximum likelihood function with the number of model parameters, k , and data points, n . Adding a penalty for the number of model parameters allows it to choose the model with the fewest parameters that best describe the data. Thus the model with the lowest BIC score is selected from a finite set of models. The BIC score is calculated based on the number of parameters k , the sample size n , and the log-likelihood function. Analysis of the mental health scores resulted in four clusters. The labeling of the clusters was as follows: excellent (score range = 0 to 3), good (score range = 4 to 7), fair (score range = 8 to 11), and poor (score range = 12 to 20). We calculated the range of the mental health quality groups using the mean and standard deviation of their associated clusters. The maximum value of the range is given by adding the mean with twice the standard deviation, whereas the minimum value of the range is obtained by subtracting twice of standard deviation from the mean.

Predictive modeling using functional network connectivity data

The prediction network for the sFNC data is designed using a one-dimensional convolutional neural network (1D-CNN) (Krizhevsky et al., 2017; LeCun et al., 2015). Our main goal was to predict mental health quality categories using the sFNC data from the UK Biobank database. The features consist of 53×53 sFNC matrices that represent inter-connectivity strengths between various ICNs. The corresponding labels are the mental health scores we computed from the self-report data.

The proposed CNN is a 16-layer network with four convolutional layers with a kernel size of three and with 16, 32, 64, and 128 filters respectively. The rectified linear unit (ReLU) (Nair & Hinton, 2010) non-linearity is used in the convolutional layers. The three fully connected layers towards the end of the model have 64, 16, and 4 nodes and on the output layer, the softmax activation function (Bridle, 1990) is used to determine the likelihood that each sample belongs to a class. The model uses four max-pooling layers with kernel size two to decrease the dimensionality of the feature maps and limit overfitting. Also, a drop-out regularization (Srivastava et al., 2014) with a probability of 0.2 and 4 layers of batch normalization was added for regularization. The Adam optimizer (Kingma & Ba, 2014) was used to train the proposed CNN for 150 iterations, with a learning rate of 0.001 and a batch size of eight. The data set underwent five-fold cross-validation in which four folds were used for training and the

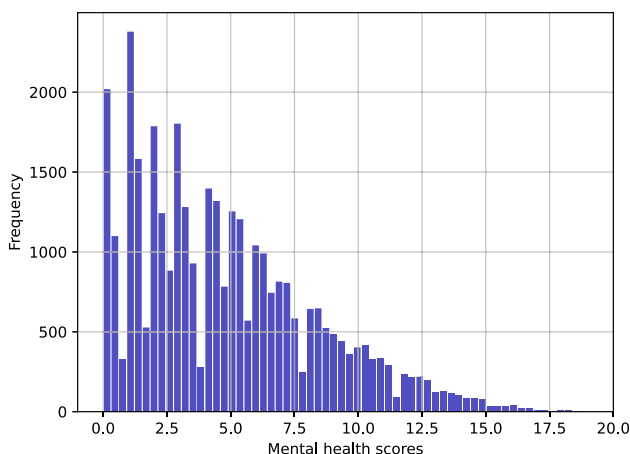


Fig. 1 Histogram of the mental health scores for 36,000 participants from the UKBiobank database

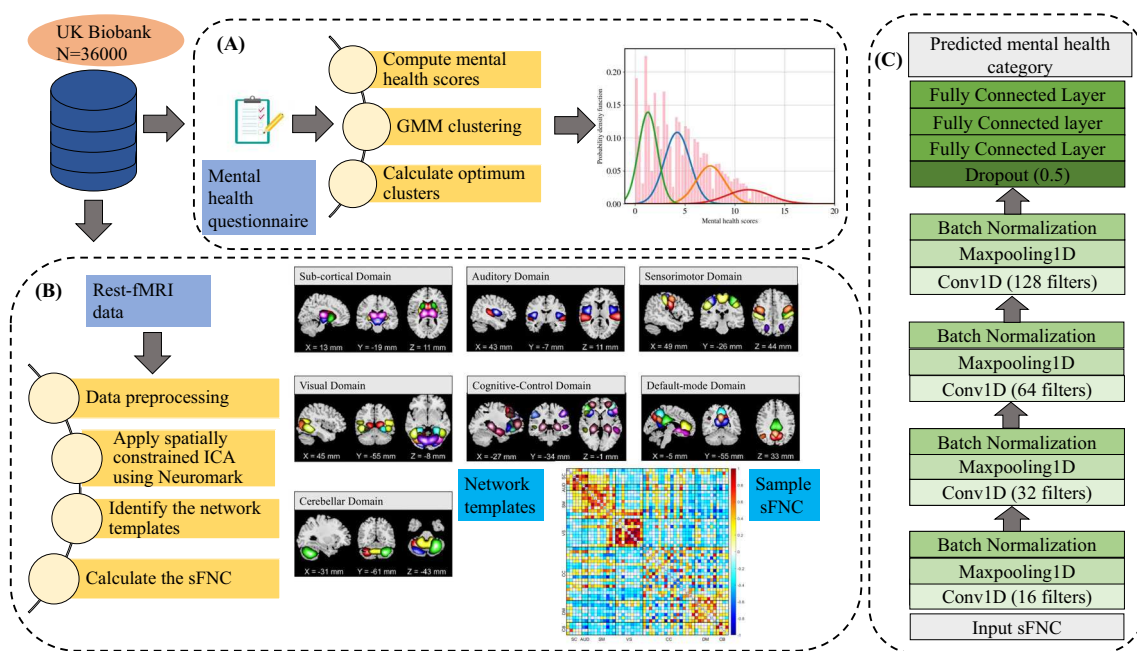


Fig. 2 The proposed deep learning architecture for mental health category prediction

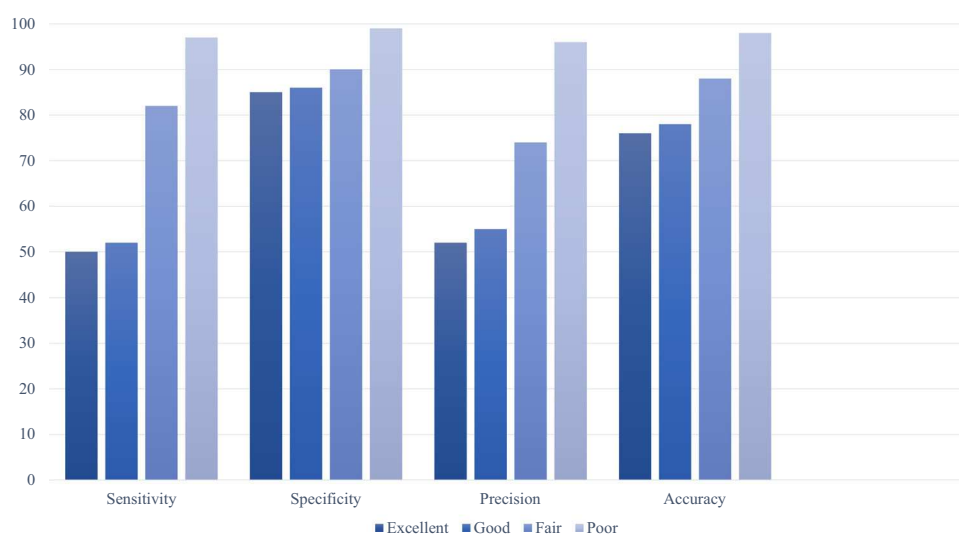
remaining fold was used for testing. The excellent, good, fair, and poor categories consisted of 16256, 11845, 5058, and 1447 participants respectively. Since the data for the four classes was imbalanced, the Synthetic Minority Over-sampling Method (SMOTE) (Chawla et al., 2002) was used to construct a balanced dataset. This is a method of over-sampling the minority class by producing synthetic samples rather than oversampling using replicated actual data values. In accordance with the amount of over-sampling needed, neighbors from the k nearest neighbors of a minority class are picked at random. Figure 2 (C) illustrates the architecture of the proposed model.

Results

Performance evaluation using various metrics

We used performance measures such as sensitivity, specificity, precision, and accuracy to analyze the multiclass classification. As illustrated in Fig. 3, the poor category had the highest values (all > 95%) for all performance measures. Also, for all participant groups, specificity was higher than sensitivity. This indicates that there are fewer false positives than false negatives present in the test set after classification. Moreover, the excellent and the good

Fig. 3 Comparison of performance metrics for the mental health categories using the proposed method



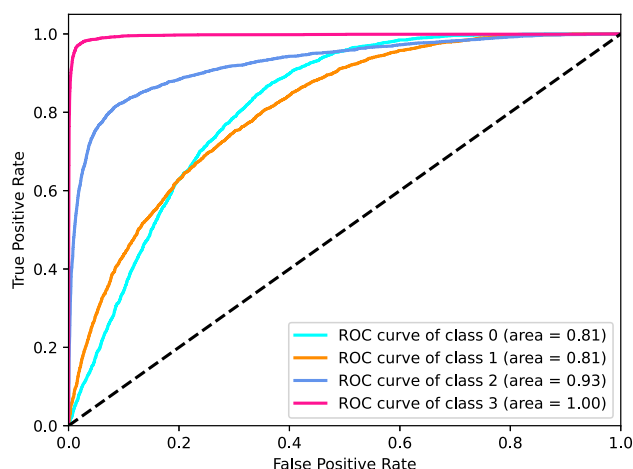


Fig. 4 ROC curve for different mental health categories. Here class 0 denotes the excellent category, class 1 is good, class 2 is fair and class 3 is the poor category

categories have lower precision values than the fair and poor categories. Hence when the proposed model was inaccurate, incorrect classifications were more likely to occur between these two categories. Additionally, for all categories, accuracy exceeded 75%.

The Receiver Operating Characteristic (ROC) curves of the proposed model were examined to measure its effectiveness in performance. The capacity of the model to discriminate between participants belonging to various groups according to different thresholds is shown by the area under ROC curves (AUC). Higher AUC values indicate that the model is more accurately classifying the participants. Figure 4 displays the multiclass ROC curve for the 1D-CNN for the four classes of mental health. The closer the curves are to the top-left corner, the better the performance of the model. We used the one vs. all methods to depict the four ROC curves for the different categories in this multiclass model. The proposed model yielded AUCs of 0.81, 0.81, 0.93, and 1 for the excellent, good, fair, and poor mental health categories respectively.

Table 3 shows the accuracy for each class and the total average accuracy for the different classifiers to evaluate the classification performance. We compared the performance of the proposed network with four other baseline models (sup-

port vector machines (SVM), multi-layer perceptron (MLP), random forests (RF), and naïve Bayes classifier (NB)). The hyperparameter tuning for these classifiers was completed using the grid search method. We can see that the proposed CNN clearly outperformed all the conventional machine learning classifiers in distinguishing among the mental health categories. It achieved the best accuracy of 76%, 78%, 88%, and 98% for excellent, good, fair, and poor, respectively, and it had an overall average accuracy of 85%. Also, the proposed model improved the average accuracy by 5%, 9%, 4%, and 18% over SVM, MLP, RF, and NB, respectively. This indicated the effectiveness of learning complex, deep characteristics from fMRI data. Moreover, the ReLU activation function was utilized in the proposed model to achieve a prediction accuracy of 85%. Although experiments were conducted with other activation functions such as LeakyReLU and PReLU, they did not yield any noteworthy enhancement compared to ReLU. Specifically, LeakyReLU achieved an accuracy of 83%, and PReLU achieved 80%. Hence the choice of ReLU was driven by its effectiveness in addressing the vanishing gradient problem and capturing non-linear relationships efficiently. While Leaky ReLU and PReLU offer ways to combat ReLU's limitations, they introduce added complexity with extra parameters, potentially not significantly improving performance in this specific context.

Table 4 offers a comprehensive examination of the average specificity, sensitivity, precision, and accuracy for all four classes, both with and without the implementation of SMOTE. The analysis emphasizes the notable enhancements addressing the data imbalance issue. From the table provided, it is evident that the application of the SMOTE has significantly improved the performance metrics compared to the case without SMOTE. Here sensitivity has increased from 25% to 70% with the use of SMOTE. This improvement suggests that SMOTE has effectively generated synthetic samples for the minority class, making it more detectable by the model. Moreover, precision has also seen an increase from 26% to 70% when SMOTE was employed. This increase implies that the model's ability to correctly classify positive instances has been notably enhanced, which is particularly crucial when dealing with imbalanced datasets. The specificity has also increased from 75% to 90% with the application of SMOTE. This rise suggests that the model's

Table 3 Accuracy comparison between the proposed model and the state-of-the-art techniques

Model	Accuracy(%)				Average accuracy (%)
	Excellent	Good	Fair	Poor	
Support Vector Machines	71	72	83	93	80
Multilayer Perceptron	70	70	74	91	76
Random Forest	72	74	83	94	81
Naive Bayes	68	69	68	64	67
Proposed model	76	78	88	98	85

Table 4 Impact of SMOTE on resolving data imbalance in performance evaluation

Model	Specificity (%)	Sensitivity (%)	Precision (%)	Accuracy (%)
Without SMOTE	75	25	26	70
With SMOTE	90	70	70	85

capability to correctly identify the negative instances has also improved, which indicates that the model is less prone to misclassifying the majority of class instances. Finally, overall accuracy has seen a notable increase from 70% to 85% with the use of SMOTE. This improvement demonstrates that SMOTE has effectively balanced the dataset, leading to better predictive performance and a more reliable model.

Finally, an ablation study was also conducted in Table 5 to determine the optimal number of layers for the proposed deep learning model. The results indicate that reducing the number of layers correspondingly decreased accuracy. In a smaller input size, such as 53×53 , employing 16 layers enhances the network's depth, facilitating the acquisition of more intricate and abstract features. Here, the way filter sizes increase in the convolutional layers is crucial, helping the neural network capture complex and hierarchical features present in the input data. The initial layers with smaller filter sizes (16) typically focus on detecting simpler features like edges, basic textures, and gradients within the input images. As the network progresses through deeper layers with larger filter sizes (32, 64, and 128), it can identify and combine these simple features into more detailed patterns. Thus, the additional layers in this 16-layer CNN help in feature extraction, prevent overfitting, and aid the network in learning more detailed representations even in a relatively smaller input size. Here, Table 5 illustrates a trend in which increasing the number of layers enhances accuracy. However, this improvement diminishes as the model gets deeper. For instance, going from 13 to 16 layers only increases accuracy by 2%, while increasing from 6 to 10 layers results in a 4% gain. While a smaller number of layers might perform decently, adding layers is aimed at fine-tuning the model to achieve higher accuracy. However, the incremental rise in accuracy beyond a certain layer count becomes marginal, signaling diminishing returns. Nevertheless, these slight enhancements hold significance,

particularly in tasks like image classification, where even minor accuracy improvements can significantly impact overall performance.

Analysis of the mental health questionnaires

Responses to 20 mental health questions were analyzed to identify which questions best discriminated the participants in each mental health quality category. We calculated the percentage of scores within each category by summing the scores of the questions for the participants assigned to each category, dividing it by the total number of participants, and multiplying it by 100. Here the higher percentage value denotes that the corresponding question contributed more to the specified category. Figure 5 presents a bar graph representing these percentage values for each mental health category. For the excellent category, all questions had very low percentage values (less than 30%). Here we observe that extremely low values can be found for symptoms such as nervous feelings, tense feelings, suffering from nerves, loneliness/isolation, frequency of depressed mood, unenthusiasm, and restlessness in the last 2 weeks and for seeing a psychiatrist for nerves, anxiety, and depression. The lack of these symptoms can be considered an indication of excellent mental health quality. When a person is not experiencing these negative emotions, it may signify that they are able to effectively regulate their emotions, which can further improve their general well-being. In contrast, for the poor category, 14 questions had more than 50% value. Hence poor mental health directly corresponds to symptoms such as mood swings, miserableness, irritability, sensitivity/hurt feelings, fed-up feelings, nervous feelings, worry/anxious feelings, tense feelings, worrying too long, suffering from nerves, loneliness/isolation, guilty feelings, frequency of tiredness and lethargy, and seeing a doctor for nerves, anxiety, and depression. The fair category shows similar patterns but with low percentage levels.

Table 5 Relationship between the number of layers and classification accuracy

Number of layers	Classification accuracy (%)
16 layers	85
13 layers	83
10 layers	81
7 layers	78
6 layers	77

Mean sFNC and connectogram analysis

Computing the mean sFNC for each mental health group allows for the measurement of differences in neural connectivity configurations among these groups. This assessment aids in the comprehension of the precise deviations in connectivity that align with distinct mental health categories. Nevertheless, it is crucial to acknowledge that specific mental

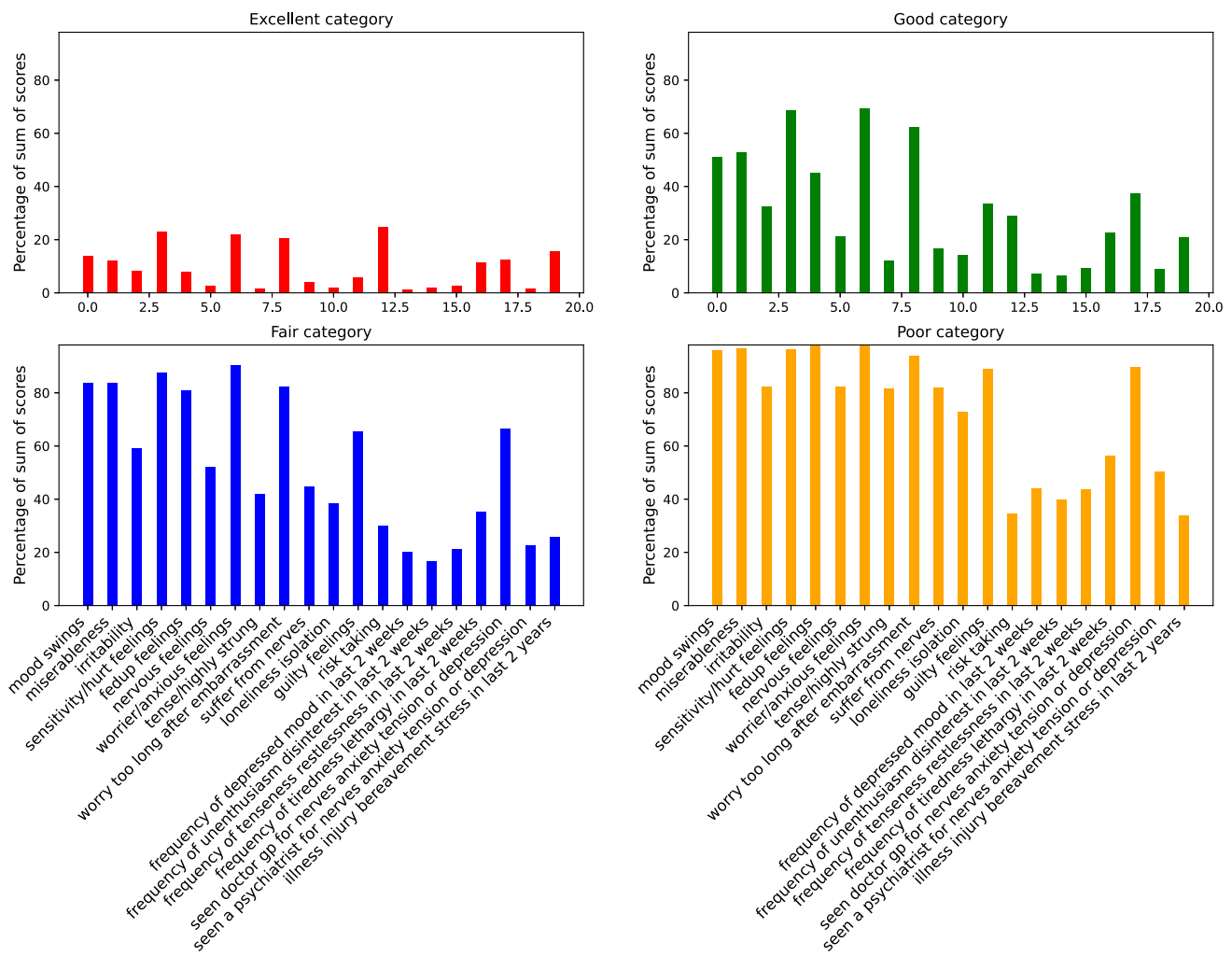


Fig. 5 Contribution of each question towards the mental health category

health conditions have been linked to modifications in brain connectivity and functional networks. Figure 6 illustrates that individuals exhibiting excellent mental health showcased heightened connectivity in the visual-sensorimotor (VIS-SM), cognitive control-sensorimotor (CC-SM) regions, cerebellar-subcortical (CB-SC) areas, and default mode (DM) regions. Enhanced connectivity in these areas can contribute to superior cognitive function and overall mental well-being. In comparison to other categories of mental health, a relatively minor reduction in connectivity was observed in specific brain regions, which doesn't significantly affect the overall functional network. For the good category, notable connectivity was observed in the pairs of sensorimotor-subcortical (SM-SC) and visual-subcortical (VIS-SC) domains. Conversely, the fair group demonstrated more intricate connectivity patterns within the subcortical (SC) domains and visual-sensorimotor (VIS-SM) domain pairs. Participants with poor mental health displayed increased connectivity in specific brain regions, such

as subcortical-cerebellar (SB-CB) and sensorimotor-visual (SM-VS) regions. Conversely, other regions, including the cerebellar-sensorimotor (CB-SM) and visual (VIS) areas, display diminished connectivity. The subcortical-cerebellar (SB-CB) region is integral to functions such as coordination, motor control, and cognitive processing, while the sensorimotor region is responsible for integrating sensory data and coordinating motor responses. Similarly, the visual region is vital for processing visual information and is crucial for visual perception. Modifications in the connections within and between these areas can profoundly impact cognitive and emotional processes, potentially influencing an individual's mental well-being. Notably, certain studies in the field (Kaiser et al., 2015; Snyder, 2013) have indicated that changes in connectivity between neural systems involved in cognitive control and those supporting salience or emotion processing may contribute to difficulties in regulating mood.

Connectograms provide a visual representation of the connectivity patterns between different brain regions, often

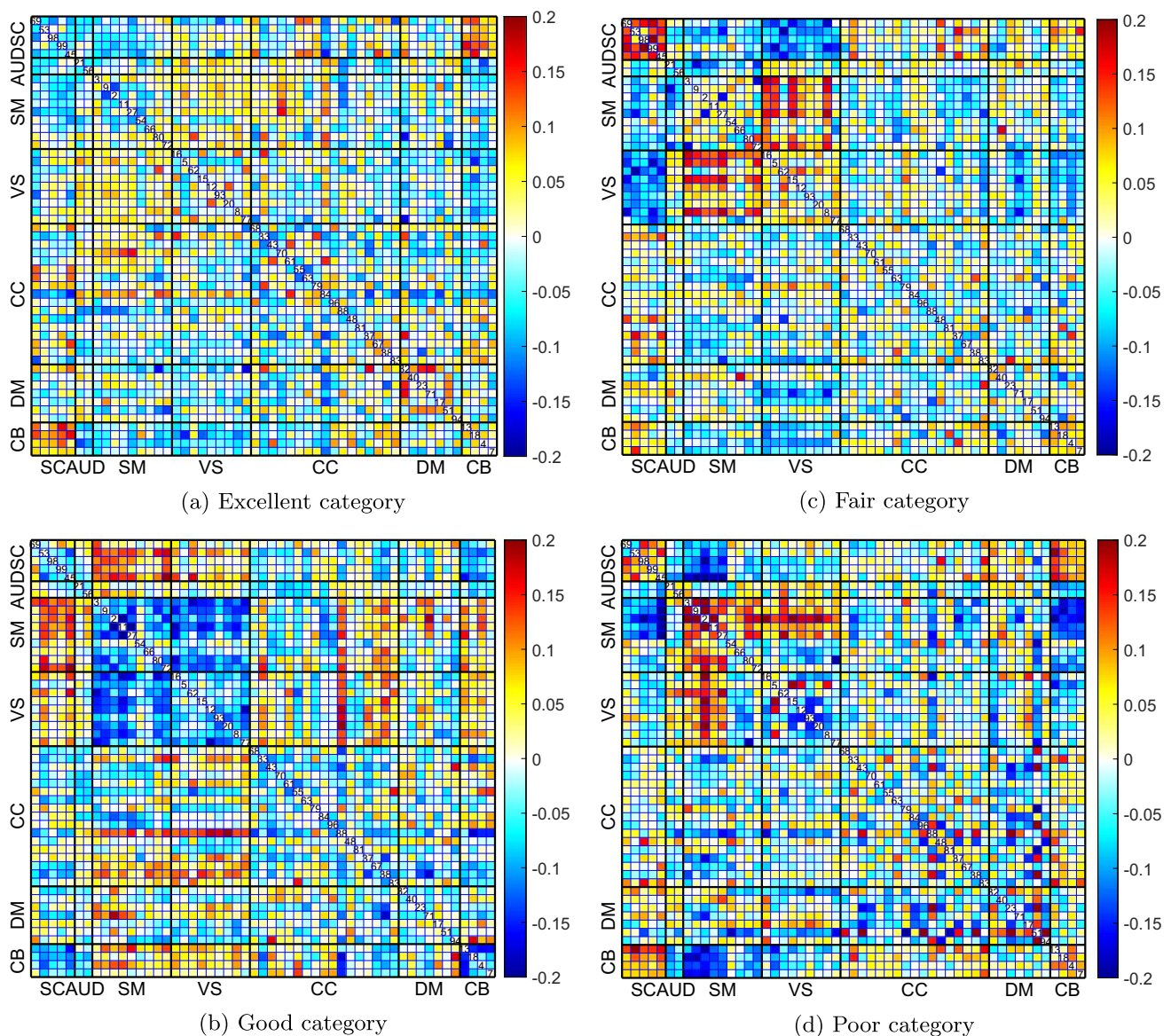


Fig. 6 Mean sFNC of the different test participants for the various categories

revealing critical insights into various cognitive and mental health conditions. Here the connectogram analysis is conducted using the ICs derived from the spatially constrained ICA of the rs-fMRI data. The given connectogram analysis, depicted in Fig. 7(a) and (b), serves as a crucial tool for illustrating the distinctions between two distinct mental health quality categories: excellent and poor. In this plot, the blue color denotes a negative correlation, while yellow denotes substantial links with a positive association. The degree of relevance is indicated by the opacity of the lines. The connectogram for the excellent category exhibited significantly positive values in connection pairs such as the cerebellar-subcortical (CB-SC), default mode-cognitive control (DM-CC), and cognitive control-sensorimotor (CC-SM)

domains. The positive connections suggest that these connections play a significant role in maintaining optimal mental well-being. For instance, the cerebellar-subcortical connections might be involved in motor coordination and emotional regulation, while the default mode-cognitive control connections may be crucial for effective decision-making and self-regulation. Similarly, the cognitive control-sensorimotor connections may be involved in integrating cognitive processes with motor functions. In comparison with the excellent group, the poor category showed a positive association in the default mode (DM) and sensorimotor (SM) domain interactions. These interactions contribute to poorer mental health outcomes, possibly leading to difficulties in emotional regulation, and integration of sensory and motor functions.

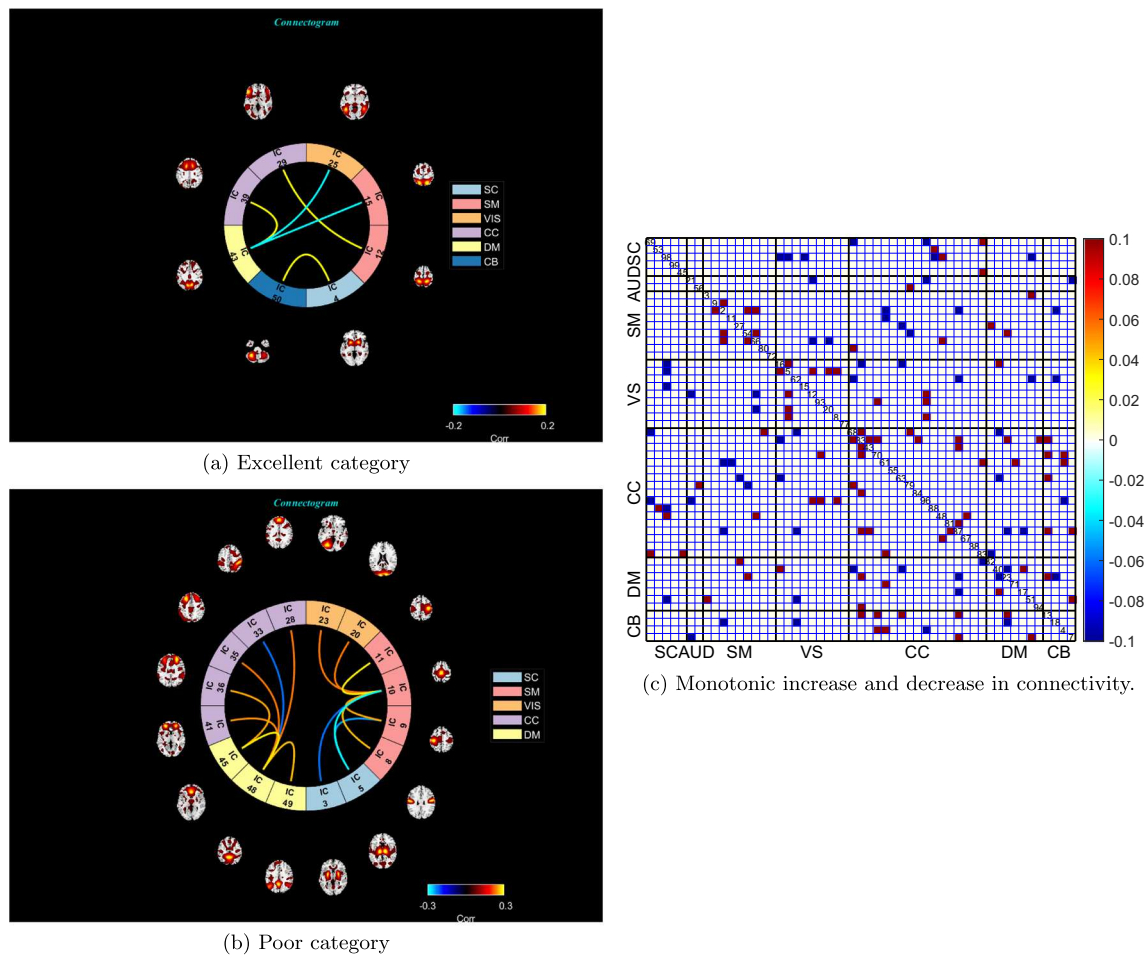


Fig. 7 (a) Connectogram of the mean sFNCs of participants belonging to the excellent category. (b) Connectogram of the mean sFNCs of participants belonging to the poor category. In both cases, warmer col-

ors represent increased connectivity and cooler colors denote decreased connectivity. (c) The red and blue colors denote a monotonic increase and a monotonic decrease in connectivity in the sFNCs

Moreover, as we visualize the progression of the connectivity pattern from excellent to poor mental health category in Fig. 7(c), certain ICs in the sensorimotor (SM), visual (VIS), and cognitive control (CC) domains show a monotonic increase in interactions. This observation suggests a potential shift in the connectivity dynamics, indicating changes in functional domains of the brain, associated with the deterioration of mental health.

Saliency maps of the sFNC

Interpretable models are critical for improving predictability and increasing clinical acceptability. We employed guided gradient class activation maps (guided Grad-CAM) (Selvaraju et al., 2017) to objectively define the prominent areas of the sFNC contributing to mental health category prediction in order to generate interpretable results for the network predictions. Then, we computed the gradients for every

prediction score with respect to the extracted feature maps from the final convolution layer. The weights for feature significance were then calculated using the global average pooling of these gradients. ReLU is applied after a weighted mixture of forward activation maps. ReLU has the advantage of highlighting features that have a positive influence on the target class. It has been shown that localization maps without ReLU may contain more information than the target class, such as negative pixels that likely correspond to other categories (Zhou et al., 2016). While the target class-specific localization map resulting from this might be significant, it could end in the loss of key context and global information.

Figure 8 illustrates the saliency maps for the four mental health categories. The localization maps were averaged across all participants to create the saliency maps. The regions that the model considers salient (darker red zones) for the prediction of mental health are highlighted by overlaying the saliency map from guided Grad-CAM. These

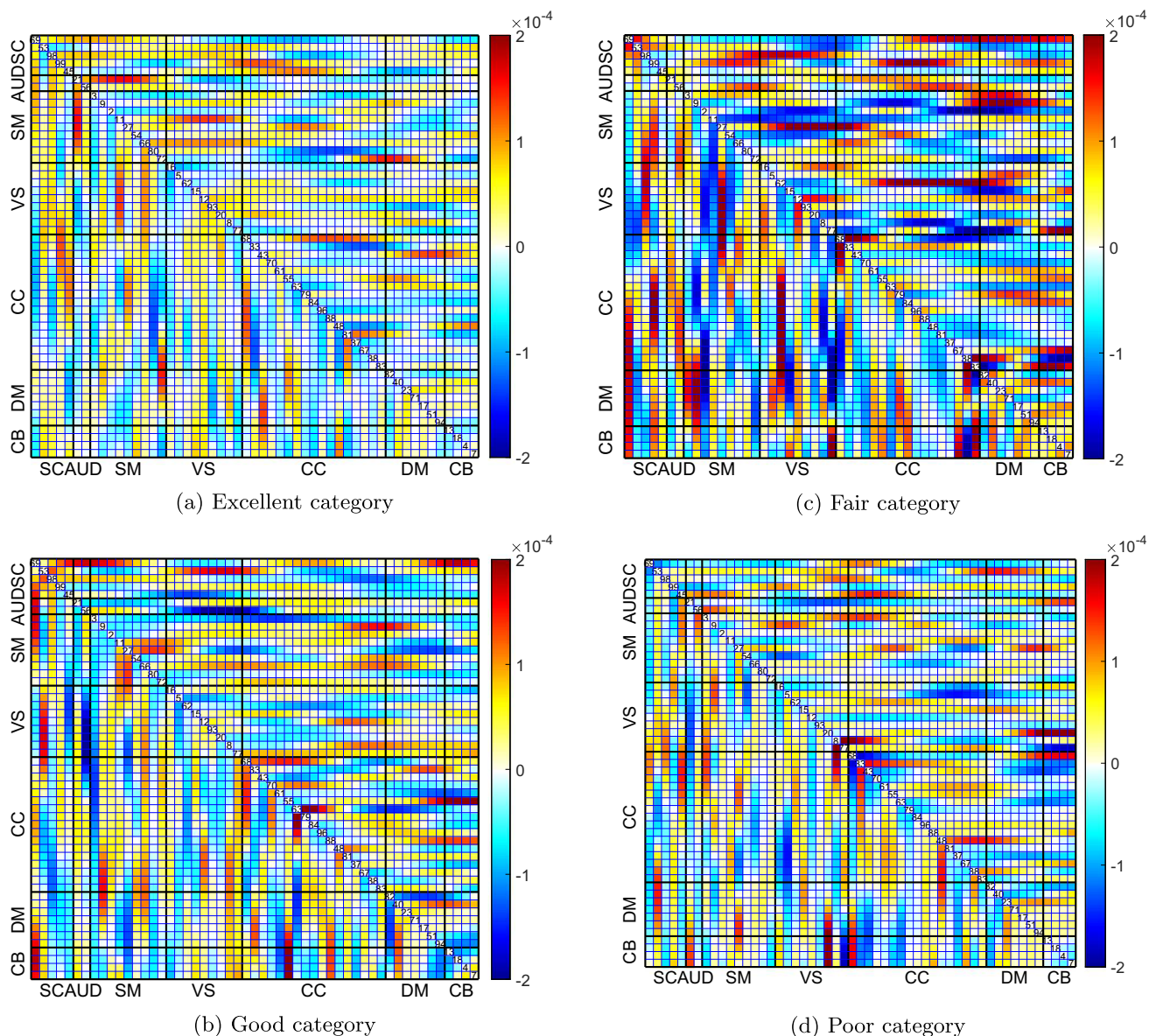


Fig. 8 Saliency maps of different test participants for the various categories

regions are likely to play a significant role in determining the mental health status of an individual. The analysis indicates that the excellent category demonstrates a higher presence of salient regions, characterized by more yellow and dark red areas. These regions, which primarily include sensorimotor-auditory (SM-AUD), visual-sensorimotor (VIS-SM), sensorimotor-subcortical (SM-SC), and parts of the cognitive control domains, are crucial for distinguishing the excellent mental health category. The presence of significant activity in these specific regions suggests that their functionality is closely associated with excellent mental health. In the case of the good category, the salient regions are predominantly concentrated in the subcortical (SC) regions, some parts of the sensorimotor (SM) regions,

and multiple regions of the cognitive control (CC) domains. These areas play a crucial role in distinguishing individuals with good mental health from other categories. The presence of salient regions in the subcortical and cognitive control domains further suggests their importance in maintaining good mental health.

The saliency analysis of the fair category demonstrates a balanced distribution of salient regions, represented by a mix of dark red and dark blue areas. The important regions for the fair category include the cerebellar (CB), default mode-cerebellar (DM-CB), subcortical (SC), sensorimotor-subcortical (SM-SC), and cognitive control-subcortical (CC-SC) regions. The balanced presence of salient regions in these specific areas suggests their moderate influence on the classi-

fication of individuals into the fair mental health category. For the poor category, the saliency map demonstrates a combination of yellow and red regions, indicating an average to high level of saliency. The important regions for the poor category include the default mode, default mode-cerebellar (DM-CB), cognitive control (CC), and some regions of cognitive control-visual (CC-VIS) and sensorimotor-visual (SM-VIS).

The specific regions highlighted in the saliency maps indicate their importance in distinguishing between different levels of mental health, ranging from excellent to poor. Understanding these crucial brain regions can provide valuable insights into the neural mechanisms associated with different mental health states, facilitating a deeper understanding of the underlying factors influencing mental health classifications based on functional network connectivity data.

Discussion

The aim of this study was to develop a framework for sFNC analysis to provide insights into the relationship between self-reported mental health assessment data and sFNCs that comprise the different brain regions. Using a deep learning technique, we discovered four distinct categories of mental health, characterized these groups based on various self-report patterns, and found potentially discriminative functional connectivity areas. The mental health questionnaire analysis also revealed which questions best distinguished individuals in each of the mental health quality categories. Importantly, all these findings were achieved on a large dataset which further hints at the robustness and generalization of the deep learning method.

Recent studies have suggested that depression, anxiety, stress, and other mental health conditions may be associated with disparities of interconnections among brain regions rather than increased or decreased activity of individual areas (Zhang et al., 2016; Wang et al., 2019). Hence, researchers studying mental health have shifted the focus of imaging studies to connections among brain areas. The data-driven connectome-based predictive models take brain connectivity data as input and generate predictions of behavioral measures in participants (Shen et al., 2017). Here the predictive model assumes a linear relationship between the connectivity data and the behavioral variable. However these models may not be optimal for capturing complex, nonlinear relationships between connectivity and behavior. Deep learning models, on the other hand, can learn complicated non-linear correlations between variables by using nonlinear activation layers. Hence in our study, we used the one-dimensional convolutional neural networks to enhance the prediction accuracy by capturing the patterns present between the sFNCs and the self-reported assessment data. Moreover, deep learning

models had shown promise in sFNC-based predictive modeling in previous research, but their lack of interpretability has remained a concern (Cwiek et al., 2022). To address this issue, we used saliency maps to highlight the most significant regions of the sFNC matrix that contribute to the model's prediction. This served as a powerful tool for the interpretability of the deep learning model.

Using the connectogram analysis, we also visualized and identified relevant brain connections contributing to the prediction model. Various studies examining stress-induced neural responses and recovery patterns through post-stress rs-fMRI scans show that the overall intra- and inter-network FC of certain core networks have been frequently reported to be altered after acute stress. These regions consist of the default mode network (DMN) (Clemens et al., 2017) which is involved in internally-directed cognition and includes the posterior cingulate cortex (PCC). Also, the salience network (SN) (W. Zhang et al., 2019; X. Zhang et al., 2019) includes the anterior insula and the dorsal anterior cingulate cortex, which detects behaviorally relevant stimuli and re-allocates the brain's neural resources. Relative to the findings in these previous studies, we also found that in the case of the poor mental health category with an increased likelihood of symptoms of stress, anxiety, and depression, there was also an increase in connectivity in the default mode and sensorimotor regions as shown in the connectogram. Overall, we demonstrated that the proposed architecture significantly improves the accuracy with which one can classify individual participants into distinct mental health quality categories. The saliency analysis also provides several sFNC pairs that exhibit a strong connection in predicting a subject's mental health. In the case of the excellent category, the significant sFNC regions are sensorimotor-auditory (SM-AUD), visual-sensorimotor (VIS-SM), sensorimotor-subcortical (SM-SC), and cognitive control-default mode (CC-DM). The sFNC is sparsely distributed and consistent with the mean sFNC plot shown in Fig. 6 (a). The good category on the other hand has dominant regions in the sensorimotor (SM), cognitive control (CC), and subcortical (SC) domains. While the fair category has a mixture of both positive and negative significant regions with the CB domain being one of the clearly contributing regions of this category. Finally, in the case of the poor category, the default mode-visual (DM-VIS), cognitive control (CC), and default mode (DM) represent the regions that have the most positive influence in this category. As a result, deep learning is a promising strategy for assisting healthcare providers in the development of neuroimaging-based biomarkers for earlier detection in clinical settings. The research conducted may also be employed to help in monitoring mental health quality and response to interventions. Additionally, the findings can serve as a guide for diagnostic testing and therapies that aim to enhance the participants' quality of life.

Conclusion

In this paper, we developed a deep learning framework based on 1D-CNN for categorizing mental health scores. On a large dataset, this model proved more effective than conventional machine learning techniques in classifying individual people accurately. In fact, this model has several advantages as it uses both sFNC and self-reported assessment data for predictive analysis. First, sFNC provides objective measurements of brain activity and connectivity, whereas using just the self-reported data to obtain mental health quality can be biased or influenced by social desirability (Abdallah et al., 2020). Secondly, sFNC data provides information about functional networks across the whole brain, whereas self-reported data may only capture information about specific symptoms. Combining this information through deep learning provides a more comprehensive view of brain function and connectivity. This can help to identify patterns of connectivity that are associated with specific symptoms associated with mental health quality. Finally, the proposed model can clearly distinguish between the different mental health quality groups with high accuracy. Once this has been trained on a dataset, it can be used to make predictions on new data without adding any additional self-reported measures. A limitation of this research is that it focuses on middle-aged and older persons, and the study includes only twenty self-reported assessments performed on the day of scanning.

Nevertheless, our promising findings offer the possibility that neuroimaging data can be leveraged to facilitate a more accurate categorization of people according to their mental health. The categories that self-reported data yielded had distinctive patterns of connectivity. Future work will focus on the development of a deep learning-based fusion model to forecast brain health by using time courses, sFNCs, and spatial maps. Also, we will evaluate whether a predictive model from neuroimaging data can outperform a predictive model based on assessment data. Additionally, we will extend our model to younger adults and include more self-reported measures that are taken at a time point that is independent of the scan date.

Author Contributions Meenu Ajith: Conceptualization, Methodology, Manuscript Writing & Editing. Dawn M. Aycock: Manuscript Writing & Editing. Erin B. Tone: Manuscript Writing & Editing. Jingyu Liu: Manuscript Writing & Editing. Maria B. Misiura: Manuscript Writing & Editing. Rebecca Ellis: Manuscript Writing & Editing. Sergey M. Plis: Manuscript Writing & Editing. Tricia Z. King: Manuscript Writing & Editing. Vonetta M. Dotson: Manuscript Writing & Editing. Vince Calhoun: Methodology, Funding acquisition, Manuscript Writing & Editing.

Funding This work was supported by the Georgia State University RISE program and NSF grant 2112455.

Data Availability All data is publicly available at the UK Biobank online database (<https://www.ukbiobank.ac.uk/>).

Declarations

Competing interests The authors declare no competing interests.

Ethical approval This research study was conducted retrospectively using human subject data made available in open access by the UK Biobank. Ethical approval was not required as confirmed by the license attached with the open-access data.

Consent to participate Informed consent was obtained from all participants.

Consent to Publish All authors have consented to publish.

References

- Abdallah, M., Farrugia, N., Chirokoff, V., & Chanraud, S. (2020). Static and dynamic aspects of cerebro-cerebellar functional connectivity are associated with self-reported measures of impulsivity: A resting-state fmri study. *Network Neuroscience*, 4(3), 891–909.
- Al-Ezzi, A., Yahya, N., Kamel, N., Faye, I., Alsaih, K., & Gunaseli, E. (2021). Severity assessment of social anxiety disorder using deep learning models on brain effective connectivity. *IEEE Access*, 9, 86899–86913.
- Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., & Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cerebral Cortex*, 24(3), 663–676.
- Andersson, J. L., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage*, 20(2), 870–888.
- Barlow, D. H., Curren, A. J., & Woodard, L. S. (2021). Neuroticism and disorders of emotion: A new synthesis. *Current Directions in Psychological Science*, 30(5), 410–417.
- Bispo Júnior, J. P. (2022). Social desirability bias in qualitative health research. *Revista de Saúde Pública*, 56.
- Bridle, J. S. (1990). Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. *Neurocomputing: Algorithms, architectures and applications* (pp. 227–236).
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321–357.
- Cheng, W., Zhao, M., Xie, X., Chen, Y., Huang, M., & Tang, Y. (2022). Abnormal intrinsic connectivity of resting-state networks in bipolar disorder: A machine learning study. *Frontiers in Psychiatry*, 13, 782698.
- Chen, G., Liu, S., Zhu, J., Gu, J., & Wang, Y. (2021). Machine learning-based prediction of cognitive decline using resting-state fmri. *Frontiers in Aging Neuroscience*, 13, 692102.
- Clemens, B., Wagels, L., Bauchmüller, M., Bergs, R., Habel, U., & Kohn, N. (2017). Alert default mode: Functional connectivity changes in the aftermath of social stress. *Scientific Reports*, 7(1), 1–9.
- Craddock, R. C., Holtzheimer, P. E., III, Hu, X. P., & Mayberg, H. S. (2009). Disease state prediction from resting state functional connectivity. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 62(6), 1619–1628.

- Cwiek, A., Rajtmajer, S. M., Wyble, B., Honavar, V., Grossner, E., & Hillary, F. G. (2022). Feeding the machine: Challenges to reproducible predictive modeling in resting-state connectomics. *Network Neuroscience*, 6(1), 29–48.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1–22.
- Du, Y., Fu, Z., Sui, J., Gao, S., Xing, Y., Lin, D., . . . others (2020). Neuromark: An automated and adaptive ica based pipeline to identify reproducible fmri markers of brain disorders. *NeuroImage: Clinical*, 28, 102375.
- Du, Y., & Fan, Y. (2013). Group information guided ica for fmri data analysis. *NeuroImage*, 69, 157–197.
- Dutt, R. K., Hannon, K., Easley, T. O., Griffis, J. C., Zhang, W., & Bijsterbosch, J. D. (2022). Mental health in the uk biobank: A roadmap to self-report measures and neuroimaging correlates. *Human Brain Mapping*, 43(2), 816–832.
- Fräley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458), 611–631.
- Goulas, A., & Margulies, D. S. (2021). Resting-state fmri: A window into human brain plasticity. *Neuroscience and Biobehavioral Reviews*, 129, 38–55.
- Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2), 825–841.
- Kaiser, R. H., Andrews-Hanna, J. R., Wager, T. D., & Pizzagalli, D. A. (2015). Large-scale network dysfunction in major depressive disorder: A metaanalysis of resting-state functional connectivity. *JAMA Psychiatry*, 72(6), 603–611.
- Kawahara, J., Brown, C. J., Miller, S. P., Booth, B. G., Chau, V., Grunau, R. E., & Hamarneh, G. (2017). Brainnetcn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146, 1038–1049.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Li, X., Yang, Z., Zhang, Q., Wei, D., Zhang, Y., Liu, C., . . . Hu, D. (2021). Exploring the potential of deep learning on rs-fmri data for accurate diagnosis of mental disorders. *NeuroImage: Clinical*, 32, 102804.
- Lin, Q.-H., Liu, J., Zheng, Y.-R., Liang, H., & Calhoun, V. D. (2010). Semiblind spatial ica of fmri using spatial constraints. *Human Brain Mapping*, 31(7), 1076–1088.
- Liu, Z., Cui, Y., Du, Y., Gao, J., Yin, H., Sun, H., & Li, X. (2022). Exploring the topological organization of resting-state functional connectivity networks in schizophrenia. *Schizophrenia Research*, 238, 23–29.
- McGee Ng, S. A., Bagby, R. M., Goodwin, B. E., Burchett, D., Sellbom, M., Ayerst, L. E., & Baker, S. (2016). The effect of response bias on the personality inventory for dsm-5 (pid-5). *Journal of Personality Assessment*, 98(1), 51–61.
- Miller, K., Alfaro-Almagro, F., Bangert, N., Thomas, D., Yacoub, E., Xu, J., & Smith, S. (2016). Multimodal population brain imaging in the uk biobank prospective epidemiological study. *Nature Neuroscience*, 19(11), 1523–1536.
- Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th international conference on machine learning (icml-10)* (pp. 807–814).
- Riaz, A., Asad, M., Al Arif, S. M. R., Alonso, E., Dima, D., Corr, P., & Slabaugh, G. (2018). Deep fmri: An end-to-end deep network for classification of fmri data. *2018 IEEE 15th international symposium on biomedical imaging (isbi 2018)* (pp. 1419–1422).
- Saba, T., Rehman, A., Shahzad, M. N., Latif, R., Bahaj, S. A., & Alyami, J. (2022). Machine learning for post-traumatic stress disorder identification utilizing resting-state functional magnetic resonance imaging. *Microscopy Research and Technique*, 85(6), 2083–2094.
- Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., & Smith, S. M. (2014). Automatic denoising of functional mri data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, 90, 449–468.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461–464.
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE international conference on computer vision* (pp. 618–626).
- Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., & Constable, R. T. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols*, 12(3), 506–518.
- Smith, D. J., Nicholl, B. I., Cullen, B., Martin, D., Ul-Haq, Z., Evans, J., . . . others (2013). Prevalence and characteristics of probable major depression and bipolar disorder within uk biobank: Cross-sectional study of 172,751 participants. *PLoS ONE*, 8(11), e75362.
- Snyder, H. R. (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: A meta-analysis and review. *Psychological Bulletin*, 139(1), 81.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of machine learning research*, 15(1), 1929–1958.
- Su, C., Xu, Z., Pathak, J., & Wang, F. (2020). Deep learning in mental health outcome research: A scoping review. *Translational Psychiatry*, 10(1), 116.
- Uyulan, C., Ergüzel, T. T., Unubol, H., Cebi, M., Sayar, G. H., Nezhad Asad, M., & Tarhan, N. (2021). Major depressive disorder classification based on different convolutional neural network models: Deep learning approach. *Clinical EEG and Neuroscience*, 52(1), 38–51.
- Wang, W., Peng, Z., Wang, X., Wang, P., Li, Q., Wang, G., . . . Liu, S. (2019). Disrupted interhemispheric resting-state functional connectivity and structural connectivity in first-episode, treatment-naïve generalized anxiety disorder. *Journal of Affective Disorders*, 251, 280–286.
- Yoshida, K., Shimizu, Y., Yoshimoto, J., Takamura, M., Okada, G., Okamoto, Y., . . . Doya, K. (2017). Prediction of clinical depression scores and detection of changes in whole-brain using resting-state functional mri data with partial least squares regression. *PLoS ONE*, 12(7), e0179638.
- Zeng, L.-L., Shen, H., Liu, L., Wang, L., Li, B., Fang, P., . . . Hu, D. (2012). Identifying major depression using whole-brain functional connectivity: A multivariate pattern analysis. *Brain*, 135(5), 1498–1507.
- Zhang, J.-T., Yao, Y.-W., Li, C.-S.R., Zang, Y.-F., Shen, Z.-J., Liu, L., . . . Fang, X.-Y. (2016). Altered resting-state functional connectivity of the insula in young adults with internet gaming disorder. *Addiction Biology*, 21(3), 743–751.
- Zhang, W., Hashemi, M. M., Kaldewaij, R., Koch, S. B., Beckmann, C., Klumpers, F., & Roelofs, K. (2019). Acute stress alters the ‘default’ brain processing. *NeuroImage*, 189, 870–877.
- Zhang, X., Huettel, S. A., O’Daniel, A., Guo, H., & Wang, L. (2019). Exploring common changes after acute mental stress and acute tryptophan depletion: Resting-state fmri studies. *Journal of Psychiatric Research*, 113, 172–180.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2921–2929).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Meenu Ajith¹ · Dawn M. Aycock² · Erin B. Tone³ · Jingyu Liu^{4,7} · Maria B. Misiura^{3,5} · Rebecca Ellis⁶ · Sergey M. Plis¹ · Tricia Z. King^{3,7} · Vonetta M. Dotson^{8,9} · Vince Calhoun¹

Dawn M. Aycock
daycock@gsu.edu

Erin B. Tone
etone@gsu.edu

Jingyu Liu
jliu75@gsu.edu

Maria B. Misiura
mmisiura1@gsu.edu

Rebecca Ellis
rellis@gsu.edu

Sergey M. Plis
splis@gsu.edu

Tricia Z. King
tzking@gsu.edu

Vonetta M. Dotson
vdotson1@gsu.edu

Vince Calhoun
vcalhoun@Rebeccagsu.edu

University, Georgia Institute of Technology, and Emory University, 55 Park Pl NE, Atlanta, GA 30303, USA

² Byrdine F. Lewis College of Nursing and Health Professions, Georgia State University, P.O. Box 4019, Atlanta, GA 30302, USA

³ Department of Psychology, Georgia State University, Atlanta, GA, USA

⁴ Department of Computer Science, Georgia State University, Atlanta, GA, USA

⁵ Departments of Neurology, Emory University, 615 Michael Street, Suite 505, Atlanta, GA 30322, USA

⁶ Department of Kinesiology and Health, Georgia State University, Atlanta, GA, USA

⁷ Neuroscience Institute, Georgia State University, Atlanta, GA, USA

⁸ Department of Psychology, Georgia State University, P.O. Box 5010, Atlanta, GA 30302-5010, USA

⁹ Gerontology Institute, Georgia State University, Atlanta, GA, USA

¹ Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State