Self-Supervised Mental Disorder Classifiers via Time Reversal

Zafar Iqbal  
TReNDS Center  
Georgia State University  
Atlanta, Ga, USA  
ziqbal5@gsu.edu

Usman Mahmood  
TReNDS Center  
Georgia State University  
Atlanta, Ga, USA  
umahmood1@gsu.edu

Zening Fu  
TReNDS Center  
Georgia State University  
Atlanta, Ga, USA  
zfu@gsu.edu

Sergey Plis  
TReNDS Center  
Georgia State University  
Atlanta, Ga, USA  
s.m.plis@gmail.com

Abstract—Data scarcity is a notable problem, especially in the medical domain, due to patient data laws. Therefore, efficient Pre-Training techniques are required to combat this problem. In this paper, we demonstrate that a model trained on the time direction of functional neuro-imaging data could help in the downstream classification tasks, for example, classifying diseases from healthy controls in fMRI data. We train a Deep Neural Network on Independent components derived from fMRI data using Independent component analysis (ICA) to learn time direction. This pre-trained model is further trained to classify brain disorders in different datasets. Through various experiments, we show that learning time direction helps a model learn some causal relation in fMRI data that helps in faster convergence and better generalization across various datasets, even with fewer data records available for training.

Index Terms—fMRI, dynamic-connectivity, ICA, Mental disorders, Neuro-Imaging.

I. INTRODUCTION

The proliferation of Medical imaging modalities such as Computed Tomography (CT) [1], Positron Emission Tomography (PET) [2] and Magnetic Resonance Imaging (MRI) [3] has made it possible to acquire large amounts of health informatics data. However, due to practical reasons, the data could not be shared publicly in most cases. Hence researchers are constrained to work with small datasets. Functional Magnetic Resonance imaging (fMRI) is a useful imaging technique widely used in investigating psychiatric behaviors and neuro-cognitive diagnosis [4]. The fMRI is broadly classified into two parts; resting state fMRI and task-based fMRI. In the former, the subject is in a task-negative state, and in the latter, the subject is asked to perform some cognitive task. Generally, fMRI is performed using a method called Echo Planar Imaging (EPI) [5]. It collects data for 2D images approximately every 60 ms at a resolution of 3.4 x 3.4 x 4 mm$^3$ voxel size. To cover the whole brain, approximately 32 slices are acquired with a repetition time of 2s/volume [6].

The high dimensional nature of the fMRI data makes it challenging to effectively train a deep learning model on it owing to high computational cost. Likewise, such high dimensional data is inherently inefficient due to redundant information (noise) in it. Therefore, reducing the dimensions of the data without losing valuable information could be helpful during the training phase of the model. Functional brain networks derived from fMRI data can serve this purpose and those can be used as potential biomarkers for mental disorders [7]. One such technique to extract functional brain networks is called Independent Component Anlaysis (ICA).

Our goal is to use ICA-based data to pre-train a model from scratch in a self-supervised way on time direction and investigate its efficacy in the downstream classification task of three diseases, namely, Schizophrenia (SZ), Autism and Alzheimer’s Disease (AD) from their respective healthy subjects. Fig. 1 shows a simplistic example where the process of breaking a jar is demonstrated with respect to time. The Jar is in different states at different time points. The direction of time gives us information about the series of events that happened. In a similar fashion, fMRI data is collected and processed. A stacked pile of 2D axial slices forms a volume representing the entire brain. This process is repeated over time which enables us to look at the functions of the brain. We have shown that learning time direction (forward and backward sequence of volumes) can help in two ways: the model can generalize well even with fewer subjects for training and the training time in the downstream task decreases considerably. We demonstrate that a model trained on forward and reverse time direction can outperform the model trained from scratch. This pretraining method is called Time Reversal (TR).

II. LITERATURE REVIEW

Deep learning models have revolutionized the machine learning domain and have had an everlasting impact in many fields including, Astronomy, Physics, finance, Bio-medicine, etc. These models use different kinds of Neural networks to train and make predictions on unseen data. Convolutional Neural Networks (CNN), Transformers, and Recurrent Neural Networks are to name a few. Reference [8] presents a malware classification method to identify IoT device malware families. They use CNNs as knowledge input to a multilayer perceptron meta-learner. This ensemble approach helps get better classification results on the image datasets under study. Similarly, [9] presents a novel CNN model for arrhythmia classification. They use a multi-channel model to concatenate spectral and spatial feature maps. Furthermore, Sequential...
Model-Based Global Optimization (SMBO) algorithm is used for hyperparameter optimization which results in more efficient predictions by the network. However, the limitation of the machine learning models is that they require large amounts of data.

Data scarcity is a well-known problem in domains where data collection is an expensive and difficult process. One such domain is Medical imaging data [10]. There are many protocols that needed to be followed before making the data public. Owing to such restrictions, it is very challenging to get one’s hands on enough medical data to train a model effectively [11]. One of the methods used to help resolve this issue is unsupervised pre-training of the machine learning model. It acts as a regularizer and could help in better generalization in the downstream task [12]. Stacked Denoising Autoencoders (SDAE) [13] and Deep Beliefs Nets [14] are two of many classical models. These methods, however, are not very popular outside the field of Natural Language Processing (NLP) [15]. In Computer Vision (CV), supervised pre-training is widely used on large imaging datasets such as COCO or Imagenet. With medical imaging, however, We do not have large datasets to pre-train a model. Therefore, self-supervised learning is a suitable method in such cases. It has been shown that self-supervised methods outperform supervised methods when small datasets are used for pretraining [16].

![Forward and reverse time direction of a broken Jar.](image)

Although there has been some research on exploring different self-supervised pretraining techniques involving time direction, no significant work has been done on FMRI data. Agrawal et al. [17] proposed a method called Order Contrastive Pretraining (OCP). They sample random pairs of time segments, switch the order for half of them and train a model to predict the correct time order. They use chest CT scan data (Progression dataset) for experimentation and show that the pretraining with OCP helps in the downstream task. They have used Bidirectional Encoder Representations from Transformers (BERT) architecture for training. Similarly, Hyvarinen et al. [18] proposed Permutation Contrastive learning (PCL), which is very similar to OCP. The key difference is in their objective functions. In OCP, the negatives are incorrectly ordered window pairs, whereas, In PCL, the negatives are random window pairs and they could be in the correct order. The positives in both the approaches are same. There are also some other studies that explored different pretraining objectives on medical time series data. Reference [19] introduces the Patient Contrastive Learning of Representations (PCLR) approach, which results in an average 51% performance increase on Electrocardiogram (ECGs) datasets. However, none of these works have explored the efficacy of Pretraining objective functions on FMRI data. In our proposed method (Time Reversal), We investigate the impact of pretraining via Time Reversal on the classification task in the FMRI data.

Owing to the advancement of medical data acquisition techniques, it is possible to acquire high dimensional data [20]. Such data encompass a large amount of information about the subject. However, processing high dimensional data is computationally expensive and also inefficient due to redundancy and noise in the data [21]. Therefore, Dimensionality Reduction techniques can be useful. One of the most widely used multivariate techniques to estimate brain functional networks from fMRI data is called independent component analysis (ICA). Unlike other techniques based on General Linear Model (GLM) [22], ICA does not depend on prior information in calculating time points. Furthermore, it can help de-noise the fMRI data by decomposing artifacts into independent components, shedding off the noise and redundant information in the volumetric data [7].

![Training Scheme.](image)

III. METHODOLOGY

The proposed method includes two phases; the first phase is pretraining of the model on (Human Connectome Project (HCP) time courses and the second phase is transfer learning, where the pretrained model is further trained on four different datasets, as described in section IV. A generic block diagram of the proposed method is shown in Fig. 2. To evaluate the performance of the model, we used Area Under the ROC curve (AUC) metric. The Receiver Operator Characteristic (ROC) is a probability curve that plots True Positive Rate (TPR) against False Positive rate (FPR) at different thresholds. The AUC score is therefore a measurement of the ability of a model to distinguish between two classes.
We have used three 1D convolutional layers followed by a fully connected layer with 256 units and a Bi-directional long short-term memory (LSTM) layer and two more fully connected layers. We feed ICA time courses to the convolutional layers in the form of a sequence of windows. The output features used for the convolutional layers are 64, 128, and 200 respectively and the kernel sizes used in the first two layers are 4 each, and in the third convolutional layer are 3. A leaky Rectified Linear Unit (Relu) is used as an activation function in all the layers except in the last layer where we used the Softmax function to convert the scores to a normalized probability distribution. The convolutional layers essentially act as an encoder. The encoder encodes the input data into latent representations z. The latent representation of the entire time series is then fed to a Bi-directional LSTM layer with 200 hidden units in sequence. The output from biLSTM layer is fed to the attention layer to get the representation of the entire time series in the form of a single vector c. The output is then passed through two fully connected layers to get the classification scores as shown in Fig. 3. Further details of the architecture can be found in this research article [23].

IV. Experiments and Results

In this section, we have described the performance of the Pre-Trained (PTR) model on the datasets described in section IV-A. The model is pretrained on the time direction of subjects from Human connectome Project (HCP) [24]. The pre-trained model is then further trained on four datasets for the downstream classification task of three abnormalities, namely, Autism, Schizophrenia and Alzheimer’s diseases from healthy controls in the respective datasets. Schizophrenia is a severe psychotic mental disorder. The symptoms include diminished emotional expressions, a lack of motivation, paucity of speech etc. Alzheimer’s Disease is also a neurological disorder which is a progressive form of Dementia. It leads to declining in memory functions and deteriorating social and behavioural functions. Autism leads to severe social-emotional reciprocity, affects non-verbal communicative behaviours and also impacts the ability to understand and maintain relationships. Interestingly, all these three mental disorders have common clinical features, as they present some kind of impairment in cognitive ability [25].

For both pre-training and the downstream task of classification, we used the same architecture as shown and described in section III. The datasets used for downstream tasks are collected from the projects of Functional Biomedical Informatics Research Network (FBIRN) [26], Centre of Biomedical Research Excellence (COBRE) [27], Autism Brain Imaging Data Exchange(ABIDE) [28] and Open Access Series of Imaging Studies (OASIS) [29]. To evaluate how well the pre-trained model performed, we also trained a model from scratch for the downstream classification task. The experimental results are discussed in section V.

A. Datasets

For pretraining, we used the ICA-based data from HCP project. It has 823 subjects in total. The pretraining was done based on the time direction in fMRI data. We labelled the normal time direction as 0 and then reversed the time direction in each component and labelled it as 1. By adding up both the classes, we got 1646 subjects with 823 representing forward time direction and the other 823 subjects representing reverse time direction. The model was able to learn the order of time with high precision.

We used four datasets to evaluate the performance of the PTR model on the classification task. For Schizophrenia classification, we used COBRE and FBIRN. For Autism we trained the model on ABIDE and for the classification of subjects with Alzheimer’s disease, we used OASIS. It is pertinent to mention here that all the datasets used in this study are Pre-Processed through a method called Independent Component Analysis (ICA). The fMRI data is processed using Statistical Parametric Mapping Working (SMP12). Subsequently, only subjects with head motion of \( \leq 3^\circ \) and \( \geq 3 \text{mm} \) were chosen [30]. For all the datasets under study, we acquired 100 ICA components and only 53 non-noise components were used for training purposes. ICA-based data is handy because of its lower dimensional nature. Unlike techniques such as General Linear Model, this approach requires no prior information in marking the regions. One important advantage of ICA-based data is that it contains less noise than the actual fMRI data which helps get good performance from the model [31].

1) Schizophrenia: For the classification of Schizophrenia (SZ), two datasets were used; FBIRN and COBRE.

In FBIRN, There are 311 subjects in total. The number of subjects possesses SZ tallies to 160 and the remaining 151 are the healthy controls (HC). Each subject has 53 non-noise components with 140 time points in each component. For experimentation, we used non-overlapping windows of size 53 X 20. The number of windows was calculated to be 7 given that there were 140-time points. To evaluate the performance of the model on the said dataset, k-fold cross-validation is applied. The validation and test data size was 59 each.

Similarly, COBRE has 157 subjects in total. The number of subjects affected by SZ totals to 89 and the remaining 68 is the healthy controls (HC). Similar to FBIRN, Each subject has 53 non-noise components with 140 time points in each component. For experimentation, we used non-overlapping windows of size 53 X 20. The number of windows was calculated to be 7 given that there were 140 time points. To evaluate the performance of the model on the said dataset, k-fold cross-validation is applied. Two holdout datasets of size 27 each were used for validation and testing of the trained model.

2) Alzheimer’s Disease: For the classification of Alzheimer’s Disease, we used OASIS dataset. It has a total of 823 subjects, out of which 651 are Healthy Controls and the remaining 172 are affected with Alzheimer’s disease. Like FBIRN or COBRE, Each subject has 53 non-noise components with 140 time points in each component. For
experimentation, we used non-overlapping windows of size 53 X 20. The number of windows is calculated to be 7 with 140 time points. The dataset is imbalanced given that the HC class has way more records than the other one. To tackle the class imbalance issue, we ran multiple trials by using all the subjects with Alzheimer’s disease and 172 subjects from HC class in sequence to make sure all the subjects are used in training and evaluating the model. To evaluate the performance of the model on the said dataset, k-fold cross-validation is applied. Two holdout datasets of size 69 each were used for validation and testing purposes.

3) Autism: The dataset ABIDE has 869 subjects. The number of subjects affected by Autism totals to 471 and the remaining 398 is the healthy controls (HC). Similar to other datasets, each subject has 53 non-noise components with 140 time points in each component. For experimentation, we used non-overlapping windows of size 53 X 20. The number of windows was calculated to be 7 given that there are 140 time points. To evaluate the performance of the model on the said dataset, k-fold cross-validation is applied. Two holdout datasets of size 237 each were used for validation and testing purposes.

V. RESULTS AND DISCUSSION

Our hypothesis was that training a model on time direction of fMRI data could help in learning hidden dynamics of the data which in turn could help in any downstream task. To prove our hypothesis, we pre-trained the model on HCP time courses. We had 53 non-noise components in each subject in the HCP dataset. We reversed the time points in each component and pre-train the model on forward and reverse time direction. We evaluated the classification ability of the pre-trained model using four datasets. We experimented with different number of subjects per class to observe how well the model performs even with fewer subjects to train.

For Schizophrenia classification, we used two datasets (FBIRN and COBRE) to further train the model. Fig. 4 shows the details of the classification results on FBIRN. It is noticeable that even with fewer number of subjects during the training phase, the pre-trained model generalizes well and the difference between the AUC scores obtained on test data in both the setups is reasonably large. When we increase the number of subjects, the pre-trained model gets more data for training and thus the performance also increases significantly. We see an increase in the mean AUC score as evident from Fig. 4. The mean and median AUC score with pretraining comes as 0.801 and 0.802 respectively, while without pretraining the respective values are merely 0.70 and 0.71. For COBRE, we used the same setup. The results adhere with the previous findings, that is, the pre-trained model was able to outperform the NPT model in SZ classification task, as shown in Fig. 5. With increasing number of subjects per class, we see a proportional increase in the AUC scores.

For Autism Vs Healthy Controls classification, we used ABIDE. In line with the previous results, the pre-trained model is able to classify better even with only a 1/3rd fraction of training data used in comparison to the model trained from
scratch. As the number of subjects increases, the model starts performing well. An important point to notice here is that even with training on the whole data, the NPT model has a low median AUC score $\approx 0.61$ in comparison to 0.67 when the pre-trained model was used. The age ranges in HCP and ABIDE are quite different with means of 29.2 and 17.04 respectively [32], which proves that the pre-trained model learned signal dynamics that helped in the downstream task even with visible differences in both the datasets. Please refer to the results in Fig. 6.

The dataset OASIS is used for classification of Alzheimer’s Disease vs Healthy Controls. As shown in Fig. 7, there is not a visible difference when we choose 1/3rd of the training data. The reason behind this, in our understanding, is the differences in both the datasets. However, with increasing number of subjects per class, the model’s performance gets better and the difference in AUC score between PTR and NPT models increases significantly.

To thoroughly investigate the efficacy of TR, we also drew several comparisons with two baseline methods called OCP and PCR. We used HCP for pretraining and FBIRN for downstream classification. We pretrained a Visual Transformer (ViT) (https://github.com/FrancescoSaverioZuppichini/ViT) on HCP dataset using the OCP and PCR sampling techniques. Fig. 8 shows various comparisons. From the baseline methods comparisons, we can conclude that the TR approach performs significantly better than both the OCP and PCR pretraining methods. MILC is performing better in comparison to ViT. We speculate that LSTM-based architecture learns better when it comes to longitudinal data.

We further aimed to measure the similarity of activations between the fine-tuned model and the model trained from scratch (NPT). We chose a representation-level measure called Centered Kernel alignment (CKA) for the said purpose. CKA is based on the Hilbert-Schmidt Independence Criterion (HSIC). It is basically a normalization of HSIC. We ran CKA analysis on the weights of the fine-tuned model (Pretrained on HCP time courses) and NPT model. The resultant heatmap is shown in Fig. 9. From the figure, we can observe that both the models are learning different representations. The difference between random and pre-trained models is significant.
layers of both the models have a higher degree of similarity as compared to the higher layers, which shows that top layers are more affected by fine-tuning than their counterparts. The inner square is the encoder part of the model whereas the two larger rectangles represent the LSTM and decode part. The low similarity of the rectangles with the inner and outer squares shows that the LSTM layers were the ones primarily responsible for the difference in AUC scores in both the models.

VI. CONCLUSION

In this paper, we have investigated the ability of the pre-trained model to classify abnormalities from health controls. We have demonstrated that a self-supervised pre-trained model on the time direction of fMRI data can learn information that could help classify better. We pretrained the model on HCP time points. The time direction was reversed and used both forward and reversed time directions for pretraining. The performance of the pretrained model is evaluated using four different datasets. We have demonstrated that training on time direction of fMRI data using ICA time courses, in a self-supervised manner, gives a significant improvement in the downstream classification task. Pre-training with time reversal provides benefits that transfer across datasets. Learning dynamics of fMRI helps to improve the performance of the model, as observed from the results discussed earlier. The Pre-trained model outperforms the NPT model significantly even with fewer number of subjects used for training. In the future, we intend to skip ICA preprocessing and work directly with fMRI data.

ACKNOWLEDGMENT

This work was supported by NIH 2R01EB006841 and in part by NSF 2112455, and NIH RF1MH121885.

REFERENCES

[1] C. J. Garvey and R. Hanlon, “Computed tomography in clinical practice,” Brj, vol. 324, no. 7345, pp. 1077–1080, 2002.
[2] V. P. Sudarshan, G. F. Egan, Z. Chen, and S. P. Awate, “Joint pet-mri image reconstruction using a patch-based joint-dictionary prior,” Medical image analysis, vol. 62, p. 101669, 2020.
[3] A. Ammar, O. Bouattane, and M. Youssfi, “Automatic cardiac cine mri segmentation and heart disease classification,” Computerized Medical Imaging and Graphics, vol. 88, p. 101864, 2021.
[4] J. Yang, S. Gohel, and B. Vachha, “Current methods and new directions in resting state fMRI,” Clinical imaging, vol. 65, pp. 47–53, 2020.
[5] P. Mansfield, “Multi-planar image formation using nmr spin echoes,” Journal of Physics C: Solid State Physics, vol. 10, no. 3, p. L55, 1977.
[6] G. H. Glover, “Overview of functional magnetic resonance imaging,” Neurosurgery Clinics, vol. 22, no. 2, pp. 133–139, 2011.
[7] M. S. Salman, Y. Du, D. Lin, Z. Fu, A. Fedorov, E. Damaraju, J. Sui, J. Chen, A. R. Mayer, S. Posse, et al., “Group ica for identifying biomarkers in schizophrenia: ‘adaptive’networks via spatially constrained ica show more sensitivity to group differences than spatio-temporal regression,” NeuroImage: Clinical, vol. 22, p. 101747, 2019.
[8] H. Naem, X. Cheng, F. Ullah, S. Jabbar, and S. Dong, “A deep convolutional neural network stacked ensemble for malware threat classification in internet of things,” Journal of Circuits, Systems and Computers, 2022.
[9] M. Jangra, S. K. Dhull, K. K. Singh, A. Singh, and X. Cheng, “O-wcnn: an optimized integration of spatial and spectral feature map for arrhythmia classification,” Complex & Intelligent Systems, pp. 1–14, 2021.
[10] I. Izoin, R. Tkachenko, K. Zuh, P. Tkachenko, et al., “A grnn-based approach towards prediction from small datasets in medical application,” Procedia Computer Science, vol. 184, pp. 242–249, 2021.
[11] V. J. Kadam and S. M. Jadhav, “Performance analysis of hyperparameter optimization methods for ensemble learning with small and medium sized medical datasets,” Journal of Discrete Mathematical Sciences and Cryptography, vol. 23, no. 1, pp. 115–123, 2020.
[12] D. Erhan, A. Courville, Y. Bengio, and P. Vincent, “Why does unsupervised pre-training help deep learning?,” in Proceedings of the thirteenth international conference on artificial intelligence and statistics, pp. 201–208, JMLR Workshop and Conference Proceedings, 2010.
[13] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, and L. Bottou, “Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion.,” Journal of machine learning research, vol. 11, no. 12, 2010.
[14] G. E. Hinton, “Deep belief networks,” Scholarpedia, vol. 4, no. 5, p. 5947, 2009.
[15] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT press, 2016.
[16] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in 2009 IEEE conference on computer vision and pattern recognition, pp. 248–255, Ieee, 2009.
[17] M. N. Agrawal, H. Lang, M. Offin, L. Gazit, and D. Sontag, “Leveraging
time irreversibility with order-contrastive pre-training,” in International
Conference on Artificial Intelligence and Statistics, pp. 2330–2353,
PMLR, 2022.

[18] A. Hyvarinen and H. Morioka, “Nonlinear ica of temporally dependent
stationary sources,” in Artificial Intelligence and Statistics, pp. 460–469,
PMLR, 2017.

[19] N. Diamant, E. Reinertsen, S. Song, A. D. Aguirre, C. M. Stultz, and
P. Batra, “Patient contrastive learning: A performant, expressive, and
practical approach to electrocardiogram modeling,” PLoS Computational
Biology, vol. 18, no. 2, p. e1009862, 2022.

[20] S. Faisal and G. Tutz, “Imputation methods for high-dimensional mixed-
type datasets by nearest neighbors,” Computers in Biology and Medicine,
vol. 135, p. 104577, 2021.

[21] S. Wang, M. E. Celebi, Y.-D. Zhang, X. Yu, S. Lu, X. Yao, Q. Zhou, M.-G. Miguel, Y. Tian, J. M. Gorriz, et al., “Advances in data preprocessing
for biomedical data fusion: an overview of the methods, challenges, and
prospects,” Information Fusion, vol. 76, pp. 376–421, 2021.

[22] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub,
K. Ugurbil, W.-M. H. Consortium, et al., “The wu-minn human connectome
project: an overview,” Neuroimage, vol. 80, pp. 62–79, 2013.

[23] J. Levine, “Clinical convergence of autism, schizophrenia, and
alzheimer’s disease: the case of social cognition,” in Neuroprotection in Autism, Schizophrenia and Alzheimer’s disease, pp. 21–53, Elsevier,
2020.

[24] A. Di Martino, C.-G. Yan, Q. Li, E. Denio, F. X. Castellanos, K. Alaerts,
J. S. Anderson, M. Assaf, S. Y. Bookheimer, M. Dapretto, et al., “The
autism brain imaging data exchange: towards a large-scale evaluation of
the intrinsic brain architecture in autism,” Molecular psychiatry, vol. 19,
nos. 6, pp. 659–667, 2014.

[25] Z. Fu, A. Caprihan, J. Chen, Y. Du, J. C. Adair, J. Sui, G. A. Rosenberg,
and V. D. Calhoun, “Altered static and dynamic functional network
connectivity in alzheimer’s disease and subcortical ischemic vascular
disease: shared and specific brain connectivity abnormalities,” Human
brain mapping, vol. 40, no. 11, pp. 3203–3221, 2019.

[26] M. S. Salman, Y. Du, D. Lin, Z. Fu, A. Fedorov, E. Damaraju, J. Sui,
J. Chen, A. R. Mayer, S. Posse, et al., “Group ica for identifying biomarkers in schizophrenia: ‘adaptive’ networks via spatially constrained
ica show more sensitivity to group differences than spatio-temporal
regression,” NeuroImage: Clinical, vol. 22, p. 101747, 2019.

[27] A. Hyvarinen and H. Morioka, “Nonlinear ica of temporally dependent
stationary sources,” in Artificial Intelligence and Statistics, pp. 460–469,
PMLR, 2017.

[28] N. Diamant, E. Reinertsen, S. Song, A. D. Aguirre, C. M. Stultz, and
P. Batra, “Patient contrastive learning: A performant, expressive, and
practical approach to electrocardiogram modeling,” PLoS Computational
Biology, vol. 18, no. 2, p. e1009862, 2022.

[29] S. Faisal and G. Tutz, “Imputation methods for high-dimensional mixed-
type datasets by nearest neighbors,” Computers in Biology and Medicine,
vol. 135, p. 104577, 2021.

[30] S. Wang, M. E. Celebi, Y.-D. Zhang, X. Yu, S. Lu, X. Yao, Q. Zhou, M.-G. Miguel, Y. Tian, J. M. Gorriz, et al., “Advances in data preprocessing
for biomedical data fusion: an overview of the methods, challenges, and
prospects,” Information Fusion, vol. 76, pp. 376–421, 2021.

[31] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub,
K. Ugurbil, W.-M. H. Consortium, et al., “The wu-minn human connectome
project: an overview,” Neuroimage, vol. 80, pp. 62–79, 2013.

[32] J. Levine, “Clinical convergence of autism, schizophrenia, and
alzheimer’s disease: the case of social cognition,” in Neuroprotection in Autism, Schizophrenia and Alzheimer’s disease, pp. 21–53, Elsevier,
2020.