Available online at http://scik.org

Commun. Math. Biol. Neurosci. 2021, 2021:39

https://doi.org/10.28919/cmbn/5565

ISSN: 2052-2541

TRANSFER LEARNING USING INCEPTION-RESNET-V2 MODEL TO THE AUGMENTED NEUROIMAGES DATA FOR AUTISM SPECTRUM DISORDER CLASSIFICATION

NICHOLAS DOMINIC[1,*], DANIEL[1], TJENG WAWAN CENGGORO[2,3], ARIF BUDIARTO[2,3], BENS PARDAMEAN[1,3]

[1]Computer Science Department, BINUS Graduate Program - Master of Computer Science Program, Bina Nusantara University, Jakarta 11480, Indonesia

[2]Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia

[3]Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta 11480, Indonesia

Abstract: From a psychiatric perspective, the detection of Autism Spectrum Disorders (ASD) can be seen from the differences in some parts of the brain. The availability of the four-dimensional resting-state Functional Magnetic Resonance Imaging (rs-fMRI) from Autism Brain Imaging Data Exchange I (ABIDE I) led us to reorganize it into two-dimensional data and extracted it further to create a pool of neuroimage dataset. This dataset was then augmented by shear transformation, brightness, and zoom adjustments. Resampling and normalization were also performed. Reflecting on prior studies, this classification accuracy of ASD using only 2D neuroimages should be improved. Hence, we proposed the use of transfer learning with the InceptionResNetV2 model on the augmented dataset. After freezing layer by layer, the best training, validation, and testing results were 70.22%, 57.75%, and 57.6%, respectively. We

*Corresponding author

E-mail address: nicholas.dominic@binus.ac.id

Received February 16, 2021

proved that the transfer learning approach was successfully outperformed the convolutional neural network (CNN) model from the previous study by up to 2.6%.

Keywords: psychiatric; rs-fMRI; autism spectrum disorder; transfer learning; classification; InceptionResNetV2.

2010 AMS Subject Classification: 62M45, 78M32, 82C32, 92B20, 92C20, 68T05, 68T45.

1. INTRODUCTION

Autism Spectrum Disorder (ASD) is an increasingly prevalent but lifelong neuro-developmental disorder (NDD) that cause individuals to experience difficulty in both verbal or nonverbal communication, lack of skill to establish a relationship [1], restricted interest, repetitive pattern of behavior, anxiety [2], irritability, aggressive, self-harming behaviors [3], sleep problems and somatic complaints [4], followed by the risk of health, of depression by 10-70%, and Attention-Deficit Hyperactivity Disorder (ADHD) by 28.2% [5]. All these symptoms emerge distinctly depending on the age, language, and cognitive capabilities of each individual [6]. Even some studies suggested that ASD was highly influenced by genetic or environmental factors, its exact cause is still unmeasurable [7]–[10]. Thus, routine screening proposed by the American Academy of Pediatrics and the "Learn the Sign. Act Early." program proposed by the Centers for Disease Control and Prevention (CDC) are worth implementing as an early diagnosis and treatment of ASD [11], [12].

ASD individuals with lower Full-scale Intelligence Quotient (FIQ) score have a significant difference within their brain cortical thickness (CT) compared to those with higher FIQ, while Autism Diagnostic Observation Schedule (ADOS) assessment does not reflect the shift in the same CT, as Bedford et al. [13] stated. This reason was reconfirmed through a study conducted by Yassin et al. [14] that CT is the only feature that could differentiate ASD, TD, and even Schizophrenia. It can be concluded that there are some parts of the brain associated with the presence of ASD and therefore the usage of some brain scans (mined from Autism Brain Imaging Data Exchange/ABIDE) to classify ASD with normal/typically developing (TD) individuals is an intriguing task.

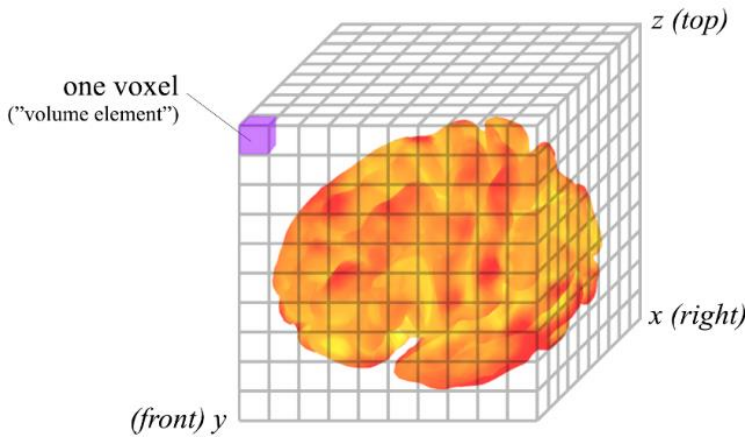

Figure 1. Brain Volume

Functional Magnetic Resonance Imaging (fMRI), one of the various methods for visualizing the human brain, generates a 4D image, where its former, MRI only generates 3D. Even though, fMRI has poor resolution compared to MRI. For ABIDE I dataset, brains that are scanned at resting state (so it is called rs-fMRI) or a vegetative state or negative activity, have the same 3D resolution of (61 × 73 × 61) but various time repetition (TR). The sample of 3D brain volume is displayed in Figure 1.

Since the use of Functional Connectivity Matrix as input is competitive in performance, our work is to figure out if the use of 2D neuroimage as input is sufficient to classify ASD and TD, by utilizing the InceptionResNetV2 model to the transfer learning purpose, which was once state-of-the-art in the image classification experiments. The use of the deep learning model here is because it has been proven to significantly enhance the performances in many tasks in the Computer Vision field [15]–[17].

2. RELATED WORKS

Since ABIDE consists of rs-fMRI neuroimage data, most researchers highly concern to utilize active brain region interconnectivity strength (connectome) as the main feature for this classification task. This connectome is extracted into what they called Functional Connectivity

Matrix or Connectivity Maps, in the form of Pearson Correlation Matrix which each cell represents the brain activation-level signal between two brain regions of interest (ROI) [18].

The result by utilizing this feature is highly competitive. For instance, the use of a Stacked Denoising Auto Encoder model yields up to 70% accuracy [19], [20]. However, other studies proved that with the same amount of data (505 ASD and 530 TD), accuracy can be achieved up to 70.22% with the simple deep convolutional neural network (CNN) [21], up to 71.98% with the classical ridge regression [22], and even up to 75.27% only with a simple deep multi-layer perceptron model [23].

Of course, many things influence the result above, such as the kind of brain atlas parcellation used, its validation scheme, and total dataset. With CC200 atlas, a basic Auto Encoder can achieve 70.3% of accuracy, while with AAL116 and TT97 atlas the accuracy decreases to 67.5% and 65.3% respectively [24].

Nevertheless, other aspects in rs-fMRI data include BOLD (Blood Oxygenated Level Dependent) signal and the 4D neuroimage itself, which is rarely explored as a major feature in ASD classification tasks. A study performed by Ke, Choi, Kang, Cheon, and Lee [25] reduced this 4D neuroimage into 2D input and has not yielded significant accuracy. Even with the 2D CNN model + the class activation mapping (CAM) and the 3D CNN + the 2D spatial transformer network (STN), the classification accuracy is still below 50% for using both models.

3. MATERIALS AND METHODS

3.1. Python Packages

Python v3.7.1 serves all processes in this research, from data fetching into model training, while pip v21.0.1 used to install related packages (run `pip install [package_name]`) as displayed in Table 1. Other dependencies include Numpy v1.15, Scipy v1.6.0, Scikit-learn v0.24.0, etc (run `pip show [package_name]`).

Table 1. Python packages used in this experiment.

No.	Package List	Purpose
1	Pandas v1.2.1	• Load ABIDE I phenotypic information • Insert available neuroimages path • Save training results into CSV file
2	Matplotlib v3.3.4	• Figure settings (size, *x* label, *y* label, etc.)
3	Nibabel v3.2.0	• Load NifTI files
4	Nilearn v0.7.0	• Fetch dataset • Convert 4D into 3D images • Plot neuroimages
5	Tensorflow-Keras v2.4.0	• Resample, augment, and normalize dataset • Define, compile, and train the model

3.2. Dataset

This classification task requires ABIDE (Autism Brain Image Data Exchange) I as the main dataset, fetched from Python's library Nilearn 0.7.0, using Connectome Computation System (CCS) (run `nilearn.datasets.fetch_abide_pcp()`) as a preprocessing pipeline. Gathered from 17 different international sites (as shown in Table 2), ABIDE I consist of 1,112 rows of phenotypic data and only 871 NifTI files, whose 403 data are labeled as ASD and 468 data are labeled as Normal/TD.

Table 2. Phenotypical information of ABIDE I grouped by international sites.

Intl' Site	Age at Scan Range (in years)	FIQ	VIQ	PIQ	ADOS Total Score
CALTECH	17 – 56.2	82.2 ± 99.24	93.77 ± 61.63	87.28 ± 73.02	15 ± 4.9
CMU	19 – 33	112.63 ± 11.28	112.18 ± 11.8	110.18 ± 10.86	12.33 ± 2.88
KKI	8.2 – 12.77	107 ± 16.03	N/A	N/A	13.75 ± 3.91
LEUVEN	15 – 24.5	112.43 ± 13.21	109.14 ± 16.31	105.86 ± 15.08	N/A
MAX_MUN	7 – 58	93.69 ± 80.32	N/A	N/A	-286.63 ± 342.53
NYU	**6.47 – 39.1**	**110.78 ± 14.86**	**109.69 ± 14.61**	**109.63 ± 15.24**	**11.27 ± 4.05**
OSHU	8 – 15.23	94.99 ± 78.88	N/A	N/A	9.25 ± 3.31
OLIN	10 – 24	85.78 ± 103.21	N/A	N/A	14.07 ± 4.16
PITT	9.33 – 35.2	110.26 ± 11.99	107.36 ± 12.21	110.34 ± 11.64	-86.84 ± 226.8
SBL	20 – 49	-216.34 ± 140.47	-39.43 ± 115.92	-70.02 ± 141.52	-286.96 ± 309.47
SDSU	8.67 – 17.15	111.96 ± 12.9	109.85 ± 12.88	111.41 ± 13.2	11.38 ± 44.4
STANFORD	7.53 – 12.93	113.96 ± 15.1	112.2 ± 17.8	112.76 ± 15.01	12.58 ± 3.2
TRINITY	12 – 25.66	109.7 ± 14.09	108.34 ± 14.67	109.07 ± 13.57	10.73 ± 2.7
UCLA	9.1 – 17.2	103.48 ± 12.9	104.7 ± 12.79	102.48 ± 13.43	N/A
UM	10.5 – 24	100.87 ± 46.42	113.515 ± 15.3	99.56 ± 36.99	N/A
USM	8.77 – 50.22	105.31 ±18.02	101.96 ± 20.3	107.16 ± 16.59	-34.91 ± 157
YALE	7 – 17.75	98.34 ± 20.21	101.03 ± 22.04	95.51 ± 18.59	11 ± N/A

FIQ stands for Full-scale Intelligence Quotient (max. value: 148), VIQ stands for Verbal Intelligence Quotient (max. value: 180), PIQ stands for Performance Intelligence Quotient (max. value: 155), and ADOS stands for Autism Diagnostic Observation Schedule. ADOS Total (max. value: 22) represents the total of Communication (max. value: 8) and Social (max. value: 14) sub-score, according to the ABIDE I Dataset Legend.

A hundred phenotypic details are available for each patient in additional ABIDE phenotypic inscribed within CSV file format, but for brevity, only the five items above are presented as an overview of the patient's condition at each site.

To reduce variability from different sites, we eventually decided to select only 172 patients' data from NYU (New York University) site. The final total neuroimage sample generated is 6,880, as later explained in section 3.3.1.

3.3. Methodology

The workflow of this research is divided into data preparation and preprocessing, continued by transfer learning using Inception-ResNetV2 and evaluate the results.

3.3.1. Data Preparation and Preprocessing

Data preparation started from downloading the ABIDE I dataset, that included NifTI files with separated CSV file of phenotypic information of the patients. There are only 871 neuroimages available out of 1,112, with a proportion of 403 ASD and 468 TD. The exploration step required Nibabel v3.2.0 library to load the NifTI files and Nilearn v0.7.0 library to plot them as Echo Planar Imaging/T2* scan (see Figure 2).

Since each NifTI data has a 4D shape, where the first 3D is spatial data (x, y, z) and 1D is the time of repetition (TR) which is time-series data, we only extracted the 2D neuroimage from the x-direction, from the brain slice 1 until 40 (x=1, x=2, …, x=40). This dataset extension was also done so the size of datasets is sufficient to be fed into very complex models [26]–[28] such as InceptionResNetV2, since it has been proven to have a significant effect on deep neural network performance [29].

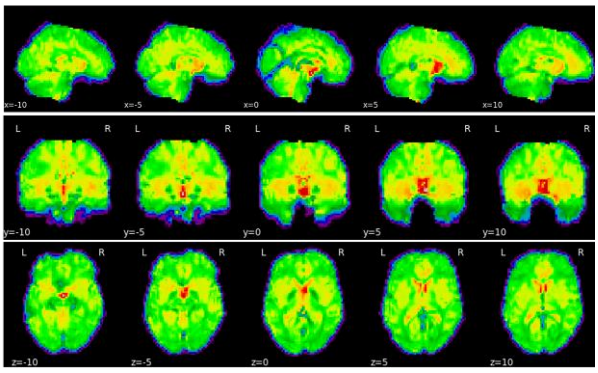

Figure 2. Echo Planar Imaging (EPI) brain scan sample, with cutting coordinates x, y, z = (-10, -5, 0, 5, 10)

To reduce computational time and variability [25], we decide to use only 172 neuroimages data from the NYU site, with a proportion of 74 ASD (×40 slices = 2,960) samples and 98 TD (×40 slices = 3,920) samples. Due to this imbalance of data, rearrangement is needed to ensure that the total of both data is balanced in the 70% train, 15% validation, and 15% test split. The distribution is shown in Table 3.

Table 3. Total samples in train, validation, and test split.

Split	ASD	TD	Total
Train	2,072 (70%)	2,744 (70%)	4,816
Validation	444 (15%)	588 (15%)	1,032
Test	444 (15%)	588 (15%)	1,032
Total	2,960	3,920	6,880

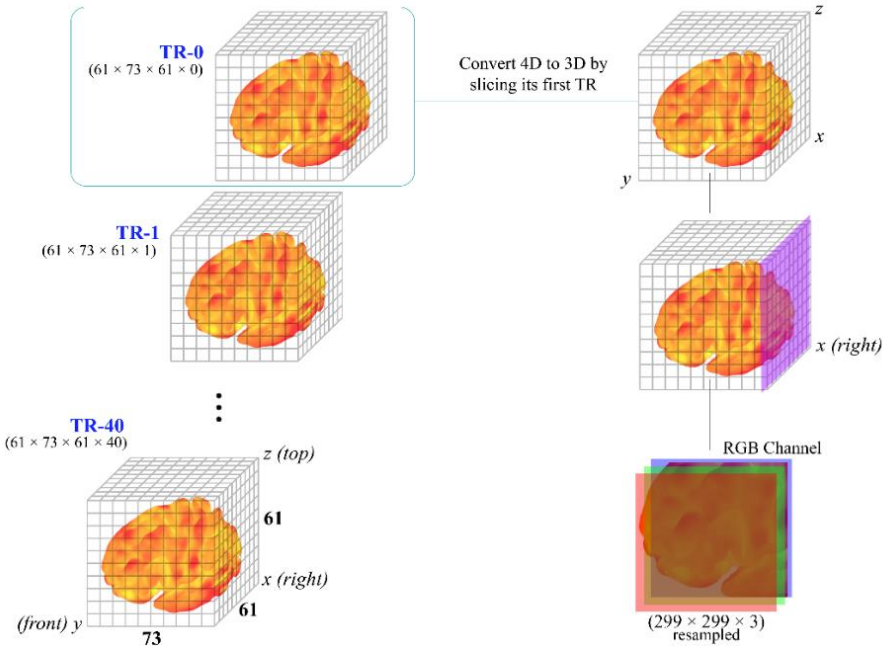

Figure 3. 2D-Neuroimage Acquisition Process

Each image resampled to the size of (299 × 299 × 3) with the nearest interpolation technique (or the "pixels replication"), where 3 here denotes an RGB channel. This method is used for enlarging the image size by multiplying the pixel, where the pixel will be filled with the contents of the adjacent pixel to keep the colors information. Details for this 2D neuroimages acquisition stages are depicted in Figure 3.

The dataset was split into 70% of training data, 15% of validation data, and 15% of test data. Both training and validation data were loaded, augmented, and normalized as BatchDataset object using Tensorflow v2.4.0 library, and finally fed into the InceptionResNetV2 model.

3.3.2. Transfer Learning and Evaluation Scheme

Before we dive further into the definition of transfer learning, let it be clear about some notations used as follows.

- The source domain is denoted as $D_S = \{(x_{S1}, y_{S1}), (x_{S2}, y_{S2}), \ \ldots, (x_{Sn}, y_{Sn})\}$.
- The target domain is denoted as $T_s = \{y_s, f(\cdot)\}$.
- The task domain is denoted as $D_T = \{(x_{T1}, y_{T1}), (x_{T2}, y_{T2}), \ \ldots, (x_{Tn}, y_{Tn})\}$.
- The task target is denoted as $T_T = \{y_T, f(\cdot)\}$.

Here, x is an input data instance and y is a target label, while $f(\cdot)$ represents the objective predictive function.

Mathematically, transfer learning is defined as an approach to improve the result in the target task T_T given the base knowledge from the source domain D_s and target domain T_s, on the condition of $D_s \neq D_T$ or $T_s \neq T_T$ [30].

The augmentation part includes counter-clockwise/vertical shear transformation (with shear factor m = 30, see equation (1)), brightness shift (in constant range of c = 0.25 to c = 0.75, see equation (2)), and zoom adjustments in range 0.8 to 1.2. This does not mean we try to extend the number of samples or randomly generate new samples like the Gibbs Sampler algorithm [31], but we randomly varied the dataset for each batch in the training process. We also let all batches in each epoch be shuffled.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ m & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (1)$$

$$f(u) = \begin{cases} u + c, & u + c \leq 255 \\ 255, & otherwise \end{cases} \quad (2)$$

In equations (1) and (2), *x* and *y* represent axes point, *m* represents shear factor, *c* represents a brightness constant, and *u* represents a pixel.

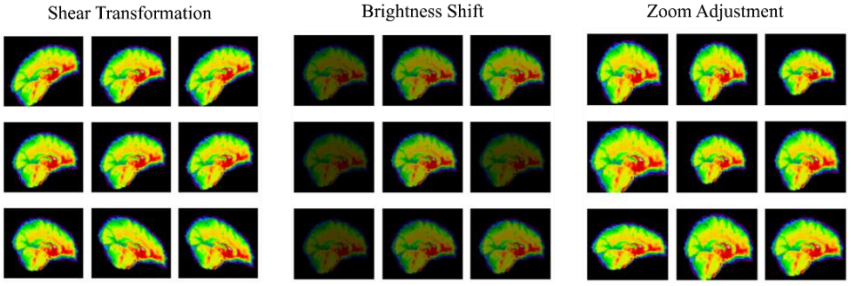

Figure 4. Neuroimages random augmentation examples

After augmentation, normalization in range (-1, +1) was applied. Since we proposed a transfer learning approach, the ImageNet [27] weights were transferred to an InceptionResNetV2 so the model will learn effectively over a smaller dataset [32], i.e., ABIDE I.

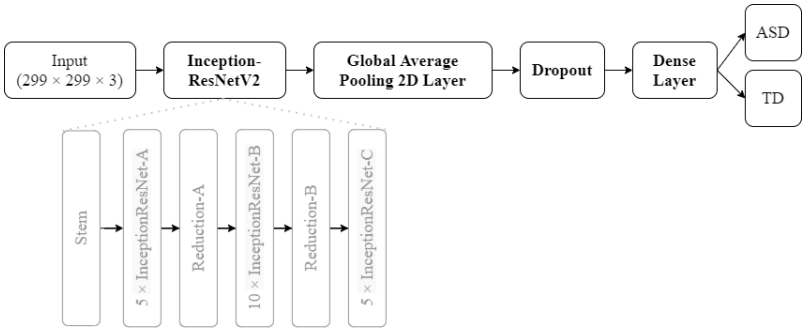

Figure 5. InceptionResNetV2 architecture

The InceptionResNetV2 model is depicted in Figure 5, where the stem layer is the same as the InceptionV4 model, and the rest consists of (a) 35 × 35 grid InceptionResNet-A module, (b) 35 × 35 to 17 × 17 Reduction-A module, (c) 17 × 17 grid InceptionResNet-B module, (d) 17 × 17 to 8 × 8 Reduction-B module, and the last (e) 8 × 8 grid InceptionResNet-C module [33].

After this main architecture, Global Average Pooling was added rather than Flatten to natively prevent overfitting in the convolutional structure due to the absence of parameters to be optimized

and by reinforcing the link between the feature importance and label category [34]. This also makes Global Average Pooling more parameter-efficient compares to the Flatten method. Afterward, a Dropout layer with a fixed value of 0.8 was added, as performed by Szegedy, Ioffe, Vanhoucke, and Alemi [33].

$$\sigma\,(x)_i = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{y_j}} \quad \textbf{(3)}$$

Lastly in the dense layer, Softmax activation function σ was applied, as shown in equation (3), where *x* and *y* point respectively to input and output, *K* points to the number of classes, and *e* itself points to the standard exponential function, i.e., $e \approx 2.718$.

$$w' = w - \alpha * \nabla(w; x^{(i)}; y^{(i)}) \quad \textbf{(4)}$$

For optimization during backpropagation, the iterative Stochastic Gradient Descent (SGD) algorithm was utilized, formulated in equation (4), where *w* denotes weight, α denotes learning rate, and $\nabla(w; x^{(i)}; y^{(i)})$ denotes the gradient to weight, input, and output/label, respectively [35].

The complete tunable hyperparameter settings are displayed in Table 4. Later these hyperparameters, i.e., decay and momentum, were fine-tuned to achieve better accuracy.

Table 4. Default hyperparameter settings.

Number of epoch(s)	:	20
Number of batch(s)	:	32
Train, val., test split	:	70:15:15
Learning rate	:	1e-3
Decay	:	1e-6
Momentum	:	0.5
Optimizer	:	SGD
Activation	:	Softmax

3.4. Performance Evaluation

Based on Confusion Matrix, in this study, (*a.*) True Positive/TP represents ASD correctly identified as ASD, (*b.*) False Positive/FP represents TD incorrectly identified as ASD, (*c.*) True Negative/TN represents TD correctly identified as TD, and (*d.*) False Negative/FN represents ASD incorrectly identified as TD.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad \textbf{(5)}$$

$$Sensitivity/Recall = \frac{TP}{TP + FN} \quad \textbf{(6)}$$

$$Specificity = \frac{TN}{TN + FP} \quad \textbf{(7)}$$

$$Precision = \frac{TP}{TP + FP} \quad \textbf{(8)}$$

$$F_\beta - score = \frac{(1 + \beta^2) \times Precision \times Recall}{\beta^2 \times (Precision + Recall)} \quad \textbf{(9)}$$

By this, five evaluation metrics were utilized, namely accuracy (see equation (5)), recall/sensitivity (see equation (6)), specificity (see equation (7)), precision (see equation (8)), F_β-score with β = 1 in common (so it called "F_1-score", see equation (9)), and Area Under Curve (AUC) of the Receiver Operating Characteristics (ROC) curve.

4. RESULTS

The result of the transfer learning approach is shown in Table 5, where the configurations are as follows:

- in Scenario 1 (Sc1) all layers in the model were frozen,

- in Scenario 2 (Sc2) the InceptionResNetV2 model was frozen, and
- in Scenario 3 (Sc3) only the last dense layer was frozen.

The layer freezing was done to keep the learned features during the training stages. Global Average Pooling and Dropout layers were excluded since they have zero parameters. For all results, the best validation accuracy is highlighted in bold.

Table 5. The result for the Transfer Learning approach. Acc. means accuracy, sen. means sensitivity, spec. means specificity, and prec. means precision.

	Train					Validation				
	Acc.	Sen.	Spec.	Prec.	AUC	Acc.	Sen.	Spec.	Prec.	AUC
Sc1	**0.7022**	**0.6862**	**0.8522**	**0.7077**	**0.7724**	**0.5775**	**0.5718**	**0.6128**	**0.5773**	**0.5547**
Sc2	0.6546	0.6349	0.7757	0.6582	0.7098	0.5756	0.5671	0.6327	0.5803	0.5561
Sc3	0.6158	0.5966	0.7221	0.6237	0.6623	0.5728	0.5699	0.6213	0.5716	0.5652

All training processes which take about 113 to 132 seconds per epoch were executed by using two Intel Xeon E5-2630 v4 Processor with 10-core running at 2.20 GHz and one NVIDIA Tesla P100 GPU with 3584 CUDA cores.

By the default hyperparameters setting as stated in Table 5, the best validation accuracy was achieved if all layers are frozen (Sc1). However, the smallest gap between train and validation was 4.3% in Sc3 where only the last dense layer was frozen, and thus yield the best test accuracy by 56.98%. Then followed by Sc2 where only the base InceptionResNetV2 model was frozen, with a 7.9% gap.

The highest validation recall, which is the measure of a model's ability to identify the actual ASD correctly, was obtained at 57.18% in Sc1. However, the highest specificity, which is a measure of a model's ability to identify the actual TD correctly, and also the highest precision,

which is a measure of a model's to precisely identify ASD against the total of True Positive and False Positive, was obtained at 63.27% and 58.03% respectively, both in Sc2.

Through all these scenarios, precision was higher than recall (1.32% higher in Sc2, 0.55% higher in Sc1, and 0.17% higher in Sc3). This indicated that the classification result has a low False Positive (FP, or commonly called "the false alarm") rate, since $Precision \propto 1/FP$. Simply put, the model was succeeded to avoid misclassification of TD as ASD. After all, in the case of an imbalanced dataset, precision is preferred over recall [36].

The F_1-score from the best validation accuracy was 57.45%. This value, which simply represents the balance between precision and recall, is also considered as a better measurement result when faced with an imbalanced dataset [37].

Table 6. The result using different learning momentums (m). Acc. means accuracy, sen. means sensitivity, spec. means specificity, and prec. means precision.

	Train					Validation				
	Acc.	Sen.	Spec.	Prec.	AUC	Acc.	Sen.	Spec.	Prec.	AUC
m=.5	0.7317	0.7194	0.8918	0.7374	0.8086	0.5890	0.5861	0.6384	0.5894	0.5652
m=.6	0.6629	0.6410	0.7936	0.6713	0.7216	0.5728	0.5499	0.6099	0.5734	0.5617
***m*=.7**	**0.6221**	**0.5943**	**0.7184**	**0.6262**	**0.6652**	**0.5899**	**0.5671**	**0.6298**	**0.5925**	**0.5645**
m=.8	0.7217	0.7086	0.8676	0.7272	0.7903	0.5709	0.5699	0.6289	0.5705	0.5392
m=.9	0.6756	0.6556	0.8199	0.6835	0.7451	0.5699	0.5699	0.6251	0.5699	0.5397

Table 7. The result using different learning decays (*d*). Acc. means accuracy, sen. means sensitivity, spec. means specificity, and prec. means precision.

	Train					Validation				
	Acc.	Sen.	Spec.	Prec.	AUC	Acc.	Sen.	Spec.	Prec.	AUC
d= 1e-5	0.7897	0.7809	0.9431	0.7947	0.8718	0.5816	0.5794	0.6432	0.5861	0.5584
***d*= lr/20**	**0.7711**	**0.7634**	**0.9318**	**0.7749**	**0.8525**	**0.5871**	**0.5814**	**0.6593**	**0.5864**	**0.5752**

Since the ImageNet does not contain medical brain images, we tried to leave all parameters to be trainable, which means they will be updated during backpropagation as the model learned the features, to see if there is any improvement in validation accuracy. The fine-tuning was attempted for learning momentum and decay, while the rests were retained as default.

As inscribed in Table 6, various learning momentums were utilized in the range .5 to .9. With m = .7, the smallest gap between training and validation accuracy was obtained, i.e., 3.22%, which was 1.03% lower than the best in the transfer learning approach. The validation accuracy also increased by 1.24% to 58.99%. Here, the F_1-score was 59.12%.

Now from the latest hyperparameter configurations (with m = .7), we tried two distinct learning decays, ie., 1e-5, and 5e-5 (derived from learning rate divided by the number of epoch). The results in Table 7 revealed that there is no more improvement in validation accuracy. It was instead decreased by 0.28% to 58.71%. Even when the second attempt seemed to have a higher validation accuracy result, it turns out that the better test accuracy was acquired from the first one, i.e., 57.56% (it means there is a 0.58% increase from the previous Sc3), while the second only yielded 55.33%.

5. DISCUSSION

Compared to the prior study [25] which used a combination model of CNN, recurrent neural network (RNN), STN, or CAM, our approach with InceptionResNetV2 delivered an improvement toward all models except the 2D CNN + 2D STN and 3D CNN + 3D STN model, which still has the difference of -1.4% and -2.4%, respectively. Nevertheless, the amount of data (ours only used data from 172 patients), input dimension (ours only used 2D image), and the model used must also be taken into consideration in observing the comparison results in Table 8.

Table 8. The comparison results with models in prior study. Acc. means testing accuracy.

Ref.	Source Dataset	Data Split	Neural Network Model	Acc.
[25]	ABIDE I+II: 1,992 patients (946 ASD and 1,046 TD)	80% train and 20% test	2D Input + 2D CNN + 2D STN	59%
			2D Input + 3D CNN + 2D STN	< 50%
			3D Input + 2D CNN + 3D STN	57%
			3D Input + 3D CNN + 3D STN	**60%**
			3D Input + 2D CNN + 3D STN + RNN	55%
			3D Input + 3D CNN + 3D STN + RNN	56%
			2D Input + 2D CNN + CAM	< 50%
			3D Input + 3D CNN + CAM	56%
Ours	ABIDE I/NYU: 172 patients (74 ASD and 98 TD)	70% train, 15% val., and 15% test	**2D Input + InceptionResNetV2**	**57.6%**

The base InceptionResNetV2 itself has 54,336,736 parameters in total, consists of many blocks of CNN [33] that make it a denser model compared to the prior research. Our result outperformed 2D CNN + 3D STN by 0.6%, 3D CNN + 3D STN + RNN and 3D CNN + CAM by 1.6%, and 2D CNN + 3D STN + RNN by 2.6%, 3D CNN + 2D STN and 2D CNN + CAM by >7.6%.

This improvement reconfirmed that the transfer learning approach is capable of delivering better performance, as stated in the introduction.

6. CONCLUSION

Our simple method to acquire and augment the only 172 neuroimages of ABIDE I dataset (NYU site) yielded an improvement from the previous method where the data used were the 1,992 neuroimages of ABIDE I and II [25]. With smaller dataset fed to the denser model such as InceptionResNetV2 can achieve up to 78.9% of training accuracy, up to 58.9% of validation accuracy, and up to 57.56% of testing accuracy. The best configurations that can be reported were by leaving the parameters untrained and by changing the momentum value from .5 to .7.

The improvement suggested that there is still ample room to prove that ASD can be detected from simpler brain scans (ours using the T2* or EPI), compared to the Functional Connectivity Matrix of associations between brain regions.

For future works, it is intriguing to expand the way of 2D/3D extraction from the origin 4D neuroimage. The development of a pre-trained model using the same base domain, i.e., brain scans/neuroimages instead of ImageNet can also be done since it has been proven to enhance performance and even lessen training time [32], [38]. Additionally, trying to construct a multimodal model where the patient phenotypic data are included may be worth further consideration.

ACKNOWLEDGEMENT

All models were trained using NVIDIA Tesla P100 GPU provided by NVIDIA – Bina Nusantara University Artificial Intelligence Research and Development Center.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders (DSM-5). 2013.

[2] B.Y. Lau, R. Leong, M. Uljarevic, et al. Anxiety in young people with autism spectrum disorder: Common and autism-related anxiety experiences and their associations with individual characteristics, Autism. 24 (2020), 1111–1126.

[3] M. Hosozawa, A. Sacker, N. Cable, Timing of diagnosis, depression and self-harm in adolescents with autism spectrum disorder, Autism. 25 (2021), 70–78.

[4] B. Restrepo, K. Angkustsiri, S.L. Taylor, et al. Developmental‐behavioral profiles in children with autism spectrum disorder and co‐occurring gastrointestinal symptoms, Autism Res. 13 (2020), 1778–1789.

[5] A. Posar, P. Visconti, Long-term outcome of autism spectrum disorder, Turk. Pediatr. Ars. 54 (2019), 207–212.

[6] S.L. Hyman, S.E. Levy, S.M. Myers, Identification, evaluation, and management of children with autism spectrum disorder, Pediatrics. 145 (2020), e20193447.

[7] P. Lanillos, D. Oliva, A. Philippsen, Y. Yamashita, Y. Nagai, G. Cheng, A review on neural network models of schizophrenia and autism spectrum disorder, Neural Netw. 122 (2020), 338–363.

[8] H. Hodges, C. Fealko, N. Soares, Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation, Transl Pediatr. 9 (2020), S55–S65.

[9] E. Karhu, R. Zukerman, R.S. Eshraghi, J. Mittal, R.C. Deth, A.M. Castejon, M. Trivedi, R. Mittal, A.A. Eshraghi, Nutritional interventions for autism spectrum disorder, Nutrit. Rev. 78 (2020), 515–531.

[10] C. Cheroni, N. Caporale, G. Testa, Autism spectrum disorder at the crossroad between genes and environment: contributions, convergences, and interactions in ASD developmental pathophysiology, Mol. Autism. 11 (2020), 69.

[11] K.A. Shaw, M.J. Maenner, J. Baio, et al. Early identification of autism spectrum disorder among children aged 4 years - early autism and developmental disabilities monitoring network, six sites, United States, 2016, MMWR Surveill. Summ. 69 (2020), 1–11.

[12] M.J. Maenner, K.A. Shaw, J. Baio, et al. Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2016, MMWR Surveill. Summ. 69 (2020) 1–12.

[13] S.A. Bedford, M.T.M. Park, G.A. Devenyi, et al. Large-scale analyses of the relationship between sex, age and intelligence quotient heterogeneity and cortical morphometry in autism spectrum disorder, Mol. Psych. 25 (2020), 614–628.

[14] W. Yassin, H. Nakatani, Y. Zhu, et al. Machine-learning classification using neuroimaging data in schizophrenia, autism, ultra-high risk and first-episode psychosis, Transl. Psych. 10 (2020), 278.

[15] T.W. Cenggoro, A. Budiarto, R. Rahutomo, B. Pardamean, Information System Design for Deep Learning Based Plant Counting Automation, in: 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), IEEE, Jakarta, Indonesia, 2018: pp. 329–332.

[16] T.W. Cenggoro, B. Mahesworo, A. Budiarto, et al. Features Importance in Classification Models for Colorectal Cancer Cases Phenotype in Indonesia, Procedia Computer Sci. 157 (2019), 313–320.

[17] K. Muchtar, F. Rahman, T.W. Cenggoro, A. Budiarto, B. Pardamean, An Improved Version of Texture-based Foreground Segmentation: Block-based Adaptive Segmenter, Procedia Computer Sci. 135 (2018), 579–586.

[18] M. Tang, P. Kumar, H. Chen, A. Shrivastava, Deep Multimodal Learning for the Diagnosis of Autism Spectrum Disorder, J. Imaging. 6 (2020), 47.

[19] A.S. Heinsfeld, A.R. Franco, R.C. Craddock, A. Buchweitz, F. Meneguzzi, Identification of autism spectrum disorder using deep learning and the ABIDE dataset, NeuroImage: Clinic. 17 (2018), 16–23.

[20] K. Sairam, J. Naren, G. Vithya, S. Srivathsan, Computer Aided System for Autism Spectrum Disorder Using Deep Learning Methods, Int. J. Psychosoc. Rehabil. 23 (2019), 418–425.

[21] Z. Sherkatghanad, M. Akhondzadeh, S. Salari, M. Zomorodi-Moghadam, M. Abdar, U.R. Acharya, R. Khosrowabadi, V. Salari, Automated Detection of Autism Spectrum Disorder Using a Convolutional Neural Network, Front. Neurosci. 13 (2020), 1325.

[22] X. Yang, M.S. Islam, A.M.A. Khaled, Functional connectivity magnetic resonance imaging classification of autism spectrum disorder using the multisite ABIDE dataset, in: 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), IEEE, Chicago, IL, USA, 2019: pp. 1–4.

[23] X. Yang, P.T. Schrader, N. Zhang, A deep neural network study of the ABIDE repository on autism spectrum classification, Int. J. Adv. Computer Sci. Appl. 11 (2020), 1-6.

[24] T. Eslami, V. Mirjalili, A. Fong, A.R. Laird, F. Saeed, ASD-DiagNet: a hybrid learning approach for detection of autism spectrum disorder using fMRI data, Front. Neuroinform. 13 (2019), 70.

[25] F. Ke, S. Choi, Y.H. Kang, K.-A. Cheon, S.W. Lee, Exploring the structural and strategic bases of autism spectrum disorders with deep learning, IEEE Access. 8 (2020), 153341–153352.

[26] B. Pardamean, H.H. Muljo, T.W. Cenggoro, B.J. Chandra, R. Rahutomo, Using transfer learning for smart building management system, J. Big Data. 6 (2019), 110.

[27] T.W. Cenggoro, F. Tanzil, A.H. Aslamiah, E.K. Karuppiah, B. Pardamean, Crowdsourcing annotation system of object counting dataset for deep learning algorithm, IOP Conf. Ser.: Earth Environ. Sci. 195 (2018), 012063.

[28] B. Pardamean, T.W. Cenggoro, B.J. Chandra, R. Rahutomo, Data annotation system for intelligent energy conservator in smart building, IOP Conf. Ser.: Earth Environ. Sci. 426 (2020), 012008.

[29] F.R. Lumbanraja, B. Mahesworo, T.W. Cenggoro, A. Budiarto, B. Pardamean, An Evaluation of Deep Neural Network Performance on Limited Protein Phosphorylation Site Prediction Data, Procedia Computer Sci. 157 (2019), 25–30.

[30] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (2010) 1345–1359.

[31] R.E. Caraka, N.T. Nugroho, S.K. Tai, R.C. Chen, T. Toharudin, B. Pardamean, Feature importance of the aortic anatomy on endovascular aneurysm repair (Evar) using boruta and bayesian mcmc, Commun. Math. Biol. Neurosci. 2020 (2020), 2020, 22.

[32] B. Pardamean, T.W. Cenggoro, R. Rahutomo, A. Budiarto, E.K. Karuppiah, Transfer Learning from Chest X-Ray Pre-trained Convolutional Neural Network for Learning Mammogram Data, Procedia Computer Sci. 135 (2018), 400–407.

[33] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning, ArXiv:1602.07261 [Cs]. (2016). http://arxiv.org/abs/1602.07261.

[34] M. Lin, Q. Chen, S. Yan, Network In Network, ArXiv:1312.4400 [Cs]. (2014). http://arxiv.org/abs/1312.4400.

[35] S. Ruder, An overview of gradient descent optimization algorithms, ArXiv:1609.04747 [Cs]. (2017).

[36] T. Saito, M. Rehmsmeier, The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating

Binary Classifiers on Imbalanced Datasets, PLoS ONE. 10 (2015), e0118432.

[37] Farida, R.E. Caraka, T.W. Cenggoro, B. Pardamean, Batik Parang Rusak Detection Using Geometric Invariant Moment, in: 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), IEEE, Jakarta, Indonesia, 2018: pp. 71–74.

[38] I.W. Harsono, S. Liawatimena, T.W. Cenggoro, Lung nodule detection and classification from Thorax CT-scan using RetinaNet with transfer learning, J. King Saud Univ. - Computer Inform. Sci. In Press, https://doi.org/10.1016/j.jksuci.2020.03.013