

## VISPNN: VGG-Inspired Stochastic Pooling Neural Network

Shui-Hua Wang<sup>1</sup>, Muhammad Attique Khan<sup>2</sup> and Yu-Dong Zhang<sup>3,\*</sup>

<sup>1</sup>School of Mathematics and Actuarial Science, University of Leicester, Leicester, LE1 7RH, United Kingdom

<sup>2</sup>Department of Computer Science, HITEC University Taxila, Taxila, 47080, Pakistan

<sup>3</sup>School of Informatics, University of Leicester, Leicester, LE1 7RH, United Kingdom

\*Correspondence Author: Yu-Dong Zhang. Email: yudongzhang@ieee.org

Received: 14 April 2021; Accepted: 28 June 2021

**Abstract:** **Aim** Alcoholism is a disease that a patient becomes dependent or addicted to alcohol. This paper aims to design a novel artificial intelligence model that can recognize alcoholism more accurately. **Methods** We propose the VGG-Inspired stochastic pooling neural network (VISPNN) model based on three components: (i) a VGG-inspired mainstay network, (ii) the stochastic pooling technique, which aims to outperform traditional max pooling and average pooling, and (iii) an improved 20-way data augmentation (Gaussian noise, salt-and-pepper noise, speckle noise, Poisson noise, horizontal shear, vertical shear, rotation, Gamma correction, random translation, and scaling on both raw image and its horizontally mirrored image). In addition, two networks (Net-I and Net-II) are proposed in ablation studies. Net-I is based on VISPNN by replacing stochastic pooling with ordinary max pooling. Net-II removes the 20-way data augmentation. **Results** The results by ten runs of 10-fold cross-validation show that our VISPNN model gains a sensitivity of  $97.98 \pm 1.32$ , a specificity of  $97.80 \pm 1.35$ , a precision of  $97.78 \pm 1.35$ , an accuracy of  $97.89 \pm 1.11$ , an F1 score of  $97.87 \pm 1.12$ , an MCC of  $95.79 \pm 2.22$ , an FMI of  $97.88 \pm 1.12$ , and an AUC of 0.9849, respectively. **Conclusion** The performance of our VISPNN model is better than two internal networks (Net-I and Net-II) and ten state-of-the-art alcoholism recognition methods.

**Keywords:** Deep learning; alcoholism; multiple-way data augmentation; VGG; convolutional neural network; stochastic pooling

### 1 Introduction

Alcoholism (also known as alcohol use disorder) is a disease that a patient becomes dependent or addicted to alcohol [1]. Patients with alcoholism continue to drink alcohol; even drinking causes negative consequences to themselves. The difference between alcoholism and alcohol abuse is people of alcohol abuse are not physically dependent on alcohol [2]. Excessive alcohol use will damage all organ systems, but it mainly influences the heart, liver, brain, pancreas, and immune system [3]. Besides, alcoholism may cause schizophrenia, bipolar disorder, depression [4], irregular heartbeat [5], Wernicke–Korsakoff syndrome [6], etc.



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

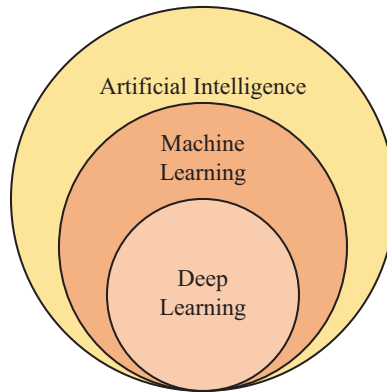
Long-term alcoholism affects the brain, i.e., the volumes of white matter and grey matter of patients are smaller than age-matched controls. The brain shrinkages and lateral ventricle enlargement [7] caused by alcoholism can be observed via magnetic resonance imaging (MRI) scanning, which facilitates doctors diagnosing alcoholism.

Nevertheless, the diagnosis of alcoholism is mainly based on manual observation of brain images in the current clinical routine, which is naturally labor-intensive and onerous. Mainly, the slight shrinkage of alcoholism brains in the early prodromal stage [8] associated with mild symptoms is susceptible to be neglected by radiologists and clinicians, which may trigger costs to the patient and his/her family. In light of the above limitations, accurate and fast diagnostic artificial intelligence (AI) models to recognize alcoholism are beneficial to patients, families, and society.

In the past, various AI models have been proposed to recognize alcoholism. Fig. 1 shows the relationship between AI with machine learning (ML) and deep learning (DL). ML is a subfield of AI, while DL is a subfield of ML. Hou [9] brought about a novel algorithm—Predator-prey Adaptive-inertia Chaotic Particle Swarm Optimization (PACPSO). The authors applied the PACPSO to identify alcoholism. Jenitta et al. [10] presented a local mesh vector co-occurrence pattern (LMVCoP) feature for assisting diagnosis. This method can be used for the application of alcoholism identification. Han [11] proposed a three-segmented encoded Jaya (3SJ) method to identify alcoholism. The authors found the 3SJ gave better performance than other optimization methods, such as multi-objective genetic algorithm, plain Jaya, bee colony optimization, particle swarm optimization, etc. Lima [12] presented a novel method utilizing Haar wavelet transform (HWT) to extract features from brain scanning images of patients. Their method achieved an accuracy of  $81.57 \pm 2.18\%$  on their dataset. Afterward, Macdonald [13] presented a wavelet energy logistic regression (WELR) model. The authors used 5-fold stratified cross-validation to verify the performances of their model. Qian [14] proposed a computer vision-based technique that utilizes cat swarm optimization (CSO), which mimics the behaviour of the cats. In their experiment, CSO was demonstrated to have better performances than four bio-inspired algorithms. Chen [15] presented a new model combining support vector machine (SVM) with genetic algorithm (GA). The combined model is abbreviated as SVMGA. The authors stated their model was effective in alcoholism detection, showing an average accuracy of  $88.68 \pm 0.30\%$ . Chen [16] presented an AI model based on a linear regression classifier (LRC) for alcoholism detection.

Recently, DL techniques have been successfully applied to alcoholism recognition. Lv [17] created a 7-layer convolutional neural network (CNN). Their experiments showed stochastic pooling (SP) provided better performance than other pooling methods. Nevertheless, their CNN structure was simple, so its expression ability was limited. Xie [18] used the AlexNet transfer learning (ANTL) model. The authors fine-tuned their model and tested five different replacement configurations.

There are some other ML methods based on different data sources. For example, Kamarajan et al. [19] used random forest and Electroencephalogram (EEG) source functional connectivity, neuropsychological functioning, and impulsivity measures to classify alcohol use disorder. Quaglieri et al. [20] harnessed functional MRI (fMRI) to analyze brain network underlying executive functions in gambling and alcohol use disorder. Many other scanning modalities and protocols may help identify alcoholism; however, we focus on MRI in this study due to its high-resolution three-dimensional imaging ability.



**Figure 1:** Relationship among AI, ML, and DL

The motivation of this paper is to propose a novel model, VGG-inspired stochastic pooling neural network (VISPNN), for alcoholism recognition with an expectation to obtain better performance than existing alcoholism identification approaches. The contributions of our study are the following four aspects.

- (a) A VGG-inspired network is used as a mainstay network.
- (b) Stochastic pooling is used to replace traditional max pooling.
- (c) Improved multiple-way data augmentation is proposed to avoid overfitting.
- (d) Our model is proven to render better performances than state-of-the-art methods.

## 2 Methodology

### 2.1 Introduction of VGG

[Tab. 1](#) displays the abbreviation list in this study for ease of reading. First, we introduce VGG, which stands for Visual Geometry Group, an academic group at Oxford University. The VGG team presented two renowned networks: VGG-16 [21] and VGG-19, encompassed as library packages of prevalent programming language platforms, e.g., Python and MATLAB.

[Fig. 2](#) displays the structure of VGG-16, which is composed of five convolutional blocks (CBs) and three fully connected layers (FCLs). The size of the input of VGG-16 is  $224 \times 224 \times 3$ . After the 1st CB, the output is  $112 \times 112 \times 64$ . Components of 1st CB are shown in [Tab. 2](#). The 1st CB can be written as

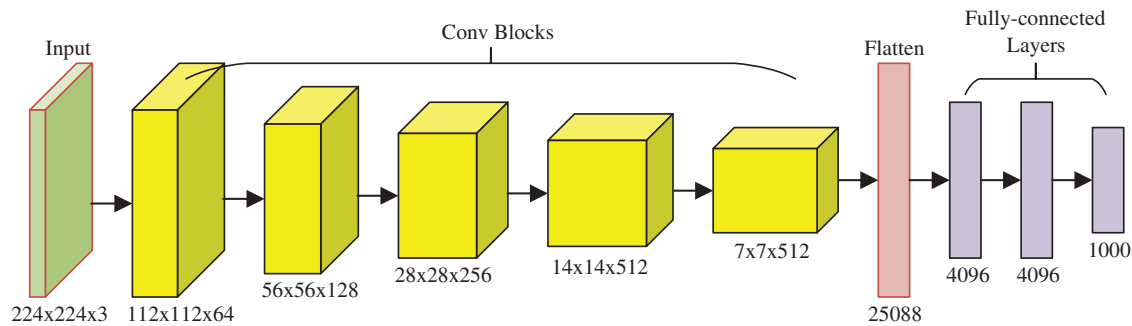
$$a \times (bc \times c) / d \quad (1)$$

which means “ $a$  repetitions of  $b$  kernels with sizes of  $c \times c$  followed by a max-pooling with a size of  $d \times d$ .”

Note that (i) the activation function: rectified linear unit (ReLU) layers are skipped in the subsequent texts as default. (ii) Stride and padding are not reported because they can be calculated easily. The five CBs are itemized in [Tab. 3](#), and the feature map (FM) of the output is displayed in the final column. After five CBs, the FM is compressed from  $7 \times 7 \times 512$  to a vector with a size of 25,088 neurons. Three FCLs with neurons of 4096, 4096, and 1000 are appended at last.

**Table 1:** Abbreviation List

Abbreviation	Full meaning	Abbreviation	Full meaning
AI	artificial intelligence	ML	machine learning
AP	average pooling	MP	max pooling
AUC	the area under the curve	MRI	magnetic resonance imaging
CB	convolutional block	MSD	mean and standard deviation
DA	data augmentation	NWL	number of weighted layers
DL	deep learning	PM	probability map
EEG	Electroencephalogram	PN	Poisson noise
FCL	fully connected layer	ReLU	rectified linear unit
FM	feature map	RLV	random location vector
FMI	Fowlkes–Mallows index	ROC	receiver operating characteristic
fMRI	functional MRI	SN	speckle noise
L2P	l2-norm pooling	SP	stochastic pooling
MCC	Matthews correlation coefficient	VGG	visual geometry group

**Figure 2:** Diagram of VGG-16 network structures**Table 2:** Components of 1st CB in VGG-16

Layer	Component
1	1 convolutional layer with 64 kernels with sizes of $3 \times 3$ and stride $[1, 1]$ and padding $[1, 1, 1, 1]$
2	1 ReLU layer
3	1 convolutional layer with 64 kernels with sizes of $3 \times 3$ and stride $[1, 1]$ and padding $[1, 1, 1, 1]$
4	1 ReLU layer
5	one max-pooling layer with a size of $2 \times 2$

## 2.2 Improvement I: Stochastic Pooling

Within the standard CNNs, pooling is an essential component followed by a convolution layer (See Layer 5 in [Tab. 2](#)) to reduce the size of FMs. Traditional pooling methods are either max-pooling (MP) or average pooling (AP).

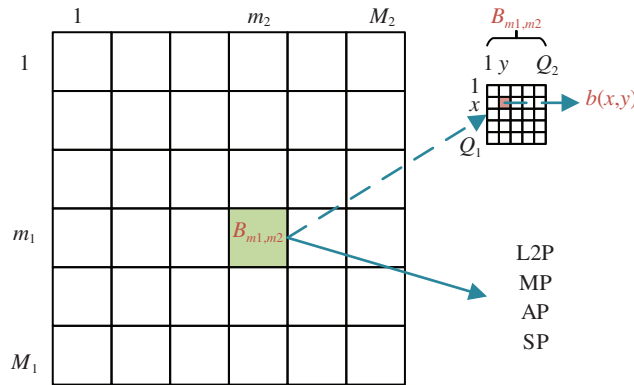
**Table 3:** Five CBs in VGG-16 Net. The definition of  $a, b, c, d$  can be found in Eq. (1)

CB	$a$	$b$	$c$	$d$	FM
1	2	64	3	2	$112 \times 112 \times 64$
2	2	128	3	2	$56 \times 56 \times 128$
3	3	256	3	2	$28 \times 28 \times 256$
4	3	512	3	2	$14 \times 14 \times 512$
5	3	512	3	2	$7 \times 7 \times 512$

Suppose we have an FM, which can be separated into  $M_1 \times M_2$  blocks, in which every block has the size of  $Q_1 \times Q_2$ . Now let us focus on the  $m_1$ -th row and  $m_2$ -th column block

$$B_{m_1, m_2} = \{b(x, y), x = 1, \dots, Q_1, y = 1, \dots, Q_2\}. \tag{2}$$

where  $1 \leq m_1 \leq M_1, 1 \leq m_2 \leq M_2$  (See Fig. 3).



**Figure 3:** Diagram of block-wise processing

The strided convolution (SC) traverses the input activation map with the strides, which equals the size of the block  $(Q_1, Q_2)$ , so here its output is set to

$$B_{m_1, m_2}^{SC} = b(1, 1). \tag{3}$$

The  $l_2$ -norm pooling (L2P), average pooling (AP), and max pooling (MP) generate the  $l_2$ -norm value, average value, and maximum value within the block  $B_{m_1, m_2}$ , respectively. Nevertheless, the AP outputs the average, downscaling the greatest values, where the essential features may lie. In contrast, MP stores the most significant value but deteriorates the overfitting obstacle.

$$\begin{cases} B_{m_1, m_2}^{L2P} = \sqrt{\frac{\sum_{x=1}^{Q_1} \sum_{y=1}^{Q_2} b^2(x, y)}{Q_1 \times Q_2}} \\ B_{m_1, m_2}^{AP} = \frac{1}{Q_1 \times Q_2} \sum_{x=1}^{Q_1} \sum_{y=1}^{Q_2} b(x, y) \\ B_{m_1, m_2}^{MP} = \max_{x=1}^{Q_1} \max_{y=1}^{Q_2} b(x, y) \end{cases} \tag{4}$$

Alternatively, stochastic pooling (SP) provides a way to the defects of AP and MP. Successful application cases are using SP in the stochastic resonance model [22], COVID-19 recognition [23], etc. SP is a four-step procedure. Step 1 generates the probability map (PM) for each entry in the block  $B_{m_1, m_2}$ .

$$\begin{cases} S(x, y) = \frac{b(x, y)}{\sum_{x=1}^{Q_1} \sum_{y=1}^{Q_2} b(x, y)} \\ s.t. \sum_{x=1}^{Q_1} \sum_{y=1}^{Q_2} S(x, y) = 1 \end{cases} \quad (5)$$

where  $S(x, y)$  stands for the PM value at pixel  $(x, y)$ .

Step 2 creates a random location vector (RLV)  $r = (x_r, y_r)$  that takes the discrete probability distribution (DPD) as

$$\begin{cases} Z[r = (1, 1)] = S(1, 1) & Z[r = (1, 2)] = S(1, 2) & \cdots & Z[r = (1, Q_2)] = S(1, Q_2) \\ Z[r = (2, 1)] = S(2, 1) & Z[r = (2, 2)] = S(2, 2) & \cdots & Z[r = (2, Q_2)] = S(2, Q_2) \\ \cdots & \cdots & \cdots & \cdots \\ Z[r = (Q_1, 1)] = S(Q_1, 1) & Z[r = (Q_1, 2)] = S(Q_1, 2) & \cdots & Z[r = (Q_1, Q_2)] = S(Q_1, Q_2) \end{cases} \quad (6)$$

where  $Z$  represents the probability.

Step 3, a sample location vector  $r_0$  is drawn from the RLV  $r$ , and we have  $r_0 = (x_{r_0}, y_{r_0})$ .

Step 4, the output of SP is at location  $r_0$ , namely

$$B_{m_1, m_2}^{SP} = b(x_{r_0}, y_{r_0}) \quad (7)$$

Fig. 4 gives an example where we can observe the 1-st row and 2-nd column block  $B_{12}$ , outlined by red rectangle. The L2P calculates

$$B_{1,2}^{L2P} = \sqrt{\frac{\sum (6.4^2 + 7.3^2 + 6.4^2 + 9.5^2 + 5.7^2 + 0.5^2 + 3.6^2 + 1.1^2 + 7.8^2)}{9}} = 6.08 \quad (8)$$

The AP and MP output the pooling values of 5.37 and 9.5, respectively. In contrast, the SP generates the PM and selects the top-left pixel  $b(1, 1) = 0.13$  in the block  $B_{12}$  and output the SP value of 6.4.

### 2.3 Improvement 2: VGG-Inspired Stochastic Pooling Neural Network

A novel VGG-inspired mainstay network is proposed. Tab. 4 shows the structure of the proposed 10-layer VGG-inspired mainstay network. The definition of  $(a, b, c, d)$  can be found in Tabs. 2 and 3. The variables  $(e, f)$  represents the weights and biases of FCL, respectively. The NWL in Tab. 4 represents the number of weighted layers. The total weighted layers in this VGG-inspired mainstay network is  $1 + 2 + 3 + 2 + 1 + 1 = 10$ . We can observe that after four CBs, the size of output FM is  $11 \times 11 \times 64$ , which is flattened to a vector of 7744, which is then sent through an FCL with 100 neurons, finally outputting two neurons indicating alcoholism or healthy.

The structure of our VGG-inspired mainstay network is displayed in Fig. 5a. If we replace the max-pooling in each CB with stochastic pooling, we can get the proposed VGG-Inspired Stochastic Pooling Neural Network (VISPNN), as shown in Fig. 5b.

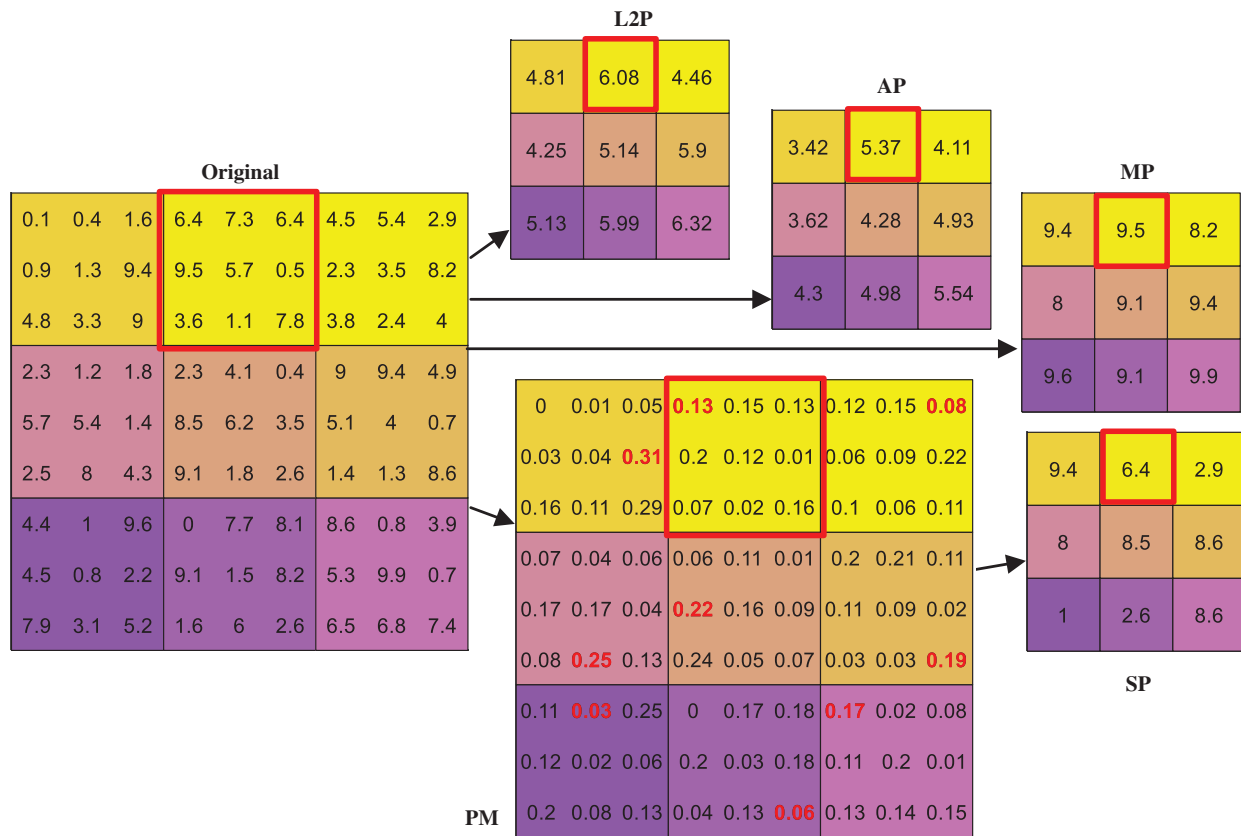


Figure 4: Comparison of different pooling methods

Table 4: Proposed 10-layer VGG-inspired mainstay network

Index	Tag	NWL	Trainable parameters	Size of FM
1	Input	0		$176 \times 176 \times 1$
2	CB-1	1	$a=1, b=3, c=16, d=2$	$88 \times 88 \times 16$
3	CB-2	2	$a=2, b=3, c=32, d=2$	$44 \times 44 \times 32$
4	CB-3	3	$a=3, b=3, c=64, d=2$	$22 \times 22 \times 64$
5	CB-4	2	$a=2, b=3, c=64, d=2$	$11 \times 11 \times 64$
6	Flatten	0		7744
7	FCL-1	1	$e=100 \times 7744, f=100 \times 1$	100
9	FCL-3	1	$e=2 \times 200, f=2 \times 1$	2

### 2.4 Improvement 3: 20-Way Data Augmentation

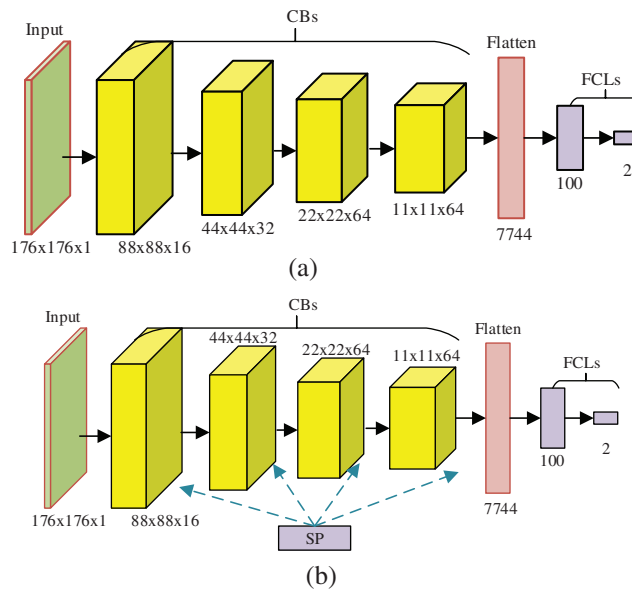
The dataset in this study was reported in Ref. [18], composed of 188 alcoholic brain images and 191 non-alcoholic brain images. Fig. 6 shows two samples of our dataset.

The relatively small dataset may breed the overfitting problem. To avoid this, data augmentation (DA) is a powerful tool because it can generate fake images on the training set. Cheng [23] presented a 16-way DA, in which 8 DA techniques were applied on both  $q(k)$  and  $q^{(h)}(k)$ . The multiple-way DA shows better performance than traditional DA. This study is based on 16-way

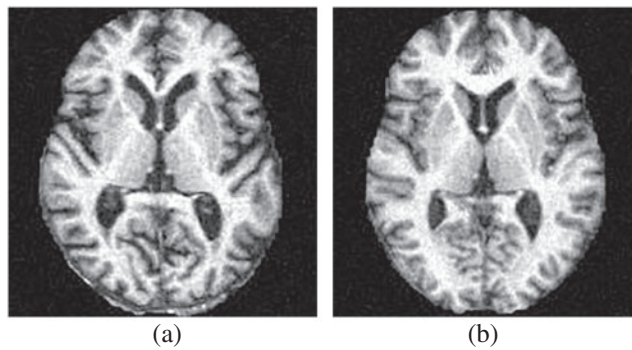
DA from Cheng [23]; furthermore, we add two new DAs on both  $q(k)$  and  $q^{(h)}(k)$ . One DA is speckle noise (SN), which alters the image as

$$q^{SN}(k) = q(k) + N_{SN} \times q(k) \quad (9)$$

where  $N_{SN}$  is uniformly distributed random noise. The mean and variance of  $N_{SN}$  are set to 0 and 0.05, respectively.



**Figure 5:** Diagram of two proposed networks (a) VGG-inspired mainstay network (b) proposed VISPNN



**Figure 6:** Samples of our dataset (a) alcoholism and (b) HC

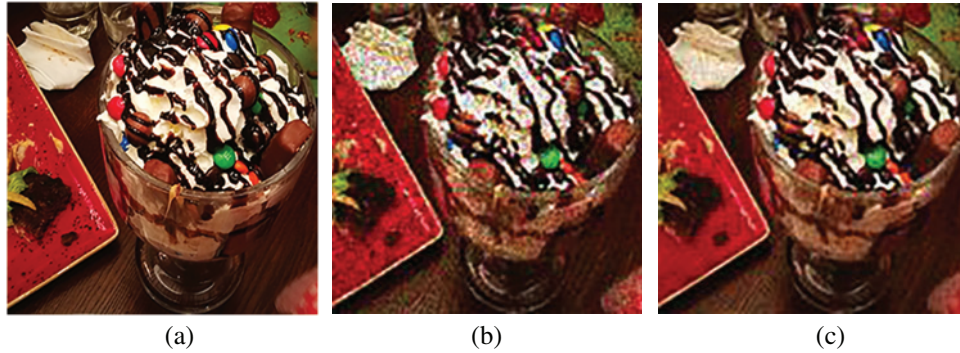
The other DA is Poisson noise (PN). In the electronics field, PN originates from the discrete nature of the electric charge. Instead of adding artificial noise to the raw image, we generate PN from the raw image. The pixel values of raw images are stored in uint 8 format; if a pixel has the value of 20, then the corresponding pixel  $q^{PN}$  of the PN altered image will be generated from a



Poisson distribution with a mean of 20. Mathematically

$$\begin{cases} q^t(k|x,y) \sim P_o[q(k|x,y)] \\ q^{PN}(k|x,y) = \min\{256, q^t(k|x,y)\} \end{cases} \quad (10)$$

where  $P_o(\lambda)$  represents a Poisson distribution with a mean of  $\lambda$ , and  $(x,y)$  are the coordinates.  $q^t$  is the temporary variable. The min function helps the final output is within the value of  $[0, 256]$ . Using a colourful natural image can observe how those two noises alter the image, as shown in Fig. 7.



**Figure 7:** Raw image and its altered images by two new Das (a) raw image  $q(k)$ , (b) SN altered  $q^{SN}(k)$  and (c) PN altered  $q^{PN}(k)$

With the help of SN and PN, we can propose a novel 20-way DA, First,  $M_1$  different DA methods as shown in Fig. 8 are applied to  $q(k)$ . Let  $\mathbf{K}_m, m = 1, \dots, M_1$  be each DA operation, we have  $M_1$  augmented datasets on raw image  $q(k)$  as:

$$\mathbf{K}_m[q(k)], m = 1, \dots, M_1. \quad (11)$$

Let  $M_2$  stand for the size of generated new images for each DA method, we have

$$|\mathbf{K}_m[q(k)]| = M_2. \quad (12)$$

Second, horizontally mirrored image is generated as:

$$q^{(h)}(k) = y_a[q(k)] \quad (13)$$

where  $y_a$  stands for horizontal mirror function.

Third, all the  $M_1$  different DA methods are performed on the mirrored image  $q^{(h)}(k)$ . We generate  $M_1$  new datasets as:

$$\begin{cases} \mathbf{K}_m[q^{(h)}(k)], m = 1, \dots, M_1 \\ |\mathbf{K}_m[q^{(h)}(k)]| = M_2, m = 1, \dots, M_1 \end{cases} \quad (14)$$

Fourth, the raw image  $q(k)$ , the mirrored image  $q^{(h)}(k)$ , all the above  $M_1$ -way results of raw image  $\mathbf{K}_m[q(k)]$ , and all  $M_1$ -way DA results of horizontal mirrored image  $\mathbf{K}_m[q^{(h)}(k)]$  are

combined. The final generated dataset from  $q(k)$  is defined as  $\mathbf{G}(k)$ :

$$q(k) \mapsto \mathbf{G}(k) = y_b \left\{ \begin{array}{cc} \underbrace{q(k)}_{M_2} & \underbrace{q^{(h)}(k)}_{M_2} \\ \underbrace{\mathbf{K}_1[q(k)]}_{M_2} & \underbrace{\mathbf{K}_1[q^{(h)}(k)]}_{M_2} \\ \dots & \dots \\ \underbrace{\mathbf{K}_{M_1}[q(k)]}_{M_2} & \underbrace{\mathbf{K}_{M_1}[q^{(h)}(k)]}_{M_2} \end{array} \right\} \quad (15)$$

where  $y_b$  stands for the concatenation function. Suppose augmentation factor is  $M_3$ , which stands for the number of images in  $\mathbf{G}(k)$ , we obtain

$$M_3 = \frac{|\mathbf{G}(k)|}{|q(k)|} = \frac{(1 + M_1 \times M_2) \times 2}{1} = 2 \times M_1 \times M_2 + 2 \quad (16)$$

Algorithm 1 summarizes the pseudocode of the proposed 20-way DA method. In this study, we set  $M_1 = 10$ , i.e., a 20-way DA. We also set  $M_2 = 30$ , thus  $M_3 = 602$ , indicating each raw training image will generate 602 images, including the raw image itself.

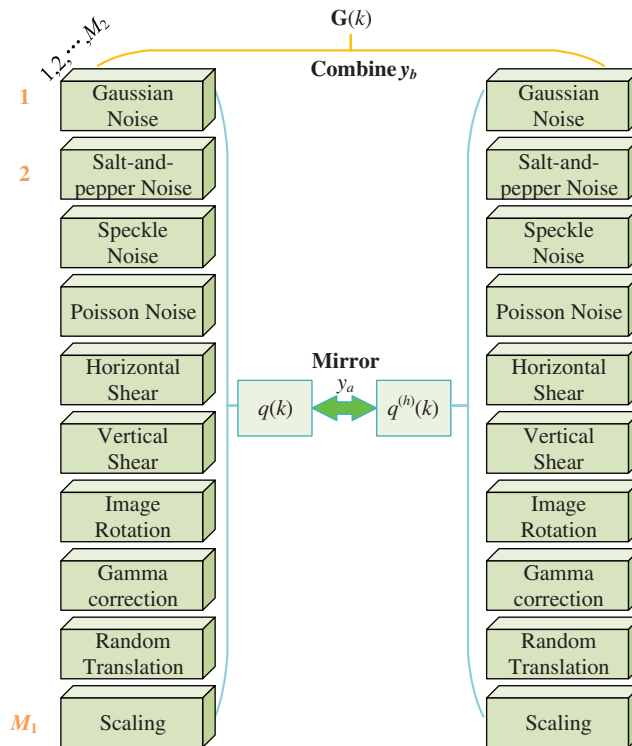


Figure 8: Diagram of proposed 20-way DA

---

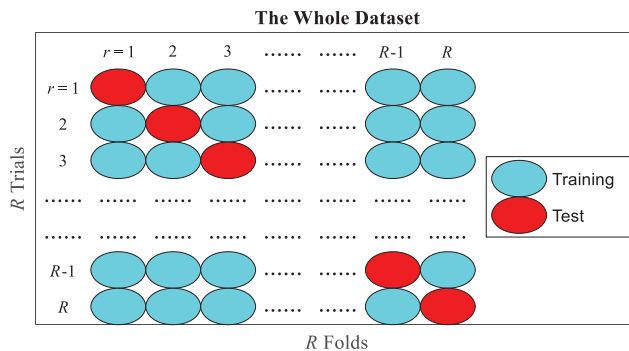
**Algorithm 1:** Pseudocode of proposed 20-way DA on the  $k$ -th training image

---

- Step 1 Import raw preprocessed the  $k$ -th training image  $q(k)$ .
  - Step 2  $M_1$  geometric or photometric or noise-injection DA transforms  $\mathbf{K}_m$  are utilized on  $q(k)$ .  
We obtain  $\mathbf{K}_m[q(k)], m = 1, \dots, M_1$ . See Eq. (11)  
Each enhanced dataset contains  $M_2$  new images. See Eq. (12).
  - Step 3 A horizontally mirrored image is generated as  $q^{(h)}(k) = y_a[q(k)]$ . See Eq. (13).
  - Step 4  $M_1$ -way data augmentation methods are carried out on  $q^{(h)}(k)$ ,  
we obtain  $\mathbf{K}_m[q^{(h)}(k)], m = 1, \dots, M_1$ . See Eq. (14).
  - Step 5 The raw image, the mirrored image, all the above  $M_1$ -way DA results of the raw image,  
and all  $M_1$ -way DA results of the horizontal mirrored image are combined via  $y_b$ .  
See Eq. (15).
  - Step 6 A new dataset  $\mathbf{G}(k)$  is generated with the number of images as  $M_3 = 2 \times M_1 \times M_2 + 2$ .  
See Eq. (16).
- 

**2.5 Implementation**

$R$ -fold cross-validation is employed. The whole dataset is divided into  $R$  folds [24]. At  $r$ -th trial,  $1 \leq r \leq R$ , the  $r$ -th fold is picked up as the test, and the rest  $R - 1$  folds:  $[1, \dots, r - 1, r + 1, \dots, R]$  are chosen as training set (Fig. 9). We let  $R = 10$ , namely a 10-fold cross validation. Furthermore, we run the 10-fold cross validation ten times.



**Figure 9:** Diagram of  $R$ -fold cross validation

**2.6 Measures**

Seven measures are used based on the confusion matrix of 10 runs of 10-fold cross-validation. Let  $I$  stands for the confusion matrix

$$I = \begin{bmatrix} I(1,1) & I(1,2) \\ I(2,1) & I(2,2) \end{bmatrix} \tag{17}$$

where  $I(1,1)$  means true positive,  $I(1,2)$  false negative (FN),  $I(2,1)$  false positive, and  $I(2,2)$  true negative. Sensitivity, specificity, precision, and accuracy are already familiarized to readers, so we will not give their definitions. Besides, we use the F1 score, Matthews correlation coefficient (MCC), and Fowlkes–Mallows index (FMI).

$$F1 = \frac{2I(1,1)}{2I(1,1) + I(2,1) + I(1,2)}, \tag{18}$$

$$MCC = \frac{I(1,1) \times I(2,2) - I(2,1) \times I(1,2)}{\sqrt{(I(1,1) + I(2,1)) \times (I(1,1) + I(1,2)) \times (I(2,2) + I(2,1)) \times (I(2,2) + I(1,2))}}, \quad (19)$$

$$FMI = \sqrt{\frac{I(1,1)}{I(1,1) + I(2,1)} \times \frac{I(1,1)}{I(1,1) + I(1,2)}}, \quad (20)$$

The receiver operating characteristic (ROC) curve is used to provide a graphical plot of measuring AI models. First, the ROC curve is produced by plotting the true positive rate against the false-positive rate at various threshold levels. Then, the area under the curve (AUC) is calculated via ROC.

### 3 Results

#### 3.1 20-Way DA Results

Fig. 10 shows the  $M_1$ -way DA results of raw image, which is chosen as Fig. 6a. Due to the page limit, we do not display the horizontally mirrored image and the corresponding  $M_1$ -way DA results.

#### 3.2 Statistical Results of Proposed Method

Tab. 5 itemizes the statistical results (10 runs of 10-fold cross-validation) of the proposed VISPNP method. The mean and standard deviation (MSD) over ten runs are displayed in the last row. It shows our model reaches a sensitivity of  $97.98 \pm 1.32$ , a specificity of  $97.80 \pm 1.35$ , an accuracy of  $97.89 \pm 1.11$ , respectively.

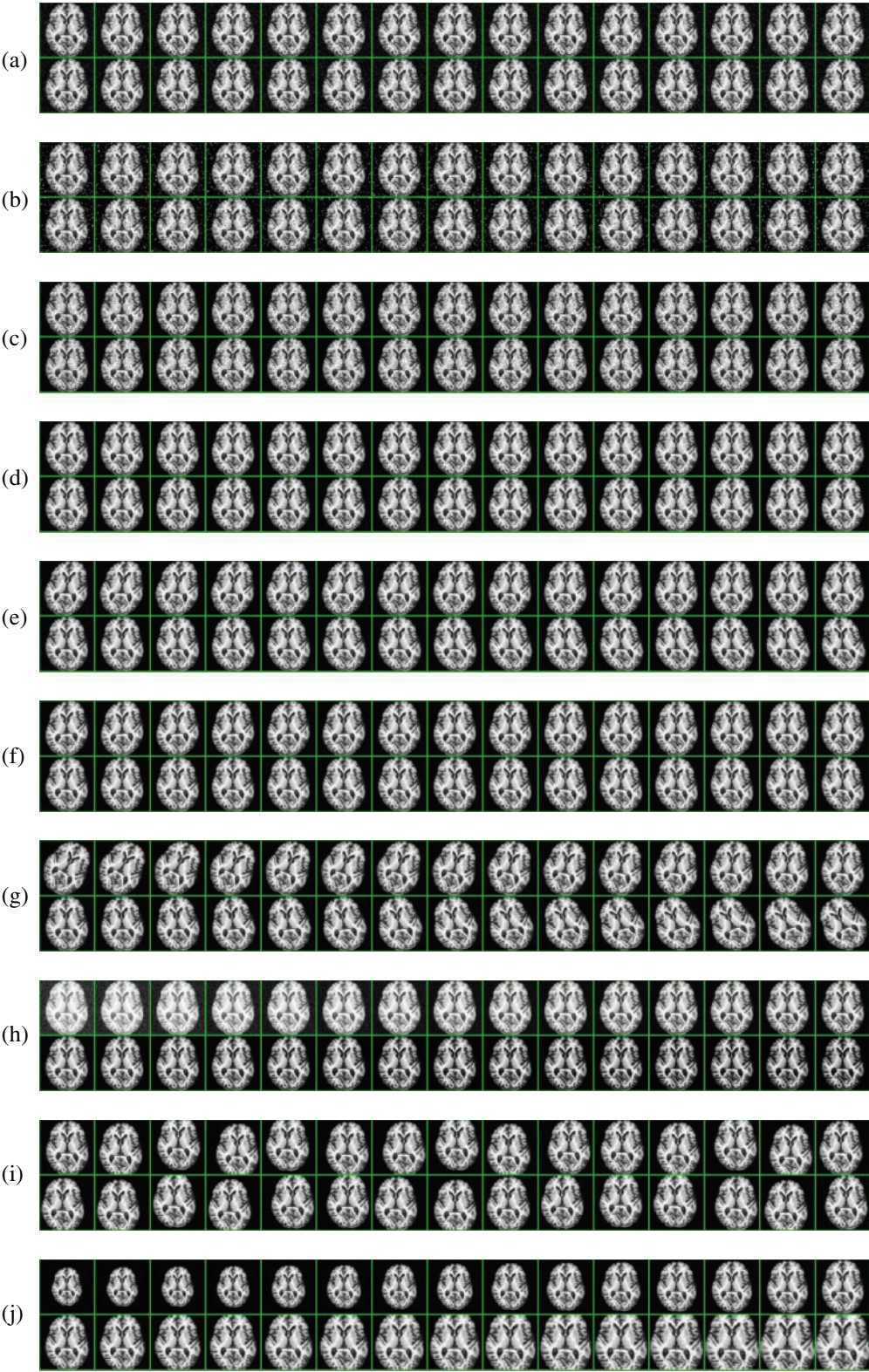
#### 3.3 Ablation Studies

An ablation study is a procedural experiment that removes a network's submodule to understand that submodule better. Two ablation studies are carried out: (i) Net-I: We remove stochastic pooling from the proposed VISPNP model and use max-pooling to replace the removed layers. Thus, the network is named Net-I. (ii) Net-II: We remove the multiple-way data augmentation. The resultant network is named Net-II. The comparison of our VISPNP model with Net-I and Net-II is shown in Tab. 6.

Fig. 11 displays the ROC curves comparison of the proposed VISPNP model with Net-I and Net-II. The blue patches correspond to the lower and upper confidence bounds. The AUC of Net-I is 0.9683, compared to that of VISPNP of 0.9849. Therefore, we can observe stochastic pooling indeed increase performances. Meanwhile, the AUC of Net-II is 0.9602, which is a significant drop from VISPNP (0.9849). This drop reflects that multiple-way data augmentation can significantly increase the prediction performance due to its ability to generate diverse "fake" training images.

#### 3.4 Comparison to Other Alcoholism Recognition Methods

This proposed VISPNP model is compared with 10 state-of-the-art alcoholism recognition methods: PACPSO [9], LMVCoP [10], WRE [11], HWT [12], WELR [13], CSO [14], SVMGA [15], LRC [16], CNNSP [17], and ANTL [18], respectively. The comparison results are itemized in Tab. 7, with the cognate bar plot shown in Fig. 12 which ranks all the methods in order of MCC.



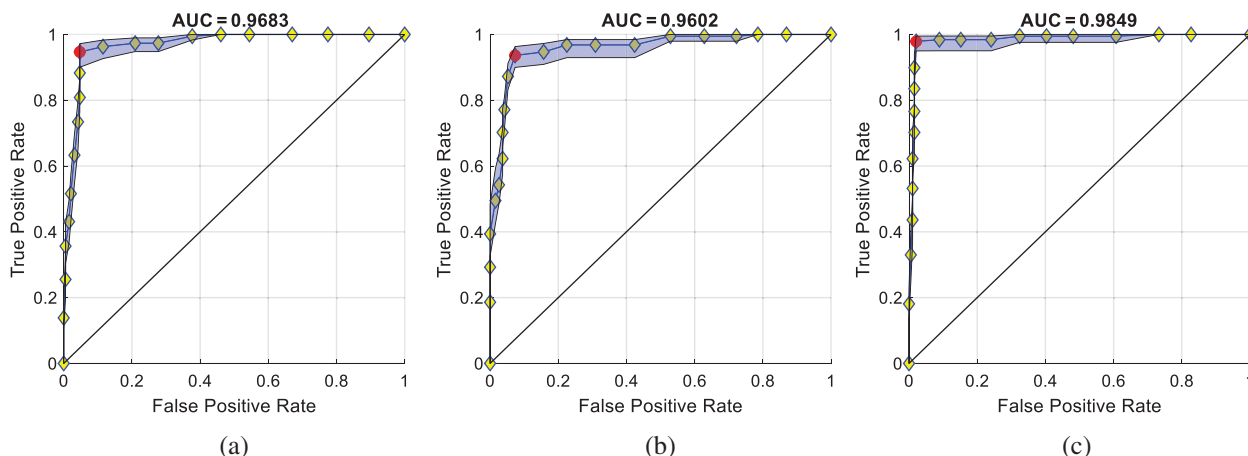
**Figure 10:** 20-way DA results (we only show 10-way DA results on the raw image here). (a) Gaussian noise (b) Salt-and-pepper noise (c) Speckle noise (d) Poisson noise (e) Horizontal shear (f) Vertical shear (g) Image rotation (h) Gamma correction (i) Random translation (j) Scaling

**Table 5:** Statistical results of proposed VISPNN method

Run	SEN	SPC	PRC	ACC	F1	MCC	FMI
1	97.34	95.81	95.81	96.57	96.57	93.15	96.57
2	97.87	97.91	97.87	97.89	97.87	95.78	97.87
3	95.74	96.86	96.77	96.31	96.26	92.62	96.26
4	98.94	96.86	96.88	97.89	97.89	95.80	97.90
5	99.47	98.95	98.94	99.21	99.20	98.42	99.20
6	100.00	100.00	100.00	100.00	100.00	100.00	100.00
7	97.87	98.95	98.92	98.42	98.40	96.84	98.40
8	97.87	96.86	96.84	97.36	97.35	94.73	97.36
9	98.40	96.86	96.86	97.63	97.63	95.26	97.63
10	96.28	98.95	98.91	97.63	97.57	95.28	97.58
MSD	$97.98 \pm 1.32$	$97.80 \pm 1.35$	$97.78 \pm 1.35$	$97.89 \pm 1.11$	$97.87 \pm 1.12$	$95.79 \pm 2.22$	$97.88 \pm 1.12$

**Table 6:** Results of ablation studies

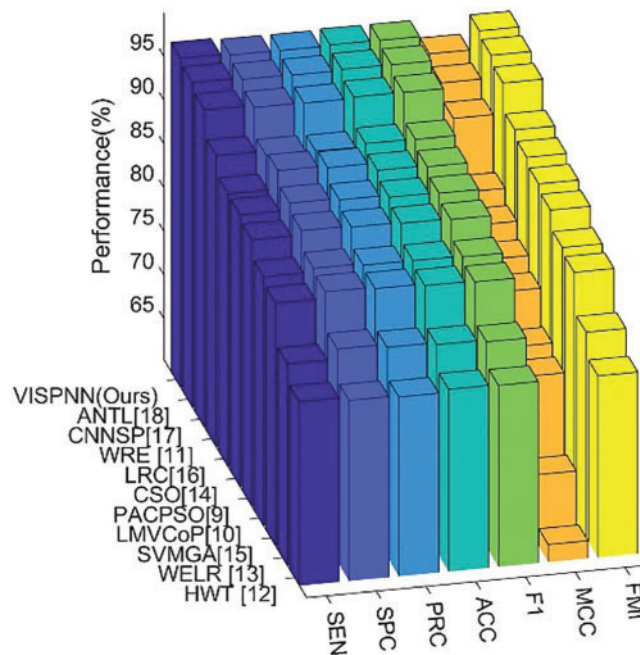
Model	SEN	SPC	PRC	ACC	F1	MCC	FMI
Net-I	$94.47 \pm 2.22$	$95.18 \pm 1.18$	$95.10 \pm 1.09$	$94.83 \pm 0.81$	$94.76 \pm 0.88$	$89.70 \pm 1.58$	$94.77 \pm 0.87$
Net-II	$93.51 \pm 1.82$	$92.72 \pm 1.89$	$92.68 \pm 1.84$	$93.11 \pm 1.66$	$93.09 \pm 1.66$	$86.24 \pm 3.31$	$93.09 \pm 1.66$
VISPNN	$97.98 \pm 1.32$	$97.80 \pm 1.35$	$97.78 \pm 1.35$	$97.89 \pm 1.11$	$97.87 \pm 1.12$	$95.79 \pm 2.22$	$97.88 \pm 1.12$

**Figure 11:** ROC comparison of ablation studies (a) Net-I, (b) Net-II and (c) VISPNN

We can observe from Fig. 12 that the proposed VISPNN model beats all the other ten state-of-the-art methods in terms of all seven measures. The reason is three folds. First, the VGG-inspired mainstay network gains many benefits by mimicking the similar structures from VGG-16. Second, stochastic pooling helps our model more robust than max pooling does. Third, the improved 20-way data augmentation generates diverse fake training images to help our model more resistant to overfitting.

**Table 7:** Comparison with state-of-the-art alcoholism recognition methods

Method	SEN	SPC	PRC	ACC	F1	MCC	FMI
PACPSO [9]	90.80 ± 1.98	91.31 ± 1.71	91.17 ± 1.52	91.06 ± 1.01	90.96 ± 1.06	82.14 ± 2.01	90.97 ± 1.05
LMVCoP [10]	88.83 ± 1.09	89.27 ± 1.42	89.08 ± 1.27	89.05 ± 0.85	88.95 ± 0.84	78.11 ± 1.71	88.95 ± 0.84
WRE [11]	93.62 ± 1.30	93.14 ± 1.32	93.08 ± 1.26	93.38 ± 1.05	93.34 ± 1.05	86.77 ± 2.10	93.35 ± 1.05
HWT [12]	81.33 ± 1.65	80.94 ± 1.53	80.79 ± 1.05	81.13 ± 0.70	81.05 ± 0.77	62.29 ± 1.38	81.05 ± 0.77
WELR [13]	83.40 ± 2.91	84.55 ± 1.87	84.21 ± 1.45	83.98 ± 1.15	83.76 ± 1.36	68.02 ± 2.29	83.78 ± 1.35
CSO [14]	91.38 ± 1.60	92.20 ± 1.17	92.03 ± 1.12	91.79 ± 1.06	91.70 ± 1.10	83.60 ± 2.11	91.70 ± 1.10
SVMGA [15]	88.09 ± 1.42	88.85 ± 2.67	88.67 ± 2.30	88.47 ± 1.20	88.35 ± 1.09	76.98 ± 2.39	88.37 ± 1.09
LRC [16]	91.54 ± 1.47	93.66 ± 1.34	93.45 ± 1.27	92.61 ± 0.81	92.48 ± 0.83	85.26 ± 1.62	92.47 ± 0.84
CNNSP [17]	96.60 ± 1.28	96.34 ± 1.18	96.29 ± 1.18	96.46 ± 1.11	96.44 ± 1.11	92.93 ± 2.22	96.44 ± 1.11
ANTL [18]	97.45 ± 1.68	97.28 ± 1.32	97.24 ± 1.33	97.36 ± 1.35	97.34 ± 1.36	94.73 ± 2.69	97.34 ± 1.36
VISPNN (Ours)	97.98 ± 1.32	97.80 ± 1.35	97.78 ± 1.35	97.89 ± 1.11	97.87 ± 1.12	95.79 ± 2.22	97.88 ± 1.12

**Figure 12:** Bar plot of ranked 11 state-of-the-art methods in order of MCC

#### 4 Conclusions

To identify alcoholism more efficiently, we propose the VISPNN model based on a VGG-inspired mainstay network, stochastic pooling technique, and an improved 20-way data augmentation. The results show that our model gains a sensitivity of  $97.98 \pm 1.32$ , a specificity of  $97.80 \pm 1.35$ , an accuracy of  $97.89 \pm 1.11$ , and an AUC of 0.9849, respectively. The performance is better than 10 state-of-the-art alcoholism recognition methods.

The limitations of this study are that this model does not go through strict clinician verification; also, the dataset is relatively small. Hence, we will try to collect more brain images of both alcoholism and healthy subjects. Meanwhile, we shall deploy our VISPNN model to the cloud server and invite clinicians and radiologists to use our web app and get their feedback to improve our model further.

**Funding Statement:** This paper is partially supported by the Royal Society International Exchanges Cost Share Award, UK (RP202G0230); Medical Research Council Confidence in Concept Award, UK (MC\_PC\_17171); Hope Foundation for Cancer Research, UK (RM60G0680); British Heart Foundation Accelerator Award, UK; Sino-UK Industrial Fund, UK (RP202G0289); Global Challenges Research Fund (GCRF), UK (P202PF11). In addition, we acknowledge the help of Dr. Hemil Patel and Dr. Qinghua Zhou for their help in English correction.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] O. I. A. Hamid, L. M. E. Sabik, K. F. Abdelfadeel and S. F. Shaban, "Tramadol aggravates cardiovascular toxicity in a rat model of alcoholism: Involvement of intermediate microfilament proteins and immune-expressed osteopontin," *Journal of Biochemical and Molecular Toxicology*, vol. 121, pp. 15, 2021.
- [2] D. A. Shirley, I. Sharma, C. A. Warren and S. Moonah, "Drug repurposing of the alcohol abuse medication disulfiram as an anti-parasitic agent," *Frontiers in Cellular and Infection Microbiology*, vol. 11, pp. 7, 2021.
- [3] M. I. Airapetov, S. O. Eresko, A. A. Lebedev, E. R. Bychkov and P. D. Shabanov, "The role of toll-like receptors in neuroimmunology of alcoholism," *Biochemistry Moscow-Supplement Series B-Biomedical Chemistry*, vol. 15, pp. 71–79, 2021.
- [4] J. L. Meyers, D. B. Chorlian, T. B. Bigdeli, E. C. Johnson, F. Aliev *et al.*, "The association of polygenic risk for schizophrenia, bipolar disorder, and depression with neural connectivity in adolescents and young adults: Examining developmental and sex differences," *Translational Psychiatry*, vol. 11, pp. 11, 2021.
- [5] N. T. Rachman, H. Tjandrasa and C. F. Ichah, "Alcoholism classification based on EEG data using independent component analysis (ICA), wavelet de-noising and probabilistic neural network (PNN)," in *Int. Seminar on Intelligent Technology and Its Applications*, Lombok, Indonesia, pp. 17–20, 2016.
- [6] T. Bordia and N. M. Zahr, "The inferior colliculus in alcoholism and beyond," *Frontiers in Systems Neuroscience*, vol. 14, pp. 19, 2020.
- [7] Q. Y. Zhao, M. Fritz, A. Pfefferbaum, E. V. Sullivan, K. M. Pohl *et al.*, "Jacobian maps reveal under-reported brain regions sensitive to extreme binge ethanol intoxication in the rat," *Frontiers in Neuroanatomy*, vol. 12, pp. 13, 2018.
- [8] C. E. Thompson, "The prodromal phase of alcoholism in Herman Melville's *Bartleby, the scrivener and cock-a-doodle-doo!*," *Explicator*, vol. 71, no. 4, pp. 275–280, 2013.
- [9] X.-X. Hou, "Alcoholism detection by medical robots based on Hu moment invariants and predator-prey adaptive-inertia chaotic particle swarm optimization," *Computers and Electrical Engineering*, vol. 63, no. 6, pp. 126–138, 2017.
- [10] A. Jenitta and R. S. Ravindran, "Image retrieval based on local mesh vector co-occurrence pattern for medical diagnosis from MRI brain images," *Journal of Medical Systems*, vol. 41, no. 10, Article ID: 157, 2017.
- [11] L. Han, "Identification of alcoholism based on wavelet Renyi entropy and three-segment encoded jaya algorithm," *Complexity*, vol. 2018, Article ID: 3198184, 2018.
- [12] D. Lima, "Alcoholism detection in magnetic resonance imaging by Haar wavelet transform and back propagation neural network," *AIP Conference Proceedings*, vol. 1955, Article ID: 040012, 2018.
- [13] F. Macdonald, "Alcoholism detection via wavelet energy and logistic regression," *Advances in Intelligent Systems Research*, vol. 148, pp. 164–168, 2018.
- [14] P. Qian, "Cat swarm optimization applied to alcohol use disorder identification," *Multimedia Tools and Applications*, vol. 77, no. 17, pp. 22875–22896, 2018.



- [15] Y. Chen, "Alcoholism detection by wavelet entropy and support vector machine trained by genetic algorithm," in *27th IEEE Int. Conf. on Robot and Human Interactive Communication (ROMAN)*, Nanjing, China, pp. 770–775, 2018.
- [16] X. Chen, "Alcoholism detection by wavelet energy entropy and linear regression classifier," *Computer Modeling in Engineering & Sciences*, vol. 127, no. 1, pp. 325–343, 2021.
- [17] Y.-D. Lv, "Alcoholism detection by data augmentation and convolutional neural network with stochastic pooling," *Journal of Medical Systems*, vol. 42, Article ID: 2, 2018.
- [18] S. Xie, "Alcoholism identification based on an alexnet transfer learning model," *Frontiers in Psychiatry*, vol. 10, Article ID: 205, 2019.
- [19] C. Kamarajan, B. A. Ardekani, A. K. Pandey, D. B. Chorlian, S. Kinreich *et al.*, "Random forest classification of alcohol use disorder using EEG source functional connectivity," *Behavioral Sciences*, vol. 10, pp. 32, 2020.
- [20] A. Quagliari, E. Mari, M. Boccia, L. Piccardi, C. Guariglia *et al.*, "Brain network underlying executive functions in gambling and alcohol use disorders: An activation likelihood estimation meta-analysis of fmri studies," *Brain Sciences*, vol. 10, pp. 19, 2020.
- [21] G. S. B. Jahangeer and T. D. Rajkumar, "Early detection of breast cancer using hybrid of series network and VGG-16," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 7853–7886, 2021.
- [22] W. Y. Zhang, P. M. Shi, M. D. Li and D. Y. Han, "A novel stochastic resonance model based on bistable stochastic pooling network and its application," *Chaos Solitons & Fractals*, vol. 145, pp. 10, 2021.
- [23] X. Cheng, "Psspnn: Patchshuffle stochastic pooling neural network for an explainable diagnosis of covid-19 with multiple-way data augmentation," *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID: 6633755, 2021.
- [24] H. Akbari, M. T. Sadiq and A. U. Rehman, "Classification of normal and depressed EEG signals based on centered correntropy of rhythms in empirical wavelet transform domain," *Health Information Science and Systems*, vol. 9, pp. 15, 2021.