Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 5792975, 17 pages
https://doi.org/10.1155/2021/5792975

Research Article

SOSPCNN: Structurally Optimized Stochastic Pooling Convolutional Neural Network for Tetralogy of Fallot Recognition

Shui-Hua Wang,[1] Kaihong Wu,[2] Tianshu Chu,[3] Steven L. Fernandes,[4] Qinghua Zhou,[1] Yu-Dong Zhang,[1,3] and Jian Sun[2]

[1]*School of Informatics, University of Leicester, Leicester LE1 7RH, UK*
[2]*The Affiliated Children's Hospital of Nanjing Medical University, Nanjing, China*
[3]*Nanjing Yirongda Institute of Intelligent Medicine and Additive Manufacturing, Nanjing, China*
[4]*Department of Computer Science, Design & Journalism, Creighton University, Omaha, Nebraska, USA*

Correspondence should be addressed to Yu-Dong Zhang; yudong.zhang@le.ac.uk and Jian Sun; sunjian67@njmu.edu.cn

Received 6 May 2021; Revised 27 May 2021; Accepted 5 June 2021; Published 2 July 2021

Academic Editor: Shan Zhong

Aim. This study proposes a new artificial intelligence model based on cardiovascular computed tomography for more efficient and precise recognition of Tetralogy of Fallot (TOF). *Methods*. Our model is a structurally optimized stochastic pooling convolutional neural network (SOSPCNN), which combines stochastic pooling, structural optimization, and convolutional neural network. In addition, multiple-way data augmentation is used to overcome overfitting. Grad-CAM is employed to provide explainability to the proposed SOSPCNN model. Meanwhile, both desktop and web apps are developed based on this SOSPCNN model. *Results*. The results on ten runs of 10-fold crossvalidation show that our SOSPCNN model yields a sensitivity of 92.25 ± 2.19, a specificity of 92.75 ± 2.49, a precision of 92.79 ± 2.29, an accuracy of 92.50 ± 1.18, an F1 score of 92.48 ± 1.17, an MCC of 85.06 ± 2.38, an FMI of 92.50 ± 1.17, and an AUC of 0.9587. *Conclusion*. The SOSPCNN method performed better than three state-of-the-art TOF recognition approaches.

1. Introduction

Tetralogy of Fallot (TOF) is a congenital defect that influences normal blood flow through the heart [1]. It is made up of 4 defects of the heart and its blood vessels [2]: (a) ventricular septal defect, (b) overriding aorta, (c) right ventricular outflow tract stenosis, and (d) right ventricular hypertrophy. Defects of TOF can cause oxygen in the blood that flows to the rest of the body to be reduced. Infants with TOF have a bluish-looking skin color [3] since their blood does not carry enough oxygen.

Traditional diagnosis of TOF is after a baby is born, often after the infant had an episode of cyanosis during crying or feeding. The most common test is an echocardiogram [4], an ultrasound of the heart that can show problems with the heart structure and how well the heart is working with this defect. Recently, computed tomography (CT) has shown its success in the differential diagnosis of TOF [5], since it can provide detailed images of many types of cardiovascular issue; besides, computed tomography (CT) can be performed even if the subject has an implanted medical device, unlike magnetic resonance imaging (MRI) [6].

Manual diagnosis on CT is lab-intensive, onerous, and needs expert skills. Besides, the manual results vary due to intraexpert and interexpert factors. Shan et al. (2021) [7] mention that "fully manual delineation that often takes hours" and the modern automatic diagnosis models based on artificial intelligence (AI) can only take seconds to minutes to get decisions, which now becomes a hot research field.

For example, Ye et al. (2011) [8] present a morphological classification (MC) method. The authors extract morphological features by registering cardiac MRI scans to a template. Later, deep learning (DL) rises as a new type of artificial intelligence (AI) technique and has shown its powerfulness in many academic and industrial fields. Within the field of

DL, convolutional neural network (CNN) is one standard DL algorithm that is particularly suitable for handling images. Giannakidis et al. (2016) [9] presented a multiscale three-dimensional CNN (3DCNN) for segmentation of the right ventricle. Tandon et al. (2021) [10] present a ventricular contouring CNN (VCCNN) algorithm.

The difference between this study to previous studies is that we simplify the problem to a binary-coded classification problem [11]; that is, given an input cardiovascular CT image, the AI model should have the ability to give a binary output, i.e., predict whether the subject is TOF or healthy. This simplification makes the AI model focus on the prediction task itself and does not need to generate human-understandable outputs (such as segmentation, contouring, etc.) in the light of the expectation to make our AI model more accurate. Furthermore, we propose a new stochastic pooling CNN (SCCNN) that uses a new pooling technique—stochastic pooling to improve the prediction performance. All in all, our contributes are fourfold:

(a) Stochastic pooling is employed to replace traditional max-pooling

(b) Structural optimization is carried out to fix the optimal structure

(c) Multiple-way DA is introduced to increase the diversity of training images

(d) Experiments by ten runs of 10-fold crossvalidation show that our method is better than three state-of-the-art approaches

The rest of this paper is structured as follows: Section 2 describes the dataset. Section 3 contains the rationale of methodology, including the preprocessing, stochastic pooling, structural optimization, multiple-way data augmentation, the implementation, Grad-CAM, and evaluation measures. Section 4 presents the experimental results and discussions. Section 5 concludes this paper.

Table 1: Abbreviation list.

Abbreviation	Meaning
AP	Average pooling
AUC	Area under the curve
BN	Normalization
CNN	Convolutional neural network
CT	Computed tomography
DA	Data augmentation
DPD	Discrete probability distribution
FCL	Fully connected layer
FMI	Fowlkes–Mallows index
Grad-CAM	Gradient-weighted class activation mapping
GUI	Graphical user interface
HS	Histogram stretching
L2P	l_2-norm pooling
MCC	Matthews correlation coefficient
MP	Max-pooling
MRI	Magnetic resonance imaging
MSD	Mean and standard deviation
PM	Probability map
ReLU	Rectified linear unit
RLV	Random location vector
ROC	Receiver operating characteristic
SC	Strided convolution
SP	Stochastic pooling
TOF	Tetralogy of Fallot

2. Dataset

This study is a retrospective research, of which ethical approval is exempted. The imaging protocol is described below: Philips Brilliance 256 row spiral CT machine, KV: 80, MAS: 138, Layer Thickness 0.8 mm, Lung Window (W: 1600 HU, L: -600 HU), Mediastinal Window (W: 750 HU, L: 90 HU), thin layer reconstruction according to the lesion display, layer thickness, and layer distance are both 0.8 mm mediastinal window images. Place the patient in a supine position, let the patient breathe deeply after holding in, and conventionally scan from the apex of the lung to the costal diaphragmatic angle. The resolutions of all images are 512 by 512 pixels. Data is available upon reasonable requests to corresponding authors.

We selected ten children with Tetralogy of Fallot who were admitted to Nanjing Children's Hospital from March 2017 to March 2020. We then used a systematic random sampling method to select ten normal children from healthy medical examiners within the same period of time. The Tetralogy of Fallot (TOF) observation group included three males and seven females, aged 4-22 months, with an average age of (8.90 ± 5.47) months. Normal children in the control group included six males and four females, aged from 3 months to 24 months, with an average age of 10.4 ± 8.14 months. Inclusion criteria for children with confirmed Tetralogy of Fallot are as follows:

(1) CT suggests Tetralogy of Fallot

(2) Surgery confirmed that the anatomical deformity of the heart is Tetralogy of Fallot

3. Methodology

3.1. Preprocessing. Table 1 lists the abbreviation list for the ease of reading. A five-step preprocessing was carried out on all the images to select the important slices, save storage, enhance contrast, remove unnecessary image regions, and reduce the image resolution.

First, four slices were chosen by radiologists using a slice-level selection method. For TOF patients, the slices showing the largest size and number of lesions were selected. For healthy control subjects, any level of the image can be selected. Now, we have in total 40 TOF images and 40 HC images.

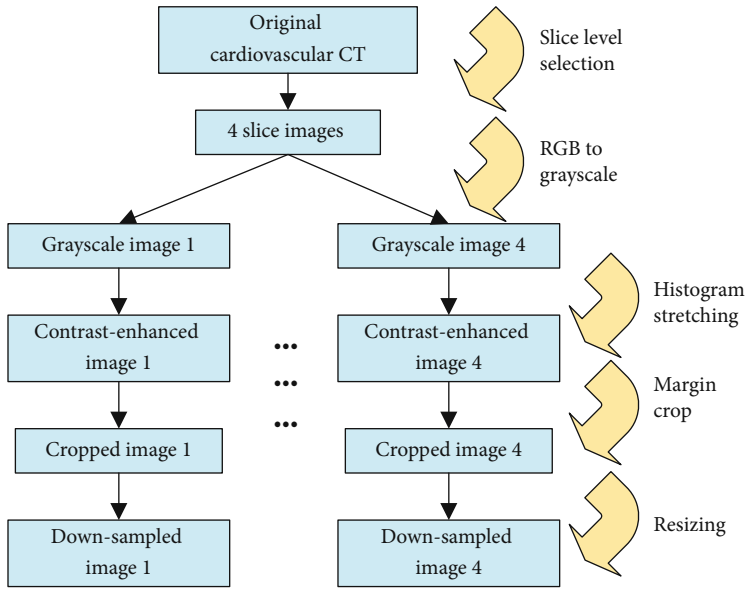
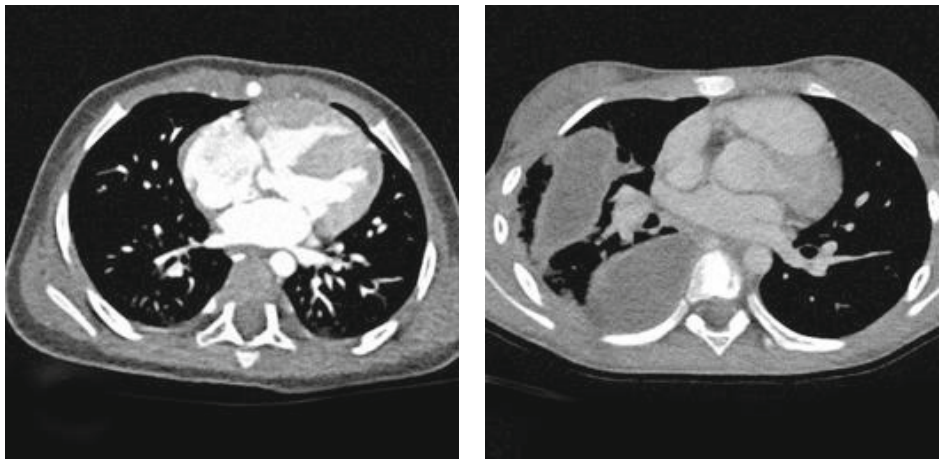


Figure 1: Diagram of preprocessing.



(a) TOF (b) HC

Figure 2: Illustration of our dataset.

Second, all the images are converted to grayscale images and stored in tiff format [12] using the compression lossless method. Third, histogram stretching (HS) was employed to enhance image contrast. Suppose the kth input and output of HS is $x(k)$ and $y(k)$. HS can be formulated as

$$y(k) = \frac{x(k) - x_{\min}(k)}{x_{\max}(k) - x_{\min}(k)}, \quad (1)$$

where $x_{\min}(k)$ and $x_{\max}(k)$ stand for the minimum and maximum grayscale values in the input $x(k)$.

Fourth, cropping was done in order to eliminate the check-up bed at the bottom, the subject's two arms at bilateral sides, the rulers at the bottom and right side, and information (hospital, scanning protocol, subject's information, image head information, and labeling) at four corners.

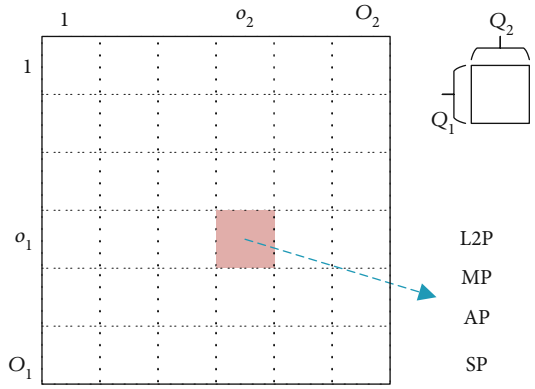


Figure 3: A diagram of block-wise pooling.

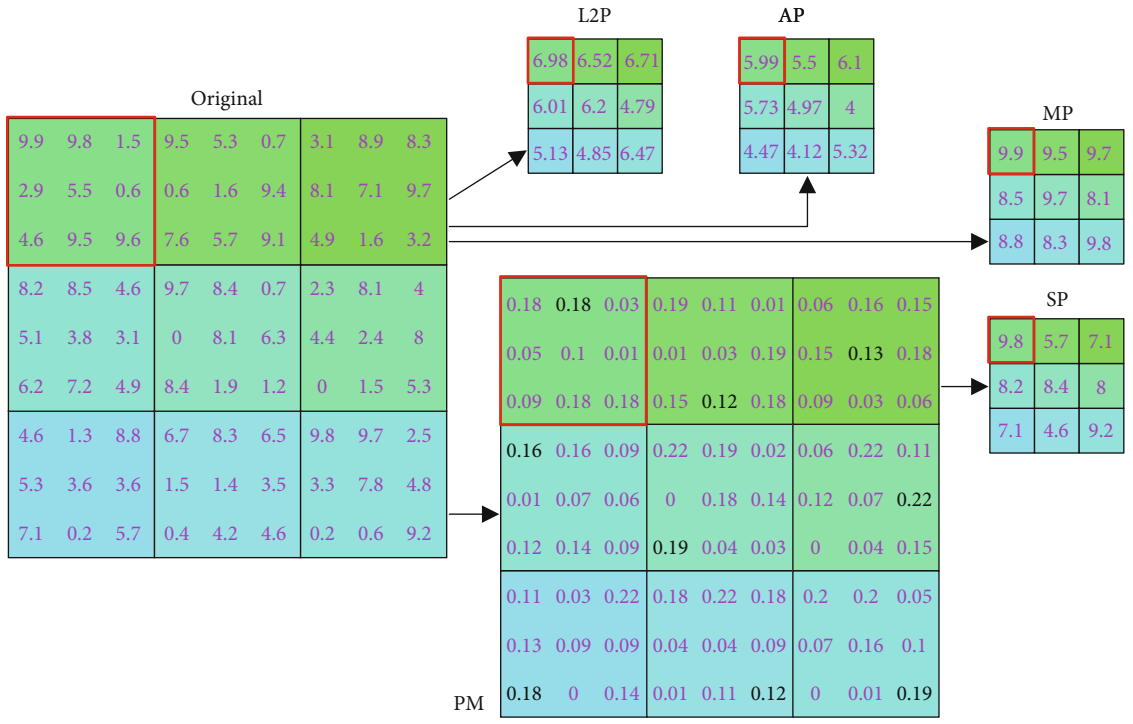


Figure 4: Comparison of four different pooling methods.

Input: block B_{o_1,o_2}.
Step 1: produce the PM for each pixel V_{PM}. See Equation (7).
Step 2: create a RLV $\vec{r} = (x_r, y_r)$. See Equation (8).
Step 3: draw a sample location vector $\vec{r}_0$ from the RLV $\vec{r}$. See Equation (9).
Step 4: output pixel at location $\vec{r}_0$. See Equation (10)
Output matrix $B^{SP}_{o_1,o_2}$.

Algorithm 1: Pseudocode of SP.

Lastly, downscaling was performed to reduce each image to the size of $[256 \times 256]$. Figure 1 displays the diagram of our preprocessing procedure. Figures 2(a) and 2(b) shows two preprocessed examples of TOF and HC, respectively.

3.2. Stochastic Pooling. Pooling is an essential operation in standard convolutional neural networks (CNNs) [13]. Two types of pooling exist. One is max pooling (MP), and the other is average pooling (AP). The objective of pooling is to down-sample an input image or feature map (FM), reducing their dimensionality (width or height) and allowing for some assumption about the features to be made in each block.

Suppose we have an input image or FM, which can be split into $O_1 \times O_2$ blocks, where every block has the extent of $Q_1 \times Q_2$. Currently, let us fix on the block B_{o_1,o_2} at o_1th row and o_2th column as shown as the red rectangle in Figure 3.

$$B_{o_1,o_2} = \{b(x,y), x = 1,\cdots,Q_1, y = 1,\cdots,Q_2\}, \quad (2)$$

Table 2: Structures of nine customized neural networks.

Configuration	No. of Conv layers	No. of FCLs
I	2	1
II	2	2
III	2	3
IV	3	1
V	**3**	**2**
VI	3	3
VII	4	1
VIII	4	2
IX	4	3

Bold means the best.

where $1 \le o_1 \le O_1, 1 \le o_2 \le O_2$, $b(x,y)$ means the pixel value at coordinate (x,y).

The strided convolution (SC) goes over the input activation map with the strides that equals the size of the block (Q_1, Q_2). The output of SC is

Table 3: Detailed structure of network of configuration V.

Layer	Parameters	FM
Input		$256 \times 256 \times 1$
Conv_1 (BN-ReLU)	32, $3 \times 3/2$	$128 \times 128 \times 32$
SP_1		$64 \times 64 \times 32$
Conv_2 (BN-ReLU)	64, $3 \times 3/2$	$32 \times 32 \times 64$
SP_2		$16 \times 16 \times 64$
Conv_3 (BN-ReLU)	128, 3×3	$16 \times 16 \times 128$
SP_3		$8 \times 8 \times 128$
Flatten		8192
FCL_1	100×8192, 100×1	100
FCL_2	2×100,2×1	2
Output		

$$B_{o_1,o_2}^{SC} = b(1,1). \tag{3}$$

The l_2-norm pooling (L2P), average pooling (AP) [14], and max pooling (MP) [15] produce the l_2-norm, average, and maximum values within the block B_{m_1,m_2}, respectively. Their formula can be written as below:

$$B_{o_1,o_2}^{L2P} = \sqrt{\frac{\sum_{x=1}^{Q_1}\sum_{y=1}^{Q_2} b^2(x,y)}{Q_1 \times Q_2}}, \tag{4}$$

$$B_{o_1,o_2}^{AP} = \frac{1}{Q_1 \times Q_2}\sum_{x=1}^{Q_1}\sum_{y=1}^{Q_2} b(x,y), \tag{5}$$

$$B_{o_1,o_2}^{MP} = \max_{x=1}^{Q_1} \max_{y=1}^{Q_2} b(x,y). \tag{6}$$

Nevertheless, the AP outputs the average, downscaling the greatest value, where the important features may lie. In contrast, MP stores the greatest value but deteriorates the overfitting obstacle. In order to solve the above concerns, stochastic pooling (SP) [15] is introduced to provide a resolution to the drawbacks of AP and MP. SP is a four-step process.

Step 1. It produces the probability map (PM) V_{PM} for each pixel in the block B_{o_1,o_2}.

$$\begin{cases} V_{\mathrm{PM}}(x,y) = \dfrac{b(x,y)}{\sum_{x=1}^{Q_1}\sum_{y=1}^{Q_2} b(x,y)}, \\ \text{s.t.} \displaystyle\sum_{x=1}^{Q_1}\sum_{y=1}^{Q_2} V_{\mathrm{PM}}(x,y) = 1, \end{cases} \tag{7}$$

where $V_{\mathrm{PM}}(x,y)$ stands for the PM value at pixel (x,y).

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

$$\begin{cases} P\left[\vec{r} = (1,1)\right] = V_{\mathrm{PM}}(1,1) & P\left[\vec{r} = (1,2)\right] = V_{\mathrm{PM}}(1,2) & \cdots & P\left[\vec{r} = (1,Q_2)\right] = V_{\mathrm{PM}}(1,Q_2), \\ P\left[\vec{r} = (2,1)\right] = V_{\mathrm{PM}}(2,1) & \Pr\left[\vec{r} = (2,2)\right] = V_{\mathrm{PM}}(2,2) & \cdots & P\left[\vec{r} = (2,Q_2)\right] = V_{\mathrm{PM}}(2,Q_2), \\ \cdots & \cdots & \cdots & \cdots \\ P\left[\vec{r} = (Q_1,1)\right] = V_{\mathrm{PM}}(Q_1,1) & P\left[\vec{r} = (Q_1,2)\right] = V_{\mathrm{PM}}(Q_1,2) & \cdots & P\left[\vec{r} = (Q_1,Q_2)\right] = V_{\mathrm{PM}}(Q_1,Q_2), \end{cases} \tag{8}$$

where P represents the probability. Shortly speaking, $P[\vec{r} = (x,y)] = V_{\mathrm{PM}}(x,y)$, $\forall 1 \le x \le Q_1 \& 1 \le y \le Q_2$ or $\vec{r} \sim V_{PM}$, namely, the distribution of RLV $\vec{r}$ has the DPD as V_{PM}.

Step 3. A sample location vector $\vec{r}_0$ is drawn from the RLV $\vec{r}$, and we have

$$\vec{r}_0 = \left(x_{r_0}, y_{r_0}\right). \tag{9}$$

Step 4. SP outputs the pixel at the location $\vec{r_0}$, namely,

$$B_{o_1,o_2}^{SP} = b\left(x_{r_0}, y_{r_0}\right). \tag{10}$$

Figure 4 shows a realistic example of four different pooling methods. Algorithm 1 presents the pseudocode of SP. Take the 3×3 block $B_{1,1}$ (The red rectangle in Figure 4) as an example, L2P generates the output as 6.98. AP and MP present 5.99 and 9.9, respectively. Meanwhile, SP first generates PM matrix

$$V_{\mathrm{PM}} = \begin{bmatrix} 0.18 & 0.18 & 0.03 \\ 0.05 & 0.1 & 0.01 \\ 0.09 & 0.18 & 0.18 \end{bmatrix}, \tag{11}$$

and a sample location vector is drawn as $\vec{r}_0 = (1,2)$. Therefore, the output of SP is $B_{1,1}^{SP} = b(1,2) = 9.8$.

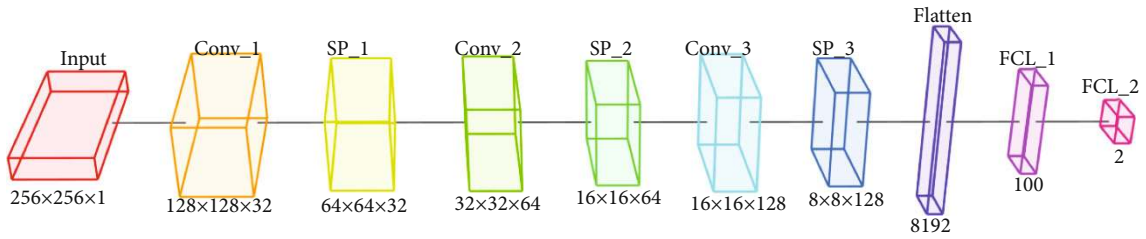


Figure 5: FM plot.

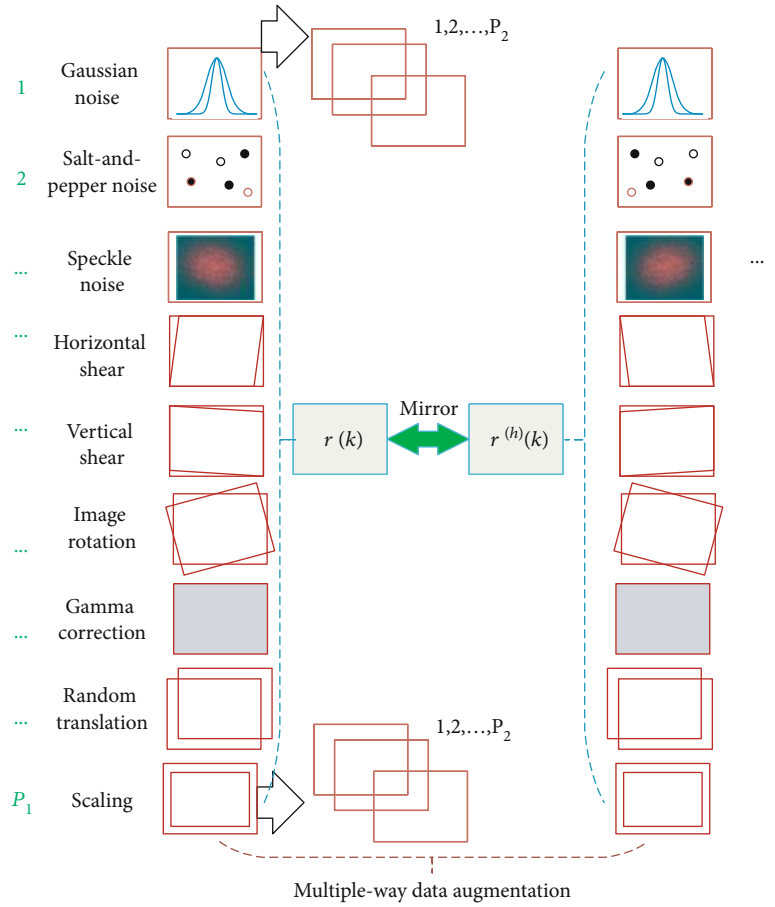


Figure 6: Diagram of multiple-way data augmentation.

Input: import raw preprocessed kth training image $r(k)$.
P_1 geometric or photometric or noise-injection DA transforms $\mathbf{Z}_p$ are utilized on $r(k)$.
Step 1: we obtain $\mathbf{Z}_p[r(k)], p = 1, \cdots, P_1$. See Equation (12)
Each enhanced dataset contains P_2 new images. See Equation (13).
Step 2: a horizontal mirror image is produced as $r^{(h)}(k) = \beta_1[r(k)]$. See Equation (14).
Step 3: M_1-way data augmentation methods are carried out on $r^{(h)}(k)$,
We obtain $\mathbf{Z}_p[r^{(h)}(k)], p = 1, \cdots, P_1$. See Equation (15).
Step 4: $r(k)$, $r^{(h)}(k)$, $\mathbf{Z}_p[r(k)], p = 1, \cdots, P_1$, and $\mathbf{Z}_p[r^{(h)}(k)], p = 1, \cdots, P_1$ are merged via β_2. See Equation (16).
Output A new dataset $\mathbf{G}(k)$ is produced. The number of images is $P_3 = 2 \times P_1 \times P_2 + 2$. See Equation (17).

Algorithm 2: Pseudocode of proposed 18-way DA on kth training image

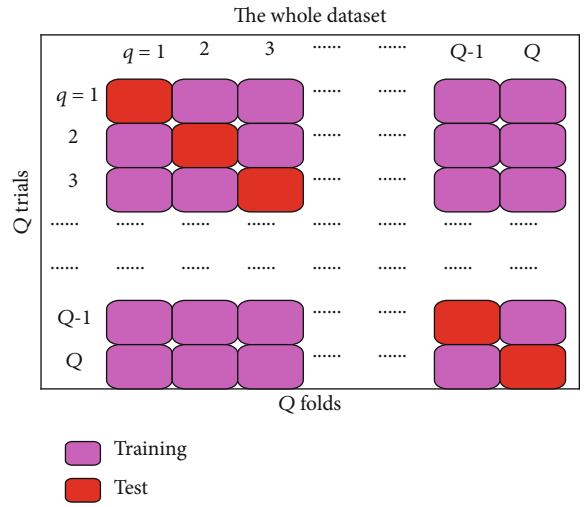


Figure 7: Q-fold crossvalidation.

Table 4: Meanings in measures.

Abbreviation	Full form	Symbol	Meaning
P	Positive		TOF
N	Negative		HC
TP	True positive	$g(1,1)$	TOF images are classified correctly.
FP	False positive	$g(2,1)$	HC images are wrongly classified as TOF.
TN	True negative	$g(2,2)$	HC images are classified correctly.
FN	False negative	$g(1,2)$	TOF images are wrongly classified as HC.

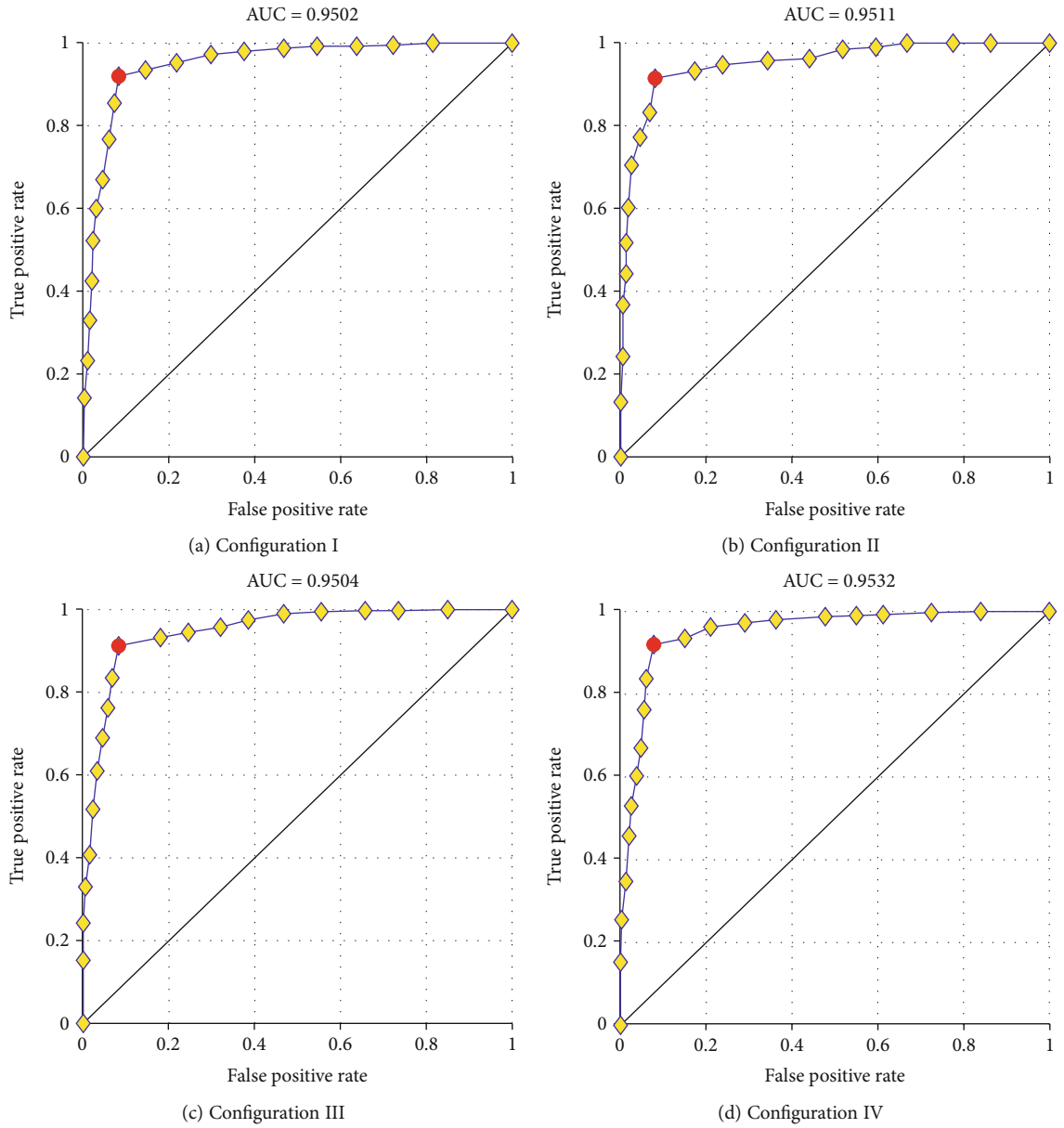
Table 5: Statistical analysis of SOSPCNN model.

Run	Sen	Spc	Prc	Acc	F1	MCC	FMI
1	95.00	92.50	92.68	93.75	93.83	87.53	93.83
2	92.50	90.00	90.24	91.25	91.36	82.53	91.36
3	95.00	92.50	92.68	93.75	93.83	87.53	93.83
4	90.00	92.50	92.31	91.25	91.14	82.53	91.15
5	90.00	95.00	94.74	92.50	92.31	85.11	92.34
6	90.00	97.50	97.30	93.75	93.51	87.75	93.58
7	92.50	95.00	94.87	93.75	93.67	87.53	93.68
8	92.50	90.00	90.24	91.25	91.36	82.53	91.36
9	90.00	92.50	92.31	91.25	91.14	82.53	91.15
10	95.00	90.00	90.48	92.50	92.68	85.11	92.71
MSD	92.25 ± 2.19	92.75 ± 2.49	92.79 ± 2.29	92.50 ± 1.18	92.48 ± 1.17	85.06 ± 2.38	92.50 ± 1.17

3.3. Structural Optimization. How to obtain the best network structure [16]? We try to design nine different configurations in this study. Their hyperparameters of structures are listed in Table 2. Two hyperparameters are considered in this study: (i) the number of Conv layers and (ii) the number of fully connected layers (FCLs). Those two types of layers are common layers in standard CNN, so we will not introduce them due to the page limit.

In the following experiment, we will observe that configuration V, a five-layer customized neural network, gives the

Figure 8: Confusion matrix of 10×10-fold crossvalidation (Here, classes 1 and 2 stand for ToF and HC, respectively).



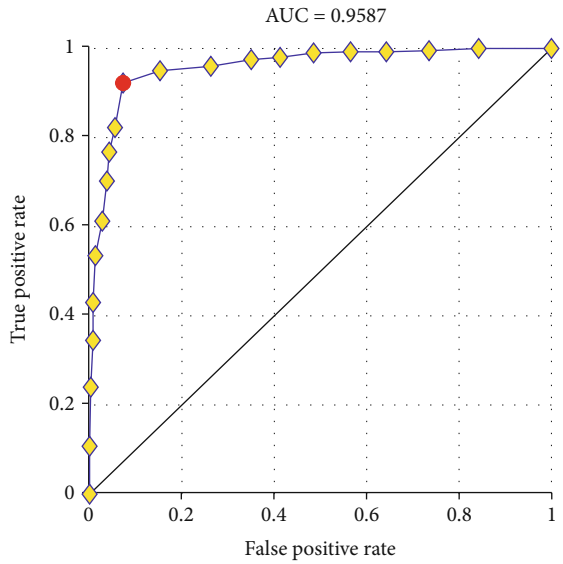
(a) Configuration I

(b) Configuration II

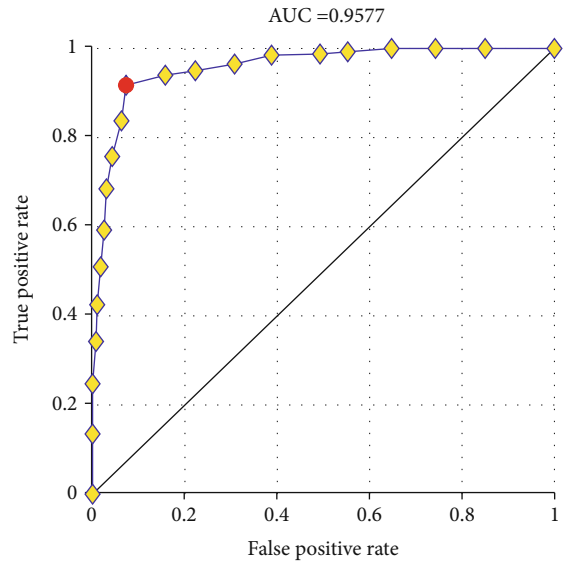
(c) Configuration III

(d) Configuration IV

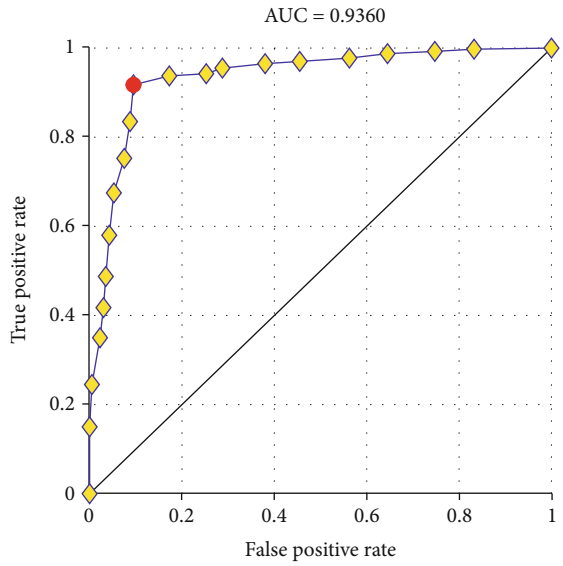
Figure 9: Continued.



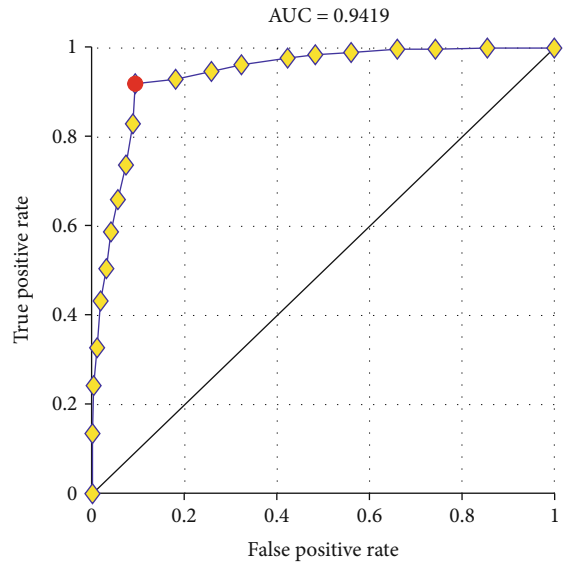
(e) Configuration V



(f) Configuration VI

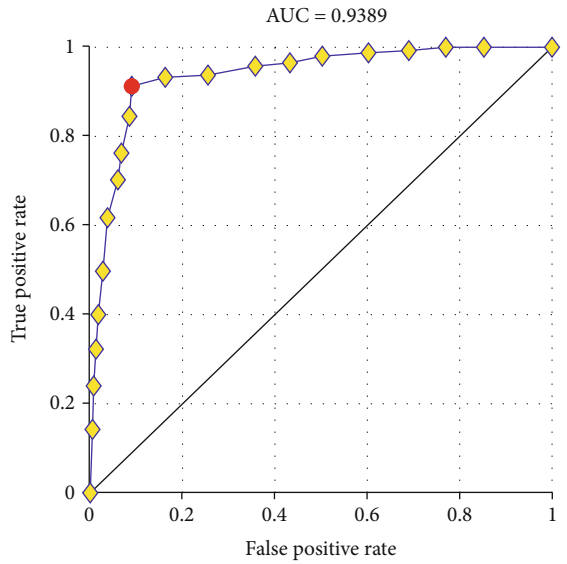


(g) Configuration VII



(h) Configuration VIII

Figure 9: Continued.



(i) Configuration IX

Figure 9: Comparison of nine configurations.

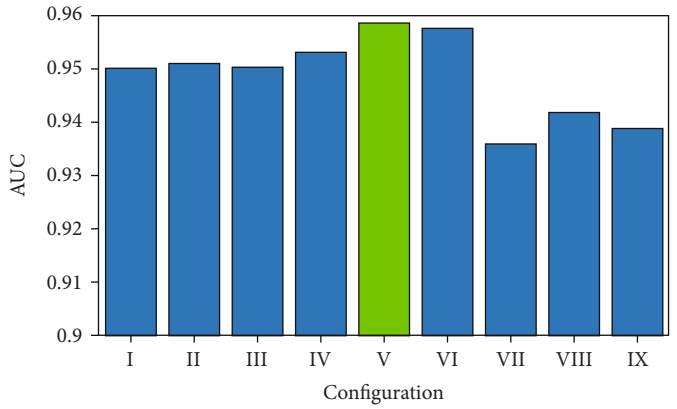


Figure 10: Bar plot of AUC against nine configurations.

best performance. Here, we briefly give its detailed structure in Table 3. The input is of size $256 \times 256 \times 1$. The first Conv layer (Conv_1) is associated with the batch normalization (BN) layer and rectified linear unit (ReLU) activation. The parameters of Conv_1 are 32 kernels with sizes of 3×3 and stride of 2. Afterward, the first SP (SP_1) reduce the FM from $128 \times 128 \times 32$ to $64 \times 64 \times 32$.

After three Conv layers and three SP layers, the size of FM is $8 \times 8 \times 128$. It is then flattened to a vector of 8192 neurons. With two FCLs of 100 and 2 hidden neurons, the neural network finally outputs whether TOF or HC. All in all, our model is termed structurally optimized stochastic pooling convolutional neural network (SOSPCNN). The FM plot is portrayed in Figure 5.

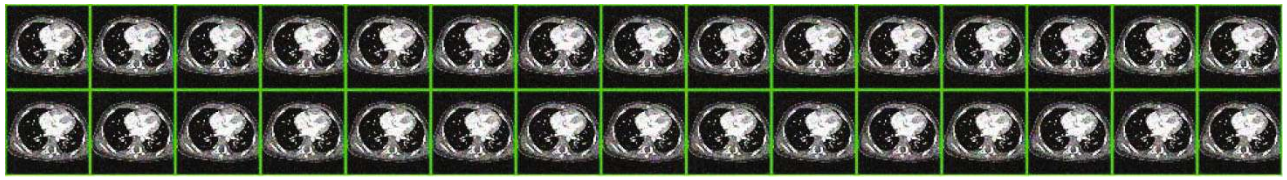
3.4. Multiple-Way Data Augmentation. The relatively small dataset (40+40 = 80 images) may bring the overfitting problem. To avoid overfitting, data augmentation (DA) [17] is a powerful tool because it can generate synthetic images on the training set [18]. Zhu (2021) [19] presented an 18-way DA method and proved this 18-way DA works better than the traditional DA approach. Its diagram is shown in Figure 6. The difference of DA and MDA is that (i) MDA uses a combination of different DA methods on training set; (ii) MDA is modular design. The users are easy to add or remove particular DA methods from a MDA.

Suppose we have the raw training image $r(k)$, where k represents the image index. First, P_1 different DA methods displayed in Figure 6 are applied to $r(k)$. Let $\boldsymbol{Z}_p, p = 1, \cdots, P_1$ be each DA operation, we get P_1 augmented datasets on raw image $r(k)$ as

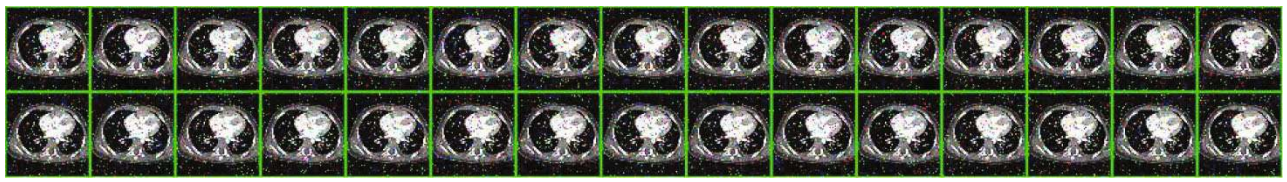
$$\boldsymbol{Z}_p[r(k)], p = 1, \cdots, P_1. \tag{12}$$

Let P_2 stands for the size of generated new images for each DA method, thus,

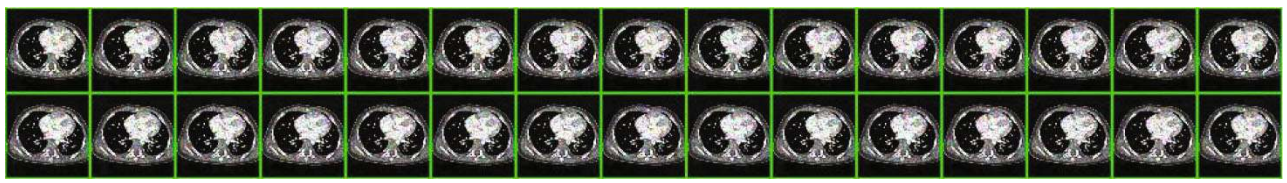
$$\left|\boldsymbol{Z}_p[r(k)]\right| = P_2. \tag{13}$$



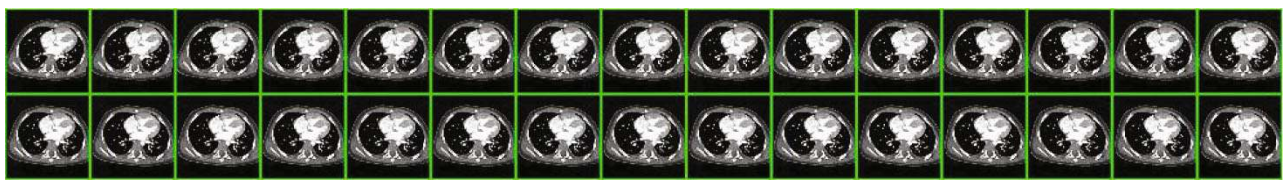
(a) Gaussian noise



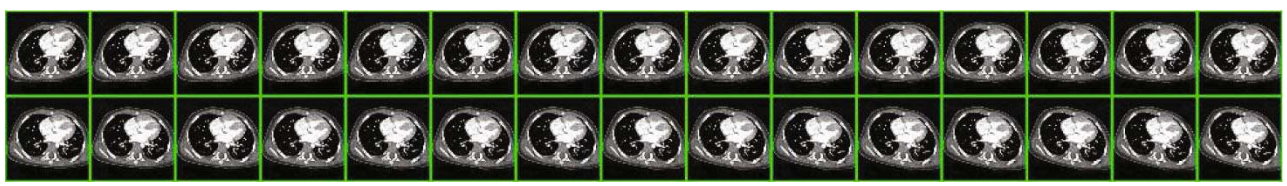
(b) Salt-and-pepper noise



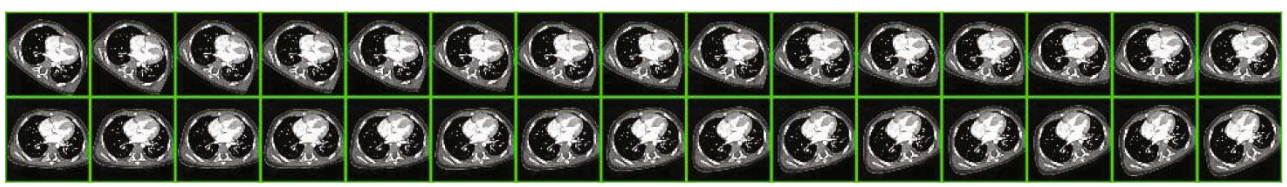
(c) Speckle noise



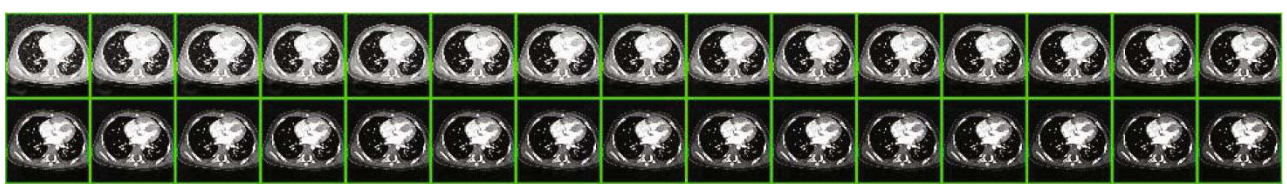
(d) Horizontal shear



(e) Vertical shear

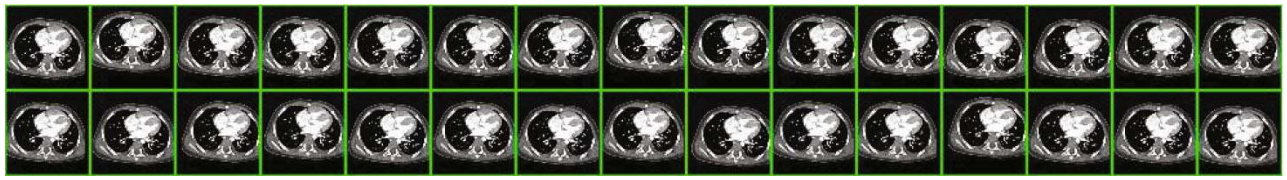


(f) Rotation

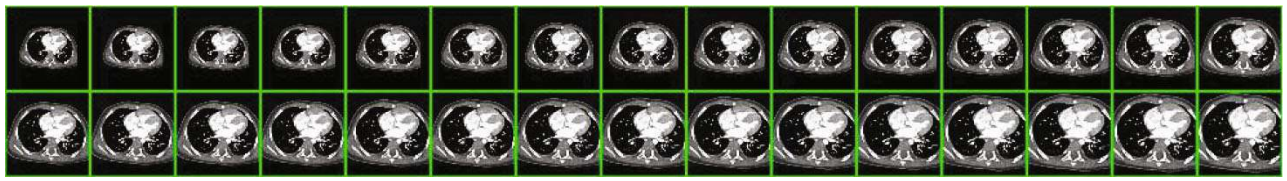


(g) Gamma correction

Figure 11: Continued.



(h) Random translation



(i) Scaling

Figure 11: Illustration of multiple-way data augmentation.

Second, horizontal mirrored image is produced by

$$r^{(h)}(k) = \beta_1[r(k)], \tag{14}$$

where β_1 means horizontal mirror function.

Third, all P_1 different DA methods are carried out on the mirrored image $r^{(h)}(k)$ and produce P_1 new datasets as

$$\begin{cases} \mathbf{Z}_p\left[r^{(h)}(k)\right], & p = 1, \cdots, P_1, \\ \left|\mathbf{Z}_p\left[r^{(h)}(k)\right]\right| = P_2, & p = 1, \cdots, P_1. \end{cases} \tag{15}$$

Fourth, the raw image $r(k)$, the mirrored image $r^{(h)}(k)$, all P_1-way results of raw image $\mathbf{Z}_p[r(k)]$, and all P_1-way DA results of horizontal mirrored image $\mathbf{Z}_p[r^{(h)}(k)]$ are combined. The final generated dataset from $r(k)$ is defined as $\mathbf{G}(k)$:

$$r(k) \mapsto \mathbf{G}(k) = \beta_2 \left\{ \begin{array}{cc} r(k) & r^{(h)}(k) \\ \underbrace{\mathbf{Z}_1[r(k)]}_{P_2} & \underbrace{\mathbf{Z}_1\left[r^{(h)}(k)\right]}_{P_2} \\ \cdots & \cdots \\ \underbrace{\mathbf{Z}_{P_1}[r(k)]}_{P_2} & \underbrace{\mathbf{Z}_{P_1}\left[r^{(h)}(k)\right]}_{P_2} \end{array} \right\}, \tag{16}$$

where β_2 stands for the concatenation function. Let augmentation factor be P_3, which means the number of images in $\mathbf{G}(k)$, we obtain

$$P_3 = \frac{|\mathbf{G}(k)|}{|r(k)|} = \frac{(1 + P_1 \times P_2) \times 2}{1} = 2 \times P_1 \times P_2 + 2. \tag{17}$$

Algorithm 2 recaps the pseudocode of this 18-way DA. We set $P_1 = 9$ to achieve an 18-way DA. We also set $P_2 = 30$, thus $P_3 = 542$, indicating each raw training image will generate 542 images, which include the raw image $r(k)$ itself.

3.5. Implementation and Grad-CAM. Q-fold crossvalidation [20] is employed. The whole dataset is divided into Q folds (see Figure 7). At qth trial, $1 \le q \le Q$, the qth fold is picked up as the test, and the rest $Q - 1$ folds: $[1, \cdots, q-1, q+1, \cdots, Q]$ are chosen as training set [21]. In this study, we set $Q = 10$, namely, a 10-fold cross validation. Furthermore, we run the 10-fold crossvalidation 10 times, i.e., 10×10-fold crossvalidation.

Gradient-weighted class activation mapping (Grad-CAM) [22] is employed to explain how our model makes its decision in classification. Grad-CAM utilizes the gradient of the classification score with respect to the convolutional features determined by the network to understand which parts of the image are most important for classification. Grad-CAM is a generalization of the class activation mapping (CAM) method [23] to a broader range of CNN models since the original CAM relies on a fully convolutional neural network structure. The output of SP_3 (see Table 3) is used as the feature layer for Grad-CAM.

Mathematically, suppose our classification network is with output y^c, standing for the score for class c. We would like to compute the Grad-CAM map for a layer with k feature maps $A^k_{i,j}$, where (i, j) stands for the indexes of pixels. We can obtain the neural importance weight as

$$\alpha^c_k = \frac{1}{N} \sum_i \sum_j \frac{\partial y^c}{\partial A^k_{i,j}}, \tag{18}$$

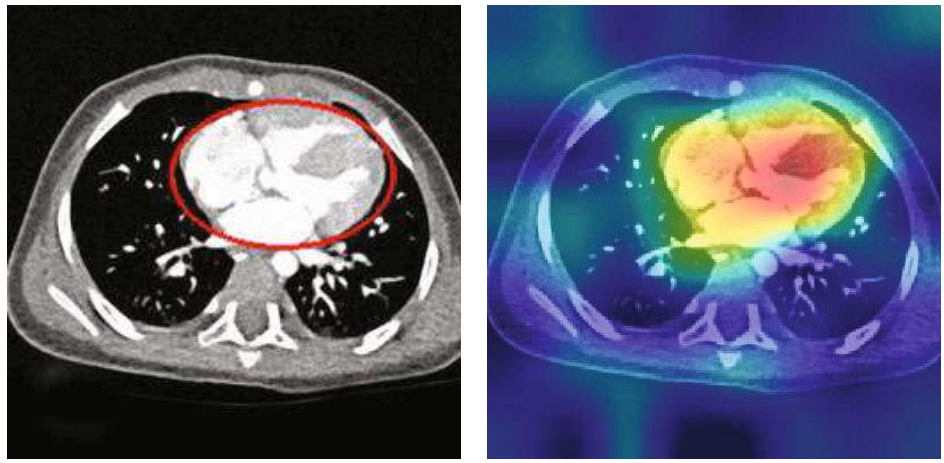
where N stands for the total number of pixels in the feature map. The Grad-CAM is a weighted combination of the feature maps with a ReLU as

$$M = ReLU\left(\sum_k \alpha^c_k A^k\right). \tag{19}$$

The Grad-CAM map M is then upsampled to the size of input data.

Table 6: Statistical analysis without multiple-way data augmentation.

Run	Sen	Spc	Prc	Acc	F1	MCC	FMI
1	87.50	90.00	89.74	88.75	88.61	77.52	88.61
2	90.00	87.50	87.80	88.75	88.89	77.52	88.90
3	87.50	87.50	87.50	87.50	87.50	75.00	87.50
4	90.00	87.50	87.80	88.75	88.89	77.52	88.90
5	85.00	90.00	89.47	87.50	87.18	75.09	87.21
6	85.00	90.00	89.47	87.50	87.18	75.09	87.21
7	82.50	92.50	91.67	87.50	86.84	75.38	86.96
8	87.50	87.50	87.50	87.50	87.50	75.00	87.50
9	85.00	90.00	89.47	87.50	87.18	75.09	87.21
10	82.50	92.50	91.67	87.50	86.84	75.38	86.96
MSD	86.25 ± 2.70	89.50 ± 1.97	89.21 ± 1.58	87.87 ± 0.60	87.66 ± 0.82	75.86 ± 1.16	87.70 ± 0.79



(a) Manually delineated (b) Heatmap

Figure 12: Heatmap of one TOF image.

3.6. Measures. The confusion matrix of 10 runs of 10-fold crossvalidation is supposed to be

$$G = \begin{bmatrix} g(1,1) & g(1,2) \\ g(2,1) & g(2,2) \end{bmatrix} = \begin{bmatrix} \mathrm{TP} & \mathrm{FN} \\ \mathrm{FP} & \mathrm{TN} \end{bmatrix}. \tag{20}$$

Note FN = FP = 0 for a perfect classification. The meaning of P, N, TP, FP, TN, and FN are itemized in Table 4.

Nine measures are used: sensitivity, specificity, precision, accuracy, F1 score, Matthews correlation coefficient (MCC), Fowlkes–Mallows index (FMI), receiver operating characteristic (ROC), and area under the curve (AUC).

The first four measures are defined as

$$\begin{cases} \mathrm{Sen} = \dfrac{g(1,1)}{g(1,1)+g(1,2)} & \mathrm{Spc} = \dfrac{g(2,2)}{g(2,2)+g(2,1)}, \\ \mathrm{Prc} = \dfrac{g(1,1)}{g(1,1)+g(2,1)} & \mathrm{Acc} = \dfrac{g(1,1)+g(2,2)}{g(1,1)+g(2,2)+g(1,2)+g(2,1)}. \end{cases} \tag{21}$$

F1, MCC [24], and FMI [25] are defined as

$$\begin{cases} F_1 = 2 \times \dfrac{\mathrm{Sen} \times \mathrm{Prc}}{\mathrm{Sen} + \mathrm{Prc}} = \dfrac{2 \times g(1,1)}{2 \times g(1,1) + g(1,2) + g(2,1)}, \\ \mathrm{MCC} = \dfrac{g(1,1) \times g(2,2) - g(2,1) \times g(1,2)}{\sqrt{[g(1,1)+g(2,1)] \times [g(1,1)+g(1,2)] \times [g(2,2)+g(2,1)] \times [g(2,2)+g(1,2)]}}, \\ \mathrm{FMI} = \sqrt{\mathrm{Sen} \times \mathrm{Prc}} = \sqrt{\dfrac{g(1,1)}{g(1,1)+g(1,2)} \times \dfrac{g(1,1)}{g(1,1)+g(2,1)}}. \end{cases} \tag{22}$$

The above measures are calculated in the mean and standard deviation (MSD) format. Furthermore, ROC is a curve to measure a binary classifier with varying discrimination thresholds [26]. The ROC curve is created by plotting the sensitivity against 1-specificity. The AUC is calculated based on the ROC curve [27].

4. Experimental Results

4.1. Statistical Analysis. The result of the SOSPCNN model using configuration V is itemized in Table 5. The model arrives at a performance with a sensitivity of 92.25 ± 2.19, a specificity of 92.75 ± 2.49, a precision of 92.79 ± 2.29, an

Table 7: Comparison with state-of-the-art approaches.

Approach	Sen	Spc	Prc	Acc	F1	MCC	FMI
MC [8]	86.25 ± 3.58	80.75 ± 3.55	81.88 ± 2.32	83.50 ± 0.79	83.92 ± 0.97	67.25 ± 1.56	84.00 ± 0.98
3DCNN [9]	91.00 ± 3.16	89.50 ± 3.29	89.77 ± 2.70	90.25 ± 1.42	90.32 ± 1.42	80.63 ± 2.79	90.35 ± 1.41
VCCNN [10]	90.75 ± 1.69	90.00 ± 2.36	90.14 ± 1.95	90.38 ± 0.84	90.41 ± 0.78	80.80 ± 1.64	90.43 ± 0.76
SOSPCNN (ours)	92.25 ± 2.19	92.75 ± 2.49	92.79 ± 2.29	92.50 ± 1.18	92.48 ± 1.17	85.06 ± 2.38	92.50 ± 1.17

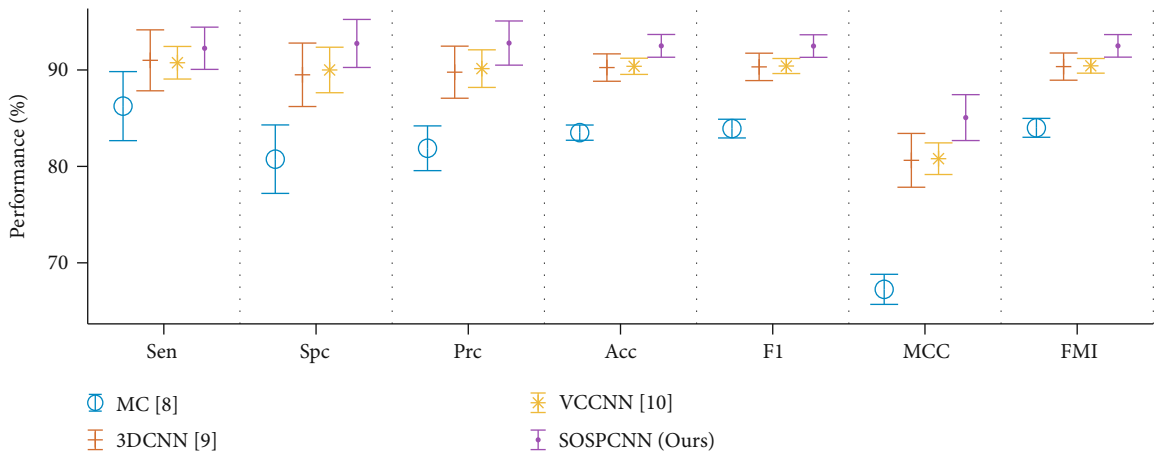


Figure 13: Error bar comparison.

accuracy of 92.50 ± 1.18, an F1 score of 92.48 ± 1.17, an MCC of 85.06 ± 2.38, and an FMI of 92.50 ± 1.17.

Figure 8 shows the confusion matrix of 10 × 10-fold crossvalidation, where we can see the TP = 369, FN = 31, TN = 371, and FP = 29, indicating 31 TOF are wrongly classified as HC while 29 HC are misclassified to TOF. Hence, the sensitivity is $369/(369 + 31) = 92.25\%$, and specificity is $371/(29 + 371) = 92.75\%$.

4.2. Configuration Comparison. We compare nine configurations (see Table 2). The validation is the same as previous experiment. Due to the page limit, the detailed statistical analysis is not shown. The ROC and AUC values are displayed in Figure 9. The AUC values of nine networks with different configurations are: 0.9502, 0.9511, 0.9504, 0.9532, 0.9587, 0.9577, 0.9360, 0.9419, and 0.9389 (as shown in Figure 10). We can observe from Figure 10 that the best network is with configuration V, whose structure is shown in Table 3.

4.3. Effect of Multiple-Way Data Augmentation. Figure 11 shows the multiple-way DA results if we take Figure 2(a) as the raw training examples. Due to the page limit, the multiple-way DA results on the horizontally mirrored image are not displayed. As we can see from Figure 11, multiple-way DA increases the diversity of the training images.

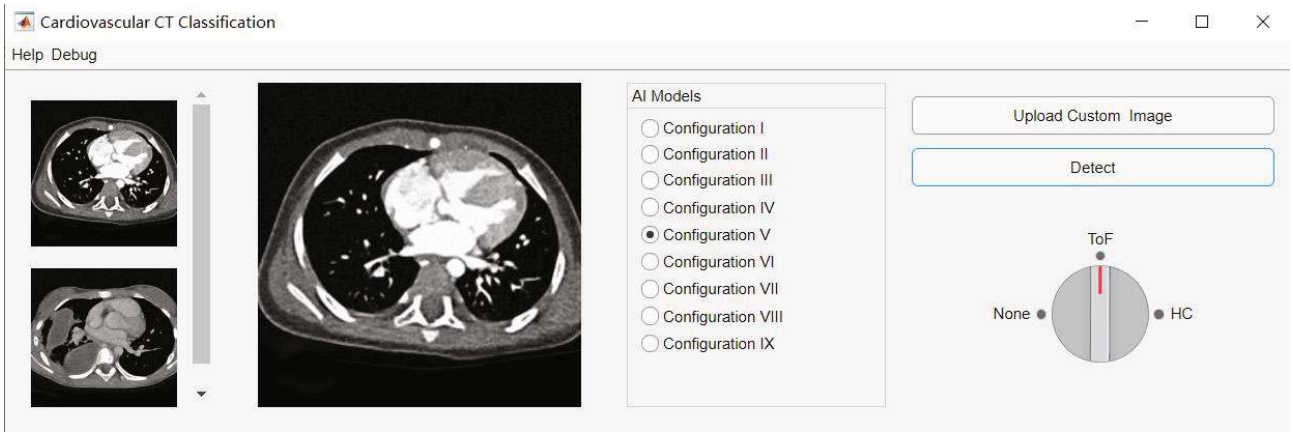
If we remove the multiple-way data augmentation from our model, the performances are decreased, as shown in Table 6, where MSD stands for mean and standard deviation. Comparing Table 5 with Table 6, we can observe multiple-way DA is efficient in improving the classification performance. The reason is that it helps our model resist overfitting by enhancing the diversity of the training set.

4.4. Explainability. Figure 12 shows the manual delineation and the heat map of Figure 2(a) via Grad-CAM described in Section 3.5. The manual delineation showed the radiologist make decisions on "TOF" diagnosis based on all the areas of the abnormal heart, while the heat map shows the proposed SOSPCNN model also puts more focus on the heart region other than the surrounding tissues and background areas.

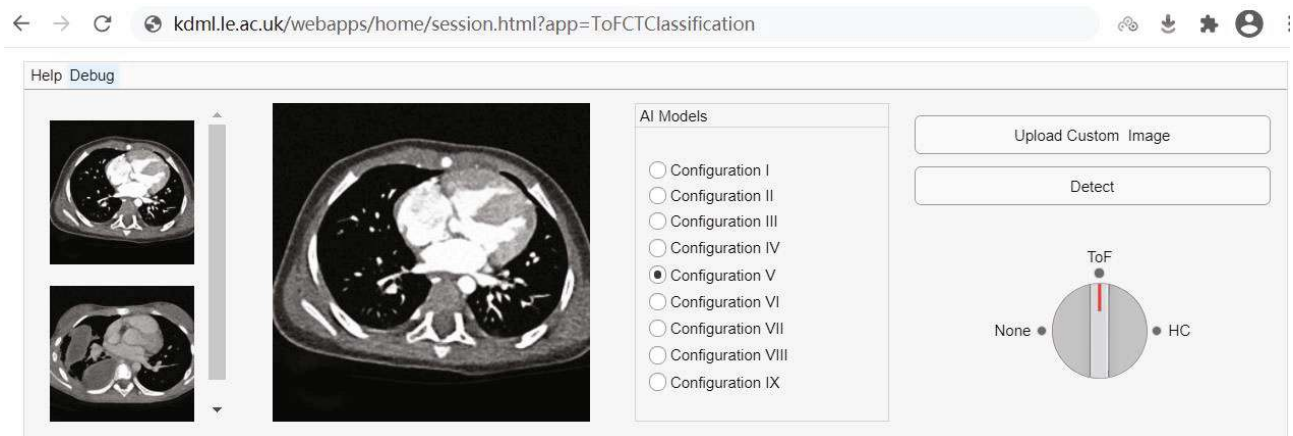
4.5. Comparison with State-of-the-Art Approaches. We compare the proposed SOSPCNN model with three other approaches: MC [8], 3DCNN [9], and VCCNN [10]. The results are shown in Table 7. Note that some comparison methods are not suitable for our dataset, so we modify them to adapt to our dataset.

The error bar comparison is drawn in Figure 13, which clearly shows that the proposed SOSPCNN outperforms all three comparative approaches. The reason is three folds: (i) we use stochastic pooling to replace traditional max-pooling; (ii) we use structural optimization to determine the optimal structure of our SOSPCNN model; (iii) multiple-way DA is included to increase the diversity of training images. In the future, more advanced techniques [28–30] will be tested and integrated into our model.

4.6. Desktop and Web Apps. MATLAB app designer is used to create a professional application for both desktop and web. The input to this web app is any cardiovascular CT image, and the aforementioned SOSPCNN model is included in our app. Figure 14(a) displays the graphical user interface (GUI) of the standalone desktop app. The users can upload their custom images, and the software can show the results by turning the knob into the correct texts: TOF, HC, or none.



(a) Desktop app



(b) Web app

Figure 14: GUI of developed apps.

Figure 14(b) shows the GUI of the web app that is accessed through a "Google Chrome" web browser. The web app is based on a client-server modeled structure [31], i.e., the user is provided services through an off-site server hosted by a third-party cloud service, Microsoft Azure in our study. Our developed online web app can assist hospital clinicians in making decisions remotely and effectively.

5. Conclusion

This paper proposes a web app for TOF recognition. Our proposed model is termed structurally optimized stochastic pooling convolutional neural network (SOSPCNN) with explainable property achieved by Grad-CAM. The results by ten runs of 10-fold crossvalidation show that this SOSPCNN model yields a sensitivity of 92.25 ± 2.19, a specificity of 92.75 ± 2.49, a precision of 92.79 ± 2.29, an accuracy of 92.50 ± 1.18, an F1 score of 92.48 ± 1.17, an MCC of 85.06 ± 2.38, an FMI of 92.50 ± 1.17, and an AUC of 0.9587. Further, we develop both desktop and web apps to realize this SOSPCNN model.

The shortcomings of our method are as follows: (i) our model is trained on a small dataset; (ii) our model does not go through strict medical verification; (iii) our model only considers TOF and HC.

Therefore, we shall attempt to solve the above three weak points in the future. We shall try to collect more TOF and HC cardiovascular CT images. We shall invite clinicians to use our web app and return feedbacks so that we can continue to improve our model. We shall try to collect data of other heart diseases, so make our model can identify more types of diseases.

Data Availability

Data is available upon reasonable requests to corresponding authors.

Conflicts of Interest

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

Authors' Contributions

Shui-Hua Wang and Kaihong Wu contributed equally to this work.

Acknowledgments

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); and Global Challenges Research Fund (GCRF), UK (P202PF11).

References

[1] D. Carli, A. Moroni, A. Zonta et al., "Atypical microdeletion 22q11.2 in a patient with tetralogy of Fallot," *Journal of Genetics*, vol. 100, no. 1, pp. 1–4, 2021.

[2] M. Ghaderian, A. Ahmadi, M. R. Sabri et al., "Clinical outcome of right ventricular outflow tract stenting versus Blalock-Taussig shunt in Tetralogy of Fallot: a systematic review and meta-analysis," *Current Problems in Cardiology*, vol. 46, no. 3, article 100643, 2021.

[3] M. Uecker, C. Petersen, C. Dingemann, C. Fortmann, B. M. Ure, and J. Dingemann, "Gravitational autoreposition for staged closure of omphaloceles," *European Journal of Pediatric Surgery*, vol. 30, no. 1, pp. 45–50, 2020.

[4] E. Cambronero-Cortinas, P. Moratalla-Haro, A. E. González-García et al., "Predictors of atrial tachyarrhythmias in adults with congenital heart disease," *Kardiologia Polska*, vol. 78, no. 12, pp. 1262–1270, 2020.

[5] T. Ashraf, F. Farooq, A. S. Muhammad et al., "Coronary artery anomalies in Tetralogy of Fallot patients undergoing CT angiography at a tertiary care hospital," *Cureus*, vol. 12, no. 9, article e10723, 2020.

[6] M. Engbersen, M. Versleijen, D. Lambregts, R. Beets-Tan, M. Tesselaar, and L. Max, "Comparison of whole-body MRI and 68Ga-DOTATATE PET-CT findings in patients with suspected peritoneal metastases from neuroendocrine tumors," *Journal of Neuroendocrinology*, vol. 33, pp. 120–120, 2021.

[7] F. Shan, Y. Gao, J. Wang et al., "Abnormal lung quantification in chest CT images of COVID-19 patients with deep learning and its application to severity prediction," *Medical Physics*, vol. 48, no. 4, pp. 1633–1645, 2021.

[8] D. H. Ye, H. Litt, C. Davatzikos, and K. M. Pohl, "Morphological classification: application to cardiac MRI of Tetralogy of Fallot," in *Functional Imaging and Modeling of the Heart*, D. N. Metaxas and L. Axel, Eds., pp. 180–187, Springer-Verlag Berlin, Berlin, 2011.

[9] A. Giannakidis, K. Kamnitsas, V. Spadotto et al., "Fast fully automatic segmentation of the severely abnormal human right ventricle from cardiovascular magnetic resonance images using a multi-scale 3D convolutional neural network," in *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pp. 42–46, New York, 2016.

[10] A. Tandon, N. Mohan, C. Jensen et al., "Retraining convolutional neural networks for specialized cardiovascular imaging tasks: lessons from tetralogy of Fallot," *Pediatric Cardiology*, vol. 42, no. 3, pp. 578–589, 2021.

[11] M. Klimo, P. Lukáč, and P. Tarábek, "Deep neural networks classification via binary error-detecting output codes," *Applied Sciences*, vol. 11, no. 8, article 3563, 2021.

[12] V. S. Alfio, D. Costantino, and M. Pepe, "Influence of image TIFF format and JPEG compression level in the accuracy of the 3D model and quality of the orthophoto in UAV photogrammetry," *Journal of Imaging*, vol. 6, no. 5, 2020.

[13] S. Schmid, J. Krabusch, T. Schromm et al., "A new approach for automated measuring of the melt pool geometry in laser-powder bed fusion," *Progress in Additive Manufacturing*, vol. 6, no. 2, pp. 269–279, 2021.

[14] S. Hegde and S. Gangisetty, "PIG-Net: inception based deep learning architecture for 3D point cloud segmentation," *Computers & Graphics*, vol. 95, pp. 13–22, 2021.

[15] T. Vrzal, M. Malečková, and J. Olšovská, "DeepReI: deep learning-based gas chromatographic retention index predictor," *Analytica Chimica Acta*, vol. 1147, pp. 64–71, 2021.

[16] J. Pokhrel and J. Seo, "Statistical model for fragility estimates of offshore wind turbines subjected to aero-hydro dynamic loads," *Renewable Energy*, vol. 163, pp. 1495–1507, 2021.

[17] K. C. Jung and S. H. Chang, "Advanced deep learning model-based impact characterization method for composite laminates," *Composites Science and Technology*, vol. 207, article 108713, 2021.

[18] A. Rahman, P. Deshpande, M. S. Radue et al., "A machine learning framework for predicting the shear strength of carbon nanotube-polymer interfaces based on molecular dynamics simulation data," *Composites Science and Technology*, vol. 207, article 108627, 2021.

[19] W. Zhu, "ANC: attention network for COVID-19 explainable diagnosis based on convolutional block attention module," *Computer Modeling in Engineering & Sciences*, vol. 127, no. 3, pp. 1037–1058, 2021.

[20] 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, no. 1, p. 9, 2021.

[21] M. Rajapandy and A. Anbarasu, "An improved unsupervised learning approach for potential human microRNA-disease association inference using cluster knowledge," *Network Modeling and Analysis in Health Informatics and Bioinformatics*, vol. 10, no. 1, 2021.

[22] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: visual explanations from deep networks via gradient-based localization," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336–359, 2020.

[23] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2921–2929, Las Vegas, NV, USA, 2016.

[24] A. Alahmadi, A. Davies, J. Royle et al., "An explainable algorithm for detecting drug-induced QT-prolongation at risk of torsades de pointes (TdP) regardless of heart rate and T-wave morphology," *Computers in Biology and Medicine*, vol. 131, article 104281, 2021.

[25] C. E. Coipan, T. J. Dallman, D. Brown et al., "Concordance of SNP- and allele-based typing workflows in the context of a large-scale international Salmonella Enteritidis outbreak investigation," *Microbial Genomics*, vol. 6, no. 3, article 000318, 2020.

[26] I. Ali and P. A. Kalra, "A validation study of the 4-variable and 8-variable kidney failure risk equation in transplant recipients in the United Kingdom," *Bmc Nephrology*, vol. 22, no. 1, p. 57, 2021.

[27] A. Wubalem, "Landslide susceptibility mapping using statistical methods in Uatzau catchment area, northwestern Ethiopia," *Geoenvironmental Disasters*, vol. 8, no. 1, pp. 1–21, 2021.

[28] Y. Zhang, S. Wang, K. Xia, Y. Jiang, and P. Qian, "Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion," *Information Fusion*, vol. 66, pp. 170–183, 2021.

[29] Y. Zhang, F. L. Chung, and S. Wang, "Clustering by transmission learning from data density to label manifold with statistical diffusion," *Knowledge-Based Systems*, vol. 193, article 105330, 2020.

[30] Y. Zhang, F. L. Chung, and S. Wang, "Fast exemplar-based clustering by gravity enrichment between data objects," *IEEE Transactions on Systems Man Cybernetics-Systems*, vol. 50, no. 8, pp. 2996–3009, 2020.

[31] H. M. Salama, M. Z. A. el Mageed, G. I. M. Salama, and K. M. Badran, "CSMCSM," *International Journal of Information Security and Privacy*, vol. 15, no. 1, pp. 44–64, 2021.