Secondary Pulmonary Tuberculosis Identification Via pseudo-Zernike Moment and Deep Stacked Sparse Autoencoder

Shui-Hua Wang · Suresh Chandra Satapathy · Qinghua Zhou · Xin Zhang · Yu-Dong Zhang ⬤

Abstract Secondary pulmonary tuberculosis (SPT) is one of the top ten causes of death from a single infectious agent. To recognize SPT more accurately, this paper proposes a novel artificial intelligence model, which uses Pseudo Zernike moment (PZM) as the feature extractor and deep stacked sparse autoencoder (DSSAE) as the classifier. In addition, 18-way data augmentation is employed to avoid overfitting. This model is abbreviated as PZM-DSSAE. The ten runs of 10-fold cross-validation show this model achieves a sensitivity of $93.33\% \pm 1.47\%$, a specificity of $93.13\% \pm 0.95\%$, a precision of $93.15\% \pm 0.89\%$, an accuracy of $93.23\% \pm 0.81\%$, and an F1 score of $93.23\% \pm 0.83\%$. The area-under-curve reaches 0.9739. This PZM-DSSAE is superior to 5 state-of-the-art approaches.

Shui-Hua Wang, Suresh Chandra Satapathy and Qinghua Zhou contributed equally to this work.

S.-H. Wang
School of Mathematics and Actuarial Science, University of Leicester, Leicester LE1 7RH, UK
e-mail: shuihuawang@ieee.org

S. C. Satapathy
School of Computer Eng, KIIT Deemed to University, Bhubaneswar, India
e-mail: sureshsatapathy@ieee.org

Q. Zhou · Y.-D. Zhang (✉)
School of Informatics, University of Leicester, Leicester LE1 7RH, UK
e-mail: yudongzhang@ieee.org

X. Zhang (✉)
Department of Medical Imaging, The Fourth People's Hospital of Huai'an, Huai'an 223002, Jiangsu Province, China
e-mail: 973306782@qq.com

1 Background and Introduction

Tuberculosis (TB) generally infects the lungs; hence, it is frequently meant as pulmonary Tuberculosis (PT); nonetheless, TB can affect any part of the human body, viz., nervous systems, glands, abdomen, etc. [1]. TB outside of the lung is dubbed extra-pulmonary tuberculosis (EPT) [2]. Globally, PT is one of the top 10 causes of death and the leading cause of a single infectious agent, as reported by WHO.

Several standard tests could be employed to confirm suspected PT or EPT. Those tests include CT or MRI or ultrasound scans, blood or urine test, endoscopy [3] or laparoscopy [4], any biopsy tests from the infected areas, etc. Sometimes, a lumbar puncture [5] is needed where a small amount of cerebrospinal fluid is taken from the base of the spine.

PT could be divided into two classes: (i) primary pulmonary tuberculosis (PPT) [6], (ii) secondary pulmonary tuberculosis (SPT) [7]. SPT is investigated

because most PT cases belong to SPT. The most widespread locales of SPT lesions include the apical and posterior segments of the upper lobe and the dorsal segment of the lower lobe [7]. The chest CT (CCT) exhibitions are enumerated as exudative patchy shadows of uneven density, miliary shadows, tree bud signs, proliferative nodular shadows, caseous pneumonia, cavities, tuberculosis bulbs, bronchial dissemination, satellite foci, calcifications and cord strips, and lots of other forms [8, 9].

In the past, manual analysis of SPT is tedious and onerous. In addition, the interpretation results are without doubt influenced by intra−/inter-expert factors. Recently, scholars favor the use of advanced artificial intelligence tools to design automatic SPT classification methods. For instance, Bagci, et al. [10] proposed a computer-aided quantification and detection model for cavitary TB. Based on the support vector machine (SVM), their team proposed a new shape-based SVM (SSVM) to detect airways and cavities. Nevertheless, this method has to manually extract shape-based features. Li, et al. [11] combined features from both convolutional neural network (CNN) and AutoEncoder (AE). Their team dubbed their method as "AECNN". However, the performances (Accuracy=0.8126, F1=0.8149, and Recall=0.8172) are insufficient in clinical usage. Based on deep learning (DL), Park, et al. [12] proposed a DL-based automatic detection (DLAD) algorithm to detect active pulmonary tuberculosis. Using the high-sensitivity cutoff, the sensitivities and specificities were 94.3%-100% and 91.1%-100%. Using the high-specificity cutoff, the figure turned to 84.1%-99.0% and 99.1%-100%. James-Reynolds, et al. [13] used a three-dimensional block-based residual deep learning (RDL) model with depth information. This method is shortened to RDL. Nevertheless, 3D models consume more computation costs and resources. They can easily overfit during training. Xie, et al. [14] presented a computer-aided system (CAS) for multiple-category PT detection in radiograms. A learning scalable pyramid structure was utilized for the faster RCNN. The weakness is that the pyramid structure in faster RCNN is still inefficient. The data augmentation implemented merely enhanced the size to 5x of the original size; thus, the improvement is restricted.

In order to improve the diagnosis performance of SPT, this paper proposed a novel artificial intelligence model based on the pseudo-Zernike moment (PZM) and deep stacked sparse autoencoder (DSSAE). The novelties of this study are four points:

(i) We are the first to apply PZM to automatic SPT diagnosis.
(ii) A novel PZM-DSSAE model is proposed to identify SPT.
(iii) 18-way data augmentation is used to prevent overfitting.
(iv) The proposed "PZM-DSSAE" model is better than 5 state-of-the-art approaches.

2 Dataset

The dataset was described in Ref. [15, 16], of which the retrospective study was exempt by the Institutional Review Board of local hospitals. The data is available upon reasonable request to the corresponding authors. All images are stored in picture archiving and communication systems (PACS) format. Besides, the detailed demographics are list in Table 1. The data is available upon reasonable request to corresponding authors.

Two junior radiologists V_a and V_b with ten years of chest-related diagnostic experiences read the radiographs collectively. V_a and V_b record the sizes, distributions, and morphological shapes of the CT manifestations of the lesions, and then make slice-wise annotations. Several slices (More than 0 and less than 5) are selected by the slice level selection method [15]. V_c is the senior radiologist. The sizes of all images are stored as 1024×1024.

Let x_0 stands for a lung-window CCT image. The labeling R is yielded via.

$$R(x_0) = \begin{cases} V_a(x_0) & V_a(x_0) = V_b(x_0) \\ \mathrm{MV}\left[V_a(x_0), V_b(x_0), V_c(x_0)\right] & V_a(x_0) \neq V_b(x_0) \end{cases}$$
(1)

Table 1 Demographics of SPT and HC

Category	NS	NI
SPT	68	144
HC	66	144

NS number of subjects, *NI* number of images

When the radiologist V_a disagrees with V_b, we will invite the senor one V_c to join and make a consensus by majority voting (MV).

3 Methodology

3.1 Preprocessing

Table 2 shows the abbreviation list. Let each raw image be x_o, the original dataset is written as $X_0 = \{x_0(i)\}$. The four preprocessing steps can be written as.

$$x_0 \overset{\text{grayscale}}{\mapsto} x_g \overset{\text{HS}}{\mapsto} x_{HS} \overset{\text{Cropping}}{\mapsto} x_{crop} \overset{\text{Resizing}}{\mapsto} x \qquad (2)$$

where x_g, x_{HS}, x_{crop}, and x are the results of grayscale operation, histogram stretch (HS), cropping, and

Table 2 Abbreviation list

Abbreviation	Meaning
AE	autoencoder
AI	artificial intelligence
AUC	area under the curve
CCT	chest computed tomography
DA	data augmentation
DCR	data compression ratio
DSSAE	deep stacked sparse autoencoder
EPT	extra-pulmonary tuberculosis
FMI	Fowlkes–Mallows index
MCC	Matthews correlation coefficient
MDA	multiple-way data augmentation
MSD	mean and standard deviation
MV	majority voting
PACS	picture archiving and communication systems
PPT	primary pulmonary tuberculosis
PT	pulmonary tuberculosis
PZF	pseudo Zernike function
PZM	pseudo Zernike moment
PZP	pseudo-Zernike polynomial
ROC	receiver operating characteristic
SAE	sparse autoencoder
SAPN	salt-and-pepper noise
SN	speckle noise
SPT	secondary pulmonary tuberculosis
SSR	space saving ratio
TB	tuberculosis

resizing. Four points should be noted: (i) All CCT images are saved in red, green, and blue (RGB) layout though those images are grayscale essentially; (ii) cropping values (z_t, z_b, z_l, z_r) are measured in terms of pixels from four directions: top, bottom, left, and right, respectively; (iii) $X = \{x(i)\}$ denotes the ultimate output of preprocessing, and (iv) the resized $[W, W]$ image set X is utilized for input to the proposed AI model.

Suppose $W = 256$ and only consider individual image size, the data compression ratio (DCR) of the whole preprocessing is deduced as:

$$DCR = \frac{\text{size}(x_0)}{\text{size}(x)} = \frac{1024^2 \times 3}{256 \times 256} = 48 \qquad (3)$$

and the space-saving ratio (SSR) is defined as.

$$SSR = 1 - \frac{\text{size}(x)}{\text{size}(x_0)} = 1 - \frac{256^2}{1024^2 \times 3} = 97.92\% \qquad (4)$$

This high DCR of 48 and SSR of 97.92% could help the AI model prediction and future cloud computing and online web service. Figure 1 shows the diagram of preprocessing procedures. Figure 2 displays two samples of the preprocessed dataset.

3.2 Pseudo-Zernike Moment

Image moment was firstly introduced by Hu [17], who employed geometric moments to generate a set of invariants. Nevertheless, geometric moments are sensitive to noise. Thus, Teague [18] introduced Zernike moments (ZMs) based on orthogonal Zernike polynomials. The orthogonal ZMs have been verified
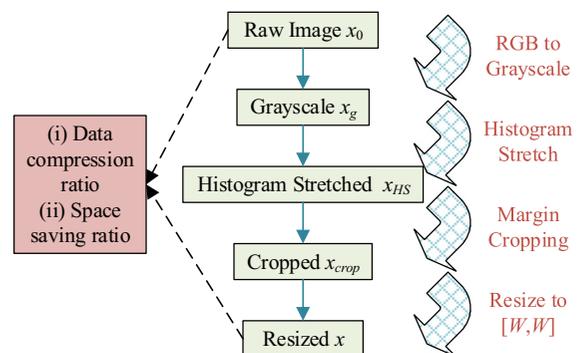

Fig. 1 Diagram of preprocessing

Fig. 2 Sample of the pre-processed dataset

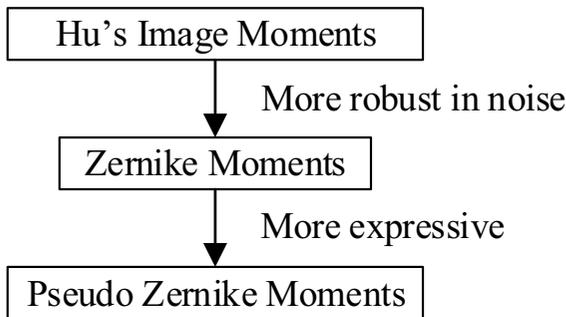(a) SPT   (b) HC

Fig. 3 Development of image moments

to behave more robust in noisy conditions, and they can yield an almost zero value of redundancy measure [19] (Fig. 3).

Later, pseudo Zernike moment (PZM) is derived from Zernike moment. PZMs have been proven to give better performances than other moment functions such as Hu moments, Zernike moments, etc. For example, for an order p, there are N_p^{PZM} linearly independent pseudo-Zernike polynomials of orders $\leq p$, while there are only N_p^{ZM} Zernike polynomials, where.

$$\begin{cases} N_p^{PZM} = (p+1)^2 \\ N_p^{ZM} = \frac{1}{2}(p+1)(p+2) \end{cases} \tag{5}$$

Hence, PZM is more expressive and offers more feature vectors than ZM. PZM has been successfully applied in breast cancer diagnosis [20], pathological brain detection [21], etc.

The kernel of PZMs is a set of orthogonal pseudo-Zernike polynomials defined inside a unit circle (UC) using the polar coordinate. The transformation between Cartesian and polar coordinates are.

$$\begin{cases} r^2 = x^2 + y^2 \\ x = r \times \cos\theta \\ y = r \times \sin\theta \end{cases}, \tag{6}$$

The PZM y_{pq} of order p with repetition q of an image $x(r, \theta)$ is defined as [22].

$$y_{pq} = \frac{p+1}{\pi} \int_{-\pi}^{\pi} \int_0^1 V_{pq}^*(r, \theta)x(r, \theta)r\mathrm{d}r\mathrm{d}\theta, \tag{7}$$

where the pseudo-Zernike polynomials (PZPs) $V_{pq}(r, \theta)$ of the order p are defined as.

$$V_{pq}(r, \theta) = Q_{pq}(r)e^{jq\theta}, j = \sqrt{-1} \tag{8}$$

where radial polynomials R_{pq} are defined as.

$$Q_{pq}(r) = \sum_{k=0}^{p-|q|} (-1)^k \frac{(2p+1-k)!}{k!(p+|q|+1-k)!(p-|q|-k)!} r^{p-k} \tag{9}$$

where $0 \leq |q| \leq p$.

PZM triangular matrix $Y = \{y_{pq}\}$, $0 \leq p \leq p_{max}$ are defined in terms of polar coordinates (r, θ) within $|r| \leq 1$. Therefore, the computation of PZM needs a linear transformation of the image plane (IP) coordinates (w_1, w_2), $1 \leq w_1 \leq W$, $1 \leq w_2 \leq W$ to the UC

Fig. 4 Two transformation

(a) Raw image plane $W \times W$ (b) IP over UC   (c) IP inside UC
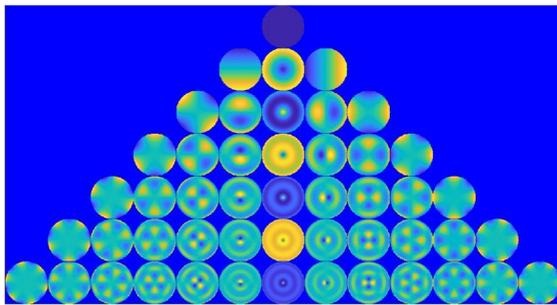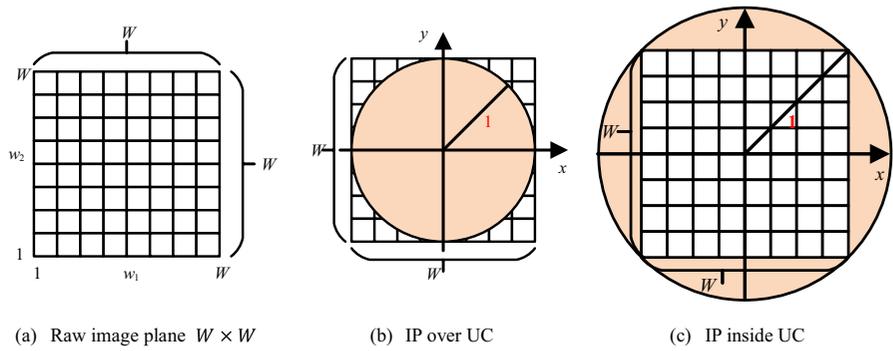

Fig. 5 PZF with maximum order $0 \leq p \leq p_{max} = 6$

domain $(x, y) \in \mathbb{R}^2$. There are two commonly used transformations as shown in Fig. 4: (i) IP over UC; and (ii) IP inside UC. In this study, we use the former transformation (IP over UC, See Fig. 4b), because the lesions of SPT will not occur within the four corners of the CCT image.

In practice, pseudo-Zernike functions (PZFs) [23] are used instead of PZPs for simplicity and fast calculation. The PZFs with orders $0 \leq p \leq p_{max}$, $p_{max} = 6$ are displayed in Fig. 5, where we can see in total $(p_{max} + 1)^2 = 49$ PZFs. An example of PZM

Fig. 6 PZM decomposition of a peak function

(a) Raw   (b) Reconstructed   (c) Difference

(d) Detailed decomposition ($p_{max} = 6$)
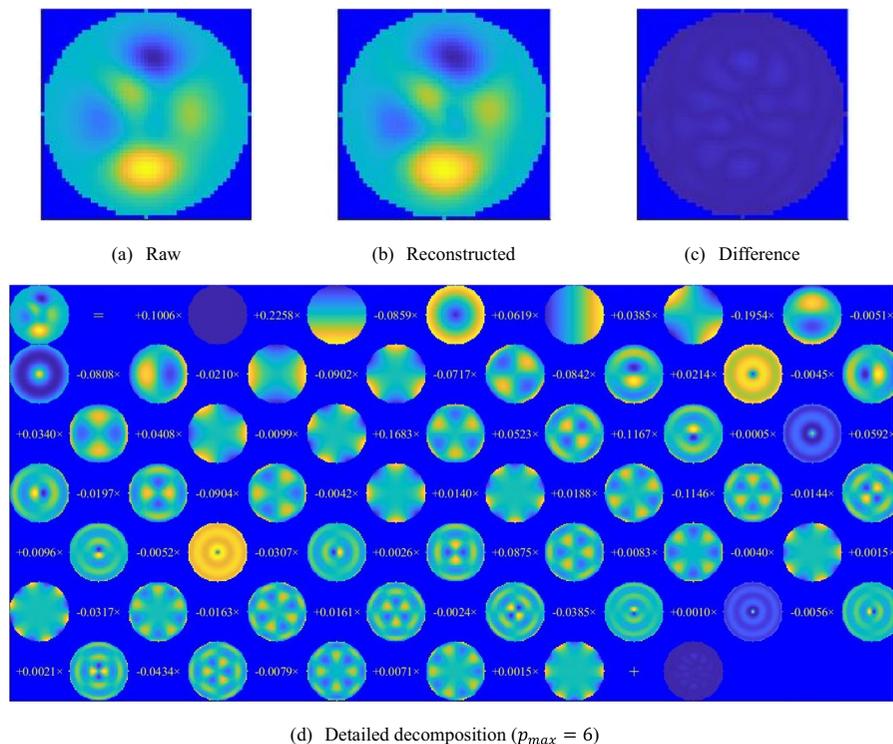
decomposition is shown in Fig. 6, in which (a-c) show the raw peak image, the reconstructed image using $p_{max} = 6$ PZFs, and the difference between the raw and reconstructed image. Figure 6(d) gives the detailed decomposition and the corresponding PZMs. There are many other excellent feature extractors which we will investigate in our future studies.

3.3 Autoencoder

The PZM triangular matrix is vectorized into a $(p_{max} + 1)^2$ vector feature Y, and then are sent to a custom deep stacked sparse autoencoder (DSSAE), which is one of the deep learning models [24, 25]. The fundamental element of DSSAE in the autoencoder (AE) [26], which is a typical neural network that learns to map its input Y to output Z.

There is an internal code output C that represents the input Y. The whole AE can be split into two parts: An encoder part (W_Y, B_Y) that maps the input Y to the code C, and a decoder part (W_Z, B_Z) that maps the code C to a reconstructed data Z, namely,

$$Y \overset{encoder}{\mapsto} C \overset{decoder}{\mapsto} Z \qquad (10)$$

We expect the output Z equals to the input Y.

The structure of AE is displayed in Fig. 7, where the encoder part is with weight W_Y and bias B_Y, and the decoder part is with weights W_Z and bias B_Z. We have.
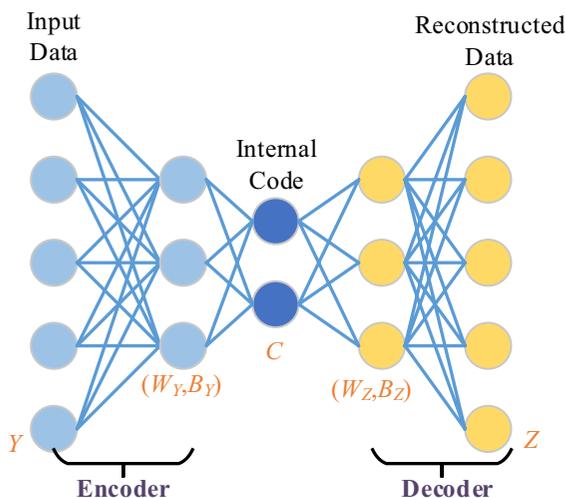

Fig. 7 Structure of an autoencoder

$$C = g_{LS}(W_Y Y + B_Y), \qquad (11)$$

$$Z = g_{LS}(W_Z C + B_Z), \qquad (12)$$

where the output Z is an estimate of input Y, and g_{LS} is the log sigmoid function [27]:

$$g_{LS}(x) = \frac{1}{1 + \exp(-x)} \qquad (13)$$

3.4 Loss Functions of Autoencoder and Sparse Autoencoder

The sparse autoencoder (SAE) [28] is a variant of AE. To minimize the error between the input vector Y and the output Z, the raw loss function of AE is deduced as:

$$I_{AE}(W_Y, W_Z, B_Y, B_Z) = \frac{1}{N_S} \|Z - Y\|^2, \qquad (14)$$

where N_S means the number of training samples. From Eqs. (11), (12), we find the output Z can be expressed in the way of

$$Z = g_{AE}(Y|W_Y, W_Z, B_Y, B_Z), \qquad (15)$$

where g_{AE} is the abstract of the AE model; therefore, Eq. (14) can be revised as

$$I_{AE}(W_Y, W_Z, B_Y, B_Z) = \frac{1}{N_S} \left\| g_{AE}(Y|W_Y, W_Z, B_Y, B_Z) - Y \right\|^2. \qquad (16)$$

Sparse autoencoder (SAE) encourages sparsity into AE. SAE only allows a small fraction of the hidden neurons to be active at the same time [29]. This sparsity forces SAE to respond to unique statistical features of the training data [30].

In practice, one L_2 regularization term Γ_w of the weights (W_Y, W_Z) and one regularization term Γ_s of the sparsity constraint are defined to avoid overcomplete mapping or trivial mapping. The loss function of SAE is written as:

$$I_{SAE}(W_Y, W_Z, B_Y, B_Z) = \frac{1}{N_S} \left\| g_{AE}(Y|W_Y, W_Z, B_Y, B_Z) - Y \right\|^2 + c_s \times \Gamma_s + c_w \times \Gamma_w, \qquad (17)$$

where c_s stands for the sparsity regulation factor, and c_w the weight regulation factor. The sparsity regularization term is defined as:

$$\Gamma_s = \sum_{j=1}^{|C|} g_{KL}(\rho, \hat{\rho}_j) = \sum_{j=1}^{|C|} \rho \log \frac{\rho}{\hat{\rho}_j} + (1-\rho) \log \frac{1-\rho}{1-\hat{\rho}_j}, \tag{18}$$

where g_{KL} stands for the Kullback-Leibler divergence [31] function, $|C|$ is the number of elements of internal code output C. $\hat{\rho}_j$ is the j-th neuron's average activation value over all N_S training samples, and ρ is its desired value, named sparsity proportion factor. The weight regularization term Γ_w is defined as.

Table 3 Layer details of proposed PZM-DSSAE model

Layer	Trainable weights	Size
Input		$1024 \times 1024 \times 3$
Preprocessing		$W \times W$
PZM		$\{y_{pq}\}$, $0 \le p \le p_{max}, -p \le q \le p$
Vectorization		$(p_{max}+1)^2$
1st SAE	$(p_{max}+1)^2 \times S_1$	S_1
2nd SAE	$S_1 \times S_2$	S_2
3rd SAE	$S_2 \times S_3$	S_3
4th SAE	$S_3 \times S_4$	S_4
Softmax	$S_4 \times S_c$	S_c

$$\Gamma_w = \frac{1}{2} \times \| W_Y \ W_Z \|_2^2. \tag{19}$$

The training procedure is set to scaled conjugate gradient descent (SCGD) [32] method.

3.5 Stacked Sparse Autoencoder

We use SAE as a building block and create the final deep stacked sparse autoencoder (DSSAE) classifier by following three operations: (i) We include input layer, preprocessing layer, PZM layer, and vectorization layer; (ii) We stack four SAEs with various numbers of hidden neurons; (iii) We append softmax layer at the end of our AI model.
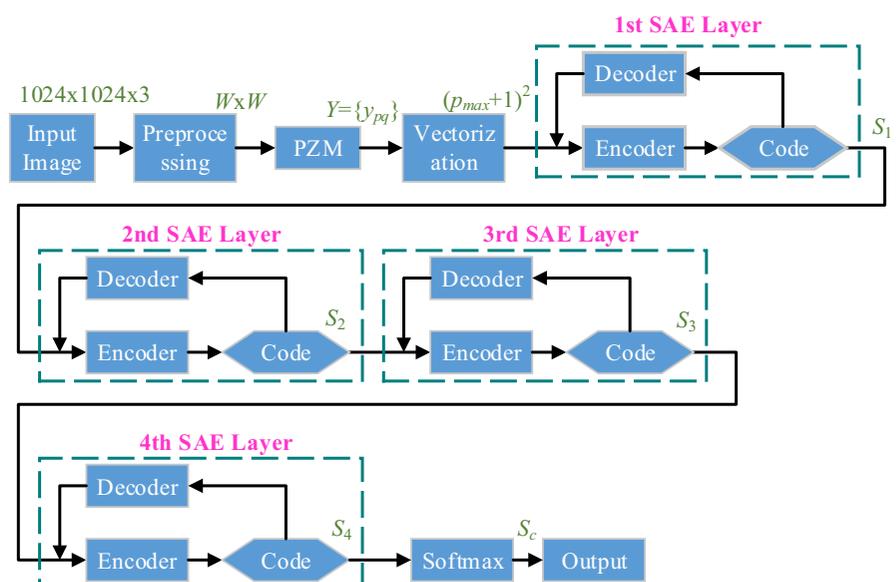
The structure of this proposed PZM-DSSAE model is itemized in Table 3 and illustrated in Fig. 8. The sizes of the input CCT images are $1024 \times 1024 \times 3$. After processing, all the CCT images are normalized to grayscaled images with the sizes of $W \times W$. Afterwards, PZM layer will generate triangular matrix $Y = \{y_{pq}\}$, $0 \le p \le p_{max}, -p \le q \le p$ (See Fig. 5). The vectorization arranges all the PZM into one vector with number of elements of $(p_{max}+1)^{\wedge}2$ (See Fig. 6d).

In the classification stage, four SAE blocks with the number of neurons of (S_1, S_2, S_3, S_4) are employed. Only the encoder parts of the four SAE are stacked (See Fig. 8). Finally, a softmax layer

Fig. 8 Structure of proposed SSAE model

with neurons of S_c is appended, where N_c means the number of classes in our classification task.

3.6 Cross-Validation

The PZM feature set Y is set to V-fold cross-validation. The whole dataset is split into V separate folds. At v-th trial, the v-th fold is used as the test set, and all the rest folds $\{1,\dots,v-1,v+1,\dots,V\}$ are merged and used as training. This procedure repeats until all the folds are used only once as the test set (See Fig. 9).

3.7 Measures

To avoid randomness, we run above V-fold cross-validation U times with different initial random seeds. Suppose the sample number of each class is $T_k(k=1,2)$. The ideal confusion matrix (CM) is.

$$F = \left\{ f^{ideal}(m,n) \right\} = U \times \begin{bmatrix} T_1 & 0 \\ 0 & T_2 \end{bmatrix}, \tag{20}$$

Note here the off-diagonal entries of ideal F are all zero, viz., $f^{ideal}(m,n)=0, \forall m \neq n$. Here suppose "P" and "N" means positive and negative, corresponding to SPT and HC, respectively. The meaning of TP, TN, FP, and FN are shown in Table 4.

Nine measures are used in this study: sensitivity, specificity, precision, accuracy, F1 score, Matthews correlation coefficient (MCC) [33], Fowlkes–Mallows index (FMI) [34], receiver operating characteristic (ROC), area under the curve (AUC). The first four measures are defined as.

$$\begin{cases} \text{Sen} = \frac{f(1,1)}{f(1,1)+f(1,2)} & \text{Spc} = \frac{f(2,2)}{f(2,2)+f(2,1)} \\ \text{Prc} = \frac{f(1,1)}{f(1,1)+f(2,1)} & \text{Acc} = \frac{f(1,1)+f(2,2)}{f(1,1)+f(2,2)+f(1,2)+f(2,1)} \end{cases} \tag{21}$$

and the middle three measures are defined as:

$$\begin{cases} F_1 = 2 \times \frac{\text{Sen} \times \text{Prc}}{\text{Sen} + \text{Prc}} = \frac{2 \times f(1,1)}{2 \times f(1,1)+f(1,2)+f(2,1)} \\ \text{MCC} = \frac{f(1,1) \times f(2,2) - f(2,1) \times f(1,2)}{\sqrt{[f(1,1)+f(2,1)] \times [f(1,1)+f(1,2)] \times [f(2,2)+f(2,1)] \times [f(2,2)+f(1,2)]}} \\ \text{FMI} = \sqrt{\text{Sen} \times \text{Prc}} = \sqrt{\frac{f(1,1)}{f(1,1)+f(1,2)} \times \frac{f(1,1)}{f(1,1)+f(2,1)}} \end{cases} \tag{22}$$
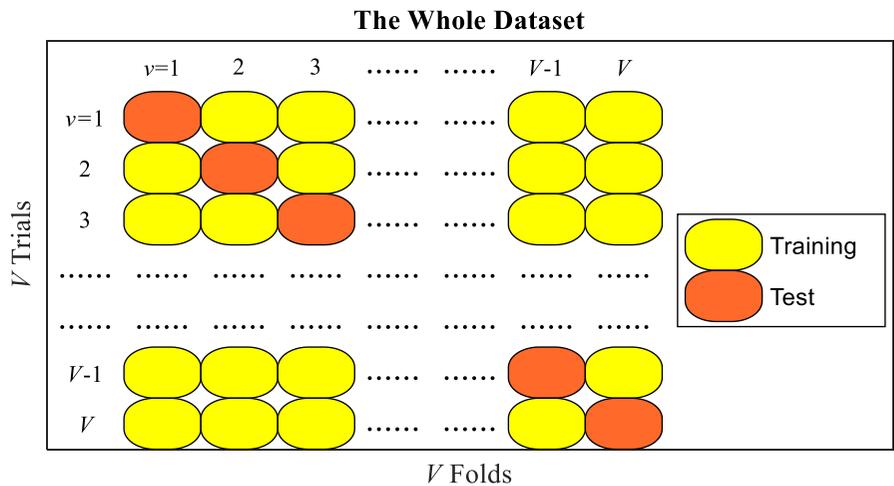
Fig. 9 Diagram of V-fold cross validation

Table 4 Meanings in measures

Abbreviation	Full Form	Symbol	Meaning
P	Positive		SPT
N	Negative		HC
TP	True Positive	$f(1,1)$	SPT images are classified correctly.
FP	False Positive	$f(2,1)$	HC images are wrongly classified as SPT.
TN	True Negative	$f(2,2)$	HC images are classified correctly.
FN	False Negative	$f(1,2)$	SPT images are wrongly classified as HC.

The above measures are calculated in the mean and standard deviation (MSD) format. Besides, ROC is a curve to measure a binary classifier with varying discrimination thresholds. The ROC curve is created by plotting the sensitivity against 1-specificity. The AUC is calculated based on the ROC curve [35].

3.8 Multiple-Way Data Augmentation

Data augmentation creates fake training images to increase the size of the training set. Recently, multiple-way data augmentation (MDA) attract scholar's research interests. Wang [36] proposed a 14-way data augmentation (DA), in which the authors employed seven different DA techniques on raw training image $x(m)$ and its mirrored image $x'(m)$, thus creating 14 new images for each raw image. Cheng [37] proposed a 16-way DA which adds salt-and-pepper noise (SAPN) to the 14-way DA.

This study adds a new DA technique—speckle noise (SN)—to the 16-way DA. Thus, we have an 18-way DA for the training set. Suppose the raw training image is $x(m)$, the SN altered image is defined as

$$x^{SN}(m) = x(m) + N_{SN} * x(m), \tag{23}$$

where N_{SN} is uniformly distributed random noise. The mean and variance of N_{SN} is set to γ_m and γ_v, respectively. The SAPN altered image $x^{SAPN}(m)$ is defined as:

$$\begin{cases} \mathbb{Q}\left(x^{SAPN} = x\right) = 1 - \gamma_d, \\ \mathbb{Q}\left(x^{SAPN} = x_{\min}\right) = \frac{\gamma_d}{2}, \\ \mathbb{Q}\left(x^{SAPN} = x_{\max}\right) = \frac{\gamma_d}{2}, \end{cases} \tag{24}$$

where γ_d stands for noise density, and $\mathbb{Q}$ the probability function. $x_{\min}$ and $x_{\max}$ correspond to the black and white colors, respectively.

Fig. 10 Diagram of proposed 18-way DA
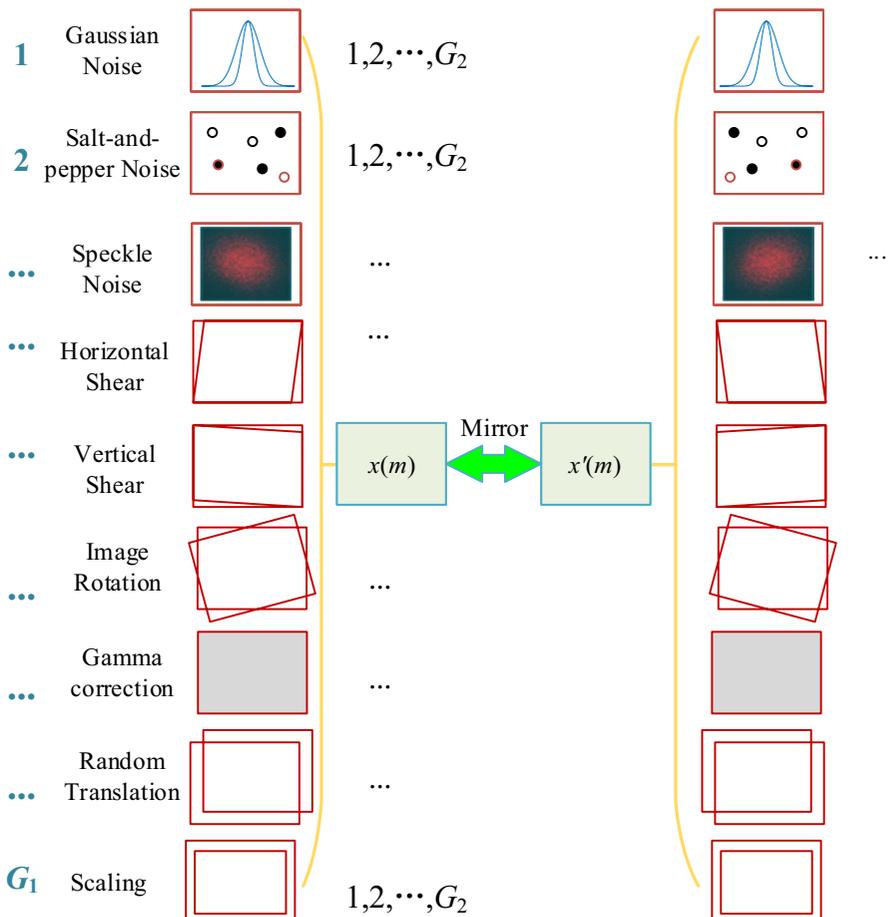
First, G_1 different DA methods as displayed in Fig. 10 are applied to raw training image $x(m)$. Let $\mathbf{H}_g$, $g = 1, \ldots, G_1$ denote each DA operation, we have the augmented datasets on raw image $x(m)$ as.

$$\mathbf{H}_g[x(m)], g = 1, \ldots, G_1 \tag{25}$$

Suppose G_2 stands for the size of generated new images for each DA method, we have.

$$\left| \mathbf{H}_g[x(m)] \right| = G_2 \tag{26}$$

Second, horizontal mirrored image is generated by.

$$x'(m) = f_{HM}[x(m)] \tag{27}$$

where f_{HM} stands for horizontal mirror function.

Third, all the G_1 different DA methods are performed on the mirror image $x'(m)$, and generate G_1 different datasets.

$$\begin{cases} \mathbf{H}_g[x'(m)], g = 1, \cdots, G_1 \\ \left| \mathbf{H}_g[x'(m)] \right| = G_2, g = 1, \ldots, G_1 \end{cases} \tag{28}$$

Fourth, the raw image $x(m)$, the mirrored image $x'(m)$, G_1-way datasets of raw image $\mathbf{H}_g[x(m)]$, and G_1-way datasets of horizontal mirrored image $\mathbf{H}_g[x'(m)]$, are combined. The final generated dataset from $x(m)$ is defined as $\mathbf{I}(m)$:

$$x(m) \mapsto \mathbf{I}(m) = f_{fuse} \left\{ \begin{array}{cc} x(m) & x'(m) \\ \underbrace{\mathbf{H}_1[x(m)]}_{G_2} & \underbrace{\mathbf{H}_1[x'(m)]}_{G_2} \\ \ldots & \ldots \\ \underbrace{\mathbf{H}_{G_1}[x(m)]}_{G_2} & \underbrace{\mathbf{H}_{G_1}[x'(m)]}_{G_2} \end{array} \right\} \tag{29}$$

where f_{fuse} is the concatenation function. Suppose augmentation factor is G_3, which stands for the number of images in $\mathbf{I}(m)$, we have.

$$G_3 = \frac{|\mathbf{I}(m)|}{|x(m)|} = \frac{(1 + G_1 \times G_2) \times 2}{1} = 2 \times G_1 \times G_2 + 2 \tag{30}$$

Algorithm 1 summarizes the pseudocode of the proposed 18-way DA method.

Algorithm 1 Pseudocode of proposed 18-way data augmentation

Input	Import raw preprocessed training image $x(m)$.
Step 1	G_1 geometric/photometric/noise-injection DA transforms $\mathbf{H}_g$ are utilized on $x(m)$.
	We obtain datasets $\mathbf{H}_g[x(m)], g = 1, \ldots, G_1$. See Eq. (25).
	Each enhanced dataset contains G_2 new images. See Eq. (26).
Step 2	The horizontal mirrored image is obtained by $x'(m) = f_{HM}[x(m)]$. See Eq. (27).
Step 3	G_1-way DA transforms are carried out on $x'(m)$,
	we obtain datasets $\mathbf{H}_g[x'(m)], g = 1, \cdots, G_1$. See Eq. (28).
Step 4	$x(m)$, $x'(m)$, $\mathbf{H}_g[x(m)], g = 1, \ldots, G_1$, and $\mathbf{H}_g[x'(m)], g = 1, \ldots, G_1$ are combined. See Eq. (29).
Output	A new dataset $\mathbf{I}(m)$ is provided with the number of images as $G_3 = 2 \times G_1 \times G_2 + 2$. See Eq. (30).

---

4 Experiments and Results

4.1 Parameter Setting

Table 5 shows the parameter setting. The cropping values of top, bottom, left, and right are all set to 200. Resized images are set to the sizes of $[256 \times 256]$.

The maximum order is set to $p_{max} = 24$, indicating we will have $(24 + 1)^2 = 625$ PZMs. The weight regularization factor $c_w = 0.001$, the sparsity regulation factor $c_s = 1.1$, and the sparsity proportion factor is set to $\rho = 0.05$. The number of neurons in the four SAEs is 300, 200, 100, and 60, respectively. The number of classes is set to $S_c = 2$, viz., SPT and HC. V is set to

Table 5 Parameter setting

Parameter	Value
(z_t, z_b, z_l, z_r)	(200,200,200,200)
W	256
p_{max}	24
c_w	0.001
c_s	1.1
ρ	0.05
S_1	300
S_2	200
S_3	100
S_4	60
S_c	2
V	10
U	10
γ_m	0
γ_v	0.05
γ_d	0.05
G_1	9
G_2	30
G_3	542

10 in V-fold cross validation. The model runs $U = 10$ times. The mean and variance of uniformly distributed random noise in SN are 0 and 0.05, respectively. The noise density in SAPN is set to 0.05. The factors in our data augmentation are set to $G_1 = 9$ and $G_2 = 30$; thus, the augmentation factor $G_3 = 524$.

4.2 Illustration of 18-Way Data Augmentation

Suppose Fig. 2a is the raw preprocessed training image $x(m)$, Fig. 11 displays the $\mathbf{H}_g[x(m)]$, $g = 1$, …, G_1 datasets on the image $x(m)$. Due to the page limit, the horizontally mirrored image $x'(m)$ and its corresponding G_1 datasets are not displayed. As shown in Fig. 11, our 18-way DA help increases the diversity of the training set.

4.3 Statistical Analysis

The statistical analysis of the proposed model is displayed in Table 6. As is shown, the ten runs of 10-fold cross-validation obtain the performances of this proposed model. Its sensitivity is

93.33% ± 1.47%, the specificity is 93.13% ± 0.95%, the precision is 93.15% ± 0.89%, the accuracy is 93.23% ± 0.81%, the F1 score is 93.23% ± 0.83%, MCC and FMI are 86.47% ± 1.62% and 93.24 ± 0.83, respectively.

4.4 Effect of MDA

An ablation study is carried out to check the effectiveness of multiple-way data augmentation. If we remove the multiple-way data augmentation from our model, the performance will decrease. The sensitivity and specificity will decrease to 90.07% and 89.24%, respectively. Table 7 clearly shows that the performances of all the seven indicators decrease if we remove MDA from our model. Figure 12 illustrates the ROC plots, from which we can see the AUC will decrease from 0.9739 to 0.9253 if MDA is not used. In future studies, MDA can be integrated into other classification models [38, 39].

4.5 Comparison to State-of-the-Art Methods

This proposed method is compared with 5 state-of-the-art methods: SSVM [10], AECNN [11], DLAD [12], RDL [13], CAS [14]. All the methods are run on the same dataset via ten runs of 10-fold cross-validation. The results are shown in Table 8.

Figure 13 shows the error bar comparison of all SPT identification methods. The figure shows that our PZM-DSSAE model achieves the highest performance among all SPT identification. The reasons for our method achieving the best performance are three folds: (a) We choose PZM as the feature extractor. (b) We introduce DSSAE as the classifier. (c) 18-way data augmentation is used to prevent overfitting.

5 Conclusion

This study proposes a novel AI model of "PZM-DSSAE" for SPT diagnosis. The PZM is used to extract image features, and the DSSAE is used as the classifier. In addition, 18-way data augmentation is used to prevent overfitting. Finally, the whole model "PZM-DSSAE" is better than 5 state-of-the-art approaches.

Fig. 11 Results of G_1-way DA

(a) Gaussian Noise

(b) SAPN

(c) SN

(d) Horizontal Shear

(e) Vertical Shear

(f) Gamma Correction

(g) Random Translation

(h) Scaling

The shortcomings of our method are four folds: (i) We manually seek the optimal order p_{max}. (ii) The structure of DSSAE is obtained by a trial-and-error method. (iii) Our model is only tested by simulation experiments. (iv) The dataset is relatively small.

Table 6 Statistical analysis of proposed model (Unit: %)

Run	Sen	Spc	Prc	Acc	F1	MCC	FMI
1	90.97	95.14	94.93	93.06	92.91	86.19	92.93
2	92.36	92.36	92.36	92.36	92.36	84.72	92.36
3	95.14	93.06	93.20	94.10	94.16	88.21	94.16
4	93.06	92.36	92.41	92.71	92.73	85.42	92.73
5	92.36	91.67	91.72	92.01	92.04	84.03	92.04
6	94.44	93.75	93.79	94.10	94.12	88.20	94.12
7	95.83	93.06	93.24	94.44	94.52	88.92	94.53
8	93.06	93.75	93.71	93.40	93.38	86.81	93.38
9	92.36	93.06	93.01	92.71	92.68	85.42	92.68
10	93.75	93.06	93.10	93.40	93.43	86.81	93.43
MSD	93.33 ±1.47	93.13 ±0.95	93.15 ±0.89	93.23 ±0.81	93.23 ±0.83	86.47 ±1.62	93.24 ±0.83

Table 7 Comparison of using MDA

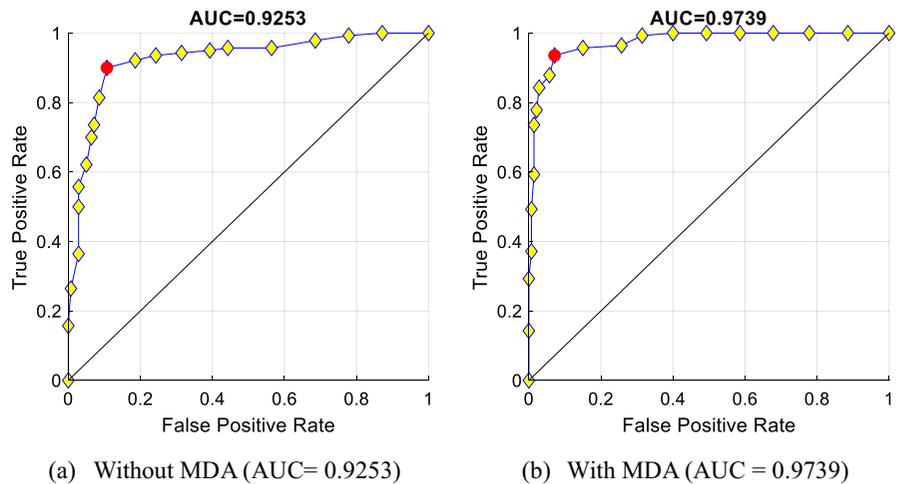| Method	Sen	Spc	Prc	Acc	F1	MCC	FMI
Without MDA	90.07 ±1.76	89.24 ±1.00	89.34 ±0.81	89.65 ±0.78	89.69 ±0.85	79.33 ±1.59	89.70 ±0.86
With MDA	93.33 ±1.47	93.13 ±0.95	93.15 ±0.89	93.23 ±0.81	93.23 ±0.83	86.47 ±1.62	93.24 ±0.83

Fig. 12 Effectiveness of multiple-way data augmentation

(a) Without MDA (AUC= 0.9253)   (b) With MDA (AUC = 0.9739)

Table 8 Comparison to State-of-the-art methods

Method	Sen	Spc	Prc	Acc	F1	MCC	FMI
SSVM [10]	83.82 ±2.07	86.53 ±1.80	86.18 ±1.51	85.17 ±1.05	84.96 ±1.13	70.41 ±2.09	84.98 ±1.13
AECNN [11]	87.57 ±2.06	88.33 ±2.47	88.30 ±2.09	87.95 ±1.37	87.91 ±1.35	75.95 ±2.72	87.92 ±1.34
DLAD [12]	91.39 ±1.40	88.06 ±2.55	88.51 ±2.05	89.72 ±0.82	89.90 ±0.68	79.54 ±1.57	89.92 ±0.66
RDL [13]	88.75 ±1.42	91.60 ±1.48	91.38 ±1.32	90.17 ±0.61	90.03 ±0.63	80.40 ±1.23	90.05 ±0.63
CAS [14]	88.47 ±1.40	91.32 ±1.19	91.08 ±1.07	89.90 ±0.68	89.75 ±0.72	79.84 ±1.36	89.76 ±0.71
PZM-DSSAE (Ours)	93.33 ±1.47	93.13 ±0.95	93.15 ±0.89	93.23 ±0.81	93.23 ±0.83	86.47 ±1.62	93.24 ±0.83

Fig. 13 Error bar of SPT identification methods
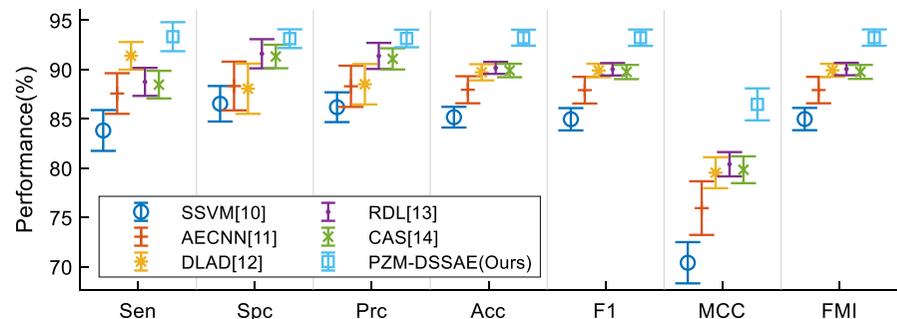
In the future, we shall try to propose automatic algorithms to determine this AI model's optimal order and structure. Besides, we shall attempt to collect more data and verify our model in a stricter medical environment.

Data Availability The data is available upon reasonable request to the corresponding authors.

Declarations

Conflict of Interest We have no conflicts of interest to disclose concerning this paper.

References

1. Willie, B., Hakim, A.J., Badman, S.G., Weikum, D., Narokobi, R., Coy, K., et al.: High prevalence of pulmonary tuberculosis among female sex workers, men who have sex with men, and transgender women in Papua New Guinea. Trop Med Health. **49**, 6, Article ID: 4 (2021)
2. Lohiya, S., Tripathy, J.P., Sagili, K., Khanna, V., Kumar, R., Ojha, A., et al.: Does Drug-Resistant Extrapulmonary Tuberculosis Hinder TB Elimination Plans? A Case from Delhi, India. Trop Med Infect Dis. **5**, 13, Article ID: 109 (2020)
3. Almeida, S.R.P., Bastos, F.Z., Barussi, F.C.M., Lessa, D.A.B., Alencar, N.X., Michelotto, P.V.: Airway endoscopy and tracheal cytology of two-year-old thoroughbred horses during the first year of race training. Comp Exer Physiol. **14**, 143–148 (2018)
4. Akhtar, A.N., Ahmad, M.S., Khokhar, M.I., Afzal, M.F.: Early experience of laparoscopy in emergency operation theatre at Lahore General Hospital, Lahore. Pak J Med Health Sci. **11**, 1291–1292 (2017)
5. Gaubert, S., Blet, A., Dib, F., Ceccaldi, P.F., Brock, T., Calixte, M., et al.: Positive effects of lumbar puncture simulation training for medical students in clinical practice. BMC Med Educ. **21**, 6, Article ID: 18 (2021)

6. Arulprakash, N., Narayanan, L., Narayanan, S.: A young patient with stroke and primary tuberculosis. Journal of Neurosciences in Rural Practice. **9**, 613–615 (2018)

7. Li, X.W., Li, X.H., Liu, Q.L., Sun, N., Zhang, B., Shi, C.D., et al.: Traditional Chinese medicine combined with western medicine for the treatment of secondary pulmonary tuberculosis A PRISMA-compliant meta-analysis. Medicine. **99**, 10, Article ID: e19567 (2020)

8. Iliyasu, G., Mohammad, A.B., Yakasai, A.M., Dayyab, F.M., Oduh, J., Habib, A.G.: Gram-negative bacilli are a major cause of secondary pneumonia in patients with pulmonary tuberculosis: evidence from a cross-sectional study in a tertiary hospital in Nigeria. Trans. R. Soc. Trop. Med. Hyg. **112**, 252–254 (2018)

9. Rai, D.K., Alok: Clinico-radiological Difference between Primary and Secondary MDR Pulmonary Tuberculosis. J. Clin. Diagn. Res. **13**, OC08–OC010 (2019)

10. Bagci, U., Kubler, A., Luna, B., Jain, S., Bishai, W.R., Mollura, D.J.: Computer-aided detection and quantification of cavitary tuberculosis from CT scans. Med. Phys. **40**, 14, Article ID: 113701 (2013)

11. Li, L.J., Huang, H.Y., Jin, X.Y.: AE-CNN Classification of Pulmonary Tuberculosis Based on CT Images. In: Ninth International Conference on Information Technology in Medicine and Education, pp. 39–42. Hangzhou, China (2018)

12. Park, S., Jin, K.N., Kim, J.I., Choi, S.Y., Lee, J.H., Goo, J.M., et al.: Development and validation of a deep learning-based automatic detection algorithm for active pulmonary tuberculosis on chest radiographs. Clin. Infect. Dis. **69**, 739–747 (2019)

13. James-Reynolds, C., Currie, E., Gao, X.H.W.: Analysis of tuberculosis severity levels from CT pulmonary images based on enhanced residual deep learning architecture. Neurocomputing. **392**, 233–244 (2020)

14. Xie, Y.L., Wu, Z.Y., Han, X., Wang, H.Y., Wu, Y.F., Cui, L., et al.: Computer-Aided System for the Detection of Multicategory Pulmonary Tuberculosis in Radiographs. J Healthc Eng. 12, Article ID: 9205082 (2020, 2020)

15. Zhang, Y.-D., Nayak, D.R., Zhang, X., Wang, S.-H.: Diagnosis of secondary pulmonary tuberculosis by an eight-layer improved convolutional neural network with stochastic pooling and hyperparameter optimization. J. Ambient. Intell. Humaniz. Comput. (2020). https://doi.org/10.1007/s12652-020-02612-9

16. Govindaraj, V.: Explainable diagnosis of secondary pulmonary tuberculosis by graph rank-based average pooling neural network. J. Ambient. Intell. Humaniz. Comput. (2021). https://doi.org/10.1007/s12652-021-02998-0

17. Hu, M.-K.: Visual pattern recognition by moment invariants. IRE Transactions on Information Theory. **8**, 179–187 (1962)

18. Teague, M.R.: Image analysis via the general theory of moments. J. Opt. Soc. Am. **70**, 920–930 (1980)

19. Kar, A., Pramanik, S., Chakraborty, A., Bhattacharjee, D., Ho, E.S.L., Shum, H.P.H.: LMZMPM: local modified Zernike moment per-unit mass for robust human face recognition. IEEE Transactions on Information Forensics and Security. **16**, 495–509 (2021)

20. Singh, S.P., Urooj, S.: An Improved CAD System for Breast Cancer Diagnosis Based on Generalized Pseudo-Zernike Moment and Ada-DEWNN Classifier. J. Med. Syst. **40**, Article ID: 105, (2016)

21. Jiang, Y.: Exploring a smart pathological brain detection method on pseudo Zernike moment. Multimed. Tools Appl. **77**, 22589–22604 (2018)

22. Chong, C.W., Raveendran, P., Mukundan, R.: The scale invariants of pseudo-Zernike moments. Pattern. Anal. Applic. **6**, 176–184 (2003)

23. Fricker, P.: *Pseudo-Zernike Functions* [MATLAB Central File Exchange]. Available: https://www.mathworks.com/matlabcentral/fileexchange/33644-pseudo-zernike-functions (2021)

24. Risco, S., Molto, G., Naranjo, D.M., Blanquer, I.: Serverless Workflows for Containerised Applications in the Cloud Continuum. Journal of Grid Computing. **19**, Article ID: 30, (2021)

25. Darabian, H., Homayounoot, S., Dehghantanha, A., Hashemi, S., Karimipour, H., Parizi, R.M., et al.: Detecting Cryptomining malware: a deep learning approach for static and dynamic analysis. Journal of Grid Computing. **18**, 293–303 (2020)

26. Emdadi, A., Eslahchi, C.: Auto-HMM-LMF: feature selection based method for prediction of drug response via autoencoder and hidden Markov model. BMC Bioinformatics. **22**, 22, Article ID: 33 (2021)

27. Alessandri, L., Cordero, F., Beccuti, M., Licheri, N., Arigoni, M., Olivero, M., et al.: Sparsely-connected autoencoder (SCA) for single cell RNAseq data mining. NPJ Systems Biology and Applications. **7**, 10, Article ID: 1 (2021)

28. Tivive, F.H.C., Bouzerdoum, A.: Clutter removal in Through-The-Wall radar imaging using sparse autoencoder with low-rank projection. IEEE Trans. Geosci. Remote Sens. **59**, 1118–1129 (2021)

29. Nguyen, C.D., Prosvirin, A.E., Kim, C.H., Kim, J.M.: Construction of a Sensitive and Speed Invariant Gearbox Fault Diagnosis Model Using an Incorporated Utilizing Adaptive Noise Control and a Stacked Sparse Autoencoder-Based Deep Neural Network. Sensors. **21**, 23, Article ID: 18 (2021)

30. Benyamin, M., Genish, H., Califa, R., Wolbromsky, L., Ganani, M., Wang, Z., et al.: Autoencoder based blind source separation for photoacoustic resolution enhancement. Sci. Rep. **10**, 7, Article ID: 21414 (2020)

31. Jelodar, R.A., Mehri-Dehnavi, H., Agahi, H.: Some properties of Tsallis and Tsallis-Lin quantum relative entropies. Physica A-Statistical Mechanics and Its Applications. **567**, 7, Article ID: 125719 (2021)

32. Abbaspour-Gilandeh, Y., Fazeli, M., Roshanianfard, A., Hernandez-Hernandez, M., Gallardo-Bernal, I., Hernandez-Hernandez, J.L.: Prediction of Draft Force of a Chisel Cultivator Using Artificial Neural Networks and Its Comparison with Regression Model. Agronomy-Basel. **10**, 14, Article ID: 451 (2020)

33. Winzer, R., Martin, R., Kuhn, J.P., Baldus, J.C., Seppelt, D., Heidrich, F.M., et al.: Adrenal glands enhancement in computed tomography as predictor of short-and intermediate term mortality in critically ill patients. Clin. Imaging. **70**, 56–60 (2021)

34. Jena, R., Pradhan, B., Alamri, A.M.: Susceptibility to Seismic Amplification and Earthquake Probability Estimation Using Recurrent Neural Network (RNN) Model in Odisha, India. Applied Sciences-Basel. **10**, 18, Article ID: 5355 (2020)

35. Shekter, D.H., Samuelson, F.W.: Efficiently calculating ROC curves, AUC, and uncertainty from 2AFC studies with finite samples. Proc. SPIE. **11316**, (2020)

36. Wang, S.-H.: Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network. Information Fusion. **67**, 208–229 (2021)

37. Cheng, X.: PSSPNN: PatchShuffle Stochastic Pooling Neural Network for an Explainable Diagnosis of COVID-19 with Multiple-Way Data Augmentation. Comput Math Methods Med. 2021 (2021, Article ID: 6633755)

38. Loyola-Gonzalez, O., Medina-Perez, M.A., Choo, K.K.R.: A review of supervised classification based on contrast patterns: applications, trends, and challenges. Journal of Grid Computing. **18**, 797–845 (2020)

39. Gupta, A., Sahu, H., Nanecha, N., Kumar, P., Roy, P.P., Chang, V.: Enhancing text using emotion detected from EEG signals. Journal of Grid Computing. **17**, 325–340 (2019)