TBNet: a context-aware graph network for tuberculosis diagnosis

Si-Yuan Lu[1], Shui-Hua Wang[1,2], Xin Zhang[3,*], Yu-Dong Zhang[1,*],

1 School of Computing and Mathematical Sciences, University of Leicester, Leicester, LE1 7RH, UK
2 School of Architecture Building and Civil engineering, Loughborough University, Loughborough, LE11 3TU, UK
3 Department of Medical Imaging, The Fourth People's Hospital of Huai'an, Huai'an, Jiangsu Province, 223002, China

* Correspondence should be addressed to Xin Zhang & Yu-Dong Zhang
Email: Si-Yuan Lu (sl672@le.ac.uk), Shui-Hua Wang (shuihuawang@ieee.org), Xin Zhang (973306782@qq.com), Yu-Dong Zhang (yudongzhang@ieee.org)

Abstract: **Background and Objective:** Tuberculosis (TB) is an infectious bacterial disease. It can affect the human lungs, brain, bones, and kidneys. Pulmonary tuberculosis is the most common. This airborne bacterium can be transmitted with the droplets by coughing and sneezing. So far, the most convenient and effective method for diagnosing TB is through medical imaging. Computed tomography (CT) is the first choice for lung imaging in clinics because the conditions of the lungs can be interpreted from CT images. However, manual screening poses an enormous burden for radiologists, resulting in high inter-observer variances. Hence, developing computer-aided diagnosis systems to implement automatic TB diagnosis is an emergent and significant task for researchers and practitioners. This paper proposed a novel context-aware graph neural network called TBNet to detect TB from chest CT images. **Methods:** Traditional convolutional neural networks can extract high-level image features to achieve good classification performance on the ImageNet dataset. However, we observed that the spatial relationships between the feature vectors are beneficial for the classification because the feature vector may share some common characteristics with its neighboring feature vectors. To utilize this context information for the classification of chest CT images, we proposed to use a feature graph to generate context-aware features. Finally, a context-aware random vector functional-link net served as the classifier of the TBNet to identify these context-aware features as TB or normal. **Results:** The proposed TBNet produced state-of-the-art classification performance for detecting TB from healthy samples in the experiments. **Conclusions:** Our TBNet can be an accurate and effective verification tool for manual screening in clinical diagnosis.

Keywords: tuberculosis; computed tomography; computer-aided diagnosis; graph neural network; random vector functional-link net

1. Introduction

Tuberculosis is a highly infectious lung disease caused by Mycobacterium tuberculosis airborne bacteria. Tuberculosis is among the top causes of human death. Although the human immune system can kill most bacteria, there are approximately ten million new confirmed cases of TB and nearly two million death tolls by the disease every year around the world. The world health organization envisions eliminating TB by 2030 [1]. Early diagnosis is crucial to control the spreading of TB. The most widely used TB diagnosis method is computed tomography (CT). The infected regions can be observed in the chest CT images.

Nevertheless, manual screening requires expertise and long working hours, and the results of different observers may be contradictory. On the other side, the development of deep learning algorithms in the last decade has achieved outstanding performance on image classification and recognition tasks, making automatic medical image analysis possible. Therefore, computer-aided diagnosis (CAD) has become a heated research field. CAD systems can analyze the medical images and produce the results automatically and rapidly, which can serve as the verification to assist the radiologists in their diagnosis. In recent years, numerous CAD methods have been proposed to identify TB.

Hooda, Sofat, Kaur, Mittal and Meriaudeau [2] constructed a deep convolutional neural network (CNN) with seven convolutional layers and three fully connected layers to detect TB from chest X-ray images. The Adam algorithm was employed to train the CNN model. The proposed method yielded an accuracy of 82.09% on the validation set in the experiments. Lakhani and Sundaram [3] proposed to ensemble the pre-trained AlexNet and GoogLeNet to classify the chest radiographs as TB or healthy. They discovered that the pre-trained models performed better than un-pre-trained models. Liu, Cao, Alcantara, Liu, Brunette, Peinado and Curioso [4] also leveraged AlexNet and GoogLeNet to design their TX-CNN by transfer learning. Cross-validation and shuffle sampling were employed to evaluate the classification performance of their method. The proposed method achieved an accuracy of 85.68% in detecting TB from normal samples. Abbas and Abdelsamea [5] transferred a pre-trained AlexNet to identify TB and healthy samples from the chest X-ray images. In their experiments, they tried to fine-tune different layers of the AlexNet to obtain the best accuracy. Dinesh Jackson Samuel and Rajesh Kanna [6] exploited the pre-trained Inception V3 as the backbone and trained it on microscopic images. Then, the Inception V3 was used as the feature extractor to obtain image features from the microscopic images. Finally, a support vector machine (SVM) was trained with these features for the identification of TB. Shabut, Hoque Tania, Lwin, Evans, Yusof, Abu-Hassan and Hossain [1] put forward a TB diagnosis method

based on plasmonic ELISA images. They computed the color moments from those images as the features and selected the random forest as the classification algorithm. Their CAD system can be deployed on mobile platforms. Vajda, Karargyris, Jaeger, Santosh, Candemir, Xue, Antani and Thoma [7] firstly employed an atlas-based segmentation algorithm to obtain regions of interest (ROIs) from the chest X-ray images. Then, they extracted three different sets of image features: shape, edge, moment, texture, size, orientation, eccentricity, bounding box, etc. Afterward, a wrapper feature selection method was used to obtain the best image features from the three feature sets. Finally, they trained a multi-layer perceptron was trained to classify the samples as TB or normal. Xiong, Ba, Hou, Zhang, Chen and Li [8] constructed a CNN model called TB-AI and pre-trained it on the CIFAR-10 dataset. The pre-trained TB-AI was transferred to the classification of microscopic images to detect TB. Diaz-Huerta, Tellez-Anguiano, Fraga-Aguilar, Gutierrez-Gnecchi and Arellano-Calderon [9] proposed a segmentation method for bacilloscopies to diagnose TB. The segmentation was based on a Bayesian classification algorithm with a Gaussian mixture. Hwang, Park, Jin, Kim, Choi, Lee, Goo, Aum, Yim, Park, Deep Learning-Based Automatic Detection Algorithm and Evaluation [10] added 12 residual connections in their CNN model with 27 layers. The CNN was trained from scratch on their chest radiograph dataset and achieved satisfactory results. Lopez-Garnier, Sheen and Zimic [11] collected a dataset of microscopic observed drug susceptibility images for TB diagnosis. A 15-layer CNN model inspired by VGG16 was trained from scratch on their dataset for TB classification. Nguyen, Nguyen, Dao, Unnikrishnan, Dhingra and Ravichandran [12] tested the backbone models pre-trained on different datasets such as the ImageNet dataset and NIH-14 dataset. They discovered that the models pre-trained on the NIH-14 dataset can generate better low-level features than those pre-trained on the ImageNet dataset for chest X-ray images. Norval, Wang and Sun [13] tested different pre-processing methods for the classification of lung X-ray images including histogram equalization, contrast enhancement, sharpening, color channel reduction, and cropping of ROIs. They trained a CNN model to compare the classification performance of those pre-processing methods and found that the combination of contrast enhancement and cropping of ROIs produced the best results. Qin, Sander, Rai, Titahong, Sudrungrot, Laah, Adhikari, Carter, Puri, Codlin and Creswell [14] compared and discussed the classification performance of three deep learning-based TB diagnosis systems. Chandra, Verma, Singh, Jain and Netam [15] proposed a two-level hierarchical feature (HF) structure to generate features from chest X-ray images. In the first level, they computed the geometric features, and in the second level, they generated statistical features from the images. The SVM classification algorithm was chosen as the identifier of TB. Tao Hwa, Bade, Hijazi and Saffree Jeffree [16] combined deep learning models with contrast-enhanced Canny edge detector for TB diagnosis in chest X-ray images. They used the Canny edge detector to get the edges in the X-ray images. Based on the edges, they obtained edge-enhanced images and edge-detected images. Then, they trained VGG16 and InceptionV3 on these two types of images. The final predictions were generated using the ensemble of the CNN models. Ul Abideen, Ghafoor, Munir, Saqib, Ullah, Zia, Tariq, Ahmed and Zahra [17] proposed a Bayesian-based CNN (B-CNN) to detect TB and non-TB samples in the chest X-ray images. Govindarajan and Swaminathan [18] employed extreme learning machine (ELM) and online sequential ELM (OSELM) to classify the chest radiographs as TB and normal. The drawbacks of these existing methods were presented in Table 1.

Table 1: drawbacks of the state-of-the-art methods

Methods	Drawbacks
Hooda, Sofat, Kaur, Mittal and Meriaudeau [2]	The validation accuracy was low, which was just over 80%.
Lakhani and Sundaram [3]	The classification performance can be improved with more advanced backbones.
Liu, Cao, Alcantara, Liu, Brunette, Peinado and Curioso [4]	The validation accuracy was low, which was just over 85%.
Abbas and Abdelsamea [5]	The classification performance can be improved with more advanced backbones.
Dinesh Jackson Samuel and Rajesh Kanna [6]	The performance for different folds fluctuated in a wide range.
Shabut, Hoque Tania, Lwin, Evans, Yusof, Abu-Hassan and Hossain [1]	Handcrafted image features were not optimal, which can be replaced by features from a CNN model.
Vajda, Karargyris, Jaeger, Santosh, Candemir, Xue, Antani and Thoma [7]	Handcrafted image features were not optimal, which can be replaced by features from a CNN model.
Xiong, Ba, Hou, Zhang, Chen and Li [8]	The dataset was too small to train a deep CNN model.
Diaz-Huerta, Tellez-Anguiano, Fraga-Aguilar, Gutierrez-Gnecchi and Arellano-Calderon [9]	Their method was evaluated on a small dataset.
Hwang, Park, Jin, Kim, Choi, Lee, Goo, Aum, Yim, Park, Deep Learning-Based Automatic Detection Algorithm and Evaluation [10]	It's time-consuming to train a deep network from scratch.
Lopez-Garnier, Sheen and Zimic [11]	It's time-consuming to train a deep network from scratch.
Nguyen, Nguyen, Dao, Unnikrishnan, Dhingra and Ravichandran [12]	They trained the models for 100 epochs, which was time-consuming.
Norval, Wang and Sun [13]	The validation accuracy was low, which was just over 90%.
Qin, Sander, Rai, Titahong, Sudrungrot, Laah, Adhikari, Carter, Puri, Codlin and Creswell [14]	The dataset was class-imbalanced.
Chandra, Verma, Singh, Jain and Netam [15]	Handcrafted image features were not optimal, which can be replaced by features from a CNN model.
Tao Hwa, Bade, Hijazi and Saffree Jeffree [16]	Handcrafted image features were not optimal, which can be replaced by features from a CNN model.
Ul Abideen, Ghafoor, Munir, Saqib, Ullah, Zia, Tariq, Ahmed and Zahra [17]	It's time-consuming to train a deep network from scratch.
Govindarajan and Swaminathan [18]	Handcrafted image features were not optimal, which can be replaced by features from a CNN model.

From these recently published TB diagnosis methods, it can be found that deep CNN models are prevalent in medical image analysis, and transfer learning is often used to transfer the pre-trained weights on medical image datasets. However, these CAD systems can be improved to obtain better classification performance. The medical image datasets are often much smaller than the ImageNet dataset, so it may cause an overfitting problem if the pre-trained CNNs are fine-tuned for classification on the medical image datasets. In addition, the relationships between the samples in the latent feature space can be beneficial for the classification, but they are neglected.

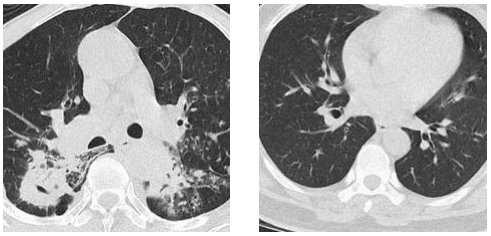
Therefore, in this work, we proposed a new TB diagnosis method called TBNet, based on the context-aware graph neural network. We employed EfficientNet as the backbone model to generate the sample-level features (SLFs). To utilize the context information of these features, we extracted novel context-aware features (CAFs) extracted based on the nearest neighbors of the SLFs in the feature space. Finally, the context-aware random vector functional-link net (CARVFLN) was trained as the classifier for TB diagnosis. Extensive experiments were implemented to evaluate the classification performance of our TBNet, and the results suggested that the TBNet was effective and accurate for the classification of TB based on chest CT images.

The remainder of this study is organized as follows. The information of the chest CT dataset in the experiments is presented in Section 2. The methodology of the proposed TBNet is provided in Section 3. Section 4 is about the experimental results, and discussion is presented in Section 5. The conclusion of this paper is given in Section 6.

2. Materials

We obtained two types of chest CT images: tuberculosis and healthy control (HC) from the Fourth Hospital of Huai'an. All the CT images were generated with a Philips spiral CT machine, and the resolution of the images was 1024 × 1024 × 3 pixels. We enrolled a total of 840 chest images to form our dataset in which half of the images are TB, and the other half are healthy samples. The ground-truth labels of these images were generated with the consensus of three radiologists.

The original chest CT images contained excessive information not related to the diagnosis of TB, such as texts and backgrounds. In addition, the grayscale and the contrast of the images can be improved. Hence, we employed a set of pre-processing operations to improve the quality of these CT images. Firstly, we converted the original CT images into grayscale images. Then, a histogram stretching algorithm was performed on the images to improve the contrast. Afterward, a margin cropping operation was employed to remove the texts and backgrounds in the images. Finally, the images were resized to meet the input resolutions of the backbone CNN models, which are often 224 × 224 × 3 pixels or 227 × 227 × 3. Two samples of both labels are shown in Figure 1.



(a) Tuberculosis (b) Healthy control

Figure 1: chest CT images in our dataset

3. Methodology

We proposed a novel CAD method called TBNet to classify TB in the chest CT images. The diagram of the proposed TBNet is shown in Figure 2. Initially, CNN models are the cutting-edge methods for feature extraction as they can automatically generate high-level image representations. However, training a deep CNN model from scratch requires much memory and time.

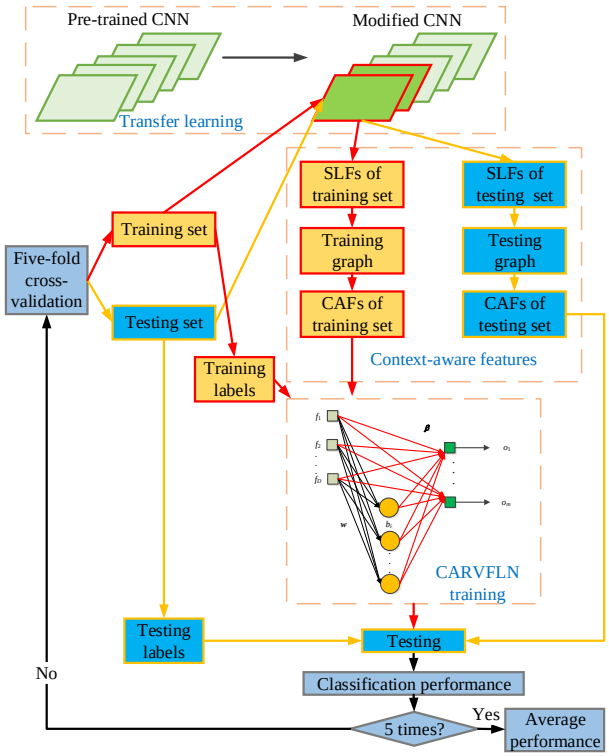


Figure 2: diagram of our TBNet (TBNet includes three main parts: transfer learning, context-aware feature extraction, and CARVFLN training.)

In addition, big labeled image datasets such as the ImageNet dataset are also needed to train the CNN models to avoid the overfitting problem but labeled medical image datasets are usually small. Hence, transfer learning was employed in our TBNet to extract sample-level features (SLFs), and the EfficientNet was used as the backbone model pre-trained on the ImageNet dataset. Then, based on the obtained SLFs, a graph was built in the SLF space by k

nearest neighbors. The SLF vectors are connected with their nearest neighbors in the feature graph. Consequently, the context information of the SLF vectors was fused into the CAFs so that can be used for classification. Finally, a context-aware random vector functional-link net (CARVFLN) was embedded in the TBNet to serve as the classifier to identify the CAFs as TB or healthy. Five-fold cross-validation was employed to evaluate the classification performance of our TBNet, and the gradient-weighted class activation mapping (Grad-CAM) was used to explain the TBNet predictions.

3.1. Sample-level features (SLFs)

Feature extraction is a necessary and significant procedure in computer vision tasks because the volumes of the images are usually too high with excessive information. The distribution of the features in the latent space directly determines the complexity of the image classification problem. Handcrafted SLFs were usually used in the CAD methods over ten years ago, such as wavelet features [19-21] before the prevalence of CNN models. As the AlexNet won the famous ImageNet competition, CNN models have become the cutting-edge solution for image classification and segmentation problems.

Because these deep models can automatically learn high-level representations from the input images so that the researchers are free from developing handcrafted SLFs, it is time-consuming to train the deep CNNs. It also requires dedicated GPU and big image datasets during the training. Hence, transfer learning is a more feasible method in specific applications than training from scratch.

In this paper, transfer learning was used only to generate the SLFs instead of classifying the chest CT images. We wanted to transfer the pre-trained deep CNN model's powerful image representation learning ability to extract effective SLFs from chest CT images by transfer learning. The backbone model was the EfficientNet [22], pre-trained on the ImageNet dataset. The structure of the EfficientNet was determined by architecture search. To extract SLFs, the EfficientNet was modified and fine-tuned on our chest CT dataset because of the difference between the ImageNet dataset and our chest CT dataset. The original pre-trained EfficientNet was designed to distinguish 1,000 categories of images, so there were 1,000 nodes in its output layer.

Nevertheless, in our dataset, there were only 2 categories of images: TB and healthy. In addition, the fully connected layers in the pre-trained EfficientNet served as the classifier, so we considered that they could not contribute to the generation of SLFs. Therefore, we removed the top three layers in the pre-trained EfficientNet and added five new layers to serve as the classifier during the fine-tuning of our chest CT images. The modifications are presented in Figure 3.

All the parameters in the pre-trained EfficientNet were preserved as the initial values except those in the last three layers. We proposed to insert a set of buffer layers: the 'fully connected_1' with 256 nodes and 'ReLU', between the 'global average pooling' layer and the 'fully connected_2' layer with two nodes because there were 1,000 nodes in the original pre-trained EfficientNet. The buffer layers can smooth the reduction procedures of the dimensions. After the fine-tuning, the SLFs can be computed in the feature layer 'fully connected_1' so that the dimension of the SLFs is 256. The 'fully connected _1' was selected as the feature layer because it was the beginning of the classifier and the end of feature learning.

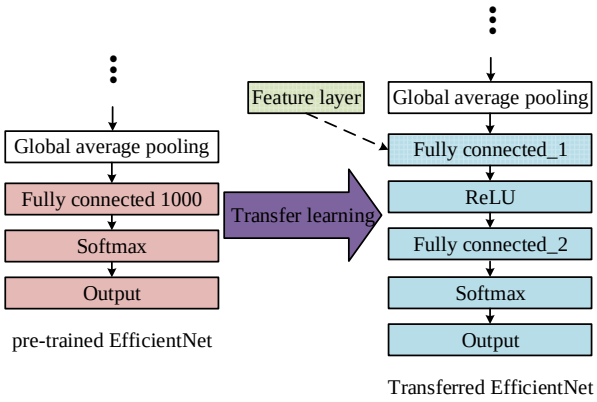


Figure 3: transfer learning by EfficientNet ('fully connected_1' was the feature layer.)

3.2. Context-aware features (CAFs)

The context information in the SLFs can be beneficial for the classification because good representation generation can ensure that the SLFs of the same labels are often distributed close to each other in the latent feature space.

To exploit the context information in these SLFs, we proposed to generate an SLF graph and fuse the context information into the SLFs. Specifically, the SLF graph was computed based on the k nearest neighbors (k-NN) algorithm, in which the distances between the SLFs were measured by Euclidean distance. Each SLF was regarded as a node and connected with its k nearest nodes in the graph. In this way, the context information was converted as the connections in the SLF graph. Given the SLF set as

$$\mathbf{F_{SL}} = [f_1, f_2, f_3, \cdots, f_N]^T \in \mathbb{R}^{N \times D} \quad (1)$$

Where N and D represent the number of samples in the SLF set and the dimension of each SLF, respectively, the distance matrix **Dis** between the SLFs and the matrix of adjacent nodes **Adj** can be calculated with the following two equations:

$$\mathbf{Dis}(m,n) = \|f_m - f_n\|, 1 \le m,n \le N \,\&\, m \ne n \quad (2)$$

$$\mathbf{Adj}(m,n) = 1, \text{ if } f_n \in knn(f_m), 1 \le m,n \le N \,\&\, m \ne n \quad (3)$$

Where the $knn(f_m)$ function produced the k nearest neighbors of the SLF f_m, and both **Dis** and **Adj** were initialized with zero values. Then, the normalized **Adj** can be obtained by:

$$\widehat{\mathbf{Adj}} = \mathbf{Deg}^{-\frac{1}{2}}(\mathbf{Adj} + \mathbf{I})\mathbf{Deg}^{-\frac{1}{2}} \quad (4)$$

Where the matrix **I** represents the identity matrix and the degree matrix **Deg** is expressed as:

$$\mathbf{Deg}(m,n) = \begin{cases} k, \text{if } m = n \\ 0, \text{if } m \neq n \end{cases}, \ 1 \leq m,n \leq N \quad (5)$$

Consequently, the context-aware feature set $\mathbf{F}_{CA}$ can be calculated as:

$$\mathbf{F}_{CA} = \widehat{\mathbf{Adj}} \cdot \mathbf{F}_{SL} \quad (6)$$

The diagram of the CAF generation is provided in Figure 4. First of all, an EfficientNet pre-trained on the ImageNet dataset was modified and trained on our chest CT images. Then, the SLFs can be obtained from the feature layer in the transferred EfficientNet. Afterward, the SLF graph was constructed to store the context information of the SLFs in the latent feature space using the k-NN algorithm. Finally, the CAFs can be generated with the graph and the SLFs.

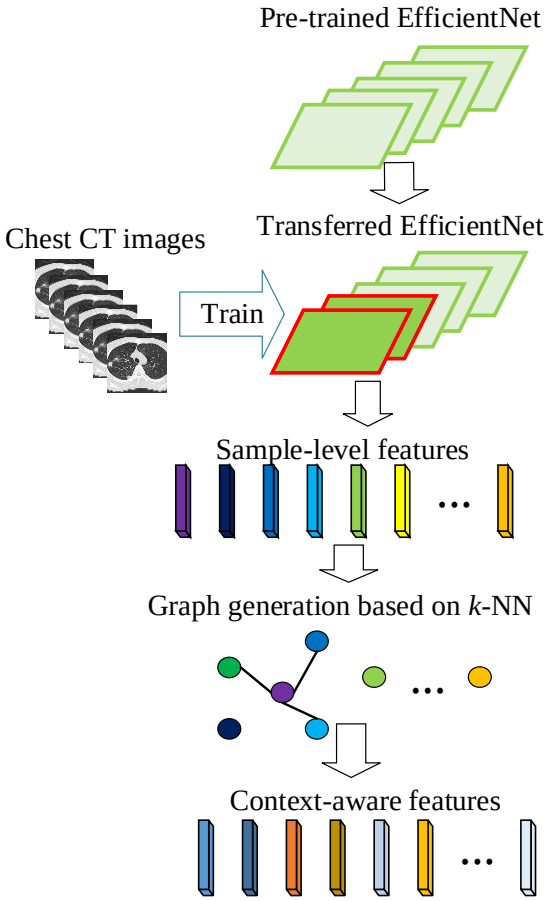


Figure 4: diagram to compute CAFs (k-NN algorithm was employed to generate the graph for sample-level feature vectors.)

3.3. Context-aware random vector functional-link net (CARVFLN)

For the classification of the CAFs, a CARVFLN was proposed as the classifier of our TBNet based on the traditional random vector functional-link (RVFL) [23-25]. The differences between the RVFL and conventional back-propagation neural networks (BPNNs) are mainly two-fold. Firstly, there is an extra shortcut between the input layer and the output layer in the RVFL, while traditional BPNNs don't have such links. Secondly, the training of an RVFL is much faster than BPNNs [26, 27]. Because the weights and biases in the random mapping layer are randomly initialized and remain fixed during the entire training process, the output weights of the RVFL can be obtained by pseudo-inverse.

The difference between the CARVFLN and the RVFL lies only in that the input of the CARVFLN is the CAF instead of conventional SLFs. The structure of the CARVFLN is presented in Figure 2.

In total, there are three parameters in the CARVFLN to be trained: the hidden mapping weights $\boldsymbol{w} = [w_1, w_2, w_3, \ldots, w_{\widehat{N}}]^T$, hidden mapping biases $b = [b_1, b_2, b_3, \ldots, b_{\widehat{N}}]^T$, and the output weights $\boldsymbol{\beta}$. $\widehat{N}$ represents the dimension of the hidden space. The training algorithm of the CARVFLN consists of three steps. First, the weights and biases from the input to the hidden mapping space are assigned with random values. Then, the output matrix M of the hidden mapping layer can be computed with the training CAFs:

$$\mathbf{M} = \sum_{i=1}^{\widehat{N}} f\left(w_i f_j + b_i\right), j = 1, \ldots, N, \quad (7)$$

where $f(x)$ represents the activation function. The obtained **M** is concatenated with the original CAFs:

$$\mathbf{F} = \text{concat}(\mathbf{F}_{CA}, \mathbf{M}) \quad (8)$$

Finally, the output weights can be calculated by the pseudo-inverse:

$$\boldsymbol{\beta} = \mathbf{F}^{\dagger}\mathrm{T}, \quad (9)$$

where the T denotes the ground-truth labels of the training samples. In this way, all the parameters in the CARVFLN are determined. It can be found that there is no iteration in the training algorithm of CARVFLN, which is often employed in gradient descent methods. Therefore, the CARVFLN can converge much faster than BPNNs, and we utilize it as the classifier of the TBNet to produce the predictions of chest CT images.

4. Results

The proposed TBNet was implemented and trained on a laptop on MATLAB 2021a with an i7 7700H processor and a GTX1060 graphic card. The EfficientNet B0 was selected as the backbone model in the proposed TBNet, considering the capacity of our experiment platform. All the statistical results were obtained using five-fold cross-validation. In the five-fold cross-validation, the entire chest CT dataset was evenly separated into five groups of samples. In every loop, one group served as the testing set, and the rest four groups served as the training set. The loop was repeated five times so that every data group was tested. The average performance can be obtained based on the results of the five runs.

4.1. Hyper-parameter settings

The hyper-parameters in our TBNet are listed in Table 2. The mini-batch size was set as 32, considering the memory size of our device. The value of max-epochs was set to be only one because it can result in overfitting if the EfficientNet is fine-tuned on the small chest CT dataset. The learning rate was 1e-4, the usual setting in most applications. The k denotes the number of neighbors in the CAFs, which was set as 4 considering the mini-batch size. The dimension of the hidden mapping layer in the CARVFLN was 400, as the dimension of the input CAFs was 256. It is beneficial for the classification to map the CAFs into higher dimensions randomly.

Table 2: hyper-parameter settings

Hyper-parameter	Value
Mini-batch size	32
Max-epochs	1
Initial learning rate	1e-4
K	4
$\hat{N}$	400

4.2. Performance of the proposed TBNet

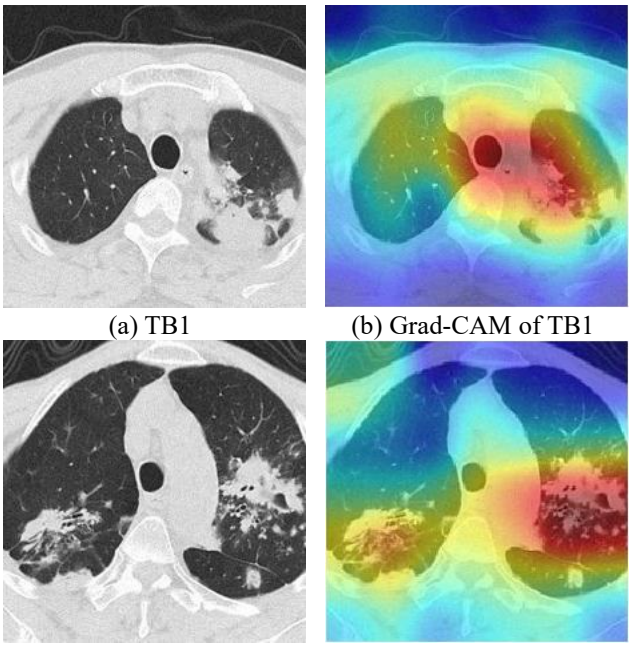
Five common measurements were employed to evaluate the classification performance of the proposed TBNet, including accuracy, sensitivity, specificity, precision, and F1 score based on five-fold cross-validation. The results are presented in Table 3. It can be observed that the proposed TBNet achieved the perfect sensitivity of 100.00% for all five folds. The average accuracy, specificity, and precision were over 97%, and the average F1 score was near 98.91%. Meanwhile, the fluctuation range of all the five indicators was around 5% in different folds of images, which suggested the robustness of our TBNet.

Table 3: classification performance of the proposed TBNet based on five-fold cross-validation

	Accuracy	Sensitivity	Specificity	Precision	F1 score
Fold 1	99.40%	100.00%	98.82%	98.81%	99.40%
Fold 2	97.02%	100.00%	94.38%	94.05%	96.93%
Fold 3	100.00%	100.00%	100.00%	100.00%	100.00%
Fold 4	98.81%	100.00%	97.67%	97.62%	98.80%
Fold 5	99.40%	100.00%	98.82%	98.81%	99.40%
Average	98.93%	100.00%	97.94%	97.86%	98.91%

4.3. Grad-CAM of the TBNet

The explanation of deep learning models is another hot issue in artificial intelligence recently because it remains a problem why these CNN models can produce accurate predictions. The CNN models are like the black box with only the input and output. Researchers try to interpret the working mechanism of the black box. The Grad-CAM provides a visual explanation of why CNN produces such results. The regions where the CNN model is concentrating on are highlighted in the Grad-CAM map, while the other areas are less related to the prediction. We provided two Grad-CAM maps of TB samples in Figure 5. The original chest CT images were on the left, while the Grad-CAM maps were on the right. The regions in red and yellow colors indicated the concentrations of the TBNet, while the regions in blue color were less important for making the predictions.



(a) TB1 (b) Grad-CAM of TB1

(c) TB2 (d) Grad-CAM of TB2

Figure 5: Grad-CAM maps of two TB samples by TBNet

4.4. Effects of backbone models in the TBNet

The pre-trained EfficientNet was selected as the backbone model for the proposed TBNet. However, there are numerous state-of-the-art pre-trained CNN backbone models to choose from, such as AlexNet, ResNet, VGG, DenseNet, and MobileNet.

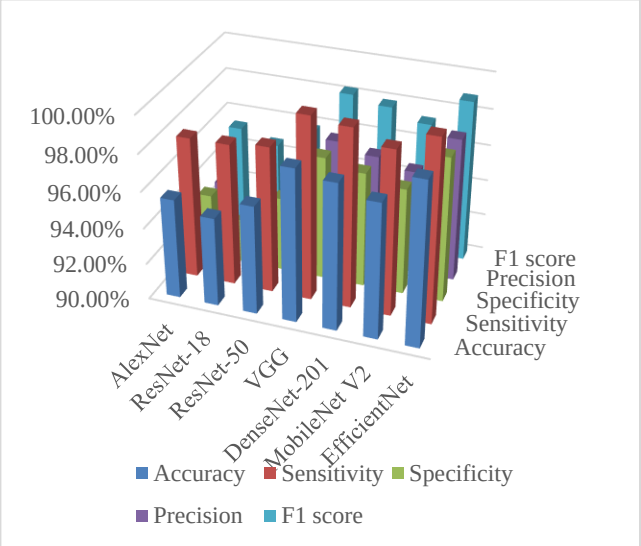


Figure 6: comparison of the TBNets performance using different backbones

Although the classification performance of these deep CNNs varies on the ImageNet dataset, there are no general rules for selecting the optimal backbone model for a specific image classification problem. Hence, we experimented with testing the performance of the proposed TBNet using different backbone models.

All the backbone models were modified with the same operation shown in Figure 3, and all the hyper-parameters were the same during the entire training and testing. The results of the TBNet using pre-trained AlexNet, ResNet-18, ResNet-50, VGG-16, DenseNet-201, and MobileNetV2 were listed in Table 4, and the comparison of the results was demonstrated in Figure 6.

Table 4: Performance of the TBNet with different transferred backbone models by five-fold cross-validation

Backbone model		Accuracy	Sensitivity	Specificity	Precision	F1 score
AlexNet	Fold 1	92.26%	97.33%	88.17%	86.90%	91.82%
	Fold 2	97.02%	98.77%	95.40%	95.24%	96.97%
	Fold 3	95.24%	96.34%	94.19%	94.05%	95.18%
	Fold 4	96.43%	98.75%	94.32%	94.05%	96.34%
	Fold 5	96.43%	97.56%	95.35%	95.24%	96.39%
	Average	95.48%	97.75%	93.49%	93.10%	95.34%
ResNet-18	Fold 1	92.86%	93.90%	91.86%	91.67%	92.77%
	Fold 2	94.05%	96.25%	92.05%	91.67%	93.90%
	Fold 3	98.81%	100.00%	97.67%	97.62%	98.80%
	Fold 4	94.05%	100.00%	89.36%	88.10%	93.67%
	Fold 5	94.64%	98.70%	91.21%	90.48%	94.41%
	Average	94.88%	97.77%	92.43%	91.90%	94.71%
ResNet-50	Fold 1	95.24%	98.72%	92.22%	91.67%	95.06%
	Fold 2	93.45%	95.06%	91.95%	91.67%	93.33%
	Fold 3	98.21%	98.80%	97.65%	97.62%	98.20%
	Fold 4	96.43%	98.75%	94.32%	94.05%	96.34%
	Fold 5	96.43%	98.75%	94.32%	94.05%	96.34%
	Average	95.95%	98.01%	94.09%	93.81%	95.86%
VGG-16	Fold 1	98.21%	100.00%	96.55%	96.43%	98.18%
	Fold 2	98.81%	100.00%	97.67%	97.62%	98.80%
	Fold 3	98.81%	100.00%	97.67%	97.62%	98.80%
	Fold 4	97.02%	100.00%	94.38%	94.05%	96.93%
	Fold 5	98.81%	100.00%	97.67%	97.62%	98.80%
	Average	98.33%	100.00%	96.79%	96.67%	98.30%
DenseNet-201	Fold 1	98.81%	100.00%	97.67%	97.62%	98.80%
	Fold 2	98.21%	100.00%	96.55%	96.43%	98.18%
	Fold 3	98.21%	100.00%	96.55%	96.43%	98.18%
	Fold 4	96.43%	98.75%	94.32%	94.05%	96.34%
	Fold 5	98.21%	100.00%	96.55%	96.43%	98.18%
	Average	97.98%	99.75%	96.33%	96.19%	97.94%
MobileNet V2	Fold 1	98.81%	100.00%	97.67%	97.62%	98.80%
	Fold 2	98.21%	100.00%	96.55%	96.43%	98.18%
	Fold 3	97.02%	98.77%	95.40%	95.24%	96.97%
	Fold 4	94.64%	96.30%	93.10%	92.86%	94.55%
	Fold 5	98.21%	100.00%	96.55%	96.43%	98.18%
	Average	97.38%	99.01%	95.86%	95.71%	97.33%

4.5. Transfer learning versus training from scratch

Table 5: Performance of the TBNet with the EfficientNet trained from scratch based on five-fold cross-validation

	Accuracy	Sensitivity	Specificity	Precision	F1 score
Fold 1	77.98%	78.31%	77.65%	77.38%	77.84%
Fold 2	74.40%	87.27%	68.14%	57.14%	69.06%
Fold 3	74.40%	77.33%	72.04%	69.05%	72.96%
Fold 4	80.95%	78.26%	84.21%	85.71%	81.82%
Fold 5	84.52%	89.19%	80.85%	78.57%	83.54%
Average	78.45%	82.07%	76.58%	73.57%	77.05%

Deep CNN models can be used in automatic medical image analysis by either transfer learning or training from scratch. In transfer learning, the weights and biases in the pre-trained model can be re-used in the target datasets, contributing to faster convergence and representation generation. However, the pre-trained dataset and the target dataset can be very different. On the other side, training a

deep model from scratch is a more straight choice, but overfitting and the long training time posed new challenges, especially for small datasets. To compare the two methods, we experimented, and the results of the TBNet with the EfficientNet trained from scratch were given in Table 5. The architecture of the EfficientNet trained from scratch was the same as the pre-trained EfficientNet. The hyper-parameters for training the EfficientNet from scratch were the same as Table 2. The comparison is shown in Figure 7.

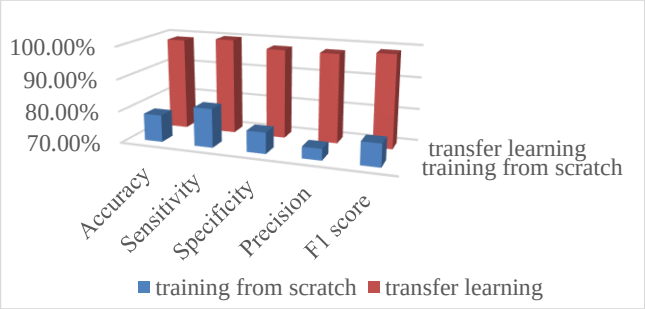


Figure 7: performance comparison of the TBNets with the transferred EfficientNet and the EfficientNet trained from scratch

4.6. Comparison with state-of-the-art methods

We compared the classification performance of the proposed TBNet with recently published state-of-the-art methods for the classification of TB, including Inception V3+SVM [6], random forests [1], HF+SVM [15], and OSELM [18]. The comparison results are listed in Table 6 and Figure 8.

This comparison revealed that the proposed TBNet achieved the best overall accuracy of 98.93% among the listed methods. Our TBNet and the random forests method both yielded 100%, but the handcrafted features used to train the random forests might be sub-optimal. OSELM attained the best performance for specificity and precision, but our TBNet was only marginally worse. Meanwhile, ten-fold cross-validation was used to evaluate the OSELM, which means that a higher proportion of images were used for training. At the same time, fewer samples were tested compared with five-fold cross-validation in the evaluation of TBNet. In addition, the TBNet also produced the highest F1-score of 98.91% in the listed methods.

Table 6: comparison with state-of-the-art TB classification approaches

Method	Accuracy	Sensitivity	Specificity	Precision	F1 score	Validation
Inception V3+SVM [6]	95.05%	-	-	-	-	Five-fold cross-validation
Random forests [1]	98.36%	**100.00%**	97.62%	95.00%	97.44%	Hold-out validation
HF+SVM [15]	95.60%	93%	-	96.90%	94.3%	Hold-out validation
OSELM [18]	-	98.7%	**98.7%**	**97.9%**	98.6%	Ten-fold cross-validation
TBNet (ours)	**98.93%**	**100.00%**	97.94%	97.86%	**98.91%**	Five-fold cross-validation

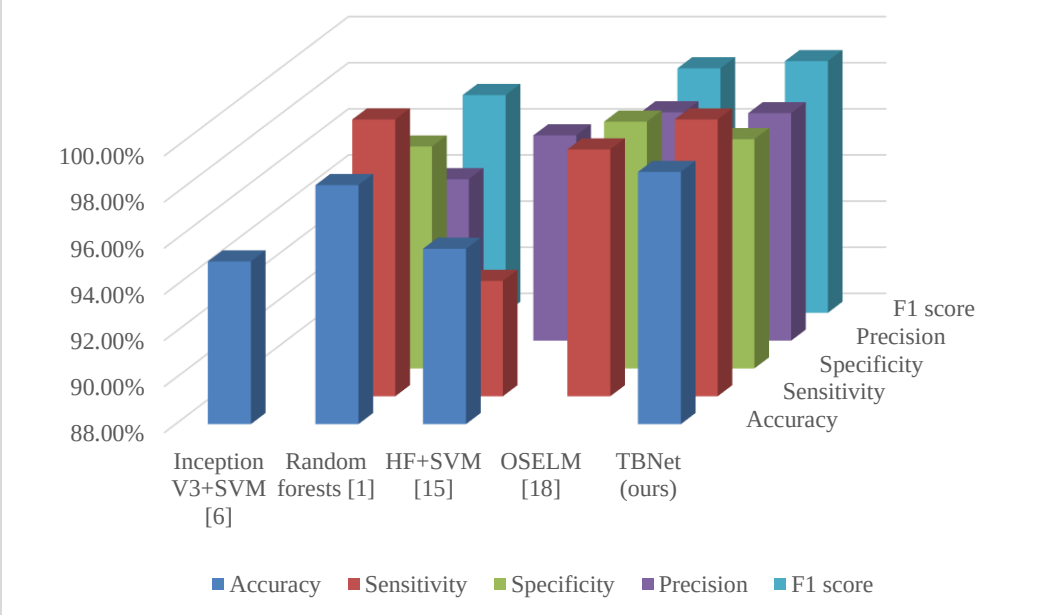


Figure 8: comparison with state-of-the-art approaches

5. Discussion

Sensitivity is an important indicator in clinical diagnosis because low sensitivity can result in the misdiagnosis of true positive cases. The proposed TBNet produced 100.00% sensitivity on five-fold cross-validation, revealing that our TBNet is an accurate and practical model to detect TB in CT scans.

The Grad-CAM maps in Figure 5 demonstrated that the TBNet could generally find the suspicious locations of TB. This observation suggested that our TBNet can make accurate predictions based on the CT scans and locate the potential lesion regions. Therefore, the TBNet can be further applied to CT segmentation in the future.

It can be seen from Figure 6 that the proposed TBNet produced over 94% accuracy with all the seven backbone models for five-fold cross-validation, and the performance of the TBNet with different backbones was close. The backbone VGG and EfficientNet performed the best among the seven models in terms of sensitivity. Considering all the five indicators, we obtained a slightly superior result with the EfficientNet. We speculated this result might be due to the architecture search algorithm for designing EfficientNet.

There showed overwhelming evidence in Figure 7 that the proposed TBNet worked substantially better with transferred EfficientNet than that with the EfficientNet trained from scratch. The gap between the two methods was around 20%. The present finding confirmed that the parameters learned in the pre-trained EfficientNet from the ImageNet were fundamental for the SLF generation, contributing to the good classification performance of our TBNet consequently.

The poor performance might be due to the small max-epochs when training the EfficientNet from scratch. Still, our TBNet worked very well with the same hyper-parameter settings by transfer learning, which both require approximately 650 seconds to finish the five-fold cross-validation. Hence, through this fair comparison, it can be revealed that transfer learning was the better option for utilizing the EfficientNet in our TBNet for the automatic diagnosis of TB in chest CT images.

The possible reasons behind the high performance of our TBNet were analyzed as follows. The good SLFs generated from the pre-trained EfficientNet provided the potential for high performance. The CAFs obtained using the fusion of SLFs and their context relationships in the latent feature space improved the discrimination ability of the image features. Finally, the CARVFLN served as the classifier in our TBNet, which was fast to train. The TBNet leveraged the representation learning ability from the EfficientNet and the excellent generalization ability of the CARVFLN on small datasets. Together, the proposed TBNet can be an effective and efficient tool to assist radiologists in the clinical diagnosis of TB in chest CT images.

6. Conclusion

In this work, we put forward a CAD method called TBNet to diagnose tuberculosis in the chest CT images. In the TBNet, we firstly extracted sample-level features from the images using a pre-trained EfficientNet. Then, we proposed to generate a feature graph based on the distribution of these sample-level features. Consequently, the context-aware features can be obtained based on the sample-level features and the graph. In this way, the context information was fused into these context-aware features. Considering the relatively small size of our chest CT image dataset, we proposed a CARVFLN as the classifier in the TBNet. The proposed TBNet achieved an overall accuracy of 98.93%, a sensitivity of 100.00%, a specificity of 97.94%, a precision of 97.86%, and an F1 score of 98.91% based on five-fold cross-validation. Extensive experiments were carried out, and the results showed that our TBNet was comparable to state-of-the-art methods in the diagnosis of tuberculosis from chest CT images.

For future research, we shall collect more images with more pulmonary diseases. Image-level fusion is another research direction because the imaging results of the same patient using different imaging machines can be different, so the fusion of different images can improve the classification performance. In addition, segmentation of the chest CT images to find the focuses of the diseases is also an important research topic.

Statements of ethical approval

This work does not contain any studies with human participants or animals performed by any authors.

Declaration of competing interests

The authors declare that they have no competing interests.

Funding

This paper is partially supported by Hope Foundation for Cancer Research, UK (RM60G0680), Royal Society International Exchanges Cost Share Award, UK (RP202G0230), Medical Research Council Confidence in Concept Award, UK (MC_PC_17171), British Heart Foundation Accelerator Award, UK (AA/18/3/34220); Sino-UK Industrial Fund, UK (RP202G0289); Global Challenges Research Fund (GCRF), UK (P202PF11). Siyuan Lu holds the CSC scholarship with the University of Leicester.

References

[1] A.M. Shabut, M. Hoque Tania, K.T. Lwin, B.A. Evans, N.A. Yusof, K.J. Abu-Hassan, M.A. Hossain, An intelligent mobile-enabled expert system for tuberculosis disease diagnosis in real time, Expert Systems with Applications, 114 (2018) 65-77.

[2] R. Hooda, S. Sofat, S. Kaur, A. Mittal, F. Meriaudeau, Deep-learning: A Potential Method for Tuberculosis Detection using Chest Radiography, 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), IEEE, Kuching, Malaysia, 2017, pp. 497-502.

[3] P. Lakhani, B. Sundaram, Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks, Radiology, 284 (2017) 574-582.
[4] C. Liu, Y. Cao, M. Alcantara, B. Liu, M. Brunette, J. Peinado, W. Curioso, TX-CNN: DETECTING TUBERCULOSIS IN CHEST X-RAY IMAGES USING CONVOLUTIONAL NEURAL NETWORK, 2017 IEEE International Conference on Image Processing (ICIP), IEEE, Beijing, China, 2017, pp. 2314-2318.
[5] A. Abbas, M.M. Abdelsamea, Learning Transformations for Automated Classification of Manifestation of Tuberculosis using Convolutional Neural Network, 2018 13th International Conference on Computer Engineering and Systems (ICCES), IEEE, Cairo, Egypt, 2018, pp. 122-126.
[6] R. Dinesh Jackson Samuel, B. Rajesh Kanna, Tuberculosis (TB) detection system using deep neural networks, Neural Computing and Applications, 31 (2018) 1533-1545.
[7] S. Vajda, A. Karargyris, S. Jaeger, K.C. Santosh, S. Candemir, Z. Xue, S. Antani, G. Thoma, Feature Selection for Automatic Tuberculosis Screening in Frontal Chest Radiographs, Journal of medical systems, 42 (2018) 146.
[8] Y. Xiong, X. Ba, A. Hou, K. Zhang, L. Chen, T. Li, Automatic detection of mycobacterium tuberculosis using artificial intelligence, Journal of Thoracic Disease, 10 (2018) 1936-1940.
[9] J.L. Diaz-Huerta, A.D.C. Tellez-Anguiano, M. Fraga-Aguilar, J.A. Gutierrez-Gnecchi, S. Arellano-Calderon, Image processing for AFB segmentation in bacilloscopies of pulmonary tuberculosis diagnosis, PLoS One, 14 (2019) e0218861.
[10] E.J. Hwang, S. Park, K.N. Jin, J.I. Kim, S.Y. Choi, J.H. Lee, J.M. Goo, J. Aum, J.J. Yim, C.M. Park, D. Deep Learning-Based Automatic Detection Algorithm, G. Evaluation, Development and Validation of a Deep Learning-based Automatic Detection Algorithm for Active Pulmonary Tuberculosis on Chest Radiographs, Clinical Infectious Diseases, 69 (2019) 739-747.
[11] S. Lopez-Garnier, P. Sheen, M. Zimic, Automatic diagnostics of tuberculosis using convolutional neural networks analysis of MODS digital images, PLoS One, 14 (2019) e0212094.
[12] Q.H. Nguyen, B.P. Nguyen, S.D. Dao, B. Unnikrishnan, R. Dhingra, S.R. Ravichandran, Deep Learning Models for Tuberculosis Detection from Chest X-ray Images, 2019 26th International Conference on Telecommunications (ICT), IEEE, Hanoi, Vietnam, 2019, pp. 381-385.
[13] M. Norval, Z. Wang, Y. Sun, Pulmonary Tuberculosis Detection Using Deep Learning Convolutional Neural Networks, Proceedings of the 3rd International Conference on Video and Image ProcessingShanghai, China, 2019, pp. 47-51.
[14] Z.Z. Qin, M.S. Sander, B. Rai, C.N. Titahong, S. Sudrungrot, S.N. Laah, L.M. Adhikari, E.J. Carter, L. Puri, A.J. Codlin, J. Creswell, Using artificial intelligence to read chest radiographs for tuberculosis detection: A multi-site evaluation of the diagnostic accuracy of three deep learning systems, Scientific Reports, 9 (2019) 15000.
[15] T.B. Chandra, K. Verma, B.K. Singh, D. Jain, S.S. Netam, Automatic detection of tuberculosis related abnormalities in Chest X-ray images using hierarchical feature extraction scheme, Expert Systems with Applications, 158 (2020).
[16] S.K. Tao Hwa, A. Bade, M.H.A. Hijazi, M. Saffree Jeffree, Tuberculosis detection using deep learning and contrastenhanced canny edge detected X-Ray images, IAES International Journal of Artificial Intelligence (IJ-AI), 9 (2020).
[17] Z. Ul Abideen, M. Ghafoor, K. Munir, M. Saqib, A. Ullah, T. Zia, S.A. Tariq, G. Ahmed, A. Zahra, Uncertainty Assisted Robust Tuberculosis Identification With Bayesian Convolutional Neural Networks, IEEE Access, 8 (2020) 22812-22825.
[18] S. Govindarajan, R. Swaminathan, Extreme Learning Machine based Differentiation of Pulmonary Tuberculosis in Chest Radiographs using Integrated Local Feature Descriptors, Computer methods and programs in biomedicine, 204 (2021) 106058.
[19] Osvaldo A. Rosso, Susana Blanco, Juliana Yordanova, Vasil Kolev, Alejandra Figliola, Martin Schu¨rmann, E.B. ¸ar, Wavelet entropy: a new tool for analysis of short duration brain electrical signals, Journal of Neuroscience Methods (2001) 65-75.
[20] S. Chaplot, L.M. Patnaik, N.R. Jagannathan, Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network, Biomedical Signal Processing and Control, 1 (2006) 86-92.
[21] Y. Zhang, Z. Dong, L. Wu, S. Wang, A hybrid method for MRI brain image classification, Expert Systems with Applications, 38 (2011) 10049-10053.
[22] M. Tan, Q.V. Le, EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, International Conference on Machine LearningLong Beach, California, 2019.
[23] Y.H. Pao, G.H. Park, D.J. Sobajic, Learning and generalization characteristics of random vector functional-link net, Neurocomputing, 6 (1994) 163-180.
[24] P.N. Suganthan, R. Katuwal, On the origins of randomization-based feedforward neural networks, Applied Soft Computing, 105 (2021).
[25] R. Katuwal, P.N. Suganthan, M. Tanveer, Random Vector Functional Link Neural Network based Ensemble Deep Learning, Pattern Recognition, 117 (2021).
[26] L. Zhang, P.N. Suganthan, A comprehensive evaluation of random vector functional link networks, Information sciences, 367 (2016) 1094-1105.
[27] Y. Ren, P.N. Suganthan, N. Srikanth, G. Amaratunga, Random vector functional link network for short-term electricity load demand forecasting, Information Sciences, 367-368 (2016) 1078-1093.