IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 6 April 2023, accepted 21 April 2023, date of publication 26 April 2023, date of current version 4 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3270774

RESEARCH ARTICLE

An Improved Densenet Deep Neural Network Model for Tuberculosis Detection Using Chest X-Ray Images

VO TRONG QUANG HUY AND CHIH-MIN LIN
Department of Electrical Engineering, Yuan Ze University, Taoyuan 320, Taiwan
Corresponding author: Chih-Min Lin (cml@saturn.yzu.edu.tw)

This work was supported in part by the Ministry of Science and Technology of Republic of China under Grant MOST 109-2811-E-155-504-MY3.

ABSTRACT Tuberculosis (TB) is a highly contagious and life-threatening infectious disease that affects millions of people worldwide. Early diagnosis of TB is essential for prompt treatment and control of the spread of the disease. In this paper, a new deep learning model called CBAMWDnet is proposed for the detection of TB in chest X-ray (CXR) images. The model is based on the Convolutional Block Attention Module (CBAM) and the Wide Dense Net (WDnet) architecture, which has been designed to effectively capture spatial and contextual information in the images. The performance of the proposed model is evaluated based on a large dataset of chest X-ray images and it is compared to several state-of-the-art models. The results show that the proposed model outperforms the other models in terms of accuracy (98.80%), sensitivity (94.28%), precision (98.50%), specificity (95.7%) and F1 score (96.35%). Additionally, our model demonstrates excellent generalization ability, with consistent performance on different datasets. In conclusion, the proposed CBAMWDnet model is a promising tool for the early diagnosis of TB, with superior performance compared to other state-of-the-art models, as evidenced by the evaluation metrics of accuracy, sensitivity, and specificity.

INDEX TERMS Chest X-ray, convolutional neural network, deep learning, disease diagnosis, tuberculosis.

I. INTRODUCTION

Infectious tuberculosis (TB) is a disease caused by the Mycobacterium tuberculosis, a bacterium which causes tuberculosis in the body. It is no doubt that tuberculosis is a major health problem throughout the world, particularly in developing countries where it is prevalent. In accordance with statistics provided by the World Health Organization [1], in 2020 worldwide TB deaths are predicted to increase from approximately 1.2 million to nearly 1.4 million in HIV-negative individuals and from 209,000 to 214,000 in individuals with HIV. Tuberculosis cases had been clinically diagnosed based on symptoms, abnormalities on the CXR and medical records in order to determine the diagnosis. Based on the projected TB incidence and mortality for 16 countries that have been modeled up to 2025, there is a strong likelihood that these impacts will be much greater in 2021 and beyond, especially on the mortality and incidence of TB in 2021 and 2022. Under these circumstances, a quick and accurate diagnosis would be of great importance when it comes to the treatment and control of the disease. Currently, the most reliable method for identifying tuberculosis cases is isolating the bacteria that cause the disease. However, it is important to keep in mind that while this method may have a high specificity and a relatively low sensitivity. This means that it may take a long time to obtain test results, potentially up to three weeks. Other techniques, such as immunological tests and molecular biology tests, also have their own advantages and disadvantages. For example, immunological tests are simple and quick to perform, but they have low sensitivity and specificity, making them less commonly used for chronic tuberculosis infections. On the other hand, the

The associate editor coordinating the review of this manuscript and approving it for publication was Ravibabu Mulaveesala.

polymerase chain reaction (PCR) is a widely used nucleic acid amplification test that can detect tuberculosis in clinical specimens such as sputum, blood, bone marrow, and biopsy samples. However, PCR techniques are expensive and may not be available at all medical institutions. It is important to consider the benefits and limitations of different techniques when determining the most appropriate method for identifying tuberculosis cases.

Although Convolutional Neural Networks (CNNs) were originally introduced more than 25 years [2], [3], improvements in computer hardware and network structure, which enabled the training of truly deep CNNs, make them become the dominant machine learning approach for visual object recognition recently. The increasing number of both the layers and the size of each layer in modern networks amplifies the differences between architectures and motivates the exploration of different connectivity patterns and the revisiting of old research ideas. From LeNet [4] with only 5 layers, Resnet [5] with more than 100 layers, even the deep networks with stochastic depth that contain more than 1200 layers [6].

Deep learning techniques are currently being employed to address numerous critical problems. Speech emotion recognition (SER) has received a lot of attention in recent years [7], [8], and researchers have achieved excellent results thanks to the use of CNNs [9], [10], [11]. Another challenge is the analysis of non-stationary signals, which are characterized by a time-varying frequency spectrum, particularly in noisy environments. Machine learning-based techniques have been utilized for a range of tasks, including denoising of gravitational-wave data, estimation of parameters, classification of detector glitches, detection of gravitational waves [12], [13], [14].

Deep convolutional networks have also shown outstanding performance in tuberculosis diagnosis, surpassing other competing approaches. Pasa et al. [15] developed a specialized neural network architecture that significantly reduced computational, memory, and power requirements. Melendez et al. [16] demonstrated that Computer-Aided Detection (CAD) techniques can be enhanced with clinical information to improve accuracy and specificity. Vajda et al. [17] developed an automatic system that detects abnormal lungs with multiple tuberculosis manifestations by selecting features based on wrappers in order to minimize classification error. Lopes and João [18] investigated the use of pre-trained CNNs for tuberculosis detection and proposed three different CNN architectures for training an SVM classifier. Rahman et al. [19] and Rajaraman and Sameer [20] have proposed a reliable deep learning-based method for detecting tuberculosis from chest X-ray images, and it can achieve high accuracy and outperforming traditional methods. They also suggested that large datasets are necessary to train profound neural networks, although this is a costly and time-consuming process.

There are several ways to improve the performance of deep neural networks, but one of the most straightforward is by increasing their size. This can be done by increasing both the depth (e.g., using more layers) and the width of the network [21]. This method is easy and relatively safe, especially when there is a large amount of labeled training data available. However, it has two main drawbacks: the increased number of parameters may make the network more prone to overfitting, particularly when the training set is small, and the increased size also requires more computational resources. Another method, which is becoming increasingly popular, is to scale up the model by increasing the resolution of the images [22]. In the past, it was common to only scale one dimension (depth, width, or image size), but scaling all three can be more effective, though it requires manual tuning and may not always achieve optimal accuracy and efficiency. For example, if you want to use 2N times more computational resources, you can increase the network depth by αN, the width by βN, and the image size by γN, where α, β, γ are constants determined by a small grid search on the original, smaller model.

Despite the potential benefits of utilizing deep learning for tuberculosis detection, there are significant challenges to its implementation. One major obstacle is the need for large datasets of labeled medical images to train the models effectively, which can be difficult to obtain due to privacy concerns. Additionally, not all deep learning models are designed specifically for tuberculosis detection, which can result in suboptimal performance. There is also a risk of bias in the data, which may lead to inaccurate predictions or unfairly discriminate against certain populations.

The purpose of this study is to give physicians a tool to help them diagnose tuberculosis by developing a deep learning architecture that is tailored to tuberculosis diagnosis. This architecture will assist them in the diagnosis process. Our paper aims to contribute some of the most significant things about deep learning in this regard, in order to make a significant contribution:

- **Data**: No matter how effective algorithms for machine learning models are, it must keep in mind that the quality of data is just as critical as the quantity. In this research, we spend a great deal of time looking for open datasets that provide both the quality and quantity of information that we are seeking. The provided dataset helps deep learning models achieve high recognizing without requiring much adjustment.
- **Deep learning model**: This paper presents a new deep-learning architecture named as CBAMWDNet that can be tailored to the diagnosis of tuberculosis. It is observed that the CBAMWDNet, by increasing the computational and memory requirements just a bit, can significantly increase classification performance.
- **Comparison**: Since applying deep learning to medicine isn't a revolutionary concept nowadays, there are quite a number of researchers working on this particular topic, and many of them have already made considerable progress. In order to show the superiority of the proposed model, it needs to compare and evaluate this model with other deep learning models to show the superior of the proposed model to other models.

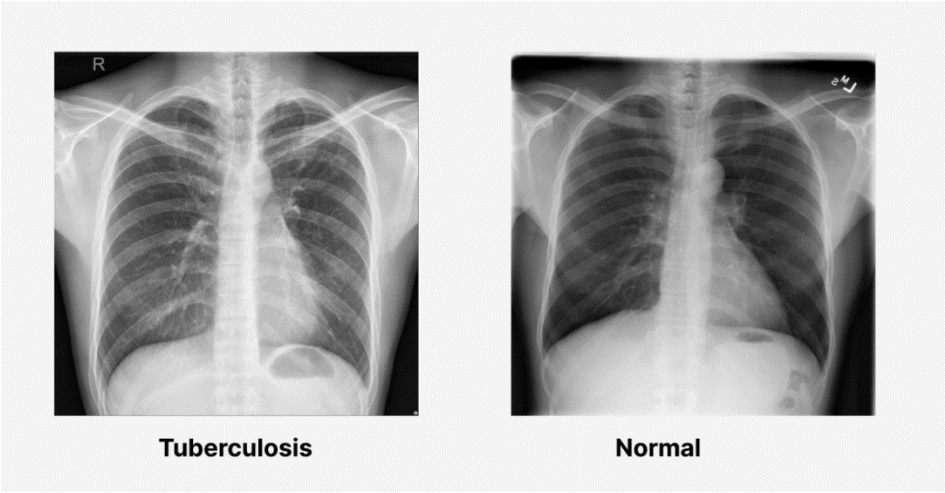

FIGURE 1. Sample of Tuberculosis CXR and Normal CXR.

TABLE 1. Details of total dataset, training set and validation set for classification problem.

Dataset	Total Images	Number of CXR Images & mask (Tuberculosis/Normal)	Train fold (Tuberculosis/Normal)	Validation Fold
Tuberculosis (TB) Chest X-ray Database	4200	700/3500	560/2800	140/700
Montgomery Dataset	138	58/80	46/64	12/16
Shenzhen Dataset	662	336/326	269/261	67/65
Total Dataset	5000	1094/3906	875/3125	219/781

This paper contributes significantly to deep learning in medical image analysis. Using its novel approach, it uses more comprehensive X-ray images for detecting tuberculosis than previous methods. Moreover, the proposed CBAMWDNet outperformed other state-of-the-art algorithms in terms of classification accuracy, sensitivity, and specificity in this study. The model also achieved satisfactory performance with fewer training epochs, resulting in significant savings in training time. Through these findings, it may be possible to improve tuberculosis detection and other medical image analysis tasks.

The paper is organized as follows: Section II provides details on the experimental setup employed in this study, including an introduction to the image dataset, the proposed CNN, and the validation methods used for the deep learning classification algorithms. Section III presents the experimental results, which are elaborated upon in detail. Finally, Section IV summarizes the concluding remarks.

II. METHODS AND MATERIALS

A. IMAGE DATASET

We populated a dataset from different publicly available data repositories as follows:

- Montgomery Dataset: CXR datasets from Montgomery County (MC) contain 138 frontal chest X-rays which are part of Montgomery County's Tuberculosis screening program. 80 of the images can be classified as ''normal'' and 58 of the images can be classified as ''TB manifestations'' [23].
- The Shenzhen dataset: In the Shenzhen dataset, there are 662 frontal CXR images that have been examined. Of these, 326 have been classified as normal and 336 have been classified as manifestations of tuberculosis [23].
- Tuberculosis (TB) chest X-ray dataset: This dataset was compiled by Qatar University, Doha, Qatar, as well as the University of Dhaka, Bangladesh, as well as their Malaysian collaborators, along with medical doctors from Hamad Medical Corporation and Bangladesh. This dataset consists of 700 images of CXRs that have been diagnosed with tuberculosis, and 3500 images of CXRs that are normal [19].

As a result, we merged and removed any corrupted images from these repositories in order to populate a dataset containing 5000 images, based on a combination of the repositories. An overview of the dataset resulting from this analysis is

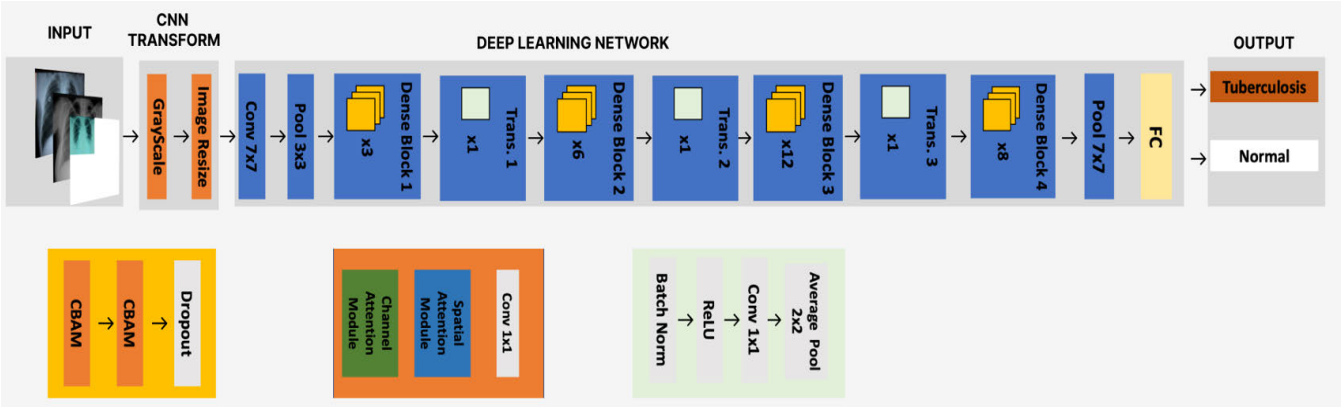

FIGURE 2. Architecture of the CBAMWDNet.

presented in Table 1. Particularly, there are 5000 CXR images in total, all of which are of high quality. In total, 3906 images are included in the Normal category and 1094 images have been included in the Tuberculosis category. In Figure 1, some examples of CXR images that have been extracted from the dataset for different categories can be seen. Afterward, the dataset was randomly divided into 80% training samples, and 20% validation samples under the 80%-20% ratio by a random number generator. More detail can be seen in Table 1.

B. THE PROPOSED CBAMWDNet

The proposed CBAMWDNet network architecture can be seen in Figure 2, which shows a schematic representation of the scheme. In this model, a feature map is constructed using a 7×7 kernel size and a stride of 2. This kernel size is large enough to capture more complex features in the input data, but not so large to be computationally expensive. By using a large kernel size, it allows the model to learn more abstract features from the input data. This will be useful for tasks such as image classification where the input data may contain a wide range of visual patterns. The stride of 2 helps to reduce the spatial resolution of the output tensor by a factor of 2, which can reduce the number of parameters in the model and improve its ability to generalize to new data. Additionally, it includes a pooling layer to reduce the number of computations required and make the network more efficient. After the initial feature map construction, we use 4 Denseblocks, each containing 2 CBAM (Convolutional Block Attention Module) layers and a 1×1 convolutional layer. These focus on the most relevant features in the input data, which can improve its performance on tasks such as image classification. Finally, the output of the Denseblocks is passed through a 7×7 convolutional layer and a fully connected layer before making predictions.

1) CBAM

Convolutional block attention module (CBAM) [24] is a type of attention mechanism that can be used in a convolutional neural network (CNN) to improve its performance. Attention mechanisms are used to allow the network to focus on certain parts of the input data and ignore the others, based on certain criteria. The basic structure of a CBAM consists of two parallel branches: a channel attention branch and a spatial attention branch. For Channel Attention in CBAM, it is similar to what has been done with the SE Module in Squeeze and Excitation Network [25], the winners of ILSVRC 2017 classification competition. But in the Squeeze section of CBAM, both GAP (Global Pooling) and GMP (Global Maximum Pooling) are applied simultaneously. In their article, the CBAM authors stated that GMP also collects relevant information about an object, but from a different perspective. Afterward, the two features that have been obtained, GMP and GAP, are all passed through the same excitation part, not two separate excitation parts. This leads to the creation of two vectors. Then, add them together, and take the sigmoid as in the SE module in order to calculate the result. Figure 3 shows the detail of the Channel Attention in CBAM. The implementation of Spatial Attention in CBAM is quite simple and it's not all that different from the implementation of Channel

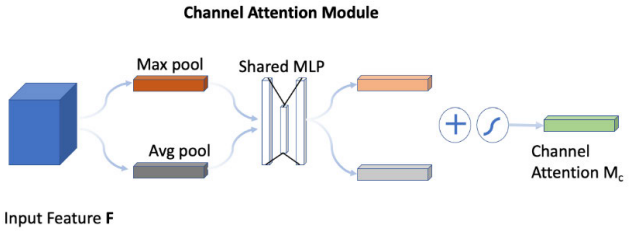

FIGURE 3. Channel Attention in CBAM.

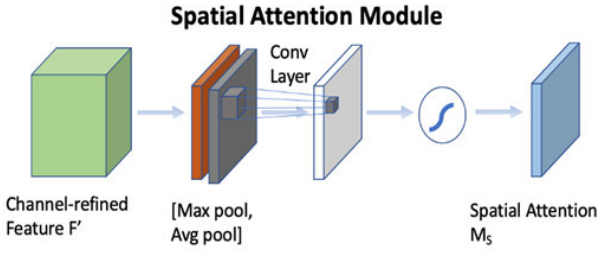

FIGURE 4. Spatial Attention in CBAM.

Attention. Instead of using GAP and GMP in the channel dimension of feature maps, GAP and GMP are processed in the spatial dimension of feature maps. Figure 4 shows the detail of the Spatial Attention in CBAM.

There are several reasons why convolutional block attention module (CBAM) can be useful in a convolutional neural network (CNN):

Improve model performance: By allowing the network to focus on certain parts of the input data and ignore others, CBAM can help improve the performance of the CNN on tasks such as image classification and object detection.

- Reduce overfitting: Attention mechanisms can help reduce overfitting by allowing the network to focus on the most relevant parts of the input data and ignore noise or other irrelevant information.
- Reduce the number of parameters: CBAM has a relatively simple structure and requires fewer parameters than some other attention mechanisms, which can make it easier to train and less prone to overfitting.
- Improve the interpretability of the model: By visualizing the attention map produced by CBAM, it is possible to understand which parts of the input data the network is focusing on and why. This can be useful for understanding the model's decision-making process and improving its transparency.

Wide: Wide ResNets are a type of deep learning neural network that have been shown to be more effective than traditional ResNets in some tasks, particularly in image classification. This is because Wide ResNets are able to capture more detailed features from the input data, resulting in better performance on the target task. One of the key differences between Wide ResNets and traditional ResNets is the number of filters used in the convolutional layers. Wide ResNets use a larger number of filters, which allows them to learn more detailed features from the input data. Additionally, Wide ResNets typically use skip connections, which help improve the flow of information through the network and make it easier for the network to learn complex patterns in the data. Our model also applies this by double the growth rate and half the number of layers.

Dense blocks: Dense block is a type of layer in a CNN that consists of multiple convolutional layers with dense connections between them. Dense connections refer to the fact that each layer in the block receives input from all of the previous layers in the block, rather than just the directly preceding layer. This allows the dense block to propagate information throughout the entire block and helps the network learn more abstract features. Some other models (like Resnet, Wide Resnet, ...) use residual block, which is also a type of layer in a CNN that consists of two or more convolutional layers with a skip connection between them (A skip connection is a shortcut that allows the output of a layer to be added directly to the output of a preceding layer, bypassing any intermediate layers). Both dense blocks and residual blocks can be used to improve the performance of CNNs, but they are used in

TABLE 2. Parameter details for cbamwdnet.

Layer			parameters
Sequential	Conv2d		9,408
	BatchNorm2d		128
	ReLU		0
	DenseBlock	DenseLayer	53,654
		DenseLayer	68,054
		DenseLayer	86,550
	Transition	BatchNorm2d	320
		ReLU	0
		Conv2d	12,800
		AvgPool2d	0
	DenseBlock	DenseLayer	60,342
		DenseLayer	76,790
		DenseLayer	97,334
		DenseLayer	121,974
		DenseLayer	150,710
		DenseLayer	183,542
	Transition	BatchNorm2d	544
		ReLU	0
		Conv2d	36,992
		AvgPool2d	0
	DenseBlock	DenseLayer	91,814
		DenseLayer	115,430
		DenseLayer	143,142
		DenseLayer	174,950
		DenseLayer	210,854
		DenseLayer	250,854
		DenseLayer	294,950
		DenseLayer	343,142
		DenseLayer	395,430
		DenseLayer	451,814
		DenseLayer	512,294
		DenseLayer	576,870
	Transition	BatchNorm2d	1,040
		ReLU	0
		Conv2d	135,200
		AvgPool2d	0
	DenseBlock	DenseLayer	206,142

TABLE 2. *(Continued.)* **Parameter details for cbamwdnet.**

		DenseLayer	245,630
		DenseLayer	289,214
		DenseLayer	336,894
		DenseLayer	388,670
		DenseLayer	444,542
		DenseLayer	504,510
		DenseLayer	568,574
	BatchNorm2d		1,032
Linear			517,000
Total params			8,159,134
Trainable params			8,159,134

different ways. Dense blocks are typically used to allow the network to learn more abstract features, while residual blocks are used to allow the network to learn residuals and improve the flow of information through the network.

CBAMWDNet is a convolutional neural network specially created for the purpose of tuberculosis classification. Due to its deep architecture, the model comprises of a large number of parameters, which amounts to 8,159,134. This extensive parameter count reflects the model's capability of effectively extracting intricate features and patterns from the input images. A detailed distribution of the parameter count across different layers of the network is presented in Table 2.

C. EVALUATION

The performance of different CNNs for the testing dataset is evaluated after the completion of the training and validation phases. This evaluation is conducted to assess the effectiveness of different CNNs in predicting the outcome or class of the samples in the testing dataset. The performance of different CNNs is compared using six performance metrics: accuracy, sensitivity, specificity, precision, negative predictive value, and F1 score. These metrics are chosen because they are widely used to evaluate the performance of classification models, and each of them captures a different aspect of the model's performance. The results of this evaluation allow researchers to compare the performance of different CNNs and determine which one is the most effective in predicting the outcome or class of the samples in the testing dataset.

Hyperparameters play a crucial role in the performance of deep learning models. However, selecting the appropriate values for these parameters can be challenging for different tasks. In this study, we have used default values for some of the most common hyperparameters in deep learning models since our focus is on the deep learning process itself. The batch size is set to 16, and we use the Stochastic Gradient Descent (SGD) optimizer with a learning rate of 0.001 and a momentum of 0.9. Additionally, we employ a decay learning rate with a step size of 7 and a gamma value of 0.1. Other hyperparameters, such as input size and dropout rate, are selected based on the specific model used for detection.

In this study, true positive (TP), true negative (TN), false positive (FP), and false negative (FN) are used to evaluate the performance of different CNNs in detecting tuberculosis. TP refers to the number of tuberculosis images that are correctly identified as tuberculosis, TN refers to the number of normal images that are correctly identified as normal, FP refers to the number of normal images that are incorrectly identified as tuberculosis images, and FN refers to the number of tuberculosis images that are incorrectly identified as normal. These performance metrics are commonly used in the evaluation of classification models, and they allow researchers to understand the model's ability to correctly classify samples as either positive or negative for the outcome or class being predicted.

1) ACCURACY

Accuracy is a commonly used measure to evaluate the performance of a deep learning model in detecting tuberculosis. It is the proportion of correctly classified samples out of all samples, and can be calculated as the number of true positive and true negative predictions divided by the total number of predictions made. For example, if a deep learning model correctly classifies 95 out of 100 samples as either positive or negative for tuberculosis, its accuracy is 95%.

$$Accuracy = \frac{(TP + TN)}{TP + FN + FP + TN} \tag{1}$$

2) SENSITIVITY OR RECALL OR TRUE POSITIVE RATE (TPR)

TPR is another important measure to consider when evaluating the performance of a deep learning model in detecting tuberculosis. It is the proportion of positive samples that are correctly classified as positive, and can be calculated as the number of true positive predictions divided by the total number of actual positive samples. For example, if a deep learning model correctly classifies 90 out of 100 samples as positive for tuberculosis, its sensitivity is 90%.

$$Sensitive\ (TPR) = \frac{TP}{TP + FN} \tag{2}$$

3) SPECIFICITY OR SELECTIVITY OR TRUE NEGATIVE RATE(TNR)

TNR is a measure that is important to consider when evaluating the performance of a deep learning model in detecting tuberculosis. It is the proportion of negative samples that are correctly classified as negative, and can be calculated as the number of true negative predictions divided by the total number of actual negative samples. For example, if a deep learning model correctly classifies 95 out of 100 samples as negative for tuberculosis, its specificity is 95%.

$$specificity(TNR) = \frac{TN}{TN + FP} \tag{3}$$

4) PRECISION OR POSITIVE PREDICTIVE VALUE (PPV)

Precision is a measure of the proportion of positive predictions that are actually correct. It can be calculated as the number of true positive predictions divided by the total number of positive predictions made. For example, if a deep learning model makes 100 positive predictions for tuberculosis and 90 of them are correct, its precision is 90%.

$$precision(PPV) = \frac{TP}{TP + FP} \tag{4}$$

5) NEGATIVE PREDICTIVE VALUE (NPV)

NPV is a measure of the proportion of negative predictions that are actually correct. It can be calculated as the number of true negative predictions divided by the total number of negative predictions made. For example, if a deep learning model makes 100 negative predictions for tuberculosis and 90 of them are correct, its negative predictive value is 90%.

$$NPV = \frac{TN}{TN + FN} \tag{5}$$

6) F1 SCORE

F1 score is a measure that combines both precision and recall, or sensitivity. It is calculated as the harmonic mean of precision and recall, and is a useful metric when there is a need to balance both false positives and false negatives. A high F1 score indicates a good balance between precision and recall.

$$F1\_score = \frac{(2 * TP)}{(2 * TP + FN + FP)} \tag{6}$$

In addition to evaluating the performance of different CNNs using these metrics, the researchers also compared the networks in terms of the processing time required for training per 50 epochs (δ_{e50}). This processing time is the amount of time it takes for a network to complete one epoch of training, and it is measured in seconds. The networks are compared for the time between the start and end times of the training epochs, where t_1 and t_2 represent the start and end times, respectively. This allows researchers to understand the efficiency of different CNNs in terms of the time required to train the model.

$$\delta_{e50} = t_2 - t_1 \tag{7}$$

III. EXPERIMENTAL RESULTS

In Figure 5 and Figure 6, the training process of our model is presented using 50 epochs for the Tuberculosis (TB) Chest X-ray Dataset and the Total dataset, which combines data from three different datasets, respectively. These figures provide a visual representation of the model's performance during training, enabling us to understand how well the model is learning and adapting to the data.

Tables 3 and 4 present the training times for 7 models using 50 epochs for the Tuberculosis (TB) Chest X-ray dataset and the Total dataset, respectively. These tables provide a summary of the training times for each model and allow us to compare the performance of different models on these datasets. The results indicate that our model may take longer to train compared to other deep learning models. This is because we have employed the use of CBAM (Convolutional Block Attention Module), a more complex architecture that has been demonstrated to enhance model performance in certain tasks. While CBAM does improve model performance, it also requires more computational resources and a longer training time due to its increased complexity.

The performance evaluations are presented in Table 5 and Table 6. Table 5 shows the overall performance of our model on the validation dataset, including metrics such as accuracy,

true positive rate, and false positive rate, positive predictive value, negative predictive value and F1 score for only Tuberculosis (TB) Chest X-ray Dataset. Table 6 shows the overall performance of our model on the validation dataset, including metrics such as accuracy, true positive rate, and false positive rate, positive predictive value, negative predictive value and F1 score for total dataset which combine from Shenzhen dataset, Montgomery dataset and Tuberculosis (TB) Chest X-ray Dataset. Both tables provide a clear and comprehensive view of the performance of our model and demonstrate its ability to effectively learn and generalize from the training data. It is believed that these results are highly encouraging the superiority of our model and the effectiveness of our training process.

The following are some important values from the experiments:

A. ACCURACY

The accuracy of our model has achieved an satisfactory accuracy of 98.80% for TB dataset and 97.00% for total dataset, which is significantly higher than the accuracy of other deep learning models. This means that our model is able to correctly classify a larger proportion of the input data, which is a key indicator of its effectiveness.

B. TRUE POSITIVE RATE

The true positive rate of our model is also significantly higher than that of other deep learning models. Our model archives 94.28% for TB dataset and 95.43% for total dataset. This means that our model is able to accurately identify a larger proportion of positive cases, which is important for tasks such as this tuberculosis diagnosis.

C. TRUE NEGATIVE RATE

In contrast to other deep learning models, our model has highest True negative rate. It got 99.71% for TB dataset and 97.43% for total dataset This means that it is able to accurately identify a large proportion of negative cases, which is important for avoiding false alarms and unnecessary interventions.

D. POSITIVE PREDICTIVE VALUE

The positive predictive value of our model is also higher than other deep learning models. Our model got 98.50% for TB dataset and 91.26% for total dataset. This means that our

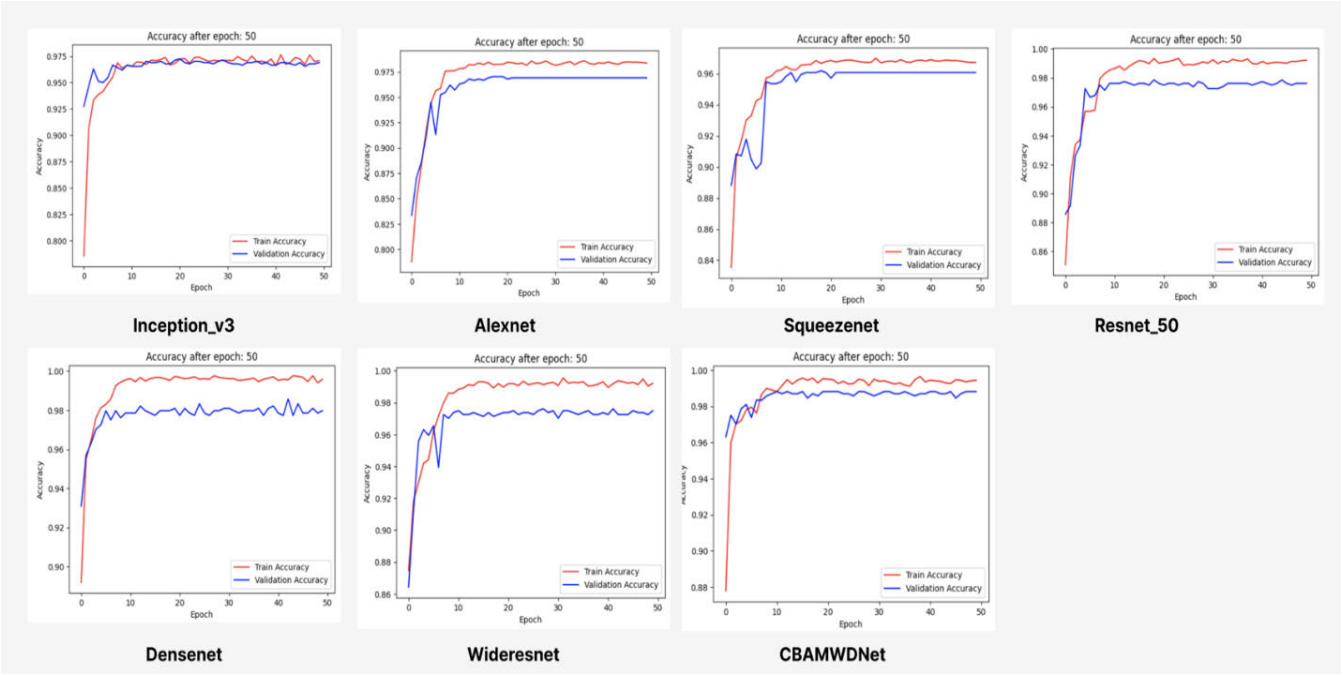

FIGURE 5. Accuracy for 50 epochs Training from Tuberculosis (TB) Chest X-ray Dataset.

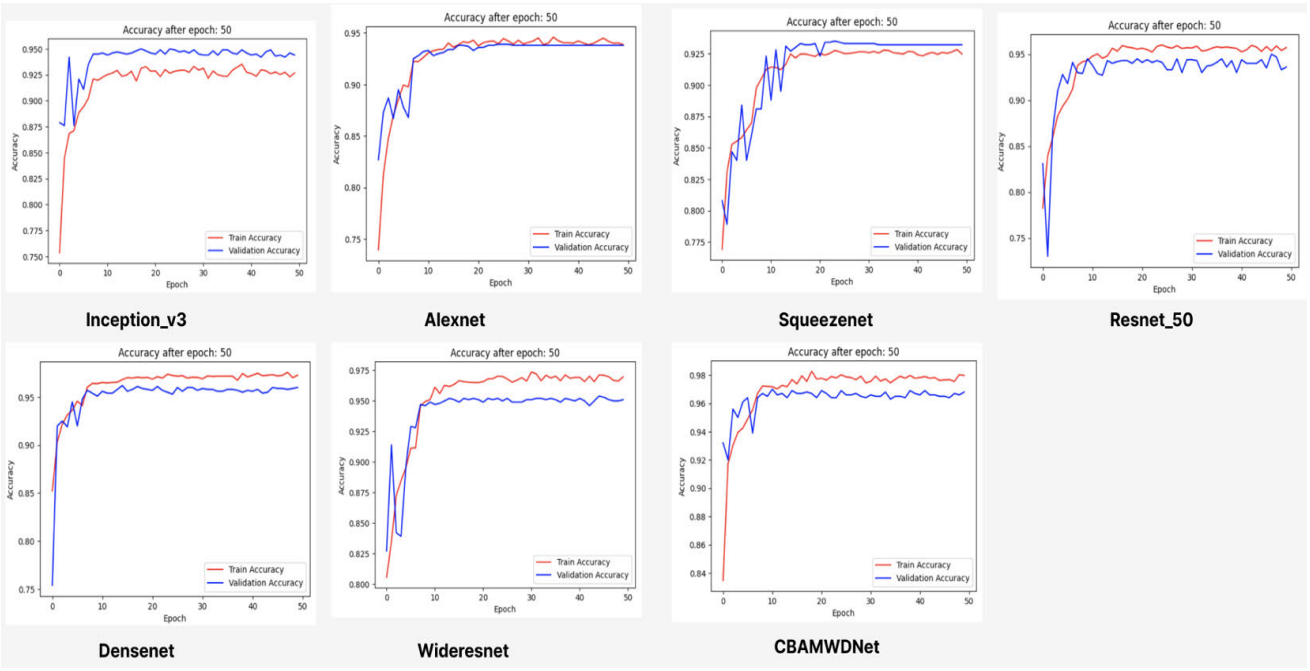

FIGURE 6. Accuracy for 50 epochs Training from total Dataset.

model is able to accurately predict a larger proportion of positive cases, which is important for tasks such as disease diagnosis like tuberculosis.

E. NEGATIVE PREDICTIVE VALUE

Our model has a very high negative predictive value, which is significantly higher than that of other deep learning models. It got 98.86% for TB dataset and 98.70% for total dataset. This means that it is able to accurately predict a large proportion of negative cases, which is important for avoiding false alarms and unnecessary interventions.

F. F1 SCORE

The F1 score of our model is also significantly higher than that of other deep learning models. It got 96.35% for TB

TABLE 3. Training time for each model in tuberculosis (TB) chest X-RAY dataset.

TB Dataset	Total Data Training (images)	Accuracy	Total Time for Training (50 epochs)
Inception_V3	3360	0.9726	30m 22s
Alexnet	3360	0.9702	21m 1s
Squeezenet	3360	0.9619	21m 44s
Resnet	3360	0.9785	26m 12s
Densenet	3360	0.9857	28m 55s
Wideresnet	3360	0.9761	31m 33s
CBAMWDNet	3360	0.9880	34m 30s

TABLE 4. Training time for each model in total dataset.

Total Dataset	Total Data Training (images)	Accuracy	Total Time for Training (50 epochs)
Inception_V3	4000	0.9440	89m 3s
Alexnet	4000	0.9380	74m 45s
Squeezenet	4000	0.9320	75m 37s
Resnet	4000	0.9360	82m 49s
Densenet	4000	0.9600	86m 27s
Wideresnet	4000	0.9510	88m 12s
CBAMWDNet	4000	0.9680	91m 19s

TABLE 5. Performance of evaluated screening strategies for tb dataset.

TB Dataset	Accuracy	TPR	TNR	PPV	NPV	F1_Score
Inception_V3	0.9726	0.8785	0.9914	0.9534	0.9760	0.9144
Alexnet	0.9702	0.8571	0.9928	0.96	0.9720	0.9056
Squeezenet	0.9619	0.8000	0.9942	0.9655	0.9613	0.8750
Resnet	0.9785	0.8857	0.9971	0.9841	0.9775	0.9323
Densenet	0.9857	0.9285	0.9971	0.9848	0.9858	0.9558
Wideresnet	0.9761	0.8785	0.9957	0.9761	0.9761	0.9248
CBAMWDNet	0.9880	0.9428	0.9971	0.9850	0.9886	0.9635

TABLE 6. Performance of evaluated screening strategies for total dataset.

Total Dataset	Accuracy	TPR	TNR	PPV	NPV	F1_Score
Inception_V3	0.9500	0.8767	0.9705	0.8930	0.9656	0.8847
Alexnet	0.9390	0.8675	0.9590	0.8558	0.9627	0.8616
Squeezenet	0.9350	0.7899	0.9756	0.9010	0.9430	0.8418
Resnet_50	0.9500	0.8630	0.9743	0.9043	0.9620	0.8831
Densenet121	0.9116	0.8949	0.9807	0.9289	0.9708	0.9620
Wideresnet	0.9540	0.8630	0.9795	0.9219	0.9622	0.8915
CBAMWDNet	0.9700	0.9543	0.9743	0.9126	0.9870	0.9330

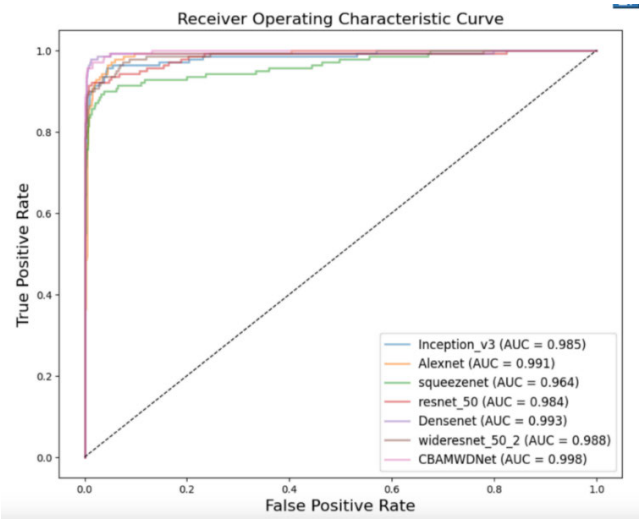

FIGURE 7. ROC curve for Total Dataset.

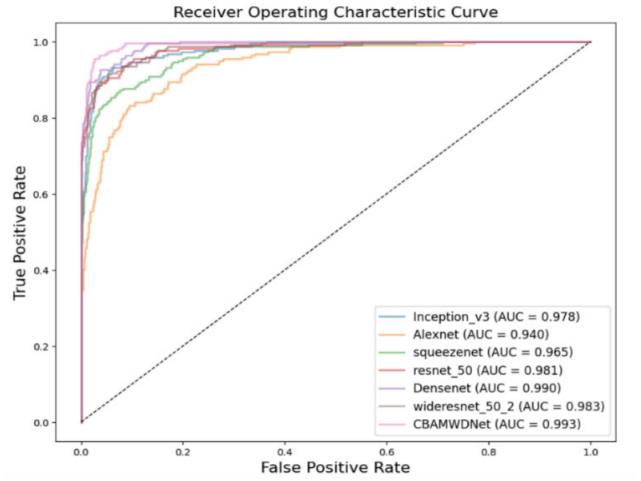

FIGURE 8. ROC curve for TB Dataset.

dataset and 93.30% for total dataset. This is a measure of the model's ability to balance precision and recall, and a high F1 score indicates that our model is able to accurately identify a large proportion of positive cases while also avoiding false alarms.

G. ROC

Figures 7 and 8 demonstrate that our model achieved the highest AUC (Area Under the Curve) compared to the other 6 deep learning models. This result shows the effectiveness of our model in accurately predicting the outcome of a binary classification task. The AUC is a widely used metric in machine learning, and a high AUC value indicates that the model can effectively distinguish between positive and negative classes.

Overall, our model has consistently outperformed other deep learning models in a wide range of metrics, including accuracy, true positive rate, false positive rate, positive predictive value, negative predictive value, F1 score and AUC. This shows the superiority of our model and the effectiveness of our training process.

IV. CONCLUSION

As a result of our study, it is found that the diagnosis performance of a supervised machine learning model depends on the dataset. This is because of the varying technical specifications of CXR images, as well as the distribution of disease severity in different populations. Not only CBAMWDNet but also other competing models maintain high diagnostic accuracy for the training, validation, and test images using the total dataset and Tuberculosis (TB) chest X-ray dataset. There is no doubt that the quality of data is as significant as the quantity of data. This is regardless of the fact that you have implemented a highly advanced algorithm. Moreover, our model has achieved satisfactory level of accuracy with our model. What makes this achievement particularly noteworthy is that we are able to achieve this level of accuracy without using any pre-trained processes and with only 50 epochs of training. This is worth mention, as it typically takes many more epochs to achieve such high levels of accuracy for other models. The fact that our model is able to learn and generalize so effectively from the training data in such a short amount of time is a testament to its efficiency and the effectiveness of our training process. With such high levels of accuracy, we are confident that the model will perform well on unseen data, and we look forward to analyzing how this model performs in real-world situations.

There are a few limitations to this study that need to be considered. The first difficulty is that CBAMWDNet is computationally intensive because of its large number of parameters, which can pose a challenge for researchers who do not have access to high-performance computing resources. Furthermore, the lack of labeled data may also pose a limitation. Limited availability of validated open-source data may hinder the performance of the model, resulting in lower accuracy and reliability. It may be possible to resolve this issue in the future when more valid open-source data is available.

For future study, the proposed CBAMWDNet can be also applied to other disease classification tasks. Alternatively, incremental learning can enable continuous training with new data while retaining knowledge from previous sessions, resulting in improved accuracy. Additionally, an enhanced weighted model ensemble strategy may be considered for further optimizing the performance of the model. Lastly, multi-objective optimization techniques may be utilized in order to determine the optimal model weights and possibly enhance the overall performance of CBAMWDNet.

DATA AVAILABILITY STATEMENT

The Tuberculosis (TB) Chest X-ray Database is collected by Rahman et al. [19] and can be access from https://www.kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset

The Montgomery dataset is collected by Jaeger et al. [23] and can be access from https://www.kaggle.com/datasets/raddar/tuberculosis-chest-xrays-montgomery

The Shenzhen dataset is also collected by Jaeger et al. and can be access from https://www.kaggle.com/datasets/raddar/tuberculosis-chest-xrays-shenzhen

REFERENCES

[1] N. Salazar-Austin, C. Mulder, G. Hoddinott, T. Ryckman, C. F. Hanrahan, K. Velen, L. Chimoyi, S. Charalambous, and V. N. Chihota, "Preventive treatment for household contacts of drug-susceptible tuberculosis patients," *Pathogens*, vol. 11, no. 11, p. 1258, Oct. 2022.

[2] S. N. Cho, "Current issues on molecular and immunological diagnosis of tuberculosis," *Yonsei Med. J.*, vol. 48, no. 3, pp. 347–359, 2007.

[3] B. Zhang, Q. Zhao, W. Feng, and S. Lyu, "AlphaMEX: A smarter global pooling method for convolutional neural networks," *Neurocomputing*, vol. 321, pp. 36–48, Dec. 2018.

[4] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," in *The Handbook of Brain Theory and Neural Networks*, vol. 3361, no. 10. Cambridge, MA, USA: MIT Press, p. 1995.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[6] G. Huang, "Deep networks with stochastic depth," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2016, pp. 646–661.

[7] M. Maithri, U. Raghavendra, A. Gudigar, J. Samanth, P. D. Barua, M. Murugappan, Y. Chakole, and U. R. Acharya, "Automated emotion recognition: Current trends and future perspectives," *Comput. Methods Programs Biomed.*, vol. 215, Mar. 2022, Art. no. 106646.

[8] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar, and T. Alhussain, "Speech emotion recognition using deep learning techniques: A review," *IEEE Access*, vol. 7, pp. 117327–117345, 2019.

[9] Z. Zhao, Z. Bao, Y. Zhao, Z. Zhang, N. Cummins, Z. Ren, and B. Schuller, "Exploring deep spectrum representations via attention-based recurrent and convolutional neural networks for speech emotion recognition," *IEEE Access*, vol. 7, pp. 97515–97525, 2019.

[10] L. Yang, K. Xie, C. Wen, and J.-B. He, "Speech emotion analysis of netizens based on bidirectional LSTM and PGCDBN," *IEEE Access*, vol. 9, pp. 59860–59872, 2021.

[11] S. Zhong, B. Yu, and H. Zhang, "Exploration of an independent training framework for speech emotion recognition," *IEEE Access*, vol. 8, pp. 222533–222543, 2020.

[12] W. Wei and E. A. Huerta, "Gravitational wave denoising of binary black hole mergers with deep learning," *Phys. Lett. B*, vol. 800, Jan. 2020, Art. no. 135081.

[13] A. J. K. Chua and M. Vallisneri, "Learning Bayesian posteriors with neural networks for gravitational-wave inference," *Phys. Rev. Lett.*, vol. 124, no. 4, Jan. 2020, Art. no. 041102.

[14] N. Lopac, F. Hrzic, I. P. Vuksanovic, and J. Lerga, "Detection of non-stationary GW signals in high noise from Cohen's class of time–frequency representations using deep learning," *IEEE Access*, vol. 10, pp. 2408–2428, 2022.

[15] F. Pasa, V. Golkov, F. Pfeiffer, D. Cremers, and D. Pfeiffer, "Efficient deep network architectures for fast chest X-ray tuberculosis screening and visualization," *Sci. Rep.*, vol. 9, no. 1, p. 6268, Apr. 2019.

[16] J. Melendez, C. I. Sánchez, R. H. H. M. Philipsen, P. Maduskar, R. Dawson, G. Theron, K. Dheda, and B. van Ginneken, "An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information," *Sci. Rep.*, vol. 6, no. 1, p. 25265, Apr. 2016.

[17] S. Vajda, A. Karargyris, S. Jaeger, K. C. Santosh, S. Candemir, Z. Xue, S. Antani, and G. Thoma, "Feature selection for automatic tuberculosis screening in frontal chest radiographs," *J. Med. Syst.*, vol. 42, no. 8, pp. 1–11, Aug. 2018.

[18] U. K. Lopes and J. F. Valiati, "Pre-trained convolutional neural networks as feature extractors for tuberculosis detection," *Comput. Biol. Med.*, vol. 89, pp. 135–143, Oct. 2017.

[19] T. Rahman, A. Khandakar, M. A. Kadir, K. R. Islam, K. F. Islam, R. Mazhar, T. Hamid, M. T. Islam, S. Kashem, Z. B. Mahbub, M. A. Ayari, and M. E. H. Chowdhury, "Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization," *IEEE Access*, vol. 8, pp. 191586–191601, 2020.

[20] S. Rajaraman and S. K. Antani, "Modality-specific deep learning model ensembles toward improving TB detection in chest radiographs," *IEEE Access*, vol. 8, pp. 27318–27326, 2020.

[21] S. Zagoruyko and N. Komodakis, "Wide residual networks," 2016, *arXiv:1605.07146*.

[22] H. Liu, J. Xu, Y. Wu, Q. Guo, B. Ibragimov, and L. Xing, "Learning deconvolutional deep neural network for high resolution medical image reconstruction," *Inf. Sci.*, vol. 468, pp. 142–154, Nov. 2018.

[23] S. Jaeger, S. Candemir, S. Antani, Y. X. Wang, P. X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quant. Imag. Med. Surg.*, vol. 4, p. 475, Dec. 2014.

[24] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," 2018, *arXiv:1807.06521*.

[25] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141, doi: 10.1109/CVPR.2018.00745.

VO TRONG QUANG HUY was born in Da Nang, Vietnam, in 1994. He received the master's degree in information management from Yuan Ze University, Taoyuan, Taiwan, in 2019, where he is currently pursuing the Ph.D. degree in electrical engineering. His research interests include deep learning, neural networks, fuzzy logic control, and intelligent control systems.

CHIH-MIN LIN was born in Changhua, Taiwan, in 1959. He received the B.S. and M.S. degrees from the Department of Control Engineering, National Chiao Tung University, Hsinchu, Taiwan, in 1981 and 1983, respectively, and the Ph.D. degree from the Institute of Electronics Engineering, National Chiao Tung University, in 1986. He is currently a Chair Professor with Yuan Ze University, Taoyuan, Taiwan. His research interests include fuzzy neural networks, cerebellar model articulation controller, intelligent control systems, adaptive signal processing, and classification problem. He serves as an Associate Editor for IEEE TRANSACTIONS ON CYBERNETICS and IEEE TRANSACTIONS ON FUZZY SYSTEMS.