Evaluation of Image Processing Technologies for Pulmonary Tuberculosis Detection Based on Deep Learning Convolutional Neural Networks

Michael J. Norval and Zenghui Wang
Department of Electrical and Mining Engineering, University of Souht Africa, Johannesburg, South Africa
Email: {mnorval, wangzengh}@gmail.com

Yanxia Sun
Department of Electrical and Electronic Engineering Science, University of Johannesburg, Johannesburg, South Africa
Email: sunyanxia@gmail.com

***Abstract*—Tuberculosis (TB) is a serious infectious disease that mainly affects the human lungs. The bacteria that cause TB are spread via minute droplets released into the air via sneezes and/or coughs. A bacterium called Mycobacterium is the root cause of TB. This paper is to investigate the precision of four factors of detecting Pulmonary Tuberculosis based on the patients' chest X-ray images (CXR) using Convolutional Neural Networks (CNN). We evaluate image dataset resolution, and then the pre-trained networks (AlexNet, VGG16 and VGG19) and various hyperparameter changes are investigated. Finally, additional sample images are tested and investigated. Simulations have been carried out based on 406 normal images & 394 abnormal images. Later an additional 239 normal images and 554 abnormal images are added. It is found that the splitting of images yielded the best results.**

***Index Terms*—artificial intelligence, DICOM, tuberculosis, pulmonary**

I. Introduction

The acronym DICOM stands for Digital Imaging and Communication in Medicine and forms a key part in digital CXR images. Digital software-driven healthcare is in use not only by radiology departments but many other clinical disciplines [1]. Once patient data has been digitized, it would allow for computerized diagnostics and artificial intelligence detection mechanisms. CNN can be used to check for the existence of tuberculosis bacteria presence in digital x-ray data. Vision detection and machine learning techniques need expertly crafted features to be manually created and defined [2]. There are too many features in medical images for this to be a viable technique [3]. A solid approach would be Deep Learning Neural Networks, coupled with supervised learning [4]. Numerous visual features are extracted with classification being a threshold or k-nearest approach. Stability, universality and reliability can be provided by Convolutional Neural Network (CNN) architectures [5]. South Africa is one of the countries with a very high infection rate. Subsequently, it is vital to detect the infection as soon as possible. By utilizing cell phone communications network, the images can be digitized and sent to a remote facility for diagnostics. The task of a Radiologist would be simplified, and they can only confirm the result obtained by CNN. This will make the whole diagnostic process much quicker and efficient. Over time, the more data received, the more the network will learn. More learning equates to better accuracy. It is thus necessary to investigate and identify the approaches yielding the higher accuracy. This research paper is to diagnose patient CXR images using CNN, and we investigate four deep learning approaches. Firstly, we evaluate image dataset resolution. Secondly, pre-trained networks are looked at, and AlexNet, VGG16 & VGG19 networks are assessed. Thirdly, hyperparameter changes are evaluated. Finally, we focus on data augmentation. The method yielding the best results is that of data augmentation.

Manuscript received October 26, 2020; revised April 2, 2021.

II. Deep Learning Convolutional Neural Network (DLCNN) Based TB Detection

A. DICOM

Figure 1. X-Ray machine [6].

doi: 10.12720/jait.12.3.253-259

The acronym DICOM stands for Digital Imaging and Communication in Medicine and forms a key part in digital CXR images. Where in the olden days an X-ray image was taken and then developed on a film to form the image, DICOM is the method where digitized image data and patient demographics can be transmitted from a modality to a workstation or Picture Archiving and Communication System (PACS) An X-Ray machine can be seen in Fig. 1 [6].

B. Pulmonary Tuberculosis

When the tuberculosis infection becomes active, the lungs are mostly affected, but other body parts outside the lungs can also be affected. Three main TB types are Miliary TB, Active TB, Latent TB. The focus of this paper is on active TB. Active TB occurs when the bacteria multiplies and invades different organs. Typical symptoms are cough, phlegm, chest pain, weakness, weight loss, fever, chills, and sweating at night. A person with active PTB may spread the disease via air. Typical symptoms are cough, phlegm, chest pain, weakness, weight loss, fever, chills, and sweating at night. A person with active PTB may spread the disease via air [7]. Not every patient demonstrates the symptoms; some patients have little or no symptoms [8]. In Fig. 2, one can see an electron microscope image of the Tuberculosis bacterium [9].

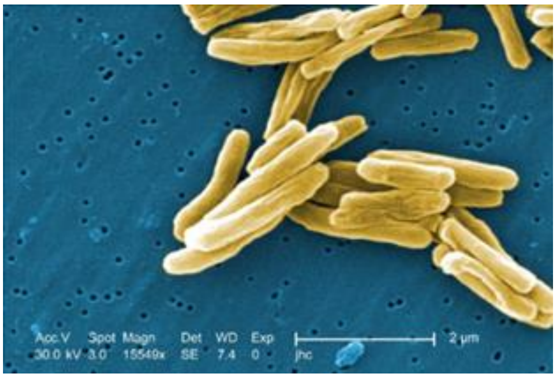

Figure 2. Image of an electron microscope enlargement of the Tuberculosis bacterium [9].

Fig. 3 below shows the flow of a digitized image from the modality where it is acquired to the store and ultimately the viewing and interpreting of the data.

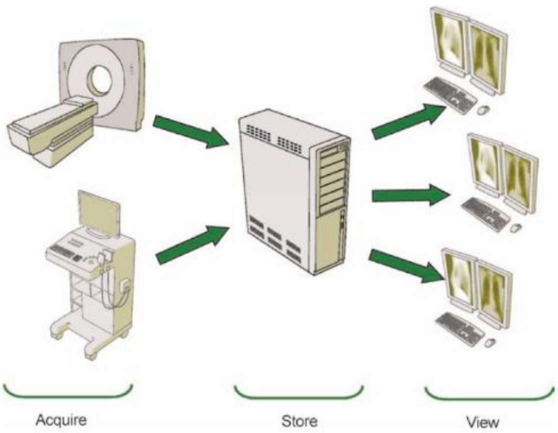

Figure 3. DICOM PACS components.

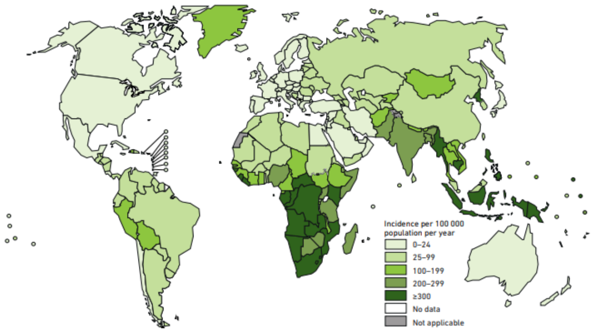

Figure 2. TB incident rates.

Fig. 4 shows the TB incident rates across the globe. Darker green indicates higher levels, whereas light green shows less infections.

C. Convolutional Neural Networks

In Multi-layer Neural Networks (MLNN), the inputs are usually in vector form. The neighbouring pixels or voxels in Medical images are another source of information. The vectorization process used in MLNN discards the voxel and neighbouring pixel information, and thus CNN's are used. The spatial information in the voxels can be much better utilized in CNN. A typical CNN is shown in Fig. 5 [10].

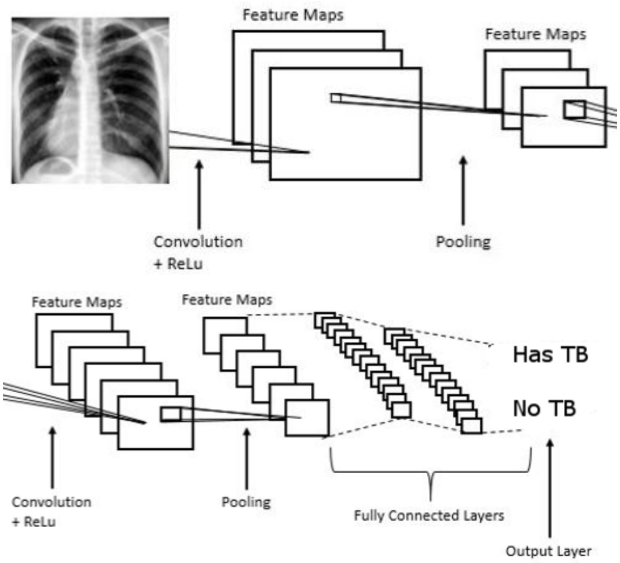

Figure 3. Multi-layer neural network [10].

III. Research Methodology

A. Resolution

Since no similar research pertaining to TB image resolution could be found, it was decided to test how is the performance using different resolution. Although some research exists, but not pertaining specifically to CXR images [11]. Input data with varying resolutions are tested in order to find the optimal resolution. The process can be seen in Fig. 6. A random generator is set, and then the original and additional dataset is used for training. The results are shown in the evaluation section.

Figure 4. Resolution change.

B. Transfer Learning

Transfer learning has been used for TB detection. Pre-trained CNN networks like ResNet121 has been evaluated [12]. In this experiment, more networks are tested. To improve detection accuracy, it is necessary to test the use of pre-trained NN / transfer learning using the steps shown in Fig. 6. A fixed random generator provides more stable results that fluctuate less. The networks tested were AlexNet, VGG16 and VGG19. Input layer was replaced with the TB dataset, and the final layers and weights recalculated. The steps can be seen in Fig. 7. Initially, the random generator is set, then the pre-trained network is loaded, relevant sections replaced, and finally, the results recorded. The results are shown in the evaluation section.

Figure 5. Transfer learning.

C. Hyperparameter Adjustment

Adjustment of specific parameters and functions like dropout has been researched extensively [13]. Thus hyperparameter changes are tested for TB recognition. Again a fixed random generator is utilized for more stable accuracy results. The process can be seen in Fig. 8. A random generator is set, and then a myriad of layers and methods are tested and the results recorded. The results are shown in the evaluation section.

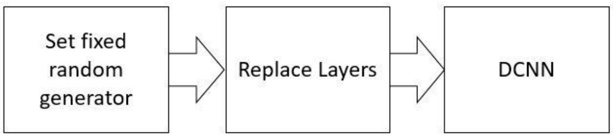

Figure 6. Hyperparameter changes.

D. Data Augmentation

Research exists but not specific to medical CXR images [14]. Various data augmentation techniques are tested. The process can be seen in Fig. 9. A random generator is set, and then the augmented training data is used to train the NN, and the results are recorded. The results are shown in the evaluation section.

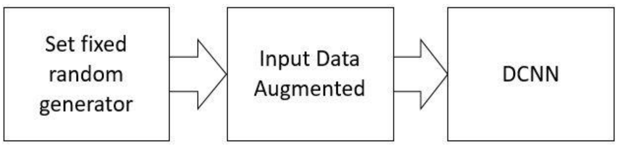

Figure 7. Data augmentation.

IV. Datasets and Settings

A. Datasets

Link to main Datasets used: https://ceb.nlm.nih.gov/repositories/tuberculosis-chest-x-ray-image-data-sets/

Shenzhen Hospital X-ray set: X-ray images in this data set have been collected by Shenzhen No.3 Hospital in Shenzhen, Guangdong province, China. The x-rays were acquired as part of the routine care at Shenzhen Hospital. The set contains images in JPEG format. There are 336 normal x-rays and 326 abnormal x-rays showing various manifestations of tuberculosis.

Montgomery: X-ray images in this data set have been acquired from the tuberculosis control program of the Department of Health and Human Services of Montgomery County, MD, USA. This set contains 138 posterior-anterior x-rays, of which 80 x-rays are normal, and 58 x-rays are abnormal with manifestations of tuberculosis. All images are de-identified and available in DICOM format. The set covers a wide range of abnormalities, including effusions and miliary patterns.

National Institutes of Health (https://www.nih.gov): National Institutes of Health supplied an additional 554 images of patients with TB and 239 of patients without TB, which was also preprocessed by cropping the ROI and contrast enhancement.

B. Settings

The software used for the simulation is Matlab. A product of Mathworks. The Image Processing and Neural Network Toolboxes of this product are used.

- All data is shuffled as to never have a repeat sequence of the same data
- A mini-batch of 80 is used
- 30 Epochs specified as a maximum
- 10% of all data is used for verification
- A fixed random generator for the initial weight selection is used. This ensures that all simulations start with the same random weight values selected. Subsequently, the results do not deviate. This method is used in all simulations.

V. Evaluation

In this evaluation section, the detection success rate of the four methods is investigated. The following methods are evaluated: transfer learning, hyperparameter changes, data augmentation, resolution and additional preprocessed training data.

A. Resolution

Image resolution equates to the number of pixels per inch PPI (Pixels Per Inch). An image resolution example can be seen in Fig. 10 and CXR image at various resolutions in Fig. 11. Various resolutions are tested to observe the impact. All simulations are repeated 3 times, and the results are recorded in Table I. (Result 1, Result 2 & Result 3).

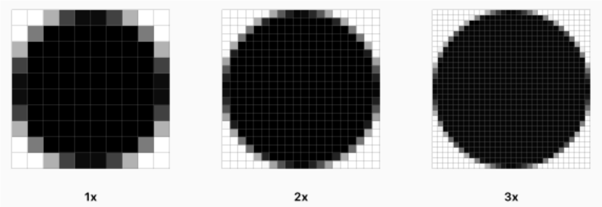

Figure 8. Image resolution.

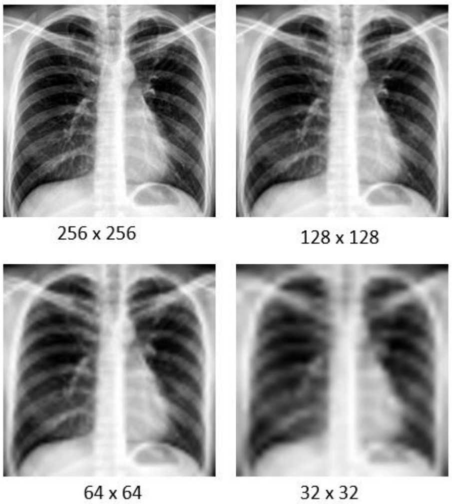

Figure 9. CXR image at various resolutions.

TABLE I. RESOLUTION RESULTS

Pre Processing	*Image resolution & color channels*	*Result 1*	*Result 2*	*Result 3*
Cropped ROI & Contrast Enhance	32×32×1	50.75	50.75	50.75
Cropped ROI & Contrast Enhance	64×64×1	91.04	91.04	91.04
Cropped ROI & Contrast Enhance	80×80×1	88.06	88.06	88.06
Cropped ROI & Contrast Enhance	100×100×1	85.07	85.07	85.07
Cropped ROI & Contrast Enhance	128×128×1	83.58	83.58	83.58
Cropped ROI & Contrast Enhance	150×150×1	80.60	80.60	80.60
Cropped ROI & Contrast Enhance	180×180×1	80.60	80.60	80.60
Cropped ROI & Contrast Enhance	256×256×1	77.61	77.61	77.61
Cropped ROI & Contrast Enhance	512×512×1	50.75	50.75	50.75

Resolution accuracy peaks at 64×64 with a colour depth of 8 bits. Fig. 12 shows a graphical depiction of color depth. Higher resolutions perform weaker, and it is hypothesized that higher resolution has a tendency to cause overfitting.

From the results, one can clearly see that the optimal resolution is 64×64. It is suspected that at this resolution, there is just enough relevant data to train the DLCNN. At higher resolutions, overfitting occurs because of the DLCNN training on irrelevant features.

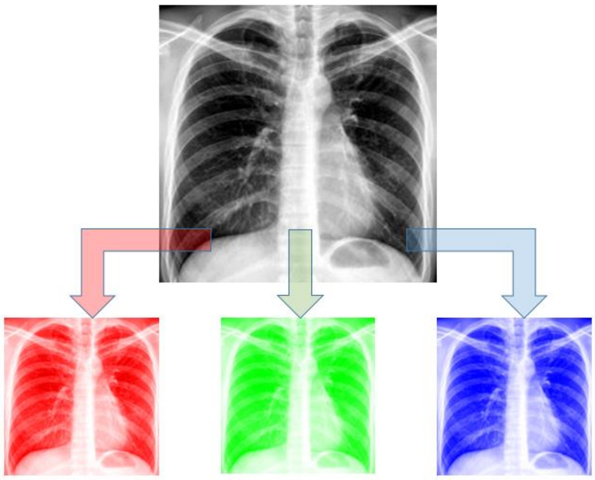

Figure 10. Color depth.

B. Transfer Learning

The pre-trained NN tested can be seen in Table II. The process of transfer learning vs training from scratch can be seen in Fig. 13 and Fig. 14. The simulation is repeated 3 times, and the results are recorded in Table III. (Result 1, Result 2 & Result 3)

TABLE II. PRE-TRAINED NETWORKS PARAMETERS

Network	Depth	Parameters	Input
AlexNET	8	61 Million	227 × 227
VGG16	16	138 Million	224 × 224
VGG19	19	144 Million	224 × 224

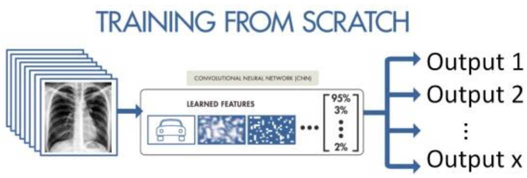

Figure 11. Normal learning.

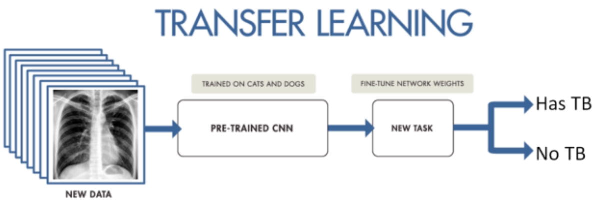

Figure 12. Transfer learning.

TABLE III. TRANSFER LEARNING RESULTS

Network Name	**Image resolution & color channels**	**Result 1**	**Result 2**	**Result 3**
AlexNet	64×64×1	89.55	89.55	88.06
VGG16	64×64×1	89.99	89.55	85.07
VGG19	64×64×1	86.57	86.57	88.06

The top-performing results are obtained from AlexNet and VGG16. These results are not higher than training the network from scratch, so no significant gain in accuracy recorded. It is hypothesized that the pre-trained networks are tuned for everyday objects and not medical images.

C. Hyper-Parameter Changes

The original NN layers can be seen in Table IV. The first column of Table V shows the change implemented before running the simulation. All simulations are repeated 3 times, and the results are recorded in Table IV. (Result 1, Result 2 & Result 3).

TABLE IV. ORIGINAL NEURAL NETWORK

1	Image Input	64×64×1 images with 'zero-centre' normalization
2	Convolution	64 5×5×1 convolutions with stride [1 1] and padding [2 2 2 2]
3	Max Pooling	3×3 max pooling with stride [2 2] and padding [0 0 0 0]
4	ReLU	ReLU
5	Convolution	64 5×5×64 convolutions with stride [1 1] and padding [2 2 2 2]
6	ReLU	ReLU
7	Average Pooling	3×3 average pooling with stride [2 2] and padding [0 0 0 0]
8	Convolution	128 5×5×64 convolutions with stride [1 1] and padding [2 2 2 2]
9	ReLU	ReLU
10	Average Pooling	3×3 average pooling with stride [2 2] and padding [0 0 0 0]
11	Fully Connected	128 fully connected layer
12	ReLU	ReLU
13	Fully Connected	2 fully connected layer
14	Softmax	softmax
15	Classification Output	crossentropyex with classes 'Adjusted_Cropped_Has_TB64' and 'Adjusted_Cropped_No_TB64'

TABLE V. HYPER PARAMETER CHANGE RESULTS

Parameter Change	**Image res & color channels**	**Result 1**	**Result 2**	**Result 3**
Change layer 7 & 10 to Max Pooling	64×64×1	80.6	80.6	80.6
Add additional Convolution and ReLU layer after layer 9	64×64×1	83.85	85.07	80.6
Change convolution layer 2,5 & 8 to have a padding value of 0.	64×64×1	88.06	86.57	86.57
Change convolution layer 2,5 & 8 to have a padding value of 1.	64×64×1	89.55	82.09	89.55
Change convolution layer 2,5 & 8 to have a padding value of 3.	64×64×1	80.60	82.09	85.07
Replace ReLU activation function in layer 4,6,9,12 with the LeakyReLU function [multiplication of 0.1].	64×64×1	88.06	85.07	86.57
Replace ReLU activation function in layer 4,6,9,12 with the LeakyReLU function [multiplication of 0.01].	64×64×1	83.58	83.58	88.06
Replace ReLU activation function in layer 4,6,9,12 with the LeakyReLU function [multiplication of 0.02].	64×64×1	86.57	83.58	83.58
Utilize Dropout Layer Hyper parameters Probability = 0.5	64×64×1	85.07	88.06	89.55

Using dropout with a 0.5 probability and padding value changes yields the highest accuracy. It is hypothesized that the network is already in an optimal state and thus changing the hyperparameters shows no significant increase in detection accuracy.

D. Data Augmentation Technique

Samples of data augmentation can be seen in Fig. 15 and Fig. 16. The types of data augmentation tested are image splitting, rotation and the addition of noise.

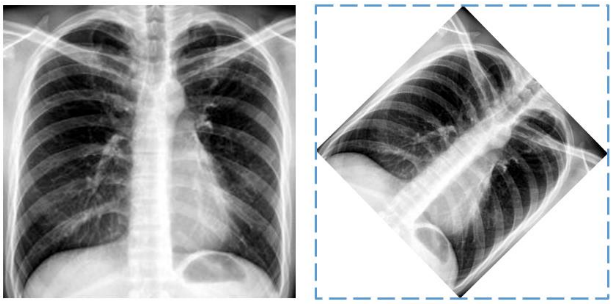

Figure 13. Data augmentation - Rotation example.

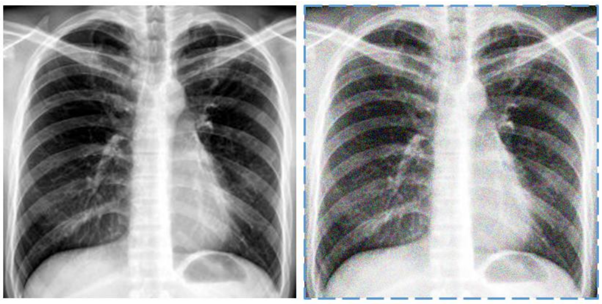

Figure 14. Data augmentation - Noise example.

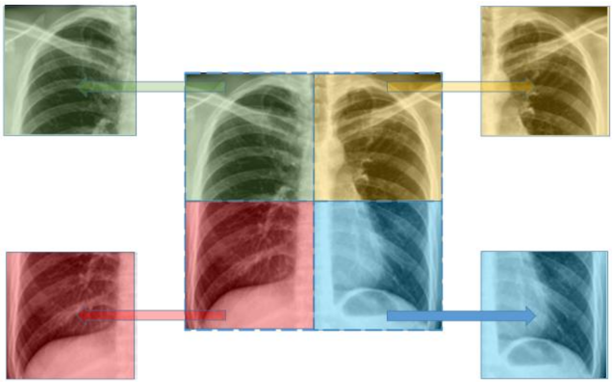

Figure 15. Data augmentation (Image splitting example - 1)

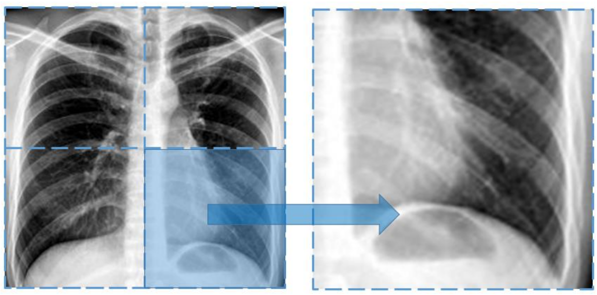

Figure 16. Data augmentation (Image splitting example - 2).

From all the augmentation techniques tested, the most successful was splitting of larger images into 4 parts. Fig. 17 and Fig. 18 shows a graphical depiction of the splitting technique. A 256×256 image is split into 4 × 64×64 images. It is suspected that the accuracy increases because of a bigger dataset. Also, the DCNN is able to extract well-defined features from each image.

Results are shown in Table VI. (Result 1, Result 2 & Result 3). All simulations are repeated 3 times.

TABLE VI. DATA AUGMENTATION RESULTS

Pre Processing	Image resolution & color channels	Result 1	Result 2	Result 3
Splitting of Images into 4 parts	64×64×1	84.85	89.99	92.54
Rotation 45 degrees	64×64×1	83.33	84.85	84.09
Rotation 90 degrees	64×64×1	81.06	77.27	82.58
Horizontal flipping	64×64×1	81.06	81.82	89.99
Vertical flipping	64×64×1	89.99	84.85	89.99
Noise	64×64×1	84.85	83.33	86.57

VI. CONCLUSION AND FUTURE WORK

The results yield solid evidence that data augmentation, coupled with an optimal resolution of 64×64 shows the highest accuracy. Firstly, image resolutions were tested, and it seems that the optimal resolution is 64×64. It is suspected that images with a too high-resolution cause overfitting. Secondly, pre-trained networks were tested. This yielded no significant improvement, and it is suspected that the reason for this is because of the fact that these networks have been training for everyday objects and not specifically for medical images. Next, Hyper-Parameter changes also yielded no significant increase. The reason being is that the network is already optimally configured. Finally, data augmentation was investigated, looking at rotation, flipping and the adding of noise. The splitting of images yielded the best results.

As a future project, we would be to obtain more clinical data from cooperating institutions. This will, over time, vastly improve the accuracy of the DCNN. In order to obtain the data, collaboration can be done with some sort of organization like a state TB hospital. Over time more data will be processed, and the system will keep on learning and improve its accuracy. This 'system' or rather portal can then be expanded to other institutions as well as assisting with the diagnosing of Tuberculosis. In conclusion, we would like to have a pre-trained NN, specializing in TB detection.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

M. J. Norval carried out the research under the supervision of Z. Wang and Y. Sun; and all authors approved the final version.

ACKNOWLEDGMENT

This research is supported partially by South African National Research Foundation Grants (No. 112108 and 112142), and South African National Research Foundation Incentive Grant (No. 95687 and 114911), Eskom Tertiary Education Support Programme Grants, Research grant from URC of University of Johannesburg.

REFERENCES

[1] W. Hruby, *Digital Revolution in Radiology: Bridging the Future of Health Care*, Vienna: Springer Verlag, 2010, pp. 1-25.

[2] P. Kamble, V. Anagire, and S. N. Chamtagoudar, "CXR tuberculosis detection using MATLAB image processing," *International Research Journal of Engineering and Technology (IRJET)*, pp. 1-3, 2016.

[3] U. K. Lopesb and J. F. Valiatia, "Pre-trained convolutional neural networks as feature extractors for tuberculosis detection," *Journal Computers in Biology and Medicine Archive*, vol. 89, pp. 135-143, October 2017.

[4] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future," *Classification in BioApps*, vol. 26, 2018.

[5] C. Liu, *et al*., "TX-CNN: Detecting tuberculosis in chest X-ray images using convolutional neural network," in *Proc. IEEE International Conference on Image Processing*, Beijing, 2017, pp. 2314-2318.

[6] X-Ray machine. [Online]. Available: https://www.imexmedical.com/product/x-ray-machine-for-medical-centre-chiropractor-etc/

[7] S. Jaeger, A. Karargyris, S. Antani, and G. Thoma, "Detecting tuberculosis in radiographs using combined lung masks," in *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012.

[8] A. Agyeman and R. Ofori-Asenso, "Tuberculosis—An overview," *Journal of Public Health and Emergency*, pp. 1-4, 2017.

[9] Mycobacterium tuberculosis and tuberculosis. [Online]. Available: http://textbookofbacteriology.net/tuberculosis.html

[10] Understanding-of-convolutional-neural-network-cnn-deep-learning. [Online]. Available: https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148

[11] S. P. Kannojia and G. Jaiswal, "Effects of varying resolution on performance of CNN based image classification an experimental study," *International Journal of Computer Sciences and Engineering*, vol. 6, no. 9, pp. 451-456, 2018.

[12] T. K. K. Ho, J. Gwak, O. Prakash, J. L. Song, and C. M. Park, "Utilizing pretrained deep learning models for automated pulmonary tuberculosis detection using chest radiography," in *Proc. 11th Asian Conference on Intelligent Information and Database Systems*, 2019, pp. 395-403.

[13] S. Park and N. Kwak, "Analysis on the Dropout Effect in convolutional neural networks," in *Proc. Asian Conference on Computer Vision*, 2016, pp. 189-204.

[14] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *Proc. International Interdisciplinary PhD Workshop*, Swinoujście, 2018, pp. 117-122.

Michael Norval obtained his B.Tech degree in Electronic Engineering, University of South Africa, South Africa, in the year of 2005. He is a master student and his research interests are Deep Learning Neural Network and Image Processing.
His research interests include Artificial Intelligence and Data Science.

Zenghui Wang received the B.Eng. degree in automation from NavalAviation Engineering Academy, China, in 2002, and the Ph.D. degree in control theory and control engineering from Nankai University, China, in 2007. Currently, he is a Professor with the Department of Electrical andMining Engineering, University of South Africa, Florida, South Africa. His research interests include model predictive control, nonlinear control, engineering optimization, image/video processing, artificial intelligence, chaos, and industry 4.0.

Yanxia Sun received her joint qualification: DTech in Electrical Engineering, Tshwane University of Technology, South Africa and PhD in Computer Science, University Paris-EST, France in 2012. She is an associate professor at University of Johannesburg. Her research interests include Engineering Optimization, Artificial Intelligence, Control Systems and so on.