A deep learning-based framework for automatic detection of drug resistance in tuberculosis patients

Rehab Duwairi *, Abdullah Melhem

Jordan University of Science and Technology, Jordan

ARTICLE INFO

ABSTRACT

Tuberculosis (TB), caused by mycobacterium tuberculosis, is one of the most severe respiratory diseases that kill thousands of people annually around the world. Late diagnosis and drug resistance adversely affect the treatment plan. This paper introduces several deep learning models for TB diagnosis from CT scans. Specifically, for the detection of multi-drug resistance and for the detection of the TB types. These models are multi-channel models as they take, as input, the image frame, the mask frame and the gender/age data to make the final classification. Transfer learning based on VGG19 and ResNet neural networks were used for feature extraction from CT scans. The best performing model, for the classification of multi-drug resistance, was a three-channel model which used VGG19 for feature extraction and a cascade of convolutional and dense layers for classification with accuracy equals to 74.13% and AUC equals to 64.2%. By comparison, the best performing model for the TB type classification, employed the ResNet for feature extraction and a cascade of convolutional and dense layers for classification with accuracy equals to 53% and Kappa index equals to 34.3%. These results outperform the reported results in the literature for the same type of tasks. The dataset was obtained from the Image CLEFF 2018 data on TB.
© 2023 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Computers and Artificial Intelligence, Cairo University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Annually, there are huge numbers of people who contact tuberculosis (TB) and the mortality rate is high among these patients [1,2]. It is estimated that the number of TB infections reaches 8 million each year. Moreover, the annual number of mortalities in TB disease ranges from 2 to 3 million people [3]. Mycobacterium tuberculosis is the bacteria responsible for causing TB. These bacteria usually attack the lungs, but sometimes other areas of the body such as the spine, brain, and kidneys. There are two main causes of the occurrence of tuberculosis namely: TB disease and latent TB infection (LTBI). If this disease is not treated properly, it leads to death.

The bacteria that causes tuberculosis are transmitted from one person to another through the air, as this disease is transmitted through breathing or saliva. The bacteria settle in the lungs and then move through the blood to other regions such as the heart, kidneys, and others [4].

The condition of the person infected with the bacteria that cause tuberculosis may develop until the patient's body becomes resistant to medicines. This condition is serious for the infected persons and is called drug-resistant TB or multidrug-resistant (MDR). Misuse of tuberculosis treatments, such as not completing the course of treatment, poor medication, or prescribing an inappropriate drug leads to multidrug-resistant conditions.

There are two types of drug resistance: Multidrug-Resistant (MDR) TB and Extensively Drug-resistant (XDR) TB. In the case of MDR, the patient's body becomes resistant to Isoniazid and Rifampin medicines. Extensively drug-resistant TB is a very rare type of tuberculosis where the patient's body becomes resistant to the most important collection of drugs that are given to a TB patient such as Isoniazid, Rifampin Amikacin, Kanamycin, and Capreomycin [5]. The treatment of tuberculosis is very expensive, although the severe cases of this disease are rare. Other diseases may follow the recovery from tuberculosis, in some cases the per-

---

Production and hosting by Elsevier

son recovering from tuberculosis suffers from diseases such as hearing loss, depression, or hepatitis.

The relationship between tuberculosis and the cost is related to the patient's body resistance to drugs. The more drug-resistant the patient's body is, the higher the treatment cost. There are a large number of health organizations concerned with treating tuberculosis and limiting its spread. These organizations include TB Epidemiologic Studies Consortium (TBESC), TB Trials Consortium (TBTC), Behavioral and social science research, and Epidemiologic and Economic Modeling [6]. Therefore, early diagnosis of TB has many advantages. X-ray and CT images, for example, help in the diagnosis of tuberculosis. Usually, images are used in the diagnosis of tuberculosis after (i) initial negative bacteriological testing, (ii) a course of broad-spectrum antibiotics, and (iii) a second negative round of bacteriological testing. Early detection of TB disease reduces the possibility of using inappropriate medicine for TB patients [7].

Recently, there is a great interest in utilizing machine learning and deep learning techniques for aiding in medical diagnosis [8,9]. Automatic diagnosis of TB, based on medical images, is vital in the cases where there is lack in human expertise or there is a large number of images that require analysis, such as in outbreaks of the disease. The World Health Organization recommends avoiding group screening of TB patients and turning to a more effective approach. The recommended approach is the use of Poster anterior chest radiography for selected groups that are classified as high-risk [10].

There is a clear weakness in diagnosing TB patients. TB patients can be classified into two different cases, namely, Multidrug resistance and Drug sensitivity (DS). In each case, the patient is treated differently. TB specialists face many challenges in differentiating between the two cases. In addition, the rate of error in differentiating between them is large. Moreover, it is difficult for TB specialists to determine the type of TB in an infected person. There are similarities in the symptoms that appear in more than one type of TB [11]. Tuberculosis appears in five types, and each type of TB has different patterns and treatment methods. The aim of determining the condition of the TB patient early and determining the type of TB disease in that patient is to reduce the cost of treatment and increase the chances of recovery from this disease.

Machine learning and deep learning have proven to be successful is several tasks in the medical field such as drug discovery [12], detection of TB severity [9], segmentation of CT and MRI images to localize abnormalities [13], drug resistance [14], drug repurposing [15], and drug delivery [16]. This work reports several deep neural networks which were designed for the task of MDR detection and the task of TB Type (TBT) classification. The main research question is to assess whether deep neural networks can provide diagnosis quality comparable to the specialist diagnosis and whether the quality of the results improves as the number of input channels to the networks increases. Secondary research questions were also considered in this paper. Specifically, what type of data preprocessing is suitable for the above task? Are standard image augmentation techniques suitable for MDR detection and TBT classification? Is transfer learning suitable for this task? Do CT images alone provide enough features for the model to be skilful? These models have been tested using the Image CLEFF-2018 dataset [17]. The built models and reported results prove that deep learning models are capable of providing high quality classification. They also show that multi-channel or multi-modal networks are superior to single modal networks. Image frame, the mask frame and gender/age data are used as three types of input channels to help in the classification of TB patients.

This paper is organized as follows. Section 2 summarizes related work. Section 3, by comparison, describes the materials and methods that were followed. Section 4 discusses the experiments and reported results. Section 5 discusses the skills and limitations of the designed deep neural networks. Finally, Section 6 presents the conclusion of this work.

2. Related work

This section reviews existing works which have targeted TB detection. These reported works mostly operate on 3D CT scans or chest X-ray images. Also, most of them agree that image masks, which are forms of image segmentation filters, are important to enhance accuracy of deep learning models. Several of these works have utilized pre-trained models such as AlexNet, VGG net or ResNet for feature extraction. Most of the datasets were taken from the Image CLEFF competition. Thus our work shares several aspects with the following works.

Pattnaik et. al [18] used 3D CT scans to classify patients with TB. Every 3D image was transformed into 20 2D slices. They reduced the size of the image to 150x150 pixels whereas the original size of the images is 512x512 pixels. Fixing the number of slices per CT scan may not be suitable for all images. In the case where the CT scan is not enough to extract 20 slices, the authors copied or duplicated slices. In our work, we varied this number depending on the length of the video corresponding to the CT scan. The work reported in [19], by comparison, segmented the image to extract regions of interest. Specifically, the frontal chest images were segmented to extract the area which shows TB infection. To extract the features from the resulting images, three pre-trained models were used, namely, GoogLeNet [20], ResNet [21] and VGG net [22]. Subsequently, the extracted features were passed to SVM for classification.

On the other hand, Liauchuk et al. [23] used 130 manually segmented images for lungs independently of Image CLEFF dataset. The authors compute a similarity measure between the target segmentation and reference segmentation in order to select the top 5 most similar reference segmentations. Then the five chosen similar masks with the corresponding mask were inserted into Elastix software [24]. The means of averaging similarities were used in order to obtain the final mask [25]. They employed all of the algorithms suggested in [26] to extract additional information from the data that Image CLEF provide to differentiate between MDR and DS. Their reported accuracy was equal to 49.3 % and their reported AUC was equal to 54.15 %.

Lakhani et al. in [27] use 1007 poster anterior chest radiographs to determine if the scan refers to TB or not. They split this data into three collections (training, validation, and testing) of three different percentages (68.0 %, 17.1 %, and 14.9 %), respectively. Two deep learning models were trained to classify these images. Specifically, they used AlexNet and GoogLeNet pre-trained models in two different settings. In the first setting, the weights of the pre-trained models were not used whereas in the second setting the weights from their training on ImageNet data were used. In the cases where the trained models produce different predictions for the same input, the input data is transferred to specialized doctors to evaluate the situation. Image rotation and contrast enhancement were used as for data augmentation. This technique improves the performance of the models [28].

The work report in [29] utilizes a 7-layer CNN, five of which were Conv2D layers, to classify images into bacillus negative and bacillus suspect. A second network, similar to the first one, was designed to classify, again, the images labelled by the first network as bacillus suspect. The second network classifies bacillus suspect images into either bacillus positive or bacillus negative images. This cascade of classifiers increases the confidence in the output of the model.

Using a small number of data (CT TB Images) published by ImageCLEF2017 [30], convolutional neural network models were constructed in collaboration with SVM to differentiate between DS and MDR [11]. The data used consist of 230 CT images as a training sample and 214 images as a test sample. The slices were extracted from the 3D images (CT scans), and these slices were passed to a preprocessing module. Slices that contain small parts of the lung were removed. Masks were used to correct the boundaries of the lungs that appear in the corresponding CT image. They use the mask to facilitate the selection process for the slices that will be included in the training process. The mask was also used to cut the inner part of the lungs, as it is the region of interest. This study is based on dividing the slices into small sections of size equal to 64x64 pixels (patches). After that, these patches are passed on to six convolutional layers, followed by SVM classifier. The classifications of the patches belonging to the same slice are combined in order to issue the final classification label. The accuracy of the system they developed was equal to 91.11 %. These results cannot be generalized as their work was limited to a small amount of data.

The work reported in [17] has presented a deep learning models to distinguish MDR from DS in TB patients. The reported accuracy was equal 61.44 % and this was the highest accuracy in the competition; the work also achieved an AUC equals to 61.14 % and this was the second best value. Med2image (Python3 Library) [31] was used to convert images from NIFTI to an easy-to-handle format (i.e. PNG or JPEG). These images have 3 planes, so they use them in the coronal plane in order to extract more slices containing areas of the abnormal lung segments. Since TB attacks the upper regions of the lungs, the coronal plane was used. The size of the resulting slice has been reset to 512x512 as the Med2image produces outputs with different dimensions. From 512 coronal images for each patient, 150 to 350 images were selected in order to exclude those that do not contain parts of the lungs. The Fastai library [32] was used to increase the amount of data (Data Augmentation). Image masks were not used in this study.

Ahmed et al. in [33] used the data available in the Image CLEF 2018 competition intelligently. Ahmed and his team extracted new data from the available data [34] and used this data to train deep learning models. The first step in Ahmed's et. al. work was identical to our work. The slices were extracted from the 3D CT image scan and from the mask corresponding to each image. In the second step, each slice was used with the corresponding mask. In order to extract a new image containing the region of interest. After that, the data were restored to their original state, but for the bronchial region only. In other words, a three-dimensional model of the bronchial tubes was built. In this way, this study was able to reach fifth place in determining the type of TB using the data of the Image CLEF 2018 competition. The use of the mask only at the stage of defining the region of interest and not using it in the inputs of the deep learning model leads to a reduction in the number of extracted features. This leads to the weakening of the model's ability to differentiate between the five types of TB.

The gray-colored images contain pixel values ranging from 0 to 255. From this point of view, Stefan et al. [35] modified the slices extracted from the 3D images. The extracted slices were converted from gray to colored slices containing the three basic colors (i.e. RGB). This has increased the number of parameters in the model. This type of boosting also leads to an increase in the number of the extracted features. The more accurate features led to greater accuracy in the performance of the models. A pre-trained model, GoogleNet, was used in order to differentiate between extracted slices. By this, the authors were able to reach the third and eighth place in Image CLEF Tuberculosis 2017 competition. It is not possible to adopt this method in general, but in this competition, it achieved clear success. This method may suffer from overfitting,

but this was avoided by using multi-scale cropping augmentation algorithm.

The aim of Cao et al. [36] study was to determine the type of TB (Categorical Classification) after determining whether the test (X-ray image) is for an infected person or a healthy person (Binary Classification: normal and abnormal). In order to reach this goal, it was necessary to find a large and real-world database containing various examples of tuberculosis tests. In order to build the database, a working group was established consisting of a group of researchers in addition to doctors specializing in Tuberculosis. A large number of real-world TB patient scans were collected by this team, in addition to the use of the image CLEF from the 2004–2013 dataset that was published as open-source data. Image CLEF from 2004 to 2013 dataset contains over 400,000 scans with diagnostic annotations, search topics, and related judgments.

Singh et al. [37] have used an anti-aliased CNN to detect TB lesions in the lungs using image segmentation. They have trained and tested their models using Image CLEFF 2019 dataset. The idea is to employ shift invariant network as it has potential to spot the abnormal regions in the CT images.

The study reported in [38] uses a specialized convolutional neural network, which was introduced in [39], to classify chest X-ray images into drug-resistant and drug-sensitive TB. The Portal dataset was used to train the models. The work reported in [40] employs four state of the art CNN-based networks to classify CT images into military, infiltrative, caseous, tuberculoma and cavitary TB types. The dataset was in-house collected CT images. In total, the dataset has 501 CT images with active TB and 501 images with inactive TB. The best performing model was the one which uses V-NET for segmentation and Inception + ResNet50 for classification with accuracy equals 93.4 %.

The work reported in [41] has experimented with three networks to classify drug resistance in TB patients. The dual attention dilated residual network (DADRN), the Bag of Visual Words and the Sparse Code Book algorithms were used. DADRN was the best performing algorithm. The dataset consists of CT images for 400 patients.

By comparison, the authors in [42] have utilized five pre-trained neural networks in addition to a new CNN network designed from scratch. These pre-trained neural models include Inception-v3, Xception, ResNet50, VGG19 and VGG16. The task is a binary classification, i.e. the goal is to determine if a subject has TB or not based on the chest X-ray images. Two publically datasets were used. The first one is provided by Shenzhen, China which includes chest X-rays for 80 healthy individuals and 58 TB patients. The second dataset was from Montgomery county, MD, USA which consists of 326 X-ray images for health subjects and 336 images for subjects with TB. The best performing models were Xception, ResNet50 and VGG16 with accuracy equal to 90 %.

Nafisah and Muhammad [43] used several CNN-based models, such as Xception, Inception-ResNet-V2, ResNeXt-50, MobileNet and EfficientNet in a binary classify task. This means the goal was to identify patients who have TB from those who do not. Three datasets were used for training the previous models. The best performing model was the Efficient Net-B3 with accuracy equal to 99.9 %.

Ravi et al [44] used EfficientNetB0, EfficientNetB1, and EfficientNetB2 pre-trained neural networks to detect lung diseases based on chest X-ray images. First, the features extracted using the previous three networks were fused together and then passed into a cascade of dense layers. An ensemble classifier, which consists of Random Forest and SVM, was used in the first stage and Logistic Regression was used in the second stage. The reported task is a multiclass classification task, where the aim is to classify patients with pneumonia, TB and Covid-19. The best reported accuracy in the case of pneumonia was 98 %, for TB the accuracy

was 99 % and for Covid-19 the accuracy equals 98 %. This work has used the Pediatric pneumonia CXR dataset from Guangzhou Women and Children's Hospital; the TB dataset from [45], and the Covid-19 dataset from [46].

The work reported in [47] uses Attention Branch ResNet neural network to detect lesions in the lungs using CT images of patients' chests. The model has several phases, where in the first phase, the CT images are screened to distinguish normal from with TB images. In the second phase, objects (i.e. lesions) are detected and localized. In the third phase, the lesions are classified into six types in addition to active/inactive TB type. The last phase assesses the severity of the TB. An in-house dataset of CT images was used for training. Three independent datasets were used for testing the suggested model.

Ying et al. [48] aimed at classifying patients as having nontuberculous mycobacteria pulmonary disease (NTM-PD) or pulmonary TB (PTB) using *T*-SPOT test (enzyme-linked test) and deep learning model (based on CT images). The deep learning model uses a U-Net network for image segmentation and Inception-ResNet for image classification. The authors report that the best results achieved when combining the *T*-SPOT and deep learning methods.

Prasitpuriprecha et al. [49] suggests a decision support system for TB/non-TB detection and drug-resistant categorization. The ensemble consists of EfficientNetB7, mobileNetV2, and DenseNet121 networks. A total of 7008 chest X-ray images were compiled from seven existing datasets. The best reported accuracy on test data is 92.6 %. However, the model's accuracy was equal to 96.8 % on random test images chosen from the compiled dataset.

To summarize, numerous works have targeted the discovery of TB drug resistance and TB types. Machine learning and deep learning are successful methods for such tasks. Moreover, most of these works agree on using the data provided by Image CLEFF and on using techniques to better capture abnormalities in the images such as smart segmentation or using image masks (as it is the case with this work). Table 1 summarizes recent related work that targeted the classification of TB using images of different modalities.

3. Materials and methods

The proposed methodology consists of the following phases:

- Data preparation and pre-processing: This phase is responsible for preparing the TB datasets to be usable by the deep neural models. This phase is described in Section 3.1.
- Feature extraction: This phase is concerned with extracting features from image and mask frames and from gender/age data. This phase is described in Section 3.2.
- Training deep neural models on the data related to the MDR task and on the data related to the TBT task. This phase is described in Section 3.3.
- Model evaluation using test data for both tasks. This phase is described in Section 4.

3.1. Dataset

The data provided by Image CLEF-2018 Tuberculosis [17] was used in this paper. It is a well-known dataset and has been used by researchers targeting building computational models for TB analysis. The data consists of 3D CT images stored in NIFTI file format with .nii.gz file extension. The 3D CT images contain image slices with size equal to 512 × 512 pixels per slice and the number of slices ranging from 50 to 400 slices. The data also contain image masks – where every 3D CT image is a companied by a 3D mask that contains additional information; mainly about the location

of the abnormal pattern. MDR data also have the age and gender of the patients. TBT data only have the age of the patients. This helped in designing multi-modal or multi-channel deep neural networks.

The labels of the test data are not available in all tasks in this competition. The predictions of the trained models must be delivered to the organizers of this competition. These are subsequently evaluated by the competition organizers and they communicate back only the accuracy of the models for 10 attempts at most. As this work aimed to investigate several deep learning models for several tasks, this form of assessment was not suitable. Thus, only the training data of CLEF-2018 was used as this subset of the data has labels. Table 2 shows the number of images used for every task. The column with the heading "CLEFF training images" represents the images taken from CLEFF-2018 for MDR detection (259 images) and for TBT classification (801 images). These are subsequently divided into training and testing subsets as indicated by the third and fourth columns. The 3D CT images were transformed into 2D images, to be suitable for the used deep learning models. The fifth and sixth columns display the number of 2D frames extracted from the CT images. This transformation was performed by first turning the CT image into a video using Mango software [50], subsequently, frames are extracted from this video; all frames have the same label as the original images from which they were extracted. The number of frames extracted from every video ranges from 29 to 31 depending on the length of the video. The lengths of these generated videos range from 3 to 10 seconds. This increased the number of frames (2D images) without distorting the original images.

This work argues that standard image augmentation techniques are not suitable for MDR detection and TBT classification. Specifically, rotation of the 3D images will create duplicates and may lead to overfitting. Flipping, on the other hand, leads to pixel distortion and this is not suitable for the targeted tasks which aim to identify the abnormal patterns in the TB CT scans. Finally, cropping divides a given image into several images. As the abnormal pattern typically consumes a small portion of the image, cropping may result in several images that do not contain this abnormal pattern and yet has the same label as the original image; this misleads the classifiers. The suggested method of image augmentation proves its superiority. The image masks were handled exactly as their corresponding images.

3.2. Preprocessing and feature extraction

After the 3D images were transformed into 2D images (2D frames), every frame is treated as a 2D matrix of pixels. The values in these pixels were normalized by diving them by 255. Also, all training frames were combined into one matrix of dimension (F, X, Y, C), where F represents the number of frames, X represents the X-dimension, Y represents the Y-dimension, and C represents the number of channels respectively. The same arrangement was followed for image masks. For feature selection, several methods have been experimented with such as an in-house designed CNN, transfer learning using VGG19, ResNet, and Xception neural networks. For the MDR detection tasks, VGG19 proved its superiority over other methods. For the TBT classification task, ResNet was superior compared to other options. The features of gender and age were extracted using a cascade of dense and dropout layers. The three types of input, namely image frame, mask frame and gender/age, were concatenated to form one vector which captures the features of every input. These vectors are subsequently fed into the subnetwork responsible for the classification task.

Table 1
A summary of recent related work on TB classification.

Related work	Year	Approach	Data	Obtained results
[37]	2020	U-Net, FPN and LinkNet for image segmentation and anti-aliased CNN for TB classification	CLEFF 2019 TB dataset	The work report only segmentation results. The Dice Coefficient equals to 93.43 % and IoU equals 88.73 %
[39]	2021	Tailored CNN	Portal dataset (2,973 images)	The best AUC was equal to 0.66 – 0.67
[40]	2021	V-NET for segmentation and Inception + ResNet50 for classification.	CT images (501 images with active TB, 501 CT images with inactive TB).	The best reported accuracy equals 93.4 %
[41]	2022	Dual attention dilatedresidual network (DADRN)	CT images of 400 patients	Best precision is 93.8 % and best recall is 86.4 %
[42]	2022	Inception-v3, Xception, ResNet50, VGG19, VGG16 and in-house designed CNN	Chest X-ray images. Shenzhen, China (80 healthy individuals and 58 TB patients) and Montgomery county, MD, USA (326 X-ray images for healthy subjects and 336 images for subjects with TB)	Best reported accuracy was equal to 90 %.
[43]	2022	Several CNN-based models, such as Xception, Inception-ResNet-V2, ResNeXt-50, MobileNet and EfficientNet	Chest X-ray images. Shenzhen, China (80 healthy individuals and 58 TB patients), Montgomery county, MD, USA (326 X-ray images for healthy subjects and 336 images for subjects with TB), and Belarus dataset (304 subjects with TB)	The best reported accuracy was achieved when using EfficientNet-B3 with accuracy equal to 99.9 %.
[44]	2022	EfficientNetB0, EfficientNetB1, and EfficientNetB2 for feature extraction. Random Forest, SVM and Logistic Regress for classification	Chest X-ray images. Pediatric pneumonia CXR dataset from Guangzhou Women and Children's Hospital (3662 normal images, 1570 images with pneumonia). The TB dataset from [45] (3800 normal images, 3800 images with TB, 800 images indicate sick but not TB). The Covid-19 dataset from [46] (4675 normal images, 2004 images with Covid-19).	The best reported accuracy: For pneumonia 98 %, for TB 99 % and for Covid-19 98 %
[47]	2022	Attention Branch ResNet	CT images. In-house dataset (dataset 1). 892 chest CT images. Dataset 2 (for testing). 99 CT images. Dataset 3 (for testing). 86 CT images. Dataset 4 (for testing). 171 CT images.	The best F1-score for dataset-2 was 96.2 %, for dataset-3 97.4 % and for Dataset-4 98.42 %
[48]	2022	T-SPOT plus U-Net (for segmentation) and Inception-ResNet (classification)	In-house CT images that includes 1049 cases (467 with NTM-PD and 582 with PTB).	The best precision was achieved when using *T*-SPOT combined with deep learning model: 91.7 % for NTM-PD and 89.8 % for PTB.
[49]	2023	ensemble consists of EfficientNetB7, MobileNetV2, and Dense-Net121 networks	Chest X-ray images compiled from Portal dataset, Montgomery County dataset, Shenzhen dataset and Kaggle dataset)	The best accuracy on test data equals 96.2 %

Table 2
Distribution of training and testing data.

Task	CLEFF training images	# of images used for training	# of images used for testing	# of extracted frames from taring data	# of extracted frames from testing data
MDR detection	259	201	58	5949	1716
TBT	classification	1008	801	200	32,459
8340					

3.3. Multi-modal deep learning models

Several neural networks were experimented with which use only the frames of the images without the masks or gender/age information. VGG16 was used for feature extraction, and several dense/dropout/batch-normalization layers were added to the model to classify the images into MDR or DS in the MDR detection task; and into Infiltrative, Focal, Tuberculoma, Miliary and Fibrocavernous in the case of the task of TBT classification. It was concluded that CT images alone are not sufficient and additional information, such as image masks and age/gender information, is needed. This leads to the development of multi-modal or multi-channel deep neural networks as will be explained next.

3.3.1. MDR classification models

This task aims to classify images to either having MDR or DS. Four deep learning models have been designed; these rely on CNN, VGG19 or Xception networks for feature extraction and rely on a set of convolution and dense layers for classification. These models take the frame, mask and gender/age as input and produce the classification as output.

Fig. 1 depicts the baseline model which have been used in this work. Here, an CNN network was used to extract features from the images and masks. Dense layers were utilized for extracting features from age and gender. Feature extraction layers were followed by several dense and dropout layers to reach the output layer – where the final class is determined (i.e. MDR or DS).

Fig. 2 depicts the second model for MDR detection. It employs VGG19 and Xception networks for feature extraction from image frames and mask frames, respectively. It also uses dense layers to extract features from age and gender data. These three modalities are combined (concatenated) and fed into several dense and dropout layers to make the final classification. This model achieved superior results when compared with the baseline model.

Fig. 3 shows the architecture of the model which employs only VGG19 for feature extraction from images and masks. It also relies on dense layers to extract features from the age and gender data. The three sets of features are concatenated and fed into the dense/-dropout layers for the final classification.

The model depicted in Fig. 3 was altered by dropping one convolutional layer from the channel that deals with image frames and by removing three convolutional layers from the channel that deals
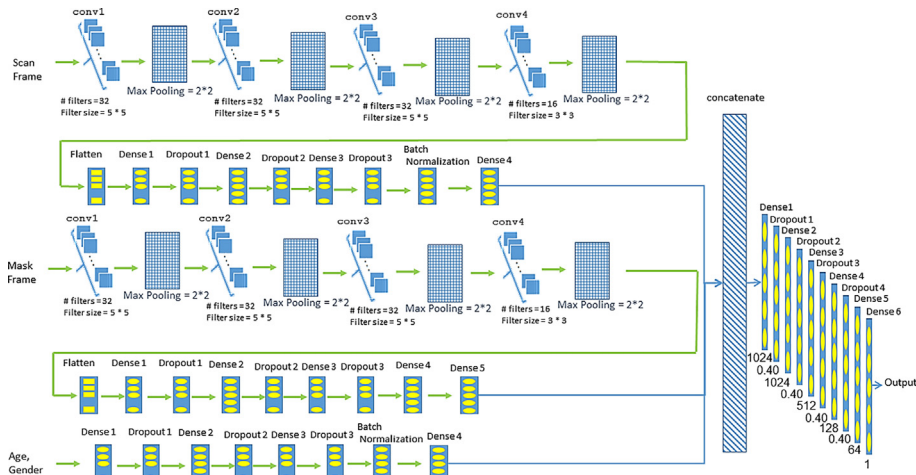
Fig. 1. The Architecture of the Multi-Channel CNN Model to detect MDR (baseline model).
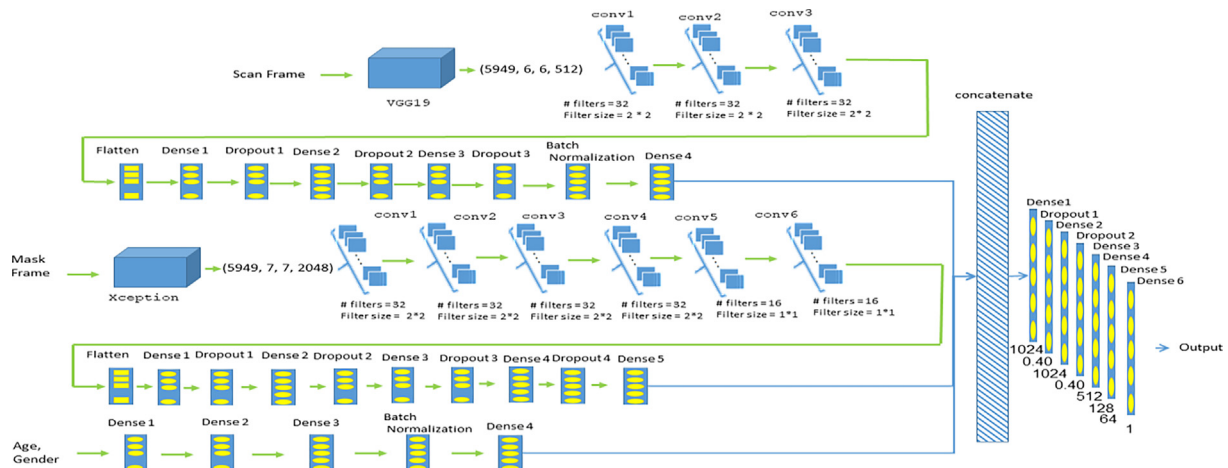

Fig. 2. The Architecture of the multi-channel Xception and VGG19 model to detect MDR.
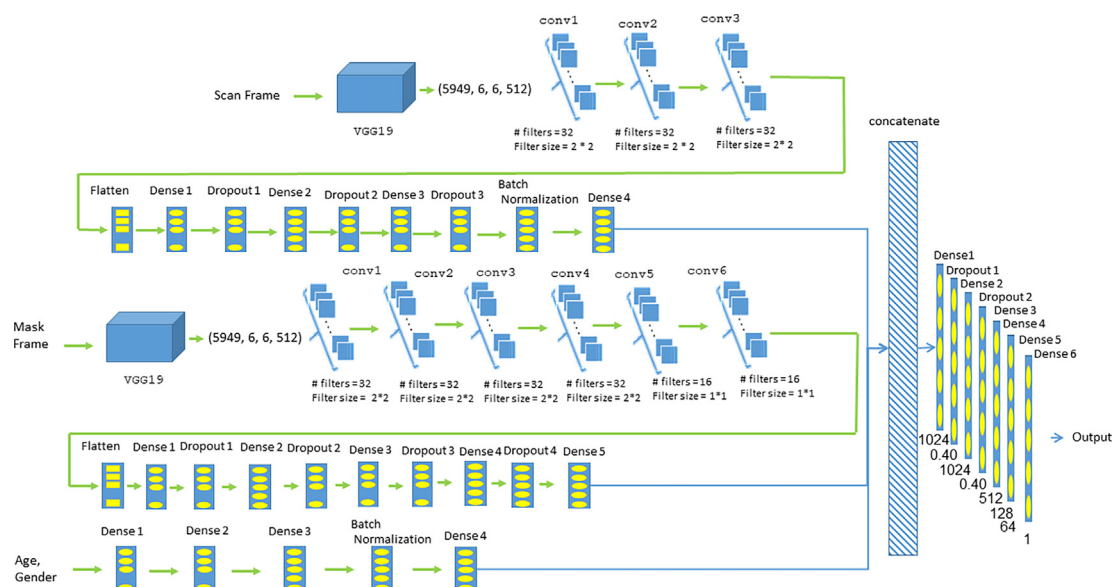

Fig. 3. The Architecture of the multi-Channel VGG19 V1 model to detect MDR.

with mask frames. This model is referred to as Multi-Channel VGG19 V2 model.

3.3.2. TBT classification

This task aims to classify an image into Infiltrative, Focal, Tuberculoma, Miliary and Fibro- cavernous classes. Four different neural networks were designed for this purpose. All of these models relied on different architectures of ResNet network for feature extraction. These models use the image and the mask frames in addition to age data. Gender information is not available for this task. ResNet50 was used in the baseline model. ResNet101 and ResNet150 were employed in the subsequent two models. The three input channels are concatenated and fed into the second half of the network for classification. Fig. 4 depicts the architecture of the best performing model; we call it multi-channel ResNet101-V2 model.

4. Experimentation and results analysis

This section describes the process of tuning the neural networks. Also, it explains the evaluation metrics which have been used to assess the skill of the models. Finally, the obtained results are reported; and comments on models' skill for both MDR detection and TBT classification tasks are presented.

4.1. Hyperparameters tuning

The skill of a given deep learning model relies on several parameters such as its architecture (number and type of layers), learning rate, optimizer, dropout rate and so on. The models were implemented using Keras API version 2.3.1. Numerous experiments were carried out to fine-tune the hyperparameters of the models. Table 3 shows the best performing hyperparameters for the MDR detection task. By comparison, Table 4 shows the values of the best performing parameters for the TB type classification task.

4.2. Evaluation metrics

In this work, the same evaluation metrics specified by CLEFF 2018 competition were used. Specifically, accuracy, AUC (Area

Table 3
The best performing hyperparameters for the MDR detection task.

MDR hyperparameters			
# of CNN layers	# of Dropout layers	# of Dense layers	# of epochs
5	6	19	20
# of filters	Percentages	# of neurons	Batch Size
32 (2*2)	0.30 & 0.40	1024 to 1	100
optimizers	Learning Rate	Loss Function	Activation Function
RMSprop	0.001	binary_crossentropy	Sigmoid

Table 4
The values of the best performing parameters for the TB type classification task.

| TBT hyperparameters			
# of CNN layers	# of Dropout layers	# of Dense layers	# of epochs
16	10	16	40
# of filters	Percentages	# of neurons	Batch Size
64 (2*2)	0.30 & 0.40	1024 to 5	50
optimizers	Learning Rate	Loss Function	Activation Function
Adam	0.001	categorical_crossentropy	Softmax

Under Curve) and Kappa index were employed. Formula (1) shows how to calculate the accuracy.

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

where TP: indicates true positives; TN: stands for true negatives; FP: means false positives; and FN: means false negatives.

AUC is the area that is located under a ROC line. This line uses true-positive-rate values at the x-axis and false-positive-rate values at the y-axis. The model is most skilful if this area reaches 1. By comparison, the model guesses the correct labels by chance when AUC equals 0.5. Finally, the model performs worse than random guessing when the AUC is less than 0.5. Formulae (2) and (3) show the equations for true-positive rate and false-positive rate.

$$True\ Positive\ Rate = \frac{TP}{TP + FN} \tag{2}$$
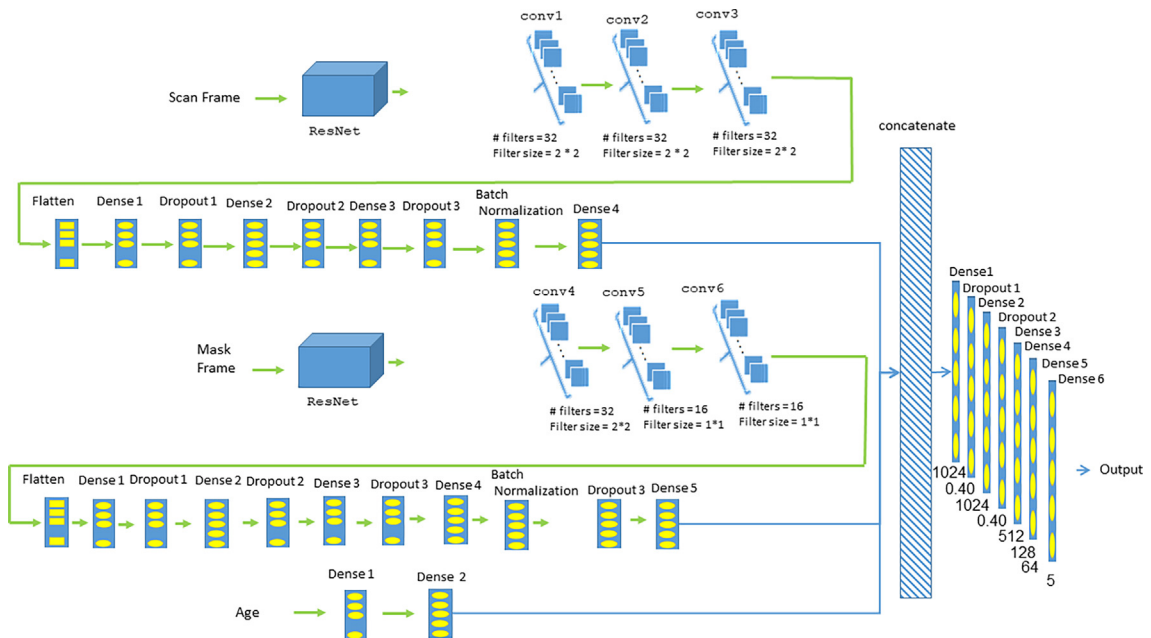
Fig. 4. The Architecture of the multi-channel ResNet101-V2 model to detect the type of TB.

$$False\,Positive\,Rate = \frac{FP}{FP + TN} \tag{3}$$

Cohen Kappa is a metric often used to assess the agreement between several annotators. When using two annotators to classify items into G mutually exclusive groups, the agreement between them can be measured using the Kappa index as stated in equations (4) and (5).

$$Kappa = P_0 - \frac{P_e}{1 - P_e} \tag{4}$$

$$P_e = P_{e1} + P_{e2} = P_{e1,target} * P_{e1,pred} + P_{e2,target} * P_{e2,pred} \tag{5}$$

where

- P0: is the overall accuracy of the model.
- Pe: the agreement between the predictions and the actual values (if happening by chance).
- Pe1: the probability of the predictions agreeing with actual values of a specific class (e.g. MDR class) by chance.
- Pe2: same as the Pe1 but on the other class (DS class).

Assuming that the two classifiers (model predictions and actual values) are independent. The Pe1 and Pe2 are computed by multiplying the proportion of the predicted class and the proportion of the actual class.

4.2.1. Evaluation of the MDR task

Table 5 shows the performance of the four models developed for this task using accuracy and AUC. The baseline model achieved an accuracy equals to 63.7 % while the best performing model was Multi-channel-VGG19-V2 with an accuracy equal to 74.13 % and AUC equal to 64.2 %. It is worth mentioning that the winners of CLEFF 2018 competition scored accuracy equals to 61.40 % ((San Diego VA HCS/UCSD team) and AUC equals to 61.78 % ((VISTA@UE-vora team). Table 5 shows that the suggested model scored higher accuracy and AUC. To be fair, these numbers are not directly comparable when the train-test split of data is different as it is the case here, nonetheless, it is still considered a strong indication to the skill of the suggested models.

4.2.2. Evaluation of TBT task

Table 6 shows the results obtained in the TBT task. The results are reported in terms of accuracy and Kappa index. The baseline model's accuracy equals to 52.2 % and its Kappa index equals to 31.8 %. Both Modified ResNet101-V1 and Modified ResNet101-V2 models achieved the same accuracy of 53 %. However, Modified ResNet101-V2′s Kappa index was higher than the first model and it was equal to 34.3 %. It is worth mentioning that the winner in image CLEFF-2018 scored an accuracy equal to 42.2 % and Kappa index equal to 23.12 % (UIIP_BioMed team). Our reported results supersede these results.

Table 5
Results for the MDR task.

Model Name	Accuracy for test data	AUC for test data
Multi-Channel CNN (Baseline)	63.7 %	50.1 %%
Multi-Channel VGG19 V1	68.95 %	53.9 %
Multi-Channel VGG19 V2	**74.13 %**	**64.2 %**
Multi-Model Xception and VGG19	70.6 %	56.6 %

Table 6
Results for the TBT task.

Model Name	Accuracy of test data	Kappa on test data
Modified ResNet50 (Baseline)	52.2 %	31.8 %
Modified ResNet101-V1	53 %	32.7 %
Modified ResNet101-V2	**53 %**	**34.3 %**
Modified ResNet152	49.5 %	28.4 %

5. Discussion and limitation

Multi-channel or multi-modal input has proved superiority to single modality. In this paper, multi-channel or multi-modal systems were built for MDR detection and TBT classification. The three modalities which were available in this dataset are: CT images of the lungs of the patients, image masks and age/gender data. The image mask is necessary to have as the traditional scene of tuberculosis (active or latent) is insufficient as tuberculosis transfers between both cases constantly. The conventional TB screening (tuberculin skin test and interferon-gamma release assay) is ineffective. Thus, it was necessary to use supplementary information and TB biomarkers to diagnose tuberculosis, which leads to the need for lung image segmentation (referred to as image mask).

Historically, interest in the gender of the TB patient was neglected, and gender was not recorded as part of TB related data. Lately, The World Health Organization issued a recommendation to record the patient's gender with his data. Research has demonstrated that males are more susceptible to tuberculosis than females. The susceptibility of TB ratio is 1:1.6 because males are more sensitive to the events that lead to infection with tuberculosis [51–53]. Also, other studies show that males over the age of 20 years have higher chances of being infected with TB [54].

Initially, deep neural models which rely only on the CT image (or image frames) were designed and investigated. However, these models did not achieve satisfactory results, and thus multi-channel neural networks were employed – where each modality is fed into a sub-network to extract the proper features. Afterwards, the three vector features are concatenated and fed to the network which is responsible for MDR detection or TBT classification.

On the other hand, the number of CT images provided by Image CLEFF-2018 is considered small and standard image augmentation techniques are not suitable for the targeted tasks. This motivated the authors to use a novel data augmentation – where the 3D CT image is converted in a video and subsequently frames are extracted from that video. This increased the number of images and at the same time did not distort the images, especially the regions which contain the abnormal patterns.

The models which have been designed have achieved higher accuracy, AUC and Kappa index values when compared with reported results in the literature.

6. Conclusion

This paper has discussed building deep neural networks for MDR-detection and TBT classification. For the MDR detection task, four neural networks were designed which process the image frame, the mask frame and the gender & age data of patients. The baseline model utilized an CNN network for feature extraction from image and mask frames, it also used dense layers to extract features from gender and age. The three sets of features are concatenated and fed into the network responsible for classifying images as DS or MDR. The subsequent models have utilized the powerful VGG19 or Xception networks for extracting features from the image and mask frames. The features extracted using VGG19 prove useful as the results on these models outperformed the base-

line model. The Multi-channel VGG19 V2 was the best performing model with accuracy equals to 74.13 % and AUC equals to 64.2 %. Our results outperform reported results in the field for this task.

By comparison, the TBT task is concerned with classifying an image as Infiltrative, Focal, Tuberculoma, Miliary, and Fibro-cavernous. To this end, four multi-channel neural networks were designed. These differ in the feature extraction phase and in the subnetwork responsible for classification. ResNet network has proven to be very successful for feature extraction. The modalities or channels for this task consist of image frame, mask frame and gender data; note that the age data is not available for this task. The best performing model was the Modified ResNet101-V2 with accuracy equals to 53 % and Kappa index equals to 34.3 %. Again these results outperform reported work on TBT task using the Image CLEFF-2018 data.

As a future work, dataset size should be increased to include new datasets which recently emerged. Other image modalities such as X-rays provide another option for future investigation. Investigating neural networks that can directly handle 3D images is an interesting extension to this work.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Orme I, Secrist J, Anathan S, Kwong C, Maddry J, Reynolds R, et al. Search for new drugs for treatment of tuberculosis. Antimicrob Agents Chemother 2001;45(7):1943–6. doi: https://doi.org/10.1128/AAC.45.7.1943-1946.2001.

[2] Xu G, Liu H, Jia X, Wang X, Xu P. Mechanisms and detection methods of Mycobacterium tuberculosis rifampicin resistance: The phenomenon of drug resistance is complex. Tuberculosis 2021;128:. doi: https://doi.org/10.1016/j.tube.2021.102083102083.

[3] Dye C, Scheele S, Dolin P, Pathania V, Raviglione MC. Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. JAMA 1999;282 (7):677–86. doi: https://doi.org/10.1001/jama.282.7.677.

[4] Brooks JV, Orme IM. Evaluation of once-weekly therapy for tuberculosis using isoniazid plus rifamycins in the mouse aerosol infection model. Antimicrob Agents Chemother 1998;42(11):3047–8. doi: https://doi.org/10.1128/aac.42.11.3047.

[5] Vergelli C, Cilibrizzi A, Crocetti L, Graziano A, Piaz VD, Wan B, et al. Synthesis and evaluation as antitubercular agents of 5-arylethenyl and 5-(hetero)aryl-3-isoxazolecarboxylate. Drug Dev Res 2013;74(3):162–72. doi: https://doi.org/10.1002/ddr.21057.

[6] Rosenthal M, Fisher B. Tuberculosis: ancient history, modern scourge. J Infectious Diseases Prev Med 2013. doi: https://doi.org/10.4172/2329-8731.1000104.

[7] Uplekar MW, Shepard DS. Treatment of tuberculosis by private general practitioners in India. Tubercle 1991;72(4):284–90. doi: https://doi.org/10.1016/0041-3879(91)90055-w.

[8] Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016;316 (22):2402–10. doi: https://doi.org/10.1001/jama.2016.17216.

[9] Xiaohong G, James-Reynolds C, Currie E. Analysis of tuberculosis severity levels from CT pulmonary images based on enhanced residual deep learning architecture. Neurocomputing 2020;392:233–44.

[10] Tamura T, Suganuma N, Hering KG, Vehmas T, Itoh H, Akira M, et al. Relationships (I) of international classification of high-resolution computed tomography for occupational and environmental respiratory diseases with the ILO international classification of radiographs of pneumoconiosis for parenchymal abnormalities. Ind Health 2015;53(3):260–70.

[11] Gao XW, Qian Y. Prediction of multidrug-resistant TB from CT pulmonary images based on deep learning techniques. Mol Pharm 2018;15(10):4326–35. doi: https://doi.org/10.1021/acs.molpharmaceut.7b00875.

[12] Carracedo-Reboredo P, Liñares-Blanco J, Rodríguez-Fernández N, Cedrón F, Novoa FJ, Carballal A, et al. A review on machine learning approaches and trends in drug discovery. Comput Struct Biotechnol J 2021;19:4538–58. doi: https://doi.org/10.1016/j.csbj.2021.08.011.

[13] Furtado P. Loss, post-processing and standard architecture improvements of liver deep learning segmentation from Computed Tomography and magnetic resonance. Inf Med Unlocked 2021;24. doi: https://doi.org/10.1016/j.imu.2021.100585.

[14] Ekpenyong ME, Edoho ME, Udo IJ, Etebong PI, Uto NP, Jackson TC, et al. A transfer learning approach to drug resistance classification in mixed HIV dataset. Inf Med Unlocked 2021;24:. doi: https://doi.org/10.1016/j.imu.2021.100568100568.

[15] Urbina F, Puhl A, Ekins S. Recent advances in drug repurposing using machine learning. Curr Opin Chem Biol 2021;15(65):74–84. doi: https://doi.org/10.1016/j.cbpa.2021.06.001. PMID: 34274565.

[16] He S, Leanse L, Feng Y. Artificial intelligence and machine learning assisted drug delivery for effective treatment of infectious diseases. Adv Drug Deliv Rev 2021. doi: https://doi.org/10.1016/j.addr.2021.113922.

[17] Cid YD, Liauchuk V, Kovalev V, Müller H. Overview of Image CLEF Tuberculosis 2018-Detecting Multi-Drug Resistance, Classifying Tuberculosis Types and Assessing Severity Scores. In CLEF (Working Notes). 2018 Sep., http://clef2018.clef-initiative.eu/index.php?page=Pages/labs_info.html#imageclef.

[18] Pattnaik A, Kanodia S, Chowdhury R, Mohanty S. Predicting Tuberculosis Related Lung Deformities from CT Scan Images Using 3D CNN. In CLEF (Working. Notes) 2019. , https://ceur-ws.org/Vol-2380/paper_144.pdf.

[19] Saxena S, Shukla S, Gyanchandani M. Pre-trained convolutional neural networks as feature extractors for diagnosis of breast cancer using histopathology. Int J Imaging Syst Technol 2020;30(3):577–91.

[20] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. "Going deeper with convolutions". In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9), 2015.

[21] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. p. 770–8.

[22] Howard J, Gugger S. Fastai: a layered API for deep learning. Information, 11(2), p.108, 10.3390/info11020108.

[23] Liauchuk V, Kovalev V. ImageCLEF 2017: Supervoxels and Co-occurrence for Tuberculosis CT Image Classification. 2017, In CLEF (Working Notes). https://ceur-ws.org/Vol-1866/paper_146.pdf.

[24] Elastix Software, https://elastix.lumc.nl/, last accessed 5-Sep-2021.

[25] Klein S, Staring M, Murphy K, Viergever MA, Pluim JP. Elastix: a toolbox for intensity-based medical image registration. IEEE Trans Med Imaging 2009;29 (1):196–205.

[26] Kalinovsky A, Liauchuk V, Tarasau A. Lesion detection in CT images using deep learning semantic segmentation technique. Int Arch Photogrammetry Remote Sens Spatial Information Sci 2017;42:13.

[27] Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology 2017;284(2):574–82.

[28] Wu R, Yan S, Shan Y, Dang Q, Sun G. Deep image: Scaling up image recognition. arXiv preprint, arXiv:1501.02876, 7(8), 2015.

[29] Kant S, Srivastava MM. Towards automated tuberculosis detection using deep learning. In: 2018 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE; 2018. p. 1250–3.

[30] Gentili A. ImageCLEF2018: Transfer Learning for Deep Learning with CNN for Tuberculosis Classification. In CLEF (Working Notes), 2018. https://ceur-ws.org/Vol-2125/paper_145.pdf.

[31] Pienaar R. (2020). https://github.com/FNNDSC/med2image. Retrieved from https://github.com/FNNDSC/med2image, last accessed 5-Sep-2021.

[32] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint, 2014, arXiv:1409.1556.

[33] Ahmed S, Md Sk O, Jayatilake M, Goncalves T, Rato L. Texture analysis from 3D model and individual slice extraction for tuberculosis MDR detection, type classification and severity scoring. Appears in collection. INF - Publicações - Artigos em Revistas Internacionais Com Arbitragem Científica 2018. , http://hdl.handle.net/10174/24879.

[34] Ionescu B, Müller H, Villegas M, de Herrera AGS, Eickhoff C, Andrearczyk V, et al. Overview of ImageCLEF 2018: Challenges, datasets and evaluation. In: International Conference of the Cross-Language Evaluation Forum for European Languages. Cham: Springer; 2018. p. 309–34.

[35] Stefan LD, Cid YD, del Toro OAJ, Ionescu B, Müller H. Finding and Classifying Tuberculosis Types for a Targeted Treatment: MedGIFT-UPB Participation in the ImageCLEF 2017 Tuberculosis Task. In CLEF (Working Notes). 2017. https://ceur-ws.org/Vol-1866/paper_152.pdf.

[36] Yu C, Liu C, Liu B, Brunette MJ, Zhang N, Sun T, et al. Improving tuberculosis diagnostics using deep learning and mobile health technologies among resource-poor and marginalized communities. In: 2016 IEEE first international conference on connected health: applications, systems and engineering technologies (CHASE). IEEE; 2016. p. 274–81.

[37] Singh J, Tripathy A, Garga P, Kumara A. Lung Tuberculosis detection using anti-aliased convolutional networks. International Conference on Smart Sustainable Intelligent Computing and Applications. Procedia Computer Science 2020;173:281–90.

[38] Ureta J, Shrestha A. Identifying Drug-Resistant Tuberculosis from Chest X-Ray Images Using a Simple Convolutional Neural Network. Journal of Physics: Conference Series, Volume 2071, International Conference on Biomedical Engineering (ICoBE 2021) 14, 15 September 2021, UniMAP, Malaysia (Virtual).

[39] Pasa F, Golkov V, Pfeiffer F, Cremers D, Pfeiffer D. Efficient deep network architectures for fast chest X-ray tuberculosis screening and visualization. Sci Rep 2019;9:6268.

[40] Li X, Zhou Y, Du P, Lang G, Xu M, Wu W. A deep learning system that generates quantitative CT reports for diagnosing pulmonary Tuberculosis. Appl Intell 2021;51:4082–93.

[41] Jiang Y, Zhao X, Fan Z. Intelligence Classification Algorithm-Based Drug-Resistant Pulmonary Tuberculosis Computed Tomography Imaging Features and Influencing Factors. Computational Intelligence and Neuroscience. Volume 2022, Article ID 3141807, 10.1155/2022/3141807, 2022.

[42] Showkatian E, Salehi M, Ghaffari H, Reiazi R, Sadighi N. Deep learning-based automatic detection of tuberculosis disease in chest X-ray images. Polish J Radiol 2022; 87: e118-e124, 10.5114/pjr.2022.113435, 2022.

[43] Nafisah S, Muhammad G. Tuberculosis detection in chest radiograph using convolutional neural network architecture and explainable artificial intelligence. Neural Comput Applic 2022. doi: https://doi.org/10.1007/s00521-022-07258-6.

[44] Ravi V, Acharya V, Alazab M. A multichannel EfficientNet deep learning-based stacking ensemble approach for lung disease detection using chest X-ray images. Clust Comput 2022. doi: https://doi.org/10.1007/s10586-022-03664-6.

[45] Liu Y, Wu Y H, Ban Y, Wang H, Cheng M M. Rethinking computer-aided tuberculosis diagnosis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2646–2655, 2020.

[46] El-Shafai W, Abd El-Samie F. Extensive COVID-19 X-ray and CT chest images dataset. Mendeley Data 2020;V3,. , https://data.mendeley.com/datasets/8h65ywd2jr.

[47] Yan C, Wang L, Lin J, Xu J, Zhang T, Qi J, et al. A fully automatic artificial intelligence–based CT image analysis system for accurate detection, diagnosis, and quantitative severity evaluation of pulmonary tuberculosis. Eur Radiol 2022;32:2188–99. doi: https://doi.org/10.1007/s00330-021-08365-z.

[48] Ying C, Li X, Lv S, Du P, Chen Y, Fu H, et al. T-SPOT with CT image analysis based on deep learning for early differential diagnosis of nontuberculous mycobacteria pulmonary disease and pulmonary tuberculosis. Int J Infect Dis 2022;125:42–50.

[49] Prasitpuriprecha C, Jantama SS, Preeprem T, Pitakaso R, Srichok T, Khonjun S, et al. Drug-resistant tuberculosis treatment recommendation, and multi-class tuberculosis detection and classification using ensemble deep learning-based system. Pharmaceuticals 2023;16:13. doi: https://doi.org/10.3390/ph16010013.

[50] Lancaster JL, Cykowski MD, McKay DR, Kochunov PV, Fox PT, Rogers W, et al. Anatomical global spatial normalization. Neuroinformatics 2010;8(3):171–82.

[51] Karim F, Ahmed F, Begum I, Johansson E, Diwan VK. Female-male differences at various clinical steps of tuberculosis management in rural Bangladesh. Int J Tuberculosis Lung Disease 2008;12(11):1336–9.

[52] Jimenez-Corona ME, Garcia-Garcia L, DeRiemer K, Ferreyra-Reyes L, Bobadilla-del-Valle M, Cano-Arellano B, et al. Gender differentials of pulmonary tuberculosis transmission and reactivation in an endemic area. Thorax 2006;61(4):348–53.

[53] Marçôa R, Ribeiro AI, Zão I, Duarte R. Tuberculosis and Gender-Factors influencing the risk of tuberculosis among men and women by age group. Pulmonology 2018;24(3):199–202.

[54] World Health Organization. WHO report 2009-global tuberculosis control epidemiology, strategy, financing. World Health Organization; 2009.