Prediction of multi-drug resistant TB from CT pulmonary Images based on deep learning techniques

Xiaohong W. Gao [1*], *Yu Qian* [2]

[1*] Department of Computer Science, Middlesex University, London NW4 4BT, UK.

x.gao@mdx.ac.uk

[2] Cortexica Vision Systems, London, UK

yu.qian@cortexica.com

ABSTRACT

While tuberculosis (TB) disease was discovered more than a century ago, it has not been eradicated yet. Quite contrary, at present, TB constitutes one of top 10 causes of death and has shown signs of increasing. To complement conventional diagnostic procedure of applying microbiological culture that takes several weeks and remains expensive, high resolution computer tomography (CT) of pulmonary images has been resorted to not only for aiding

clinicians to expedite the process of diagnosis but also for monitoring prognosis when administrating antibiotic drugs. This research undertakes the investigation of predicting multi-drug resistant (MDR) patients from drug sensitive (DS) ones based on CT lung images to monitor the effectiveness of treatment. To contend with smaller datasets (i.e. in hundreds) and the characteristics of CT TB images with limited regions capturing abnormities, patch-based deep convolutional neural network (CNN) allied to support vector machine (SVM) classifier is implemented on a collection of datasets from 230 patients obtained from ImageCLEF 2017 competition. As a result, the proposed architecture of CNN+SVM+patch performs the best with classification accuracy rate at 91.11% (79.80% in terms of patches). In addition, hand-crafted SIFT based approach accomplishes 88.88% in terms of subject and 83.56% with reference to patches, the highest in this study, which can be explained away by the fact that the datasets are in small numbers. Significantly, during the Tuberculosis Competition at ImageCLEF 2017, the authors took part in the task of classification of 5 types of TB disease and achieved top one with regard to averaged classification accuracy (i.e. ACC = 0.4067), which is also premised on the approach of CNN+SVM+patch. On the other hand, when the whole slices of 3D TB datasets are applied to train a CNN network, the best result is achieved through the application of CNN coupled with orderless pooling and SVM at 64.71% accuracy rate.

Keywords: Deep learning, SVM, classification, patch-based image classification, Tuberculosis (TB), multi-drug resistant TB.

INTRODUCTION

Tuberculosis (TB) is a bacterial infectious disease caused by Mycobacterium (M.) Tuberculosis through inhaling tiny droplets from the coughs or sneezes of an infected person and remains one of the top 10 causes of death worldwide. In 2015, 10.4 million people fell ill with TB, among them 1.8 million died of the disease [1], including 0.4 million of HIV patients. While most of the TB cases take place in developing countries, this Victorian disease has not been eradicated in the developed countries. Quite contrary, the rate of the disease has been risen in some parts of western countries recently, for example, in London UK, as a result of a number of reasons, including drug abuse and sleeping rough.

Although TB remains a serious contagious condition, it can be cured if treated timely with the right antibiotics. For different forms of TB which are resistant to certain antibiotics, several different antibiotics can be administrated, which however, can lead to multidrug-resistant (MDR), extensively drug-resistant TB, HIV-associated TB, and weakening health systems. To detect drug resistant TB, clinically, the most definitive method is to undertake microbiological culture that can last up to several months albeit being an expensive procedure. Therefore there is an urgent clinical need for additional methods that can determine TB forms of either drug resistant or drug sensitive (DS) in a speedy, accurate and at the same time economic fashion. One of such approaches is to apply high resolution Computed Tomography (CT) imaging tool to assist clinicians to analyze, diagnose and deliver optimal treatment for TB patients.

This paper focuses on the implementation of state of the art deep learning technique to analyse CT pulmonary images and is organised in the following structure. Section 2 reviews the fundamentals of tuberculosis disease and image analysis techniques, in particular deep learning.

In Section 3 the datasets and proposed methodology to classify them are entailed. In Section 4, the implementation details are specified together with experimental results. Section 5 summaries the research work, discusses the limitations and stipulates future directions.

BACKGROUND

a. Tuberculosis infection

Mycobacterium tuberculosis (M. TB) was discovered 130 years ago and is of an aerobic, non-motile, non-spore-forming rod that is highly resistant to drying, acid, and alcohol. This bacterium transmits from person to person via droplet nuclei containing the organism through the air mainly by coughing. A person with active but untreated TB is estimated to go on to infect 10 to 15 other people per year, depending on the number of droplets expelled by the carrier, the duration of exposure, and the virulence of the M. TB [3].

While TB can affect any other part of the body, from tummy (abdomen) glands, bones to nervous system, it mainly affects the lung. To begin the infection, TB mycobacteria firstly reach the pulmonary alveoli, where they invade and replicate within alveolar macrophages. To combat this presence of foreign bacteria, human immune system starts to responed to phagocytise inhaled mycobacteria by alveolar macrophages to allow them interact with T-lymphocytes, a subtype of white blood cell. As a result, cells of epithelioid histiocytes [4] are assembled and proceed to team up with lymphocytes to aggregate into small clusters. As a result, a mass of tissue of granulomas is constructed, upon which cell division process, i.e. cytokinesis, begins producing proteins, such as interferon-γ, generated by CD4 T-lymphocytes (effector T cell) secrete, and sets to activate macrophages to destroy the infectious bacteria. In addition, generated CD8 T lymphocytes (cytotoxic T cell) can also directly destroy infected cells [5]. However, bacteria are

not always eliminated from the concerned granuloma. At many cases, they become inactive and dormant, which leads to a latent infection and undermines the human immune system.

b. Tuberculosis diagnosis

The signs with active pulmonary TB are revealed at varying stages, ranging from initial symptomless mild or progressive dry cough to symptomatic fever, fatigue, weight loss, night sweats, and a cough with blooded sputum, which have led the diagnosis of TB a very challenge task.

Clinically, the definitive diagnosis of active tuberculosis is the detection of presence of bacterium of M. TB, the causative microorganism of TB, which can be conducted through microbiological culture of human specimens [6]. In practice, however, the culture growth of M. TB may take 2 or more weeks on average. Hence to expedite diagnosis of active TB, an array of combined approaches set to rely on, including tuberculin skin test (TST), blood test, amplification of M. TB nucleic acids and/or pathological examinations from biological specimens. While these methods benefits, they are not specific. For example, the common practice is to detect the presence of acid-fast bacilli (AFB) on sputum smears [7], however only 44% of all new cases (and only 15–20% of children) can be identified. As a result, the ad hoc decision to initiate anti-tuberculosis treatment is made difficult in those cases where AFB does not manifest on sputum smear microscopy despite the clinical suspicion of TB.

Since pulmonary TB presents characteristics patterns in the lung, radiological imaging constitutes an inseparable tool to assist diagnosis. While conventional chest X-ray remains the most commonly employed method for screening, diagnosis and the follow up of treatment responses in patients with pulmonary TB, high-resolution computed tomography (CT) on chest appears to be more sensitive than X-ray in identifying early parenchymal lesions, detecting mediastinal lymph node enlargements and determining disease activity in TB [8-10]. On the other hand, radiographic features on CT that are suggestive of active TB comprise cavitation, parenchymal abnormality, centrilobular nodule and tree-in-bud pattern [6].

c. Tuberculosis in pulmonary CT images

TB can be classified into primary and post-primary categories with the former occurring in patients for the first time exposing to mycobacterium TB whereas post-primary or reactivated TB refers to the cases whereby patients have been previously infected and have hence developed a certain degree of acquired immunity. Clinically, diagnosis of TB is based on high index of suspicion [11]. If TB is detected timely and fully treated, people with the disease can quickly become noninfectious and eventually cured. Therefore early diagnosis and thereafter treatment remain crucial for both maintaining patients' health and reducing proliferation of TB to the public.

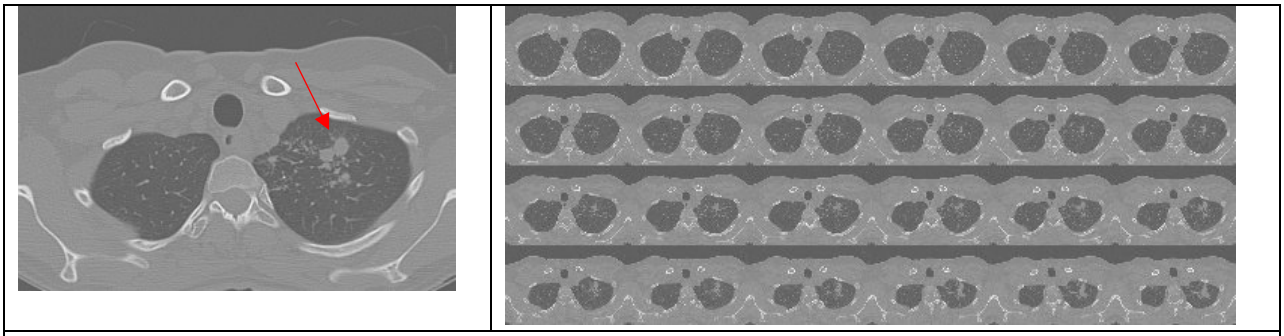
At radiology, primary tuberculosis manifests as four main entities: parenchymal disease, lymphadenopathy, miliary disease, and pleural effusion [12], whereas post-primary tuberculosis, referring to both reinfection with and reactivation of TB, may manifest as parenchymal disease,

airway involvement, and pleural extension. While primary TB is usually self-limiting, post-primary tuberculosis is progressive, with cavitation as its hallmark, resulting in haematogenous dissemination of the disease as well as disease spread throughout the lungs. Healing usually occurs with fibrosis and calcification. However, the features of primary and post-primary tuberculosis may overlap. The distinguishing features of post-primary tuberculosis characterise a predilection for the upper lobes, the absence of lymphadenopathy, and cavitation.

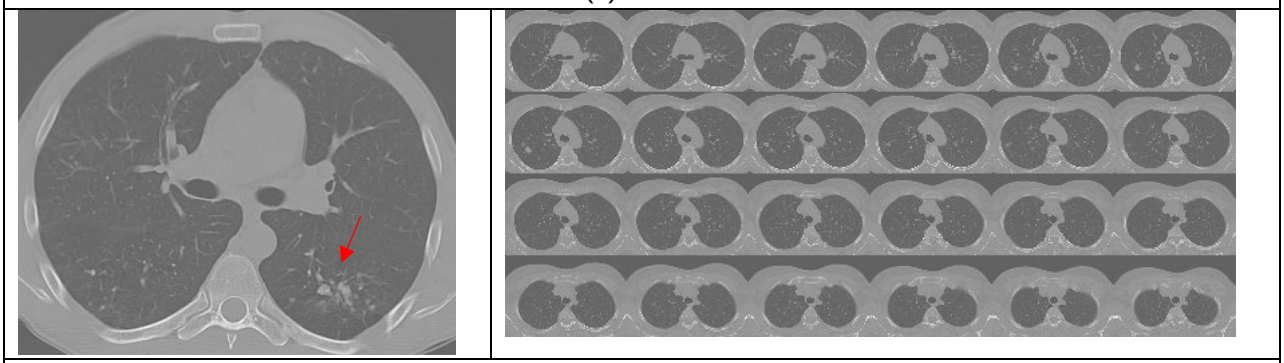
In addition, classical radiographic abnormalities have bearings of primary TB with unilateral hilar lymph node enlargement, parenchymal airspace consolidation and/or pleural effusion, whereas features of reactivation TB, focal or patchy heterogeneous consolidation involving the apical and posterior segments of the upper lobes and the superior segments of the lower lobes, poorly defined nodules, linear opacities and cavitations. [13].

While the classification in *primary* and *reactivation* TB remains a widely researched topic, evidence from genotype fingerprinting studies confirms that the radiographic feature in TB following recent and remote infection are very similar and that integrity of the immune system predicts the appearance of the patterns of active TB on chest imaging as that immune compromised individuals (e.g. those with advanced HIV infection) possessing the appearance of *primary* TB and immunocompetent individuals containing the appearance of *reactivation* TB [14].

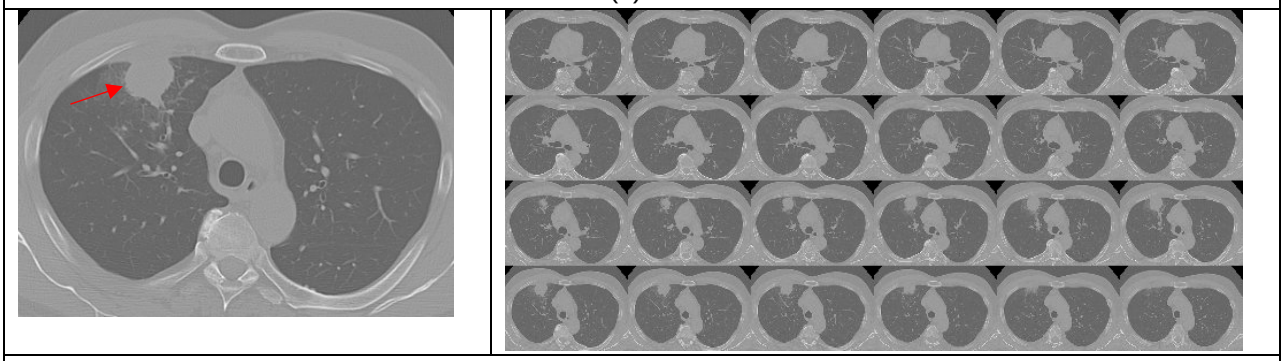
Figure 1 demonstrates five disease types of post-primary TB, which are Infiltrative, Focal, Tuberculoma, Miliary and Fibro-cavernous together with their montage form of 3D volumes. The abnormities are pointed with red arrows.



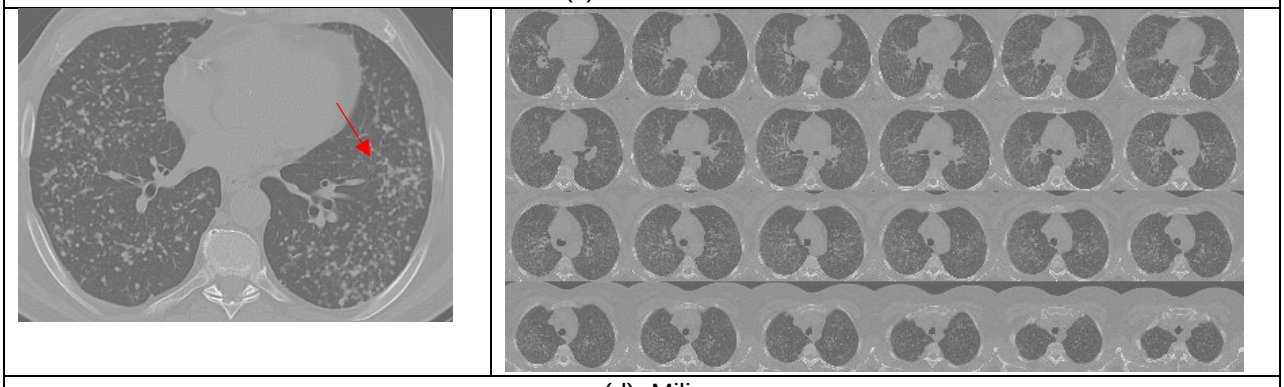
(a) Infiltrative



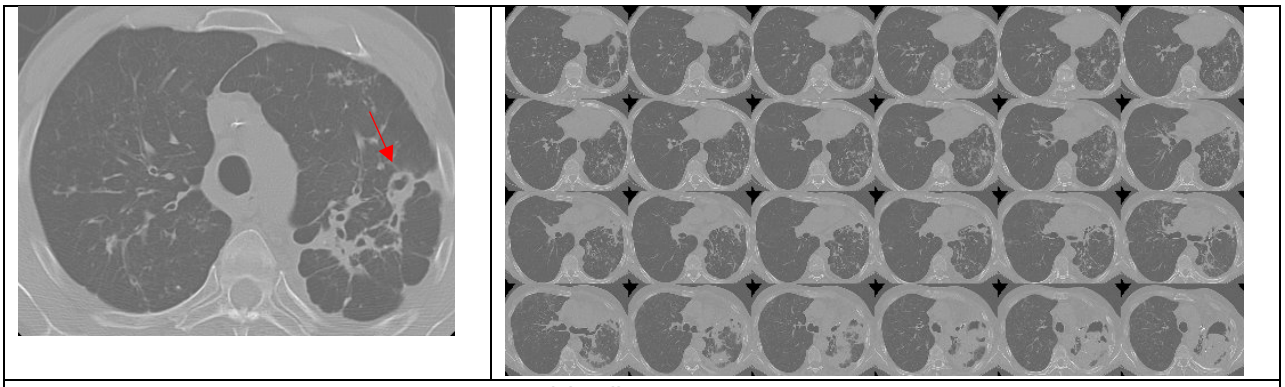
(b) Focal



(c) Tuberculoma



(d) Miliary



(e) Fibro-cavernous

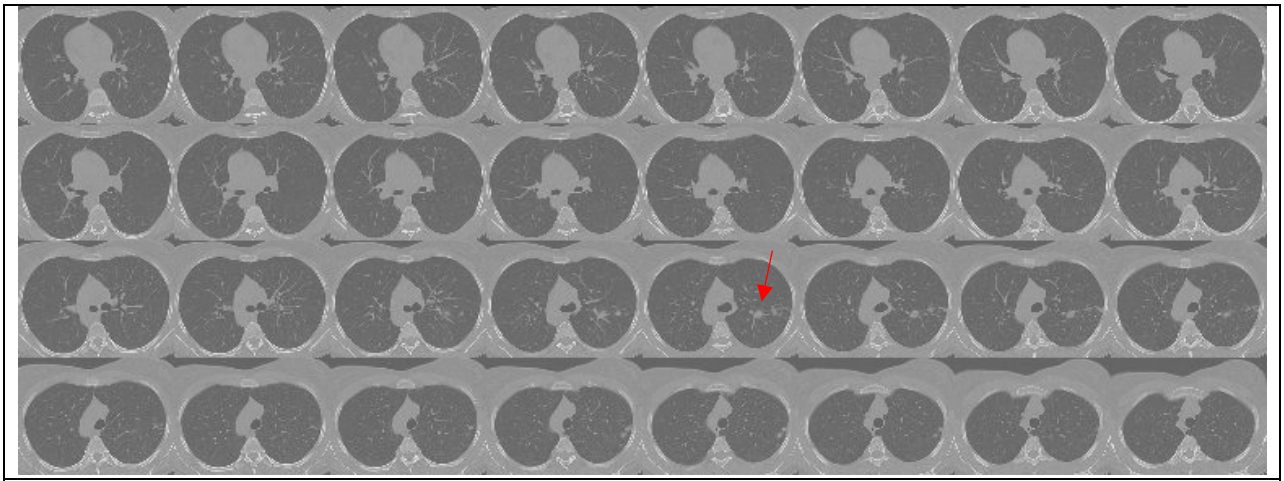
Figure 1. Illustration of five most common types of post-primary TB. Left: a slice of a 3D CT images; Right: montages of 3D volumes. From top to bottom: (a) Infiltrative; (b) Focal; (c) Tuberculoma; (d) Miliary; and (e) Fibro-cavernous.

The most serious setback that a patient with TB faces is that their organisms may become resistant to two or more standard drugs since every drug with an activity against M. TB can induce microbial resistance. As a matter of fact, the more active a drug retains, the more likely it will instigate clinical resistance [15]. Hence, multidrug-resistant (MDR) TB, defined as bacillary resistance to at least isoniazid and rifampicin *in vitro*, is posing a significant challenge to the control of TB worldwide [16]. More challengingly, genetically determined drug resistance (DR) in *M. TB* arises from spontaneous chromosomal mutation, whereas the transmission acquired DR bears the brunt of overcrowding, diagnostic delays and deficient infection control measures implemented in TB [1].

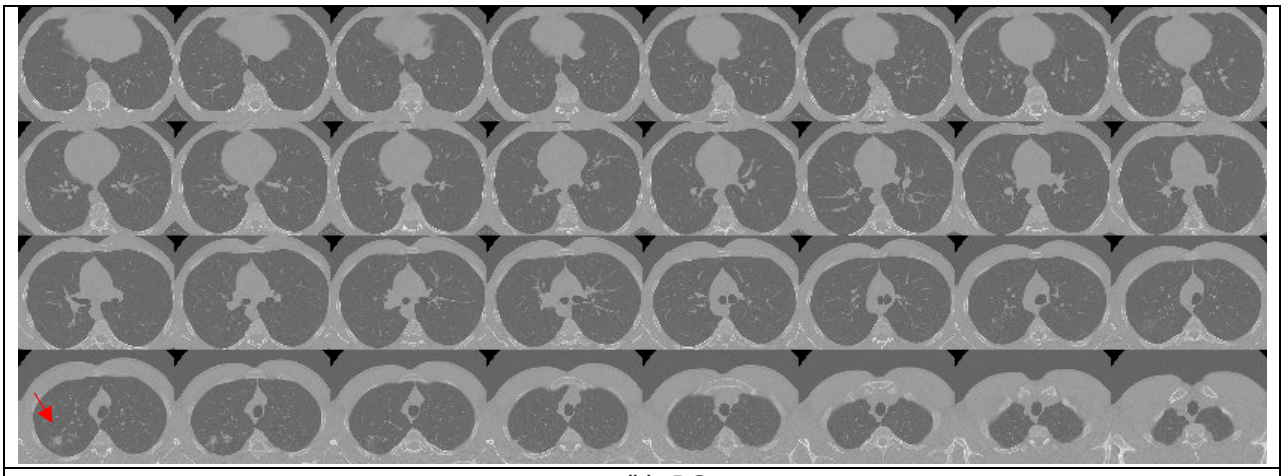
In contrast to drug sensitive (DS) TB, multi-drug resistant (MDR) form tends to be much more difficult and expensive to recover from. As a serious infection, MDR requires prolonged administration of more toxic second-line drugs that are usually correlated with higher morbidity

and mortality rates. Additionally, patients will remain infectious for a longer period once treatment starts, with associated higher risk to infect others [14]. Thus, early detection of drug resistance (DR) is of paramount in sustaining public welfare.

Figure 2 exemplifies the montages of CT lung images from both MDR and DS patients. In appearance, MDR TB appears to display lesions similar to one of the 5 types of TB shown in Figure 1 whereas DS TB tends to have less abnormal features. While DS data manifest fewer lesions, some slices do unveil similar features to those of TB types since DS patients are in the process of recovering, giving rise to the challenging task that classification and prediction of MDR sustain. A number of approaches have since developed, including texture graph-model, superpixel method, and deep learning techniques. In the *Tuberculosis Competition of ImageCLEF 2017*, the best accuracy rate (ACC) for classification of MDR and DS is 51.64% [17], further emphasizing the level of difficulty this task exhibits since statistically, a random guess rate is around 50% when dealing with only two clusters.



(a) MDR



(b) DS

Figure 2. Illustration of 3D CT images for patients with MDR (a) and DS (b), where arrows pointing to abnormal regions.

In this work, classification of MDR from DS in CT image are conducted via the application of start of the art deep learning technique implemented on 2D patches.

d. Deep learning in medical applications

Deep learning neural networks refer to a class of computing machines that can learn a hierarchy of features by establishing high-level features from low-level ones and is pioneered by Fukushima [18] based on biologically inspired human vision systems. One of these models is the convolutional neural network (CNN) developed by LeCun et al. [19]. Consisted of a set of algorithms in machine learning, CNN comprises several (deep) layers of processing involving learnable operators (both linear and non-linear), and hence has the ability to learn a hierarchy of information by building high-level information from low-level one, thereby automating the process of construction of discriminative information [20]. In addition, recent advances of

computer hardware technology (e.g., Graphics Processing Unit (GPU)) have propitiated the implementation of CNNs in representing images.

Conventionally, training a DL model requires large datasets and substantial training time. For example, the pre-trained CNN classifier, Alexnet [21], is built on 7 layers, simulating 659K neurons with 60 million (M) parameters and 630M connections, and trained on a subset (1.2M with 1K categories) of ImageNet [22] (15M 2D images of 22K categories), taking up 16 days on a CPU and 1.6 days on GPU.

While deep learning (DL) oriented approaches are widely applied on images that have large quantities, i.e. in millions, they have recently been applied on medical images in a number of domains and achieved state of the art results. In particular, CNN based approaches have won a number of competitions, including Kaggle competition on detection of diabetic retinopathy [23] and segmentation of brain tumours from MRI images [24]. In addition, in medical domain, not only the number of datasets is limited (usually in hundreds), but also images are in multiple dimensions ranging from 2D to 5D (e.g. a moving heart at a specific location). Hence additional measures have to be taken into account. For example, to classify 3D echocardiography video images, Gao et al [25] design a fused CNN architecture to incorporate both unsupervised CNN and hand crafted features to leverage the shortage of datasets. In addition, to capitalise on the information that a medical image proffers, they integrate two networks that are implemented for 2D and 3D respectively for classification of CT brain images [26]. Another way to increase the amount of datasets is to divide each slice into smaller segments or patches as implemented by

Janowczyk et al. [27] applying patch-based deep learning technique to analysis of pathology images.

In this study, patch-based deep learning network together with SVM has been implemented to tailor to the characteristics of the collected CT TB datasets. As illustrated in Figures 1 and 2, in a 3D CT volume, lung content appears less and less from top to bottom with more background information coming into view. As a direct result, if a 3D TB dataset is considered as a cube, there will be more than one third of volume of the cube filled with background information. On the other hand, on each slice, a number of abnormities usually materialise on a small region as depicted in Figure 1(c) in the case of Tuberculoma (arrow) that spreads into only a few consecutive slices whilst leaving the rest (~100 slices) taking shape of normal features. Hence in this study, to distinguish MDR subjects from DS, 2D patch based deep learning network is implemented coupled with the technique of support vector machine (SVM).

Theoretically, CNN can be conveyed as a process of minimising a cost function between the ground truths and predictions. Towards this end, with a set of training data $\left(x^{(i)}, y^{(i)}\right)$, where image $x^{(i)}$ is in three-dimension (inclusive of RGB channel as the 3rd dimension. Note: DL is a general approach and treats any input image as a colour data with dimensions of (row, column, RGB)= (M,N,3) whereby (M,N,1) is red, (M,N,2) is green and (M,N,3) is blue. For a grey image, 2D representation using (M,N) should be sufficient.) and $y^{(i)}$ the indicator vector of affiliated class of $x^{(i)}$, i.e. the ground truth, a CNN network is to solve the equation expressed in Eq. (1). In doing so, the feature maps of an image, namely, $w_1, \ldots, w_L$, will be learnt, coined deep learning.

$$\underset{w_1,\ldots,w_L}{argmin}\frac{1}{n}\sum_{i=1}^{n} l\left(f\left(x^i; w_1, \ldots, w_L\right), y^i\right) \quad (1)$$

where l refers to a suitable loss function (e.g. the hinge or log loss) and f the selected classifier.

To obtain these feature maps computationally, in a 2D CNN, convolution is conducted at each layer to extract features from local neighbourhood on the feature maps acquired in the previous layer. Then an additive bias is applied and the result is passed through a sigmoid function as formulated in Eq. (2) mathematically in order to obtain a newly calculated feature value v_{ij}^{xy} at position (x, y)on thej_{th} feature map in the i_{th}layer.

$$v_{ij}^{xy} = tanh\left(b_{ij} + \sum_{m}\sum_{p=0}^{P_i-1}\sum_{q=0}^{Q_i-1} w_{ijm}^{pq}\, v_{(i-1)m}^{(x+p)(y+q)}\right) \quad (2)$$

where the notations of those parameters in Eq. (2) are explained in Table 1.

Table 1. Notations of parameters in Eq. (2).

Parameter	**Notation**
$tanh(.)$	hyperbolic tangent function
m	index over the set of feature maps in the $(i-1)th$ layer
b_{ij}	bias for the feature map f in Eq. (1).
w_{ijk}^{pq}	value at the position (p, q) of the kernel connected to the k_{th} feature map
(p, q)	2D position of a kernel

P_i, Q_i	height and width of the kernel

As a result, CNN architecture can be constructed by stacking multiple layers of convolution and subsampling in an alternating fashion. The parameters of CNN, such as the biasb_{ij} and the kernel weight w_{ijk}^{pq} are trained using unsupervised approaches [28] whereby their initial values are setup randomly.

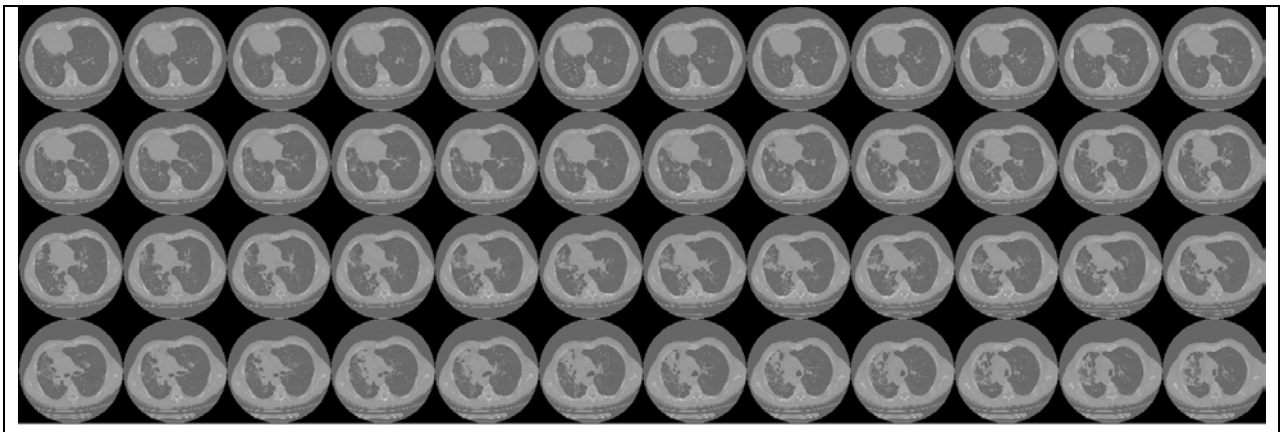
METHODOLOGY

a. Datasets

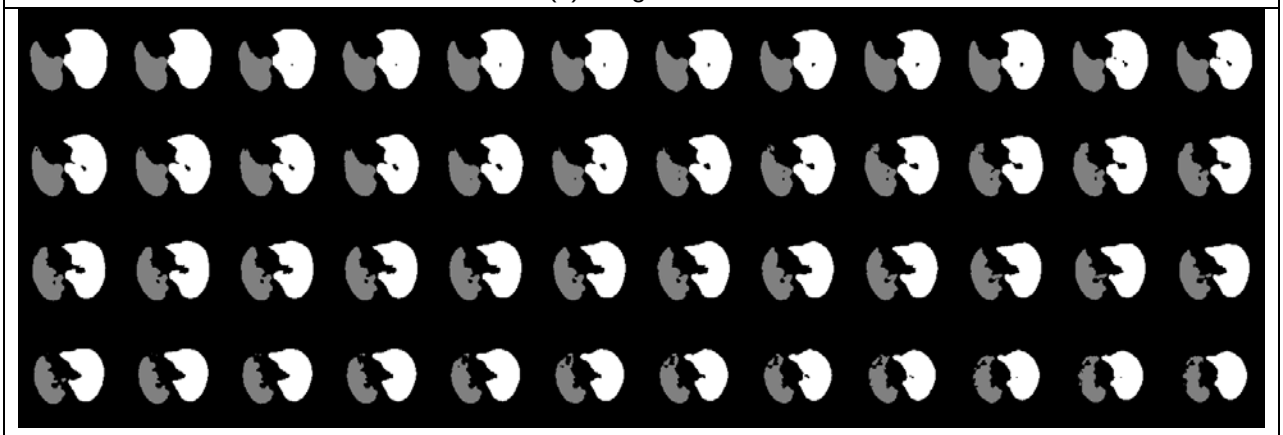
The datasets in this study are collected from the *Tuberculosis Competition* put forward by ImageCLEF [29], part of CLEF conference taking place in Dublin [30, 31], in September 2017. During the competition, the authors took part in task 2 of *Detection of TB Types* of 5 categories and have achieved top one with reference to accuracy (ACC) and ranked 5 while calculated using Kappa algorithm [32] among 23 participating teams.

This paper elaborates the work that has been conducted on Task 1 of the competition on *Detection of Multi-drug Resistance* (MDR) applying the same patch-wise CNN architecture as advanced in [32]. While there are 444 subject datasets in total with 230 for training and 214 for testing, only the training sets of 230 subjects (with 134 of DS and 96 of MDR) have known ground truth. Hence the results of this paper are based on these datasets of 230 3D CT lung images with resolution of 512×512×depth (Figure 2) where depth varies ranging from 55 to 263 slices. These datasets are then divided into training, validation and testing sub-dataset.

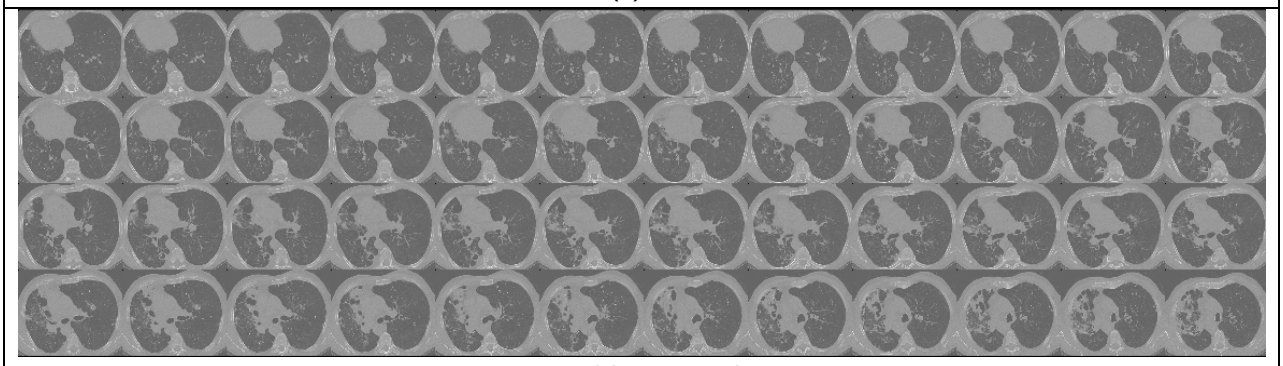
For the training data, each 3D dataset firstly undergoes pre-processing stage, whereby artefacts are removed, e.g. slices that contain little lung content will be excluded. As a result, each data volume comprises frames between 50 and 170 slices. Then upon each slice, patches with the size of 64×64 pixels are generated based on the lung boundary that is created from its corresponding mask that can either apply the existing file given by [33] or simply utilise a thresholding method. Those patches have overlapping contents of 10, 16 and 32 pixels. Figure 3 elaborates this pre-processing stage figuratively. Firstly, a 3D dataset (a) are cropped through the incorporation of their masks (b) to obtain mainly lung regions (c). Then patches of sizes of 64×64 are generated with strides of 10, 16 and 32 pixels respectively. In this way, lesions of whole sizes can be captured. Figure 3(d) demonstrates some of the generated patches. In addition, the confirmation of each patch is constrained by its corresponding mask patch that should contain at least 80% of lung contents (i.e. $\sum_{i,j} x'(i,j)/(6464) \geq 0.8$ where mask file $x'(i,j)$ is 1 for lung and 0 for background), which is done automatically. Finally, the selection of patches of interest takes place (Figure 3(e)), which is performed manually. This manual selection process can also be made semi-automatic by firstly training a small group of patches using SIFT approach to classify lesioned and normal patches. Then this trained SIFT classifier is applied to cluster the rest patches. And finally, manual check takes place to reassign wrongly classified patches. For MDR group, only patches with visible abnormalities are selected for training and test (Figure 3(e)), whereas for DS group, all the patches are employed.



(a) Original dataset



(b) Mask



(c) Cropped

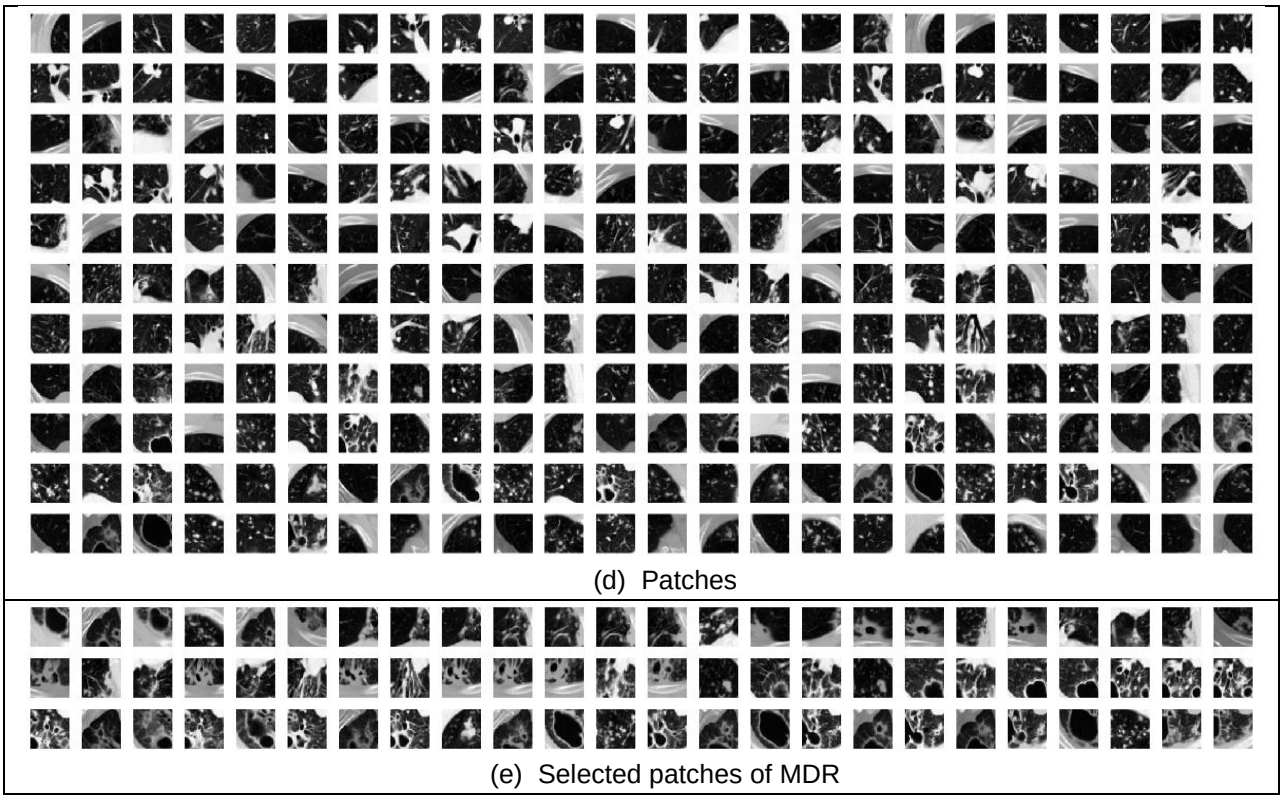


Figure 3. Pre-processing steps from raw image volumes to patches. (a) Raw image volume; (b) its mask; (c) cropped volume; (d) patches; and (e) selected patched for training and evaluation.

Table 2 lists the number of patches and the associated subject numbers that are applied in this investigation. In order to maintain the relative balance between DS and MDR class numbers, DS patches are randomly selected automatically from all the available DS patch pool for training and validation. It should be noted that all the patches from any one single subject are taken as an entity and assigned together for each activity, e.g. training or testing or validation in order to maintain their own inherent characteristics. The main purpose of validation is to adjust parameters as specified in Eq. (1).

Table 2. The overview of numbers of data that are applied in this study.

	DS	MDR	Total

	Patch	Subject	Patch	Subject	Patch	Subject
Training	5000	65	3000	85	8000	150
Validation	1167	31	179	4	1346	35
Testing	1520	38	500	7	2020	45
Total	7687	134	3679	96	11366	230

b. The Architecture of Deep Learning Convolution Neural Network (CNN) with SVM

Figure 4 describes the applied CNN architecture implemented in this study, which is built upon matConvNet package written using Matlab software [34]. Six layers of CNN are designed with input data of 64 × 64 pixels. The filter sizes for each layer are of (4, 4), (3, 3), (3, 3), (2, 2), (2, 2), and (3, 3) respectively.

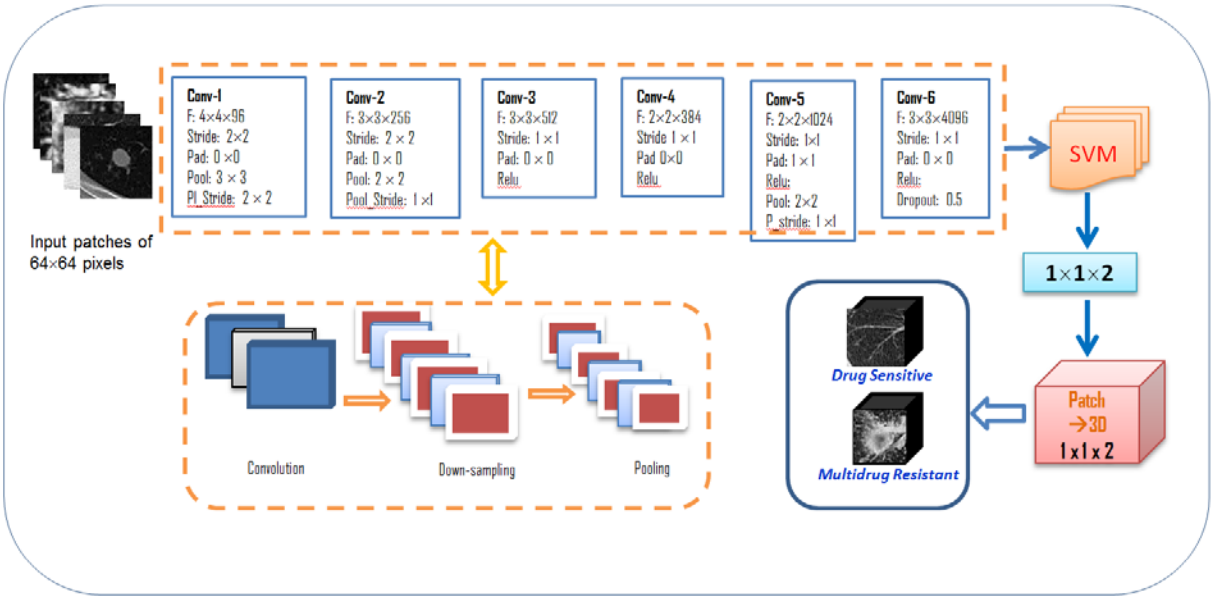


Figure 4. The CNN architecture implemented in this study.

At each layer, to learn jointly, both forward and backward processing are staged composed of several operators in an end-to-end manner. As such, a forward neural network tends to be the composition of a number of functions as formulated in Eq. (3) [32].

$$y = f(x) = f_L(\dots f_2(f_1(x; w_1); w_2) \dots; w_L) \quad (3)$$

Each function f_l takes a datum x_l as input that has a size of $M \times N$ pixels $\times K$ channels (default of K being 3 representing R, G, and B colour channels) and a parameter vector w_l, then produces an output datum x_{l+1}. The very first input of $x = x_0$ indicates a patch whereas the rest of x_l ($l > 0$) are intermediate feature maps. For each convolutional layer, the initial input filter bank of w_i is randomly generated but with pre-defined filter sizes. For example, in Figure 4 top graph, of Conv-1, the filter size is set as $4 \times 4 \times 3$, generating 96 filter banks. The output of the convolution with this bank of filters, y, is assessed in Eq. (4).

$$y_{i'j'k'} = \sum_{ijk} w_{ijkk'} x_{i+i',j+j',k} \quad (4)$$

where $k' = 96$, $k = 3$, $i' = 4$, and $j' = 4$ for the first Conv layer. In other words, each convolutional operator generates K' dimensional map of y by Eq. (4). For example, for layer 1 where $x_0 = (64,64,3)$ with the original frame size, feature map $x_1 = (15,15,96)$ is generated after layer-1 convolutional operator. Since the images are in grey, the third dimension representing RGB colour channels is ignored at this paper, i.e., $x_0 = (64,64,3)$ being replaced by $x_0 = (64,64)$. The calculation of the size of feature map follows the rule set out in Eq. (5).

$$x_{i+1} = \left(\frac{x_i - F_i + 2Pad}{Stride} + 1\right) \quad (5)$$

Additionally, each component or pixel of a feature map is subject to a non-linear gating process to legitimize the processed data. In this study, the simplest approach of rectified linear unit (ReLU) is conveyed in Eq. (6) that thresholds the data with zero.

$$y_{ijk} = max\{0, x_{ijk}\} \tag{6}$$

This operator however does not change the size of each feature map. To down size the feature map, pooling is employed to coalesce nearby feature values into one downsized samplings and reduce the influence of noise while operating on each individual feature channel. The most commonly used choice of pooling remains to be max-pooling to select the largest component within a neighborhood as manifested in Eq. (7),

$$y_{ijk} = max\{y_{i'j'k} \ : i \leq i' < i + p, j \leq j' < j + q\} \tag{7}$$

whereas the downsize rate is controlled by pooling stride (P-Stride).

Another operator remains *Dropout* [35] to remedy the overfitting tendency in a CNN network. In doing so, randomly dropout units (along with their connections) from the neural network during training stage are selected and discarded. The dropout rate in this study is set to be 0.5, i.e. half the data units are randomly selected and deleted at the corresponding layer. For example, in Figure 4 at layer 6 of Conv-6, after dropout operation, the data unit is 2048 (=4096/2).

Once each layer of forward processing is completed, backward process proceeds to ensure that the parameters of feature maps, $w = (w_1, \dots, w_L)$, are learned in such a way that the overall function of $z = f(x, w)$ sustains a minimum loss, $l(z, \hat{z})$, where $(z_1, \dots, z \mid i, \dots)$ corresponds with the output value of x_i and $\hat{z}_i$ the ground truth of x_i in the training datasets. Therefore the loss function can be determined below in Eq. (8).

$$L(w) = \frac{1}{n}\sum_{i=1}^{n} l\big(z_i, f(x_i, w)\big) \tag{8}$$

Understandingly, there exist a number of algorithms for minimising L, such as recently developed stochastic configure network (SCN) leaner model proposed by Wang et al [36]. In this research, the approach of *gradient descent* is employed which quantifies the gradient of *L* at a current solution $w^t$and then updates t along the direction of fastest descent of *L* as revealed in Eq. (9).

$$w^{t+1} = w^t - \eta_t \frac{\partial f}{\partial w}(w^t) \tag{9}$$

where η_t refers to the learning rate that is usually pre-defined and is within the range of (0,1).

Substantially, while filter sizes can be of any size within the limit of data size and are chosen manually in advance, e.g. in Figure 4, the filter sizes are 3×3×96, 3×3×256, 3×3×512, 2×2×384, 2×2×1024, 3×3×4096 for the six layers respectively, the dimension of the output layer at the end of CNN architecture must be 1×1×2, which reduces the full input image into a single vector of class scores (in this case, class number is 2), arranged along the depth dimension and can be computed using Eq. (5) completed with the values of pooling stride.

At layer 6 of Figure 4, instead of scoring features into one of the two classes using *Softmax* approach (to be given below in Eq. (11)) that employs cross-entropy loss interpreting the scores as (un-normalised) log probabilities for each class, this study applies the algorithm of support vector machine (SVM) [37] (Eq. (12) that adapts hinge loss to encourage the correct class to have a score higher by a margin than the other class scores. In this way, each class has a distinguishing boundary. The final score for each subject is calculated based on all patch scores where a threshold is applied. In other words, if a 3D dataset is classified as DS, more than 90% (or threshold) patches should belong to DS category. Otherwise, this subject has MDR.

To apply Softmax classifier, the loss function in Eq. (8) can be written as Eq. (10).

$$L_i = -f_{z_i} + log \sum_j e^{f_j} \tag{10}$$

whereas Eq. (11) is called Softmax function.

$$f_j(z) = \frac{e^{z_j}}{\sum_k e^{z_k}} \tag{11}$$

On the other hand, to train a support vector machine (SVM), linear optimisation can be applied to Eq. (8), which minimises formula of Eq. (12) .

$$L_w = \left[\frac{1}{n}\sum_{i=1}^{n} max\big(0,1 - y_i(w \cdot x_i - b)\big)\right] + \lambda\|w\|^2 \tag{12}$$

SVMs [37] are a set of supervised learning models with analyse and classify data applying associated learning algorithms. There are linear and non-linear SVMs. While in a linear SVM that is employed in this study, any hyperplane can be written as the set of points x satisfying

$$w \cdot x - b = 0 \quad (13)$$

RESULTS

While each patch can be classified into either MDR or DS class, the determination of a subject with multiple patches is based on the collection of all patch scores accommodating both MDR and DS clusters. Therefore the cut-off number or threshold has to be found out first as what is the maximum number of MDR patches a DS subject should have. According to the training datasets, a subject is classified as having DS TB if and only if there are less than 10% patches that are labeled as MDR. Otherwise, this subject is labeled as a MDR sufferer.

As a result, Table 3 lists the final classification results where both loss functions of Softmax (Eq. (11)) and SVM classifier (Eq. (12)) are put in an application. For CNN training, Softmax loss function is employed as a built-in. After the training is completed, the feature maps from the CNN model is extracted and fed into SVM to conduct an additional training, which takes a few minutes to yield a new classifier. As illustrated in Table 3, CNN with SVM performs better, in particular in terms of subject where 41 (34 + 7) out of 45 subjects (91.11%) are correctly detected as either having TB of DS or MDR type. When applying a typical CNN architecture with Softmax as a cost function, 38 out 45 subjects are corrected diagnosed (84.11%)

Table 3. Test results in the form of confusion matrix with CNN and CNN+SVM with reference to both patch and subject.

Methods		DS		MDR		Accuracy Rate		Average	
		Patch	Subject	Patch	Subject	Patch	Subject	Patch	Subject
CNN + Softmax	DS	1321	32	199	6	0.8684	0.8421	0.7930	0.8444
	MDR	209	1	281	6	0.5620	0.8571		
CNN + SVM	DS	1295	34	225	4	0.8520	0.9210	0.7980	0.9111
	MDR	183	0	317	7	0.6340	1.000		

To distinguish the difference between SVM and Softmax visually, Figure 5 epitomizes the dissimilarity where those patches that belong to MDR have been correctly identified by CNN+SVM as MDR and misclassified by CNN+Softmax as DS.

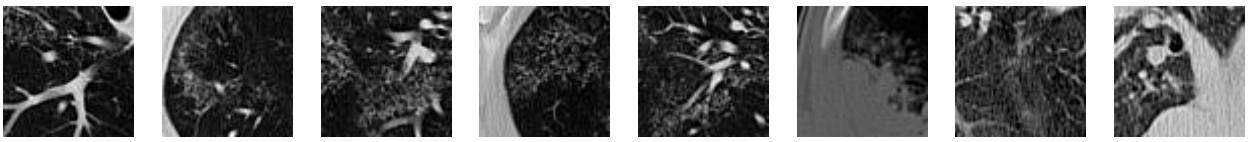


Figure 5. The patches from MDR subjects that are identified as MDR by CNN+SVM and as DS by CNN+Softmax**.**

While a CNN deep layers may be considered as a black box, the trained filter maps, i.e. , $w_1, \ldots, w_{L,}$ in Eq. (1) can be visually meaningful in the first few layers that may represent edges, texture, etc.. Figure 6 exhibits visually the weights of 96 in the first layer in this trained CNN model.

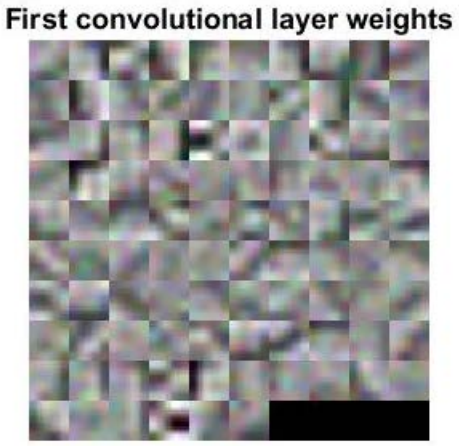


Figure 6. The visual appearance of first 96 filters.

In addition, in this investigation, the image batch size is set to be 256, i.e. the system takes 256 images in one go to process whereas GPU is employed. Furthermore, the learning rate is set to be 0.01 with initial bias being 0.1. Although the filter or kernel size is fixed in each layer, e.g. (4, 4), (3, 3), (3, 3), (2, 2), (2, 2), and (3, 3) respectively for the 6 layers, the initial values of each filter are randomly generated to start the learning process. During the training, although the epoch (iteration) number of the cycles that network loops is set to be 130, the system convergences at epoch 50 as monitored in Figure 7 for the classification of 5 TB types.

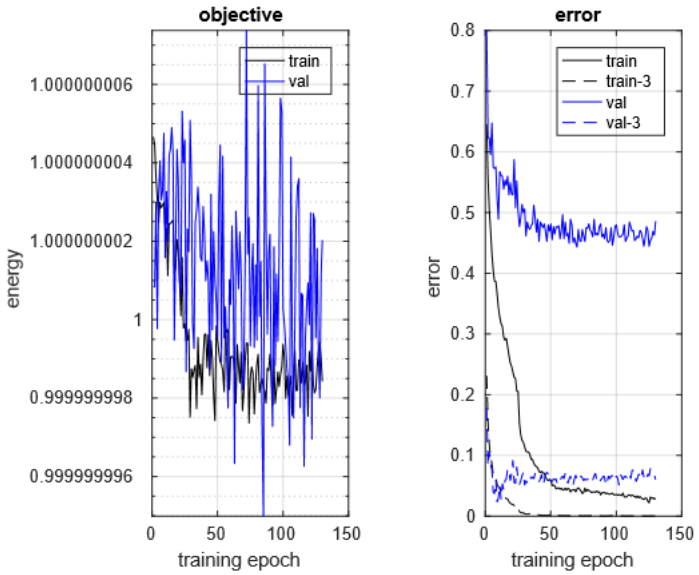


Figure 7. The learning information for training of 5 types TB when running 130 epochs.

Because of the characteristics of TB images showing structured texture, texture based approaches are frequently employed in representing them [38], by which SIFT appears to be one of the most popular one. Table 4 provides the confusion matrix obtained in this study using SIFT dense based texture approach, which also employs SVM classifier upon SIFT features. As presented in Table 4, with reference to subject, this SIFT based approach delivers an averaged accuracy rate of 88.88% and 83.56% when only considering patch classifications.

Table 4. SIFT dense results +SVM

	DS		**MDR**		**Accuracy Rate (%)**	
	patch	**subject**	**patch**	**subject**	**Patch**	**subject**
DS	1382	33	138	5	90.92%	86.64
MDR	194	0	306	7	61.2%	100
Overall					83.56%	**88.88%**

On the other hand, the evaluation is also conducted applying another pre-trained model AlexNet. As addressed in Section 2, AlexNet is built on non-medical images therefore its application follows concatenation of all patch features calculated using AlexNet into one hot feature vector that is then clustered again into 2 groups using Softmax. Table 5 gives the confusion matrix of AlexNet where the accuracy result is 91.11% with reference of subject.

Table 5. AlexNet + softMax

	DS		MDR		Accuracy Rate (%)	
	Patch	Subject	Patch	Subject	Patch	Subject
DS	1239	34	281	4	81.51	89.47
MDR	130	0	370	7	74.00	100
Overall patches					79.65	91.11

To summary, Table 6 provides the information of all the evaluation results conducted in this study. In relation to subject, the proposed network architecture of CNN+SVM appears to perform the same as Alexnet (91.11%), the best in this study. While taking into consideration of patch results, CNN+SVM (79.80%) performs slightly better than Alexnet (79.65%). Since the number of training datasets are not very large (13,000 patches in total) from 150 subjects, hand-crafted approach of dense SIFT with SVM accomplishes also very well with 88.88% accuracy rate, 4.4% higher than a typical CNN architecture and the best (83.56%) with regard to patches. It is expected that more datasets will improve this trend.

In addition, Table 6 contains a result from the conventional CNN that takes whole 2D slice into processing instead of patches. In doing so, order less pooling of filter bank response as described in [39] is adapted to simulate textures. It has been approved in [40] that in deep CNN, the convolutional layers are akin to non-linear filter banks, and appear to be better for texture

descriptions. To apply this deep learning based texture extraction, a volume of 3D TB is down sampled from slices into patches. The local deep features of each patch are then extracted from the last convolutional layer of a pre-trained deep learning model of VGG-M [41]. Subsequently, to render a representation at 3D level, after concatenating all slice features, fisher vector is applied to encode all local descriptors [40, 42] as one long feature vector. Finally, SVM is applied to assign TB to either MDR or DS. As shown in Table 6, this approach gives an accuracy rate of 64.71%

Table 6. Comparison with the Dense SIFT approaches.

Methods	Accuracy rate in patch (%)	Accuracy rate in subject (%)
CNN + orderless pooling + SVM + slice		64.71
CNN + patch	79.30	84.44
Alexnet with Softmax + patch	79.65	91.11
Dense SIFT + SVM +patch	83.56	88.88
CNN + SVM + patch	79.80	91.11

CONCLUSION AND FUTURE DIRECTIONS

This research proposes to apply patch based deep learning technique to classify multiple drug resistance from drug sensitive tuberculosis (TB) disease and has achieved the best performance with 91.11% accuracy. Early diagnosis of multi-drug resistance TB plays a crucial role not only in providing timely optimal treatment for each individual, but also in maintaining public welfare to prevent TB proliferation. High resolution CT images appear to provide a convenient, inexpensive yet important assistant tool for speedy detection of TB. While there exist many image classification techniques, deep convolutional neural networks (CNN) with patches appear to perform the best, delivering the highest classification accuracy rate. Similar to many other medical studies that fall short of datasets, this research ameliorates this predicament with only

230 datasets by using patches instead of volumes, leading to enlarging datasets from hundreds to thousands. In comparison with hand-crafted approaches that are popular with TB studies, in this study, SIFT texture based technique is employed and appears to perform well with 88.88% accuracy rate and outperform the others in terms of patches (83.56%).

Since the ground truth for the testing datasets from ImageCLEF 2017 competition is not known, the obtained results in this study may not be comparable directly to that published in [15]. However, the fact that this premised patch based classification architecture has achieved competitive result for classification of 5 TB types with top one accuracy result in the same competition [30] but different task signifies the effectiveness of this proposed approach. Additionally, it outperforms both the state of the art texture based CNN and pre-trained model of AlexNet.

Due to fact that the testing datasets only cater for 45 subject samples with 7 belonging to MDR cluster, the remaining work is to evaluate the results from real testing datasets when the ground truth is acquired. In order to obtain higher accuracy, it is recommended that medical knowledge should be embedded, including the patterns of MDR disease. Additionally, 3D segments should be also considered to further enhance the characteristics that 3D datasets entail.

The output of this CNN led architecture is a classification system or a model that bundles up every parameter for CT pulmonary images. Although the initial development and training of this system may take days or weeks depending on the volume of collected data (2 days in our case), the system/model will operate in real-time mode once the classifier or classification model is

established. In other words, once a new 3D dataset is made available and sent to the system, it takes a couple of seconds or minutes (depending on the depth of the volume) to give out the classification result of the data in probability, e.g. 89% belongs to MDR, 11% belongs to DS and outcome of MDR. Furthermore, similar to any other software systems, updating this classification system or model can be conducted at regular basis whenever new dataset/information/evidence is at our disposal.

REFERENCES:

[1] WHO, Tuberculosis, Fact Sheet, March 2017. http://www.who.int/mediacentre/factsheets/fs104/en/. Retrieved in June 2017.

[2] BBC, Parts of London have higher TB rates than Iraq or Rwanda. **http://www.bbc.co.uk/news/uk-england-london-34637968**. Retrieved in June 2017.

[3] Jeong YJ, Lee KS. Pulmonary tuberculosis: up-to-date imaging and management. AJR Am. J. Roentgenol. 2008*;* 191*:* 834–44.

[4] Houben EN, Nguyen L, Pieters J. Interaction of pathogenic mycobacteria with the host immune system. *Curr Opin Microbiol* 2006; 9:76–85.

[5] Kaufmann SH. Protection against tuberculosis: cytokines, T cells, and macrophages. *Ann Rheum Dis* 2002; 61[suppl 2]:ii54–ii58.

[6] Lange C, Mori T, Advances in the diagnosis of tuberculosis. *Respirology 2010;* 15: 220–240.

[7] WHO. Global Tuberculosis Control 2009. Epidemiology, Strategy, Financing, Geneva, 2009.

[8] Lee KS, Im JG. CT in adults with tuberculosis of the chest: characteristic findings and role in management. *AJR Am. J. Roentgenol.* 1995; 164: 1361–7.

[9] McGuinness G, Naidich DP, Jagirdar J, Leitman B, McCauley DI. High resolution CT findings in miliary lung disease. *J. Comput. Assist. Tomogr*. 1992; 16: 384–90.

[10] Pastores SM, Naidich DP, Aranda CP *et al.* Intrathoracic adenopathy associated with pulmonary tuberculosis in patients with human immunodeficiency virus infection. *Chest* 1993; 103: 1433–7.

[11] Krysl J, Korzeniewska-Kosela M, Muller NL *et al.* Radiologic features of pulmonary tuberculosis: an assessment of 188 cases. *Can. Assoc. Radiol. J.* 1994; 45: 101–7.

[12] Leung AN. Pulmonary tuberculosis: the essentials. *Radiology* 1999; 210: 307–22.

[13] Cardinale L, Parlatano D, Boccuzzi F, Onoscuri M, Volpicelli G, Veltri A. The imaging spectrum of pulmonary tuberculosis. *Acta radiologica*. 2015; 56(5):557-564.

[14] Burrill J, Williams CJ, Bain G, Conder G., Hine AL, Misra RR. Tuberculosis: a radiologic review. *Radiographics 2007;* 27*:* 1255–73.

[15] Canetti G. Present aspects of bacterial resistance in tuberculosis. *Am. Rev. Respir. Dis.* 1965; 92: 687–703.

[16] Yew WW, Chau CH. Management of multidrug-resistant tuberculosis: Update 2007, Respirology, 13:21-46, 2008.

[17] Cid YD, Kalinovsky A, Liauchuk V, Kovalev V, Müller H, Overview of the ImageCLEF 2017 Tuberculosis Task - Predicting Tuberculosis Type and Drug Resistances (2017). *In: CLEF2017 Working Notes 1866*. http://CEUR-WS.org/Vol-1866 .

[18] Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cyb.* 1980; 36: 193–202.

[19] LeCun Y, Bottou L, Bengio Y, and Haffner P. Gradient-based learning applied to document recognition, *Proceedings of the IEEE 1998*; 86(11): 2278–2324.

[20] LeCun Y, Huang FJ , Bottou L, Learning methods for generic object recognition with invariance to pose and lighting, *Processings of Computer Vision and Pattern Recognition (CVPR)* 2004; 2: II-97-104.

[21] Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems 2012*. NIPS 2012.

[22] ImageNet, http://www.image-net.org/.

[23] Kaggle Competition, https://www.kaggle.com/c/diabetic-retinopathy-detection/discussion/15801.

[24] Pereira S, Pinto A, Alves V, and Silva C. Brain tumour segmentation using convolutional neural networks in MRI images, *IEEE transactions on medical imaging 2016*; 35(5):1240-1251.

[25] Gao X, Li W, Loomes M, Wang L, A fused deep learning architecture for viewpoint classification of echocardiography, *Information Fusion 2017; 36:103-113.*

[26] Gao X, Hui R, Tian Z, Classification of CT images based on deep learning networks, *Computer Methods and Programs in Biomedicine 2017;* 138:49-56.

[27] Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases, *Journal of Pathology Informatics* 2016; 7(1):1-29.

[28] LeCun *Y*, Bengio Y, Hinton G, Deep Learning, *Nature 2015*; 521: 436-444.

[29] Müller H, Villegas M., Arenas H, Boato G, Dang-Nguyen D, Dicente C, Eickhoff C, Garcia Seco de Herrera A, Gurrin C, Islam B, Kovalev V, Liauchuk V, Mothe J, Piras L, Riegler M, Schwall I. Overview of ImageCLEF 2017: Information extraction from images, *CLEF 2017 Proceedings, Lecture Notes in Computer Science (LNCS) 2017;* 10439.

[30] Jones GJF, Lawless S, Gonzalo J, Kelly L, Goeuriot L, Mandl T, Cappellato L, Ferro N. (ed.), Proceedings of Experimental IR Meets Multilinguality, Multimodality, and Interaction, *8th International Conference of the CLEF Association, CLEF 2017*, LNCS 10439.

[31] Cappellato L, Ferro N, Goeuriot L, and Mandl T. (ed.), CLEF 2017 Labs Working Notes, *CEUR-WS Proceedings*, 2017. http://ceur-ws.org/Vol-1866.

[32] Gao X, Qian Y, Application of Deep Learning Neural Network for Classification of TB Lung CT Images Based on Patches, in Jones G. J. F., et al. (eds.), *Proceedings of Experimental IR Meets Multilinguality, Multimodality, and Interaction 8th International Conference of the CLEF Association, CLEF 2017 Working Notes 2017*, 1866.

[33] Cid Y, Jiménez-del-Toro OA, Depeursinge A, Müller H, Efficient and fully automatic segmentation of the lungs in CT volumes. In: Goksel, O., et al. (eds.) *Proceedings of the VISCERAL Challenge at ISBI. No. 1390 in CEUR Workshop Proceedings* (Apr 2015).

[34] Vedaldi A, Lenc K. MatConvNet Convolutional Neural Networks for MATLAB , *Proceedings of the 23rd ACM international conference on Multimedia* 2015, 689-692.

[35] Srivastava N, Hinton GE, Krizhevsky A, Sutskever I. Salakhutdinov, R., Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning 2014;* 15: 1929-1958.

[36] Wang D, Cui C. Stochastic Configuration Networks Ensemble for Large-Scale Data Analytics, *arXiv preprint arXiv:1707.00300, 2017*.

[37] Cortes C, Vapnik V. Support-vector networks. *Machine Learning* 1995; 20 (3): 273–297.

[38] Cid Y, Batmanghelich K, Müller H., Textured Graph-model of the Lungs for Tuberculosis Type Classification and Drug Resistance Prediction: Participation in ImageCLEF 2017, *CEUR working notes 2017*, 1866.

[39] Gong Y, Wang L, Guo R, Lazebnik S. Multi-scale Orderless Pooling of Deep Convolutional Activation Features 2014. *ECCV 2014*, Part VII. 392–407.

[40] Cimpoi M, Maji S, Vedaldi A.. Deep Filter Banks for Texture Recognition and Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition* 2015.

[41] Chatfield K, Simonyan K, Vedaldi A, Zisserman A. Return of the devil in the details: Delving deep into convolutional nets 2014. *In Proc. BMVC 2014*.

[42] Cimpoi M, Maji S, Kokkinos I, Mohamed S, Vedaldi A. Describing textures in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014*.