Abstract

Tuberculosis disease remains a global threat and a leading cause of death. The world health organization stressed on increasing the decline rate of the disease. Early diagnosis and evaluation of the TB severity stage are important for determining the right treatment and eventually avoiding the death of curable patients cases. The voluminous amount of medical images available and the proven performance of deep learning in medical diagnosis are a motive for an automatic medical diagnosis to tackle the high demand for radiologists and reduce the costs of diagnosis. Furthermore, there has been considerable growth in recent years in the field of deep learning, which allows performing hard classification tasks. Particularly in computer vision fields, it has been tested and proven that deep convolutional neural networks (CNN) are very promising algorithms for various visual tasks. Thus, in this report, we are interested in an automatic TB severity scoring by applying deep learning techniques on annotated chest CT scans. Our contribution is based on 3 deep learners namely, Resnet50, InceptionV3Resnet, and Lungnet. Our submissions on the test corpus reached AUC value of about 65% in ImageCLEF2019 SVR sub-task. We believe that our contributions could be further improved and might give better results if applied properly and in an optimized way.

Keywords : Tuberculosis, Computed Tomography, Image Classification, Deep learning, Severity Scoring, 3D Data Analysis, Lungnet.

Contents

Abbreviations

3D 3 Dimensions

AUC Area Under Curve

BK Bacillus Koch

CNN Convolutional Neural Networks

CT Computed Tomography

CTR CT report task

Dicom Digital Imaging and Communications in Medicine

FN False Negatives

FP False Positives

HIV Human Immunodeficiency Virus

JPEG Joint Photographic Experts Group

LTBI Latent Tuberculosis Infection

Minc Medical Imaging NetCDF

ML Machine Learning

MPEG Moving Picture Experts Group

Nifti Neuroimaging Informatics Technology Initiative

RLE Run-Length Encoding

RMSE Root Mean Squared Error

ROC Receiver Operating Characteristic

SVM Support Vector Machines

SVR Severity scoring task

TB Tuberculosis

TN True Negatives

TP True Positives

UPN Unsupervised Pretrained Networks

WHO World Health Organization

Introduction

In recent decades, medical imaging has become indispensable in the diagnosis and therapy of diseases. With the enhancement of medical imaging databases, new methods are required to better handle this huge volume of data. However, because of the large variations and complexity of medical imaging data, it is generally difficult to deduce analytical solutions or simple methods to describe and represent objects such as lesions and anatomies in data. Therefore, medical imaging tasks require learning from the examples, and this is one of the key interests of the machine learning field.

Machine learning has become one of the main tools for medical image analysis. Machine learning techniques are solutions for developing tools to help physicians diagnose, predict, and prevent the risk of disease before it becomes too late in less time. Deep Learning is a new component in the field of machine learning that encompasses a wide range of network architectures designed to perform multiple tasks. The first use of neural networks for medical image analysis goes back more than twenty years, their use has increased by several orders of magnitude over the last five years. different, articles [1, 2, 3, 4, 5] have highlighted the application of deep learning to a wide range of medical image analysis tasks (segmentation, classification, detection, recording, image reconstruction, enhancement, etc. . .).

Tuberculosis is an infectious disease caused by a bacterium called Bacillus mycobacterium tuberculosis [6]. In 2018, 10 million people fell ill with TB, and 1. 6 million died from the disease [6]. This disease remained one of the top ten leading causes of death in the world in 2018 [6]. Tuberculosis attacks the lungs but can also affect other parts of the body [7]. Accurate and rapid diagnosis is the key to controlling this disease, but traditional TB tests produce inaccurate or time-consuming results to be definitive. Researchers have been interested in this disease, particularly in the context of the ImageCLEF 2018[8] international challenge [8] where two tasks have been reserved for it. Algorithms involving deep learning have been tested to diagnose the presence or absence of tuberculosis. The results obtained were interesting. Indeed, the algorithms have achieved an impressive accuracy rate up to 96% [9, 10] a result that is better than the intervention of many radiologists.

The goal of our project is to automatically give a score of TB severity via Computed Tomography (CT) scan. One of the possible applications of this study is to accelerate the diagnosis of the disease from a radiology image without resorting to expensive medical tests. This work is part of the ImageCLEF2018 task of classifying types of TB, which has shown more promising results. We explore in this paper the different work and different concepts that link with this problem. Starting by giving an overview of tuberculosis and its types in chapter 1. Then, discuss the relationship between artificial intelligence and

medical images then give definitions of some important parts of these fields in chapter 2. In chapter 3, the ImageCLEF Tasks are described and the related work of tuberculosis severity scoring is reviewed. Chapter 4, will be dedicated to a description of our contribution, the methods used and a presentation of the results of the performance on the validation dataset and the test dataset provided by ImageCLEF2019[8].

Chapter 1

Pulmonary Tuberculosis and its types

Tuberculosis is a chronic infectious disease caused by a bacterium. called " Mycobacterium Tuberculosis" or Bacillus Koch (Bacillus Koch (BK)). Its most current and most common form (85% of cases) is pulmonary tuberculosis, but there are also extrapulmonary forms such as bone tuberculosis, ganglion tuberculosis, and renal tuberculosis.

Tuberculosis can develop rapidly after the first contact with the microbe, but it can also appear several years later.

1.1 Latent TB infection (LTBI)

LTBI is the presence of tubercle bacilli within the body without manifestation of the disease. LTBI carriers are by definition non-contagious and pose no risk to those around them.

1.2 Active tuberculosis

Active tuberculosis is a condition in which the body's immune system is unable to fight off or defend against the Mycobacterium tuberculosis bacterium. This inability causes an infection of the lungs, which is the most common presentation, or other parts of the body (tuberculosis is a multisystemic disease). Apart from the respiratory system, the organ systems most commonly affected include the gastrointestinal system, the musculoskeletal system, the lymphoreticular system, and the reproductive system, as well as the skin and the liver.

Globally, the best estimate is that 10.0 million people (range, 9.0–11.1 million) developed TB disease in 2017: 5.8 million men, 3.2 million women and 1.0 million children [11]. There were cases in all countries and age groups, but overall 90% were adults (aged 15 years), 9% were people living with Human Immunodeficiency Virus (HIV) (72% in Africa) and two thirds were in eight countries: India (27%), China (9%), Indonesia (8%), the Philippines (6%), Pakistan (5%), Nigeria (4%), Bangladesh (4%) and South Africa (3%). These and 22 other countries in World Health Organization (WHO)'s list of 30 high TB burden countries accounted for 87% of the world's cases[11]. Only 6% of global

cases were in the WHO, European Region (3%) and WHO Region of the Americas (3%) [11].

Estimated TB incidence rates are shown in Figure 1.1.

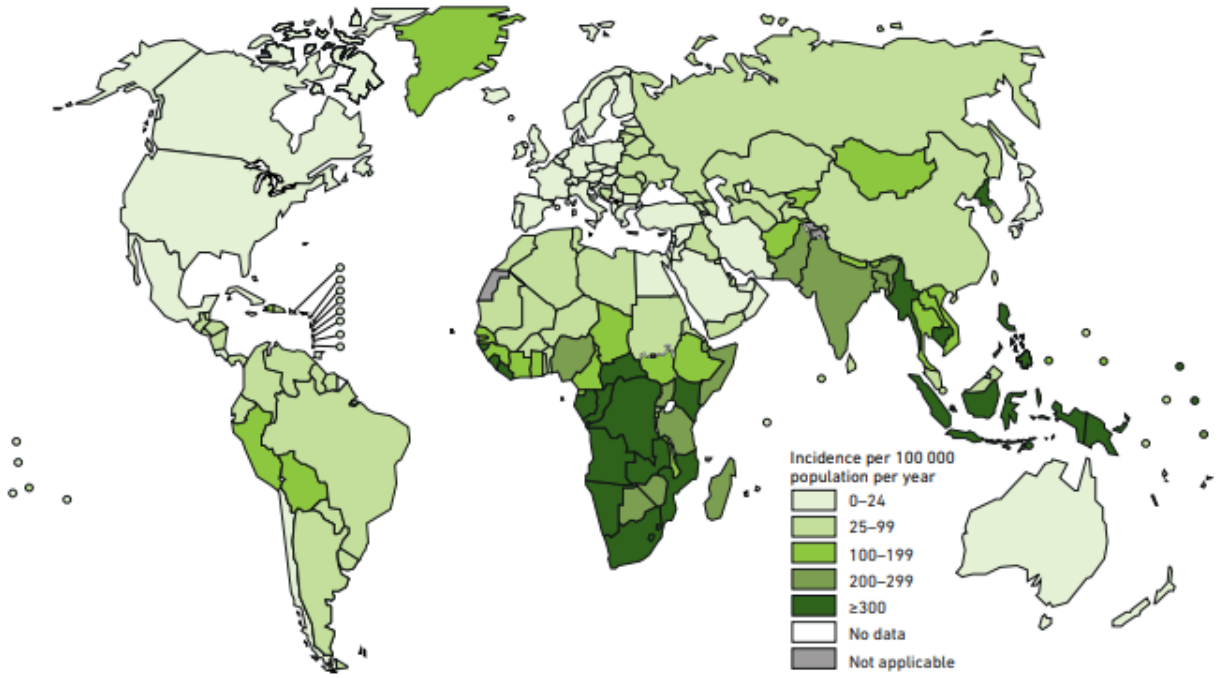


FIGURE 1.1 – Estimated TB incidence rates, 2017 [11]

1.3 Pulmonary Tuberculosis

Pulmonary TB is caused by the bacterium Mycobacterium tuberculosis (M tuberculosis). It affects the lungs and it may spread to other organs. Pulmonary TB is highly contagious as a person may get TB by breathing in air droplets of an infected person [12].

1.4 Types of Pulmonary Tuberculosis

There exist five types of pulmonary tuberculosis namely (infiltrating, focal, tuberculoma, miliary and fibro-cavernous). A 2D slice of CT scan for each type of pulmonary TB is shown in Figure 1.2.

Miliary TB Miliary tuberculosis (TB) is the widespread dissemination of Mycobacterium tuberculosis via hematogenous spread. Classic miliary TB is defined as millet-like (mean, 2 mm; range, 1-5 mm) seeding of TB bacilli in the lung, as evidenced on chest radiography. This pattern is seen in 1-3% of all TB cases. Symptoms may include fever, night sweats, and weight loss. It can be difficult to diagnose because the initial chest x-ray may be normal. Patients who are immunosuppressed and children who have been exposed to the bacteria are at high risk for developing miliary TB.[13, 14, 15, 16]

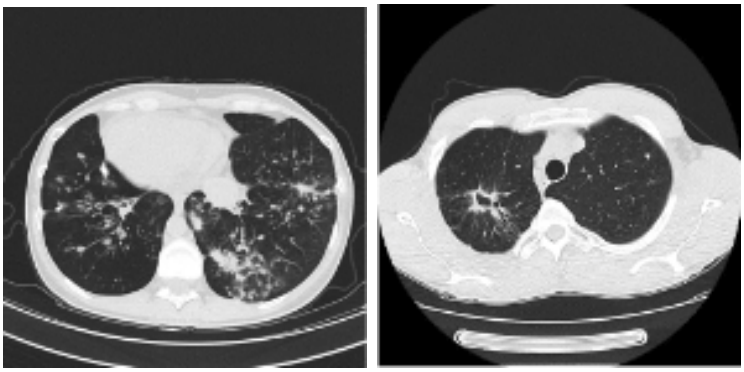
Focal TB Focal TB is diagnosed in the event of lesions not exceeding two lung segments. A small number of pathological foci with a diameter of about 1 cm are formed in the lung, which occurs at different times.[17]

Infiltrative Tuberculosis infiltrative tuberculosis is characterized by the fuzzy outlines of shadows on radiography. This often happens when forming new cavities. infiltration appears on radiography as rounded, or shade of homogeneous structure. All spots are divided into small, medium and large. The small ones have a size of 1 to 2 cm, the means 2-4 cm, the large ones up to 6 cm.[18]

Tuberculoma TB A tuberculoma is a clinical manifestation of tuberculosis which unites tubercles into a solid chunk, and so can mimic cancer tumors of many types in medical imaging studies.[19, 20]

Fibrous-cavernous TB Distinctive features of cavernous form of lung tuberculosis are the presence of the thin-walled cavity located on a background of slightly changed lung tissue at the absence of the expressed infiltrative and fibrotic changes. Cavernous tuberculosis develops among patients with, disseminated, focus lung tuberculosis, at the disintegration of tuberculomas; at the late revealing of disease, when the phase of disintegration is finished by the formation of cavities, and the attributes of the initial form disappear. At radiographic examination rounded cavity is defined, with a thin two-layer wall and usual localization in subclavicular area.[21]

(A) Infiltrative (B) Focal (C) Tuberculoma



(D) Miliary (E) Fibro-cavernous

FIGURE 1.2 – CT slices of the five pulmonary TB types.[8]

Summary

As shown in this chapter, tuberculosis is chronic and may result in the death of the infected person. Furthermore, TB has several types that are diagnosed and treated differently. In the next chapter, the different medical imaging technologies used to diagnose TB are presented along with the role of machine learning to analyze medical imaging. A particular focus will be put on deep learning and its architectures and different evaluation metrics.

Chapter 2

Medical Imaging and Machine Learning

2.1 Medical imaging

Medical imaging encompasses the technologies and processes of creating visual representations out of an anatomical volume in a form of images to be used for clinical diagnosis, medical intervention, and disease monitoring.

2.1.1 Medical imaging technologies

Common medical imaging technologies include ones which use electromagnetic radiations such as X-ray imaging and Computed Tomography (CT) imaging, others use sound waves and magnetic field such as ultrasound an magnetic resonance imaging.

X-ray imaging this technology is the oldest and the most frequently used, it works on wavelengths and frequencies that can penetrate through the skin creating a visualization of the inner body. It used to detect skeletal system malfunctioning, cancer through mammography, and other diagnoses that involve the visualization of the inner body. This technique comes with risks associated with the use of X-ray radiation.

CT imaging is a form of X-ray imaging that produces a 3D visualization of for diagnosis, providing greater quality and detailed imaging of the internal organs, bones, blood vessels, soft tissues within the body. it also inherits the risk of X-ray imaging whereas the benefits exceed its risk where in many cases the use of CT scans prevents the need for exploratory surgery.

Ultrasound imaging uses High-frequency sound waves that are transmitted from the probe to the body via the conducting gel, those waves then bounce back when they hit the different structures within the body and that is used to create an image for diagnosis. This technology is considered the safes without any recorded side effects of its usage and is the most cost-effective. Due to its low risk, it is the first choice for pregnancy.

Magnetic resonance imaging uses a strong magnetic field and radio waves it enables an in-depth view of the inside of a joint or ligament to be seen, rather than just the outside as in the case of CT scans and X-ray. It has risks associated with the use of a

strong magnetic field where any kind of metal implant, the artificial joint could be moved or heated up within the magnetic field.

2.1.2 Medical imaging Archiving and Recording

Due to the nature of a medical image, storing it is different from storing regular images. Medical image data set consists typically of one or more images representing the projection of an anatomical volume onto an image plane (projection or planar imaging), a series of images representing thin slices through a volume (tomographic or multislice two-dimensional imaging), a set of data from a volume (volume or three-dimensional imaging), or multiple acquisitions of the same tomographic or volume image over time to produce a dynamic series of acquisitions (four-dimensional imaging).[22]

There exist several medical images file formats all of them sharing the goal of standardizing medical images storage and transmission. Major file formats widely used in medical imaging are Analyze, Neuroimaging Informatics Technology Initiative (Nifti), Minc, and Digital Imaging and Communications in Medicine (Dicom).

Analyze Analysis 7.5 was created in the late 1980s as a format used by the Analyze commercial software developed at the Mayo Clinic in Rochester, MN, USA. For more than a decade, the format was the standard for post-processing medical imaging. The big point of view of the Analyze format is that it was designed for multidimensional (volume) data. Indeed, it is possible to store 3D or 4D data in a file (the fourth dimension being typically the temporal information). An Analyze 7.5 volume includes two binary files: an image file with the extension .img which contains the raw voxel data, and a header file with the extension .hdr which contains the metadata (number of pixels in the three dimensions, voxel size, and data type). The header has a fixed size of 348 bytes and is written as a structure in C programming language. Reading, and editing the header requires a utility software. The format is now considered "old" but it is still widely used and supported by many processing software, viewers, devices, and conversion utilities.[22]

Nifti Nifti is a file format created in the early 2000s with the intention to create a format that preserves compatibility of the Analyze format but solving its weaknesses. Nifti may be considered a revised 'Analyze' format. NIfTI uses the "empty space" in the ANALYZE 7.5 header to add several new features such as image orientation with the intention of avoiding left-right ambiguity in the brain study. In addition, Nifti includes unsupported data type in the Analyze format as an unsigned 16-bit format. Although the format also allows the storage of header and pixel data in separate files, the images are usually saved as a single '.nii' file in which the header and data are stored. pixels are merged. The header has a size of 348 bytes in the case of data storage '.hdr' and '.img' and a size of 352 bytes in the case of a single file '.nii'. This difference in size is due to the presence of four additional bytes at the end, essentially to make the size a multiple of 16, and also to provide space for storing additional metadata.[23, 22]

Minc The Minc file format was developed in 1992 to provide a flexible data format for medical imaging. The first version of the Minc format (Minc1) was based on the standard common network format (NetCDF). Subsequently, to overcome the large data file support constraint and provide new features. Minc's development team chose to

upgrade from NetCDF to Hierarchical Data Format version 5 (HDF5). This new version which is not compatible with the previous one was called Minc2.[24, 22]

Dicom Dicom (Digital Imaging and Communications in Medicine), is the international standard for transmitting, storing, retrieving, printing, processing, and displaying medical imaging information. Dicom can only store pixel values as an integer. However, it supports various types of data, including floats, to store metadata. Whenever the values stored in each voxel are to be scaled, Dicom uses a scaling factor using two fields in the header defining the slope and the intercept of the linear transformation to be converted in real values. Dicom supports compressed image data through a mechanism that encapsulates a non-Dicom document into a Dicom file. The compression systems supported by Dicom are JPEG, Run-Length Encoding (RLE), JPEG-LS, JPEG-2000, and MPEG2 / MPEG4.[25, 22]
Table 2.1 shows a summary of file formats characteristics

Format	**Header**	**Extension**	**Data types**
Analyze	Fixed-length: 348 byte binary format	.img and .hdr	Unsigned integer (8-bit), signed integer (16-, 32-bit), float (32-, 64-bit), complex (64-bit)
Nifti	Fixed-length: 352 byte binary format (348 byte in the case of data stored as .img and .hdr)	.nii	Signed and unsigned integer (from 8- to 64-bit), float (from 32- to 128-bit), complex (from 64- to 256-bit)
Minc	Extensible binary format	.mnc	Signed and unsigned integer (from 8- to 32-bit), float (32-, 64-bit), complex (32-, 64-bit)
Dicom	Variable-length binary format	.dcm	Signed and unsigned integer, (8-, 16-bit; 32-bit only allowed for radiotherapy dose), float not supported

TABLE 2.1 – Summary of file formats characteristics [22]

2.2 Machine Learning (ML)

Machine learning is a branch of artificial intelligence that encapsulates the methods and algorithms that automates analytical and mathematical model building. ML is based on the idea that machines can learn from given data to solve or react to a certain input without being explicitly programmed[26]. One major task of machine learning, pattern recognition, and data mining is to construct good models from data sets.

Thanks to the rise of powerful and affordable computing power and the appearance of big data. machine learning is witnessing an exciting evolution. Nowadays, machine

learning techniques are being used in almost every field of research and application ranging from healthcare, medicine, agriculture, and finance . . . , etc. As a result of these advances, systems which only a few years ago performed at noticeably below-human levels can now outperform humans at some specific tasks and complex games such as chess and Shogi[27, 28].

Machine learning algorithms are organized in categories based on the type of learning and whether data is available, labeled or not. Thus, we distinguish supervised learning, semi-supervised learning, unsupervised learning, Reinforcement learning algorithms.

2.2.1 Supervised learning

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs[29]. In supervised learning, the inference of a function is done from labeled training data [30]. Consequently, the availability of labeled training examples is mandatory.

There exist numerous supervised learning algorithms each has its strengths and weaknesses. for now, there isn't an algorithm that can perform best in every situation[31]. Support vectors machine (SVM) and decision trees are well-known examples of supervised learning algorithms and will be discussed next.

Support Vectors Machine (SVM)

A Support Vector Machine (SVMs, also support-vector networks) is a discriminative classifier formally, it constructs a hyperplane or set of hyperplanes in a high or infinite-dimensional space, which can be used for classification, regression, or other tasks like outliers detection. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. SVMs are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.[32, 33] SVMs are widely used in pattern recognition and classification tasks [34, 35, 36, 37, 38] and nonlinear regressions [39, 40]

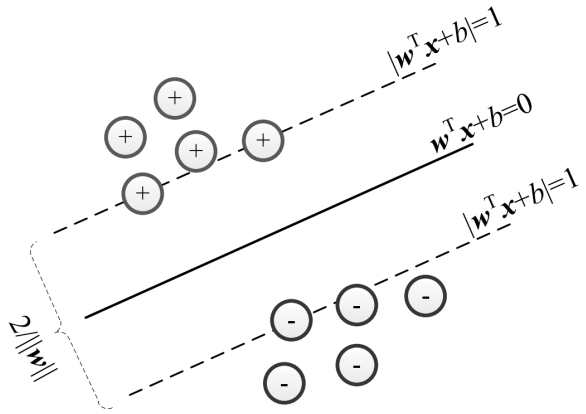


Figure 2.1 – An illustration of an SVM machine.

Decision Trees

A decision tree represents a function that takes as input a vector of attribute values and returns a "decision"—a single output value[29]. It is one of the predictive modeling approaches used in statistics, data mining and machine learning. It works as a decision support tool. A decision tree is a flowchart-like structure, works by recursively splitting training data into subsets based on the value of a single attribute. Each split corresponds to a node in the tree representing a test on the attribute/features, each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes). A basic structure of a decision tree is shown in Figure 2.2.[41, 42] There exist different algorithms for building a decision tree, namely ID3 (Iterative Dichotomiser 3), C4.5 (successor of ID3), CART (Classification And Regression Tree), CHAID (Chi-squared Automatic Interaction Detector) and MARS. The algorithms and techniques of building decision trees are an exciting research domain, where the speed, efficiency, and accuracy are improved [42]. Decision trees are used in many fields including medecine[43, 44], astronomy[45], and genetics[46]... etc.

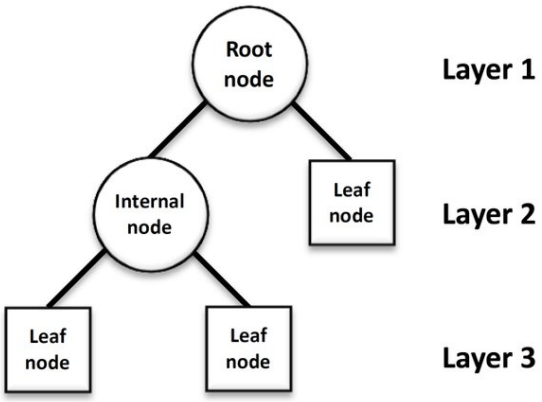


FIGURE 2.2 – Basic structure of a decision tree.

2.2.2 Ensemble learning

In ordinary machine learning algorithms. the intention is to learn one hypothesis from training data. Whereas, in ensemble methods[47] multiple learners that are called base learner. Base learners are trained to solve the same problem. Thus, a collection of hypothesizes is constructed and being combined for a prediction. For instance, during cross-validation, we might generate twenty different decision trees, and have them vote on the best classification for a new example. Ensemble learning is also called committee-based learning or learning multiple classifier systems[29]

Figure 2.3 shows a common architecture of ensemble learning. There exist two types of ensembles, those that use a single base learning algorithm to produce homogeneous base learners, leading to homogeneous ensembles, i.e, an ensemble of decision trees. The second type consists of the use of different base learning algorithm to produce heterogeneous base learners, leading to heterogeneous ensembles, i.e, an ensemble of neural networks and SVMs.

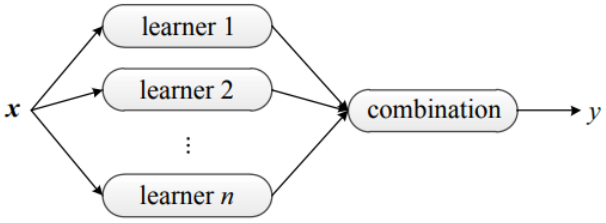


FIGURE 2.3 – A common ensemble architecture.

The use of an ensemble leads often to a much stronger generalization than the use of a single learner. Ensemble methods are proven to boost weak learners and turn them to strong learners with more accuracy.[48]

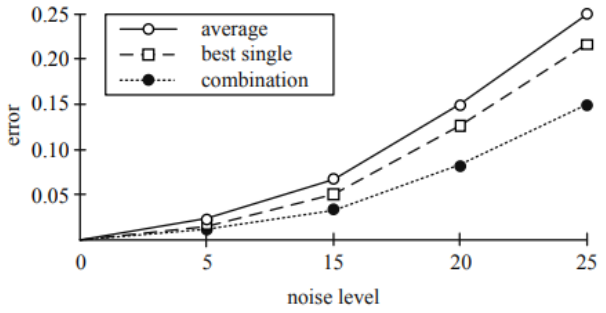


FIGURE 2.4 – A simplified illustration of Hansen and Salamon [1990]'s observation: Ensemble is often better than the best single.

Boosting and bagging are methods of ensemble learning which are used to improve the stability and accuracy of base learners. They're briefly described below.

Bagging stands for bootstrap aggregation which is a method for generating multiple versions of a predictor and using these to get an aggregated predictor[49]. Random forest is a well-known example of an ensemble learning algorithm. It uses the combination of bagging technique[50] and random selection of features on the decision tree as the base learner to produce a collection of decision trees thus constructing a random decision forest.

Boosting consists of iteratively learning weak classifiers with respect to a distribution and adding them to a final strong classifier. Moreover, boosting is based on the question "Can a set of weak learners create a single strong learner?"[49].

Stacking it is a technique that is used to ensemble a diverse group of strong learners by training a second-level machine learning algorithm called a "meta-learner" to learn the optimal combination of predictions of the base learners.

2.2.3 Deep learning

Deep learning is a method of machine learning based on learning data patterns and representations for the goal of artificial intelligence it is also known as deep structured

learning or hierarchical learning. The hierarchy of nature enables the computer to learn complicated concepts by building them out of simpler ones through the use of layers. the graphical representation of these concepts shows deep graphs, with many layers. That is why it's called deep learning.

There exist many architectures involving deep learning such as deep neural networks, deep belief networks, deep Boltzmann machines and recurrent neural networks which have been applied in a wide range of domains, like Computer vision, natural language processing, education, finance ... etc. due to the vast domain of deep learning, A particular focus is given to deep neural networks architecture in this paper.[51]

Deep neural networks have shown a state of the art performance that has beaten many ML algorithms, especially in computer vision. Deep networks could be categorized into four major network architectures groups: Unsupervised Pretrained Networks (UPNs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks, Recursive Neural Networks[52]. Widely used examples of CNN architectures are GoogleNet[53], ResNet[54], VGGNet[55], AlexNet[56].

2.2.4 Transfer learning

Transfer learning is a machine learning method where a model developed and trained for a task is reused as the starting point for building and training a model on another task. Transfer learning is meaningful to be used when the target Task dataset is relatively smaller than the dataset that the model was pre-trained on and when the inputs are homogeneous, i.e, images and X-ray scans. It helps to take advantage of the pre-learned structures and features and decrease the time of training the model.

2.2.5 Evaluation metrics

Evaluation metrics are used to measure the performance of a machine learning model. There exist different evaluation metrics used for benchmarking or fine-tuning a model. Classification accuracy, confusion matrix and receiver operating characteristic (ROC) are briefly described in the next section.

Classification Accuracy Classification Accuracy is the ratio of the number of correct predictions to the total number of input samples. It is calculated by Equation 2.1.

$$Accuracy = \frac{\text{Number of Correct predictions}}{\text{Total number of predictions made}} \tag{2.1}$$

Confusion Matrix A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") over a set of test data for which the true values are known. In the case of a binary classifier, we get Table 2.2.

- true positives (TP): These are cases in which the predicted class is 1, and the actual class is 1.
- true negatives (TN): These are cases in which the predicted class is 0, and the actual class is 0.

		Predicted	
		1	**0**
Actual	**1**	TP	FN
	0	FP	TN

TABLE 2.2 – A confusion matrix of a binary classification.

- false positives (FP): These are cases in which the predicted class is 1, and the actual class is 0.
- false negatives (FN): These are cases in which the predicted class is 0, and the actual class is 1.

Root Mean Square Error (RMSE) Root Mean Square Error (RMSE) is the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are; RMSE is a measure of how spread out these residuals are. In other words, it tells you how concentrated the data is around the line of best fit. Root mean square error is commonly used in regression analysis to verify experimental results and machine learning models. Its formula is:

$$\mathrm{RMSE}_{fo} = \left[\sum_{i=1}^{N} \left(z_{f_i} - z_{o_i} \right)^2 / N \right]^{1/2} \tag{2.2}$$

ROC curve The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) which are calculated from the output of confusion matrix at various threshold settings. The true-positive rate is also known as sensitivity, recall or probability of detection in machine learning. The false-positive rate is also known as the fall-out or probability of false alarm and can be calculated as (1 specificity).

2.3 Application of deep learning in medical imaging

Recent advancements in deep learning have revolutionized medical image analysis by allowing discovering morphological and/or textural patterns in images directly from data. As deep learning models have achieved state of the art performance over different medical applications. This achievement took the attention of researcher and practitioners where many methods and frameworks gathering the field of medical imaging and deep learning were created. Niftynet is one example of the platforms that we will discuss next.

2.3.1 Niftynet

NiftyNet is a deep learning platform for medical imaging built on top of The TensorFlow framework[57]. It comprises several modular components that ease and abstracts the model building pipeline. The NiftyNet works by connecting four components: a Reader to load data from files, a Sampler to generate appropriate samples for processing, a Network to process the inputs, and an output handler (comprising the Loss and Optimizer during training and an Aggregator during inference and evaluation). These components are briefly depicted in Figure2.5.[58]

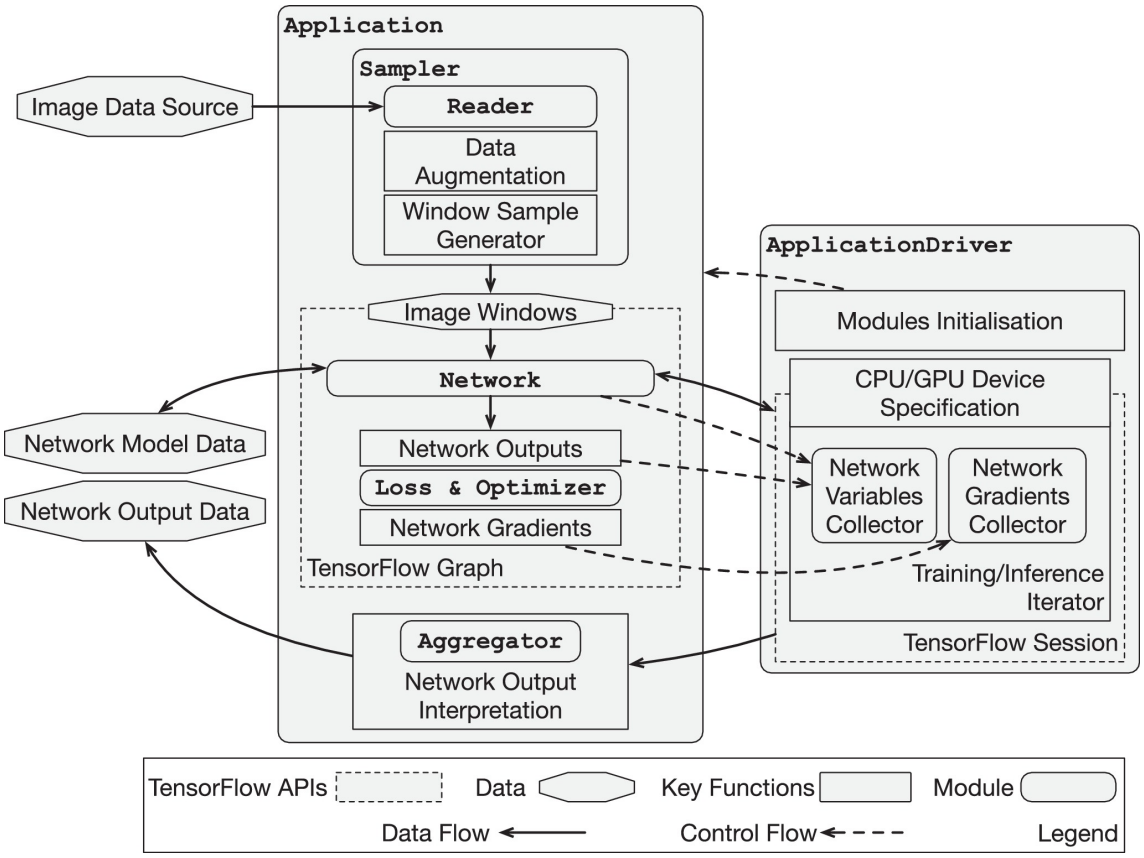


FIGURE 2.5 – A brief overview of NiftyNet components[58].

Summary

In this chapter, the different medical imaging technologies used to diagnose TB were presented along with the role of machine learning in medical imaging analysis. A particular focus was put on deep neural networks and its architectures. In the next chapter, the severity scoring subtask organized by the evaluation campaign ImageCLEF will be presented along with the organizer. Moreover, the dataset, evaluation metrics used will be looked at with f the top 3 and MostaganemFSEI teams approaches to automatic tuberculosis severity scoring.

Chapter 3

Tuberculosis severity scoring using machine learning

The severity score is a cumulative score of severity of TB case assigned by a medical doctor. Originally, the score varied from 1 ("critical/very bad") to 5 ("very good"). In this subtask, the score value is simplified so that values 1, 2 and 3 correspond to "high severity" class, and values 4 and 5 correspond to "low severity". In the process of scoring, the medical doctors considered many factors like the pattern of lesions, results of microbiological tests, duration of treatment, patient's age and some other.

This task was proposed by ImageCLEF which is an evaluation campaign that is being organized as part of the CLEF initiative labs. SVR task lays under the ImageCLEFmed Tuberculosis category of tasks along with CT report task (CTR) which consists of generating an automatic report based on the CT image of a patient.

This task possesses different difficulties and challenges. The small amount of dataset and examples for the use of deep learning, the large dimensions of the image slices without the availability of an expert to assist in reducing the irrelevant slices and decreasing the dimensions. The lack of powerful computation resources makes it difficult to train a machine learning model, particularly a deep neural network.

In the next sections, we will describe the dataset being used for ImageCLEF 2019 SVR task and discuss the participating teams' approaches and the results obtained for the severity scoring of ImageCLEF 2018.

3.1 Dataset

The dataset contains 335 chest 3D CT images with an image size per slice of 512*512 pixels and number of slices varying from about 50 to 400 of TB patients. All the CT images are stored in NIFTI file format with .nii.gz file extension (gzipped .nii files). In addition to the CT images, a set of clinically relevant metadata was given. The selected metadata includes the following binary measures: disability, relapse, symptoms of TB, comorbidity, bacillary, drug resistance, higher education, ex-prisoner, alcoholic, smoking, severity. Moreover, automatically extracted lung masks were provided along with a set of tools for loading/saving nifti files. 218 patients are used for training and 117 for the test dataset.

3.2 Participation and results

In 2018 there were 7 teams that participated and submitted their result in the SVR task. Each team used its own approach to build a solution for the SVR task. the participants were allowed to submit up to 10 runs for the task. A run consists of submission of the results of the developed model when applied on a test set provided by imageCLEF. The runs submitted for the severity scoring subtask were evaluated in two ways. One used the original severity scores from 1 to 5 and the task was to predict those numerical scores as precise as possible (a regression problem). Here, Root Mean Square Error (RMSE) was computed between the ground truth severity and the predicted scores provided by the participants. Alternatively, the original severity score was transformed into two classes, where scores from 1 to 3 correspond to "high severity" and the 4 and 5 scores correspond to the "low severity" class. In this case, the participants had to provide the probability of TB cases to belong to the "high severity" class. The corresponding results were evaluated using AUC. in this section we present the 3 top performing approaches in terms of Root Mean Squared Error (RMSE) along with the approach used by MostaganemFSEI team.

Top RMSE achieved was 0.7840, by the team UIIP_BioMed with a single run. A Coder-Decoder Convolutional Neural Network trained on a third-party dataset of 149 CT scans with lesions labeled by a qualified radiologist was used as a Lesion-based TB-descriptor extraction method. The lesion-based TB-descriptors extracted from the patients CT scans were used to generate a random forest of 100 trees as a classifier[59]. This approach had the best RMSE 0.7840 and an AUC of 0.7708. The second best RMSE was 0.8513, achieved by the team MedGIFT among 9 runs. A graph model of the lung based on regional 3D texture features was used in an SVM classifier [60]. This method resulted in a better AUC than the previous approach. VISTA@UEvora team ranked 3rd with an RMSE of 0.8883 among 7 runs. The team used a 3D modeling and further extraction of texture patterns approach with a multi-layer perceptron algorithm[61]. Finally, the team MostaganemFSEI used an approach that consists of extracting a single image semantic descriptor for each CT scan/patient instead of considering all the slices as separate samples [62]. The semantic descriptor was then passed as an input to bagging of a set of Random forest learners to perform a hierarchical classification of the severity score. With this approach, the team achieved an RMSE of 0.9721 and an AUC of 0.5987.

		RMSE		**AUC**	
Group Name	**Run**	**RMSE**	**Rank**	**AUC**	**Rank**
UIIP_BioMed[59]	SVR_run_TBdescs2_zparts3_thrprob50_rf100.csv	0.7840	1	0.7025	6
MedGIFT[60]	SVR_HOG_std_euclidean_TST.csv	0.8513	2	0.7162	5
VISTA@UEvora[61]	SVR-Run-07-Mohan-MLP-6FTT100.txt	0.8883	3	0.6239	21
MostaganemFSEI[62]	SVR_mostaganemFSEI_run3.txt	0.9721	12	0.5987	25

Table 3.1 – Results obtained by the top 3 participants and MostaganemFSEI team in the SVR subtask [8].

Chapter 4

Contributions

Introduction In this chapter, we will present our contribution by describing the approach used and discussing the results of the performance of our models. The experiments described were also submitted for the ImageCLEF 2019 Tuberculosis severity scoring competition. n addition, the work resulted in the submission of a conference paper. The ranking and the performance of the models resulted by the submissions to ImageCLEF 2019 SVR will be presented.

We proposed the system presented in Figure 4.1. The latter goes through two essential steps: input data pre-processing and learning a classification model. A third optional step is added in order to improve the performance of the first learning step. The latter includes a second learning stage by using a recurrent neural network (LSTM) or by generating semantic features and exploiting them through a learner or deep learner. We will detail our proposed system in the following.

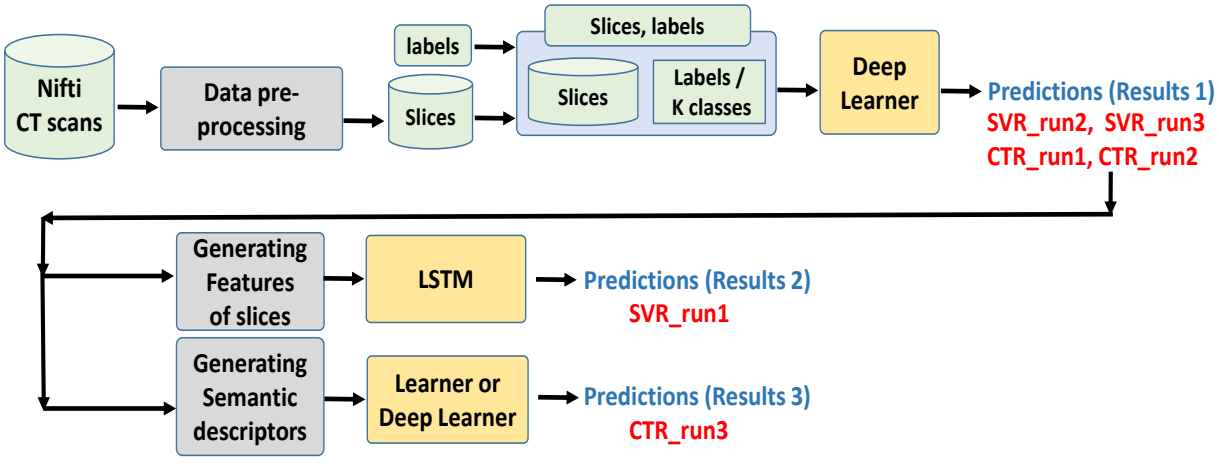


FIGURE 4.1 – The architecture of the overall proposed system

4.0.1 Input data pre-processing

3D CT scans are provided in compressed Nifti format. Firstly, we decompress the files and extract the slices. In the end, we have three sets of slices corresponding to the three dimensions of the 3D image. For each dimension and for each Nifti image we obtain a number of slices ranging according to the dimension considered (512 images for the Y and X dimensions, from 40 to 250images for the Z dimension).

The visual content of the images extracted from the different dimensions is not similar. Indeed, the images of each dimension are taken from a different view angle. We noticed from our experiments that the slices of the -Z- dimension give better results compared to the two others (X and Y). This is why we used in our work the Z-dimension. However, all steps can be applied to slices of any of the three dimensions.

After choosing the dimension to consider, we propose to filter the slices of each patient. Indeed, we can notice that many slices do not necessarily contain relevant information that could help to classify the samples due to their resemblance along the three dimensions. This is why we add a step to filter and select a number of slices per patient. For this, we propose two filtering approaches:

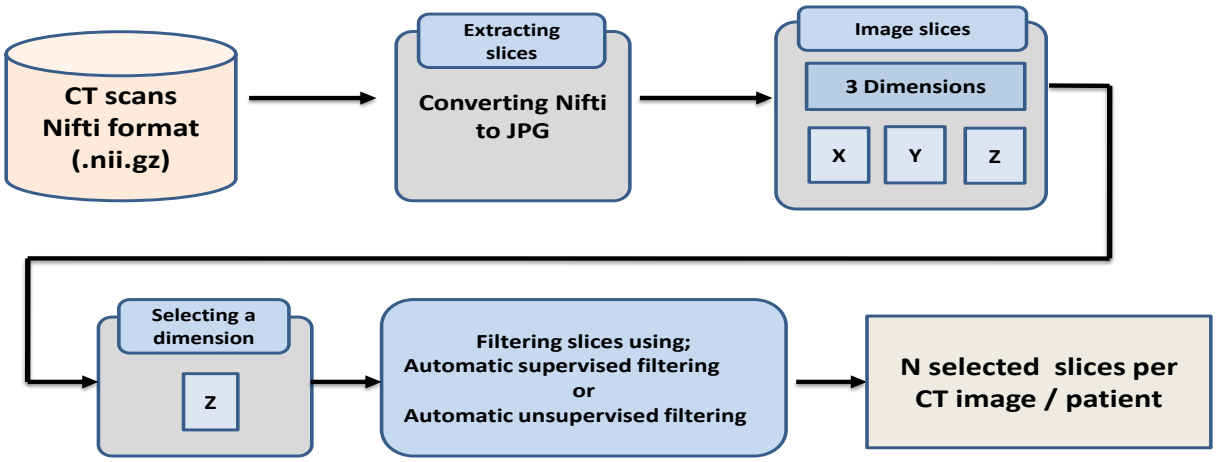


FIGURE 4.2 – The architecture of the overall proposed system

Automatic supervised filtering in this approach, we select a set of patients from each of the considered classes (the five degrees of severity for the SVR task). Then, a human expert selects for each patient, the slices likely to contain relevant information indicating the presence of Tuberculosis. The resulting set of slices constitutes a filtering group. Given a new patient, we compare each of its slices to the filtering group by calculating a distance measure: a weighted sum of distances between the slice and those of the filtering group. We selected at the end N slices that are judged to be the most similar to the filtering group. So, in the end, each patient is represented by the N filtered slices instead of all its extracted images. We think that this would reduce the noise introduced by the consideration of all slices. We tested in our contributions the value N=10.

Automatic unsupervised filtering we noticed that there is usually a maximum of 50/60 slices visually informative. Since the slices are ordered, these most informative slices are usually at the center of the list. We propose then to keep only the N middle ones. This is not optimal but we opted for this choice for a fully automatic and unsupervised approach. This choice can be improved by performing manual filtering with the intervention of a human expert, preferably with medical skills on TB disease.

Figure 4.2 summarizes the pre-processing steps.

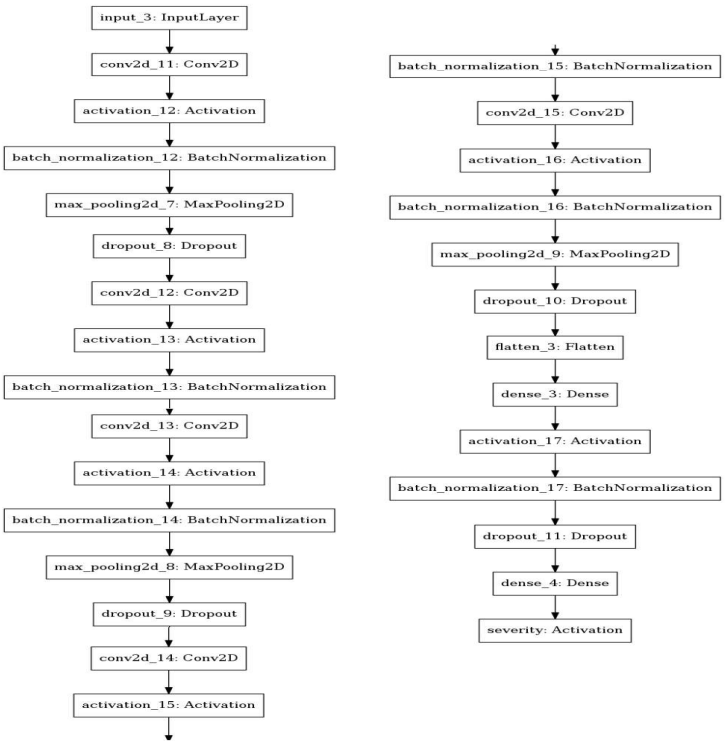


FIGURE 4.3 – The architecture of the LungNet Deep Learner

4.0.2 Deep learning models for CT image classification

As a deep learner, we chose to use Resnet-50 architecture because of its good results in the context of the same problematic in last Tuberculosis task editions [63] and the google inceptionV3Resnet model. On the other hand, we developed a model that we called LungNet. We present more details about this model in the following section.

LungNet Deep Learner we proposed and developed our deep learner architecture for CT Image Analysis that we called LungNet. The input to the latter is an RGB image of size 119x119, followed by five convolutional layers and two fully connected layers. After each convolutional layer a "relu" activation is applied followed by a local normalization and max pooling. the first, the second and third convolution blocks have dropout layers to reduce overfitting. The sigmoid activation function is applied to the output layer in order to predict values in the range of 0 to 1. Figure 4.3 illustrates the architecture of the Lungnet model.

4.1 Experiments and results

We describe in the following sections the different experiments performed for the SVR tasks.

We used in our experimental work the following tools:

- med2image [64] for the conversion of nifti medical images to the classic Jpeg format;
- Tensorflow frawework [57] and Keras library for deep learning;
- scikit-learn [65] library for testing several machine learning techniques.

Table 4.1 – Dataset given for Tuberculosis SVR and CTR tasks [67].

	Train Collection	Test Collection
Number of patients	218	117

- tools and scripts developed by us [66].

We chose to use slices of the -Z- dimension because our experiments showed that they are more suitable than dimensions -X- and -Y- and got better results.

Dataset:

The dataset used in the SVR task includes chest CT scans of TB patients along with some metadata regarding a set of 19 classes. 2 classes concern the SVR task, six other classes concern CTR task. The other values are considered as additional information regarding the patients that could be used as contextual information.

Table 4.1 summarizes the number of CT scans for train and test collections.

The same dataset is given for the CTR task. The samples are labeled regarding eight main target classes:

1. Target classes for SVR Task:

   (a) SVR_severity (binary class: HIGH and LOW). Another label called md_Severity is given (Five discrete values ranging from 1 to 5). We remind that values of md_Severity (1, 2 and 3) belong to the "HIGH" Severity class. The other two values (4 and 5) correspond to the "LOW" Severity class.

Experimental protocol:

We used the train collection provided by the organizers and we split it into two sub-collections: training and validation sets. The performed experiments are described below:

- **Resnet**: results of resnet50 deep learner trained on 50% of data. Each patient was represented by 50 slices filtered using the automatic unsupervised filtering approach that was described in section 4.0.1;

- **Lungnet**: results of LungNet deep learner trained on 80% of data. Each patient was represented by 10 slices filtered using the automatic supervised filtering approach that was described in section 4.0.1;

- **InceptionV3Resnet**: results of InceptionV3Resnet deep learner trained on 80% of data. Each patient was represented by 50 slices filtered using the automatic unsupervised filtering approach that was described in section 4.0.1;

In the case of Resnet50 and InceptionV3Resnet models, a hierarchical classification problem was considered.whereas, in the case of LungNet model, it was a direct binary classification.

	Accuracy		AUC		Ranking
	Validation	Test	Validation	Test	
Resnet50	0.6300	0.6154	0.3400	0.6510	22
LungNet	0.5220	0.6103	0.5260	0.5983	33
InceptionV3Resnet	0.5	0.4701	0.4431	0.4933	48

TABLE 4.2 – Results on validation data and test set for SVR task (ImageClef submissions)

Hierarchical classification firstly, the probabilities of the 5 classes were generated by the model for n slice of the patient's CT scan. We will end up with a list of arrays holding the 5 class probabilities for each n slice. Then, an Argmax function is applied to each probability array to get the highly predicted class for each slice. This will result in a vector of values that represent the highly predicted class for each slice. we then choose the class that has the highest occurrence to be qualified as an elected class.the final step is to choose from the probability arrays list, the arrays that have the same argmax value as the elected class. Then, a column-wise mean is calculated for the list of elected arrays, resulting in a mean array of probabilities. Summing up the probabilities of 1st, 2nd and 3rd classes is what we considered as the probability of High severity class.

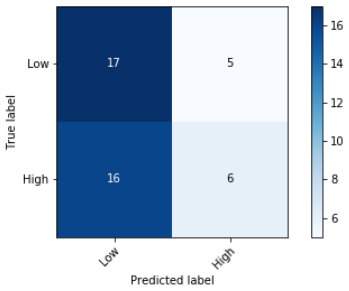
Binary classification the model was designed and trained to predict the probability of High severity class directly.

Our tools and scripts used in our experiments are accessible in [66].

Results:

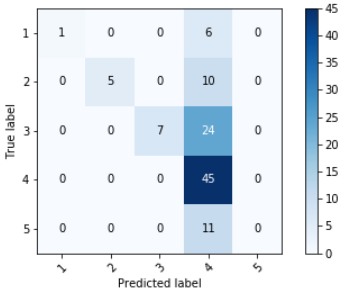
Table 4.2 shows the results in terms of AUC and accuracy obtained by our experiments on the evaluation performed by the ImageCLEF committee on test collection and our validation data.

Confusion matrices:

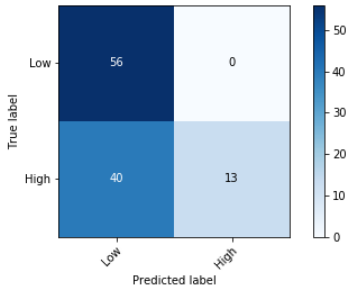


(A) High-Low classes

FIGURE 4.4 – Lungnet validation confusion matrix

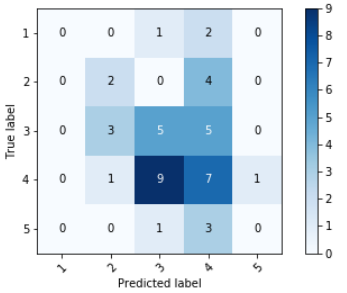


(A) 5-degrees classes

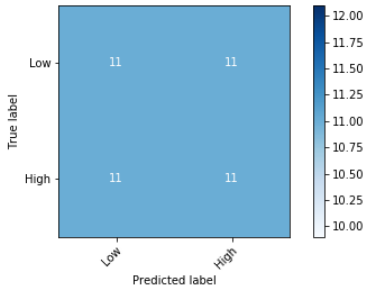


(B) High-Low classes

FIGURE 4.5 – Resnet validation confusion matrices



(A) 5-degrees classes



(B) High-Low classes

FIGURE 4.6 – InceptionV3Resnet validation confusion matrices

The Resnet50 model got the best accuracy in validation and test datasets, followed by our developed model Lungnet with a small difference in term of test accuracy. the two models have outperformed the InceptionV3Resnet model. It is important to state that the LungNet model is a way smaller model than the resnet50 and InceptionV3Resnet and has been trained on less data than both models and yet achieved good results compared to those achieved by the other models. it would be interesting to see the performance of the Lungnet model after training it on 50 slices in order to make a detailed comparison with the perform of the other two models.4.7 illustrates the ranking of the participating teams of the SVR task in terms of AUC and accuracy. We believe that our models could give better results after a more applying advanced data preprocessing including the use of masks, samples selection and data augmentation.

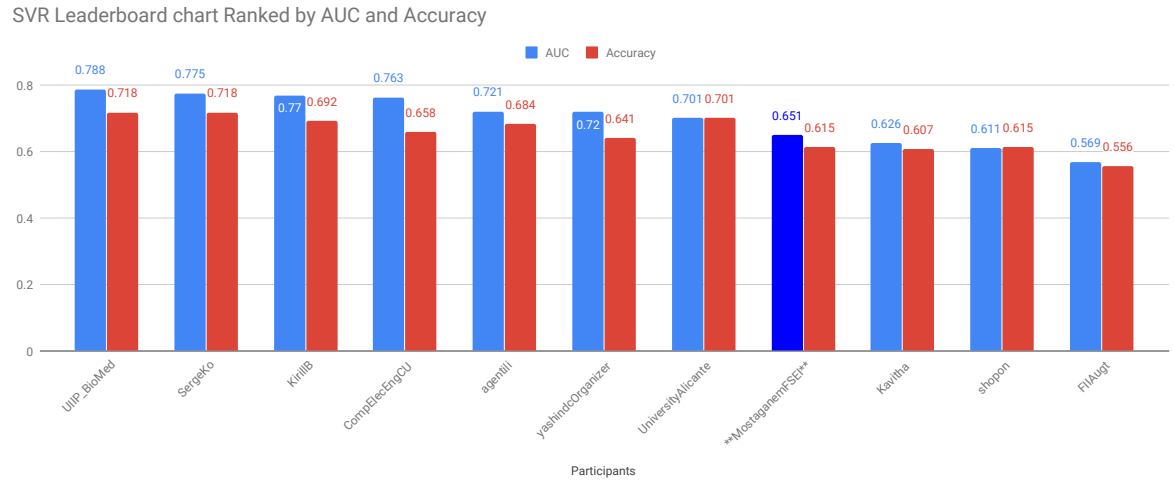


FIGURE 4.7 – Ranking of SVR task participants

4.2 Conclusion

We have described in this chapter our experiments of making a machine learning model to score the severity of a Tuberculosis case from a patient's CT scan and presented our participation submissions results in the SVR task of ImageCLEFmed Tuberculosis 2019. We proposed to use deep learner after preprocessing the CT scans in order to perform classification. We used for that, a Resnet-50, inceptionV3Resnet, and our proposed Lungnet architecture. Although our proposals had not been the best, the results obtained showed that these approaches could be much more efficient and give more interesting results if it is applied in an optimized way which includes advanced data-processing and the involvement of expert radiologists.

As perspectives, we plan to adopt enrichment strategies and learning samples selection. In addition, we noticed during the sub-sampling of our data that the deletion or addition of some samples had an impact on the results. On the other hand, filtering slices in an optimized way is a key idea that could further improve system performance. Furthermore, we noticed in our experiments that there is a difference in terms of precision achieved for each studied class. Indeed, some classes are more difficult to identify than others. This is

also an interesting track to study. The use of masks and tracking the visual contents of the lungs could also be a good trail to explore.

General Conclusion and future works

We have presented in this report the description of a severity scoring system for pulmonary tuberculosis. As a first step, we conducted a literature review on TB to get a clear idea about this disease its types and impact on the environment. Then, we broadly described various concepts related to our problematic, concerning medical imaging and machine learning.

We also presented the work that was done in the context of the International ImageCLEF 2018 challenge concerning the tuberculosis severity scoring task, which has shown promising results.

We have proposed as a contribution an approach that consists of using 3 different deep learners, namely Resnet50, InceptionResnet and our deep model LungNet. We participated to test our approach in the International Challenge ImageCLEF 2019 in SVR sub-task.

Using the Resnet50 deep learner We achieved an AUC of 0.6510 as our best result which is ranked 22nd out of 54 submissions. LungNet model achieved an AUC of 0.6103 ranking 33rd. We note that these results were achieved despite the lack of advanced preprocessing and filtering of slices.

However, the results obtained show that this approach could be much more efficient and give more potent results if it is applied correctly. As prospects, we plan to adopt augmentation strategies and selection of learning samples. Indeed, one of the characteristics of the problem addressed is the nature of the datasets provided, which are small and noisy due to the presence of many slices that do not contain useful information. Our prioritization and sub-sampling strategies adopted in our experiments confirm that.

In addition, we noticed during the downsampling of our data that the deletion or addition of certain samples had an impact on the results. On the other hand, the effective filtering of slices to keep only those that are really informative is a key idea that could further improve the performance of the system. Moreover, we noticed in our experiments that there is a difference of precision for each severity class studied which arises the hypothesis of the classes having varying difficulties to be identified by the model. Indeed, some classes are more difficult to identify than others. It is also an interesting track to study.

Abbreviations

3D 3 Dimensions

AUC Area Under Curve

BK Bacillus Koch

CNN Convolutional Neural Networks

CT Computed Tomography

CTR CT report task

Dicom Digital Imaging and Communications in Medicine

FN False Negatives

FP False Positives

HIV Human Immunodeficiency Virus

JPEG Joint Photographic Experts Group

LTBI Latent Tuberculosis Infection

Minc Medical Imaging NetCDF

ML Machine Learning

MPEG Moving Picture Experts Group

Nifti Neuroimaging Informatics Technology Initiative

RLE Run-Length Encoding

RMSE Root Mean Squared Error

ROC Receiver Operating Characteristic

SVM Support Vector Machines

SVR Severity scoring task

TB Tuberculosis

TN True Negatives

TP True Positives

UPN Unsupervised Pretrained Networks

WHO World Health Organization

List of Figures

List of Tables

Bibliography

[1] J. Allard and N. Atalla, "Automatic segmentation of liver structure in ct images using a neural network," *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications, and Computer Sciences*, vol. E77-A, no. 11, p. 892–1895, 1994.

[2] H. A, L. SCB, L. JS, F. MT, and M. SK, "A shift-invariant neural network for the lung field segmentation in chest radiography," *A Shift-Invariant Neural Network for the Lung Field Segmentation in Chest Radiography*, vol. 18, no. 3, pp. 241–250, 1998.

[3] O. Mehmed, D. B. M, and M. Robert, "Neural- network-based segmentation of multimodal medical images: A comparative and prospective study," *IEEE Transactions on Medical Imaging*, vol. 12, no. 3, pp. 534–544, 1993.

[4] K. J. E, N. FD, J. TK, and K. DL, "Abdominal organ segmentation using texture transforms and a hopfield neural network," *IEEE Transactions on Medical Imaging*, vol. 18, no. 7, pp. 640–648, 1999.

[5] R. WE, G. JO, C. EN, and et al, "Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks," *IEEE Trans Med Imaging*, vol. 16, no. 6, 1997.

[6] W. H. Organization, "Tuberculosis fact sheet," September 2018. `https://www.who.int/en/news-room/fact-sheets/detail/tuberculosis/` [Online; posted 18 September 2018].

[7] A. L. Association, "Learn about tuberculosis," December 2018. `https://www.lung.org/lung-health-and-diseases/lung-disease-lookup/tuberculosis/learn-about-tuberculosis.html` [Online;Updated December 14, 2018].

[8] Y. Dicente Cid, V. Liauchuk, V. Kovalev, , and H. Müller, "Overview of ImageCLEFtuberculosis 2018 - detecting multi-drug resistance, classifying tuberculosis type, and assessing severity score," in *CLEF2018 Working Notes*, CEUR Workshop Proceedings, (Avignon, France), CEUR-WS.org <http://ceur-ws.org>, September 10-14 2018.

[9] F. M. B. e. Puneet Sharma, *Image Analysis: 20th Scandinavian Conference, SCIA 2017, Tromsø, Norway, June 12–14, 2017, Proceedings, Part II*. Lecture Notes in Computer Science 10270, Springer International Publishing, 1 ed., 2017.

[10] A. Menegola, M. Fornaciali, R. Pires, S. E. F. de Avila, and E. Valle, "Towards automated melanoma screening: Exploring transfer learning schemes," *CoRR*, vol. abs/1609.01228, 2016.

[11] W. H. Organization, "Global tuberculosis 2018 report," 2018. http://apps.who.int/iris/bitstream/handle/10665/274453/9789241565646-eng.pdf.

[12] U. N. L. of Medicine, "Pulmonary tuberculosis." https://medlineplus.gov/ency/article/000077.htm.

[13] M. A, B. M, T. F, O. R, O. R, and S. H. et al, "Miliary tuberculosis: clinical manifestations, diagnosis and outcome in 38 adults," *Respirology*, vol. 6, p. 217–24, Sept. 2001.

[14] S. SK, M. A, S. A, and M. DK, "Miliary tuberculosis: new insights into an old disease," *Lancet Infect Dis*, vol. 5, pp. 415–30, July 2005.

[15] B. HM, B. WJ, C. RE, D. CL, E. SC, and F. L. et al, "American thoracic society/centers for disease control and prevention/infectious diseases society of america: treatment of tuberculosis," *Am J Respir Crit Care Med*, vol. 6, pp. 603–62, Feb. 2003.

[16] R. L. Hunter, "Pathology of post primary tuberculosis of the lung: an illustrated critical review.," *Tuberculosis*, vol. 91 6, pp. 16–17, 2011.

[17] M. Format, "Focal pulmonary tuberculosis - causes, symptoms, diagnosis and treatment." http://medicalformat.com/1194-focal-pulmonary-tuberculosis-causes-symptoms-diagnosis-and-treatment.html.

[18] Respiratory, "infiltrative tuberculosis - causes, symptoms, diagnosis and treatment." http://respiratory.vsebolezni.com/tuberkulez/infiltrativnyj-tuberkulez-legkix-chto-eto-takoe.html.

[19] S. D. Pitlik, V. Fainstein, and G. P. Bodey, "Tuberculosis mimicking cancer 2014;a reminder," *The American Journal of Medicine*, vol. 76, pp. 822–825, May 1984.

[20] S. Vento and M. Lanzafame, "Tuberculosis and cancer: a complex and dangerous liaison," *The Lancet Oncology*, vol. 12, pp. 520–522, jun 2011.

[21] S. Jeffrey, H. Shen, K. Vadim, and J. Chinnawsamy, "Pathogenesis of post primary tuberculosis: immunity and hypersensitivity in the development of cavities," *Annals of clinical and laboratory science*, vol. 44, pp. 365–87, jun 2015.

[22] M. Larobina and L. Murino, "Medical image file formats," *Journal of Digital Imaging*, vol. 27, pp. 200–206, Dec 2013.

[23] M. Jenkinson, "Nifti-1 data format," Oct 2007. https://nifti.nimh.nih.gov/nifti-1.

[24] R. D. Vincent, L. Baghdadi, and V. S. FONOV, "Minc/softwaredevelopment/minc2.0 file format reference," Oct 2012. https://en.wikibooks.org/wiki/MINC/SoftwareDevelopment/MINC2.0_File_Format_Reference [Accessed; posted 24 January 2019].

[25] O. S. Pianykh, *Digital Imaging and Communications in Medicine: A Practical Introduction and Survival Guide*. Springer Publishing Company, Incorporated, 1 ed., 2008.

[26] J. R. Koza, F. H. Bennett, D. Andre, and M. A. Keane, *Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming*, pp. 151–170. Dordrecht: Springer Netherlands, 1996. https://doi.org/10.1007/978-94-009-0279-4_9.

[27] M. Campbell, A. J. Hoane, Jr., and F.-h. Hsu, "Deep blue," *Artif. Intell.*, vol. 134, pp. 57–83, Jan. 2002.

[28] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, T. Lillicrap, K. Simonyan, and D. Hassabis, "Mastering chess and shogi by self-play with a general reinforcement learning algorithm," 2017.

[29] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach (3rd Edition)*. Pearson, 2009.

[30] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning (Adaptive Computation and Machine Learning series)*. The MIT Press, 2012.

[31] D. H. Wolpert, "The lack of a priori distinctions between learning algorithms," *Neural Computation*, vol. 8, pp. 1341–1390, Oct 1996.

[32] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, Sept. 1995.

[33] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Berlin, Heidelberg: Springer-Verlag, 1995.

[34] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," *Trans. Neur. Netw.*, vol. 10, pp. 1055–1064, Sept. 1999.

[35] H. Drucker, D. Wu, and V. N. Vapnik, "Support vector machines for spam categorization," *Trans. Neur. Netw.*, vol. 10, pp. 1048–1054, Sept. 1999.

[36] S.-K. Woo, C.-B. Park, and S.-W. Lee, "Protein secondary structure prediction using sequence profile and conserved domain profile," in *Advances in Intelligent Computing* (D.-S. Huang, X.-P. Zhang, and G.-B. Huang, eds.), (Berlin, Heidelberg), pp. 1–10, Springer Berlin Heidelberg, 2005.

[37] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods*. New York, NY, USA: Cambridge University Press, 2000.

[38] R. Kumar, A. Kulkarni, V. K. Jayaraman, and B. D. Kulkarni, "Symbolization assisted svm classifier for noisy data," *Pattern Recogn. Lett.*, vol. 25, pp. 495–504, Mar. 2004.

[39] S. Mukkamala, A. H. Sung, and A. Abraham, "Intrusion detection using an ensemble of intelligent paradigms," *J. Netw. Comput. Appl.*, vol. 28, pp. 167–182, Apr. 2005.

[40] R. K. Prasoona, A. Jyoti, Y. Mukesh, S. Nishant, N. S. Anuraj, and J. Shobha, "Optimization of gaussian kernel function in support vector machine aided qsar studies of c-aryl glucoside sglt2 inhibitors," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 5, pp. 45–52, Mar 2013.

[41] O. Z. Maimon and L. Rokach, *Data Mining with Decision Trees: Theroy and Applications (Machine Perception and Artificial Intelligence)*. World Scientific Pub Co Inc, 2007.

[42] C. Manapragada, G. I. Webb, and M. Salehi, "Extremely fast decision tree," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery ; Data Mining*, KDD '18, (New York, NY, USA), pp. 1953–1962, ACM, 2018.

[43] V. Podgorelec, P. Kokol, B. Stiglic, and I. Rozman, "Decision trees: An overview and their use in medicine," *Journal of Medical Systems*, vol. 26, no. 5, pp. 445–463, 2002.

[44] E. Cantú-Paz and C. Kamath, "Using evolutionary algorithms to induce oblique decision trees," in *Proceedings of the 2Nd Annual Conference on Genetic and Evolutionary Computation*, GECCO'00, (San Francisco, CA, USA), pp. 1053–1060, Morgan Kaufmann Publishers Inc., 2000.

[45] Q. Ding, Q. Ding, and W. Perrizo, "Decision tree classification of spatial data streams using peano count trees," in *Proceedings of the 2002 ACM Symposium on Applied Computing*, SAC '02, (New York, NY, USA), pp. 413–417, ACM, 2002.

[46] P. W. H. Wong, T. W. Lam, Y. C. Mui, S. M. Yiu, H. F. Kung, M. Lin, and Y. T. Cheung, "Filtering of ineffective sirnas and improved sirna design tool," in *Proceedings of the Second Conference on Asia-Pacific Bioinformatics - Volume 29*, APBC '04, (Darlinghurst, Australia, Australia), pp. 247–255, Australian Computer Society, Inc., 2004.

[47] T. G. Dietterich, "Ensemble methods in machine learning," in *Proceedings of the First International Workshop on Multiple Classifier Systems*, MCS '00, (London, UK, UK), pp. 1–15, Springer-Verlag, 2000.

[48] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*. Chapman & Hall/CRC, 1st ed., 2012.

[49] M. Kearns, "Thoughts on hypothesis boosting." (Unpublished), Dec. 1988.

[50] T. K. Ho, "Random decision forest," pp. 14–16, Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, Aug. 1995.

[51] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[52] J. Patterson and A. Gibson, *Deep Learning: A Practitioner's Approach*. O'Reilly Media, Inc., 1st ed., 2017.

[53] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *CoRR*, vol. abs/1409.4842, 2014.

[54] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.

[55] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. http://arxiv.org/abs/1409.1556.

[56] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, (USA), pp. 1097–1105, Curran Associates Inc., 2012.

[57] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from http://tensorflow.org/.

[58] E. Gibson, W. Li, C. Sudre, L. Fidon, D. Shakir, G. Wang, Z. Eaton-Rosen, R. Gray, T. Doel, Y. Hu, T. Whyntie, P. Nachev, M. Modat, D. C. Barratt, S. Ourselin, M. J. Cardoso, and T. Vercauteren, "Niftynet: a deep-learning platform for medical imaging," vol. 158, pp. 113–122, 2018.

[59] V. Liauchuk, A. Tarasau, E. Snezhko, V. Kovalev, A. Gabrielian, and A. . I. Rosenthal, "Lesion-based TB-descriptor for CT image analysis," *CLEF2018 Working Notes*, September 10-14 2018. CEUR-WS.org http://ceur-ws.org.

[60] D. Cid, M. Y., and H., "Texture-based graph model of the lungs for drug resistance detection, tuberculosis type classification, and severity scoring: Participation in ImageCLEF 2018 tuberculosis task," *CLEF2018 Working Notes*, September 10-14 2018. CEUR-WS.org http://ceur-ws.org.

[61] M. S. Ahmed, S. M. Obaidullah, M. Jayatilake, T. Goncalves, and L. Rato, "Texture analysis from 3D model and individual slice extraction for tuberculosis MDR detection, type classification and severity scoring," *CLEF2018 Working Notes*, September 10-14 2018. CEUR-WS.org http://ceur-ws.org.

[62] A. Hamadi and D. E. . I. Yagoub, "Semantic descriptors for tuberculosis CT image classification," *CLEF2018 Working Notes*, September 10-14 2018. CEUR-WS.org http://ceur-ws.org.

[63] J. Sun, P. Chong, Y. X. M. Tan, and A. Binder, "Imageclef 2017: Imageclef tuberculosis task - the sgeast submission," in *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum, Dublin, Ireland, September 11-14, 2017.*, 2017.

[64] "med2image: https://github.com/fnndsc/med2image. Last check: 30/05/2018.."

[65] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[66] "Resnet-50 and lungnet for tuberculosis severity scoring. mostaganem university at imageclefmed 2019 : tools to run experiments.." `https://github.com/anouar1991/imageCLEFfsei`.

[67] B. Ionescu, H. Müller, R. Péteri, Y. D. Cid, V. Liauchuk, V. Kovalev, D. Klimuk, A. Tarasau, A. B. Abacha, S. A. Hasan, V. Datla, J. Liu, D. Demner-Fushman, D.-T. Dang-Nguyen, L. Piras, M. Riegler, M.-T. Tran, M. Lux, C. Gurrin, O. Pelka, C. M. Friedrich, A. G. S. de Herrera, N. Garcia, E. Kavallieratou, C. R. del Blanco, C. C. Rodríguez, N. Vasillopoulos, K. Karampidis, J. Chamberlain, A. Clark, and A. Campello, "ImageCLEF 2019: Multimedia retrieval in medicine, lifelogging, security and nature," in *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, Proceedings of the 10th International Conference of the CLEF Association (CLEF 2019), (Lugano, Switzerland), LNCS Lecture Notes in Computer Science, Springer, September 9-12 2019.