Extracting and Classifying Salient Fields of View From Microscopy Slides of Tuberculosis Bacteria

Marios Zachariou[1,4], Ognjen Arandjelović[1], Evelin Dombay[2], Wilber Sabiiti[2], Bariki Mtafya[3], and Derek Sloan[2]

[1] School of Computer Science, St Andrews, Scotland, KY16 9SX
[2] School of Medicine, St Andrews, Scotland, KY16 9TF
[3] Mbeya Medical Research Center, Mbeya, Tanzania
[4] marios.zachariou@hotmail.com

Abstract. Tuberculosis is one of the most serious infectious diseases, and its treatment is highly dependent on early detection. Microscopy-based analysis of sputum images for bacilli identification is a common technique used for both diagnosis and treatment monitoring. However, it a challenging process since sputum analysis requires time and highly trained experts to avoid potentially fatal mistakes. Capturing fields of view (FOVs) from high resolution whole slide images is a laborious procedure, since they are manually localized and then examined to determine the presence of bacteria. In the present paper we propose a method that automates the process, thus greatly reducing the amount of human labour. In particular, we (i) describe an image processing based method for the extraction of a FOV representation which emphasises salient, bacterial content, while suppressing confounding visual information , and (ii) introduce a novel deep learning based architecture which learns from coarsely labelled FOV images and the corresponding binary masks, and then classifies novel FOV images as salient (bacteria containing) or not. Using a real-world data corpus, the proposed method is shown to outperform 12 state of the art methods in the literature, achieving (i) an approximately 10% lower overall error rate than the next best model and (ii) perfect sensitivity (7% higher than the next best model).

Keywords: Whole slide images · Fluorescence microscopy · Image processing · Artificial intelligence · Medicine · Infection · Respiratory system

1 Introduction

Tuberculosis (TB) is the biggest infectious disease-related cause of mortality globally [10]. *Mycobacterium tuberculosis* (Mtb) is the causative bacterium of TB, spread by droplet and aerosol, with up to 85 percent of cases affecting the lungs [38]. Other organs or tissues, such as the brain, kidneys, bone, and skin, can be infected by these pathogens. The present work focuses on microscopy images for identifying Mtb bacilli, with an emphasis on pulmonary tuberculosis. According to the WHO, up to 2 billion people worldwide have Mtb bacteria in their bodies, with up to 10 million instances of active illness and 2 million

deaths every year [38]. Since the 1940s, TB was treatable, but things began to deteriorate with the advent of Drug Resistant TB variations such as Multi Drug Resistant, eXtensive Drug Resistant, and Total Drug Resistant forms of the bacterium [12]. The largest burden of morbidity and death from tuberculosis occurs in poor and middle-income nations, where healthcare resources are limited [30]. Early TB testing improves patient chances for treatment and recovery while also assisting in the prevention of disease spread, and lowering the probability of drug resistant pathogen emergence [38, 7, 20].

Sputum smear microscopy has traditionally been the primary method for diagnosing tuberculosis. Sputum samples from symptomatic individuals are heat-fixed onto slides and stained using laboratory techniques that identify acid-fast bacteria (AFB) like Mtb cells. For light microscopy (typically at $\times 1000$ magnification), the older Ziehl-Neelsen treatment stains AFB red on a blue background, but newer Auramine-based protocols stain it yellow-green against a black background for fluorescence microscopy (usually at $\times 400$ magnification). Semi-quantitative grading methods have been created to measure the bacterial burden in a patient's lungs. The findings of sputum smear microscopy are often described as 'negative', 'scanty', '1+', '2+', or '3+' [34], in ascending order of disease severity.

1.1 Importance of microscopy

Many centres throughout the globe have switched their attention away from smear microscopy and towards new tools (such as the Xpert MTB/RIF test) for TB diagnosis in recent years [17]. Sputum smear gradings, on the other hand, remain effective for triaging disease severity and prognosis, with implications for therapeutic individualisation [34] and treatment response where new assays are not currently suggested [38].

Smear microscopy provides data considerably faster than waiting for Mtb to develop in culture in clinical microbiology practice [31]. When properly done, it has a high specificity (99%) for detecting Mtb cells [34]. Smear microscopy has become more sensitive since switching from classic Ziehl-Neelsen to fluorescent Auramine-based microscopy (from 0.34-0.94 to 0.52-0.97 according to one systematic review) [3, 32]. Although microscopy laboratory materials are typically affordable, the method is time-consuming, which has an impact on laboratory staffing costs. The large ranges of diagnostic sensitivity reported for TB smear microscopy also reflect the complexity and subjectivity of the process.

Disadvantages of microscopy and motivation for computer based automatic detection There are obstacles to using microscopy effectively for clinical patient treatment and scholarly research on Mtb. As a microscopist, maintaining a high level of skill necessitates a consistent commitment of time. To stay proficient, practitioners should study at least 25 slides every day according to general guidelines [22]. Each slide is divided into small regions that are examined sequentially, with human error (e.g. due to weariness) reducing the specificity and

sensitivity of analysis [25]. Some slides are also difficult to interpret because AFB can exhibit unusual appearance or because non-bacterial components (artefacts) inside the sputum matrix resemble Mtb cells. A promising direction of addressing these challenges lies in the removal of the human from the loop, that is by employing modern artificial intelligence approaches [36].

2 Related Work

To the best of our knowledge, there is no published work that is specific to the collecting of enlarged FOV images containing probable microorganisms. Most datasets used in tuberculosis research comprise images that have been manually magnified and cropped by a microbiologist specialist.

The work of Forero *et al.* [6] was one of the earliest attempts at the use of automatic methods for the analysis of fluorescence microscopy slides with Mtb bacteria. The authors' primary objective was to develop a diagnostic tool, and they used autofocus functions to crop FOVs from microscopic slides. Autofocus was accomplished by a two-pass algorithm that determines whether or not a specific area is void of bacterial content before bringing the image into focus [18]. The initial run of the algorithm analyses slides at three z-axis points to assess if there is sufficient variance between them to signal the presence of salient content in the field. As the authors note, Mtb bacteria occupy extremely small areas of the image, i.e. most of the image is taken up by the background, and their experiments demonstrate that a narrow scanning window (256×256 pixels) must be used for accurate FOV localization. Using auramine stained slides, out of the four focusing methods examined by the authors, two, namely the wave and auto-correlation based ones, produced promising results.

A more recent attempt to make use of autofocus functions is that of Zhai *et al.* [41]. The primary distinction between theirs and previous work lies in the use of conventional rather than fluorescent microscopy. Amongst others, a notable difference as compared with the approach introduced in the present paper lies in the scanning process. In particular, Zhai *et al.* employ a row-wise scanning strategy whereas herein we proceed in a spiral manner. The authors employed three different autofocus measurement, namely the sum of gray-level differences, the Laplacian, and the Tenengrad function with the Sobel operator, and found that the latter outperforms the former two. Nevertheless, the reported empirical accuracy of the method is much lower than that of other methods in the literature that perform diagnosis without collecting FOVs [24, 37, 19].

Kant and Srivastava's work processes entire slides in a bottom-up manner, that is by aggregating information extracted from small patches [12]. They used a five-layer patch-wise classifier to load each tile from a microscopic slide and a 20×20 pixel window which moves through the FOV to assess the presence of germs. Although the authors claim 99.8% accuracy, this number is rather misleading. The reason stems from the observation that the vast majority of the area of a microscopic slide is occupied by the background, resulting in high accuracy owing to accurate background classification (i.e. in effect, false negative

errors are deprioritized). Indeed, when sensitivity and specificity are considered, the reported rates of 83.8% and 67.6% respectively are comparable to or *worse* than those of other methods discussed in the present paper.

3 Proposed method

We begin our process of cropping the slide by putting each tile into memory (as the entire slide is too big $\approx$ 19GB) and cropping 200×400 pixel patches from each tile. Our slides include a total of 2700 tiles which are anisotropically scaled by a factor of 4.83 in the x direction and 3.24 in the y direction, so as to match FOVs created manually by a specialist.

3.1 Discrimination enhanced representation extraction

A human specialist detects Mtb microorganisms by inspecting the green channel of a FOV. The more acid fast a bacterium is, the more prominent is its appearance [5]. Additionally, Mtb bacteria that store non-polar lipids intracellularly are classified as lipid-rich (LR) cells, as opposed to lipid-poor (LP) cells. It is hypothesised that resistant LR bacteria play a critical role in patient relapse [21, 27]. As a result, their acid resistance begins to deteriorate, and they become less visible in the green channel. Fortunately, in this case they become more apparent in the red channel [23, 8] using Nile red staining as opposed to Auramine-O. While the evidence that the presence of LR Mtb bacteria indeed does predict poor treatment outcome is still insufficiently strong, there are numerous studies on non-polar lipids in Mtb bacteria and Nile red staining which point in this direction [2, 14, 13, 4]. Thus, both the red (Nile red staining) and the green (Auramine-O) channel remain of relative importance in both research and clinical terms.

Additionally, the ability to generate high-quality microscopic images is contingent upon the quality of clinical samples collected and the details of smear preparation and staining processes. Thick smears from highly mucous samples can have an excessive amount of background staining, which makes bacteria harder to localize. To increase the robustness of our method, as well as to reduce the complexity of the learning task, we propose a pre-processing stage that enhances image content of interest while at the same time suppressing confounding information e.g. in the form of staining artefacts [39].

Considering that the bacteria of interest form largely straight, thin, and elongated structures, we employ a ridge detector [1]. In particular, we make use of the Hessian matrix approximation (at the scale of 2×2 pixels in the present work). Its eigendecomposition allows for a differentiation between different kinds of local image behaviour leading to a straightforward process of distinguishing between blob-like structures, uniform regions, and elongated structures of interest herein [15]. Considering that bacilli form elongated structures, we are interested in the loci which exhibit significant change in one principal direction (perpendicular to a bacterium) and little change in the other (along a bacterium), and

these can be readily identified using the corresponding Hessian matrix eigenvalues. In particular, to create an enhanced image (in the context of our end goal), each pixel in the original image is replaced with the absolute value of the lower magnitude value of the Hessian eigenvalue computed at the locus. The at first sight appealing alternatives which take into account both eigenvalues, such as the use of the ratio of the two eigenvalues, were found unsuitable due to an increase in noise and the dynamic range.

3.2 The learning

Our end goal is to classify cropped FOVs as positive or negative for Mtb microorganisms. By doing so, that is by filtering out uninteresting FOVs, the burden of the manual workload required from a lab worker is dramatically reduced. In addition, the aforementioned classification can also be used to simplify and improve further automatic processing, e.g. diagnostic inference or bacterial culture analysis.

Proposed model Evidence from previous work on non-automatic smear analysis, that is analysis performed by humans, suggests that for the detection of Mtb bacteria the use of both texture and shape information is superior to the use of either of the two in isolation [37, 6, 29]. Herein we introduce a network that reflects this finding by employing two encoders to generate two separate feature maps. One of these is trained on the discrimination enhanced representation of FOVs introduced in Section 3.1, while the other is trained on the binary masks corresponding to the FOVs, which distinguish between the objects of interest (bacteria) and the uninteresting content (background and artefacts). The encoder outputs are concatenated to generate the input matrix for another smaller network ($16 \times 32 \times 512$); see Figure 1. The weights of the two encoders are frozen, and no gradient computation is done during the training of the smaller network. As a result, the smaller network makes an effort to infer the probability distribution from the two encoders which independently infer texture and shape. To train the two encoders, a further layer with adaptive max pooling and a linear layer leading to a single output unit with a sigmoid activation function were added. The same environmental and hyper-parameters as previously employed were utilized to train the two encoders and the smaller network.

4 Evaluation

4.1 Data

The data used in the present work consists of microscopic slides captured from a clinical cohort study in Mbeya, Tanzania. Between February 2017 and March 2018, 46 persons with sputum smear positive pulmonary tuberculosis (40 newly

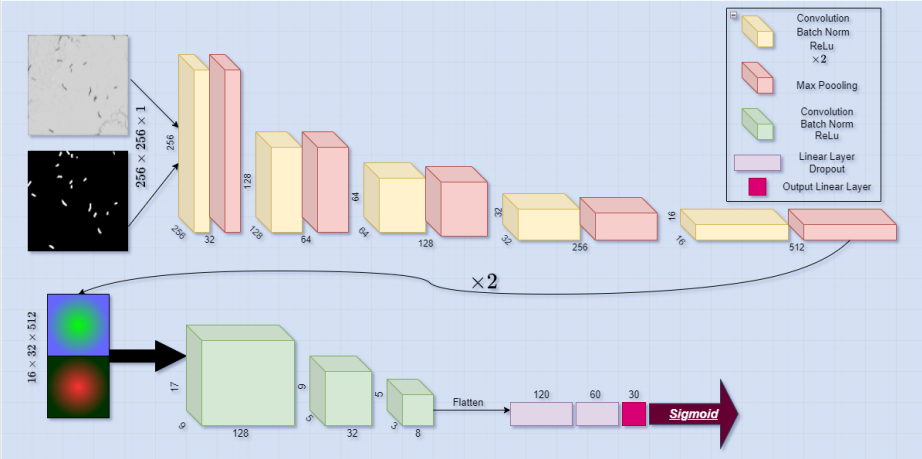

Fig. 1. Diagram of information flow through the architecture proposed in the present paper: following an encoding process, information passes through a convolutional network, leading to the eventual inference of the bacterial presence in a FOV.

diagnosed and 6 previously diagnosed) were recruited and tracked until the conclusion of a 6-month course of conventional tuberculosis therapy. Sputum samples were taken pre-treatment and at the end of months 2, 4, and 6 of therapy. Smears on microscope slides were made from the sputum samples. The slides were dyed using normal Auramine-O procedures, and the smears were systematically scanned through a fluorescein isothiocyanate filter using a Leica DMLB epifluorescence microscope at $\times 1000$ magnification by an experienced microscopist. A digital camera was used to capture and save all fields having auramine-stained, yellow-green AFB. A total of 230 slides were inspected, and for each AFB positive slide 30 images were created.

The training dataset for this experiment included 46 patients with around 150 FOV images per patient from diagnosis through therapy completion. Around 800 FOVs were randomly chosen from the Tanzanian corpus . To verify that the automated image analysis method being developed is not affected by changes in the morphology of Mtb cells during or after TB therapy, images were picked across all time periods of sample collection. These images were re-examined by a microscopist who was not involved in the original experiment. Each FOV was then assessed as positive or negative by a microscopy specialist.

Two additional slides from a separate facility and technician were used to evaluate our method. The test set consists of 130 FOVs extracted from the slides using the approach proposed in the present paper, chosen using balanced random sampling that ensures that the set is balanced in terms of positive and negative examples.

4.2 Competitor models

We compare the proposed method with a number of state of the art models from the VGG family [26], the ResNet family [9] (including Wide-Resnet [40]), the Densenet family [11], and InceptionV3 [33]. All models were pre-trained using the ImageNet dataset, which contains 1000 target classes and three input channels. Regardless of their initial configuration, each model's first convolutional layer was replaced with one that comprises a single input channel, kernel of size 3×3, stride value of 1, and padding of size 3×3. The alterations are motivated by the fact that our slide representation is monochrome (i.e. single channel) and the objects of interest are thin, elongated structures that frequently appear near the image boundary. The last modification is to the final linear layer, which is replaced with one that retains the same input features but has just one output node (in InceptionV3, this change is applied to its auxiliary classifier). The output weights of the last linear layer are passed through the sigmoid function.

4.3 Hyper-parameter learning

To ensure a fair comparison, all models, including the proposed one, were trained using the same set of hyper-parameters. To begin, the batch size was chosen to be 16 in order to achieve a balance between generalization, accuracy, and computing speed. Adam optimizer with the values of the β parameters (that is, the initial decay rates used when estimating the first and second moments of the gradient) equal to 0.50 and 0.99 was used. For the training process, following evidence from prior research [28], the base and maximum learning rates were set to 0.00001 and 0.0004, respectively, and the learning scheduler used was the novel circular scheduler with a step size equal to five times the size of the dataset (which varies according to batch size). Finally, binary cross entropy was used as the loss function; see Figure 2.

4.4 Results and discussion

A microbiological specialist classified 73 of the 130 FOV images in our test set as positive and 57 as negative. The same expert manually created binary masks corresponding to all of the 130 FOVs, which were treated as the ground truth. In order to facilitate a comprehensive and nuanced comparison between models [35], we assess performance using a number of metrics, namely overall accuracy, recall (sensitivity) and precision (specificity), receiver operating characteristics (ROC), and the area under the ROC curve (AUC).

A summary of our experimental results is shown in Table 1. To start with, consider the overall performance metric in the form of the classification error (in the rightmost column of the table) and observe that the proposed method achieved the best performance of all 13 methods compared. The error rate of the next best model, namely ResNet50, is more than 11% greater. InceptionV3 and the best DenseNet family model, DenseNet201, performed next best (23% higher error rate than the proposed model). While there is significant variation

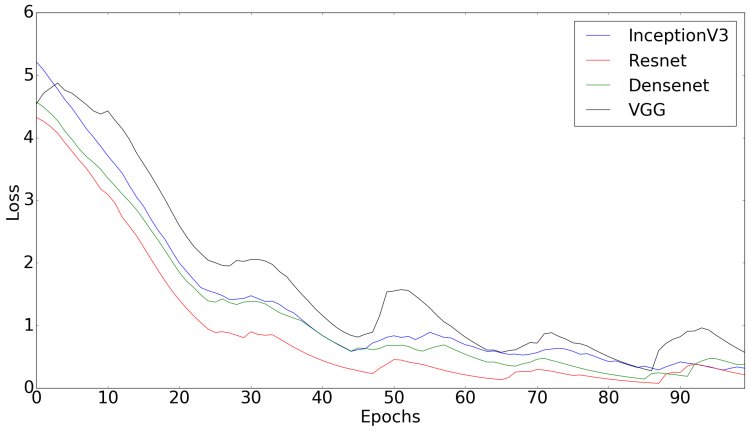

Fig. 2. With the exception of VGG family, all models converge around the 65th epoch and exhibit the same overall learning behaviour. The poor performance of the VGG family suggests that the present task requires greater architectural sophistication than that achieved by merely stacking convolutional layers.

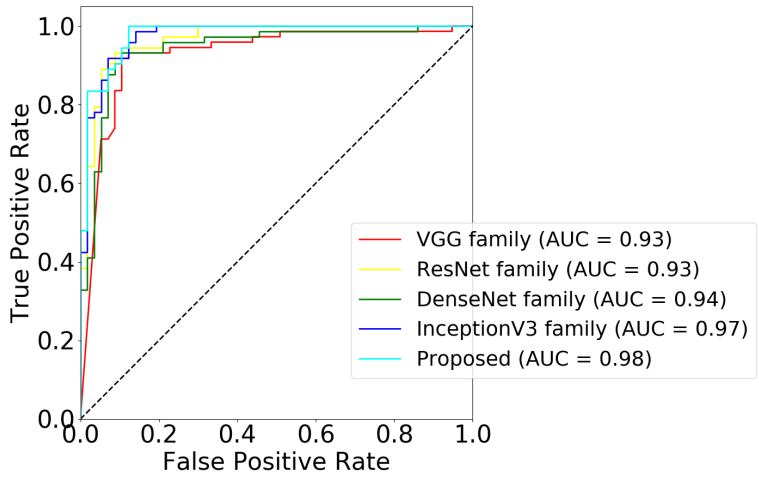

Fig. 3. Comparison of ROC curves and the areas under the curves, both show that the proposed solution outperforms the current state of the art. For the sake of clarity, each model family is illustrated using the average performance of its evaluated models.

between different specific models within all families, generally speaking ResNet performed better than DenseNet, and VGG networks fared the worst (45–101% higher error rate than the proposed model).

Table 1. Summary of results. All VGG models were trained using batch normalization (BN). The best performing model with respect to each statistic is shown in bold.

Model name	**True +ve**	**False +ve**	**True -ve**	**False -ve**	**Error rate**
VGG11 (w/ BN)	68	5	49	8	0.100
VGG13 (w/ BN)	67	6	47	10	0.123
VGG16 (w/ BN)	67	6	46	11	0.131
VGG19 (w/ BN)	63	10	50	7	0.131
ResNet18	65	8	51	6	0.108
ResNet34	66	7	51	6	0.100
ResNet50	68	5	**52**	**5**	0.077
ResNet50-Wide	68	5	49	8	0.100
DenseNet121	66	7	51	6	0.100
DenseNet169	67	6	45	12	0.139
DenseNet201	68	5	51	6	0.085
InceptionV3	67	6	**52**	**5**	0.085
Proposed	**73**	**0**	48	9	**0.069**

A more nuanced insight into the behaviour of different models can be gained by examining the specific error types (in columns 3 and 5 of Table 1; also see Figure 3). Importantly, note that the proposed method performed best in terms of the false positive error rate – indeed, it made no incorrect positive calls at all. This is extremely important in this context for reasons already noted in Section 2, to wit the vast majority of FOVs do not contain bacteria and it is of paramount importance that these are filtered out as a means of reducing expert human labour thereafter. No other method comes close to ours, with DenseNet201, ResNet50, ResNet50-Wide, and VGG11 erring in approximately 7% of the cases. On the other hand, the proposed method was not superior in terms of the false positive rate. In the context of this metric InceptionV3 and ResNet50 performed best, achieving the error rate of approximately 9%. However, here it is important to observe the asymmetry of the importance of type I and type II errors on the task at hand [16]. As we noted earlier, the former are of primary importance as low type I error rate means that the vast amount of irrelevant information is not passed on further for human analysis which is where the practical bottleneck lies. On the other hand, type II errors, while of course undesirable (as any error is), are far less important, as clinically salient information about a bacterial culture can be readily derived from only a sample of bacteria, without there being a need for the entirety of the culture to be examined. Of course, this is predicated on the sample being representative which is why the type II error rate must not be excessively high. As witnessed by our results in Table 1 this is not a major challenge here as none of the methods compared produced a high number of false negatives.

5 Conclusion

Although sputum smear microscopy is being phased out in many settings in favour of Xpert MTB/RIF and other molecular tests, it still serves a function in some aspects of disease severity assessment and therapy monitoring. Since microscopic examination of Mtb cells is critical as a research tool, work on the development of improved automated tools for standardising and expediting image analysis remains important. To this end, in this paper we described a novel solution based on a newly crafted deep learning based architecture tailored specifically for the task, which learns from coarsely labelled FOV images and the corresponding binary masks, and then classifies novel FOV images as bacteria containing or not. The fully automated nature of the model and its empirically evidenced vanishing false positive rate, demonstrate the potential of the proposed method to significantly reduce human expert labour.

Acknowledgements We will like to express our appreciation to the McKenzie Institute for providing the necessary funding to complete this work.

References

1. Arandjelović, O., Cipolla, R.: A new look at filtering techniques for illumination invariance in automatic face recognition. pp. 449–454 (2006)
2. Baron, V.O., Chen, M., Clark, S.O., Williams, A., Hammond, R.J.H., Dholakia, K., Gillespie, S.H.: Label-free optical vibrational spectroscopy to detect the metabolic state of M. tuberculosis cells at the site of disease. Scientific Reports **7**(1), 1–9 (2017)
3. Costa Filho, C.F.F., Costa, M.G.F., Júnior, A.K.: Autofocus functions for tuberculosis diagnosis with conventional sputum smear microscopy. Current Microscopy Contributions to Advances in Science and Technology pp. 13–20 (2012)
4. Daniel, J., Kapoor, N., Sirakova, T., Sinha, R., Kolattukudy, P.: The perilipin-like PPE15 protein in Mycobacterium tuberculosis is required for triacylglycerol accumulation under dormancy-inducing conditions. Molecular Microbiology **101**(5), 784–794 (2016)
5. Deb, C., Lee, C.M., Dubey, V.S., Daniel, J., Abomoelak, B., Sirakova, T.D., Pawar, S., Rogers, L., Kolattukudy, P.E.: A novel in vitro multiple-stress dormancy model for mycobacterium tuberculosis generates a lipid-loaded, drug-tolerant, dormant pathogen. PLoS ONE **4**(6), e6077 (2009)
6. Forero, M.G., Sroubek, F., Cristóbal, G.: Identification of tuberculosis bacteria based on shape and color. Real-Time Imaging **10**(4), 251–262 (2004)
7. Gele, A.A., Bjune, G., Abebe, F.: Pastoralism and delay in diagnosis of TB in Ethiopia. BMC Public Health **9**(1), 1–7 (2009)
8. Greenspan, P., Fowler, S.D.: Spectrofluorometric studies of the lipid probe, nile red. Journal of Lipid Research **26**(7), 781–789 (1985)
9. He, D., Xia, Y., Qin, T., Wang, L., Yu, N., Liu, T.Y., Ma, W.Y.: Dual learning for machine translation. Advances in Neural Information Processing Systems **29**, 820–828 (2016)

10. Holmes, C.B., Hausler, H., Nunn, P.: A review of sex differences in the epidemiology of tuberculosis. International Journal of Tuberculosis and Lung Disease **2**(2), 96–104 (1998)
11. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely Connected Convolutional Networks. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 4700–4708 (2017)
12. Kant, S., Srivastava, M.M.: Towards Automated Tuberculosis detection using Deep Learning. In: IEEE Symposium Series on Computational Intelligence. pp. 1250–1253 (2019)
13. Kayigire, X.A., Friedrich, S.O., van der Merwe, L., Donald, P.R., Diacon, A.H.: Simultaneous staining of sputum smears for acid-fast and lipid-containing Myobacterium tuberculosis can enhance the clinical evaluation of antituberculosis treatments. Tuberculosis **95**(6), 770–779 (2015)
14. Kennedy, J.A., Baron, V., Hammond, R.J.H., Sloan, D.J., Gillespie, S.H.: Centrifugation and decontamination procedures selectively impair recovery of important populations in Mycobacterium smegmatis. Tuberculosis **112**, 79–82 (2018)
15. Kumar, N.C.S., Radhika, Y.: Optimized maximum principal curvatures based segmentation of blood vessels from retinal images. Biomedical Research **30**(2) (2019)
16. Lomacenkova, A., Arandjelović, O.: Whole slide pathology image patch based deep classification: an investigation of the effects of the latent autoencoder representation and the loss function form. In Proc. IEEE International Conference on Biomedical and Health Informatics (2021), DOI: 10.1109/BHI50953.2021.9508577
17. Mehta, P.K., Raj, A., Singh, N., Khuller, G.K.: Diagnosis of extrapulmonary tuberculosis by PCR. FEMS Immunology & Medical Microbiology **66**(1), 20–36 (2012)
18. Merchant, F.A., Castleman, K.R.: Computer-assisted microscopy. In: The Essential Guide to Image Processing, pp. 777–831 (2009)
19. Panicker, R.O., Kalmady, K.S., Rajan, J., Sabu, M.K.: Automatic detection of tuberculosis bacilli from microscopic sputum smear images using deep learning methods. Biocybernetics and Biomedical Engineering **38**(3), 691–699 (2018)
20. Peter, J.G., van Zyl-Smit, R.N., Denkinger, C.M., Pai, M.: Diagnosis of TB: state of the art. European Respiratory Monograph **58**, 123–143 (2012)
21. Phillips, P.P.J., Mendel, C.M., Burger, D.A., Crook, A., Nunn, A.J., Dawson, R., Diacon, A.H., Gillespie, S.H.: Limited role of culture conversion for decision-making in individual patient care and for advancing novel regimens to confirmatory clinical trials. BMC Medicine **14**(1), 1–11 (2016)
22. Rieder, H.L., Van Deun, A., Man Kam, K., Jae Kim, S., Chonde, T.M., Trebucq, A., Urbanczik, R.: Priorities for Tuberculosis Bacteriology Services in Low-Income Countries. International Union Against Tuberculosis and Lung Disease (2007)
23. Rumin, J., Bonnefond, H., Saint-Jean, B., Rouxel, C., Sciandra, A., Bernard, O., Cadoret, J.P., Bougaran, G.: The use of fluorescent Nile red and BODIPY for lipid measurement in microalgae. Biotechnology for Biofuels **8**(1), 1–16 (2015)
24. Sadaphal, P., Rao, J., Comstock, G.W., Beg, M.F.: Image processing techniques for identifying Mycobacterium tuberculosis in Ziehl-Neelsen stains. The International Journal of Tuberculosis and Lung Disease **12**(5), 579–582 (2008)
25. Shea, Y.R., Davis, J.L., Huang, L., Kovacs, J.A., Masur, H., Mulindwa, F., Opus, S., Chow, Y., Murray, P.R.: High sensitivity and specificity of acid-fast microscopy for diagnosis of pulmonary tuberculosis in an African population with a high prevalence of human immunodeficiency virus. Journal of Clinical Microbiology **47**(5), 1553–1555 (2009)
26. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

27. Sloan, D.J., Mwandumba, H.C., Garton, N.J., Khoo, S.H., Butterworth, A.E., Allain, T.J., Heyderman, R.S., Corbett, E.L., Barer, M.R., Davies, G.R.: Pharmacodynamic modeling of bacillary elimination rates and detection of bacterial lipid bodies in sputum to predict and understand outcomes in treatment of pulmonary tuberculosis. Clinical Infectious Diseases **61**(1), 1–8 (2015)
28. Smith, L.N.: Cyclical learning rates for training neural networks. In: IEEE Winter Conference on Applications of Computer Vision. pp. 464–472 (2017)
29. Sotaquira, M., Rueda, L., Narvaez, R.: Detection and quantification of bacilli and clusters present in sputum smear samples: a novel algorithm for pulmonary tuberculosis diagnosis. In: International Conference on Digital Image Processing, pp. 117–121 (2009)
30. Spence, D.P., Hotchkiss, J., Williams, C.S., Davies, P.D.: Tuberculosis and poverty. British Medical Journal **307**(6907), 759–761 (1993)
31. Steingart, K.R., Henry, M., Laal, S., Hopewell, P.C., Ramsay, A., Menzies, D., Cunningham, J., Weldingh, K., Pai, M.: A systematic review of commercial serological antibody detection tests for the diagnosis of extrapulmonary tuberculosis. Postgraduate Medical Journal **83**(985), 705–712 (2007)
32. Steingart, K.R., Henry, M., Ng, V., Hopewell, P.C., Ramsay, A., Cunningham, J., Urbanczik, R., Perkins, M., Aziz, M.A., Pai, M.: Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. Lancet Infectious Diseases **9**(6), 570–581 (2006)
33. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 1–9 (2015)
34. Toman, K.: Toman's tuberculosis: case detection, treatment and monitoring. Questions and answers. World Health Organization (2004)
35. Valsson, S., Arandjelović, O.: Nuances of Interpreting X-ray Analysis by Deep Learning and Lessons for Reporting Experimental Findings. Sci **4**(1), 1–13 (2022)
36. Vente, D., Arandjelović, O., Baron, V., Dombay, E., Gillespie, S.: Using machine learning for automatic counting of lipid-rich tuberculosis cells in fluorescence microscopy images. In Proc. AAAI Conference on Artificial Intelligence Workshop on Health Intelligence pp. 57–68 (2019)
37. Veropoulos, K., Learmonth, G., Campbell, C., Knight, B., Simpson, J.: Automated identification of tubercle bacilli in sputum: A preliminary investigation. Analytical and Quantitative Cytology and Histology **21**(4), 277–282 (1999)
38. World Health Organisation: Global Tuberculosis Report. Tech. rep. (2018), https://apps.who.int/iris/bitstream/handle/10665/274453/9789241565646-eng.pdf
39. Zachariou, M., Arandjelović, O., Sloan, S., Sabiiti, W., Mtafya, B.: Tuberculosis bacteria detection and counting in fluorescence microscopy images using a multi-stage deep learning pipeline. Information **13**(2), 96 (2022)
40. Zagoruyko, S., Komodakis, N.: Wide residual networks. arXiv preprint arXiv:1605.07146 (2016)
41. Zhai, Y., Liu, Y., Zhou, D., Liu, S.: Automatic identification of mycobacterium tuberculosis from ZN-stained sputum smear: Algorithm and system design. In: IEEE International Conference on Robotics and Biomimetics. pp. 41–46 (2010)