Artificial intelligence based automatic acid-fast bacilli identification: Colentina University Hospital and Zaya Artificial Intelligence experience

Sabina Zurac[1,2], M Trascau[2*], C Mogodici[2*], Cristiana Popp[1,2*], Luciana Nichita[1,2], Mirela Cioplea[1,2], T Poncu[2], B Ceachi[3], Liana Sticlaru[1,2], Alexandra Cioroianu[1,2], M Busca[1,2], Oana Stefan[1], Irina Tudor[1], A Voicu[2], Daliana Stanescu[2], Alexandra Bastian[1]*

**Authors with equal contribution*

1 Colentina University Hospital
2 Zaya Artificial Intelligence SRL

Tuberculosis is an ancient infection (with hominid infections up to 3.000.000 years ago) afflicting approximately 10 million people worldwide; it represents one of the first 10 causes of death and first cause of death by an infectious disease. TB diagnosis requires several investigations, the most reliable one being Mycobacteria identification; on tissue, Mycobacterium tuberculosis is identified in Ziehl Nielsen (ZN) stained-slides as a minute red colored rod of 2-4/0.2-0.5 microns on a blue background; its searches in 2-6 cm^2 of tissue with 40X microscopic objective (0.5mm diameter) takes hours, thus automatic artificial intelligence-based identification is a viable solution in this field.

Our team includes pathologists from Colentina University Hospital, machine learning engineers from University of Polytechnics Bucharest and experts in IT and economics from Zaya Artificial Intelligence SRL. We prepared a training data set of 1.500.000 bounding boxes (500.000 positive and 1.000.000 negative) annotated on scans of 81 ZN-stained slides (whole slide images = WSI) originating in 81 paraffin blocks (21 cases positive, 60 cases negative). Several image augmentation techniques coupled with different custom computer vision architectures were used. The results were tested internally and in pathology laboratory settings on a testing sample of 46 cases (23 positive and 23 negative ZN-stained slides).

WSIs automatic analysis was followed by a report indicating areas more likely to present Mycobacteria. We compared the results obtained by 4 teams of pathologists (with different level of experience) by separately evaluating a) the original slides; b) the scanned images of the slides and c) the results given by automatic analysis of WSIs. Our architecture showed very good results (0.9020 on ROC curve). The results of final test (algorithm versus pathologist evaluating WSIs, algorithm versus pathologist evaluating slides and pathologist evaluating WSIs versus slides) presented accuracy over 90% in all cases.

AI-based software for automated analysis of ZN-stained slides shows reliable results, being a helpful tool to the pathologist.