A Deep Learning Approach to Investigate Tuberculosis Pathogenesis in Nonhuman Primate Model: Combining Automated Radiological Analysis with Clinical and Biomarkers Data

By

FAISAL YASEEN
THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

HEALTH INFORMATICS

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Dr. Imran H Khan

Dr. Nicholas Anderson

Mark J Carrol

Committee in Charge

2023

Contents

List of Figures

List of Tables

Acknowledgements

I could not have done this without the unwavering support from my family, the UCD MHI faculty and staff, and my colleagues at UC Davis Pathology department.

First and foremost, I would like to express my sincere gratitude to my research supervisor, Imran H. Khan, Ph.D., M.B.A, Professor, and director of proteomics core lab for his professional mentorship and support. He gave me the opportunity to work on this project and learn under his supervision and created opportunities for me to learn the wet lab work. For this, I will be forever grateful.

I would like to convey my sincere appreciation for my thesis chair, Mark Carroll, Lecturer - UC Davis School of Medicine Division of Public Health Sciences, for constant support and insightful guidance throughout the graduate program. I would also like to express my deepest gratitude to Nick Anderson, Ph.D., Cardiff Professor of Informatics and Yong Choi, PhD, MPH, Assistant Professor, University of Pittsburgh for their consistent support and invaluable guidance during this program and for my next journey as well. It has been an amazing experience to be part of the MHI program.

Finally, I would like to dedicate this dissertation to my lovely wife, Dr. Heba Aly, thank you for your patience, constant support, and sacrifices you made throughout this journey, and above all your love. This would not have been possible without you. I am so blessed to have you in my life.

I also dedicate this my parents, Mina and Yaseen, and my parents-in-law, Zakia and Aly, thank you so much for all your prayers and love.

Abstract

Tuberculosis (TB) kills approximately 1.6 million people yearly. Among the three top infectious killers, TB is rising worldwide while HIV-AIDS and malaria are trending down, despite the fact anti-TB drugs are generally curative. The problem lies in the inefficient detection of this complex disease. It is hypothesized utilizing informatics methods such as deep learning (DL) approaches to machine learning (ML) analysis of radiological data, combined with clinical, microbiological, and immunological data, delivered as clinical decision support (CDS), can not only afford better diagnostics but also improved determination of pathogenesis and severity, and monitoring efficacy of therapy. This study proposes a comprehensive approach for efficient disease detection employing informatics methods including but not limited to information retrieval, visualization, ML and artificial intelligence (AI), tested in nonhuman primate (NHP) model which captures human TB most closely, as a prelude to TB patient studies. A group of six rhesus macaques were experimentally inoculated with *Mycobacterium tuberculosis (M. tb*., Erdman strain) in the right lower lung. Animals were followed at regular intervals over 24 weeks by Computed Tomography (CT) imaging. A DL algorithm was developed and trained for automated scoring of lung CT images and compared head-to-head with radiologists' scores. Correlations of ML scores with several other TB indicators were also performed. DL model afforded early disease detection as compared to radiologists. Importantly, ML analysis demonstrated greater consistency over multiple runs compared to scoring by two radiologists. ML scores also exhibited strong correlations with granuloma and total TB-lesion volumes at necropsy, and disease-signs and blood biomarkers throughout pathogenesis. ML-based analysis of radiological imaging enabled early and consistent disease detection and assessment of severity, enabling a noninvasive and objective approach. ML and AI approaches can improve early detection and understanding of the disease. In addition, the multimodality approach described here is valuable in monitoring efficacy of therapy.

Background and Introduction

Tuberculosis (TB) is one of the world's deadliest infectious diseases [1]. Two billion people are latently infected with the etiological agent of TB, *Mycobacterium tuberculosis* (*M. tb*.), and 10 % of these develop active TB in their lifetime, leading to approximately 10 million new TB patients, and 1.6 million deaths annually [1]. Most of the TB endemic countries are developing nations where attempts to eradicate TB are challenged by inefficient diagnostics, limited understanding of disease pathogenesis, poor healthcare facilities, and lack of medical infrastructure. TB is generally curable if diagnosed properly and timely. However, frontline diagnostic tests (e.g., Acid-Fast Bacilli (AFB) microscopy) lack sensitivity while newer, state-of-the-art molecular diagnostics are not readily available universally [2-5]. Also, there is dearth of expert radiologists for sound assessment of radiological images to properly diagnose TB at the earliest possible stage. Poor TB detection rates and lack of health infrastructure facilities as a result cause improper care of infected infections, additional infection transmission, and perhaps even the emergence of medication resistance. In high burden TB countries such as Pakistan, India, Nigeria etc. healthcare systems will substantially benefit from rapid, accurate and cost-effective TB detection. Because pulmonary TB is highly contagious, it is of utmost urgency to devise new strategies for early and efficient diagnosis.

Automated solutions for TB diagnosis can play a major role in alleviating these problems, reducing the workload of hospitals, especially in developing countries. One approach to automate this is to devise and utilize informatics algorithms that automatically perform image analysis on digital medical images such as X-Rays, CT scans or microscopy images of sputum samples, etc.

We propose a comprehensive approach to TB diagnosis employing Deep Neural Networks (DNNs) architectures to combine multimodal data such as clinical, immunological, and radiological data. The

most effective way is to combine various data sets, including radiological image analysis by machine learning (ML), followed by integration with multiple diagnostic modality data through artificial intelligence (AI) (Figure 1).

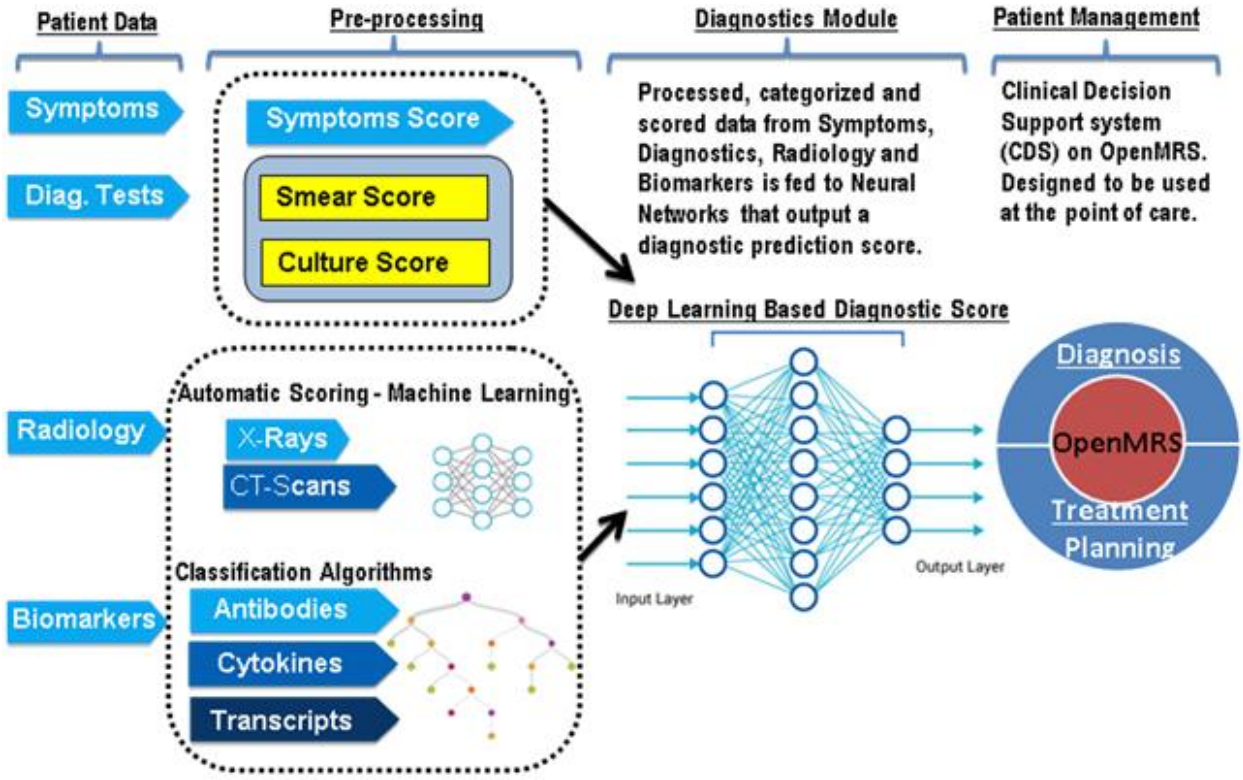

Figure 1: Proposed Clinical Decisions Support System via Open Medical Record System (Open-MRS).

Proposed computational modeling of TB data and Clinical Decision Support System (CDSS) to be delivered to TB physician through Open Medical Record System (Open-MRS) to improve detection and understanding of the disease for diagnosis and monitoring efficacy of therapy.

In the past, research on image analyses focused mainly on classical image processing - using image processing for extracting certain manually defined ('hand-crafted') features and then using those features for classification using machine learning techniques such as random forests [6], support vector machines or (shallow) neural networks, etc.

In recent years, due to new research and availability of much more computational power, researchers have been able to utilize deep neural network architectures [6-7], which surpass the performance of shallow architectures. Deep neural networks learn manifold stages of representation and abstractions of high dimensional data enabling us to train machine learning models without using hand-crafted features and with minimal preprocessing. Furthermore, the use of convolutional layers in these architectures, especially for image data, has given rise to what are called Convolutional Neural Networks (CNNs) [10-11]. CNNs enable us to make the network take spatial context into account (for images, etc.) while also greatly reducing the computational complexity.

Consequently, a lot of research has been carried out to develop machine learning models using CNNs for problems related to detecting and diagnosing diseases from medical images. They have been found to be very effective, usually surpassing classical image processing-based methods.

While earlier CNN architectures consisted of a single linear pathway of stacked layers; more recently, researchers have started employing residual connections (direct connection of a layer with a layer deeper in the network, skipping one or more layers) and different parallel pathways. This is a departure from networks with a single linear pathway and has produced extraordinary results.

As a prelude to the development of computational models for ML generated radiological scoring of lung TB lesions in chest CT scans, and to evaluate a strategy for integration of above clinical and immunological markers, we demonstrate utility in a well-controlled non-human primate (NHP) model of TB which captures human TB most closely. Importantly, NHPs are outbred, reflecting genetic diversity similar to human populations, vulnerable to natural infection and disease in TB outbreaks [13]. CT scan images analyzed through ML algorithms for automated scoring can be integrated with clinical markers, microbiological markers, and multiplex immune markers (e.g., panels of antibodies against *M.tb.* antigens, cytokines/chemokines etc., previously identified as valuable in TB diagnostics) [14-16].

Materials and Methods

Data Acquisition

Animal Housing Conditions

A group of six male rhesus macaques MMU35414, MMU35446, MMU35603, MMU35710, MMU36365, and MMU36727, were housed at the California National Primate Research Center (CNPRC) at the Animal BioSafety Level-3 (ABSL-3) facility, University of California, Davis (UC Davis), and cared for according to the American Association for Accreditation of Laboratory Animal Care (AALAS) guidelines. Experiments were performed under approval from Institutional Animal Care and Use Committees (IACUC: 15307) at UC Davis. NHPs were maintained in steel cages in pairs whose sizing was scaled to the size of the animal in a temperature-controlled vivarium with continuous monitoring of temperature and humidity. Veterinarians, animal health technicians, and staff technicians conducted daily clinical assessments of animals. This included monitoring weight, temperature, behavior, diarrhea, and opportunistic infections. Animals were fed a balanced commercial macaque chow twice daily with fresh produce twice weekly, with free access to water 24 hours per day. Supplemental food was provided when clinically indicated. Environmental enrichment was provided daily, including manipulanda (forage boards, mirrors, puzzle feeders) and novel foodstuffs. Veterinarians at the CNPRC have established procedures to minimize pain and distress using several approaches. Animals were anesthetized by intramuscular injection (i.m) of ketamine-HCl at 10 mg/kg of body weight prior to all procedures. For M.tb. inoculation and prior to transport for CT scans, animals were additionally anesthetized with 0.03 mg/kg medetomidine HCl injected i.m., and anesthesia was reversed with 0.15 mg/kg atipamezole HCl injected i.m. Analgesics were given to minimize pain and discomfort at the discretion of the veterinary staff, and nutritional supplements were administered as necessary**.**

When euthanasia was necessary, animals were humanely euthanized at the end of the study by a barbiturate overdose, and necropsy procedures were performed by veterinary pathologists and support staff. If an animal's physical condition deteriorated prior to the scheduled endpoint, the animal was

euthanized following the Guidelines for Humane Euthanasia of Animals on Projects (GHEAP) at the CNPRC. Criteria for assessments of health and well-being of the animals were as follows: weight loss of >20%, dyspnea for 24 hours, moderate tachypnea (40-60 bpm) for 36 hours, severe tachypnea (>60) for 12 hours, lethargy longer than 24 hours, or anorexia longer than 36 hours. The fixed live-phase end point was 24 wk postinfection. All animals (tuberculin skin test (TST) negative) were experimentally inoculated with 500 colony-forming units (CFU) of *M. tb*., Erdman strain by bronchial instillation in lower right lung, as previously described [17]. As expected, due to the outbred nature of NHPs, disease outcome was variable. Two animals MMU36365 and MMU36727 became moribund and were humanely euthanized at 15 weeks post-inoculation. The rest of the animals were carried till the conclusion of the study at 24 weeks. Necropsy and pathology procedures were performed by veterinary pathologists [17]. TB was confirmed by postmortem examination including histopathology and *M. tb*. culturing from the homogenized lung tissue [17].

Clinical Data Acquisition

Daily clinical assessment of the animals was performed as previously described [17]. This included physical examinations (e.g., weight, temperature), appetite scores, behavior (breathing abnormalities, cough), diarrhea, and opportunistic infections. Pre- and post-inoculation procedures included blood collections for cryopreservation of plasma, clinical chemistry panel, complete blood counts with erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) measurements. EDTA blood samples were collected at week (WK) 0 (baseline) and longitudinally at 2, 4, 8, 12, 16, 20 and 24 weeks (WK). Plasma samples were collected and stored in aliquots at −80°C.

Immunological Data Acquisition

A panel of ten antigens comprising of RV 3875(ESAT6), Rv3874(CFP10), Rv0934, Rv1886c (Ag85b), Rv3881, Rv2031 (HSPX), Rv3841 (Bfrb), Rv2878c and Rv3619, Rv2875(MPT 70) and one fusion of

Rv3874–Rv3875 (CFP10-ESAT) were used in multiplex microbead immunoassay (Luminex, Austin, TX) as previously described [18,19, 20, 21]. Plasma antibody data from each antigen-coated microbead were reported as median fluorescence intensities (MFI). Cutoff values were calculated in non-infected NHPs (N=405) as previously described (Cutoff = Mean MFI + (3 × standard deviation) [18].

Radiological Data Acquisition

All procedures were performed under ABSL-3 conditions as previously described [17]. Briefly, animals were anesthetized, intubated, and had an intravenous catheter placed. Under the direction of a veterinary radiologist, CT imaging was performed using iodinated contrast medium delivered intravenously at WK 0, 2, 4, 8, 12, 16, and 20. Contiguous transverse images of the thorax were acquired using a 16-slice helical CT scanner and the following parameters: helical acquisition mode; 3 mm collimation; 1.5 pitch; 0° gantry tilt; 120 kVp; 200 mA; 0.6 second acquisition, standard bone/thorax reconstruction algorithm. The anatomic volume of acquisition was determined from initial scout views and varied depending on subject size and relative position within the CT gantry. Following completion of the initial scanning, nonionic iodinated contrast medium (3 ml per kg, Isovue 370, Bracco Diagnostics, Princeton NJ) was administered as an intravenous bolus by manual injection over 1 to 2 minutes and CT images were again acquired using the same acquisition parameters.

CT scan scoring by radiologists

Two radiologists (T.G. and S.Z.) studied the chest CT images of all animals, at all the time points (0, 2,4, 8, 12, 16, 20 and 24 WK's). The extent of disease was assessed in each lung and scored from 0 to 100; the scores are subjective depending upon the degree of lung involvement. The scores for each monkey are presented in Table 1, Table 2, Table 3, and Table 4. The average scores of both radiologists and their standard deviations were calculated.

Disease-signs scoring

Disease-signs were scored from 0 to 4, arbitrarily. Cough was taken as the clearest sign of TB, and given a score of 4. Abnormal and harsh breath, a score of 3. Diarrhea, a score of 2. All other signs (rash face, swollen face, oral lesion, and eyes red) were given a score of 1. No signs, given a score of 0.

Data Analysis

Data augmentation and preprocessing

Due to a limited number of subjects (6 animals), artificial expansion of data set and variation in body posture from animal to animal was compensated by image data augmentation, a common practice in medical image analysis to increase the size of the training dataset, as previously demonstrated [22]. Images were rotated at random angles between 0 to 45 degrees, generating 21 rotations per image. Each image was resized to 256×256 pixels for faster training, and to reduce complexity.

ML Model

ML model in each group of experiments (whole, right, and left lung) was trained for 200 epochs or until the validation loss stopped improving, whichever came first. We used stochastic gradient descent [23] with learning rate $\eta = 0.01$ as the optimization algorithm using a batch size of 10 and categorical cross entropy as the loss in function [24]. Our proposed model , 'TB-Net' derived from 'Inception- ResNet-v2', was optimized by trimming it (removed 24 redundant convolution layers to improve its generalization and performance) to suit our dataset [25]. The proposed network was trained on a machine with GeForce RTX 2080 graphics card with 8GB RAM, taking approximately 18 hours, completing approximately 21,600 epochs.

Training Data Strategy

Each CT scan was sub-divided into seven different groups of optical slices such that every 10^{th} optical slice belonged to each group (3mm in thickness; 'slice' refers to the number of rows of detectors in the z-axis of a CT) and included in the training from optical slice 20 to 80, covering the entire lung (Figure 2). In addition, one optical slice above and one below the said optical slice was included in the analysis for each group. Hence, each group comprised of three different optical slices. ML model was trained randomly where one out for 6 animals was used for testing and the remaining 5 for training at each time point. Following this strategy, models for each of 6 animals were generated for whole, right, and left lung such that 18 models per optical slice per group were generated. For the seven different groups, 18×7 = 126 models were trained (Table 5 and Figure 2). WK0 was considered as 'No TB'. All other time points (WK 4, 8, 12, 16 & 20) in all animals had 'TB' (Table 1 and Table 2). Since majority of the time points correspond to active-TB, class weights and cross entropy function were used to handle the imbalance [24,26]. Training data (data of 5 animals used for training) was divided into 80% for training and 20% for validation (after augmentation), while the images of the sixth animal were used only for testing, based on the trained model. The training methodology was kept consistent for the whole, right, and left lung. The Figure 7 explains why the CT scan slices between 20 and 80 were not include in the training.

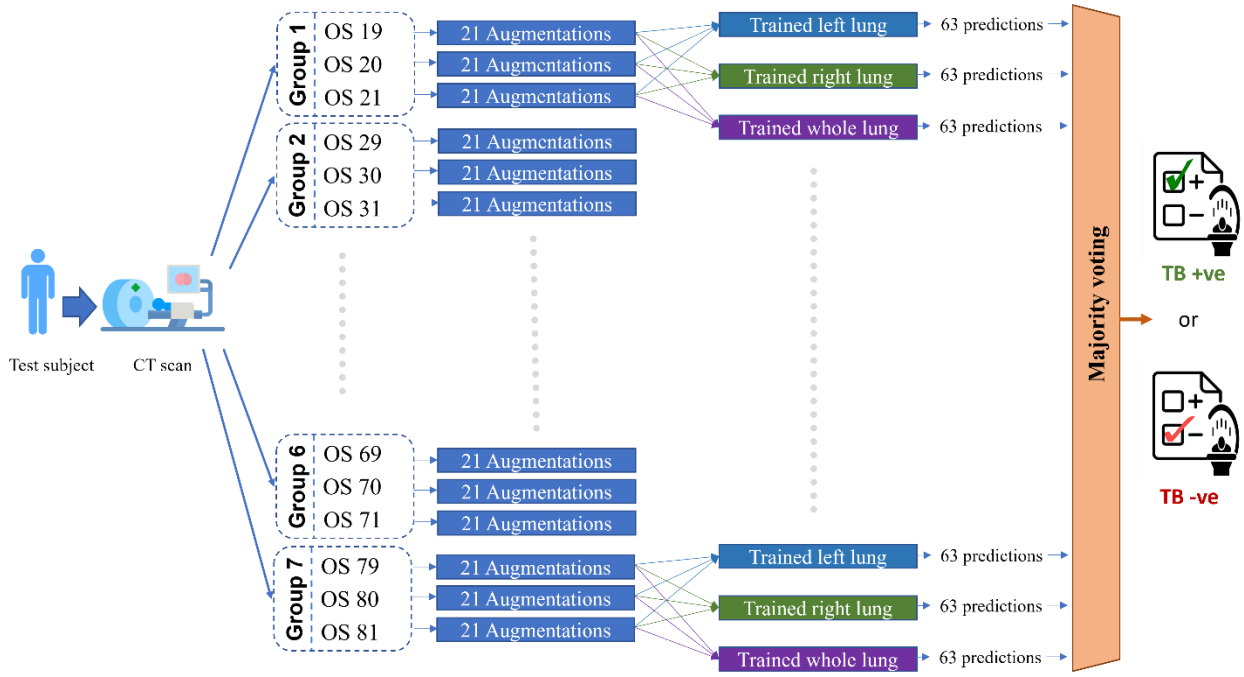

Figure 2: Proposed testing strategy for imaging analysis.

Overall testing strategy of the proposed model for machine learning (ML) analysis of CT scans. Optical slices in CT scan were divided into seven different groups. For convenience, four groups (Groups 1 and 2, and 6 and 7) are shown. Each group contained 3 optical slices such that every 10th optical slice (OS) was included in the analysis starting from Slice 20 to 80, capturing the entire lung. One OS above and one below were also included in the analysis. Augmentation (one CT optical slice equals 21 optical slices as described in the Materials and Methods section) of optical slices was performed for the right, left, and whole lung. Models for TB or NTB (no TB) were trained on augmented OS using well defined TB positive or negative scans (see Materials and Methods). Decision on disease was made based on the majority of slices (out of total 441) predicted TB positive or negative.

CT Scans omitted from Training

MMU36727 at WK0 for the left lung image analysis was not included in the training to avoid the incorrect training because of an abnormality detected by the two radiologists in the lower left lung (Figure 3, marked in red circle). In animal, MMU36365, WK0 and 2 were not included in training for the whole,

right, and left lung to avoid the incorrect training because of a TB lesion detected by the two radiologists (Figure 3). In addition, MMU35414 at WK 2 was not included in raining (whole, right, and left lung) because of a low quality CT scan.

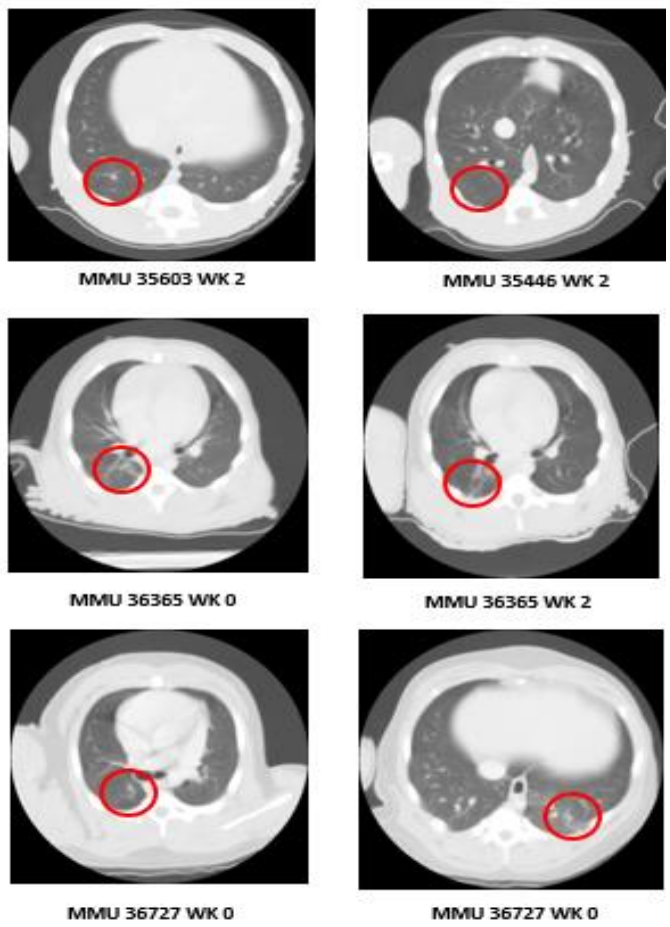

Figure 3: Explanation of discrepant results.

Unexpected and discrepant results in Figure 4 (right lung). In the top left image (MMU35603 at WK2), depicting slight abnormality in the lower right lung (red circle; site of inoculation), suggests inititation of TB which was missed by radiologists, but detected by machine learning, is likely a true positive. Top right image (MMU35446 at WK 2) depicts a slight abnormality in mid right lung (red circle) which is a possible reason for positive prediction by ML but called negative by radiologists; likely a false positive. In the middle row both images (MMU36365 at WK0 and WK2), contained TB related abnormalities, missed by TST, therefore, images at these time points were not included in ML image analysis, avoiding

incorrect training; radiologists' scores are as shown in Figure 4. In MMU36727 at WK 0 (bottom row), first image shows the non-TB abnormality confirmed by radiologists in the lower right lung, it was included in the analysis for efficient detection of TB. The second image shows the TB abnormality confirmed by radiologists, this was not included in the analysis to avoid the incorrect training.

Testing Data Strategy

The above augmentation process was also performed for testing strategy. The augmented optical slices were tested against the trained whole, right and left lung models for each optical slice, in each of the 7 groups. Predicted model output was binary, TB, or No TB (NTB), in a given augmented optical slice, based on the majority voting during testing. The number of predictions (TB or NTB) were counted for each of the 7 groups such that each group had 63 augmentations, therefore, for 7 different groups there were 63×7 = 441 predictions. For each test animal, at a particular time point, the decision of TB or NTB was made based on majority prediction out of a total of 441 predictions (Figure 2).

Outcomes of ML Image Analysis

The metrics used for performance comparisons are defined below.

- **True Positive (TP):** This is defined as if our ML model predicted TB positive and its label was TB positive, e.g., predicted optical slice is TB positive and its true.
- **True Negative (TN):** This is defined as if our ML model predicted TB negative and its label was TB negative, e.g., predicted optical slice is TB negative and its true.
- **False Positive (FP):** This is Type I error which is defined as: if our ML model prediction on the optical slice is TB positive, but its label was TB negative.
- **False Negative (FN):** This is Type II error which is defined as: if our ML model prediction on the optical slice is TB negative, but its label was TB positive.

- **ML Prediction (P):** It gives binary output TB negative (NTB = No TB), or TB positive based on the majority voting. For example, If TN >FP then it was considered No TB, otherwise TB. Also, if TP > FN, it was considered as TB, otherwise No TB (S4A Table, right lung analysis, and S4B Table, left lung analysis. The positive predicted optical slices number was used to calculate the ML ratio, which is defined as the positive predicted optical slices divided by the total number of predictions. Finally, background (0.5) was subtracted from the ML ratio to obtain positive ML ratio. This ratio represents calculated ML Score.

Results

Lung CT image analysis by ML

Right lung analysis

In the initial assessment by the radiologists (TG, and SZ.), no TB lesions were visualized at WK 0, and it was assumed there would be none at WK2. At WK4, TB lesions clearly initiated at the sight of inoculation in all animals and progressed further at the later time points. Right lung ML scores (blue line (Figure 4); average of three experiments) displayed strong correlation (ranging from R = 0.79 to R = 0.99; Pearson Correlation) with radiologists' scores (red line, average of two; Figure 4). Both types of scores show disease progression over time in all animals except MMU35710 where disease burden reduced after WK12 (Figure 4). A clear advantage of ML based scoring over radiologists' manual scoring of CT images is the machine's high consistency across three runs, in all animals, at most time points. In contrast, the error between two radiologists is high.

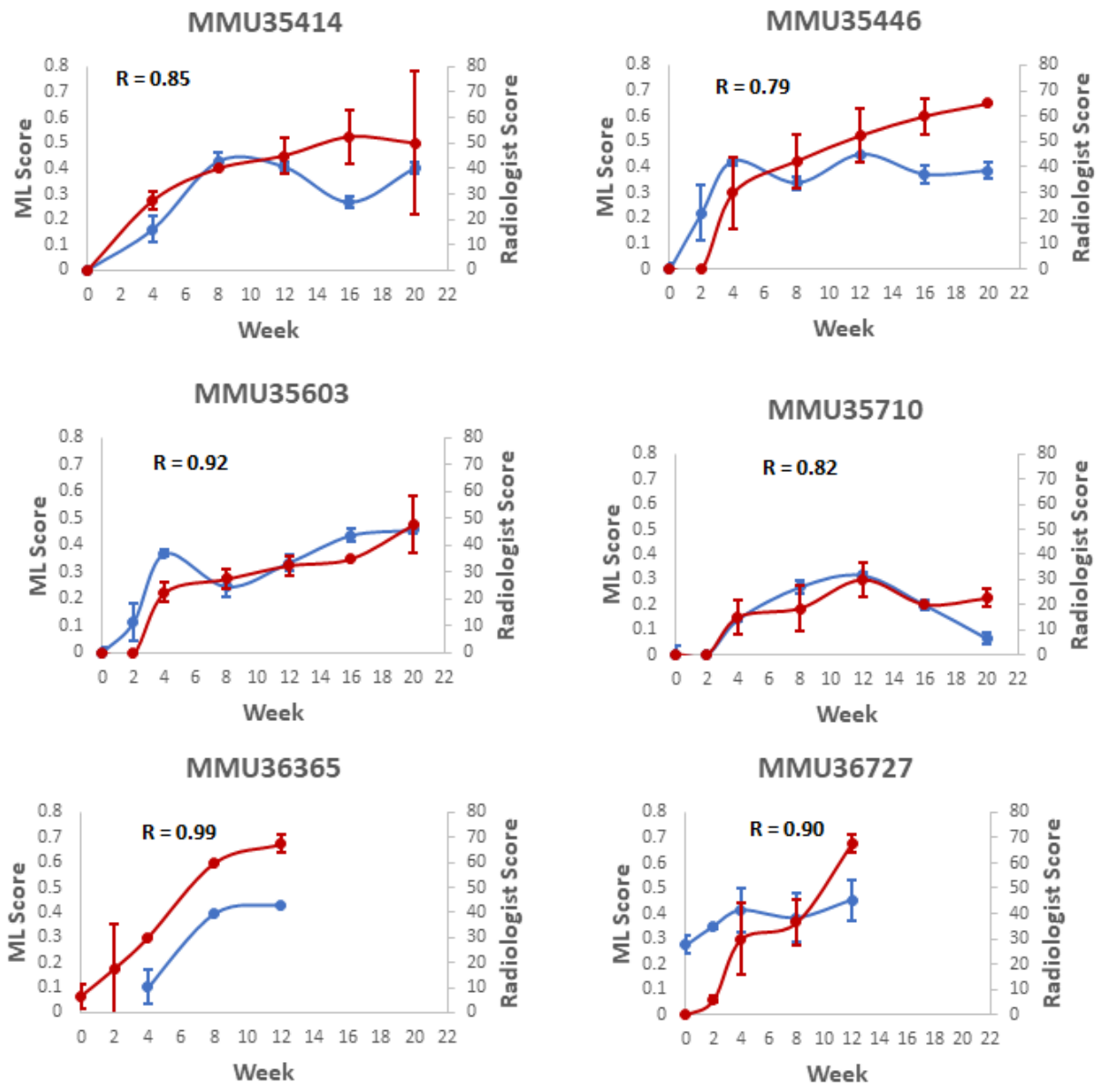

Figure 4: Right lung CT Scan imaging analysis.

ML score of CT images, weeks post-inoculation (right lung) of six rehsus macaques with *M.tb.* (Erdman strain). Blue line represents average score of three ML runs, and red line average of two radiologists' scores. Error bars represent standard deviation. Pearson correlation coefficient (R) is shown.

Left lung analysis

The extent of disease is limited in left lung except at the later time points in MMU36365 and MMU36727. Because of a low occurrence of TB lesions, correlations between the ML scores and radiologists' manual scoring are low (Figure 5). Nevertheless, a salient feature of the comparison is once again that ML scores are very consistent across three runs while radiologists' manual scores are not (Figure 5).

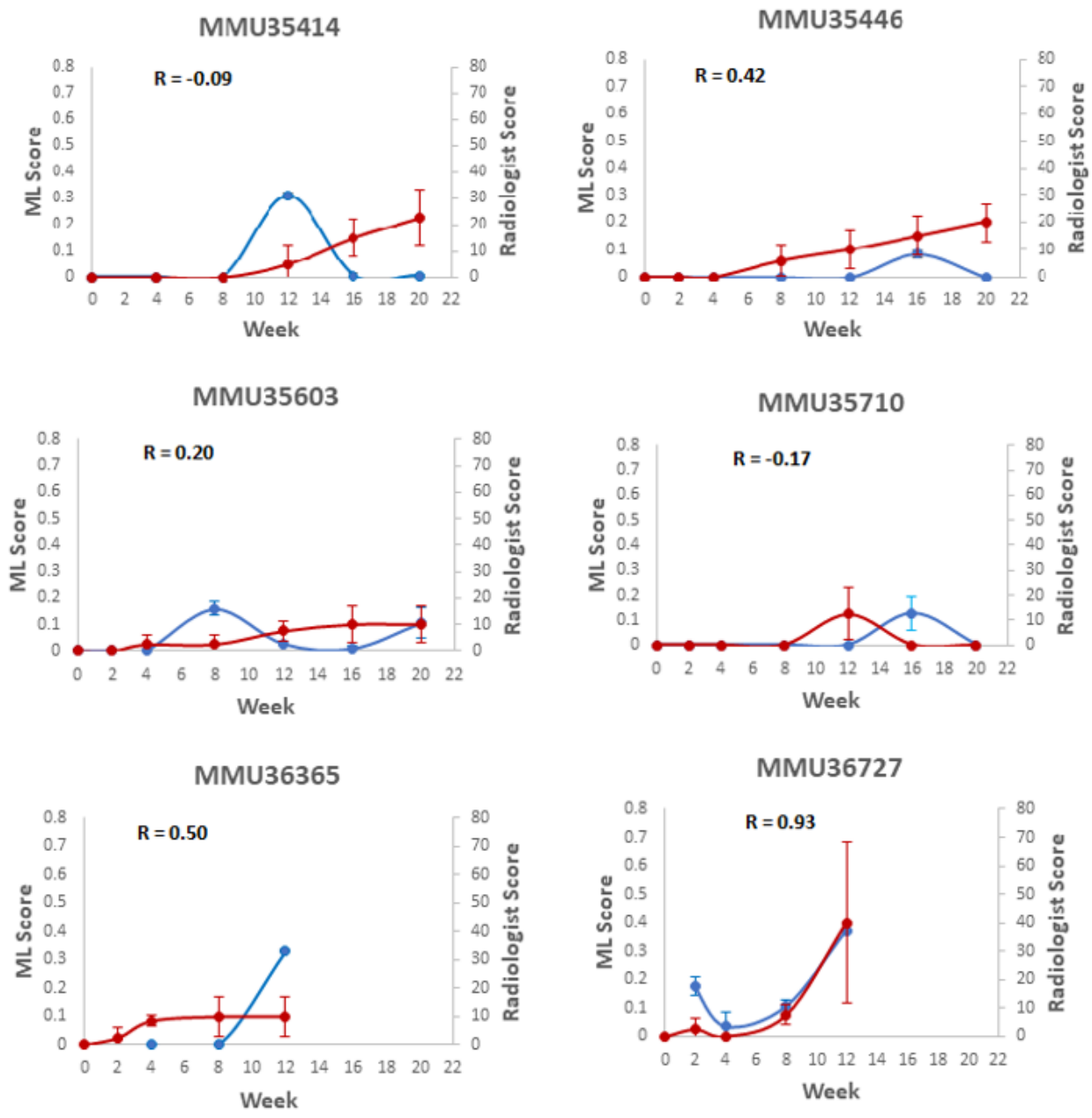

Figure 5: Left lung CT Scan imaging analysis.

ML analysis for left lung, as described for Figure 4. The scale for ML Score and Radiologist Scores are kept the same as for the right lung analysis in Figure 4 for direct comparison.

Whole lung analysis

ML and radiologists' scores for the whole lung were similar to those for the right lung (initial infection site), demonstrating right lung scores' dominance in this study (Figure 8).

Correlation of ML analysis (right lung) with TB granuloma and total TB-lesion volumes

ML scores of CT scans available at the latest time points (WK12 for MMU36365 and MMU36727; WK20 for the rest) were compared with TB granuloma and total TB-lesion volumes at the time of necropsy (WK15 for MMU36365 and MMU36727; and WK24 for the rest). A high degree of correlations were achieved between the ML scores and granuloma volumes, as well as total lesion volumes, (Pearson Correlation, R = 0.68 and R = 0.69, respectively, Figure 6) strongly suggesting ML based scoring faithfully captures disease severity.

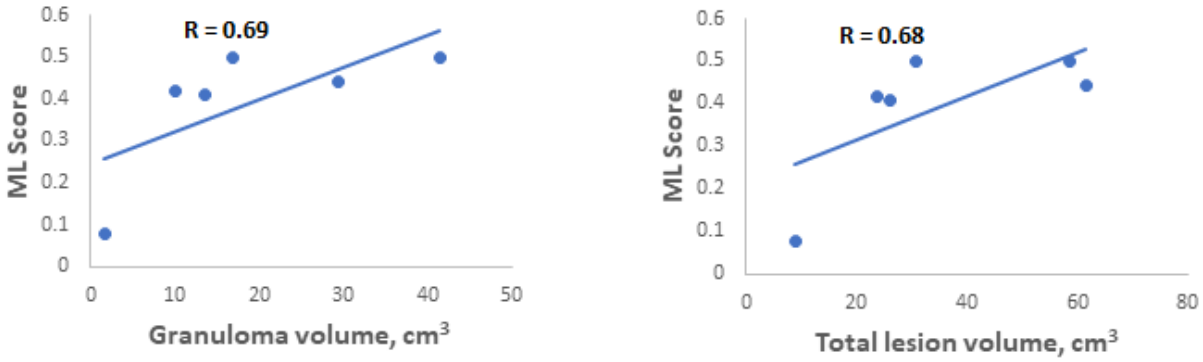

Figure 6: Correlation analysis of ML score with Granuloma and Total lesion volume, cm^3.

Correlation of ML score to TB lesion volumes (right lung). The left scatterplot shows ML score versus granuloma volume, and on the right is ML score versus total TB-lesion volume (includes non-granuloma lesions), at necropsy, in all animals. The latest CT scans available for all animals were taken 3 to 4 weeks prior to necropsy. Pearson correlation coefficients 'R' are shown.

Correlation of CT scores by ML (right lung) with clinical and blood biomarkers data

Disease signs

A high degree of correlation between ML score and signs (R = 0.76 to 0.98) was observed in all animals except MMU35710 (R = 0.34) in which disease remained mild over 24 weeks (Figure 9). This shows disease severity demonstrated by CT ML score may reflect signs' severity in the NHP model of TB, suggesting this approach may be valuable in the analysis of symptoms in TB patients with regards to disease severity.

Blood markers

Correlations of various blood markers with CT ML score were assessed and were generally found to be strong as follows: ESR (R = 0.84); C-reactive protein (0.58); total protein: (R = 0.70); plasma protein (R = 0.64); and lactate dehydrogenase (R = 0.46) (Figure 10). Among the above markers ESR, total protein, and CRP reflect inflammatory responses which in TB patients and macaques infected with *M. tb.* may serve as markers to identify correlates of TB disease progression [27-31].

Anti-M. tb. Antibodies

A multiplex panel of ten antibodies (against ten *M. tb.* Antigens) displayed variable immune responses, as in humans [18,19]. Antibodies against one antigen, CFP10-ESAT6 fusion protein was consistently strong in 5 of 6 animals over 24 weeks, first appearing at WK4, showing good correlations in four animals (R = 0.36 to 0.58) (Figure 11).

Discussion and Conclusion

Among the top three global infectious killers, for the last many decades (TB, HIV-AIDS and malaria), yearly TB deaths are on the rise (1.6 million), while in the other two (HIV-AIDS: 1.0 million, and malaria: 0.4 million) are on the decline. Such roaring rise in global TB is surprising considering anti-TB drugs are generally curative. The problem lies in the dearth of globally available efficient diagnostic approaches. Majority of diagnostics, including even state-of-the-art molecular tests introduced in the recent times, are limited to sputum as a specimen while a complex disease like TB requires multi-modality approaches for a holistic examination of the disease state and progression, as described here. AI based approaches offer Deep Neural Networks (DNNs) architectures with ready capacity to perform complex integrations in an automated fashion for comprehensive solutions, combining multimodal data [6, 7]. DNNs learn multifaceted representations and abstractions as in multi-dimensional data more efficiently, enabling training of machine learning computational models without using hand-crafted features, with minimal data preprocessing.

As a prelude to human studies, we demonstrate the utility of such high-powered computational models in a well-defined, high fidelity, and well-controlled rhesus macaque model of TB. Radiological (CT scan) imaging data were analyzed employing CNN based convolutional layered architectures yielding automated ML scores which correlated remarkably well with manual scoring by radiologists. If fact, the ML scores were much more consistent from run to run in comparison to the radiologists' scores which displayed a high degree of error (Figure 4). Additionally, after the preliminary ML runs, in two animals, MMU36365 and MMU36727, lesions resembling TB were discovered in pre-infection (WK0) CT scans (TST (skin test) is low sensitivity and specificity TB test compared to CT imaging), prompting the radiologists to take a more probing look at the CT scans, confirming the presence of such lesions; ML training models were optimized accordingly in final runs. Incidentally, MMU36365 and MMU36727 not only displayed a rapid onset of disease but also more severe disease (Figure 4 and Figure 5), requiring

early euthanasia at WK 15. This episode emphasizes the enormity of human error introduced even by trained radiologists.

The ultimate confirmation of TB lesions is achieved through histological examination of the affected tissues, as previously described [32,33]. In this study, we validated our ML based image analysis against histologically determined granuloma and total TB-lesion volumes, demonstrating a high degree of correlation (R = 0.68 and 0.69, respectively). This result demonstrates the accuracy of our ML based CT image analysis despite two key limitations: 1) Latest CT scans were obtained 3 to 4 weeks before the necropsied lung specimens, and 2) Histological determinations are marred by technical tediousness and human subjectivity [32-35].

We also found a high degree of correlation between ML based CT scores and data from multimodality approaches: (1) Clinical signs (e.g., cough, abnormal breath etc.), 2) Blood markers (e.g., CRP, ESR, and other inflammatory markers valuable in understanding TB disease progression), and 3) Anti-*M. tb.* antibodies which have been shown by us to be of high diagnostic value [18,19]. Some of the other blood parameters such as total white blood cells, specific blood cell types, (e.g., neutrophils, eosinophils etc.) and other analytes (creatinine, sodium, potassium etc.) were variable among animals with inconsistent results.

This study provides a model for undertaking human TB clinical studies, which the authors have already initiated. In the ongoing human study, ML based image analysis of CTs and chest X-rays (CXR) are under further optimization for complete automation. Radiological imaging data, and other multimodality data, from TB patients is under acquisition at different time points covering 6 months of directly observed therapy short course (DOTS) as follows: at diagnosis (Month 0), Month 2, Month 4, and Month 6). CT scan and CXR image scoring by ML, current diagnostic test scores (microbiologically validated by culture score), and upcoming blood biomarker scores (including immunological biomarkers), will be

integrated employing deep learning AI approaches. The eventual goal is to obviate the exclusive reliance on sputum specimen which is not only variable from patient to patient but also day to day. The AI based integrated score so obtained will be delivered to the physician as clinical decision support (CDS) via open medical record system (Open-MRS). In addition, individual scores on multimodal data sets, contributing to total score integrated by AI, will be made readily accessible on the physician's PC for detailed clinical investigations as needed, at the click of a mouse (Figure 1).

Challenges and Limitations

Noisy and limited data

To reduce the respiratory motion artifacts during a CT scan process, patients are asked to hold their breath. Unlike humans it is difficult to do so in the non-human primates, which leads to the noisy and multidirectional CT scans for same animal at different time points. This led us to define our alignment criteria before imaging analysis. We consider the following four parameters for the alignment of the CT scan images.

- where vertebral column is maximum.
- where the diaphragm has the same area (shape)
- where the white portion from vertebral column to the
- diaphragm is the same in all images.
- where the black portion is same in all images.

The Figure 12 shows the aligned images at different time points for animal MMMU35446.
Another challenge was having a limited dataset (n =6) for training the ML models. This challenge was solved through data augmentation and choosing the less complex model to avoid overfitting.

Generalizability of informatics methods

One of the limitations of the study is the generalizability of the informatics methods i.e., ML/DL techniques applied to the imaging analysis in the NHP models for tuberculosis. When a ML model is trained on small dataset (n =6) in this study, it may not be able to capture all the patterns because it does not represent the entire population. The proposed model may not perform well when applied to the dataset

outside of the study. To overcome this issue, transfer learning can be employed by fine tuning a pre-trained model on a small dataset.

Future Work

Human TB Patients Analysis

The automated solution for tuberculosis detection developed in this for NHP will be replicated for the human TB patients in combination with biomarkers. The human TB patient's data (CT's and X-rays) have been analyzed and annotated from two radiologists independently. The radiological, clinical and microbiological data of TB patients have been collected at 0 months, 2 months, 4 months, and 6 months. The number of samples at these time points varies.

We plan to perform the following analysis using the data mentioned above at each time point and across all time points.

- Correlation analysis of clinical and microbiological
- Correlation analysis of clinical and radiological
- Correlation analysis of clinical and immunological
- Correlation analysis of microbiological and radiological
- Correlation analysis of microbiological and immunological

Based on the results above, we will provide a TB diagnostic solution for humans utilizing computational modelling for clinical, radiological, and immunological variables.

Synthetic CT generation

Deep learning (DL) methods for synthetic CT (sCT) generation have been used in the literature recently. For example, generation of synthetic CT Images from MRI [36-37]. A systematic review of DL based methods for generating synthetic CT generation in radiotherapy and PET can be found here [38].

sCT scans for tuberculosis detection can be achieved using deep learning-based methods. By generating sCT scans that closely resemble real CT scans, deep learning algorithms can be trained on a diverse and large dataset of medical images for the purpose of medical image analysis, computer-aided diagnosis, and medical education, with the rationale that these algorithms can learn complex relationships between the input data and their associated labels, such as the presence or absence of tuberculosis.

Synthetic CT scans can be useful in cases where the real data is limited or difficult to obtain, e.g., shortage of radiologist in specific areas or patients unwilling or unable to undergo CT scans due to concern about radiation exposure. Synthetic CT scan generation has the potential to improve medical imaging and diagnosis, leading to better patient outcomes and more efficient healthcare systems.

The following steps can be performed for synthetic CT scan generation for tuberculosis detection.

- Gathering the dataset of CT scans that include both normal and tuberculosis-infected lung images. Preprocessing of the collected data to ensure that all CT scans are in the same format and resolution.
- Training a deep learning network, e.g. convolutional neural network (CNN), to learn the features that distinguish normal and tuberculosis-infected CT scans.
- Use the above trained model to generate synthetic CT scans. This can be achieved through a technique called generative adversarial networks (GANs), which involves training two neural networks in combination: one to generate synthetic images, and the other to discriminate between real and synthetic images. The two networks are trained in a competition, with the generator trying to create increasingly realistic synthetic images, and the discriminator trying to distinguish between the synthetic and real images.
- Evaluate the performance of the synthetic CT scans by comparing them to real CT scans using various metrics, such as accuracy, precision, recall, and F1 score.

Transfer Learning

The model trained in this study will be used as a pre trained model for imaging analysis in the human TB patients. This will reduce the amount of data required for calculating the CT based ML score and improve the speed and accuracy of the learning process.

Contributions to health informatics

The contribution of this thesis to the health informatics field includes, but not limited to the following:

- Informatics methods

  The informatics methods developed in this thesis to automatically score the CT images based on the severity of the disease. These methods will be publicly available to fine-tune and use for the similar problems in the health informatics field.

- Data availability

  Radiological, clinical, and microbiological data collected for non-human primates and human TB patients during this study will be publicly available to the community for further research purpose after the publication.

- Solutions to the noisy and limited data

  Another contribution to the health informatics is solutions provided in this study to the noisy and limited dataset e.g., alignment criteria and less complex novel deep learning method to avoid overfitting.

References

1. World Health Organization. Global tuberculosis report 2013. World Health Organization; 2013.

2. Bhalla AS, Goyal A, Guleria R, Gupta AK. Chest tuberculosis: Radiological review and imaging recommendations. Indian J Radiol Imaging. 2015;25(3):213-25. doi: 10.4103/0971-3026.161431. PubMed PMID: 26288514; PubMed Central PMCID: PMCPMC4531444.

3. Jeong YJ, Lee KS. Pulmonary tuberculosis: up-to-date imaging and management. AJR Am J Roentgenol. 2008;191(3):834-44. doi: 10.2214/ajr.07.3896. PubMed PMID: 18716117.

4. Safianowska A, Walkiewicz R, Nejman-Gryz P, Grubek-Jaworska H. [Two selected commercially based nucleic acid amplification tests for the diagnosis of tuberculosis]. Pneumonol Alergol Pol. 2012;80(1):6-12. PubMed PMID: 22187175.

5. Steingart KR, Ng V, Henry M, Hopewell PC, Ramsay A, Cunningham J, et al. Sputum processing methods to improve the sensitivity of smear microscopy for tuberculosis: a systematic review. Lancet Infect Dis. 2006;6(10):664-74. doi: 10.1016/s1473-3099(06)70602-8. PubMed PMID: 17008175.

6. Goodfellow I, Bengio Y, Courville A. Deep learning: MIT press; 2016.

7. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. science. 2006;313(5786):504-7.

8. He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.

9. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Medical image analysis. 2017;42:60-88.

10. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556. 2014.

11. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al., editors. Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition; 2015.

12. Zhou K, Greenspan H, Shen D. Deep learning for medical image analysis: Academic Press; 2017.

13. Flynn JL, Gideon HP, Mattila JT, Lin PL. Immunology studies in non-human primate models of tuberculosis. Immunol Rev. 2015;264(1):60-73. doi: 10.1111/imr.12258. PubMed PMID: 25703552; PubMed Central PMCID: PMCPMC4339213.

14. Hussainy SF, Zaffar F, Zaffar MA, Khaliq A, Khan IH, Ahmad R. Decision-tree inspired classification algorithm to detect Tuberculosis (TB). 2017.

15. Khan IH, Ravindran R, Krishnan VV, Awan IN, Rizvi SK, Saqib MA, et al. Plasma antibody profiles as diagnostic biomarkers for tuberculosis. Clin Vaccine Immunol. 2011;18(12):2148-53. Epub 20111005. doi: 10.1128/cvi.05304-11. PubMed PMID: 21976221; PubMed Central PMCID: PMCPMC3232686.

16. Ravindran R, Krishnan VV, Khanum A, Luciw PA, Khan IH. Exploratory study on plasma immunomodulator and antibody profiles in tuberculosis patients. Clin Vaccine Immunol.

2013;20(8):1283-90. Epub 20130612. doi: 10.1128/cvi.00213-13. PubMed PMID: 23761664; PubMed Central PMCID: PMCPMC3754529.

17. Luciw PA, Oslund KL, Yang XW, Adamson L, Ravindran R, Canfield DR, et al. Stereological analysis of bacterial load and lung lesions in nonhuman primates (rhesus macaques) experimentally infected with Mycobacterium tuberculosis. Am J Physiol Lung Cell Mol Physiol. 2011;301(5):L731-8. Epub 20110826. doi: 10.1152/ajplung.00120.2011. PubMed PMID: 21873450; PubMed Central PMCID: PMCPMC3213984.

18. Khan IH, Ravindran R, Yee J, Ziman M, Lewinsohn DM, Gennaro ML, et al. Profiling antibodies to Mycobacterium tuberculosis by multiplex microbead suspension arrays for serodiagnosis of tuberculosis. Clinical and Vaccine Immunology. 2008;15(3):433-8.

19. Ravindran R, Krishnan VV, Dhawan R, Wunderlich ML, Lerche NW, Flynn JL, et al. Plasma antibody profiles in non-human primate tuberculosis. Journal of medical primatology. 2014;43(2):59-71.

20. Khan IH, Kendall LV, Ziman M, Wong S, Mendoza S, Fahey J, et al. Simultaneous serodetection of 10 highly prevalent mouse infectious pathogens in a single reaction by multiplex analysis. Clinical and Vaccine Immunology. 2005;12(4):513-9.

21. Khan IH, Mendoza S, Yee J, Deane M, Venkateswaran K, Zhou SS, et al. Simultaneous detection of antibodies to six nonhuman-primate viruses by multiplex microbead immunoassay. Clin Vaccine Immunol. 2006;13(1):45-52. doi: 10.1128/cvi.13.1.45-52.2006. PubMed PMID: 16425999; PubMed Central PMCID: PMCPMC1356626.

22. Chlap P, Min H, Vandenberg N, Dowling J, Holloway L, Haworth A. A review of medical image data augmentation techniques for deep learning applications. Journal of Medical Imaging and Radiation Oncology. 2021;65(5):545-63.

23. Robbins H, Monro S. A stochastic approximation method. The annals of mathematical statistics. 1951:400-7.

24. Aurelio YS, de Almeida GM, de Castro CL, Braga AP. Learning from imbalanced data sets with weighted cross-entropy function. Neural processing letters. 2019;50(2):1937-49.

25. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA, editors. Inception-v4, inception-resnet and the impact of residual connections on learning. Thirty-first AAAI conference on artificial intelligence; 2017.

26. Zhu M, Xia J, Jin X, Yan M, Cai G, Yan J, et al. Class weights random forest algorithm for processing class imbalanced medical data. IEEE Access. 2018;6:4641-52.

27. Foreman TW, Mehra S, LoBato DN, Malek A, Alvarez X, Golden NA, et al. CD4+ T-cell–independent mechanisms suppress reactivation of latent tuberculosis in a macaque model of HIV coinfection. Proceedings of the National Academy of Sciences. 2016;113(38):E5636-E44.

28. Mehra S, Foreman TW, Didier PJ, Ahsan MH, Hudock TA, Kissee R, et al. The DosR regulon modulates adaptive immunity and is essential for Mycobacterium tuberculosis persistence. American journal of respiratory and critical care medicine. 2015;191(10):1185-96.

29. Mehra S, Golden NA, Dutta NK, Midkiff CC, Alvarez X, Doyle LA, et al. Reactivation of latent tuberculosis in rhesus macaques by coinfection with simian immunodeficiency virus. Journal of medical primatology. 2011;40(4):233-43.

30. Sulochana S, Siddartha JR, Fathima J. Clinical Significance of Erythrocyte Sedimentation Rate in Tuberculosis. Research Journal of Pharmacy and Technology. 2022;15(1):245-9.

31. Samanta S, Sharma A, Das B, Mallick AK, Kumar A. Significance of total protein, albumin, globulin, serum effusion albumin gradient and LDH in the differential diagnosis of pleural effusion secondary to tuberculosis and cancer. Journal of clinical and diagnostic research: JCDR. 2016;10(8):BC14.

32. Gibson-Corley KN, Olivier AK, Meyerholz DK. Principles for valid histopathologic scoring in research. Veterinary pathology. 2013;50(6):1007-15.

33. Phillips BL, Mehra S, Ahsan MH, Selman M, Khader SA, Kaushal D. LAG3 expression in active Mycobacterium tuberculosis infections. The American journal of pathology. 2015;185(3):820-33.

34. Klopfleisch R. Multiparametric and semiquantitative scoring systems for the evaluation of mouse model histopathology-a systematic review. BMC veterinary research. 2013;9(1):1-15.

35. Rousselet MC, Michalak S, Dupré F, Croué A, Bedossa P, Saint-André JP, et al. Sources of variability in histological scoring of chronic viral hepatitis. Hepatology. 2005;41(2):257-64.

36. Gupta D, Kim M, Vineberg KA, Balter JM. Generation of synthetic CT images from MRI for treatment planning and patient positioning using a 3-channel U-net trained on sagittal images. Frontiers in oncology. 2019 Sep 25;9:964.

37. Fritz J. Automated and radiation-free generation of synthetic CT from MRI data: does AI help to cross the finish line?. Radiology. 2021 Feb;298(2):350-2.

38. Spadea MF, Maspero M, Zaffino P, Seco J. Deep learning based synthetic-CT generation in radiotherapy and PET: a review. Medical physics. 2021 Nov;48(11):6537-66

Appendix A – Supplementary Material

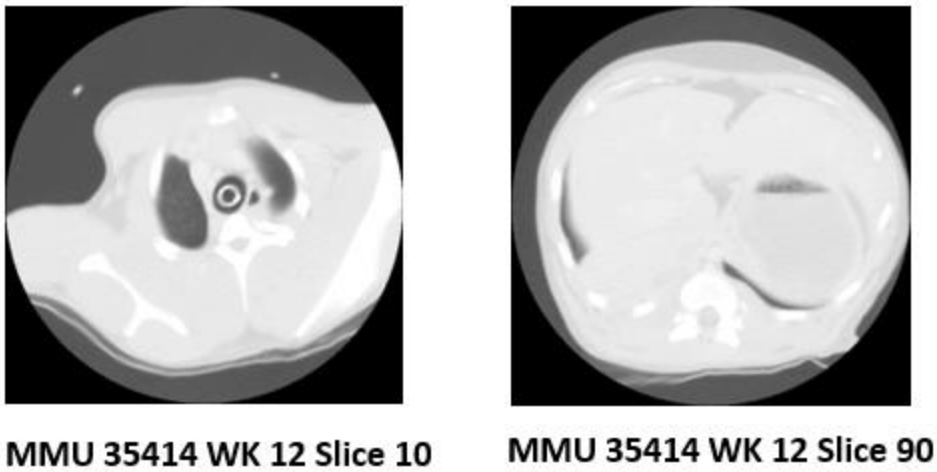

Figure 7: Optical slices excluded from imaging analysis.

The optical slices before 20 and after 80 were not included in image analysis as they display little lung information in CT scans.

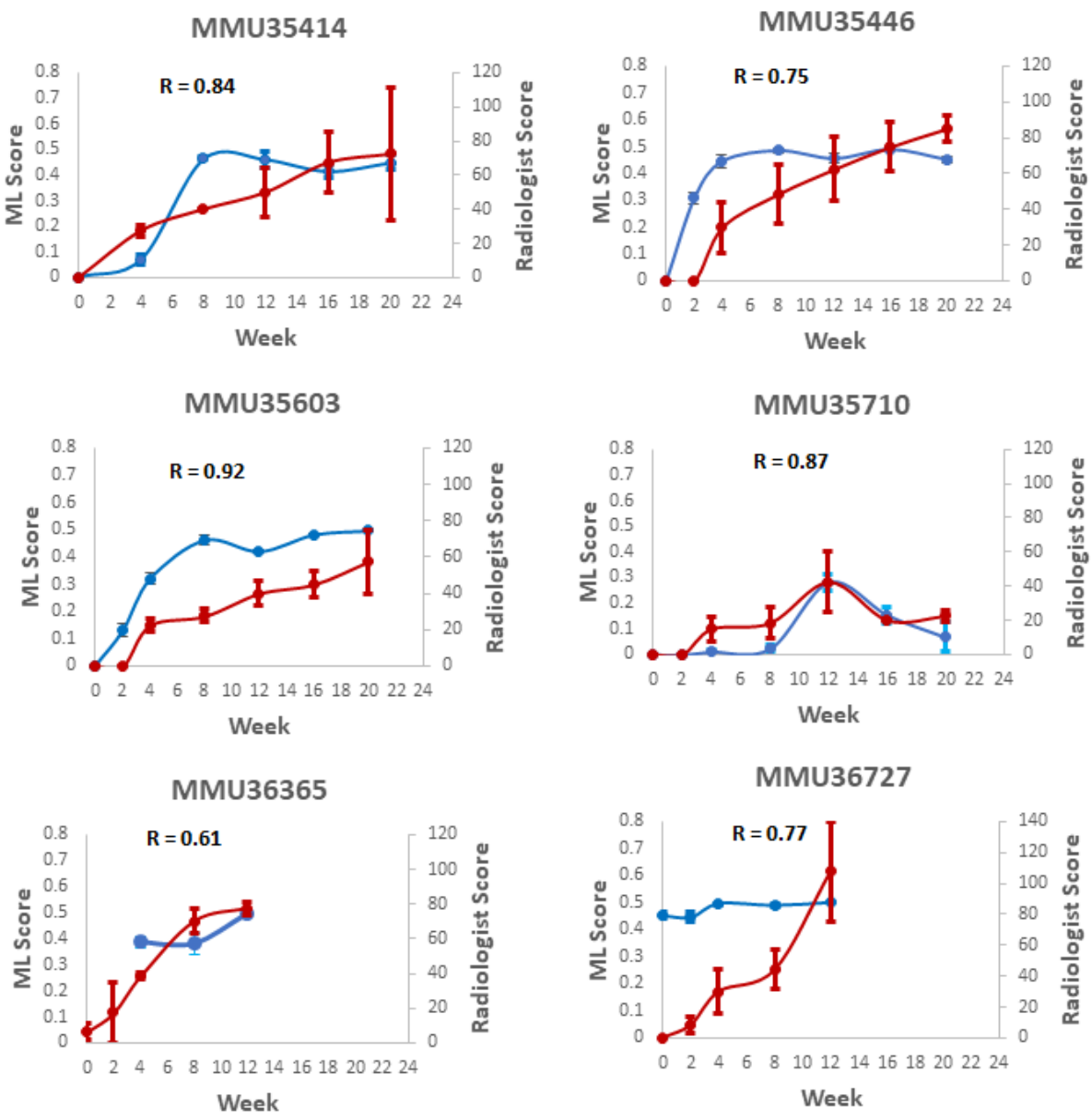

Figure 8: Whole lung CT Scan imaging analysis.

Whole lung image analysis (average of two experiments) of CT scans at various time points post-inoculation of six rehesus macaques with pathogenic *M. tb.* Blue line represents average score of three ML runs, and red line two radiologists' score. Error bars show standard deviation. Pearson correlation coefficient (R) is shown.

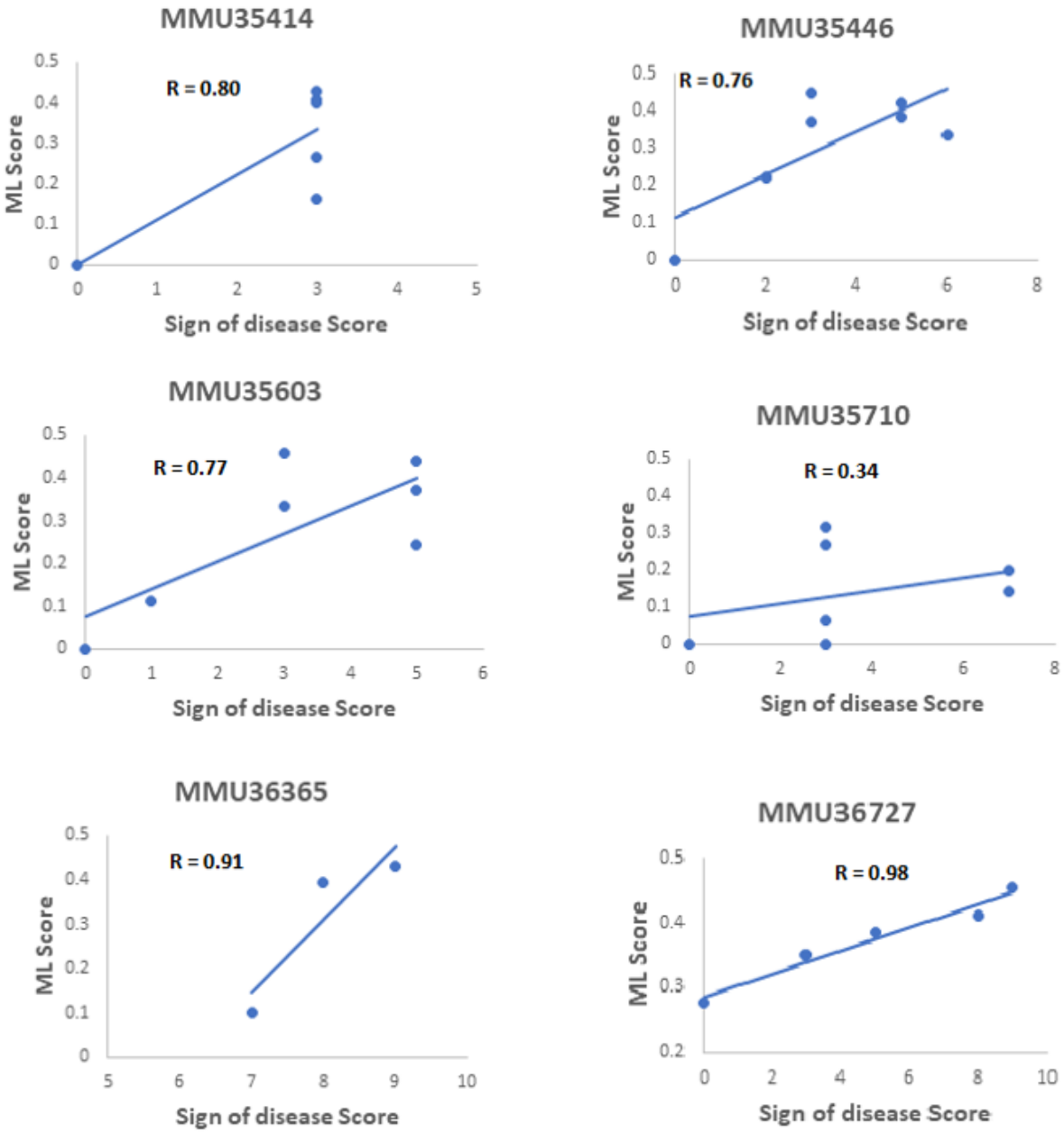

Figure 9: Correlation analysis disease signs.

ML Score (right lung) versus symptom scores for each animal. The Pearson correlation coefficient is shown.

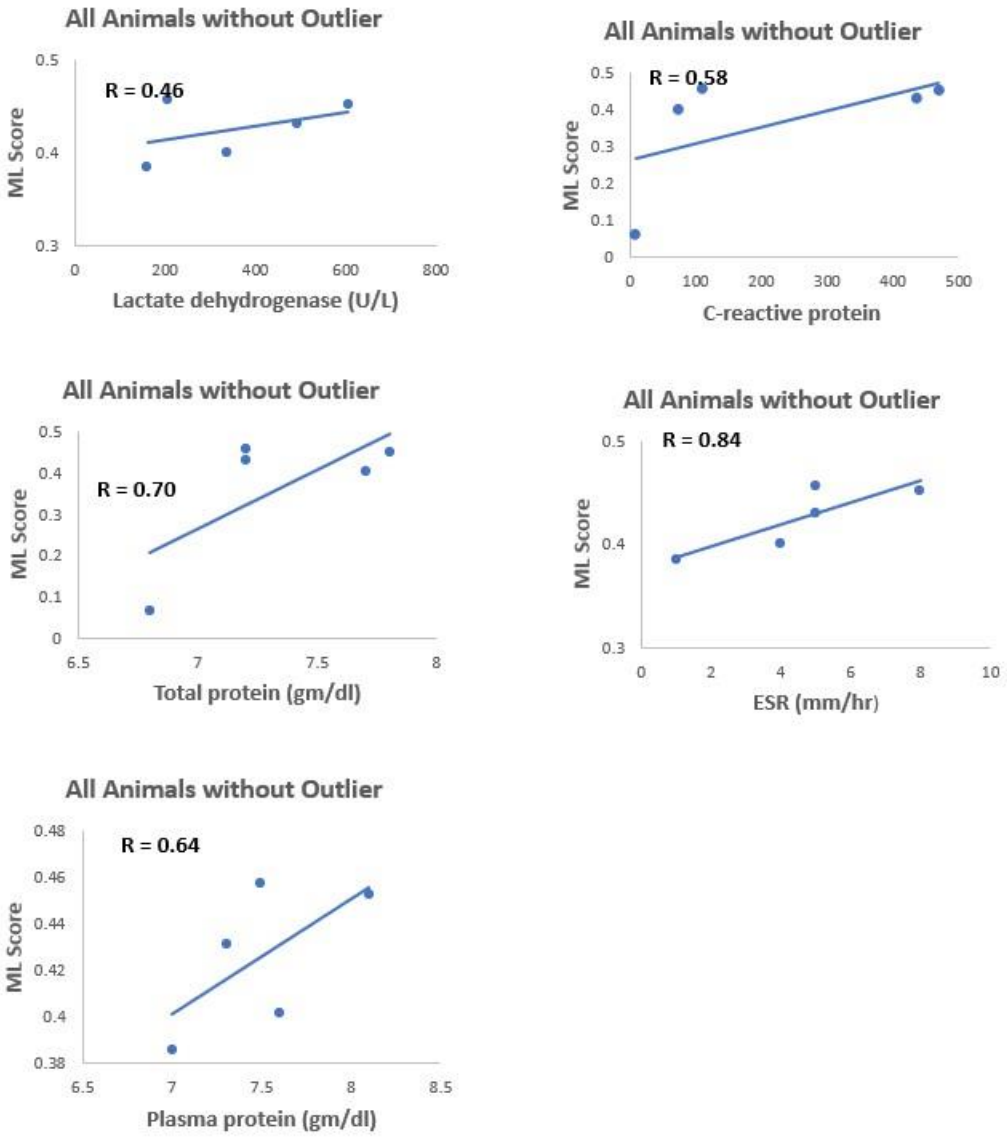

Figure 10: Correlation analysis with blood markers.

Scatterplots of ML Score (right lung) versus blood markers (LDH, C-reactive protein, total protein, ESR, plasma protein) for all animals at the time point right before necropsy. Pearson correlation coefficients are shown. For C-reactive protein, total protein, the outlier animal is MMU 35446. For ESR, LDL and plasma protein, the outlier animal is MMU 35710.

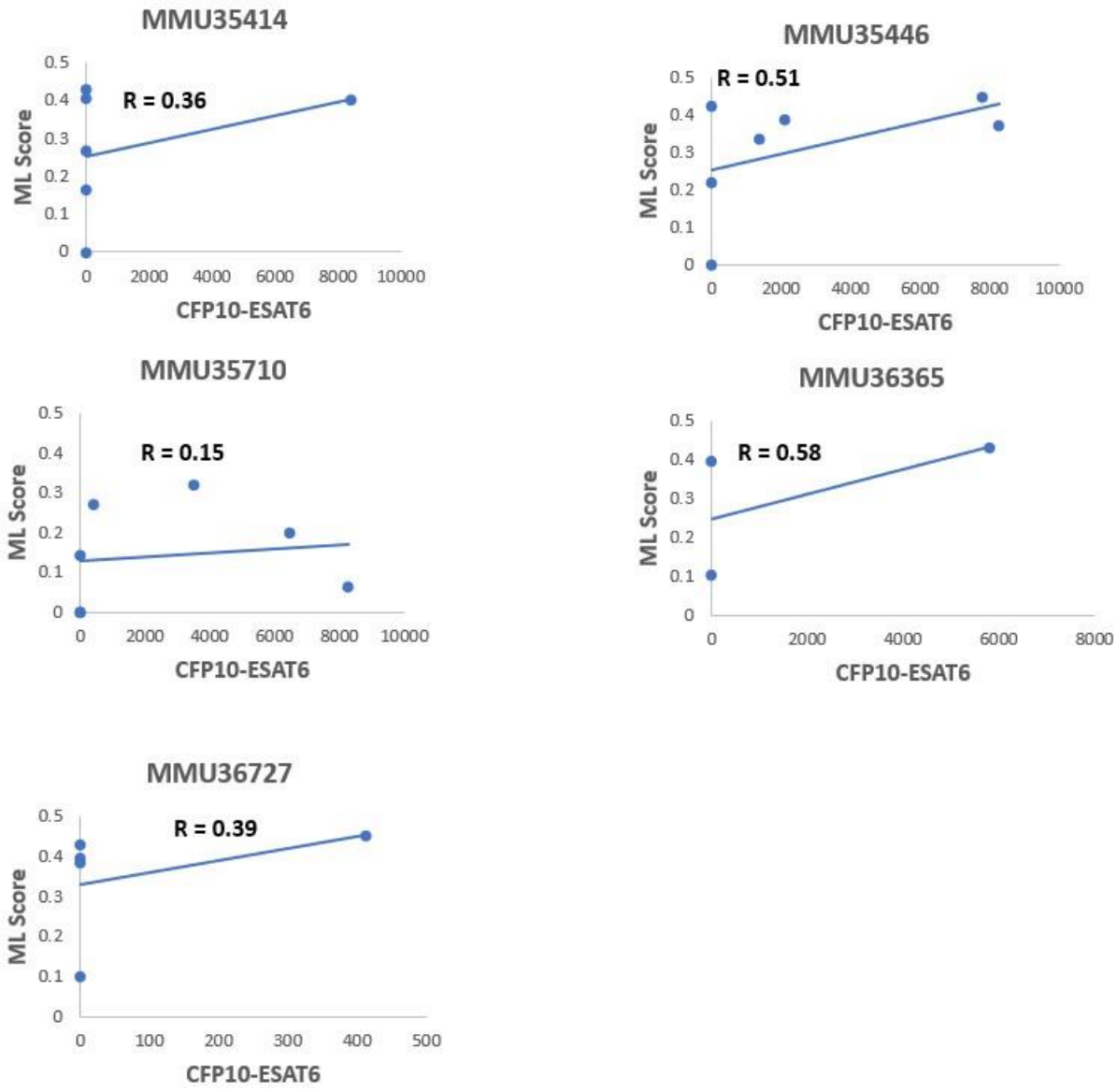

Figure 11: Correlation analysis with Antibodies.

Scatterplot of ML Score (for right lung) versus CFP10-ESAT6 antibodies detected in 5 of 6 animals. Median Fluorescence Intensities (MFIs) and Pearson correlation coefficients are shown.

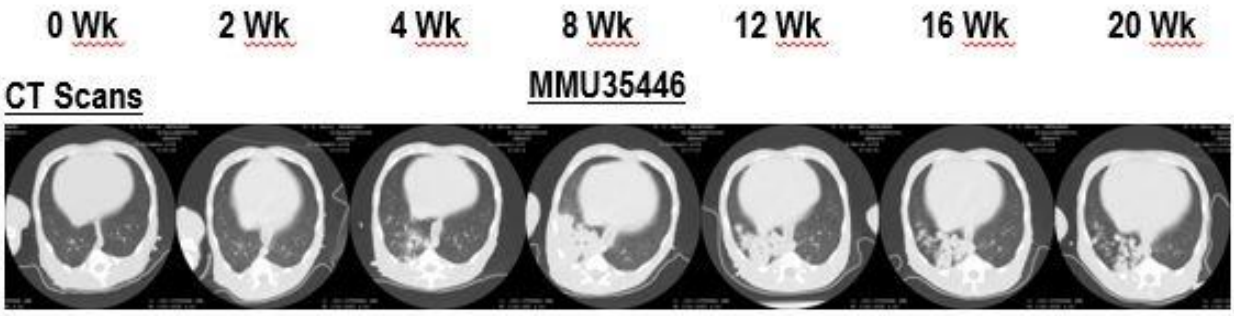

Figure 12: Images after alignment at different time points for animal MMU35446.

Table 1: Right lung CT scoring by radiologist, SZ.

The right lung score for each animal (at each respective time point), out of a total score of 100 given by the radiologist, SS.

Week	MMU35414	MMU35446	MMU35603	MMU35710	MMU36365	MMU36727
0	0	0	0	0	10	0
2	0	0	0	0	30	5
4	30	40	25	20	30	20
8	40	50	30	25	60	30
12	50	60	35	35	65	70
16	60	65	35	20		
20	70	65	55	25		

Table 2: Left lung CT scoring by radiologist, SZ.

The left lung score for each animal (at each respective time point), out of a total score of 100 given by the radiologist, SS.

Week	MMU35414	MMU35446	MMU35603	MMU35710	MMU36365	MMU36727
0	0	0	0	0	0	0
2	0	0	0	0	0	0
4	0	0	0	0	10	0
8	0	10	0	0	15	5
12	10	15	10	20	15	60
16	20	20	15	0		
20	30	25	15	0		

Table 3: Right lung CT scoring by radiologist, TG.

The right lung score for each animal (at each respective time point), out of a total score of 100 given by the radiologist, TG.

Week	MMU35414	MMU35446	MMU35603	MMU35710	MMU36365	MMU36727
0	0	0	0	0	3	0
2	0	0	0	0	5	7
4	25	20	20	10	30	40
8	40	35	25	12	60	43
12	40	45	30	25	70	65
16	45	55	35	20		
20	30	65	40	20		

Table 4: Left lung CT scoring by radiologist, TG.

The left lung score for each animal (at each respective time point), out of a total score of 100 given by the radiologist, TG.

Week	MMU35414	MMU35446	MMU35603	MMU35710	MMU36365	MMU36727
0	0	0	0	0	0	0
2	0	0	0	0	5	5
4	0	0	5	0	7	0
8	0	2	5	0	5	10
12	0	5	5	5	5	20
16	10	10	5	0		
20	15	15	5	0		

Table 5: Training strategy for whole (W), right (R), and left (L) lung imaging analysis.

Training strategy for the whole (W), right (R), and left (L) lung image analysis. The × sign indicates that the respective training primate was not included in the training and that was a testing animal. The ✓ indicates that the respective primate was used in the training.

Models	TP1	TP2	TP3	TP4	TP5	TP6	Testing Animal
M1-35414-W	×	✓	✓	✓	✓	✓	MMU35414
M2-35414-R	×	✓	✓	✓	✓	✓	MMU35414
M3-35414-L	×	✓	✓	✓	✓	✓	MMU35414
M1-35446-W	✓	×	✓	✓	✓	✓	MMU35446
M2-35446-R	✓	×	✓	✓	✓	✓	MMU35446
M3-35446-L	✓	×	✓	✓	✓	✓	MMU35446
M1-35603-W	✓	✓	×	✓	✓	✓	MMU35603
M2-35603-R	✓	✓	×	✓	✓	✓	MMU35603
M3-35603-L	✓	✓	×	✓	✓	✓	MMU35603
M1-35710-W	✓	✓	✓	×	✓	✓	MMU35710
M2-35710-R	✓	✓	✓	×	✓	✓	MMU35710
M3-35710-L	✓	✓	✓	×	✓	✓	MMU35710
M1-36365-W	✓	✓	✓	✓	×	✓	MMU36365
M2-36365-R	✓	✓	✓	✓	×	✓	MMU36365
M3-36365-L	✓	✓	✓	✓	×	✓	MMU36365
M1-36727-W	✓	✓	✓	✓	✓	×	MMU36727
M2-36727-R	✓	✓	✓	✓	✓	×	MMU36727
M3-36727-L	✓	✓	✓	✓	✓	×	MMU36727

1

[1] TP: training primate; TP1: MMU 35414; TP2: MMU 35446; TP3: MMU 35603; TP4: MMU 35710; TP5: MMU 36365; TP6: MMU 36727; M: model

Table 6: Right lung ML imaging analysis.

Right lung image analysis i.e., the number of positive predicted optical slices (average of three experiments) by the ML model at each time point for each animal and the prediction (NTB or TB) made based on majority voting. The optical slices predicted to be positive for TB were used to calculate the ML ratio, as described in materials and methods.

MMU#	WK 0			WK 2			WK 4			WK 8			WK 12			WK 16			WK 20		
	TN	FP	P	TN	FP	P	TP	FN	P	TP	FN	P	TP	FN	P	TP	FN	P	TP	FN	P
35414	391	350	NTB	NI			292	149	TB	409	32	TB	400	41	TB	338	103	TB	398	43	TB
35446	336	105	NTB	123	318	TB	407	34	TB	369	72	TB	419	22	TB	384	57	TB	391	50	TB
35603	249	192	NTB	171	270	TB	384	57	TB	329	112	TB	368	73	TB	414	27	TB	422	19	TB
35710	316	125	NTB	224	217	NTB	283	158	TB	340	101	TB	360	81	TB	308	133	TB	249	192	TB
36365	NI			NI			278	163	TB	404	37	TB	418	23	TB	NX			NX		
36727	99	342	TB	*376	*65	TB	402	39	TB	390	51	TB	420	21	TB	NX			NX		

2

[2] TN: True Negative; FP: False Positive; P: Prediction; TP: True Positive; FN: False Negative; NTB: No Tuberculosis; TB: Tuberculosis; NI: Not Included; NX: Necropsy

*376 and *65 for animal MMU 36727 at WK 2 was assigned positive score by radiologists, therefore they are not TN and FP, they are TP and FN respectively.

Table 7: Left lung ML imaging analysis.

Left lung image analysis i.e., the number of positive predicted optical slices (average of three experiments) by the ML model at each time point for each animal and the prediction (NTB or TB) made based on majority voting. The optical slices predicted to be positive for TB were used to calculate the ML ratio, as described in materials and methods.

MMU#	WK 0			WK 2			WK 4			WK 8			WK 12			WK 16			WK 20		
	TN	FP	P	TN	FP	P	TN	FP	P	TN	FP	P	TP	FN	P	TP	FN	P	TP	FN	P
35414	272	169	NTB	NI			357	84	NTB	264	177	NTB	*83	*358	NTB	223	218	TB	222	219	TB
35446	295	146	NTB	258	183	NTB	353	88	NTB	355	86	NTB	207	234	NTB	257	184	TB	167	274	NTB
35603	245	196	NTB	260	181	NTB	253	188	NTB	150	291	TB	223	213	TB	222	219	TB	174	267	NTB
35710	365	76	NTB	306	135	NTB	291	150	NTB	238	203	NTB	179	262	NTB	*165	*276	NTB	*251	*190	NTB
36365	NI			NI			*133	*308	NTB	*187	*254	NTB	366	75	TB	NX			NX		
36727	NI			143	298	TB	*220	*221	TB	*266	*175	TB	384	57	TB	NX			NX		

3

[3] TN: True Negative; FP: False Positive; P: Prediction; TP: True Positive; FN: False Negative; NTB: No Tuberculosis; TB: Tuberculosis; NI: Not Included; NX: Necropsy

*133, *308, *187 and *254 for animal MMU 36365 and *220, *221, *266 and *175 for animal MMU 36727 at WK 4 and WK 8 were assigned a positive score by radiologists, therefore they are not TN and FN, they are TP and FN.

*83 and *358 for animal MMU 35414 at WK 12 were assigned a zero score by radiologists, therefore they are not TP and FN, they are TN and FP respectively.

*165, *276, *251 and *190 for animal MMU 35710 at WK 16 and WK 20 were assigned a zero score by radiologists, therefore they are not TP and FN, they are TN and FP.