

Phenotyping Brain Changes in HIV and non-HIV Patients with Tuberculosis Meningitis

Anonymous

Abstract—Tuberculous Meningitis (TBM) is a severe form of tuberculosis that leads to the inflammation of the brain's protective membranes. The prognosis and diagnosis of TBM poses significant challenges due to its varied and often non-specific manifestations. This complexity is heightened in patients with specific pre-conditions, such as HIV, making disease characterization even more intricate. The project's main objective is to identify relevant features that can characterize TBM severity and its phenotypes, especially in patients previously diagnosed with HIV. For that purpose, a multi-modal model was designed, which leverages both imaging data, such as MRI imaging features extracted from T1w images, and non-imaging clinical data. This model aims not only to distinguish between the various lesion types in TBM, but also to correlate imaging features to the TBM grade, indicating disease severity. The results demonstrate no significant difference in TBM grades between HIV+ and HIV- groups. Further, a random forest model employing 58 selected features achieved an R^2 score of .65 in characterizing disease severity. While the model's success in predicting TBM grades reached up to 85% accuracy, its ability to predict lesion types was less substantial (38%). A weak association was identified between lesion types and TBM grade with a Cramér's V value of $V = .24$. The findings indicate the importance of both imaging and clinical data in predicting TBM severity, and highlight the potential complexities in distinguishing specific lesion types within the context of TBM and HIV co-infection. Identified limitations in label distribution and inconsistencies in evaluation inform future research directions, including the need for clinical labels and a more thorough investigation of imaging features.

Index Terms—Tuberculous Meningitis, HIV, MRI Imaging, Multimodal Model, Lesion Types, Disease Severity, Random Forest

I. INTRODUCTION

A. Background on Tuberculosis Meningitis

Tuberculosis Meningitis (TBM) is recognized as the most critical manifestation of extrapulmonary tuberculosis, characterized by a high global mortality rate of 48% in 2019 [1]. Therefore, despite its rarity, accounting for merely 1% of all tuberculosis cases, the severity of this disease underscores its global significance [2]. Several diagnostic tests, such as Xpert MTB/RIF Ultra and metagenomic next-generation sequencing (mNGS), have been employed to facilitate early diagnosis of TBM [3], [4]. However, the diagnosis of TBM remains challenging due to the non-specificity of clinical symptoms and the low sensitivity of diagnostic tests such as Polymerase Chain Reaction (PCR) [5]. Symptomatically, TBM manifests through fever, headache, nausea, vomiting, and altered mental status, but these signs can be varied and non-specific [6]. In fact, a study in India found the median time to TBM diagnosis to be around 21 days, indicating a substantial delay in detection and intervention [4]. The prognosis in TBM is also challenging due to the heterogeneous nature of the disease and individual variability in patient response to treatment. While the outcome can vary widely among individuals, there is an overall trend towards poor prognosis, particularly in late-diagnosed cases [7]. Factors contributing to this difficulty in predicting outcomes include varying host immune response, strain virulence, and the timing and effectiveness of therapeutic interventions [8].

B. Advancements in TBM research

Advancement in TBM research and treatment have been made in the use of imaging technology and the discovery of biomarkers, thus allowing a better understanding of TBM, particularly in relation to its severity and disease progression [9], [10]. For instance, MRI (Magnetic Resonance Imaging) has been integrated into TBM research and is used to evaluate features such as meningeal enhancement, tuberculomas, infarction, hydrocephalus, and abscesses [9]. The features provided by MRI include significant associations between an increase in acute cerebral infarctions (ACI), old cerebral infarctions (OCI), and Evan's ratio with poorer prognosis in TBM patients [9]. Additionally, the role of clinical biomarkers, specifically those detected in cerebrospinal fluid (CSF), has emerged as a valuable asset in diagnosing and prognosticating TBM, where inflammatory molecules present in the CSF hold diagnostic value when measured as profiles [10].

C. Impact of HIV on TBM

TBM is clinically associated with Human Immunodeficiency Virus (HIV) infections. A strong correlation exists between the prevalence of HIV and TBM, with some geographical regions (i.e., Smolensk and Kaluga) reporting correlations as high as $r = .95$ and $r = .98$ [11]. Pormohammad *et al.* [12] using a pooled analysis of 20 studies, indicated a high prevalence of HIV infections of 38% among adult patients with TBM and 6% in children, as compared to the estimated 0.7% of the global average [13], further demonstrating a relatively high prevalence of HIV in patients with TBM.

The prevalence correlation between HIV and TBM is likely associated with the alteration of the pathogenesis of TBM by HIV. In brief, HIV alters the pathogenesis of extrapulmonary TB, including TBM, by affecting the phenotype and function of MTB-specific CD4+ T cells at the disease site, exhibiting changes such as a less differentiated phenotype and the presence of polyfunctional CD4+ T cells expressing TNF, IL-2, and IFN- γ [14]. A study [15] conducted on HIV-1-infected patients with TBM in Durban, South Africa, revealed that TB co-infection within the Central Nervous System (CNS) is associated with enhanced localized HIV-1 viral replication. Specifically, TBM patients had a significantly higher median CSF viral load. Elevated soluble immunological markers were identified in the CNS of TBM participants, contributing to the increased risk of developing TBM at all stages of HIV-1 infection.

This distinct alteration in pathogenesis leads to worse prognosis for patients suffering from both HIV and TBM, leading to higher mortality rate [16], where HIV infection is a strong independent predictor of death from TBM. Accurate clinical prognosis is challenging in co-infected patients, as HIV also alters TBM's clinical features, creating marked differences between HIV+ and HIV- patients. Among HIV+ patients, cognitive dysfunction, unique changes in inflammatory response, and blood vessel pathology are common. These manifest in ways such as minimal meningeal enhancement in CT scans and conditions linked to poor prognosis [17]. Differences also exist in symptoms like temperature, seizures, and hydrocephalus [18]. A Vietnam study found higher antituberculous drug resistance rates in HIV+ patients, specifically 13% mono-resistance, 32.6%

poly-resistance, and 8.7% multidrug resistance, with a death rate of 67.2%, and factors such as increased TBM severity, low serum sodium, and decreased lymphocytes in the CSF linked to higher mortality [19].

Thus, the intertwining complexities between HIV and TBM, from pathogenesis to clinical characteristics, necessitate a multidimensional approach to research.

D. Biomarkers for Prognosis and Phenotyping

On the clinical side, research has explored predicting TBM outcomes using clinical parameters such as age, stage of meningitis, and raised intracranial pressure, with factors like age and stage associated with disease outcomes [20]–[23]. However, prognosis remains challenging due to unspecific clinical features, necessitating imaging data as discussed in Section I-B. Many studies using imaging data are limited to a single modality, possibly creating an information bottleneck [24]. In contrast, medical practice considers both images and clinical data like demographics, patient assessments and lab results [25].

Indeed, some more recent studies such as Modi *et al.* [26] have tried to delineate clinical and radiological data, but there is still no consensus exists on predictors of poor TBM outcome. Limitations include small sample size, biased human-crafted biomarkers, and insufficient analysis of imaging and non-imaging biomarkers of prognosis.

At the time of writing, a systematic study using multi-modal models for TBM prognosis, especially with HIV, is absent, highlighting a need for a comprehensive comparison between clinical and imaging data. Utilizing both can enhance understanding of TBM and potentially improve diagnosis and prognosis strategies.

E. Project Objectives

With this context, the project sets the following objectives:

- Identifying relevant features that characterize TBM and its phenotypes (lesion types) for both populations and given different data modalities
- Investigate the contributions of imaging and non-imaging data by developing a multi-modal model, and comparing its performance against single modal models.
- Assess the model's efficacy in predicting TBM severity, especially in relation to prior HIV diagnoses, and its potential clinical applicability.

II. METHODOLOGY

A. Data Collection and Preprocessing

1) Participants: The data used in the present study are from two prospective longitudinal studies conducted by the Oxford University Clinical Research Unit (OUCRU) [27], [28]. The trials were conducted across two hospitals in Ho Chi Minh City, Vietnam, and two hospitals in Jakarta, Indonesia. The inclusion criteria included:

- Aged 18+.
- Diagnosed with TBM. The diagnoses were made by senior physicians specializing in infectious diseases or tuberculosis and lung diseases, all of whom received additional diagnostic training and followed a standard operating procedure.
- Planned or initiated anti-tuberculosis chemotherapy.

The exclusion criteria include factors such as additional brain infections, prior treatment with drugs active against TBM, and lack of consent from the participant or family member. For further information on the participants inclusion, refer to [27].

This project is focused on the Vietnam segment data due to its relatively high prevalence rates of both TBM and HIV. The participants in the study were diagnosed with either or both of these diseases, providing a foundation for investigating the interrelations and co-occurrences of these two conditions. The initial dataset includes 302 participants (108 HIV+ and 194 HIV- patients), but some were excluded due to lack of imaging data and poor quality - imaging artefacts, resulting in 255 participants (89 HIV+ and 166 HIV- patients). The distribution of TBM grades within these groups is shown in Figure 1.

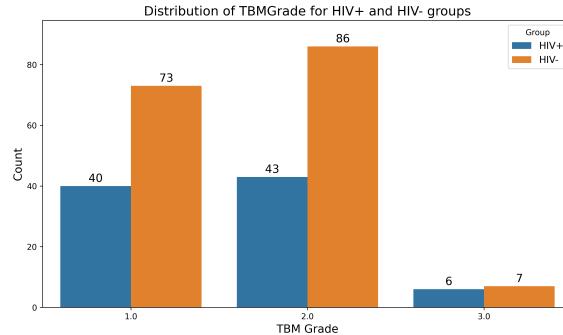


Fig. 1. Distribution of TBM grades among HIV-positive and HIV-negative groups. The Mann-Whitney U test revealed no statistically significant difference in TBM grades between the groups ($U = 7423$, $p = .94$, $r = .50$, $n = 255$).

Participants were partitioned into training, validation, and test groups at a ratio of 4:1:1 (i.e., 168, 44, 43 respectively), using Algorithm 1, with the K-S statistic below .05 to ensure comparable populations across sets. This ensured a similar distribution of demographic factors such as gender, age, HIV status, TBM grade, height, and weight, minimizing confounding factors and reducing bias risk in the analyses. The distribution of these variables between the groups is shown in Figure 8.

Algorithm 1 Iterative Algorithm for Ensuring Demographic Similarity Between Groups

- 1: **repeat**
- 2: Make a backup copy of the indices to allow recovery if the new arrangement does not improve the statistics.
- 3: Calculate the current best sum of the mean and maximum Kolmogorov-Smirnov statistics.
- 4: Select two random subsets of indices and choose a random batch size.
- 5: Exchange the selected number of samples between the two random subsets.
- 6: Recalculate the Kolmogorov-Smirnov statistics.
- 7: **if** the new statistics indicate a better arrangement **then**
- 8: Keep the new arrangement.
- 9: **else**
- 10: Restore the backup indices.
- 11: **end if**
- 12: **until** the maximum number of iterations is reached or the statistics are below the threshold

2) Severity Analysis between HIV and Non-HIV Populations:

A Shapiro-Wilk test was conducted to check for normality, revealing that the distribution of TBM grades, as a measure of disease severity, is not normally distributed ($p > .05$). Therefore a Mann-Whitney U test was utilized to compare the TBM grades as a measure of severity between the two groups (HIV- and HIV+).

3) Clinical Non-Image Data Preprocessing: In the OUCRU TBM dataset, patients diagnosed with TBM and HIV undergo clinical assessments and monitoring at 7 to 30 time intervals within follow-up periods of 24 months for HIV-negative participants and 12 months for HIV-positive participants [28]. Assessments include consciousness levels via GCS, HIV monitoring, liver function tests, lumbar puncture, and specific blood tests. Reviews are conducted inpatient and outpatient, with monitoring for adverse events and treatment effects like dexamethasone. The protocol also covers specialized studies on hyponatremia, intracranial pressure, and TBM pathophysiology, comprising 386 clinical variables.

These variables were curated in a two-fold process to retain only those with an established link to TBM prognosis. For example, gender was kept as a variable due to males having a 53% higher risk of TB and poorer treatment outcomes [29], [30]. In contrast, variables such as "prolonged coughing," associated with pulmonary tuberculosis, and cholesterol levels, not directly linked to TBM prognosis, were excluded.

After manual curation, 209 variables were identified as clinically relevant to TBM prognosis, especially with HIV co-infection. Variables with non-random missing values or those missing more than 50% of their values were eliminated to reduce bias and error that could skew data and result in inaccurate predictions [31]. The selection process is in Table I. After curation, the final dataset contained 70 clinical variables.

Step	Name of the Step	Variables Left
1	Initial variables	387
2	Manual curation:	209
3	Dropping variables without data file:	88
4	Dropping variables not in data:	84
5	Dropping systematically missing variables:	70
6	Dropping variables with < 50% data available	70

TABLE I

SELECTION STEPS SHOWING THE ORDER OF THE VARIABLE SELECTION PROCESS, WHICH INCLUDES MANUAL CURATION, CUTTING VARIABLES WITHOUT AVAILABLE DATA AND DROPPING SYSTEMATICALLY MISSING VARIABLES

4) Image Data: In the Donovan *et al.* [28], the image data comprised MRI images, chest X-ray images, and CT scans. The current project has specifically focused on the MRI scans in Vietnam. These scans were conducted at baseline (± 7 days), 60 days (± 7 days), and 12 months (-0/+1 month). The selection of MRI scans, specifically the T1-weighted scans, were chosen over other modalities such as CT due to their higher sensitivity to details and intricacies in brain structure. For the purposes of this analysis, one earliest image per patient was used, resulting in a total of 255 images, as mentioned in Section II-A.1.

5) Image Data Preprocessing: As a pre-processing step, the T1w scans were rigidly registered to the Montreal neurological institute (MNI) brain atlas [32] space using the *NiftyReg* open-source software [33]. During training, augmentations such as affine and elastic transformations, and Gaussian noise are applied using MONAI [34] to emulate data variability, enhancing model generalization [35]. Images are padded to size (193, 229, 193), the largest size in the dataset. The augmentation configurations are available at the Github repository ¹, and examples of the augmented images are in Figure 15 in Appendix I.

6) Image Data Labeling: The final image dataset consists of 255 scans. Each image undergoes a manual classification process, where they are visually assessed and labeled according to the lesion type by

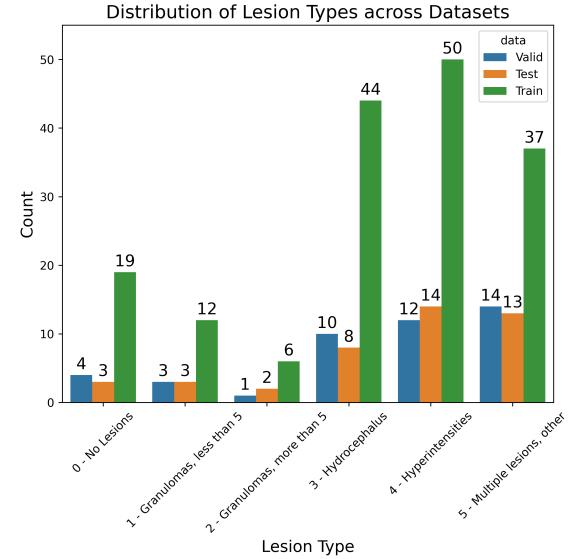


Fig. 2. Distribution of lesion types across train, validation, and test sets.

an engineering expert working on the OUCRU project. The lesions includes the following:

- 0 - No Significant Lesions
- 1 - Granulomas, less than 5
- 2 - Granulomas, more than 5
- 3 - Hydrocephalus
- 4 - Hyperintensities
- 5 - Multiple lesions, other

Figure 2 reveals imbalanced classes within the dataset, with categories like "Granulomas, more than 5" having only 6 training samples compared to the "Hyperintensities" class with 50. This imbalance could bias predictions and affect model performance on underrepresented classes. To address this, a weighted loss strategy and adjusted accuracy measures were used during training, detailed in Section II-C.2.

7) Association between Lesion Types and Disease Severity: The analysis of the correlation between lesion type and the disease severity, measured using TBM grade, was undertaken using a chi-square test. This statistical method was specifically chosen due to the nominal nature of the lesion type variable.

B. Feature Extraction and Selection

1) Non-Imaging Biomarkers: The assessment of clinical features linked to disease severity started with a handling of missing data with Multiple Imputation by Chained Equation (MICE) [36]. The process entails:

- **Analysis of Missing Data:** Most of the features in the dataset were complete, with a mean missing percentage of only 5.1%, while there is no systematically missing values as discussed in Section II-A.3. To visualize the pattern of missing values, a heatmap was generated, as depicted in Figure 12 in Appendix I. The dark blocks in the figure represent missing values, and as the figure illustrates, the absent values are relatively infrequent.
- **Imputation Process:** The missing values were imputed using the Multiple Imputation by Chained Equations (MICE) approach, employing a KNeighborsRegressor algorithm. The process was carried out over 1,000 iterations, prioritizing features with fewer missing values first.

¹<https://github.com/Don-Yin/TBM-HIV>

- Generation of Imputed Data:** A total of five different imputed datasets were generated, each starting from different random seed. The selection of five imputations was guided by Rubin's rule [37].

A random forest model was used as feature selection method, aiming at the identification of clinically relevant biomarkers among the initial 70 variable, as detailed in Section II-A.3. Random forest is model-free, suitable for numerical and ordinal data, and can accommodate multi-modal distributions that do not comply with Gaussian assumptions. This is particularly applicable to clinical data, where such relationships are common [38]. The random forest is used for identifying the clinical biomarkers and determine the optimal number of features to incorporate into a multi-modal model via feature importance metric. This approach can be summarized as in Algorithm 2.

Algorithm 2 Random Forest-based Clinical Feature Selection

- 1: Train random forest model on imputed dataset containing all 70 clinical features.
- 2: Rank features using importance scores from random forest, from most to least important.
- 3: **repeat**
- 4: Remove the least important feature in a step-wise manner.
- 5: Train a new random forest model with remaining features across all imputed datasets.
- 6: Calculate and record the average R^2 score for these features across datasets.
- 7: Construct an elbow plot with x-axis as number of remaining features and y-axis as change of average R^2 scores.
- 8: **if** optimal number of features is identified (elbow point) **then**
- 9: Proceed to next step.
- 10: **end if**
- 11: **until** optimal number of features is identified
- 12: Identify clinical biomarkers and determine the best number of features for multi-modal model.

C. Multi-Modal Model

1) Model Architecture: The study included the implementation of three distinct models as following:

- Model I - Only Image: When only image input is present, the architecture comprises a 3D DenseNet, which is connected directly to the fusion model.
- Model II - Only Clinical: When only the clinical input is present, the architecture employs a multilayer perceptron (MLP) connected directly to the fusion model.
- Model III - Both: When both image and clinical inputs are present, the architecture integrates the 3D DenseNet and the MLP, with their concatenated latent representations feeding into the fusion model.

The output of the fusion model represents the final prediction. During the training phase, these configurations were systematically tested to evaluate the role of imaging and clinical data in the prediction process. The architecture of the entire system is depicted in Figure 3.

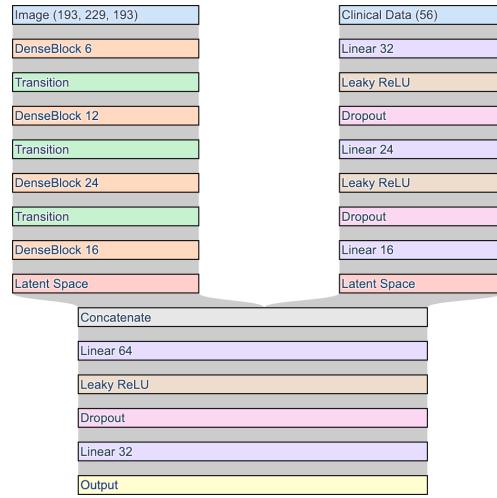


Fig. 3. The overall model architecture showing the 3D DenseNet image model, non-image multilayer perceptron, and the fusion multilayer perceptron

2) Model Training and Evaluation: The training of the proposed models has two objectives, both treated as multi-class classification problems with a cross-entropy loss with an Adam optimizer:

- Predicting the lesion type (6 classes)
- Predicting the TBM grade (disease severity; 3 classes)

Due to an imbalance in the images for the lesion types, as indicated in Figure 2, a weighted cross-entropy loss by the inverse of the class frequency is applied. Furthermore, weighted accuracy is utilized as the metric during training to mitigate the imbalance issue.

A grid search was conducted to identify the optimal hyperparameters for each model. The hyperparameters investigated include the learning rate, weight decay, batch size, and number of epochs, among others. The final set of hyperparameters that provided the best performance on the validation set are listed in Table IV. The training process was conducted on a set comprising 168 scans, utilizing 48 Nvidia v100 GPUs and lasted 8 hours.

An additional set of runs were executed with the objective of predicting the severity of the disease, represented by the TBM grade (ranging from 1 to 3, with higher values indicating increased severity). Since the lesion types are not clinical labels, their prediction does not fully validate the model's utility in a medical context. The severity of the disease, on the other hand, is determined by a clinician and is a clinical label. Therefore, the model's performance on severity prediction serves as an essential validation metric. It validates the model's capability to reliably extract relevant features from the images used in the current study and aids in establishing the reliability of the model.

Model performance was assessed using confusion matrices, with specific metrics such as F1 score and sensitivity, in addition to overall accuracy. This allowed for the examination of type I and type II errors for each individual class. The models' performance between the HIV and non-HIV groups was compared, contributing to understanding how the model's performance may vary depending on patients' HIV pre-condition.

D. Code Availability

All analyses in this study were conducted using Python (version 3.10.6). Essential packages and their respective versions used for the development, training, and evaluation of the proposed model are detailed as follows:

- `scipy` (version ~1.11.1) for various computations and statistical analyses
- `torch` (version ~2.0.1) as the deep learning framework
- `scikit-learn` (version 1.3.0) for imputation
- MONAI (version 1.2.0) for DenseNet construction and image augmentation

Further discussions regarding the specific functionalities or detailed implementation may be available upon request. Additional code and resources related to this study are at this Github repository ².

III. RESULTS

A. HIV- and HIV+ Groups Characterization

The Mann-Whitney U test was employed to assess the differences in severity between the two groups (HIV+ and HIV-) concerning TBM grade. The results obtained revealed no statistically significant difference in TBM grades between the groups ($U = 7423, p = .94, r = .50, n = 255$).

To further investigate the difference between HIV+ and HIV- groups, six separate Wilcoxon tests were performed for each lesion type. Among these comparisons, only one significant difference was observed, which pertained to label 4 (hyperintensities). For all other lesion types, the p-values exceeded .05, indicating no significant differences in the models' capacity in predicting the lesion type in HIV- and HIV+ scans.

B. Clinical Features Linked to Disease Severity

Using Algorithm 2, the optimal number of clinical features (and the features themselves) were assessed for inclusion in the model, as demonstrated in Figure 14 in Appendix I. Through the application of the elbow method, an optimal R^2 score was achieved with 58 features. The R^2 scores using these 58 selected features were found to be consistent across all imputed datasets, with a value of .65. This uniformity of scores across different imputations indicates a robust selection of features. These 58 features were then incorporated into the multi-modal model. A list of these features can be accessed at Table III in Appendix I.

C. Models Performance

1) Lesion Type: The baseline for predicting lesion type is 16.7%, corresponding to random guessing among 6 classes. The best performing model was able to achieve an accuracy of 38% on both the test set and the validation set, reflecting a advancement over chance prediction. For a complete distribution of all models' accuracy on the test set with lesion type as the label during hyperparameter search, refer to Figure 11 in Appendix I.

Interestingly, the models that used only clinical data outperformed those using only imaging data and those employing both clinical and imaging data, as depicted in Figure 11. A comparison of the models' accuracy across different input modes (clinical, imaging, both) when using lesion type as the label during training and model optimisation is provided in Figure 13 in Appendix I.

An exploration into the models' accuracy was conducted by comparing the predictions for HIV-positive and HIV-negative patients. The comparison across different lesion types reveals some intriguing patterns, as shown in Figure 4. The model demonstrates generally higher accuracy in predicting lesion types 3, 4, and 5. This tendency is likely attributed to these lesion types being more common within the dataset, as illustrated in Figure 2.

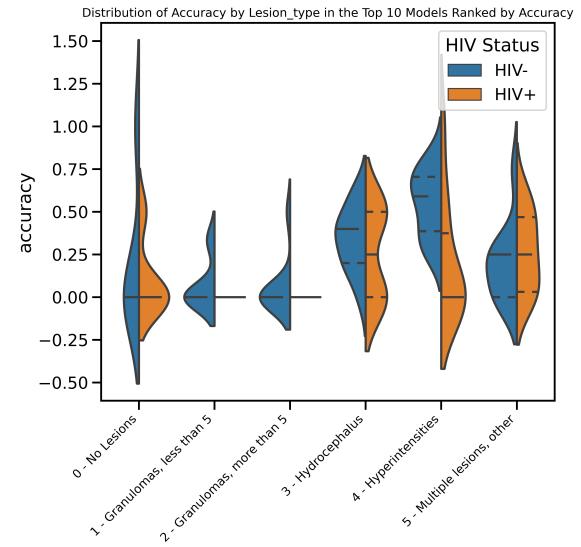


Fig. 4. Comparison of model's accuracy in predicting lesion types for HIV-positive and HIV-negative patients, based on the top model configurations in hyperparameter search. The accuracy values exceeding one in the violin plot are the result of smoothing applied by the kernel density estimation

The accuracy of 0 for the HIV-positive group in classifying label 1 and class 2 (the two types of granulomas) is due to the complete absence of images for these labels in the test set. This scarcity extends to the training and validation sets as well, accounting for the observed discrepancy. See Figure 2 for additional context.

A detailed examination of the best-performing model's ability to predict lesion types can be found in the confusion matrix shown in Figure 5. The figure shows the model predicts lesion types 4 and 5 with higher accuracy, but faces difficulty with lesion types 0, 1, and 2. Generally speaking, the model appears balanced on the rest of the classes, with the F1 scores for each class presented in Table II. The model's inability to predict lesion types 0, 1, and 2 (no lesion and granulomas) is attributable to the small number of samples in the training and validation sets and the absence of these labels in the test set.

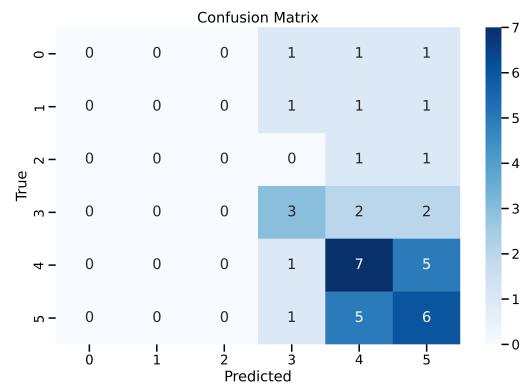


Fig. 5. Confusion matrix for the best performing model on predicting lesion types.

The class-wise sensitivity and F1-score are summarized in the following table:

2) Disease Severity: As oppose to when the lesion type is used as the labels, the models' performance on predicting TBM grades reveals substantial success. The accuracy notably exceeds the baseline

²<https://github.com/Don-Yin/TBM-HIV>

Class	Sensitivity	F1-score
Class 0	.0	.0
Class 1	.0	.0
Class 2	.0	.0
Class 3	.4	.5
Class 4	.64	.54
Class 5	.25	.17
Accuracy	.38	

TABLE II

SENSITIVITY AND F1-SCORES FOR THE BEST-PERFORMING MODEL BY CLASS ON THE TESTING SET

of 33.33% (random guessing among three classes for TBM), reaching up to 85%. Figure 6 shows the distribution of accuracies in HIV- and HIV+ groups.

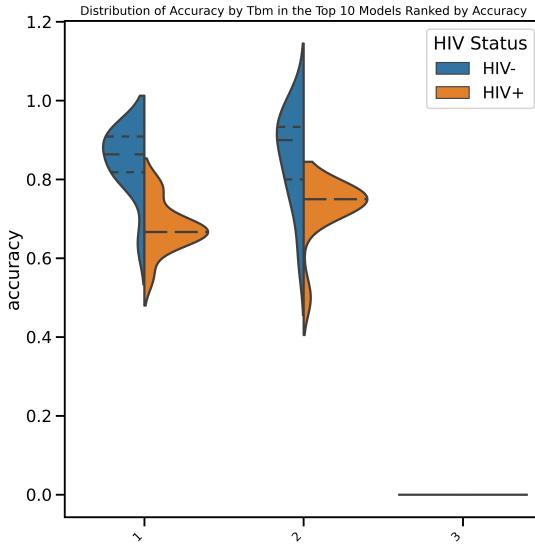


Fig. 6. Distribution of accuracies for predicting TBM grades on the test set. The max value exceeds 1 due to the kernel density smoothing.

It's to note that the test accuracy for TBM grade 3 is 0, due to the presence of only one image in the test set bearing this grade. The top-performing model misclassified this particular instance, which resulted in the null accuracy for this grade, as depicted in Figure 9 in Appendix I. This confusion matrix shows the model's capability in predicting each class. The model achieved a F1-scores of .86 for Class 1, .86 for Class 2, and 0 for Class 3.

In contrast to the models' performance when using lesion types as labels, a notable enhancement is observed when both types of data are employed for predicting TBM grades, as shown in Figure 10 in Appendix I. For TBM grade, both modalities seem to provide unique and complementary information, leading to improved accuracy.

D. Association between Lesion Types with TBM Severity

The association between lesion types and TBM severity was assessed using the chi-squared test, yielding a chi-squared statistic of $\chi^2 = 30.22$, a p-value of $p < .005$, and Cramér's V value of $V = .24$. In this analysis, a Cramér's V value of $V = .24$ suggests a weak association between lesion type and TBM grade.

The results are visualized in the contingency table depicted in Figure 7. The weak association can be further understood by examining the specific relationship between lesion types and TBM grades. For instance, lesion type 0 (indicating no lesion) is associated with

TBM grade 1, the least severe grade. This result is consistent with expectations, as the absence of a lesion would likely correspond to a lower severity of TBM.

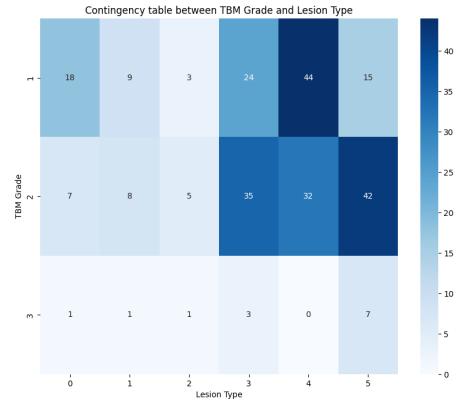


Fig. 7. Contingency table showing the association between lesion types and TBM severity.

IV. DISCUSSION

A. HIV Sub-group Characterization

This lack of difference in TBM grade severity between the HIV+ and HIV- groups contrasts with previous findings such as those reported in Diedrich *et al.* [39]. While the current study does not show differences between the groups. This does not conclusively indicate that no difference exists. Factors such as the sample size might have been too small to detect the differences.

B. Models Performance

The models' prediction accuracy for lesion types reaches a peak performance of 38%, compared to a random baseline of 16.66%. Models based on imaging data perform less effectively than those relying on clinical data, possibly due to a high degree of collinearity between imaging and clinical data. This strong interdependence may limit the ability to extract distinctive features, impacting prediction capabilities for lesion types.

This performance may imply either that imaging features and lesion types are indistinguishable, or that visual assessment and lesions labelling is inaccurate. Additionally, the alignment of labels for lesion types with visual features in the imaging data may be imperfect, leading to difficulties in prediction.

In the context of results in Section III-C.2, the model's proficiency in predicting TBM severity contrasts with its struggles in lesion type classification. This additionally suggests the issue may lie with that the image data for lesion type are unbalanced and single handedly labelled by an engineer, emphasizing the necessity for high-quality and well-balanced data to achieve robust performance.

Therefore, future work should focus on the development of augmentation strategies with the objective of increasing the training sample size for less prevalent labels, or the removal of labels that do not correspond to clinical experience. However, the situation regarding the current lesion type prediction task is more complex. The relatively lower accuracy in the lesion type prediction compared to TBM (with better labels) indicates that the lesion types might be mislabeled or not well-defined, thereby confusing the model. This not only underlines the need for obtaining clinically validated labels but also the utmost need for strong collaboration with medical experts

to ensure that the labels truly correspond to the observable features in the images. It may also require a more thorough examination of the underlying patterns and correlations within the imaging data to discern whether the model's struggles stem from indistinct visual features or from inconsistencies in labelling.

C. Influence of HIV Pre-condition

As discussed in Section III-C.1, the unique significance found in predicting hyperintensities in the HIV group could be interpreted in several ways. The presence of the HIV pre-condition might make hyperintensities more difficult to predict, leading to a more clinically complicated scenario. This could be aligned with the findings that prognosis being generally worse in co-infected patients and the potential for altered clinical features [40].

Additionally, the absence of between-group differences in other lesion types might be attributed to the small sample size (e.g., 6 images for Granulomas more than 5 in the training set), hindering the model's ability to identify other lesion types. Both possibilities highlight the need for further investigation with larger and more diverse datasets to gain a more accurate understanding of the impact of HIV pre-condition on lesion prediction.

Similarly, the models' performance on TBM grade prediction was observed to be lower in the HIV-positive groups. This observation may also be explained by the potential alteration of clinical features in individuals with the HIV precondition. Specifically, the HIV-positive status may introduce variability in the expression of TBM, complicating the task of accurate severity grading [40]. This phenomenon could be congruent with the known interactions between HIV and other pathologies, where the immunosuppressive nature of HIV may modify the disease manifestations and prognosis [41]. Further research with larger sample sizes and varied HIV-positive cohorts may clarify the underlying mechanisms and provide more robust predictive models.

D. Clinical and Imaging Biomarkers

In this study, the clinical features have been identified that characterize TBM and its phenotypes, such as lesion types, for both populations and across different data modalities. In addition, findings reveal that the clinical features have been more effective in predicting lesion types than the imaging data. However, it's important to note that the imaging features were not included in this particular investigation due to time constraints. This limitation presents an avenue for future exploration, where techniques such as grad-CAM, occlusion saliency maps, or guided backpropagation could be employed to analyze and interpret the imaging features.

E. Limitations and Future Directions

As discussed in Section IV-B, a limitation is recognised in the current study regarding the distribution of labels for various lesion types. Specifically, the test set contains only one label for TBM grade 3 and 3, 2 images for lesion types 0, 1, and 2 respectively. The imbalance of lesion types in both the training and testing sets could influence the training process and the models' performance in other populations. Future work should focus on the model assessment on datasets with more balanced labels. Additional efforts to train the models on datasets with higher prevalence of rare lesions should also be conducted.

Notably, the imbalance of classes for both lesion types in the current project are significant. However, the accuracy on the test dataset was evaluated using overall accuracy rather than balanced accuracy. Again, this issue was identified at the summarizing stage

of the project and was not addressed due to time constraints, thus remaining a direction for future work.

An additional limitation in this study is at testing of models, where the clinical data for each subject is randomly selected from the 5 imputed datasets for each scan, even though the subjects are consistent in the test data. This approach is not optimal, as when comparing different models, the divergence of data from various imputed datasets may lead to inconsistencies in the evaluation process. The method to address this issue involves ensuring that for each image in the test set, the clinical data is derived from the same imputation for all models under consideration. The solution is straightforward as implementing a fixed random seed for each image during testing. The reason this limitation was not addressed in the project was due to time constraints and the discovery of this issue only at the summarizing stage of the project.

Additional future directions include refining the approach in assessing the distribution of the accuracy, as presented in Figure 11 in Appendix I. Instead of the current method, which involves sampling from different configurations of the model during optimization, future work should focus on employing bootstrap techniques, where sampling from different populations using the same model with optimized hyperparameters is performed. This approach will involve training the model using the best hyperparameters on various populations and subsequently testing it on different test populations to obtain a more representative bootstrap distribution of accuracy.

V. CONCLUSION

The study identified that clinical features were more effective in characterizing the lesion types of TBM, with an accuracy of 38% compared to a random baseline of 16.66%. In contrast, for the characterization of TBM severity, the combination of imaging and clinical features was identified as the best method, exceeding the performance achieved using clinical features alone.

The model's interpretability of imaging features was not examined in current investigation due to time constraints. The implications of these findings may be attributed to the lack of clinical labels, small sample size, or high collinearity between imaging and clinical features, or a combination of these factors. All these require further investigation.

When developing a multi-modal model to investigate the contributions of imaging and non-imaging data, the findings revealed that the combined model's efficacy was more pronounced for HIV-groups compared to HIV+ groups in predicting TBM severity. This observation could be associated with the smaller sample size or increased clinical complexity within the HIV+ group, which requires further exploration.

Overall, the current study promoted a better understanding of TBM severity and disease phenotyping. While the investigation contributes valuable insights, limitations persist with sample size and the lack of inclusion of imaging data. These constraints emphasize areas for further rigorous investigation in other populations with different demography and TBM brain lesions prevalence. As such, the main conclusions are not yet ready to be deployed to clinics and may serve as a preliminary guide for future research.

REFERENCES

- [1] P. J. Dodd, M. Osman, F. V. Cresswell, A. M. Stadelman, N. H. Lan, N. T. T. Thuong, M. Muzyamba, L. Glaser, S. S. Dlamini, and J. A. Seddon, "The global burden of tuberculous meningitis in adults: A modelling study," *PLOS Global Public Health*, vol. 1, p. e0000069, Dec. 2021.

- [2] L. Charlie, S. Abay, A. Tesfaye, R. Mlera, S. Mwango, and M. Goretti, "Safety and efficacy of high-dose rifampicin in the management of tuberculosis meningitis: Systematic review and meta-analysis," *International Journal of Mycobacteriology*, vol. 10, no. 3, p. 312, 2021.
- [3] Y. Shen, G. Yu, W. Zhao, and Y. Lang, "Efficacy of xpert MTB/RIF ultra in diagnosing tuberculosis meningitis," *Medicine*, vol. 100, p. e26778, July 2021.
- [4] G. Yu, W. Zhao, Y. Shen, P. Zhu, and H. Zheng, "Metagenomic next generation sequencing for the diagnosis of tuberculosis meningitis: A systematic review and meta-analysis," *PLOS ONE*, vol. 15, p. e0243161, Dec. 2020.
- [5] V. M. S. Brienze, Â. P. Tonon, F. J. T. Pereira, E. Liso, W. A. Tognola, M. A. A. dos Santos, and M. U. Ferreira, "Low sensitivity of polymerase chain reaction for diagnosis of tuberculous meningitis in southeastern brazil," *Revista da Sociedade Brasileira de Medicina Tropical*, vol. 34, pp. 389–393, Aug. 2001.
- [6] N. F. Maranchick, M. H. Alshaer, A. G. C. Smith, T. Avaliani, M. Gujabidze, T. Bakuradze, S. Sabanadze, Z. Avaliani, M. Kipiani, C. A. Peloquin, and R. R. Kempker, "Cerebrospinal fluid concentrations of fluoroquinolones and carbapenems in tuberculosis meningitis," *Frontiers in Pharmacology*, vol. 13, Dec. 2022.
- [7] T. Boyles, A. Stadelman, J. P. Ellis, F. V. Cresswell, V. Lutje, S. Wasserman, N. Tiffin, and R. Wilkinson, "The diagnosis of tuberculous meningitis in adults and adolescents: protocol for a systematic review and individual patient data meta-analysis to inform a multivariable prediction model," *Wellcome Open Research*, vol. 4, p. 19, Jan. 2021.
- [8] J.-J. Sheu, R.-Y. Yuan, and C.-C. Yang, "Predictors for outcome and treatment delay in patients with tuberculous meningitis," *The American Journal of the Medical Sciences*, vol. 338, pp. 134–139, Aug. 2009.
- [9] X. Cao, Q. Yang, X. Zhou, K. Lv, Z. Zhou, F. Sun, Q. Ruan, J. Zhang, L. Shao, and D. Geng, "Cerebral infarction and evan's ratio on MRI affect the severity and prognosis of tuberculosis meningitis patients," *Diagnostics*, vol. 12, p. 1264, May 2022.
- [10] C. E. Teunissen, U. Rohlwink, D. Pajkrt, and P. J. W. Naudé, "Biomarkers of tuberculous meningitis and pediatric human immunodeficiency virus on the african continent," *Frontiers in Neurology*, vol. 13, May 2022.
- [11] T. V. Myakisheva and I. S. Lapshina, "Tuberculosis meningitis in HIV-positive and HIV-negative patients," *Tuberculosis and Lung Diseases*, vol. 100, pp. 38–44, July 2022.
- [12] A. Pormohammad, M. J. Nasiri, S. M. Riahi, and F. Fallah, "Human immunodeficiency virus in patients with tuberculous meningitis: systematic review and meta-analysis," *Tropical Medicine & International Health*, vol. 23, pp. 589–595, May 2018.
- [13] "Global Tuberculosis Report 2022 — who.int." <https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022>. [Accessed 04-08-2023].
- [14] K. Matthews, M. Ntsekhe, F. Syed, T. Scriba, J. Russell, K. Tibazarwa, A. Deffur, W. Hanekom, B. M. Mayosi, R. J. Wilkinson, and K. A. Wilkinson, "HIV-1 infection alters cd4⁺ memory t-cell phenotype at the site of disease in extrapulmonary tuberculosis," *European Journal of Immunology*, vol. 42, pp. 147–157, Dec. 2011.
- [15] I. D. Seipone, R. Singh, V. B. Patel, A. Singh, M. L. Gordon, D. M. Muema, K. Dheda, and T. Ndung'u, "Tuberculous meningitis is associated with higher cerebrospinal HIV-1 viral loads compared to other HIV-1-associated menigitides," *PLOS ONE*, vol. 13, p. e0192060, Feb. 2018.
- [16] N. T. T. Thuong, D. Heemskerk, T. T. B. Tram, L. T. P. Thao, L. Ramakrishnan, V. T. N. Ha, N. D. Bang, T. T. H. Chau, N. H. Lan, M. Caws, S. J. Dunstan, N. V. V. Chau, M. Wolbers, N. T. H. Mai, and G. E. Thwaites, "Leukotriene a4 hydrolase genotype and HIV infection influence intracerebral inflammation and survival from tuberculous meningitis," *The Journal of Infectious Diseases*, vol. 215, pp. 1020–1028, Apr. 2017.
- [17] S. M. Katrak, P. K. Shembalkar, S. R. Bijwe, and L. D. Bhandarkar, "The clinical, radiological and pathological profile of tuberculous meningitis in patients with and without human immunodeficiency virus infection," *Journal of the Neurological Sciences*, vol. 181, pp. 118–126, Dec. 2000.
- [18] S. K. Bandyopadhyay, R. Bandyopadhyay, and A. Dutta, "Profile of tuberculous meningitis with or without HIV infection and the predictors of adverse outcome," *West Indian Med. J.*, vol. 58, pp. 589–592, Dec. 2009.
- [19] M. E. Torok, T. T. H. Chau, P. P. Mai, N. D. Phong, N. T. Dung, L. V. Chuong, S. J. Lee, M. Caws, M. D. de Jong, T. T. Hien, and J. J. Farrar, "Clinical and microbiological features of HIV-associated tuberculous meningitis in vietnamese adults," *PLoS ONE*, vol. 3, p. e1772, Mar. 2008.
- [20] I. Freiman and J. Geehuysen, "Evaluation of intrathecal therapy with streptomycin and hydrocortisone in tuberculous meningitis," *The Journal of Pediatrics*, vol. 76, pp. 895–901, June 1970.
- [21] H. Smith, R. Vollum, and H. Cairns, "TREATMENT OF TUBERCULOUS MENINGITIS WITH STREPTOMYCIN," *The Lancet*, vol. 251, pp. 627–636, Apr. 1948.
- [22] M. Humphries, R. Teoh, J. Lau, and M. Gabriel, "Factors of prognostic significance in chines children with tuberculous meningitis," *Tubercle*, vol. 71, pp. 161–168, Sept. 1990.
- [23] U. Misra, J. Kalita, M. Srivastava, and S. Mandal, "Prognosis of tuberculous meningitis: a multivariate analysis," *Journal of the Neurological Sciences*, vol. 137, pp. 57–61, Apr. 1996.
- [24] R. Abdelmalek, F. Kanoun, B. Kilani, H. Tiouiri, F. Zouiten, A. Ghoubantini, and T. B. Chaabane, "Tuberculous meningitis in adults: MRI contribution to the diagnosis in 29 patients," *International Journal of Infectious Diseases*, vol. 10, pp. 372–377, Sept. 2006.
- [25] Y. Huo, Y. Zhan, G. Liu, and H. Wu, "Tuberculosis meningitis: Early diagnosis and treatment with clinical analysis of 180 patients," *Radiology of Infectious Diseases*, vol. 6, pp. 21–25, Mar. 2019.
- [26] M. Modi, K. Sharma, S. Prabhakar, M. Goyal, A. Takkar, N. Sharma, A. Garg, S. Faisal, N. Khandelwal, P. Singh, J. Sachdeva, R. Shree, V. Rishi, and V. Lal, "Clinical and radiological predictors of outcome in tubercular meningitis: A prospective study of 209 patients," *Clinical Neurology and Neurosurgery*, vol. 161, pp. 29–34, Oct. 2017.
- [27] J. Donovan, N. H. Phu, N. T. H. Mai, L. T. Dung, D. Imran, E. Burhan, L. H. B. Ngoc, N. D. Bang, D. C. Giang, D. T. M. Ha, J. Day, L. T. P. Thao, N. T. Thuong, N. N. Vien, R. B. Geskus, M. Wolbers, R. L. Hamers, R. van Crevel, M. Nursaya, K. Maharani, T. T. Hien, K. Baird, N. H. Lan, E. Kestelyn, N. V. V. Chau, and G. E. Thwaites, "Adjunctive dexamethasone for the treatment of HIV-infected adults with tuberculous meningitis (ACT HIV): Study protocol for a randomised controlled trial," *Wellcome Open Research*, vol. 3, p. 31, June 2018.
- [28] J. Donovan, N. H. Phu, L. T. P. Thao, N. H. Lan, N. T. H. Mai, N. T. M. Trang, N. T. T. Hiep, T. B. Nhu, B. T. B. Hanh, V. T. P. Mai, N. D. Bang, D. C. Giang, D. T. M. Ha, J. Day, N. T. Thuong, N. N. Vien, R. B. Geskus, T. T. Hien, E. Kestelyn, M. Wolbers, N. V. V. Chau, and G. E. Thwaites, "Adjunctive dexamethasone for the treatment of HIV-uninfected adults with tuberculous meningitis stratified by leukotriene a4 hydrolase genotype (LAST ACT): Study protocol for a randomised double blind placebo controlled non-inferiority trial," *Wellcome Open Research*, vol. 3, p. 32, Mar. 2018.
- [29] M. Humayun, J. Chirenda, W. Ye, I. Mukeredzi, H. A. Mujuru, and Z. Yang, "Effect of gender on clinical presentation of tuberculosis (TB) and age-specific risk of TB, and TB-human immunodeficiency virus coinfection," *Open Forum Infectious Diseases*, vol. 9, Oct. 2022.
- [30] T. J. JARVIS, "Implications of gender for alcohol treatment research: a quantitative and qualitative review," *Addiction*, vol. 87, pp. 1249–1261, Sept. 1992.
- [31] R. Singhal and R. Rana, "Intricacy of missing data in clinical trials: Deterrence and management," *International Journal of Applied and Basic Medical Research*, vol. 4, no. 3, p. 2, 2014.
- [32] J. C. Mazziotta, A. W. Toga, A. Evans, P. Fox, and J. Lancaster, "A probabilistic atlas of the human brain: Theory and rationale for its development," *NeuroImage*, vol. 2, pp. 89–101, June 1995.
- [33] M. Modat, D. M. Cash, P. Daga, G. P. Winston, J. S. Duncan, and S. Ourselin, "Global image registration using a symmetric block-matching approach," *Journal of Medical Imaging*, vol. 1, p. 024003, Sept. 2014.
- [34] M. J. Cardoso, W. Li, R. Brown, N. Ma, E. Kerfoot, Y. Wang, B. Murray, A. Myronenko, C. Zhao, D. Yang, V. Nath, Y. He, Z. Xu, A. Hatamizadeh, A. Myronenko, W. Zhu, Y. Liu, M. Zheng, Y. Tang, I. Yang, M. Zephyr, B. Hashemian, S. Alle, M. Z. Darestani, C. Budd, M. Modat, T. Vercauteren, G. Wang, Y. Li, Y. Hu, Y. Fu, B. Gorman, H. Johnson, B. Genereaux, B. S. Erdal, V. Gupta, A. Diaz-Pinto, A. Dourson, L. Maier-Hein, P. F. Jaeger, M. Baumgartner, J. Kalpathy-Cramer, M. Flores, J. Kirby, L. A. D. Cooper, H. R. Roth, D. Xu, D. Bericat, R. Floca, S. K. Zhou, H. Shuaib, K. Farahani, K. H. Maier-Hein, S. Aylward, P. Dogra, S. Ourselin, and A. Feng, "Monai: An open-source framework for deep learning in healthcare," 2022.
- [35] R. Atienza, "Improving model generalization by agreement of learned representations from data augmentation," in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, IEEE, Jan. 2022.
- [36] M. J. Azur, E. A. Stuart, C. Frangakis, and P. J. Leaf, "Multiple imputation by chained equations: what is it and how does it work?,"

- International Journal of Methods in Psychiatric Research*, vol. 20, pp. 40–49, Feb. 2011.
- [37] M. Mera-Gaona, U. Neumann, R. Vargas-Canas, and D. M. López, “Evaluating the impact of multivariate imputation by MICE in feature selection,” *PLOS ONE*, vol. 16, p. e0254720, July 2021.
 - [38] S. Janitza, G. Tutz, and A.-L. Boulesteix, “Random forest for ordinal responses: Prediction and variable selection,” *Computational Statistics & Data Analysis*, vol. 96, pp. 57–73, Apr. 2016.
 - [39] C. R. Diedrich and J. L. Flynn, “How does hiv-1 exacerbate tuberculosis?,” *Infection and Immunity*, vol. 79, pp. 1407–1417, Apr. 2011.
 - [40] S. Marais, D. J. Pepper, B. J. Marais, and M. E. Török, “HIV-associated tuberculous meningitis – diagnostic and therapeutic challenges,” *Tuberculosis*, vol. 90, pp. 367–374, Nov. 2010.
 - [41] G. Zandman-Goddard and Y. Shoenfeld, “HIV and autoimmunity,” *Autoimmunity Reviews*, vol. 1, pp. 329–337, Dec. 2002.

APPENDIX I APPENDIX A

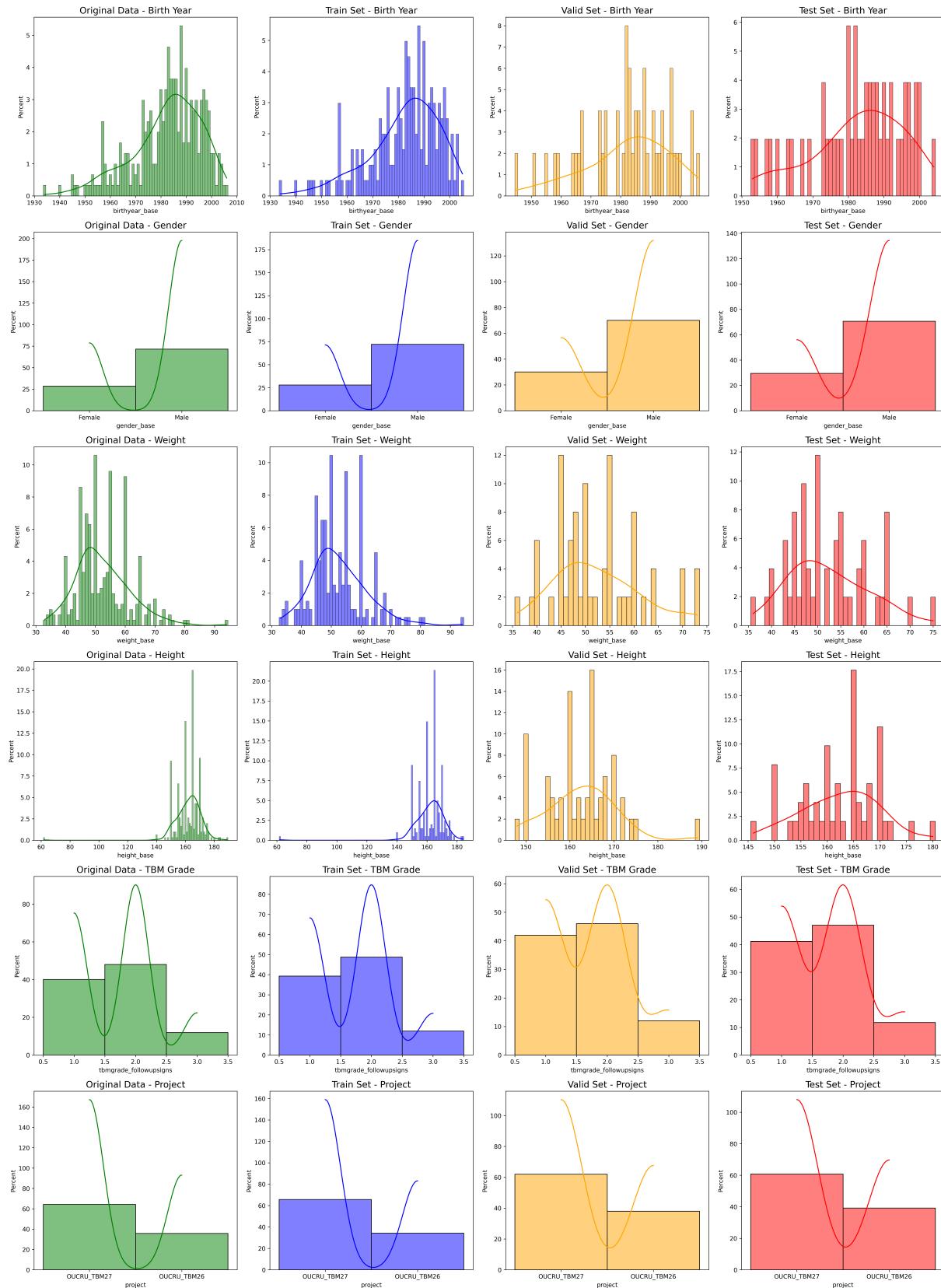


Fig. 8. Distribution of demographic variables across the training, validation, and test groups.

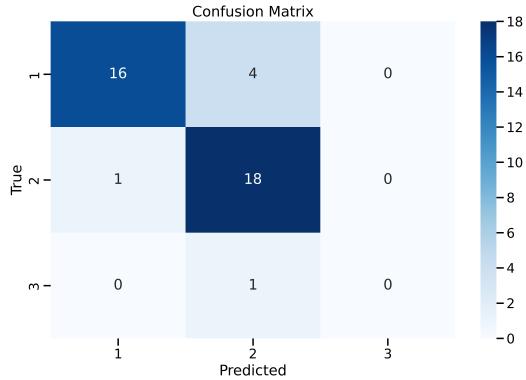


Fig. 9. Confusion matrix for the top-performing model predicting TBM grades.

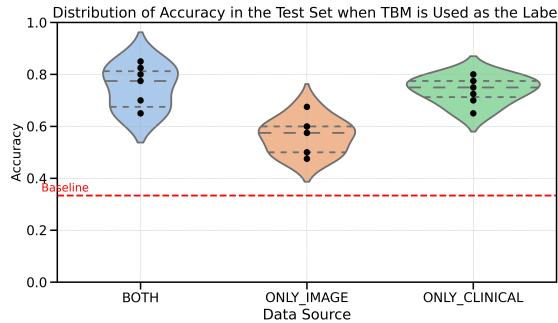


Fig. 10. Performance comparison showing the impact of using both types of data for predicting TBM grades

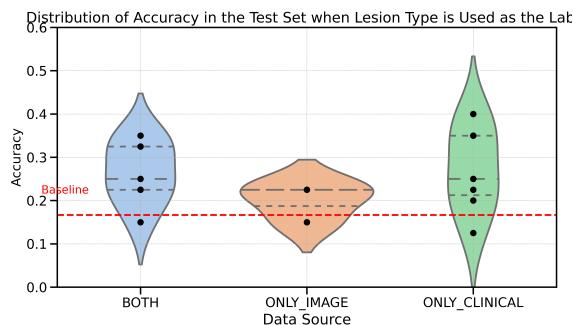


Fig. 11. Distribution of accuracy for different configurations of the model during hyperparameter search predicting lesion type, applied on the same test set.

TABLE III

THE SELECTED CLINICAL VARIABLES BY RANDOM FORESTS

Data Name	Question Title	Feature Importance
GCS_Total	GCS overall score	.1397
GCS_Verbal	GCS: Verbal score	.1226
ConsciousnessDe_has	Does the patient have decreased consciousness?	.0588
Neu_pct	Proportion of neutrophils	.0381
GCS_Motor	GCS: Motor score	.0354
Lym_pct	Proportion of lymphocytes	.0345
MRS	Level of symptoms	.0273
SysBP	Systolic blood pressure	.0237
Na	Blood Sodium	.0228
CSF_Bio_Lactate	CSF Lactate	.0204
CSF_Bio_CSF_Glu	CSF Glucose	.0195
CranialPalsy_Present	Does the patient have cranial nerves palsy?	.019
CSF_Cyto_WhiteCells	White blood cell count	.0188
CSF_Bio_Protein	CSF Protein	.0181
WBC	White blood cell count	.0176
BirthYear	Year of Birth	.0174
Creatinin	Creatinine	.0173
OP	CSF Opening Pressure	.0166
PLT	Platelet count	.0163
NeuroDay	Number of days of symptoms	.0162
CSF_Bio_BloodGlu	Paired blood Glucose	.016
Headaches_Days	How many days does the patient have headaches?	.0151
Fever_Days	How many days does the patient have fever?	.015
CSF_Cyto_Lym	Proportion of lymphocytes	.0148
CSF_Cyto_Neu	Proportion of neutrophils	.0146
Height	Height	.0144
Urea	Urea (or BUN):	.0144
FASTGLUC	Fasting blood glucose	.0144
Hb	Haemoglobin	.0142
Mono_pct	Proportion of monocytes	.0141
RequireHelp	Does the patient require help from anybody for everyday activities?	.0136
Paraplegia	Does the patient have paraplegia/paresis?	.0134
IsBaseline_RANGLUC	Random blood glucose	.0134
Neckstiffness_Days	How many days does the patient have neck-stiffness?	.0121
Temp	Highest temperature in last 24 hours:	.0121
Lethargy_has	Does the patient have lethargy?	.0099
Weight	Weight	.009
Appearance	CSF Appearance	.0077
Temp	Highest temperature in last 24 hours:	.0066
UrinaryRetention	Does the patient have urinary retention?	.0052
DisProblem	Has the illness left the patient with any other problems?	.0043
Vomiting_has	Does the patient have vomiting?	.004
Hemiplegia_Present	Does the patient have hemiplegia/paresis?	.004
CSF_Cyto_Eosinophil	Proportion of eosinophils	.0035
ReceivedTB_Status	Did the patient receive TB treatment in the past?	.0033
CranialPalsy_L6	L6	.0033
CranialPalsy_R6	R6	.0033
Irritability_has	Does the patient have irritability?	.0031
Gender	Gender at birth	.0026
Hct_IsDone	Was Haematocrit measured?	.0021
CranialPalsy_L7	L7	.002
Seizures_has	Does the patient have seizures?	.0019
BCG_Vaccine_Status	Was the patient vaccinated?	.0018
WeightLoss	Does the patient lose weight?	.0015
Neckstiffness_has	Does the patient have neckstiffness?	.0014
LIVEUNGTB_Contact	History of recent (within past year) close contact with an individual with pulmonary tuberculosis?	.0014

Hyperparameter	Value
densenet_spatial_dims	[3]
densenet_in_channels	[1]
densenet_out_channels	[128]
densenet_init_features	[64]
densenet_growth_rate	[32]
densenet_block_config	[(6, 12, 64, 48)]
densenet_bn_size	[4]
densenet_act	[Act.LEAKYRELU]
densenet_norm	["batch"]
densenet_dropout_prob	[0.15]
batch_size	[4]
learning_rate	[3e-4]
use_weighted_loss	[True]
augment	[True]
fusion_model_linear_layer_size	[[64, 32]]
use_label	["LESION_TYPE", "TBM"]

TABLE IV

HYPERPARAMETERS FOR THE BEST PERFORMING MODEL CONFIGURATION

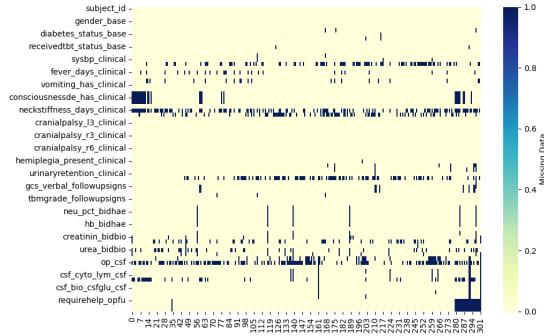


Fig. 12. Heatmap displaying the pattern of missing values. The dark symbols represent the locations where values are missing. The figure shows that the missing values are relatively scattered and infrequent.

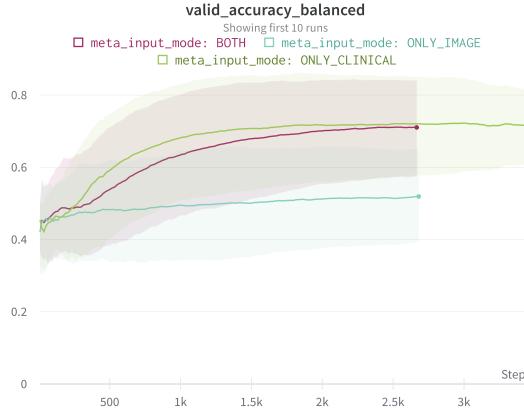


Fig. 13. Comparison of input modes in models' accuracy for lesion type prediction during training for model optimisation

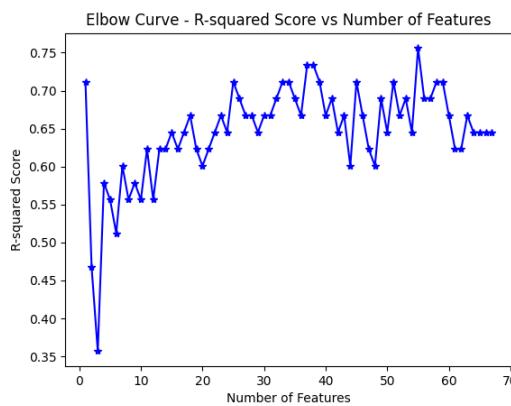


Fig. 14. Plot illustrating the relationship between the number of features and the R^2 score derived from the random forest model. Using the elbow method, 58 was identified as the optimal number of features.

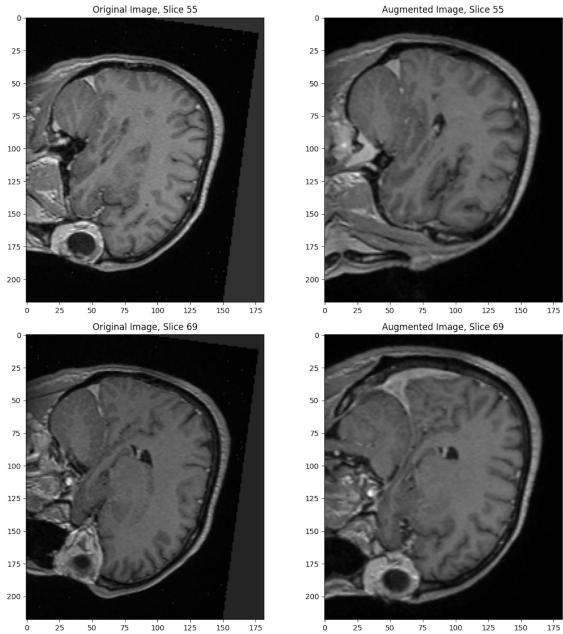


Fig. 15. Example slices of the MRI scan before (left) and after (right) augmentations. The augmentations include affine transformation, elastic transformation, and Gaussian noise. Augmentations are applied to increase the diversity of the dataset, improving the model's ability to generalize from the training data, thereby enhancing robustness and performance.