SpeChrOmics: A Biomarker Characterization Framework for Medical Hyperspectral Imaging

Ajibola S. Oladokun[1] [0000-0002-1627-5549], Bessie Malila[1], Victor M. Campello[2], Muki Shey[3], Tinashe E.M. Mutsvangwa[1,4]

[1] Division of Biomedical Engineering, University of Cape Town, Cape Town, South Africa
oldaji001@myuct.ac.za
[2] Artificial Intelligence in Medicine Lab (BCN-AIM), Universitat de Barcelona, Barcelona, Spain
[3] Department of Medicine, CIDRI-Africa, and Institute of Infectious Disease and Molecular Medicine, University of Cape Town, South Africa
[4] Data Science Department, IMT Atlantique, Brest, France

Abstract. We propose SpeChrOmics, a characterization framework for generating potential biomarkers of pathologies from hyperspectral images of body tissue. We test our model using a novel clinical application – hyperspectral imaging for the diagnosis of latent tuberculosis infection (LTBI). This is a neglected disease state predominantly prevalent in sub-Saharan Africa. Our model identified water, deoxyhemoglobin, and pheomelanin as potential chromophore biomarkers for LTBI with mean cross validation accuracy of 96%. Our framework can potentially be used for pathology characterization in other medical applications.

Keywords: Medical hyperspectral imaging, latent tuberculosis infection, biomarker characterization, tuberculin skin test indurations, chromophores.

1 Introduction

Medical optical imaging (MOI) involves use of the visible and near infrared (VNIR) range (380 to 2500 nm) of the electromagnetic spectrum to capture images of body tissue [1]. Red-Green-Blue (RGB) imaging and hyperspectral imaging (HSI) are two of the main sub modalities of MOI [2]. A medical hyperspectral (HS) image (also called a hypercube) is composed of two-dimensional (2D) spatial axes, and a spectral axis, with the imaging system capturing the spectrum of VNIR light reflected by body tissue and organizing it along the spectral axis [3]. Conversely, an RGB image possesses three channels of three broad color bands in place of a spectral axis. The spectral axis of HSI enables it to facilitate a richer characterization of body tissue than RGB images [4]. Chromophores are molecules found in body tissue and other substances which absorb and reflect selected wavelengths of light. They are responsible for the color appearance of body tissue [5]. Currently, the main method for estimating the concentration of chromophores in a sample of body tissue, as a means of pathology characterization in vivo or in vitro, is spectral unmixing using the modified Beer-Lambert model [6, 7]. This method facilitates the generation of 2D chromophore maps that can be used to

characterize body tissue. Most medical hyperspectral imaging (HSI) studies focus on either qualitative analysis, where they visually interpret chromophore maps, or on quantitative analysis that does not take into account the spatial distribution of these chromophores, to identify and characterize different pathologies. While effective for clinical applications where differences in disease states can easily be visualized in the chromophore maps, current methods may fall short in cases where these delineations are not immediately obvious to a human observer from the maps. Consequently, there is a need for an advanced framework that integrates robust spatial feature extraction with existing spectral feature extraction techniques. Such a framework would enhance the characterization of chromophore maps, particularly in clinical scenarios where a purely qualitative approach falls short. Combining radiomics, a technique for extracting spatial features from radiological images [8], with spectral unmixing could enhance the extraction of robust features that capture both chromophore information and tissue spatial details. While existing studies have yet to demonstrate this combination, our study introduces 'SpeChrOmics,' a novel framework that merges spectral unmixing with radiomics. It incorporates cross-validation and a classification mechanism to selectively identify features that most effectively characterize imaged pathologies.

Latent tuberculosis infection (LTBI) is a neglected disease state which is predominantly prevalent in sub-Saharan Africa and parts of south-east Asia [9]. The diagnosis of LTBI through the tuberculin skin test (TST) represents a clinical challenge, as it cannot be easily characterized using traditional qualitative analysis of optical images. [10-12]. This is due to the presence of erythema at the site of the test regardless of the LTBI status of the subject [13]. The TST is the most common test for LTBI in low- and middle- income countries (LMICs) [14]. TST is administered by injecting the tuberculin antigen into the dermis of a subject. An induration (lump) may form at the site of the injection 48-72 hours after the injection [14]. Diagnosis of LTBI is clinically obtained by comparing the manually measured diameter of an induration with consensus thresholds [15]. The Mantoux reading technique (manual pen and ruler measurement) currently used by clinicians is subjective as it is dependent on the reading technique of the clinician [16]. This introduces the need for a modality that addresses the subjectivity of TST indurations. Research by [10, 11] explored the possibility of using RGB imaging and 3D reconstruction to mitigate the subjectivity associated with Mantoux TST readings. The approach struggled with real TST indurations [12] as the subtle indurations lying beneath the skin surface were challenging to detect. Given its rich feature detection and subdermal imaging capabilities (up to a depth of 5 mm in skin), HSI emerges as a promising solution to reduce the subjectivity associated with TST interpretations. The interferon-gamma release assay (IGRA) is a more costly in vitro blood test for LTBI compared to TST [17], utilizing ESAT-6, CFP-10, and TB 7.7 antigens—components also found in tuberculin [18]. Theoretically, IGRA-related reactions could coincide with TST's induration-forming reactions at the injection site. Therefore, employing spectral unmixing to estimate chromophores in HS imaging could help distinguish between chromophores indicative of IGRA and those predictive of TST within the induration site. This positions HSI as a modality capable of bridging the diagnostic gap between TST and IGRA, which do not always yield concordant LTBI results for some subjects, as indicated in [19].

This study offers two significant contributions. First, we introduce a novel characterization framework for medical HSI, 'SpeChrOmics,' which merges spectral and spatial feature extraction techniques to generate validated, robust features. These features have the potential to serve as biomarkers, offering more precise discrimination between tissue sample disease states than traditional qualitative analysis. Second, we present a novel clinical application of HSI: the diagnosis of LTBI. This application serves as a practical use case to demonstrate the efficacy of the SpeChrOmics framework.

2 Methodology

A schematic of SpeChrOmics, as shown in figure 1, is composed of three blocks – spectral feature extraction, spatial feature extraction, and feature selection & ranking. The input data to SpeChrOmics is the reflectance hypercube $H_i(x, y, \lambda)$ captured from tissue sample i. The output data of our model is a set of ranked features F_{SFS}^{ranked} which could serve as biomarkers to characterize all the tissue samples $i = 1, 2, \dots, n$ based on the common pathology across the samples. The reflectance hypercube is composed of spatial axes x, y, and spectral axis λ which is defined in terms of the wavelength of light (in nm) reflected by a tissue sample. Adhering to the Beer-Lambert model, which underpins the spectral unmixing function $\mathcal{L}$, each sample's reflectance hypercube is converted into an absorbance hypercube $\widehat{H}_i(x, y, \lambda)$ [20] by taking the natural logarithm of its reciprocal. The spectral axis of $\widehat{H}_i(x, y, \lambda)$ describes the spectra of light absorbed across wavelengths of light by each pixel area of the sample. The goal of the spectral feature extraction block of our approach is to transform the continuous wavelength spectral axis of a hypercube to a discrete axis of chromophore maps based on the inherent molecular composition of captured tissue. This assumes that chromophore-based features are more indicative of specific pathologies than wavelength-based features, thereby offering superior diagnostic value.

A q-channel array of normalized chromophore maps $C_e^i = [C_1^i, C_2^i, \dots, C_q^i]$, for sample i, can be estimated from an absorbance hypercube using the Beer-Lambert model. We consider the Beer-Lambert model as a non-linear regression function g which estimates the relationship between independent spatial and spectral variables x, y and λ, and the chromophore map dependent variables $C_\varphi^i = [C_e^i, C_r^i] = [C_1^i, C_2^i, \dots, C_q^i, C_r^i]$, as shown in equation 1. Here, C_r^i is a 2D map of residual chromophores (which are distinct from those estimated in C_e^i) and unmodelled determinants of $\widehat{H}_i(x, y, \lambda)$. The number of chromophores, q, to be estimated from a sample is based on a priori knowledge of the chromophore composition of the sample. The chromophores can be preselected from a library of standard chromophores. The Beer-Lambert model can be considered as a sum of the products of the extinction coefficient $\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_q]$ and corresponding apparent estimated chromophore map C_e^i for each of the considered q chromophores [21], as shown in equation 2. The extinction coefficient of a chromophore is a unique constant measure per chromophore that defines the absorptivity of the chromophore at a given wavelength [7]. The non-linear regression function g also includes C_r^i as an additive element. The spectral unmixing function $\mathcal{L}$ transforms the per sample absorbance hypercube to a φ-channel array of normalized chromophore maps $C_\varphi^i(x, y)$ which

is composed of the estimated chromophore maps C_e^i and the residual chromophore map C_r^i. As shown in equation 3, we define the spectral unmixing function as a function that estimates the φ-channel chromophore map array that minimizes the squared error between the measured absorbance hypercube $\hat{H}_i(x, y, \lambda)$ and a modelled absorbance hypercube $H_i^{pred}(x, y, \lambda)$ generated by the Beer-Lambert model. Concatenating the chromophore map arrays, with φ channels, for n samples, along the discrete φ axis, results in I. This serves as the φ-channel chromophore composite of all samples. This completes the spectral feature extraction block of SpeChrOmics. The spectral feature extraction block of our model also performs the task of dimensionality reduction, as $\varphi < \lambda$ where the number of wavelength bands λ in a hypercube can be in the hundreds and the number of chromophore channels φ is typically less than 10. It is assumed that I encapsulates the most pertinent information from the original spectral axis across all samples.

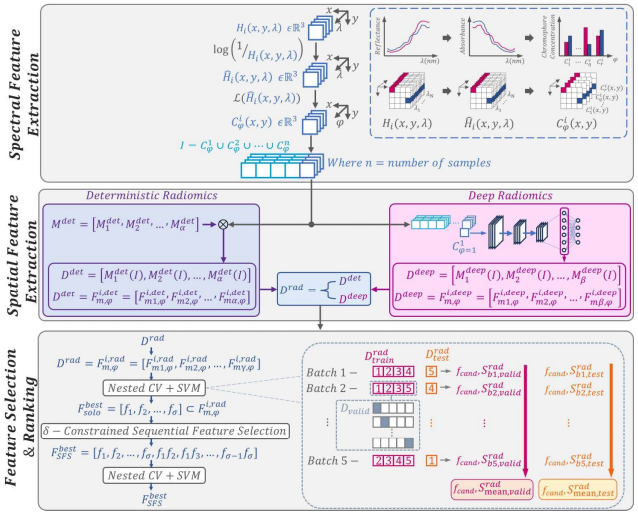

Fig. 1. A schematic of SpeChrOmics

$$\hat{H}_i^{pred}(x, y, \lambda) = g(x, y, \lambda, C_e^i) + C_r^i \tag{1}$$

$$\hat{H}_i^{pred}(x, y, \lambda) = \varepsilon_1(\lambda) \cdot C_1^i(x, y) + \varepsilon_2(\lambda) \cdot C_2^i(x, y) + \cdots + \varepsilon_q(\lambda) \cdot C_q^i(x, y) + C_r^i(x, y) \tag{2}$$

$$\mathcal{L}\left(\hat{H}_i(x, y, \lambda)\right) = \arg\min_{C_\varphi^i} \left(\hat{H}_i(x, y, \lambda) - \hat{H}_i^{pred}(x, y, \lambda)\right)^2 \tag{3}$$

The output I of the spectral feature extraction block feeds into the spatial feature extraction block of SpeChrOmics which has two sub blocks – the deterministic radiomics and deep radiomics spatial feature extraction blocks. The deterministic radiomics branch involves the application of a set of α radiomic models $M^{det} = [M_1^{det}, M_2^{det}, \dots, M_\alpha^{det}]$ from the work of [22] on the chromophore array I of all samples.

This process yields a set of deterministic radiomics features, D^{det} comprising attributes that encapsulate the spatial relationships among the chromophore pixels within each sample, as delineated in Equation 4. A constituent feature $F_{m1,\varphi=1}^{i,det}$ is defined as the output of applying radiomic model M_1^{det} on the first chromophore map in C_φ^i,the

chromophore array of sample i. In the deep radiomics branch of SpeChrOmics, the chromophore map for each chromophore array C_{φ}^{i} corresponding to each sample in I is inputted into a pretrained deep learning model. Each of the β nodes in the penultimate layer, immediately preceding the classifier layer of a pretrained model, can be represented as a set of β models $M^{deep} = [M_1^{deep}, M_2^{deep}, \dots, M_{\beta}^{deep}]$ applied to each chromophore map in I to produce a set of deep radiomics features D^{deep}, as shown in equation 5. In SpeChrOmics the deterministic and deep radiomics branches converge such that D^{rad} is the selector between D^{det} and D^{deep}. With this approach, our framework offers the option of two kinds of spatial features from chromophore maps. The spatial pixel relationships captured by the extracted features in D^{det} are more explainable than the extracted features in D^{deep}. On the other hand, the robustness of the feature extraction process of deep learning models suggests that the features in D^{deep} could better capture the spatial relationships in chromophore maps than the features in D^{det}.

$$D^{det} = F_{m,\varphi}^{i,det} = \left[M_1^{det}(I), M_2^{det}(I), \dots, M_{\alpha}^{det}(I)\right] = [F_{m1,\varphi}^{i,det}, F_{m2,\varphi}^{i,det}, \dots, F_{m\alpha,\varphi}^{i,det}] \quad (4)$$

$$D^{deep} = F_{m,\varphi}^{i,deep} = \left[M_1^{deep}(I), M_2^{deep}(I), \dots, M_{\beta}^{deep}(I)\right] = \left[F_{m1,\varphi}^{i,deep}, F_{m2,\varphi}^{i,deep}, \dots, F_{m\beta,\varphi}^{i,deep}\right] \quad (5)$$

$$F^{best} = (f_{cand}, S_b^{rad} | S_b^{rad} = CV(SVM(f_{cand}, y), k); S_b^{rad} > \tau, f_{cand} \epsilon F_{m,\varphi}^{i,rad}) \quad (6)$$

The array D^{rad} of extracted spatial features is fed to the feature selection & ranking block which utilizes nested k-fold cross validation and a support vector machine (SVM) classifier (for its soft margin ability to classify samples near a decision boundary) to produce a set of features (potential biomarkers) ranked based on mean validation and test scoring metrics $S_{mean,valid}^{rad}$ and $S_{mean,test}^{rad}$. Here, y is the set of target class labels per sample and F^{best} in equation 6 is a sorted set of candidate features and metric scores.

We generated a novel in-house HSI dataset of the TST indurations site for 37 human subjects of African descent. The Specim IQ HSI camera [23] and a broadband LED were used to acquire the 37 hypercubes. We generated an estimated RGB image from each hypercube by selecting the 750 nm, 560 nm, and 410 nm band images of each hypercube as the red, green, and blue channel images, respectively. These choices of wavelengths were empirically determined by Specim [23] and are an attempt to mimic the peak wavelengths of the color filters used in RGB cameras [24]. The Specim IQ possesses an RGB viewfinder camera, next to the HSI sensor, that captures RGB images of a scene concurrently with captured hypercubes. This results in a sample size of 37 hypercubes, estimated RGB images, and camera based RGB images of the TST skin tissue area of 37 subjects. Each hypercube, estimated RGB image, and camera based RGB image has a spatial dimension of 200 x 200 pixels around the TST injection site per subject. The hypercubes have a spectral dimension of 204 wavelength bands from 397 nm to 1003 nm.

Two sets of binary class labels y_{TST} and y_{IGRA}, based on clinically measured Mantoux TST and IGRA readings, respectively, for the 37 subjects, were utilized in the analysis and classification of the dataset in this study. Ten samples were TST positive based on clinically measured TST induration size $> 10\ mm$. Consequently, 27 samples were TST negative leading to a class imbalance in the dataset. Nine of the subjects were

IGRA positive and 28 were IGRA negative which also resulted in class imbalance. Thus, balanced accuracy was the metric of choice in this study to evaluate and rank candidate features. There was a 46% agreement between the TST and IGRA based LTBI diagnosis of the subjects. The tissue chromophores estimated in this study were oxyhemoglobin (OX), deoxyhemoglobin (DX), eumelanin (EM), pheomelanin (PM), bilirubin (BL), water (WT), and fat (FT). These are the most significant chromophores in the dermis region of the human skin [4], which is where a TST is administered. The extinction coefficients for these seven chromophores were obtained from [25, 26]. We applied our dataset of 37 hypercubes, estimated RGB images, and camera based RGB induration images to SpeChrOmics to generate potential TST and IGRA based biomarkers. The goal of this was to analyze the differences in efficacy of HSI and RGB imaging in the characterization LTBI using our proposed approach as the evaluation platform. The estimated and camera based RGB images were fed to only the spatial feature extraction and the feature selection & ranking blocks of SpeChrOmics as 3-channel versions of C_{φ}^{i}. Three deep learning models ResNet50, DenseNet121, and InceptionV3, which had been pretrained on the RadImageNet [27] dataset of 1.35 million medical images across several radiological modalities, was each used in this study to generate the deep radiomics features D^{deep}. Hyperspectral images are currently not included in RadImageNet nor any other public medical imaging dataset repository. The choice of these three pre-trained models assumes that these models have learned spatial features from other medical imaging modalities that could help them generate spatial features from the chromophore maps of TST indurations that accurately characterize the indurations.

3 Results

The camera based RGB images, estimated RGB images, estimated chromophore maps C_{OX}^{i}, C_{DX}^{i}, C_{EM}^{i}, C_{PM}^{i}, C_{BL}^{i}, C_{WT}^{i}, C_{FT}^{i}, and map of residuals C_{r}^{i} for some of the subjects are as shown in figure 2. For the hypercubes, the wavelength range of 450 – 1003 nm was utilized, rather than the full 397 – 1003 nm range, during the spectral unmixing stage as this was the widest wavelength range which minimized the regression error. This is due to the contrast between high peaks in the extinction coefficients of OX and DX from 397 – 450 nm, and the relatively low reflectance of the human skin in that range. The best ranked feature combinations and their scoring metrics after applying the 37 hypercubes, camera based RGB images, and estimated RGB images of indurations are as shown in table 1. These features are the potential optical imaging-based biomarkers for TST and IGRA from in vivo induration samples. The best ranked deterministic radiomics features natively possess explainability unlike deep radiomics features. The visualization of the deterministic radiomics features that best characterize TST indurations based on SpeChrOmics are as shown in figure 3. The skewness of WT and PM were the best features from $\boldsymbol{D^{det}}$ and their skewness contribution maps were generated.

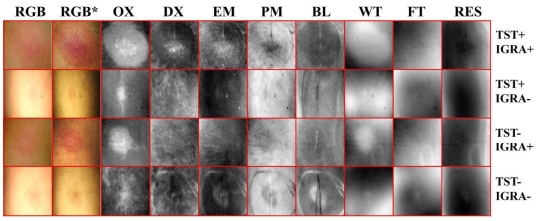

Fig. 2. Camera-based RGB images, estimated RGB images (RGB*), and chromophore maps of selected subjects with the corresponding true TST and IGRA classes.

Table 1. The scoring performance of the best features. Where BSF: best single feature, BFC: best feature combination, (MV, MT): mean cross validation, mean test balanced accuracy, (RCB, GCB, BCB): red, green, blue channel-based features, and (WTB, DXB, PMB, EMB, RSB): features based on WT, DX, PM, EM, and RES chromophores

		Deterministic Radiomics		ResNet50		InceptionV3		DenseNet121	
	Target label, *y*	**TST**	**IGRA**	**TST**	**IGRA**	**TST**	**IGRA**	**TST**	**IGRA**
RGB	**BSF**	RCB	RCB	RCB	GCB	RCB	RCB	RCB	RCB
	BSF (MV, MT)	67%, 77%	75%, 70%	65%, 73%	44%, 45%	77%, 83%	49%, 60%	90%, 91%	71%, 68%
	BFC	GCB, BCB	RCB, GCB	GCB, BCB	RCB, RCB	RCB, GCB	RCB, BCB	RCB, GCB	GCB, GCB
	BFC (MV, MT)	65%, 65%	78%, 68%	82%, 86%	91%, 90%	87%, 91%	75%, 60%	94%, 91%	79%, 68%
E-RGB	**BSF**	RCB	RCB	RCB	RCB	GCB	RCB	RCB	BCB
	BSF (MV, MT)	67%, 74%	73%, 82%	70%, 70%	59%, 64%	77%, 78%	72%, 83%	78%, 81%	82%, 83%
	BFC	RCB, GCB	RCB, RCB	RCB, RCB	RCB, RCB	RCB, GCB	RCB, GCB	RCB, BCB	GCB, GCB
	BFC (MV, MT)	67%, 83%	75%, 58%	79%, 84%	**89%, 95%**	88%, 93%	**88%, 98%**	96%, 95%	**94%, 97%**
HSI	**BSF**	WTB	PMB	WTB	PMB	WTB	PMB	DXB	DXB
	BSF (MV, MT)	71%, 81%	73%, 85%	82%, 80%	51%, 74%	79%, 81%	69%, 75%	82%, 83%	87%, 88%
	BFC	WTB, RSB	PMB, EMB	WTB, PMB	WTB, PMB	WTB, PMB	WTB, DXB	WTB, DXB	DXB, BLB
	BFC (MV, MT)	**80%, 80%**	**76%, 83%**	**85%, 90%**	78%, 78%	**92%, 94%**	83%, 87%	**96%, 100%**	86%, 100%

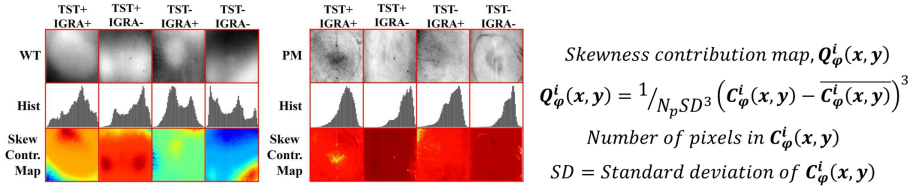

Fig. 3. Visualization of the best deterministic features for the TST and IGRA classes

4 Discussion

The camera based and estimated RGB images in figure 2 show that erythema is present at the site of the TST regardless of the resulting TST and IGRA diagnosis. The presence of erythema poses a challenge to visual TST readings which necessitated the use of tactual exploration by clinicians. The aim of SpeChrOmics in this study's novel clinical application is to extract subdermal features from TST indurations, that are typically only accessible through tactual exploration, for the Mantoux reading, and in vitro test, for the IGRA reading. Visually, the OX chromophore shows the highest contrast among the chromophores, likely due to the erythema. None of the chromophores exhibit visually observable features that easily discriminate between positive and negative TST or IGRA labels. This justifies the use of the spatial feature extraction block of SpeChrOmics for this clinical application. The performance of SpeChrOmics on the TST induration dataset is summarized in table 1, where BSF = best single feature, BFC = best

feature combination, MV = mean cross validation balanced accuracy, MT = mean test balanced accuracy, (RCB, GCB, BCB) = red, green, blue channel-based features respectively, and (WTB, DXB, PMB, EMB, RSB) = features based on WT, DX, PM, EM, and RES chromophores. The deep radiomics features generated better classification accuracies for all three groups of images than the deterministic radiomic features.

This suggests that deep learning models are more adept at extracting spatial features from TST indurations. The estimated RGB induration images (E-RGB) generated higher mean balanced accuracies across all deterministic and deep models than the camera based RGB induration images. This could be due to the demosaicking effect caused by the color filter array on the sensors of typical RGB cameras [24]. Due to demosaicking, some pixels may not display their true color, potentially affecting the accuracy of spatial feature extraction. Since HS cameras capture the true spectra of each pixel, every pixel in E-RGB contains true color information. The BFCs performed better than the BSFs across the dataset and models. The RCBs were the best performing group of features across all tests done on the RGB camera-based images and E-RGBs. This suggests that, of the three color-channel bands, the red color channel of an RGB image characterizes TST indurations the most. The order of performance of the selected deep learning models, in increasing order for both TST and IGRA labels, across all image types, is ResNet50, InceptionV3, and DenseNet121 [27]. This suggests that the DenseNet architecture, with a 96% MV, is adept at extracting LTBI-related features from optical images of TST indurations.

For the hypercube dataset, water-based features (skewness of WT for $\boldsymbol{D^{det}}$) seem to be most prominent among other chromophore features. Also, PM and DX based features (skewness of PM for $\boldsymbol{D^{det}}$) seem to contribute to the characterization TST indurations. The DX based features seem to be best at predicting IGRA diagnosis from TST indurations though IGRA tests are in vitro and involve a more specific reagent than the TST. The table columns indicate that water-based chromophores yield the best results for HS images, whereas features from the red color band are most effective for RGB images. This could be because the wavelength of peak absorption of water is closer to the red color band than other bands [4]. This appears to also be the case with DX and GCB as DX has a peak absorption wavelength in the green color band [28]. The deep radiomics features extracted from the E-RGBs outperformed the ones from the hypercubes for the IGRA classification and vice versa for the TST. This suggests that only a few wavelength bands could be required to predict IGRA diagnosis from TST indurations. The chromophore biomarkers WT, DX, and PM identified by SpeChrOmics for LTBI from the images of TST indurations suggests that less expensive HS cameras, which only capture at the peak wavelengths of these chromophores, can be used for this clinical application. Cytokines released by macrophages during the formation of TST indurations could promote localized vasodilation and vascular permeability [29] which may explain the accumulation of water and hemoglobin at the induration site. Also, the cytokines at induration site can activate local melanocytes [30] which could cause them to produce more melanin at the site. The limitations of this study include a small dataset size, the absence of an established evaluation framework, and the lack of a baseline comparison. These limitations stem from the neglected nature of LTBI as a disease condition and the under-representation of research in HSI-based biomarker

identification. This paper serves as a good starting point for further research in this field, and serves as a foundation for future validation of SpeChrOmics on larger and more diverse datasets.

5 Conclusion

This study presents a novel framework, SpeChrOmics, for extracting biomarkers from the hyperspectral images of body tissue with pathology. The framework was evaluated using a novel clinical application of HSI for LTBI, which is a neglected disease condition. Chromophore-based biomarkers were generated for aiding the diagnosis of LTBI in the TST and IGRA methods. The resulting features, with up to 96% validation accuracy, offer new quantitative image-based readings for TST and IGRA that are precise and repeatable. A limitation of this study is its small dataset size, which will be addressed in the next phase of the study. SpeChrOmics may potentially be used to characterize pathologies in other medical HSI applications such as skin cancer [31], endoscopy [3], cervical cancer [32], and wound treatment [33].

Acknowledgements. This document has been produced with the financial assistance of the European Union (EU) and co-funding from the Carnegie Foundation of New York (Grant no. DCI-PANAF/2020/420-028), through the African Research Initiative for Scientific Excellence (ARISE), pilot programme. ARISE is implemented by the African Academy of Sciences (AAS) with support from the European Commission and the African Union Commission (EUC). The statements made and views expressed are solely the responsibility of the authors. The research is also supported by the University of Cape Town International Scholarship, and the South African Research Chairs Initiative (SARChI) of the NRF and the Department of Science and Technology (grant no 98788).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Deng, L., Sun, J., Chen, Y., Lu, H., Duan, F., Zhu, L., Fan, T.: M2H-Net: A Reconstruction Method For Hyperspectral Remotely Sensed Imagery. ISPRS Journal of Photogrammetry and Remote Sensing **173** 323-348 (2021)
2. Karim, S., Qadir, A., Farooq, U., Shakir, M., Laghari, A.A.: Hyperspectral Imaging: A Review and Trends towards Medical Imaging. Curr Med Imaging **19**(5), 417-427 (2022)
3. Moulla, Y., Buchloh, D.C., Köhler, H., Rademacher, S., Denecke, T., Meyer, H.-J., Mehdorn, M., Lange, U.G., Sucher, R., Seehofer, D., Jansen-Winkeln, B., Gockel, I.: Hyperspectral Imaging (HSI)—A New Tool to Estimate the Perfusion of Upper Abdominal Organs during Pancreatoduodenectomy. Cancers **13**(11), 2846 (2021)
4. Vasefi, F., MacKinnon, N., Farkas, D.L.: Chapter 16 - Hyperspectral and Multispectral Imaging in Dermatology. In: Hamblin, M.R., Avci, P., Gupta, G.K. (eds.) Imaging in Dermatology, pp. 187-201. Academic Press, Boston (2016)

5. Shukla, R., Dubey, A., Pandey, V., Golhani, D., Jain, A.P.: Chromophore-an utility in uv spectrophotometer. Inventi Rapid: Pharm Ana & Qual Assur **2012**(3), (2012)
6. Kocsis, L., Herman, P., Eke, A.: The modified Beer-Lambert law revisited. Phys Med Biol **51**(5), N91-98 (2006)
7. Chauvin, J., Akhbardeh, A., Brunnemer, R., Vasefi, F., Bearman, G., Huong, A., Tavakolian, K.: Simulated Annealing-Based Wavelength Selection for Robust Tissue Oxygenation Estimation Powered by the Extended Modified Lambert–Beer Law. Applied Sciences **12**(17), 8490 (2022)
8. Koçak, B., Durmaz, E., Ateş, E., Kılıçkesmez, Ö.: Radiomics with artificial intelligence: a practical guide for beginners. Diagn Interv Radiol **25**(6), 485-495 (2019)
9. Gong, W., Wu, X.: Differential Diagnosis of Latent Tuberculosis Infection and Active Tuberculosis: A Key to a Successful Tuberculosis Control Strategy. Frontiers in Microbiology **12** (2021)
10. Dendere, R., Mutsvangwa, T., Goliath, R., Rangaka, M.X., Abubakar, I., Douglas, T.S.: Measurement of Skin Induration Size Using Smartphone Images and Photogrammetric Reconstruction: Pilot Study. JMIR Biomed Eng **2**(1), e3 (2017)
11. Naraghi, S., Mutsvangwa, T., Goliath, R., Rangaka, M.X., Douglas, T.S.: Mobile phone-based evaluation of latent tuberculosis infection: Proof of concept for an integrated image capture and analysis system. Computers in Biology and Medicine **98** 76-84 (2018)
12. Maclean, S.: Image analysis for a mobile phone-based assessment of latent tuberculosis infection. University of Cape Town (2020)
13. Toivgoogiin, A., Toyota, M., Yasuda, N., Ohara, H.: Validity of using tuberculin skin test erythema measurement for contact investigation during a tuberculosis outbreak in schoolchildren previously vaccinated with BCG. Journal of epidemiology **15**(2), 56-64 (2005)
14. Carranza, C., Pedraza-Sanchez, S., de Oyarzabal-Mendez, E., Torres, M.: Diagnosis for Latent Tuberculosis Infection: New Alternatives. Frontiers in Immunology **11** (2020)
15. Dinkele, R., Gessner, S., McKerry, A., Leonard, B., Seldon, R., Koch, A.S., Morrow, C., Gqada, M., Kamariza, M., Bertozzi, C.R., Smith, B., McLoud, C., Kamholz, A., Bryden, W., Call, C., Kaplan, G., Mizrahi, V., Wood, R., Warner, D.F.: Capture and visualization of live Mycobacterium tuberculosis bacilli from tuberculosis patient bioaerosols. PLOS Pathogens **17**(2), e1009262 (2021)
16. Moayedi-Nia, S., Barss, L., Oxlade, O., Valiquette, C., Ly, M.-X., Campbell, J.R., Lan, Z., Nsengiyumva, P., Fregonese, F., Lisboa Bastos, M., Sampath, D., Winters, N., Menzies, D.: The mTST – An mHealth approach for training and quality assurance of tuberculin skin test administration and reading. PLOS ONE **14**(4), e0215240 (2019)
17. Kiazyk, S., Ball, T.B.: Latent tuberculosis infection: An overview. Can Commun Dis Rep **43**(3-4), 62-66 (2017)
18. Stavri, H., Bucurenci, N., Ulea, I., Costache, A., Popa, L., Popa, M.I.: Use of recombinant purified protein derivative (PPD) antigens as specific skin test for tuberculosis. Indian J Med Res **136**(5), 799-807 (2012)
19. Santos, J.A., Duarte, R., Nunes, C.: Tuberculin skin test and interferon-γ release assays: Can they agree? The Clinical Respiratory Journal **17**(2), 109-114 (2023)
20. Calin, M.A., Manea, D., Savastru, R., Parasca, S.V.: Mapping the Distribution of Melanin Concentration in Different Fitzpatrick Skin Types Using Hyperspectral Imaging Technique. Photochem Photobiol **99**(3), 1020-1027 (2023)
21. Jakovels, D., Spigulis, J.: 2-D mapping of skin chromophores in the spectral range 500 – 700 nm. Journal of Biophotonics **3**(3), 125-129 (2010)

22. van Griethuysen, J.J.M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R.G.H., Fillion-Robin, J.-C., Pieper, S., Aerts, H.J.W.L.: Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Research **77**(21), e104-e107 (2017)
23. Behmann, J., Acebron, K., Emin, D., Bennertz, S., Matsubara, S., Thomas, S., Bohnenkamp, D., Kuska, M.T., Jussila, J., Salo, H.: Specim IQ: Evaluation of a new, miniaturized handheld hyperspectral camera and its application for plant phenotyping and disease detection. Sensors **18**(2), 441 (2018)
24. Gunturk, B.K., Glotzbach, J., Altunbasak, Y., Schafer, R.W., Mersereau, R.M.: Demosaicking: color filter array interpolation. IEEE Signal processing magazine **22**(1), 44-54 (2005)
25. Prahl, S.: Bilirubin, https://omlc.org/spectra/PhotochemCAD/html/119.html, last accessed 2024/07/03
26. Prahl, S.: Assorted Spectra, https://omlc.org/spectra/index.html, last accessed 2024/07/03
27. Mei, X., Liu, Z., Robson, P.M., Marinelli, B., Huang, M., Doshi, A., Jacobi, A., Cao, C., Link, K.E., Yang, T., Wang, Y., Greenspan, H., Deyer, T., Fayad, Z.A., Yang, Y.: RadImageNet: An Open Radiologic Deep Learning Research Dataset for Effective Transfer Learning. Radiology: Artificial Intelligence **4**(5), e210315 (2022)
28. Prahl, S.: Optical Absorption of Hemoglobin, https://omlc.org/spectra/hemoglobin/, last accessed 2024/07/03
29. Kumar, N.P., Moideen, K., Banurekha, V.V., Nair, D., Babu, S.: Plasma Proinflammatory Cytokines Are Markers of Disease Severity and Bacterial Burden in Pulmonary Tuberculosis. Open Forum Infect Dis **6**(7), ofz257 (2019)
30. Fu, C., Chen, J., Lu, J., Yi, L., Tong, X., Kang, L., Pei, S., Ouyang, Y., Jiang, L., Ding, Y., Zhao, X., Li, S., Yang, Y., Huang, J., Zeng, Q.: Roles of inflammation factors in melanogenesis (Review). Mol Med Rep **21**(3), 1421-1430 (2020)
31. Leon, R., Martinez-Vega, B., Fabelo, H., Ortega, S., Melian, V., Castaño, I., Carretero, G., Almeida, P., Garcia, A., Quevedo, E., Hernandez, J.A., Clavo, B., G, M.C.: Non-Invasive Skin Cancer Diagnosis Using Hyperspectral Imaging for In-Situ Clinical Support. J Clin Med **9**(6), (2020)
32. Schimunek, L., Schöpp, K., Wagner, M., Brucker, S.Y., Andress, J., Weiss, M.: Hyperspectral imaging as a new diagnostic tool for cervical intraepithelial neoplasia. Arch Gynecol Obstet **308**(5), 1525-1530 (2023)
33. Kounas, K., Dinh, T., Riemer, K., Rosenblum, B.I., Veves, A., Giurini, J.M.: Use of hyperspectral imaging to predict healing of diabetic foot ulceration. Wound Repair and Regeneration **31**(2), 199-204 (2023)