Iowa Initiative for Artificial Intelligence

Final Report

Project title:	AI-assisted non-invasive identification of melanoma pathologies and healthy nevus via IR spectroscopy	
Principal Investigator:	PI: Fatima Toor, Ph.D. Co-PIs: Jeffrey Johnson, M.D. , and Mohammed Milhem, MBBS	
Prepared by (IIAI):	Yanan Liu	
Other investigators:		
Date:		
Were specific aims fulfilled:	Y/N	
Readiness for extramural proposal?	Y/N	
If yes … Planned submission date		
Funding agency		
Grant mechanism		
If no … Why not? What went wrong?		

Brief summary of accomplished results:

We have developed and validated a Random Forest model to accurately predict tumor histologic subtype using Fourier Transform Infrared (FTIR) spectroscopy data. The prediction accuracy of 4 class is 84% and the prediction accuracy of 2 class (tumor vs nevus) is 90%.

Research report:

Aims (provided by PI):

Develop an AI/ML model that takes in clinical data obtained from the BioMER for each tissue specimen that is collected based on the ABCDE (the lesion asymmetry, irregular border, color, diameter, or evolution over time) rule of early melanoma detection[1, 2] and FTIR spectra of the melanoma and nevus samples for robust diagnoses of a common cancer type.

Data:

Melanoma tissue samples were collected from 12 different patients representing the three common melanoma pathologies and five different patient samples for health nevus types. These slides were characterized using the FTIR in the Toor lab to generate the data to support the AI/ML data analysis. IR slides was mounted on a micrometer x,y translational slide holder and several scans for each patient sample were displaced 0.5 mm to 5 mm (dependent on tissue size) in both x,y directions to study whether tissue heterogeneity results in spectral signatures. Overall, 419 IR plots were generated (nodular melanoma: 162; superficial spreading melanoma: 67; desmoplastic melanoma: 60; nevus:130).

AI/ML Approach:

In this study, supervised machine learning algorithm was implemented for prediction using Python. As many extracted features may be noisy, or highly correlated with each other, Random Forest (RF) algorithm was selected to predict predict tumor/nevus group and performance assessed using 5-fold-cross-validation (Figure 1). Accuracy and F1-score were calculated to compare performance of different tests.

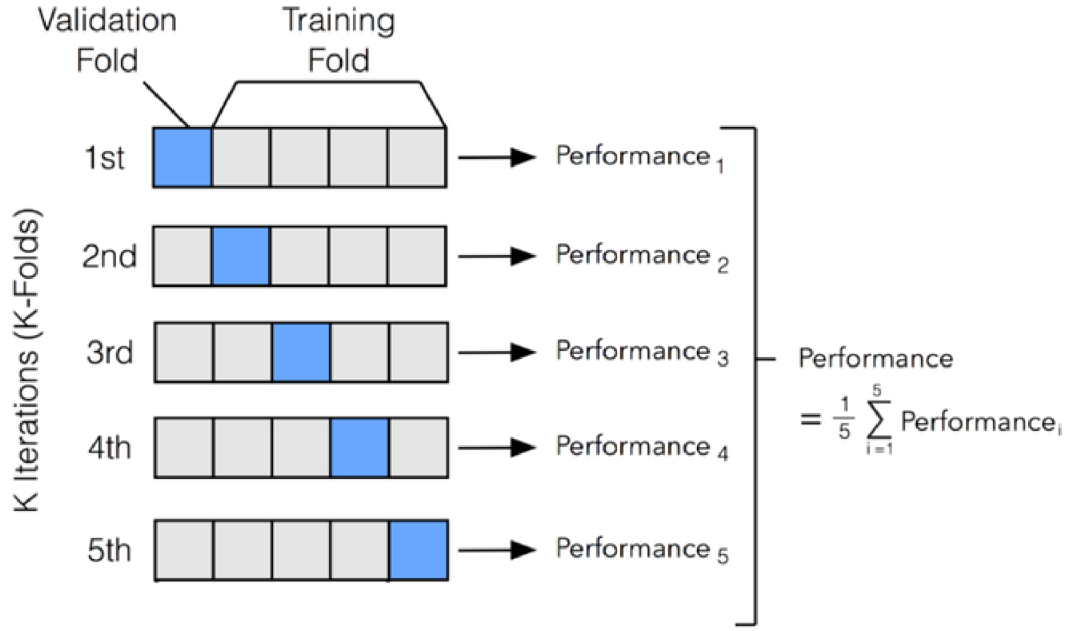

Figure 1. Representation of 5-fold cross-validation technique

Experimental methods, validation approach:

For each IR plot (Figure 2), 1869 wavelength number were used as input for Random Forest. Each IR plot was treated independently to increase input size for machine learning model.

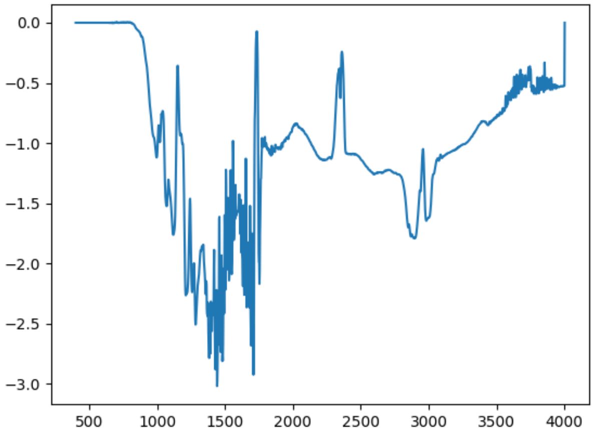

Figure 2. Representation of IR plot.

Results:

Tumor subtype

Accuracy and F1 score of different tumor type/ nevus were calculated(table 1).

Table 1 Accuracy and F1 score of different tumor/nevus group

	Accuracy (mean±SD)	F1 score
Nodular Melanoma	0.87±0.11	0.88
Superficial Spreading Melanoma	0.98±0.02	0.97
Desmoplastic Melanoma	0.96±0.03	0.93

Binary classification(tumor vs nevus) accuracy was 90% and F1 score was 0.92. Multiclass classification(3 tumor subtpye and nevus) accuracy was 84% and multicalss classification for 3 tumor subtpye accuracy was 85%.

Pathologic staging

Each subject has pathologic staging informaiton(stage 1, stage 2, stage 3). We got 103 stage 1, 97 stage 2 and 89 stage 3 regardless tumor subtype. The accuracy of prediting pathologic staging for melanoma was 85%. Due to small number of IR plots in Superficial Spreading Melanoma and Desmoplastic Melanoma group, we only separated stage information for Nodular Melanoma patients. We got 52 stage 1, 42 stage 2 and 68 stage 3 for Nodular Melanoma. The accuracy of prediting pathologic staging for Nodular Melanoma was 93%.

The findings of this study should be confirmed in a large-scale study.