University of Pittsburgh | School of Computing and Information

Dissertation Defense
Doctor of Philosophy in Information Science

"Adaptive Ensemble Learning for Anomaly Detection in Hyperspectral Imaging"
by **Bradley J. Wheeler**

Date: November 5, 2024
Time: 12:00 – 2:00 p.m.
Place: 502 Information Science Building, N Bellefield Ave, Pittsburgh PA 15213

Committee:

- Dr. Hassan A. Karimi, Advisor, Professor, Department of Informatics and Networked Systems, School of Computing and Information
- Dr. Daqing He, Professor, Department of Informatics and Networked Systems, School of Computing and Information
- Dr. Vladimir Zadorozhny, Professor, Department of Informatics and Networked Systems, School of Computing and Information
- Dr. Xiaowei Jia, Assistant Professor, Department of Computer Science, School of Computing and Information

Abstract:
Hyperspectral images provide detailed information about reflected electromagnetic radiation that range across the ultraviolet, visible, and infrared wavelength spectrums. This information can be critical in instances where visible light spectrums along may be insufficient in capturing the nuances required for completing complex analysis in remote sensing. This has implications in domains such as agriculture, surveillance, disaster recovery, and environmental monitoring to name a few. An important step in leveraging hyperspectral data in these applications is anomaly detection. Hyperspectral anomaly detection (HAD) has been a long-standing focus in research, leading to the generation of a substantial body of literature aimed at addressing this challenging problem. HAD algorithms are designed to exploit the nuanced information embedded in hyperspectral images to identify each spectral signature within an image as an anomaly or part of the background.

While many algorithms exist for detecting anomalies in hyperspectral images, most originate from a few main methods, and accordingly share one of a few main modeling biases about the nature of the hyperspectral images. This trend underpins three key problems in HAD: selection bias, performance disparities, and singular modeling bias. The limited range of modeling biases employed in HAD make correlations between datasets and algorithms likely, therefore care is required in their selection to prevent providing any algorithm an inherent advantage, skewing results and leading to misleading conclusions. These limited modeling biases additionally enable some algorithms to excel on individual datasets while other algorithms generalize better across a diverse set of datasets, creating a disparity between specialized and generalized performance across datasets. Lastly, HAD algorithm implement a singular modeling bias, limiting model flexibility and creating issues with underfitting or overfitting in various scenarios.

To address these challenges, I focus on three key tasks, each centered around the modeling biases commonly employed in HAD algorithms. First, I develop a framework to identify significant correlations between these biases and hyperspectral datasets. This framework

enables the prediction of which modeling bias is likely to perform best on a given dataset, helping to reduce selection bias when choosing datasets and algorithms for HAD tasks. Next, I design an adaptive ensemble learning algorithm that integrates information from multiple diverse HAD modeling biases. This approach demonstrates how combining multiple biases can mitigate disparities between specialized and generalized performance across datasets. Finally, I conduct a systematic study on how different error quantification methods in ensemble learning influence the extent to which each modeling bias contributes to the overall solution. This provides valuable insights into the utility of various modeling biases across different datasets. Together, these contributions highlight the importance of incorporating diverse modeling biases in HAD and demonstrate the pivotal role that ensemble learning plays in effectively integrating them.