Announces the Ph.D. Dissertation Defense of

**John T. Hancock III**

for the degree of Doctor of Philosophy (Ph.D.)

“Fraud Detection in Highly Imbalanced Big Data with Novel and Efficient Data Reduction Techniques”

April 8, 2024, 10:30 a.m.
In-person Room EE-405

**DEPARTMENT:**
Electrical Engineering and Computer Science

**ADVISOR:**
Taghi M. Khoshgoftaar, Ph.D.

**Ph.D. SUPERVISORY COMMITTEE:**
Taghi M. Khoshgoftaar, Ph.D., Chair
Man Chon U, Ph.D.
DingDing Wang, Ph.D.
Mehrdad Nojoumian, Ph.D.

**ABSTRACT OF DISSERTATION**

The rapid growth of digital transactions and the increasing sophistication of fraudulent activities have necessitated the development of robust and efficient fraud detection techniques, particularly in the financial and healthcare sectors. This dissertation focuses on the use of novel data reduction techniques for addressing the unique challenges associated with detecting fraud in highly imbalanced Big Data, with a specific emphasis on credit card transactions and Medicare claims. The highly imbalanced nature of these datasets, where fraudulent instances constitute less than one percent of the data, poses significant challenges for traditional machine learning algorithms. This dissertation explores novel data reduction techniques tailored for fraud detection in highly imbalanced Big Data. The primary objectives include developing efficient data preprocessing and feature selection methods to reduce data dimensionality while preserving the most informative features, investigating various machine learning algorithms for their effectiveness in handling imbalanced data, and evaluating the proposed techniques on real-world credit card and Medicare fraud datasets.

This dissertation covers a comprehensive examination of datasets, learners, experimental methodology, sampling techniques, feature selection techniques, and hybrid techniques. Key contributions include the analysis of performance metrics in the context of newly available Big Medicare Data, experiments using Big Medicare data, application of a novel ensemble supervised feature selection technique, and the combined application of data sampling and feature selection. The research demonstrates that, across both domains, the combined application of random undersampling and ensemble feature selection significantly improves classification performance.

The contributions presented advance the field of fraud detection. They add to the development of fraud detection systems. The proposed data reduction techniques offer practical solutions for tackling the challenges associated with imbalanced and high-dimensional data, offering a means to reduced financial losses, improved patient care, and increased trust in critical financial and healthcare domains. This dissertation offers practical implications for the financial and healthcare sectors by improving the detection of fraudulent transactions and Medicare Insurance claims. The proposed techniques may also be considered for general machine learning applications.
BIOGRAPHICAL SKETCH

Born in Florida, USA
B.S., University of Maryland University College 2009
M.S., Florida Atlantic University, Boca Raton, Florida 2015
Ph.D., Florida Atlantic University, Boca Raton, Florida 2024

CONCERNING PERIOD OF PREPARATION
& QUALIFYING EXAMINATION

Time in Preparation: 2019-2024
Qualifying Examination Passed: Spring 2019

Published Papers: