



COLLEGE OF ENGINEERING
AND COMPUTER SCIENCE
FLORIDA ATLANTIC UNIVERSITY

Announces the Ph.D. Dissertation Defense of

Zahra Salekshahrezaee

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

“Robust Anomaly Detection under Data Imbalance, Noise, and Label Scarcity”

November 24, 2025, 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

Imadeldin Mahgoub, Ph.D.

Mehrdad Nojournian, Ph.D.

DingDing Wang, Ph.D.

ABSTRACT OF DISSERTATION

In today's data-driven landscape, large volumes of data are generated continuously, often containing imperfections such as noise, missing data, or unreliable labeling. These real-world datasets are typically high-dimensional, sparsely labeled, and imbalanced, creating substantial challenges for both supervised and unsupervised learning. These challenges are especially prevalent in the task of anomaly detection, where instances belonging to the class of interest are rare and underrepresented compared to normal instances. This dissertation proposes and evaluates robust frameworks for anomaly detection that address these data challenges and improve model performance and robustness using real-world datasets, including credit card transactions and cognitive assessments.

Supervised learning requires labeled data which can be costly, hard to produce, and prone to mislabeling. We propose a reconstruction error-based method to identify and correct mislabeled samples, thereby improving the quality of labeled data. To address imbalance and high dimensionality, we combine deep feature extraction using convolutional autoencoders, an unsupervised learning technique, with class rebalancing strategies to improve classification performance. Then, we examine how the order of preprocessing steps affects downstream ensemble learners. For unlabeled data, we propose a novel hybrid unsupervised framework that integrates convolutional autoencoders for representation learning with Isolation Forest for anomaly detection (CAE-IF). CAE-IF demonstrates robust performance on unlabeled, high-dimensional, and imbalanced data across cognitive and fraud detection domains, relative to common baselines such as Isolation Forest and Local Outlier Factor. In addition, we apply an instance-based iterative cleaning method that uses reconstruction error to remove likely outliers and improves representation quality for downstream detection without requiring manual annotation. The results demonstrate that our proposed approaches improve model robustness in various imperfect data conditions. Collectively, these contributions provide a practical and generalizable toolkit for anomaly detection, addressing the core challenges of class imbalance, label noise, and label scarcity across both supervised and unsupervised settings.

BIOGRAPHICAL SKETCH

Born in Isfahan Iran

B.S., Isfahan University, Iran 2011

M.S., Alzahra University, Iran 2014

Ph.D., Florida Atlantic University, Boca Raton, Florida 2025

CONCERNING PERIOD OF PREPARATION & QUALIFYING EXAMINATION

Time in Preparation: 2020-2025

Qualifying Examination Passed: Spring 2020

Published Papers:

Salekshahrezaee, Zahra, Joffrey L. Leevy, and Taghi M. Khoshgoftaar. "A reconstruction error-based framework for label noise detection." *Journal of Big Data* 8, no. 1 (2021): 57. 2021

Salekshahrezaee, Zahra, Joffrey L. Leevy, and Taghi M. Khoshgoftaar. "Feature extraction for class imbalance using a convolutional autoencoder and data sampling." In *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 217-223. IEEE, 2021.

Salekshahrezaee, Zahra, Joffrey L. Leevy, and Taghi M. Khoshgoftaar. "A class-imbalanced study with feature extraction via PCA and convolutional autoencoder." In *2022 IEEE 23rd International Conference on Information Reuse and Integration for Data Science (IRI)*, pp. 63-68. IEEE, 2022

Kennedy, Robert KL, Zahra Salekshahrezaee, and Taghi M. Khoshgoftaar. "A novel approach for unsupervised learning of highly imbalanced data." In *2022 IEEE 4th International Conference on Cognitive Machine Intelligence (CogMI)*, pp. 52-58. IEEE, 2022.

Salekshahrezaee, Zahra, Joffrey L. Leevy, and T. M. Khoshgoftaar. The effect of feature extraction and data sampling on credit card fraud detection. *Journal of Big Data* 10 (1). 2023.

Kennedy, Robert KL, Zahra Salekshahrezaee, Flavio Villanustre, and Taghi M. Khoshgoftaar. "Iterative cleaning and learning of big highly imbalanced fraud data using unsupervised learning." *Journal of Big Data* 10, no. 1 (2023): 106. 2023

Kennedy, Robert KL, Zahra Salekshahrezaee, and Taghi M. Khoshgoftaar. "Unsupervised anomaly detection of class imbalanced cognition data using an iterative cleaning method." In *2023 IEEE 24th International Conference on Information Reuse and Integration for Data Science (IRI)*, pp. 303-308. IEEE, 2023.

Kennedy, Robert KL, Flavio Villanustre, Taghi M. Khoshgoftaar, and Zahra Salekshahrezaee. "Synthesizing class labels for highly imbalanced credit card fraud detection data." *Journal of Big Data* 11, no. 1 (2024): 38. 2024

Leevy, Joffrey L., Zahra Salekshahrezaee, and Taghi M. Khoshgoftaar. "A Review of Unsupervised Anomaly Detection Techniques for Health Insurance Fraud." In *2024 IEEE 10th International Conference on Big Data Computing Service and Machine Learning Applications (BigDataService)*, pp. 141-149. IEEE, 2024.

Salekshahrezaee, Zahra, and Taghi M. Khoshgoftaar. "The Impact of Class Imbalance on Unsupervised Deep Anomaly Detection for Cognitive Data." In *2025 IEEE International Conference on Information Reuse and Integration and Data Science (IRI)*, pp. 222-227. IEEE, 2025.

Salekshahrezaee, Zahra, and Taghi M. Khoshgoftaar. "Unsupervised Cognitive Impairment Detection Using Convolutional Autoencoders and Isolation Forest." In *2025 IEEE International Conference on Information Reuse and Integration and Data Science (IRI)*, pp. 295-300. IEEE, 2025.

Salekshahrezaee, Zahra, Mary Anne Walauskis, and Taghi M. Khoshgoftaar. "Unsupervised Feature Extraction using Convolutional Autoencoder for Credit Card Fraud Detection." In *2025 IEEE International Conference on Machine Learning and Applications (ICMLA)*. (Accepted, Forthcoming 2025).