Chapter 10
QINN-Based Approach to Detect Anomalies in High-Dimensional Secure Data

V. Thamilarasi
https://orcid.org/0000-0002-5866-0074
Sri Sarada College for Women, India

Nitendra Kumar
https://orcid.org/0000-0001-7834-7926
Amity University, Noida, India

P Ganesh Kumar
https://orcid.org/0000-0001-8681-8169
College of Engineering, Anna University, Chennai, India

G. Sivaraman
M.G.R. College (Arts and Science), India

R. RajiniGanth
SNS College of Engineering, India

Janaki Sivakumar
https://orcid.org/0000-0003-0412-1515
Global College of Engineering and Technology, Oman

ABSTRACT

Anomaly discovery is a crucial aspect of modern data analysis to finding unusual trends or behaviours in datasets across various domains like cybersecurity and finance and the healthcare. However, overfitting, in which the model becomes overly adapted to training data, results in false negatives and misclassifications, makes it difficult to target optimal detection capabilities. To get around this, training data must be carefully sanitized, eliminating unknown and irregular anomalous instances to guarantee precise anomaly detection. It is already difficult to accomplish optimal and significant feature extraction when working with noisy, high-dimensional data, like that found in network traffic. When features are too similar, overfitting and

DOI: 10.4018/979-8-3373-3551-3.ch010

incorrect classifications may occur. However, classification accuracy may suffer if important features are removed. By improving likelihood estimation and feature separability, unsupervised techniques can assist in finding and keeping pertinent features to enhance model performance.

I. INTRODUCTION

Finding patterns and data points that drastically deviate from expected norms is made possible by the crucial data analysis technique known as anomaly detection. With the growing complexity and size of datasets, its significance has increased (Swathi, Altalbe, & Kumar, 2024). In domains like healthcare, cybersecurity, industrial processes, and fraud, anomalies may indicate issues like mistakes, surprises, or danger. In these domains, effective anomaly detection is crucial to exposing information and averting damage. Anomaly detection uses a range of techniques, including deep learning, machine learning models, and statistical techniques; the choice is based on the type of data, automation level, and desired outcomes (Thi & Nguyen, 2024). Efficient anomaly detection identifies likely problems like vulnerabilities, frauds, errors, and future errors so that businesses can look for insights, lessen risks, and improve decision-making. Anomaly detection is invaluable in data accuracy and the data's strength in insights, especially in risk-critical uses such as cybersecurity and finance where finding abnormal patterns of behaviour or malicious behaviour is capable of terminating cyberattacks and avoiding financial loss and thus strengthening data analysis and decision-making resilience (Maddali, 2024) (Pham & Raahemi, 2025). Real-time anomaly detection has arisen with growing volume and velocity of data, utilizing methods such as online learning, adaptive models, and ensemble methods to learn dynamically evolving patterns and identify anomalies in high-volume, high-dimensional, complex data sets with many features. It is faced with increased computational complexity, visualization issues, and sensitivity to noise (Muneer et al., 2022). High-dimensional data is defined as having a large feature set, or number of dimensions, compared to the number of observations. In conventional statistics, Data sets that include a great number of variables are considered highly dimensional. Space and time complexity are frequently used to quantify the computational complexity of algorithms. Algorithm efficiency is hampered in high-dimensional spaces by the exponential growth in the number of calculations and memory needed. In order to effectively address the computational demands, this increase in complexity calls for the investigation of specialized algorithms and parallel computing architectures. Data sets with a large number of variables are referred to as highly dimensional in traditional statistics. Space and time complexity are frequently used to quantify the computational complexity of