IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 25 January 2024, accepted 2 March 2024, date of publication 4 March 2024, date of current version 8 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3373694

RESEARCH ARTICLE

Enhancing Insider Threat Detection in Imbalanced Cybersecurity Settings Using the Density-Based Local Outlier Factor Algorithm

TAHER AL-SHEHARI[1], DOMENICO ROSACI[2], MUNA AL-RAZGAN[3], TAHA ALFAKIH[4], MOHAMMED KADRIE[1], HAMMAD AFZAL[5], (Senior Member, IEEE), AND RAHEEL NAWAZ[6]

[1]Department of Self-Development Skill, Common First Year Deanship, King Saud University, Riyadh 11362, Saudi Arabia
[2]Department of Information Engineering, Infrastructure and Sustainable Energy (DIIES), Mediterranea University of Reggio Calabria, 89122 Reggio Calabria, Italy
[3]Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11345, Saudi Arabia
[4]Department of Information Systems, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[5]National University of Sciences and Technology, Islamabad 44000, Pakistan
[6]Executive Group, Staffordshire University, ST4 2DE Stoke-on-Trent, U.K.

Corresponding author: Taher Al-Shehari (talshehari.c@ksu.edu.sa)

This work was supported by King Saud University, Riyadh, Saudi Arabia, through the Researchers Supporting Project under Grant RSP2024R206.

ABSTRACT In today's interconnected world, cybersecurity has emerged as a critical domain for ensuring the integrity, confidentiality, and availability of digital assets. Within this sphere, insider threats represent a unique and particularly insidious class of security risks, originating not from external hackers but from within the organization itself. These threats are perpetrated by individuals with inside information concerning the organization's security practices, data, and computer systems. Traditional security measures like firewalls, intrusion detection systems, and antivirus software are often inadequate for tackling insider threats effectively, owing to their focus on external threats. This inadequacy underscores the urgent need for the development and implementation of more sophisticated, targeted detection techniques for insider threats. In response to this challenge, our research introduces an innovative approach that employs the Density-Based Local Outlier Factor (DBLOF) algorithm, fine-tuned to specifically tackle the challenges posed by the imbalanced nature of the CERT r4.2 insider threat dataset. This dataset is characterized by a highly skewed distribution, with a significant majority of benign instances and only a minimal proportion of malicious activities. Conventional detection algorithms often fail to effectively identify these rare but dangerous instances, leading to a high rate of false negatives. Our methodology capitalizes on the algorithm's ability to focus on the local density deviation of data points, thereby enabling the precise identification of outliers that are indicative of potential insider threats. Through rigorous testing and validation processes, we have achieved outstanding results, with an of F-score 98%. These remarkable outcomes not only affirm the effectiveness of the DBLOF algorithm as a powerful tool for combating insider threats but also contribute valuable insights to the broader academic and professional discourse on cybersecurity. Importantly, our findings have practical implications, offering organizations actionable recommendations for boosting their internal security mechanisms against the complex and evolving landscape of insider threats.

INDEX TERMS Machine learning, insider threat detection, local outlier factor algorithm, data imbalance addressing.

The associate editor coordinating the review of this manuscript and approving it for publication was S. Chandrasekaran.

I. INTRODUCTION

The introduction should briefly place the study in a broad context and highlight why it is important. It should define the purpose of the work and its significance. The current state of the research field should be carefully reviewed and key publications cited. Please highlight controversial and diverging hypotheses when necessary. Finally, briefly mention the main aim of the work and highlight the principal conclusions. As far as possible, please keep the introduction comprehensible to scientists outside your particular field of research. References should be numbered in order of appearance and indicated by a numeral or numerals in square brackets—e.g., [1] or [2], [3], or [4], [5], [6]. See the end of the document for further details on references.

The digital age, characterized by an ever-increasing reliance on technology, has brought forth a plethora of opportunities and advancements across various sectors. However, with these advancements come challenges, particularly in the domain of cybersecurity. While external threats have long been a focal point of security measures, insider threats, malicious activities originating from within an organization, have emerged as a pressing concern. These threats, often perpetrated by trusted employees or associates with access to sensitive information, can lead to significant damages, both in terms of financial loss and reputational harm. Insider threats are a significant issue in the cybersecurity sector since they are more difficult to identify and categorize because insiders interact with the system like ordinary users. According to studies, insider attacks, conducted by malevolent, irresponsible, or dissatisfied individuals inside organizations, represent 79% of cybersecurity concerns [1].

The CERT r4.2 insider threat dataset, a benchmark in the field, encapsulates the complexity of this challenge. It presents a highly imbalanced distribution, with a vast majority of instances being benign and only a minuscule fraction representing malicious activities. This imbalance mirrors real-world scenarios where malicious insider activities are rare but extremely consequential. Traditional detection methods, when applied to such datasets, often yield unsatisfactory results, primarily due to their inability to effectively distinguish between rare malicious instances and the overwhelming benign ones. Numerous challenges (e.g., complexity, sparsity, high dimensionality, and heterogeneity) are present when collecting data about user activity inside an organization. Algorithms for machine learning (ML) presume that the processed data are balanced, which means that the classes of the dataset are about equal. In contrast, user activity data observations are common in real-world situations, whereas harmful observations are uncommon. With such uneven data observations of dataset classes, the majority class is detected with high accuracy while the minority class is detected with low accuracy. This scenario is inappropriate in the realm of threat detection since the minority class is essential to detection [2], [3]. Researchers have suggested a variety of strategies to address this imbalance issue. The most well-known strategy is to use data augmentation approaches to rebalance the class distribution of datasets [4]. The Synthetic Minority Oversampling Technique (SMOTE) technique [5] is a straightforward approach that involves randomly repeating examples of the minority class until the dataset classes are balanced. To equalize the minority class, another tactic is to under-sample the majority occurrences. Tragically, these methods only consider local information, which leads to unitary synthetic samples and overfitting.

In light of these challenges, there is a pressing need for innovative approaches that can accurately detect insider threats, even when they are few and far between. This research paper delves into the application of the DBLOF algorithm as a potential solution to this conundrum. The DBLOF algorithm, by virtue of its design, emphasizes the local density deviation of data points, making it adept at identifying outliers in imbalanced datasets. By treating malicious activities as outliers in a predominantly benign environment, we posit that the DBLOF algorithm can offer superior detection capabilities. The goal of this study is to solve the imbalance problem in the CERT's dataset by proposing an insider threat detection model based on the DBLOF method. By using the DBLOF, our technique used the algorithm level as opposed to the prior approaches that used the data level (oversampling and under-sampling). Instead of modifying the data via duplication and removal to more effectively identify insider threats, the goal is to analyze the data as it is in a real-world context to represent the genuine situation. The following contributions are desired for that goal: focusing on identifying significant insider threats made by a bad insider prior to that person leaving a company; Making use of the DBLOF method to deal with the dataset's severely skewed class distribution; Using many runs to validate the proposed model and fine-tuning the contamination hyper parameters of the algorithm to establish a more accurate model; Applying the baseline (supervised machine learning methods) and the DBLOF method for insider threat identification to show how the proposed model is superior to both the baseline and current models.

A crucial issue in detecting insider threats is the inherent imbalance in datasets. Typically, the number of benign instances vastly outnumbers the malicious instances, making it challenging for machine learning models to adequately learn the characteristics of the minority class. Traditional algorithms are often biased towards the majority class, reducing the effectiveness of threat detection systems. The primary objective of this research is to address the significant challenges posed by insider threats in imbalanced cybersecurity environments. In tackling this issue, we focus on leveraging advanced machine learning techniques to improve detection rates and reduce false positives. Specifically, we explore the efficacy of the DBLOF algorithm, a variant of the Local Outlier Factor (LOF) algorithm, designed to work effectively on imbalanced datasets.

The crux of this research lies in its rigorous testing and validation of the DBLOF algorithm on the CERT r4.2 datasets. The results, highlighted by an impressive 99% F-score, not only underscore the algorithm's efficacy but also pave the way for its potential adoption in real-world scenarios. Through this paper, we aim to contribute to the broader discourse on insider threat detection, offering both theoretical insights and practical solutions to one of the most pressing cybersecurity challenges of our time. The objectives of this research are as follows:

- To assess the scalability of the DBLOF algorithm in larger and more complex datasets, examining its potential for broader applications. This is by developing a novel methodology specifically tailored for the CERT r4.2 insider threat dataset, which is characterized by a highly imbalanced distribution of benign and malicious instances.
- To minimize the rate of false positives while maintaining high detection accuracy, ensuring the model's applicability in real-world scenarios. This is by employing the DBLOF algorithm for detecting anomalous behaviors indicative of insider threats, focusing on the local density deviation of data points.
- To offer guidelines and best practices for implementing a DBLOF-based insider threat detection system in organizational settings. This can provide actionable recommendations and insights for organizations aiming to strengthen their internal security mechanisms against insider threats.

This research is dedicated to making a substantial impact in the domain of cybersecurity and machine learning, with a particular emphasis on addressing the difficulties associated with imbalanced data and the subtle challenges posed by insider threats. The structure of the study is organized as follows: The study begins with introduction Section I. An extensive literature review discussed in Section II. This section investigates the existing body of work related to insider threat detection, providing a critical summary of past research efforts. It also sheds light on the challenges faced when dealing with imbalanced datasets, a common complex issue in this field. This review sets the stage for the subsequent sections by establishing the current state of research and identifying gaps that this study aims to address.

In Section III, the focus shifts to the methodology employed in this research. This section is dedicated to a detailed explanation of the applied DBLOF algorithm. It explains on how the DBLOF algorithm is applied within the context of this research, detailing the specific processes and techniques used. This includes an explanation of how the algorithm is tailored to identify and analyze insider threats effectively, especially in scenarios characterized by imbalanced data.

Section IV presents the experimental results of the study. It offers a thorough evaluation of the model's performance metrics. It provides an in-depth analysis of how the model behaves under various conditions and its effectiveness in detecting insider threats. The results are presented in a manner that allows for clear interpretation and understanding of the model's performance.

Section V involves a critical discussion of the study's findings. It compares the performance of the model developed in this research with baseline models and previous studies in the field, offering a contextual understanding of its efficiency. This section also acknowledges the limitations encountered in the study, providing an open reflection on areas that could not be addressed entirely. Importantly, it proposes how these limitations could be overcome in future research, paving the way for continued advancements in this area. Finally, Section VI concludes the study by summarizing the key results and highlighting potential avenues for future investigation, suggesting how this research could be extended or applied in other contexts within cybersecurity and machine learning.

II. RELATED WORK

Insider threat research has attracted the attention of the cybersecurity research community for many years. Multiple approaches are proposed for insider threat detection. The recent survey in [6] summarized the insider threat detection models in the literature. It categorized and compared insider threat detection techniques into several factors: datasets, feature domains, classification techniques, simulated scenarios, and performance and accuracy metrics. It highlights the factors that reflect the methodology and performance of the reviewed approaches from various empirical perspectives. Another recent survey [7] addresses the taxonomy and categorization on insider threats.

Many insider threat detection approaches aim to identify outliers and anomalies in very large datasets to combat insider threats. The approaches in [8] and [9] applied different anomaly detection techniques (e.g., hidden Markov model (HMM), Gaussian Mixture Model (GMM), etc.). They are implemented on insiders' activity log data to detect various indicators of insider threats. In [10], a hybrid anomaly method was proposed for detecting blend-in outliers and unusual change anomalies. It employed hybrid detectors of anomalous insiders' activities on collected data from various domains. An anomaly and quitter detection mechanism were proposed to compact insider threats by Gavai et al. [11] They employed various machine learning-based techniques on insiders' activity log data throughout the organization. In their approach, multiple indicators of potential malicious user-related activities are considered. Rashid et al. [12] applied a hidden Markov model (HMM) in a recent study for insider anomalous detection. They utilized HMMs to model the sequences of users' activities within an organization on a weekly basis. The possible anomalous activities of malicious insiders were detected by the subtle changes that were below a given threshold. In [13], a model for categorizing and presenting malicious and normal users' activities within an organization was proposed. The authors employed self-organizing graphs and unsupervised machine learning techniques. Their proposed model presented the potential for providing cybersecurity analysts and decision makers with a visual tool to detect malicious activities of an organization's network's users. A framework for insider threat problem formulation was proposed by Legg et al. [14].

The proposed framework depends on behavioral and psychological observations of employees within an organization. It enables cybersecurity analysts and decision-makers to formulate hypothesis trees that detect potential insider threats based on the reasoning in the users' activity log data.

The One-Class Support Vector Machine (OCSVM) method was suggested by Parveen et al. [15] for the purpose of classifying data into normal and abnormal categories, specifically for the detection of harmful insider threats. The system obtains discrete portions of data (such as daily logs) from an uninterrupted flow of data. It then proceeds to develop a new model for each portion, and consistently enhances the ensemble by including the k models that exhibit the lowest prediction error. The findings demonstrated that the OCSVM algorithm outperformed the two-class Support Vector Machine (two-class SVM) in its ability to recognize threats. Moreover, the ensemble-based OCSVM including the updating stream approach exhibits superior performance compared to the OCSVM without updating. The authors then expanded upon their previous research in a subsequent publication [16], whereby they used the ensemble technique to investigate the application of unsupervised Graph-Based Anomaly Detection (GBAD). The ensemble-OCSVM demonstrated superior performance compared to ensemble-GBAD in terms of false positives (FPs).

In the existing procedures, insider hazards were often identified using machine learning techniques. While machine learning techniques are promising for this purpose, they might provide misleading results if the dataset employed is overly biased. The subject of insider threat dataset imbalance has not been adequately or openly addressed in the study. For the sake of parity in the dataset, the conventional method either oversamples examples from underrepresented classes or under-samples instances from overrepresented classes. Since this approach alters the raw data and produces an inaccurate representation of the dataset, it cannot be implemented. To successfully address such data imbalance difficulties, this study provides a machine learning-based anomaly detection technique that deals with the unbalanced classification problem at the algorithm level rather than the data level. The popular DBLOF detection method is used for this purpose. This helps establish if an out-of-the-ordinary incident is malicious or just an anomaly. When the minority class is highly disorganized and made up largely of noisy examples or small disjoints, anomaly detection strategies excel [17]. When compared to existing insider threat detection models, experimental results show that the proposed technique improves insider threat identification performance and provides a superior solution for addressing the imbalance issue of insider threat occurrences.

III. METHODOLOGY

The purpose of this research is to identify instances of insider threats and find solutions to the imbalance issue. The DBLOF detection method is used to evaluate the effectiveness of the machine learning-based anomaly detection model in spotting insider threats throughout a company's network. To do this, we use the methods shown in Figure 1. This section lays out the theoretical underpinnings and key components of the proposed approach for identifying suspicious insider threats. In Section A, we see the procedures that were followed to collect and organize the data. In Section B, some solutions to the problem of skewed datasets are addressed. The popular supervised machine learning algorithms (XGBoost, Random Forest, Decision Tree, and K-Nearest Neighbor) are used as benchmarks against which the proposed model's performance may be fairly evaluated. Then, we use the DBLOF detection method. Results are compared to both the original and modern approaches. The proposed model is shown in its entirety in Figure 1.

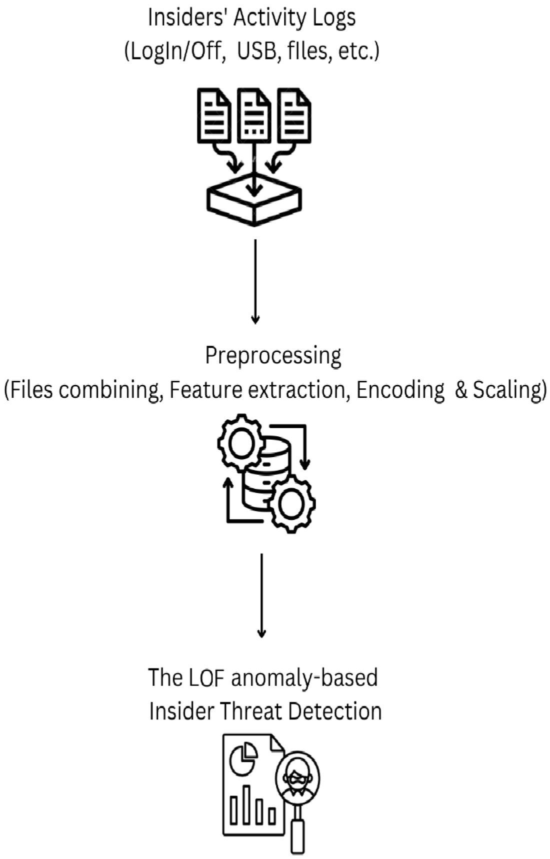

Insiders' Activity Logs
(LogIn/Off, USB, files, etc.)
Preprocessing
(Files combining, Feature extraction, Encoding & Scaling)
The LOF anomaly-based
Insider Threat Detection

FIGURE 1. An overview of the DBLOF insider threat detection technique.

A. DATASET AND PRE-PROCESSING

Data preparation is crucial for making machine learning algorithms more efficient. Experts in the field of cybersecurity may also benefit from this information. Finding an actual insider threat dataset is a major issue for the cybersecurity research community. Because of privacy constraints, there isn't a working dataset with which to test the proposed insider threat detection algorithms. The most widely used and easily available synthetic dataset for evaluating an insider threat detection system is the CERT dataset [18], which has been utilized in several insider threat detection studies such as [19], [20], [21], [22], and [23]. The CERT dataset's creation process is described in great length in [24]. This research used the most widely-used dataset from the scientific literature, CERT r4.2. The collection includes simulated data from 1,000 users, of whom 7% are classified as malevolent. This provides us with more flexibility in conducting anomaly detection tests, which in turn helps us evaluate the efficacy of the proposed model.

The CERT r4.2 insider threat dataset was chosen for validating the DBLOF model for insider threat detection due to its comprehensive and realistic simulation of user behavior within an organizational setting. This particular version of the dataset offers a balanced combination of normal and malicious activities among 1,000 employees, including over 32 million activities and 7,323 malicious instances. This rich and diverse data environment is ideal for testing the DBLOF model's efficiency in detecting outliers that indicate potential insider threats. The dataset's size and complexity provide a robust platform for evaluating the model's sensitivity and specificity, ensuring that the DBLOF model is validated in a scenario that closely simulates real-world conditions. In addition, as stated by the dataset owners, this dataset is characterized as ''dense needle'' and has a significant number of red team situations [25]. The choice of CERT r4.2, with its detailed and varied data, facilitates a comprehensive assessment of the model's ability to detect sophisticated insider threats, making it a suitable choice for this research.

The dataset contains 500 days' worth of data on the daily routines of a thousand employees at a single company. Logs of logins, emails, file transfers, and usage of portable media like USB drives are included. The data utilized from the insider threat dataset is listed as: The logon.csv file stores information about user logins and logouts. The file contains information on computer use, such as the average length of a session, the number of logins that occur outside of normal business hours, and more. This data exposes the user's access patterns; Logs of internal file operations are found in the logon.csv file. Insiders' file access habits may be examined with the use of this information; The use of removable disks on personal computers is shown in the device.csv file via a variety of different actions, such as attaching and removing various devices. The information that the user's day-to-day interactions with linked devices generate is saved in the log file named Device.csv. It discloses the times and dates that a person visited any device, regardless of whether they did so at their place of employment or elsewhere; The email.csv file keeps a record of the exchanges that users have had over email. Additionally, it is made clear if an internal or external actor was responsible for sending or receiving the email; The user's whole browsing history is stored in the http.csv file on their computer. Each and every HTTP record is composed of four different bits of data: the user, the machine, the URL, and the page content. An HTTP log file will keep a record of the addresses of any websites that employees of the firm have accessed while connected to the corporate network.

The CERT r4.2 insider threat dataset shows imbalanced characteristics primarily due to the unequal distribution of instances across different classes of insider threat behaviors. Specifically, within the dataset, instances of normal/benign user activities much outnumber instances of malicious/anomalies insider activities. This imbalance creates a skewed class distribution, where the majority class (normal behavior) dominates the dataset, while the minority class (malicious behavior) is significantly underrepresented. Furthermore, the imbalanced nature of the dataset is reflected by the inherent rarity of insider threats in real-world cybersecurity situations. Malicious insider actions occur infrequently compared to normal user behaviors, resulting in a scarcity of relevant instances for model training and evaluation. Therefore, it is essential to address the imbalance in the dataset to prevent the predictive model from being biased towards the majority class and to ensure effective detection of insider threats. Our research paper focuses on developing a technique to mitigate the imbalanced nature of the CERT r4.2 dataset and enhance the performance of insider threat detection algorithms in such imbalanced cybersecurity settings.

There are a sizable number of procedural stages that come before the creation of the first data set. After that, the feature matrix of the assault scenario is sent to the algorithms that are responsible for anomaly identification. In order to get started, we merged the files of the insider activity logs. These files each comprise various components of the insider threat scenario that we are concentrating on. The next thing we do is further clean up the dataset by using techniques such as extracting features, encoding it, and scaling it. The last phase involves training and verifying the required machine learning algorithms by employing a matrix of characteristics that simultaneously reflects both beneficial and negative insider occurrences.

The several files that included the insider activity logs have been consolidated into one single, comprehensive collection. When a corporation is accumulating logs of user activity from distributed sensors around the organization, it is possible that some of the information may be inaccurate in some manner (for example, because there are gaps in the data). This occurs often due to the fact that the data is acquired from several sensors spread out around the company, not all of which will always function as expected. Because the categorization process places a significant emphasis on the quality of the data, the merged files are analyzed to determine whether or not they include any values deemed null.

After the dataset files have been merged and polished, the characteristics that point to a scenario comprising an insider data leak assault are extracted. This happens after the files have been combined. It would be impractical to include all of the information that is included in the dataset since doing so might reduce the efficiency of the machine learning algorithms that are being used due to the existence of noisy data. The criteria that are used to identify which attributes are most significant have an influence on the efficacy of machine learning algorithms. All of the following are significant:

features, timestamps, user identities, vectors, events (such as logging in and out, connecting and disconnecting USB devices, going to other websites, etc.), and targeting, whether benign or malicious.

The observations used to spot an insider threat are those that are appropriate for the situation. In this piece of research, the potential risk is analyzed through the lens of the situation in which "a user who does not typically use an external USB drive comes to use it more after hours of work and then leaves the organization thereafter." Therefore, some crucial characteristics that represent the assault scenario are the number of times an individual logged in or out of their account outside of normal working hours, the number of times an individual used a USB device, and the number of times an individual visited the WikiLeaks website. The process of performing feature extraction on the collected data leads to the production of numerical vectors, which are commonly referred to as data instances and describe user activities. Each vector is constructed from user input, the vast majority of which consists of category data that is represented numerically for the purpose of providing machine learning algorithms with context.

The preprocessing phase is crucial in ensuring that the data is suitable for effective analysis using the DBLOF algorithm. Given the unique characteristics of the CERT r4.2 dataset, specific preprocessing steps were undertaken to address its inherent challenges, particularly the imbalance issue. Below is the summary of the preprocessing steps accompanied by relevant equations.

1) DATA CLEANING

The Handling Missing Values: Any records with missing values were identified and addressed.

Take the mean of Xi if Xi is continuous.

Take the mode of Xi if Xi is categorical.

Where Mi is the imputed value for the missing data in feature Xi.

2) FEATURE EXTRACTION

Relevant features that provide insights into user behavior, such as login/logout timestamps, file access patterns, and network activity, were extracted from the raw data.

$$F = \{f1, f2, \ldots, fn\} \tag{1}$$

where F represents the set of extracted features.

3) FEATURE SCALING

Given the diverse range of values in the dataset, feature scaling was applied to standardize the feature set.

$$Si = \frac{xi - min(X)}{max(X) - min(X)} \tag{2}$$

where Si is the scaled value of data point xi in feature X.

4) ONE-HOT ENCODING

Categorical features were transformed using one-hot encoding to convert them into a binary matrix. Assign a value of 1 if xi is equal to cj, otherwise assign a value of 0. Where Oij represents the binary value for data point xi corresponding to category cj.

5) DATA PARTITIONING

The dataset was partitioned into training and testing subsets to validate the performance of the DBLOF algorithm.

$$Dtrain = \alpha \times D \tag{3}$$

$$Dtest = (1 - \alpha \times D \tag{4}$$

where D is the entire dataset and α is the proportion allocated to the training set. By meticulously executing these preprocessing steps, the CERT r4.2 dataset was transformed into an optimized format, primed for effective analysis using the DBLOF algorithm. These steps ensured that the inherent challenges of the dataset, particularly its imbalance, were adequately addressed, paving the way for robust insider threat detection.

We firstly implemented a range of algorithms alongside the DBLOF model for insider threat detection to provide a comprehensive comparative analysis, ensuring the robustness and effectiveness of the DBLOF model. The chosen algorithms for comparison include XGBoost (XGB), Random Forest, Decision Trees, and K-Nearest Neighbors (KNN), each offering unique strengths in data analysis and pattern recognition.

XGBoost (XGB) is a highly efficient and scalable implementation of gradient boosting. It is known for its performance in classification tasks and its ability to handle large and complex datasets. In the context of insider threat detection, XGB can effectively capture non-linear relationships and interactions between features. For implementation, the algorithm is trained on the CERT dataset, with hyper-parameters like learning rate, max depth of trees, and the number of estimators tuned to optimize performance. The model's ability to handle imbalanced data, a common challenge in insider threat detection, makes it a valuable tool for comparison with the DBLOF model.

Random Forest is an ensemble learning method that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes of the individual trees. This method is particularly effective due to its ability to reduce overfitting, a common problem in decision tree algorithms. In implementing Random Forest for insider threat detection, numerous trees are trained on different subsets of the data, and their collective decision is used for the final classification. This approach is beneficial for understanding how the DBLOF model compares against ensemble methods in detecting complex patterns of insider threats.

Decision Trees are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. For insider threat detection, a decision tree is built on the CERT dataset, where nodes represent decisions based on different features, and branches represent outcomes of these decisions. This model's simplicity and interpretability make it a useful baseline to compare against the more complex DBLOF model, providing insights into the necessity and effectiveness of more sophisticated approaches in the dataset.

K-Nearest Neighbors (KNN) is a simple, instance-based learning algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been chosen for its simplicity and effectiveness in classification problems. In the context of our study, KNN is used to classify a user's behavior as normal or malicious based on the behaviors of its 'nearest neighbors' in the dataset. The implementation involves selecting the optimal 'k' value that determines the number of neighbors to consider, which is crucial for the model's accuracy. Comparing KNN with the DBLOF model provides insights into the effectiveness of distance-based methods in identifying insider threats.

These algorithms provide a diverse set of approaches for insider threat detection, allowing for a comprehensive evaluation of the DBLOF model's performance in various aspects, such as accuracy, sensitivity to imbalanced data, and computational efficiency. This comparative analysis is crucial for validating the effectiveness of the DBLOF model in a real-world scenario.

B. THE DBLOF MODEL

Outliers and anomalous data are uncommon in comparison to normal data, and they do not fit cleanly into the dataset's distribution. Anomaly detection is the process of identifying outliers in data. The field of machine learning provides techniques for outlier and anomaly detection. It refers to unsupervised machine learning methods that use normal examples to categorize new cases as normal or abnormal. Such techniques may be utilized for binary classification when the distribution of dataset classes is substantially skewed. They are trained on the dataset's majority class's input instances and then tested on a test dataset. The minority class "malicious instances" outnumber the majority class "regular instances" in the CERT insider threat dataset used in this study.

As a consequence of the occurrence of such an extreme class distribution, the minority class might be counted as outliers, causing the classification results to skew. Traditional supervised classification approaches require gathered data to be paired with labels that represent all classes of interest in order to develop a model that can identify all classes. When data labeling is expensive or one class happens considerably less often than others, this strategy may present certain concerns [26]. However, a novel method to avoid such problems is to use one-class classification systems that do not mimic all classes. A model that can identify whether or not a given new item of data belongs to a class is developed using just data from the class of interest in one-class classification approaches. As a result, in this research, we use a one-class or anomaly-based IF algorithm for insider threat identification on CERT insider threat datasets with extremely skewed class distributions. Hence, instead of traditional supervised classification algorithms, anomaly classification algorithms are used. In the next section, we demonstrate the applied anomaly-based IF machine learning algorithm [27]. To our knowledge, the anomaly-based IF algorithm has not been applied to address the severely unbalanced CERT r4.2 dataset when detecting insider data leakage attacks.

The level of abnormality of each observation in a dataset is reflected by the Outlier Factor, a score estimated by the DBLOF [28]. The k-Nearest Neighbors technique is used in the dataset since it operates under the assumption of a local density. Then, a locality is assigned to each data instance, which is used to calculate the density of clusters. First, a hyper-parameter k must be used to determine the number of neighbors. Since a small k has a tighter focus but more errors when dealing with noisy input, choosing an acceptable value for k is essential. The outliers, on the other hand, can all be included in a large k.

The DBLOF technique employs the k-distance function, which measures the distance between an instance and its k-neighbors. The reachability distance (*RD*), which is the highest value between two instances and the k-distance of the second instance, is determined using this k-distance [28].

$$RD(a,b) = max\{k\_distance(b), distance(a,b)\} \tag{5}$$

RD (a, b) is equal to the k-distance of b if an instance of a is one of $b's$ k neighbors; otherwise, it is equal to the actual distance between a and b. The local reachability density, a different function, is counted using this *RD* function (*LRD*). The *RD* of an instance a to all of its k-neighbors (*Nk(a)*) is calculated in order to determine the *LRD* of that instance. The following equation shows how to count the average of the calculated values.

$$LRD\,(a) = \left(\sum\nolimits_{n}^{N_k(a)} \frac{RD\,(a,n)}{k}\right)^{-1} \tag{6}$$

After calculating its *LRD*, each point in the dataset compares it to its k-neighbors, which is measured as the DBLOF, which is the average of those ratios. Equation 5 demonstrates that if the DBLOF is larger than 1, the *LRD* of a point is generally greater than the LRD of its k-neighbors, as shown below.

$$LOF\,(a, N_k\,(a)) = \begin{cases} \approx 1 & Similar\ density\ as\ neighbors \\ > 1 & Lower\ density\ than\ neighbors\ \ (Outlier) \\ < 1 & Higher\ density\ than\ neighburs\ (Inlier) \end{cases} \tag{7}$$

The most well-known density-based local-outlier detection algorithm is the DBLOF [28], which also introduced the idea of local outliers [29]. It only considers a small number of neighbors of dataset instances. Each data instance is assigned a level of being an outlier. Local outliers are anomalies compared to neighborhood densities. For each data instance, the local densities are calculated using the k-nearest neighbors. The DBLOF has the benefit of being able to detect all types of outliers, even those that distance-based algorithms are unable to detect [30]. Additionally, the DBLOF surpasses global outlier approaches in that it can spot local outliers in certain data sets that may not otherwise be recognized as outliers overall [28]. Therefore, we employ it to address the highly imbalanced issue of the CERT dataset for detecting insider data leakage incidents efficiently.

The DBLOF algorithm is an advanced method for outlier detection, particularly effective in datasets with varying densities. It operates on the principle that outliers are data points significantly different in density from their neighbors. In the context of insider threat detection, these outliers represent anomalous user behaviors that deviate from typical patterns. From implementation point of view, we can summarize the implementation of the model in the following steps:

1) PREPARING DATA FOR DBLOF

Before applying DBLOF, we ensure the data is in a format conducive to the algorithm. This involves organizing the CERT dataset into a matrix where each row represents a user's behavior (a data point) and each column represents a feature (like login times, USB usages, etc.). The data should be normalized to ensure that all features contribute equally to the analysis.

2) CALCULATING LOCAL DENSITY

The DBLOF starts by calculating the local density of each data point. This is achieved by measuring the distances between a data point and its neighbors within a specified range. The DBLOF then computes a density score for each point, which reflects how tightly filled its neighbors are around it. In the CERT dataset, this translates to understanding how closely a user's behavior aligns with the behaviors of others in similar roles or departments.

3) DETERMINING THE LOCAL OUTLIER FACTOR

Once the local densities are calculated, DBLOF determines the Local Outlier Factor (LOF) for each data point. The LOF is a ratio of the local density of a point to the densities of its neighbors. A high LOF score indicates that the data point is an outlier. In practical terms, if a user's behavior (data point) has a significantly lower density compared to its neighbors, it suggests anomalous activity.

4) IDENTIFYING OUTLIERS

The DBLOF then identifies outliers based on a threshold LOF score, which can be predefined or dynamically determined based on the dataset's characteristics. In the CERT dataset, users with high LOF scores would be flagged as potential insider threats. These are the individuals whose behavior patterns significantly differ from the norm.

5) CONTEXTUALIZING RESULTS

It's important to note that the DBLOF provides a mathematical identification of outliers, and interpreting these outliers requires context. In the CERT dataset, flagged behaviors are examined in the context of the organization's operational norms, specific roles of the users, and other situational factors.

6) FINE-TUNING DBLOF PARAMETERS

The effectiveness of DBLOF in detecting insider threats can be influenced by its parameters, like the radius for neighbor consideration and the threshold for outlier identification. These parameters can be fine-tuned to optimize the balance between detecting true threats and minimizing false positives.

With the data prepared and relevant features identified, the DBLOF algorithm is applied. This algorithm calculates the local density deviation of each data point (user behavior in this case) with respect to its neighbors. In the context of the CERT dataset, it would assess how much an individual's behavior deviates from the norm within the simulated organization. The DBLOF algorithm is particularly effective in datasets with varying densities, making it suitable for the diverse activities recorded in the CERT dataset. By applying the DBLOF algorithm in this detailed manner, we effectively examine through the CERT 4.2 dataset to detect potential insider threats based on deviations in data density, which signify anomalous insider behaviors. This step is critical in the broader process of leveraging machine learning techniques for cybersecurity purposes.

IV. EXPERIMENTAL EVALUATION

In this section, we demonstrate how we evaluated our predicted model for identifying threats from insiders in organizational networks when anomalies were detected. Our model aims to identify risks coming from inside the organization. This is accomplished through the use of conventional machine learning methods. Following that, we use the DBLOF approach that is based on anomalies, and then we compare our results to both the raw data and the models that were already in place. The procedures for data preparation that are outlined in (Section A, III) are first performed on the CERT r4.2 insider risk dataset [18]. Second, the DBLOF detection approach is used in order to rectify the issue of the classes in the CERT dataset having an inappropriately equal distribution. In the last part of this report, we discuss our analysis of the results of the experiment as well as how they compare to the baseline and other approaches that are currently available.

Python 3.10.4, along with the Scikit-learn module [18], is what's being utilized to put the experiments into action. The Google Colaboratory, sometimes known as Colab for short, is an online service that is made accessible via the Google Cloud platform. It is constructed on Jupyter notebooks.

Because of the many advantages it provides for the facilitation of the usage of code to handle a number of issues [31], businesses are increasingly embracing Google's Colab as an ICT for data science. This is due to the fact that it offers a wide range of benefits. As a direct consequence of this, we run our examinations on Google Cloud. In accordance with the information presented in (Section A, III), we apply many phases of preprocessing to the dataset. These stages include consolidation, management of missing data, polishing, extraction of features, and encoding. After that, the input information feature matrix is sent to the machine learning algorithms. Not only is the proposed model evaluated based on its accuracy, but also on its recall and F-score, which are the two metrics that are used the most commonly. The recall, formerly known as the detection rate or true positive rate, is a representation of the proportion of insider incidents that are properly identified. It is also known as the rate of recall. The F-score is extensively utilized not just because of its expressive capacity but also because it is a notable metric for imbalanced classification [32]. This research addresses the issue of the CERT dataset for detecting insider risks having a severely skewed class distribution, which is a problem that has been brought to light. Figure 2 presents a condensed description of the class distribution, which demonstrates how extremely skewed the dataset is.

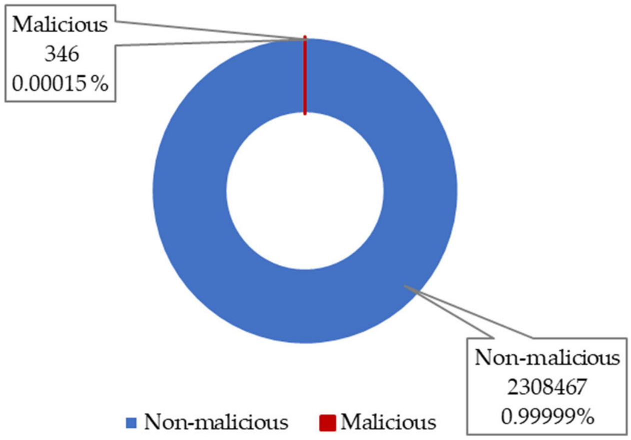


FIGURE 2. The pie chart of how the classes in the CERT insider threat dataset are distributed quite unevenly.

As it is shown in the pie chart, the dataset is extremely imbalanced, so the objective of this study is to develop the DBLOF detection of insider threats model with the intention of resolving the severely unbalanced problem that exists within the CERT dataset.

A. SUPERVISED INSIDER THREAT DETECTION (BASELINE)

To find potentially harmful events that a worker or other insider may have committed, the supervised machine learning algorithms (XGB, RF, DT, and KNN) serve as a baseline. The dataset is used in the process of instructing and fitting an analytical machine learning model. During the assessment phase, the model will learn the pattern of the dataset with the intention of distinguishing between regular events and harmful ones based on data that has not yet been observed. The dataset is divided such that 80% of it will be used for training and 20% will be used for testing. The data set that was utilized to fit the model is commonly referred to as training data, and it includes 1846773 instances of normal data and 277 instances of malicious data. After the model has been trained, it is used to make predictions about previously unknown data, which is referred to as test data. The test data includes 461694 normal cases and 69 malicious instances. In this manner, the constructed model is able to recognize new observations independently of any human involvement. Figure 3 illustrates the outcomes that were achieved via the use of the supervised machine learning algorithms.

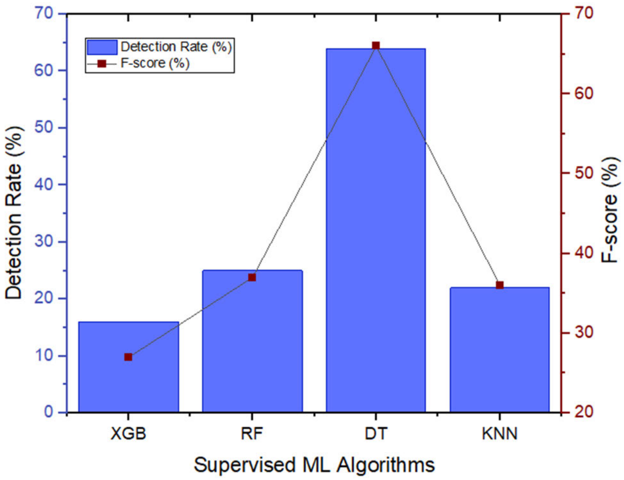


FIGURE 3. The results of applying the supervised machine learning algorithms.

The results in Figure 3 represent a baseline for insider threat detection. The model in the XGB algorithm has a precision of 1.0, indicating that every instance it flagged as an insider threat was actually a threat. This suggests a very conservative model that's careful to avoid false positives. At 0.159, the detection rate is the lowest among the algorithms. This means it misses a significant number of actual threats. The F1 score of 0.274 indicates that the model is not well-balanced in terms of recall and precision. It leans heavily towards precision at the cost of recall. Although the accuracy is extremely high (99.99%), this is misleading due to the likely class imbalance in the data.

The model with the Random Forest gets a precision of 0.708, which is less conservative than XGB algorithm but more balanced. The detection rate of 0.246 is better than XGB algorithm but still leaves room for improvement. At 0.365, the F1 score suggests a more balanced model than XGB algorithm but still not ideal. Similar to XGB algorithm, the high accuracy is likely influenced by class imbalance.

The model in the DT algorithm has the precision of 0.688 which is slightly less than that of Random Forest, indicating a slightly higher chance of false positives. It gets the highest Detection Rate among all at 0.637. This model is better at identifying threats but at the cost of higher false positives. The F1 score is the highest at 0.660, indicating that this model is the most balanced among the four in terms of precision and recall. Again, the high accuracy is likely a result of class imbalance.

The model with the K-nearest neighbors' algorithm likes the XGB algorithm, it has a precision of 1.0 but misses a significant number of actual threats. It obtains a Detection Rate with a rate of 0.217, it's better than XGB algorithm but not by much. It also gets the F1 Score 0f of 0.357, which is almost similar to Random Forest but slightly less balanced. The high accuracy is again likely due to class imbalance.

B. ANOMALY-BASED INSIDER THREAT DETECTION

In this section, the results of applying the DBLOF algorithm for insider threat detection are illustrated. Different results are obtained by tuning the contamination parameter of the algorithm for anomaly detection. We carried out several experiments by setting contamination parameters into different ratios (0.00, 0.02, 0.04, 0.06, 0.08, and 0.1) with the aim of estimating the distribution of real data in an unsupervised manner. The contamination hyper-parameter estimates the anomalies that can be found in the dataset. As it is shown in Figure 4, the contamination parameter with a ratio of 0.02 makes the highest improvement on the accuracy, recall, and F-score detection results with 0.98, 0.98, and 0.99, respectively. This is due to the actual contamination ratio/minority class of the dataset, which is around 0.03 percent. The test set of the dataset contains 461763 instances (461694 instances are normal, and 69 instances are anomalies). The detection results of applying the DBLOF algorithm are shown in Figure 4.

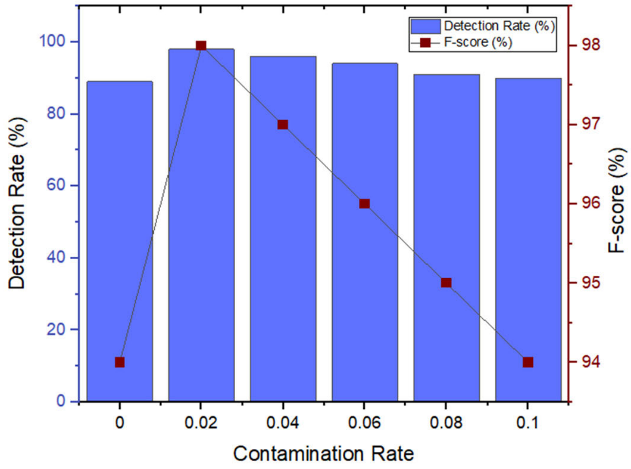


FIGURE 4. The results of applying the DBLOF algorithm.

As shown in Figure 4, the insider threat detection results are different based on the selection of the contamination ratio. At the beginning, the contamination ratio is set to 0.00 as a baseline for the DBLOF model, and it obtains an accuracy and detection rate of 95% and an F-score of 97%. The highest detection results are achieved when the contamination ratio is set to 0.02 with an accuracy and detection rate of 99% and an F-score of 99% due to the contamination ratio being close to the minority class or outliers in the dataset. On the other hand, when the contamination ratio is set to 0.1, the results decrease by about 9%, with an accuracy of 90% and a detection rate and F-score of 95%. It is noticed that when the contamination ratio is increased by 0.2, the accuracy and detection rate decrease by 2%, while the F-score decreases by 1%. For example, when the contamination ratio is increased to 0.04, both the accuracy and detection rate decrease to 96%, and the F-score decreases to 98%. When the contamination ratio is increased to 0.06, both the accuracy and detection rate decrease to 94%, and the F-score decreases to 97%. When the contamination ratio is increased to 0.08, both the accuracy and detection rate decrease to 92%, and the F-score decreases to 96%. This indicates that the contamination parameter has a high impact on the detection results for the DBLOF algorithm. It is observed that in order to improve the insider threat detection results, it is recommended to set the contamination parameter to be around the actual minority class in the dataset. The details of the obtained results are presented in Table 1.

TABLE 1. The results of applying the DBLOF algorithm utilizing various contamination ratios.

Contamination Ratio	Accuracy	Detection Rate	F-score
0.00	0.95	0.95	0.97
0.02	0.98	0.98	0.99
0.04	0.96	0.96	0.98
0.06	0.94	0.94	0.97
0.08	0.92	0.92	0.96
0.1	0.90	0.90	0.95

The LOF algorithm is an unsupervised anomaly detection method. It measures the local density deviation of a data point compared to its neighbors to identify outliers or anomalies. The "Contamination Rate" parameter helps in setting the threshold for what is considered an outlier. A higher contamination rate implies a higher proportion of outliers in the data.

As it is shown in Table 1, in the lower contamination rates (0.00 - 0.02), the model achieves high accuracy and detection rates, ranging from 95% to 98%. The F-score is also notably high (97% to 99%), indicating a balanced model in terms of precision and recall. At these rates, the algorithm is highly conservative. It assumes almost no contamination (anomalies) in the data. This results in high accuracy and detection rates but may miss some less-obvious threats. The high F-score (97% to 99%) suggests that the model is extremely well-fitted to the data. However, this could also indicate a risk of overfitting, especially when the contamination rate is zero.

In the medium contamination rates (0.04 - 0.06), the model's performance starts to decline slightly with accuracy and detection rates around 94% to 96%. The F-score is still relatively high (97% to 98%), suggesting that the model is still balanced. These rates offer a more balanced approach. The algorithm assumes a moderate level of anomalies in the data, which could be more realistic in a typical organizational setting. There is a slight degradation in performance metrics, but they are still robust. The F-score indicates the model remains balanced in terms of precision and recall.

With the higher contamination rates (0.08), we see a more noticeable drop in performance metrics. Accuracy and detection rates go down to 92%, and the F-score decreases to 96%. At this rate, the algorithm is more liberal in flagging activities as anomalies. This could result in more false positives, which is reflected in the lower accuracy and detection rates. A noticeable drop in all metrics suggests that the model starts to lose its efficacy at this contamination level. This could lead to unnecessary investigations or alerts in a real-world application.

V. DISCUSSION

We compare the newly proposed the DBLOF detection algorithm with the baseline model as well as the models that already exist in order to validate the effectiveness of the model. The DBLOF detection algorithm achieved the highest detection rate and F-score compared with the baseline algorithms, with 98% and 99%, respectively. This was in comparison to the baseline, which had a detection rate of 95%. When the contamination parameter of the DBLOF detection algorithm is set to 0.02, close to the real anomalies present in the dataset (0.03), these results are obtained by the algorithm. Regarding the accuracy metric, we found that the baseline algorithms (XGB, RF, DT, and KNN) all achieved an accuracy score of 99%, whereas the suggested DBLOF detection algorithm approach only achieved an accuracy score of 98%. This suggests that the accuracy metric does not provide an accurate measurement of the effectiveness of the detection model in the event that the dataset being utilized is excessively skewed. To put it another way, even when the dataset is severely skewed in one direction or another, the detection rate and the f-score metrics provide an accurate evaluation of the performance of the model.

The subpar results that the baseline algorithms produced are as follows, which make this abundantly clear: The XGB algorithm achieved a detection rate of 15.90% and an f-score of 27.40%; the RF algorithm achieved a detection rate of 24.60% and an f-score of 36.50%; the DT algorithm achieved a detection rate of 63.70% and an f-score of 66.00%; and the KNN algorithm achieved a detection rate of 21.70% and an f-score of 35.70%. The proposed DBLOF detection algorithm, on the other hand, achieved high results, with a detection rate of 98% and an f-score of 98.90%, respectively. The fact that the suggested DBLOF detection algorithm was able to acquire high detection rates (98%) and f-score (99%) indicates that it is resistant to the extremely skewed nature of the dataset. This is in contrast to the baseline algorithms, which were unable to understand the imbalance problem that the dataset provided. Table 2 presents the results of Baseline Algorithms vs. Best DBLOF Configuration.

TABLE 2. The comparison between baseline algorithms and DBLOF algorithm.

Algorithm	Detection Rate	F1 Score
XGB	0.159	0.27
Random Forest	0.246	0.36
Decision Tree	0.637	0.66
K-nearest neighbors	0.217	0.36
Best DBLOF	0.980	0.98

Table 2 shows that the best LOF configuration outperforms all the baseline supervised algorithms in terms of detection rate. It has a detection rate of 0.98, which is significantly higher than even the best-performing baseline algorithm (Decision Tree, 0.637). The F1 Score for the best DBLOF configuration is 0.99, which is also considerably better than the highest F1 Score among the baseline algorithms (Decision Tree, 0.66). This suggests that DBLOF not only identifies more threats but also maintains a good balance between precision and recall. The accuracy for DBLOF is 0.98, which, although slightly lower than the baseline algorithms, is still very high. It's important to note that the extremely high accuracy in the baseline models is likely skewed due to class imbalance, making them less reliable for comparison.

The DBLOF algorithm achieved the highest F-score (99%) due to the advantage of using it, contrary to the other algorithms, which is that only a normal dataset is used for training the model. It indicates that anomaly instances corresponding to malicious insiders are in a low-density region compared

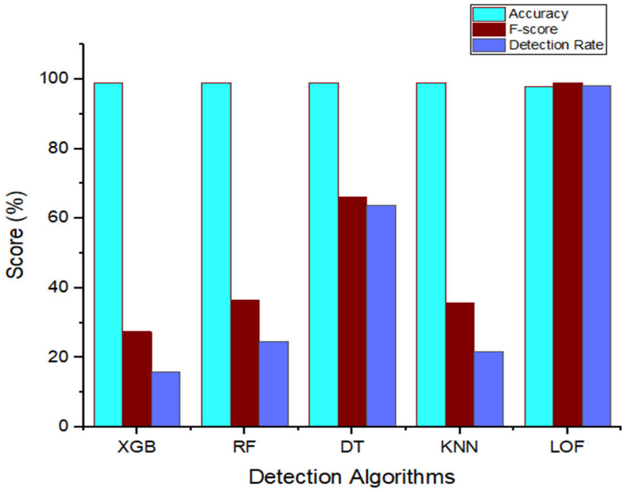


FIGURE 5. Comparison between the baseline algorithms and the proposed DBLOF strategy.

with normal instances and are well distinguished by the DBLOF algorithm. Figure 5 depicts both the results of the baseline techniques (XGB, RF, DT, and KNN) as well as the suggested DBLOF detection approach.

After we compare the proposed method with the baseline, we also compare the insider threat detection results of the proposed DBLOF detection algorithm with the baseline and existing methods that are validated on the same version of the insider threat dataset (CERT r4.2). This is to show that the proposed model is better than what has been done before. Orizio et al. [33] proposed an anomalous insider threat detection model based on constraint network techniques. The model did not consider the highly imbalanced issue of the CERT r4.2 insider threat dataset. The experimental results showed that the model got an accuracy score of 99.91%, a False Positive Rate (FPR) of 0.06%, a precision of 99.84%, and an F-measure of 55.00%. The model proposed by Gayathri et al. [23] utilized a classification approach by applying the Convolutional Neural Networks (CNN). They applied the random under sampling method to address the imbalance issue in the dataset and obtained an accuracy score of 99%, precision of 99.32%, and recall of 99.32%.

In [34], a classification model for insider threat detection was proposed based on the RF. The classification process was applied after balancing the CERT dataset classes using the under/over sampling method. Utilizing the RF classification algorithm and data sampling method, an accuracy score of 98% is obtained. Al-Shehari and Alsowail [35] proposed an insider data leakage model utilizing classification algorithms (LR, DT, RF, NB, KNN) combined with the SMOTE technique to address the imbalance issue of the dataset. This method achieved an F-score for the employed algorithms as (LR: 67%, DT: 99%, RF: 99%, NB: 85%, KNN: 98%). Recently, an insider threat detection model was proposed in [36] by applying the DT and KNN classification algorithms with Under-Sampling, Over-Sampling and Hybrid-Sampling methods for addressing the dataset skewedness issue. The model achieved the following ROC-AUC results: DT with Under-Sampling as 94%; KNN with Over-Sampling as 87%; KNN + Hybrid-Sampling as 87%. In contrast to the existing approaches, in our study we employ the DBLOF detection algorithm that addresses the highly imbalanced issue of the CERT dataset on an algorithmic level using the contamination parameter of the DBLOF detection algorithm. This is to apply the insider threat detection scenario to real data rather than the existing models that manipulate the data using the duplication and removal of dataset instances, which don't reflect real-world insider threat detection. The proposed model achieved a DR of 98% and an F-score of 98.90%. Through our experiments and the obtained results for insider threat detection, we conclude the following observations of the baseline supervised algorithms and the DBLOF algorithm.

As it is shown in Table 3, the model complexity of the baseline supervised algorithms have high F1 scores (max 0.66 for Decision Tree), these algorithms might be creating complex boundaries to fit the data. In the DBLOF algorithm achieved an F1 score of 0.99, suggesting that a simpler model could be just as effective if not more so. The hyper parameter tuning of the supervised algorithms have varying F1 scores (from 0.274 to 0.66) suggest that these models could benefit from further hyper parameter tuning, while in the DBLOF algorithm the F1 score was highest at a contamination rate of 0.02, indicating that even a single hyper parameter like contamination rate can have a significant impact. The validation in the supervised algorithms with high accuracy (around 99.99%) might indicate overfitting if not validated correctly. The high F1 score and detection rate of the DBLOF algorithm, validating on unseen data is crucial to ensure it isn't overfitting. The operational Factors of the supervised algorithms have high accuracy but varied F1 and detection rates mean that the cost of false negatives or positives will vary between models. In the DBLOF algorithm, the high F1 and detection rate at an accuracy of 98% suggest a well-balanced model, but the cost associated with false classifications should be evaluated.

TABLE 3. Experimental observations of the baseline supervised algorithms and proposed DBLOF algorithm.

Consideration	Baseline Algorithms	Proposed DBLOF algorithm
Model Complexity	Higher complexity due to ensemble techniques; more computational resources required.	Lower complexity; generally less computationally intensive.
Hyper-parameter Tuning	Multiple hyper-parameters like tree depth, learning rate; requires rigorous cross-validation.	Fewer key hyper-parameters like the number of neighbors and Contamination Rate; quicker tuning process.
Validation	Requires separate validation set; risk of overfitting if not validated properly.	Validation on unseen data is crucial to avoid overfitting; no need for labeled data.
Operational Factors	Computational cost and interpretability vary; cost-sensitive applications may prefer high precision.	Generally lower computational requirements; interpretability may be lower compared to some supervised models.

The achievement of a remarkable F-score of 98% underscores the effectiveness and reliability of the proposed methodology in detecting insider threats within imbalanced cybersecurity situations. The F-score, which is a harmonic mean of precision and recall, provides a comprehensive measure of the algorithm's performance in accurately detecting both normal and anomalous activities, thereby minimizing false positives and false negatives.

In real-world applications, such high accuracy holds significant implications for enhancing cybersecurity and protecting critical assets against insider threats. There are some potential scenarios where achieving a F-score of 98% is crucial:

- By accurately detecting insider threats with slight false alarms, organizations can proactively prevent malicious activities before they increase, thereby reducing the risk of insider threats.

- The high accuracy of the detection algorithm enables early identification of anomalous behaviors indicative of insider threats, allowing security teams to investigate and respond promptly to suspicious activities, thereby minimizing potential losses.
- Many industries are subject to stringent regulatory requirements mandating the protection of sensitive information and data privacy. Achieving a high F-score demonstrates compliance with regulatory standards and helps organizations to mitigate insider threats effectively.
- Insider threats pose not only financial and operational risks but also undermine organizational trust and reputation. By maintaining a high level of accuracy in insider threat detection, organizations can maintain trust among their clients.
- In sectors such as healthcare, finance, and energy, insider threats can have severe consequences on public safety, national security, and economic stability. A robust detection algorithm with a F-score of 98% plays a vital role in protecting critical infrastructure and minimizing disruptions to essential services.

Therefore, achieving a remarkable F-score of 98% signifies the effectiveness and reliability of the proposed insider threat detection methodology in real-world cybersecurity settings. By minimizing false positives and false negatives, the algorithm enables organizations to proactively identify and mitigate insider threats, thereby enhancing cybersecurity and protecting critical assets against potential risks and vulnerabilities.

A. LIMITATION AND FUTURE WORK

The DBLOF model for insider threat detection faces several key limitations that necessitate future research and development. These limitations include a high sensitivity to the 'contamination rate' hyper parameter, requiring advanced optimization techniques; concerns about the model's generalizability and risk of overfitting, calling for rigorous validation methods like k-fold cross-validation; and issues related to interpretability and scalability, which can impact organizational adoption. To address these challenges, future work should focus on enhancing model interpretability through methods like Local Interpretable Model-agnostic Explanations (LIME), testing scalability under diverse conditions, and conducting a comprehensive cost-benefit analysis that accounts for the financial and operational implications of false classifications. Table 4 summarizes the key limitations of the DBLOF model for insider threat detection and outlines potential avenues for future work to address these limitations.

Table 4 outlines critical limitations and corresponding future work avenues for the DBLOF model in insider threat detection. One of the major limitations is the model's sensitivity to the 'contamination rate' hyper-parameter, which calls for advanced optimization techniques like grid search or Bayesian optimization to ensure robust performance. Additionally, there are concerns about the model's generalizability and risk of overfitting, which could be mitigated by implementing rigorous validation techniques such as k-fold cross-validation. The model's lack of straightforward interpretability and its untested scalability in high-dimensional or real-time scenarios add layers of complexity to its potential deployment in organizational settings.

TABLE 4. The critical limitations and corresponding future work avenues for the DBLOF model in insider threat detection.

Consideration	Limitation	Points for Future Work
Hyper-parameter Sensitivity	Efficacy tightly coupled with 'contamination rate'; may not generalize well.	Utilize techniques like grid search or Bayesian optimization for hyperparameter tuning.
Validation Concerns	Generalizability to other datasets or real-world scenarios remains untested; risk of overfitting.	Implement k-fold cross-validation or time-based validation to assess model's robustness.
Interpretability	May lack straightforward interpretability in terms of feature importance or decision rationale.	Investigate methods like LIME to improve model interpretability.
Scalability	Scalability in high-dimensional or real-time scenarios is unclear.	Test the model's scalability under different conditions, including high-dimensional and real-time data.
Cost Analysis	Lacks in-depth cost-benefit analysis considering financial and operational implications.	Conduct a thorough financial and operational analysis considering false positives and negatives.

Regarding future work, the table suggests several strategic directions. For example, improving model interpretability is critical for organizational adoption, and methods like LIME could be explored to that end. Similarly, assessing the model's scalability under varied conditions can provide insights into its real-world applicability. Given that the model's current evaluation lacks a comprehensive cost-benefit analysis, future studies should also focus on financial and operational implications, particularly considering the costs associated with false positives and negatives. This will provide a holistic view of the model's effectiveness, facilitating more informed decision-making for its deployment.

VI. CONCLUSION

Insider threats pose a significant risk to networked systems in both business and public sector environments. Addressing this challenge is compounded by the presence of highly unbalanced datasets and a lack of comprehensive ground truth. To tackle these issues, we developed the DBLOF model, leveraging the CERT r4.2 insider threat dataset. Our model excels in identifying malicious insider activities by treating them as outliers within the data landscape, effectively separating them from benign activities. A thorough adjustment of the contamination hyper-parameter led to an optimized performance, achieving a 98% detection rate and a 98.9% F1 score at a contamination rate of 0.02. Compared to existing methodologies, the DBLOF model stands out for its efficacy in a variety of contexts and its ability to handle skewed datasets efficiently. The model is highly tuned to the contamination rate, necessitating cautious selection for applicability across different datasets. The model has been validated on a specific dataset (CERT r4.2), and its ability to generalize to other data is not yet verified, raising concerns about overfitting. While effective, the model's lack of interpretability and untested scalability could hinder its broader implementation, especially in real-time or high-dimensional data scenarios. The study lacks a comprehensive cost-benefit analysis, particularly the economic and operational ramifications of false positives and false negatives. Future research should focus on utilizing advanced hyper-parameter optimization techniques like grid search or Bayesian optimization.

To establish the model's robustness, it is imperative to perform cross-dataset validation and k-fold cross-validation. Methods like the LIME could be employed to improve the model's performance. The model's efficiency in high-dimensional and real-time environments should be investigated to assess its scalability. A detailed analysis considering the financial and operational implications of false classifications is crucial for a more rounded evaluation of the model's utility. The DBLOF model demonstrates promising capabilities for insider threat detection, but addressing these limitations and integrating the suggested future works can significantly enhance its robustness and applicability.

REFERENCES

[1] Informa. (2022). *Greatest Threat*. Accessed: Dec. 26, 2022. [Online]. Available: https://www.darkreading.com/vulnerabilities—threats/greatest-threat/d/d-id/1269416

[2] A. Azaria, A. Richardson, S. Kraus, and V. S. Subrahmanian, "Behavioral analysis of insider threat: A survey and bootstrapped prediction in imbalanced data," *IEEE Trans. Computat. Social Syst.*, vol. 1, no. 2, pp. 135–155, Jun. 2014, doi: 10.1109/TCSS.2014.2377811.

[3] S. Yuan and X. Wu, "Deep learning for insider threat detection: Review, challenges and opportunities," *Comput. Secur.*, vol. 104, May 2021, Art. no. 102221, doi: 10.1016/j.cose.2021.102221.

[4] A. Fernández, V. López, M. Galar, M. J. del Jesus, and F. Herrera, "Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches," *Knowl.-Based Syst.*, vol. 42, pp. 97–110, Apr. 2013, doi: 10.1016/j.knosys.2013.01.018.

[5] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[6] R. A. Alsowail and T. Al-Shehari, "Empirical detection techniques of insider threat incidents," *IEEE Access*, vol. 8, pp. 78385–78402, 2020, doi: 10.1109/ACCESS.2020.2989739.

[7] I. Homoliak, F. Toffalini, J. Guarnizo, Y. Elovici, and M. Ochoa, "Insight into insiders and it: A survey of insider threat taxonomies, analysis, modeling, and countermeasures," *ACM Comput. Surv.*, vol. 52, no. 2, pp. 1–40, 2018.

[8] H. Goldberg, W. Young, M. Reardon, B. Phillips, and T. Senator, "Insider threat detection in PRODIGAL," in *Proc. 50th Hawaii Int. Conf. Syst. Sci.*, 2017, pp. 1–10, doi: 10.24251/hicss.2017.320.

[9] T. E. Senator, H. G. Goldberg, A. Memory, W. T. Young, B. Rees, R. Pierce, D. Huang, M. Reardon, D. A. Bader, E. Chow, and I. Essa, "Detecting insider threats in a real corporate database of computer usage activity," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, Aug. 2013, pp. 1393–1401, doi: 10.1145/2487575.2488213.

[10] H. Eldardiry, E. Bart, J. Liu, J. Hanley, B. Price, and O. Brdiczka, "Multi-domain information fusion for insider threat detection," in *Proc. IEEE CS Secur. Privacy Workshops*, 2013, pp. 45–51.

[11] G. Gavai, K. Sricharan, D. Gunning, R. Rolleston, J. Hanley, and M. Singhal, "Detecting insider threat from enterprise social and online activity data," in *Proc. 7th ACM CCS Int. Workshop Managing Insider Security Threats*, New York, NY, USA, Oct. 2015, pp. 13–20, doi: 10.1145/2808783.2808784.

[12] T. Rashid, I. Agrafiotis, and J. R. C. Nurse, "A new take on detecting insider threats: Exploring the use of Hidden Markov Models," in *Proc. Int. Workshop Manag. Insider Secur. Threats, Co-Located*, New York, NY, USA, Oct. 2016, pp. 47–56, doi: 10.1145/2995959.2995964.

[13] D. C. Le and A. N. Zincir-Heywood, "Evaluating insider threat detection workflow using supervised and unsupervised learning," in *Proc. IEEE Secur. Privacy Workshops (SPW)*, San Francisco, CA, USA, May 2018, pp. 270–275, doi: 10.1109/SPW.2018.00043.

[14] P. A. Legg, O. Buckley, M. Goldsmith, and S. Creese, "Automated insider threat detection system using user and role-based profile assessment," *IEEE Syst. J.*, vol. 11, no. 2, pp. 503–512, Jun. 2017, doi: 10.1109/JSYST.2015.2438442.

[15] P. Parveen, Z. R. Weger, B. Thuraisingham, K. Hamlen, and L. Khan, "Supervised learning for insider threat detection using stream mining," in *Proc. IEEE 23rd Int. Conf. Tools Artif. Intell.*, Nov. 2011, pp. 1032–1039, doi: 10.1109/ICTAI.2011.176.

[16] P. Parveen, N. Mcdaniel, Z. Weger, J. Evans, B. Thuraisingham, K. Hamlen, and L. Khan, "Evolving insider threat detection stream mining perspective," *Int. J. Artif. Intell. Tools*, vol. 22, no. 5, Oct. 2013, Art. no. 1360013, doi: 10.1142/s0218213013600130.

[17] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, *Learning From Imbalanced Data Sets*. Cham, Switzerland: Springer, 2018, doi: 10.1007/978-3-319-98074-4.

[18] CERT and ExactData LLC. (2020). *Insider Threat Test Dataset*. Software Engineering Institute, Carnegie Mellon University. Accessed: Sep. 14, 2021. [Online]. Available: https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=508099

[19] M. N. Al-Mhiqani, R. Ahmed, Z. Zainal, and S. N. Isnin, "An integrated imbalanced learning and deep neural network model for insider threat detection," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 1, 2021, doi: 10.14569/ijacsa.2021.0120166.

[20] D. C. Le and N. Zincir-Heywood, "Exploring anomalous behaviour detection and classification for insider threat identification," *Int. J. Netw. Manag.*, vol. 31, no. 4, p. e2109, Jul. 2021, doi: 10.1002/nem.2109.

[21] D. C. Le, N. Zincir-Heywood, and M. I. Heywood, "Analyzing data granularity levels for insider threat detection using machine learning," *IEEE Trans. Netw. Service Manage.*, vol. 17, no. 1, pp. 30–44, Mar. 2020, doi: 10.1109/TNSM.2020.2967721.

[22] D. C. Le and N. Zincir-Heywood, "Anomaly detection for insider threats using unsupervised ensembles," *IEEE Trans. Netw. Service Manag.*, vol. 18, no. 2, pp. 1152–1164, Jun. 2021, doi: 10.1109/TNSM.2021.3071928.

[23] R. G. Gayathri, A. Sajjanhar, and Y. Xiang, "Image-based feature representation for insider threat classification," *Appl. Sci.*, vol. 10, no. 14, p. 4945, Jul. 2020, doi: 10.3390/app10144945.

[24] J. Glasser and B. Lindauer, "Bridging the gap: A pragmatic approach to generating insider threat data," in *Proc. IEEE CS Secur. Privacy Workshops*, May 2013, pp. 98–104, doi: 10.1109/SPW.2013.37.

[25] A. Gamachchi, L. Sun, and S. Boztas, "A graph based framework for malicious insider threat detection," 2018, *arXiv:1809.00141*.

[26] S. S. Khan and M. G. Madden, "A survey of recent trends in one class classification," in *Proc. Irish Conf. Artif. Intell. Cogn. Sci.* (Lecture Notes in Computer Science), 2010, pp. 188–197, doi: 10.1007/978-3-642-17080-5_21.

[27] S. Hariri, M. C. Kind, and R. J. Brunner, "Extended isolation forest," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 4, pp. 1479–1489, Apr. 2021, doi: 10.1109/TKDE.2019.2947676.

[28] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," *ACM SIGMOD Rec.*, vol. 29, no. 2, pp. 93–104, Jun. 2000, doi: 10.1145/335191.335388.

[29] M. Goldstein and S. Uchida, "A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data," *PLoS ONE*, vol. 11, no. 4, Apr. 2016, Art. no. e0152173, doi: 10.1371/journal.pone.0152173.

[30] L. Ertoz, E. Eilertson, A. Lazarevic, P. N. Tan, V. Kumar, J. Srivastava, and P. Dokas, "Minds-minnesota intrusion detection system," in *Next Generation Data Mining*, 2004.

[31] A. Y. Wang, A. Mittal, C. Brooks, and S. Oney, "How data scientists use computational notebooks for real-time collaboration," *Proc. ACM Hum.-Comput. Interact.*, vol. 3, pp. 1–30, Nov. 2019, doi: 10.1145/3359141.

[32] C. Ferri, J. Hernández-Orallo, and R. Modroiu, "An experimental comparison of performance measures for classification," *Pattern Recognit. Lett.*, vol. 30, no. 1, pp. 27–38, Jan. 2009, doi: 10.1016/j.patrec.2008.08.010.

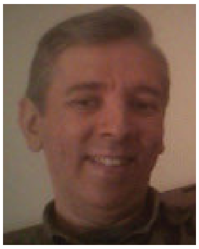
[33] R. Orizio, S. Vuppala, S. Basagiannis, and G. Provan, "Towards an explainable approach for insider threat detection: Constraint network learning," in *Proc. Int. Conf. Intell. Data Sci. Technol. Appl. (IDSTA)*, Oct. 2020, pp. 42–49, doi: 10.1109/IDSTA50958.2020.9264049.

[34] D. Noever, "Classifier suites for insider threat detection," 2019, *arXiv:1901.10948*.

[35] T. Al-Shehari and R. A. Alsowail, "An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques," *Entropy*, vol. 23, no. 10, p. 1258, Sep. 2021, doi: 10.3390/e23101258.

[36] T. Al-Shehari and R. A. Alsowail, "Random resampling algorithms for addressing the imbalanced dataset classes in insider threat detection," *Int. J. Inf. Secur.*, vol. 22, no. 3, pp. 611–629, Jun. 2023, doi: 10.1007/s10207-022-00651-1.

TAHER AL-SHEHARI received the B.S. degree in computer science from King Khalid University, Saudi Arabia, in 2007, and the M.S. degree in computer science from the King Fahd University of Petroleum and Minerals (KFUPM), in 2014. From 2011 to 2014, he was a Research Assistant with KFUPM. Since 2015, he has been a Senior Lecturer and a Researcher with King Saud University. He is the author of several articles that are published in prestige journals. His research interests include information security and privacy, insider threat detection and prevention systems, machine learning models, and data analysis. His awards and honors include the Honor Award from King Khalid University's Rector and the Best Designed Curriculum Award from CFY's Dean, KSU.



DOMENICO ROSACI received the Ph.D. degree in electronic engineering, in 1999. He is currently an Associate Professor of computer science with the Department of Information, Infrastructures and Sustainable Energy Engineering, Mediterranea University of Reggio Calabria, Italy. His research interests include distributed artificial intelligence, multiagent systems, trust, and reputation in social communities. He is a member of a number of conference PCs. He is an Associate Editor of the *Journal of Universal Computer Science* (Springer).

MUNA AL-RAZGAN received the Ph.D. degree in information technology from George Mason University, Fairfax, VA, USA. She is currently an Associate Professor in software engineering with the College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. Her research interests include data mining, machine learning, artificial intelligence, educational data mining, and assistive technologies.

TAHA ALFAKIH received the B.S. degree in computer science from the Department of Computer Science, Hadhramout University, Yemen, the M.Sc. degree from the Department of Computer Science, King Saud University (KSU), Riyadh, Saudi Arabia, and the Ph.D. degree from the Department of Information System, KSU. He is currently a Researcher with the College of Computer Science, KSU. His research interests include machine learning, mobile edge computing, and the Internet of Things (IoT).

MOHAMMED KADRIE received the B.S. degree in mathematics and informatics from the Faculty of Sciences, University of Aleppo, Aleppo, Syria, in 1990, the D.E.A. degree in informatics from the ESSI High School, University of Nice Sophia Antipolis (UNSA), Nice, French, in 1998, and the Ph.D. degree in informatics from UNSA, in 2002. From 2010 and 2015, he was a Teacher in informatics with the Teachers Colleges, King Saud University, Riyadh, Saudi Arabia. His research interests include the theory of codes, automata, and artificial intelligence.

HAMMAD AFZAL (Senior Member, IEEE) received the M.Sc. degree (Hons.) in advanced computing science from the Department of Informatics, The University of Manchester, U.K., and the Ph.D. degree from the School of Computer Science, The University of Manchester. He was a Research Intern with the Insight Center for Data Analytics, University of Ireland, Galway. He is currently a Professor and also heading "The Center of Data and Text Engineering and Mining" (CoDTeEM), a research center with the National University of Sciences and Technology (NUST), Pakistan. He was awarded the Program Prize of the Year from the Department of Computation, The University of Manchester, for acquiring the highest grades in the M.S. courses. He has also attained several awards, such as the Best Researcher with MCS, NUST, in 2019, the Peter Jones Prize, in 2005, and attained the highest grades in all the M.Sc. courses (Equivalent to Gold Medal).

RAHEEL NAWAZ is currently a Pro-Vice-Chancellor with Staffordshire University, U.K. He is also a leading Researcher in artificial intelligence and digital education. He holds several adjunct professorships and scientific directorships across Asia and North America. He sits on the boards of research and charitable organizations, such as the National Centre for Artificial Intelligence, Pakistan; TechSkills, U.K.; and NTF, U.K. He has advised national policy organizations, including the Prime Minister's Task Force on Science and Technology, Pakistan. He has authored over 150 peer-reviewed research articles and his career grant capture stands at over 14 million. He has graduated 19 Ph.D. students so far. According to Google Scholar, he is among the top 10 most cited scholars in the world in the fields of digital transformations, applied artificial intelligence, and educational data science.