

# **ANOMALY DETECTION USING GENERATIVE ADVERSARIAL NETWORKS**

**Thesis Submitted  
in Partial Fulfillment of the Requirements for the  
Degree of**

**MASTER OF TECHNOLOGY**

**in**

**ARTIFICIAL INTELLIGENCE**

**by**

**SHIKHAR ASTHANA**

**(2K22/AFI/24)**

**Under the Supervision of**

**Dr. ANURAG GOEL**

**Assistant Professor, Department of Computer Science and Engineering  
Delhi Technological University**



**Department of Computer Science and Engineering**

**DELHI TECHNOLOGICAL UNIVERSITY**

**(Formerly Delhi College of Engineering)**

**Bawana Road, Delhi 110042**

**June, 2024**

**DELHI TECHNOLOGICAL UNIVERSITY**  
(Formerly Delhi College of Engineering)  
Bawana Road, Delhi-110042

**ACKNOWLEDGMENT**

I wish to express my sincerest gratitude to **Dr. Anurag Goel** for his continuous guidance and mentorship that he provided during research work. He showed me the path to achieving targets by explaining all the tasks to be done and explained to me the importance of this work as well as its industrial relevance. He was always ready to help me and clear our doubts regarding any hurdles in this project. Without his constant support and motivation, this work would not have been successful.

**Place: Delhi**

**SHIKHAR ASTHANA**

**Date:**

**DELHI TECHNOLOGICAL UNIVERSITY**  
(Formerly Delhi College of Engineering)  
Bawana Road, Delhi-110042

**CANDIDATE'S DECLARATION**

I, **Shikhar Asthana 2K22/AFI/24**, of **M.Tech. (AI)**, hereby certify that the work which is being presented in the thesis entitled “**Anomaly Detection Using Generative Adversarial Networks**” in partial fulfillment of the requirement for the award of the degree of Master of Technology in Artificial Intelligence, submitted in the Department of Computer Science and Engineering, Delhi Technological University is an authentic record of my own work carried out during the period from to under the supervision of Dr. Anurag Goel.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other institute.

**Candidate's Signature**

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis and the statement made by the candidate is correct to the best of our knowledge.

**Signature of Supervisor**

**Signature of External Examiner**

**DELHI TECHNOLOGICAL UNIVERSITY**  
(Formerly Delhi College of Engineering)  
Bawana Road, Delhi-110042

**CERTIFICATE BY THE SUPERVISOR**

Certified that **Shikhar Asthana (2K22/AFI/24)** has carried out their research work presented in this thesis entitled “**Anomaly Detection Using Generative Adversarial Networks**” for the award of **Master of Technology in Artificial Intelligence** from the department of Computer Science and Engineering, Delhi Technological University, Delhi under my supervision. The thesis embodies results of original work, and studies are carried out by the student himself and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

(Dr.ANURAG GOEL)

(Assistant Professor)

(Department of Computer Science and Engineering)

(Delhi Technological University)

Date:

# **ANOMALY DETECTION USING GENERATIVE ADVERSARIAL NETWORKS**

## **SHIKHAR ASTHANA**

### **ABSTRACT**

Anomaly detection (AD) has emerged as a critical application across various domains, especially where identifying abnormal behaviors or events is crucial. The advent of deep learning techniques has significantly advanced AD methods, enabling the handling of complex and high-dimensional data. However, these advancements pose the challenge of explainability, requiring approaches that address the 'black box' nature of deep learning models. This thesis builds upon a comprehensive review of recent AD techniques, emphasizing their explainability within the realm of Explainable AI (XAI). Key insights include the importance of interpretability in AD systems, the versatility of deep learning architectures, and emerging trends such as graph-based AD using deep learning.

Building on this theoretical foundation, the thesis also explores practical enhancements through the implementation of the Skip-GANomaly model with novel modifications to its loss function, incorporating contrastive learning to improve semi-supervised AD. Contrastive learning involves training a model to distinguish between positive and negative sample pairs, leading to robust representation learning. Experimental results demonstrate that these modifications yield significant performance improvements across various datasets. By integrating a thorough exploration of XAI in AD and proposing an effective semi-supervised AD approach, this thesis aims to advance the field, providing valuable insights and paving the way for future research and applications.

## LIST OF PUBLICATIONS

- Shikhar Asthana and Anurag Goel, “Unveiling Anomalies: A Review of Anomaly Detection Through Lens of Explainable AI” accepted to be published in “3rd International Conference on ‘Smart Technologies and Systems for Next Generation Computing’” to be held on July 18 - 19, 2024 at IFET College Of Engineering, Tamil Nadu, India.
- Shikhar Asthana and Anurag Goel, “ConGANomaly: A Contrastive Learning Approach Of Anomaly Detection Using Generative Adversarial Networks” accepted to be published in “The 15th International IEEE Conference on Computing, Communication and Networking Technologies (ICCCNT)” to be held on June 24 - 28, 2024 at IIT - Mandi, Himachal Pradesh, India.

# TABLE OF CONTENTS

<b>Acknowledgement</b>	<b>ii</b>
<b>Candidate’s Declaration</b>	<b>iii</b>
<b>Certificate by the Supervisor</b>	<b>iv</b>
<b>Abstract</b>	<b>v</b>
<b>List of Publications</b>	<b>vi</b>
<b>Table of Contents</b>	<b>viii</b>
<b>List of Tables</b>	<b>ix</b>
<b>List of Figures</b>	<b>x</b>
<b>List of Symbols, Abbreviations</b>	<b>xi</b>
<b>1 INTRODUCTION</b>	<b>1</b>
1.1 Problem Statement and Objective . . . . .	2
1.1.1 Comprehensive Review of AD Techniques Using XAI . . . . .	2
1.1.2 Enhancing AD Using GAN-Based Model . . . . .	3
1.2 Motivation . . . . .	3
<b>2 LITERATURE REVIEW</b>	<b>5</b>
2.1 Rise of Autoencoders . . . . .	5
2.2 Adversarial Autoencoders . . . . .	6
2.3 GANomaly . . . . .	6
2.4 Skip-GANomaly . . . . .	6
2.5 Contrastive Learning . . . . .	7
2.6 SOTA Techniques for Anomaly Detection . . . . .	7
2.6.1 Traditional Anomaly Detection Methods . . . . .	8
2.6.2 Deep Learning Anomaly Detection Methods . . . . .	9
2.6.3 Graph Based Anomaly Detection Methods . . . . .	11
2.7 Explainable AI for Anomaly Detection . . . . .	12
2.8 SOTA Techniques Through Lens of Explainable AI . . . . .	13
<b>3 METHODOLOGY</b>	<b>16</b>
3.1 Why Choose GAN Based Model? . . . . .	16

3.2	Proposed Approach . . . . .	17
3.2.1	Generator Sub-Network . . . . .	18
3.2.2	Discriminator Sub-Network . . . . .	18
3.2.3	Individual Loss Functions . . . . .	19
3.2.4	Contrastive Loss . . . . .	20
3.3	Experimental Setup . . . . .	22
3.3.1	Datasets . . . . .	22
3.3.2	Setup Configurations . . . . .	23
<b>4</b>	<b>RESULTS &amp; ANALYSIS</b>	<b>24</b>
4.1	Results Obtained . . . . .	24
4.2	Analyzing Challenges Faced . . . . .	26
<b>5</b>	<b>CONCLUSION, FUTURE SCOPE AND SOCIAL IMPACT</b>	<b>27</b>
5.1	Conclusion . . . . .	27
5.2	Future Scope and Social Impact . . . . .	28
	<b>References</b>	<b>29</b>
	<b>Proof of Publishing</b>	
	<b>Plagiarism Report</b>	
	<b>Curriculum Vitae</b>	



## List of Tables

2.1	AD Approaches through Explainable AI . . . . .	14
4.1	AUC Results Overview . . . . .	24
4.2	AUC results for UBA and FFOB dataset . . . . .	25

## List of Figures

2.1	Contrastive Learning Intuition . . . . .	7
3.1	Proposed Architecture with Contrastive Learning Data Pairings . . . . .	18
3.2	Dataset Sample Images - (a) MNIST (b) CIFAR-10 (c) UBA (d) FFOB . . . . .	23
4.1	MNIST Performance Comparison . . . . .	25
4.2	CIFAR-10 Performance Comparison . . . . .	25

## List of Symbols

$x_i$	$i^{th}$ element/instance
$x'$	Generated instance of x
$\bar{x}$	Mean of x instances
$\hat{y}_i$	Predicted $i^{th}$ instance
$z_i$	Z Score of the $i^{th}$ element/instance, Bottleneck features
$\mu$	Mean
$\sigma$	Standard deviation
$s$	Expected deviation
$\sum_{i=1}^N$	Summation from $i = 1$ to $i = N$
$\mathbb{E}$	Error function
$L_{xyz}$	Loss function of xyz approach
$\lambda$	Learning rate, Balancing factor, Weight factor, Regularisation parameter
$T_c$	Temperature constant, Correct class identifier
$v_c$	Length of vector capsule
$m^+, m^-$	Margins
$W_i$	$i^{th}$ weightage parameter, Temporal smoothness
$\gamma$	Weight factor
$\alpha$	Learning rate, Hyperparameter, Weight factor
$E_A, E_S$	Supervised Loss, Auxillary/Regularization Loss
$P_i$	$i^{th}$ pairing
$G()$	Generator function
$C_k$	$k^{th}$ class of dataset
$\varepsilon$	Belongs to operator
$\tau$	Temperature factor
$sim()$	Similarity function

# Chapter 1

## INTRODUCTION

The ever-changing and dynamic world is full of patterns and sequences waiting to be detected. With these comes the opportunity to observe and detect anomalies in these patterns which play a very vital role in the world. The field of Artificial Intelligence (AI) which is concerned with the identification and observation of detecting anomalies is called anomaly detection (AD). AD often is also coupled with a great opportunity for business or industrial impacts like – pre-emptive detection of potential threats or irregularities, better planning for inventory-based operations, and many others, which are key in domains such as cybersecurity, finance, healthcare, and industrial systems [1, 2, 3, 4, 5]. Modern datasets have increasing levels of complexity and dimensionality which are problem areas for the traditional AD approaches, which leads to a growing interest in finding ways to leverage deep learning techniques for more effective AD. However, the widespread adoption of deep learning approaches in AD raises concerns about the interpretability and transparency of these models, often tokenized as a "black box" problem. As a result, there is a pressing need for research development, innovations and development efforts centered around improving the interpretability of deep learning-based AD systems. In recent years, the landscape has also witnessed the remarkable efficacy of generative models [6], particularly GANomaly [7] and Skip-GANomaly [8], in the domain of semi-supervised AD. While these models have demonstrated significant success, the research work presented in this thesis endeavors to amplify their capabilities by seamlessly integrating contrastive learning into their frameworks. The fusion of the inherent strengths of generative adversarial networks (GANs) and contrastive learning [9, 10] is poised to empower the models to discern subtle distinctions between normal and anomalous instances, contributing to the continuous evolution of robust anomaly detection methodologies.

As part of the introduction, it is essential to grasp a brief overview of traditional methods in anomaly detection (AD). Traditional AD methods predominantly relied on statistical-based techniques, encompassing clustering, density estimation, distance-based methods, and others, to identify instances significantly deviating from the majority of the data [11]. While these approaches offer high explainability, being easy to interpret and understand, they exhibit limitations in coping with the increasing complexity and high dimensionality of modern datasets prevalent in industries. In such scenarios, where anomalies may manifest subtly or in complex patterns, traditional methods may falter. Moreover, traditional AD often includes a lot of extensive manual

interventions in the machine learning pipeline, ranging from data pre-processing and annotation to algorithm selection and hyperparameter tuning. These manual interventions introduce the possibility of human errors, potentially affecting AD performance.

In recent times, with the trifactor rise in data, computational power, and better deep learning algorithms, the industry has seen a huge boost in the adoption of deep learning techniques and how these techniques have revolutionized the field of AD [12]. This revolution was brought by offering more sophisticated and data-driven approaches to scrutinize and search out anomalies in complex datasets [13]. Deep learning models, over traditional AD methods, have demonstrated, in multiple scenarios, how remarkable their capabilities are in automatically learning from unprocessed data, the hierarchical based representations, allowing them to seize baroque patterns and relationships that may not be apparent if viewed from the lenses of the traditional methods [14]. These techniques are also more versatile in terms of data modalities on which AD can be performed – including images, time series, text, and graph-structured data.

Real-world applications of social networks, biological networks, cybersecurity networks, and other such emerging fields have garnered considerable attention recently, which brings with it a mounting heap of graph-structured data and a crucial need for specialized AD methodology for these datasets – Graph Anomaly Detection (GAD), a specialized area within AD is the answer [15]. The reason for a specialized approach is because traditional AD techniques fail to consider inherent structural information and relationships among entities which causes them to ineffectively handle graph data. Graph Neural Networks (GNN), a deep learning based approach, have opened up new avenues for AD in graph-based data [16]. GNNs can recognize complex structural dependencies and node interactions in graphs, making them well-suited for AD tasks in graph-structured data. This intersection of deep learning and graph anomaly detection holds great promise for addressing challenges in anomaly detection in complex networked systems.

## **1.1 Problem Statement and Objective**

Anomaly detection (AD) is a quintessential application in numerous domains where identifying abnormal behaviors or events is crucial for maintaining security, safety, and operational efficiency. The increasing complexity and high dimensionality of data have necessitated the adoption of advanced deep learning techniques, which, despite their effectiveness, often suffer from a lack of transparency and interpretability. This dual-fold problem statement addresses two significant challenges in the field of anomaly detection: the need for explainable AI (XAI) in AD techniques and the enhancement of AD models using generative adversarial networks (GANs). Let us explore these in the below subsections.

### **1.1.1 Comprehensive Review of AD Techniques Using XAI**

The first part of the problem addresses the challenge of explainability in anomaly detection methods. While deep learning techniques have significantly advanced all

the capabilities of AD, their 'black box' nature poses a substantial barrier to their widespread adoption and trust. There is a critical need for a comprehensive review of current AD techniques with a focus on their explainability. This thesis aims to tackle this by trying to achieve the following objectives:

- Provide an exploratory overview of recent advancements in AD techniques.
- Highlight the importance of XAI parameters in AD systems.
- Examine emerging trends, such as graph-based AD from the lens of XAI.
- Identify the challenges, evaluation metrics, and future directions in making AD techniques more explainable and interpretable for practitioners and researchers.

### **1.1.2 Enhancing AD Using GAN-Based Model**

The second part of the problem focuses on improving the performance of AD models, especially in the area of Semi-supervised AD. Semi-supervised AD is vital in scenarios where labeled data is scarce, and leveraging both labeled and unlabeled data can significantly enhance model performance. This research specifically investigates the enhancement of the Skip-GANomaly model, a GAN-based architecture, by incorporating novel modifications to its loss function through contrastive learning. This new model has been termed ConGANomaly. The goals of this enhancement are to:

- Integrate contrastive learning into the Skip-GANomaly model to improve the robustness and effectiveness of the learned representations.
- Implement the proposed modified ConGANomaly model practically and train it across various datasets
- Evaluate the proposed modified ConGANomaly model's performance across various datasets and other AD models to validate the improvements and establish its efficacy in semi-supervised AD tasks.

By addressing these two interrelated problems, this thesis aims to advance the field of anomaly detection both theoretically and practically. The comprehensive review of AD techniques with a focus on XAI will provide valuable insights for future research and development. At the same time, the enhanced ConGANomaly model will offer a novel and effective approach to semi-supervised anomaly detection.

## **1.2 Motivation**

Motivating my pursuit is firsthand exposure to the burgeoning demands within the data science industry. Drawing upon years of immersive experience in the industry, and witnessing the dynamic landscape of data-driven decision-making, I, and my able mentor, discern a growing need for robust anomaly detection methods. As organizations increasingly leverage AI for complex tasks, ensuring the integrity of datasets becomes

paramount. Anomalies, deviations from the norm, can carry crucial insights or signify potential issues, making their effective detection an imperative component in real-world applications. The motivation to bridge the gap between industry demands and cutting-edge research fuels our commitment to exploring innovative approaches in the realm of generative anomaly detection. The motivation behind integrating contrastive learning into anomaly detection models stems from its proven success in enhancing feature representations and discriminative capabilities. Contrastive learning has exhibited remarkable results in various implementations of machine learning based domains, such as those relating to natural language processing and the highly popular computer vision. By instilling the ability to learn from the differences between data points, contrastive learning complements the generative nature of GANs. This synergy aims to address challenges associated with discerning nuanced anomalies, providing the models with a more nuanced understanding of normal and abnormal patterns. The motivation lies in harnessing the amalgamation of generative adversarial networks and contrastive learning to propel anomaly detection models beyond their current capabilities.

The intersection of GANs and contrastive learning presents a promising avenue for advancing the state-of-the-art in semi-supervised anomaly detection. I, under the esteemed guidance of my mentor, aim to contribute to a more nuanced understanding of normal and anomalous patterns by leveraging generative capabilities and refining feature representations through contrastive learning. This exploration is not merely an academic pursuit but a direct response to the evolving needs of industries seeking robust solutions to the challenges posed by anomalies in diverse datasets. Thus, the research presented through this thesis represents a strategic stride, fusing established generative models with innovative contrastive learning paradigms. This synthesis not only fortifies the foundations of anomaly detection but also promises to elevate its applications across diverse domains, marking a significant advancement in the ongoing pursuit of more sophisticated and adaptable anomaly detection methodologies.

The literature review is explained in Chapter 2 while Chapter 3 holistically presents and explains the methodology, proposed approach, and experimental setup. Results obtained, along with challenges faced, are analyzed in Chapter 4 and Chapter 5 concludes the thesis.

## Chapter 2

### LITERATURE REVIEW

Anomaly detection, a fundamental facet of machine learning, encompasses a diverse array of techniques designed to identify deviations or irregularities within datasets [11]. Landscape of anomaly detection within the domain of machine learning has witnessed significant evolution, driven by the continuous demand for robust methods capable of identifying irregular patterns in diverse datasets. The integration of semi-supervised learning techniques within the broader context of anomaly detection represents a key paradigm shift in machine learning [17]. This chapter provides an insightful exploration of the historical context and contemporary advancements in anomaly detection, laying the foundation for the subsequent examination of specific methodologies and models. Notably, the rise of generative models, such as GANomaly and Skip-GANomaly, has introduced novel dimensions to semi-supervised anomaly detection. Additionally, this chapter will explore the paradigm of contrastive learning. Post laying a strong foundation in the understanding of AD techniques, we will also try to tackle the first portion of our problem statement in this chapter.

#### 2.1 Rise of Autoencoders

The ascendancy of autoencoders has been instrumental in reshaping the landscape of unsupervised learning and feature representation. Autoencoders, a class of neural networks, are aimed to learn an efficient way of representing the input data by trying to encode it into a lower-dimensional space and then reconstructing the same. In the realm of anomaly detection, autoencoders have demonstrated efficacy in capturing underlying patterns in both normal and anomalous instances.

One prominent variant of autoencoders is the Variational Autoencoder (VAE) [18, 19], which introduces a probabilistic element into the encoding process. VAEs learn not just a deterministic representation but also a probability distribution over possible representation. This capability is particularly advantageous in applications such as object detection. For instance, in object detection, VAEs can effectively model the variability in object appearances, enabling more robust feature representation[20]. However, it is essential to note that while VAEs exhibit success in capturing intricate patterns, their downfall lies in potentially over-smoothing learned representations, diminishing their discriminative power.



## 2.2 Adversarial Autoencoders

Adversarial autoencoders (AAEs) [21] represent a fusion of generative adversarial networks (GANs) and autoencoders, introducing adversarial training to enhance the generative capabilities of the latter. The primary concept involves training an autoencoder in conjunction with a discriminator that distinguishes between generated (fake) and actual (real) data. This adversity in the process of training encourages the autoencoder to generate realistic reconstructions, making AAEs particularly powerful in capturing complex data distributions. A notable example of adversarial autoencoders is Bidirectional Generative Adversarial Networks (BiGAN) [22]. BiGAN extends the concept of AAEs by introducing an additional inference network that maps data points to their latent representations. This bidirectional training further refines the learned representations, enhancing the discriminative capabilities of the model. Applications of adversarial autoencoders span various domains, including image synthesis, anomaly detection, and data generation. In the context of anomaly detection, the adversarial training inherent in AAEs facilitates the creation of more nuanced representations, enabling the model to discern anomalies more effectively [23].

## 2.3 GANomaly

GANomaly, introduced by Akcay et al. [7], represents a pioneering model in the landscape of semi-supervised anomaly detection. The model leverages the power of Generative Adversarial Networks (GANs) to enhance anomaly detection capabilities. In the GANomaly architecture, an autoencoder is combined with a GAN, consisting of a generator and a discriminator. The data samples are created by the generator, both normal and anomalous, while the distinction between real and generated samples is handled by the discriminator. The autoencoder is simultaneously trained to reconstruct the input data, reinforcing the learning of normal data patterns. The key contribution of GANomaly lies in its ability to leverage the adversarial form of training of GANs to increase the robust nature of anomaly detection. By introducing adversarial training, GANomaly refines the learned representations, enhancing the discrimination between normal and anomalous instances. This integration of adversarial training and autoencoder-based reconstruction sets GANomaly apart as a powerful tool in semi-supervised anomaly detection.

## 2.4 Skip-GANomaly

Skip-GANomaly [8] represents a noteworthy extension and improvement upon the GANomaly model, addressing limitations and further enhancing anomaly detection capabilities. The key innovation introduced by Skip-GANomaly is the incorporation of skip connections within the generator network. Skip connections facilitate straight flow of information of initial layers to later ones, granting model the capability to capture hierarchical features more effectively. The inclusion of skip connections in Skip-GANomaly contributes to improved feature learning and enables capturing of sophisti-

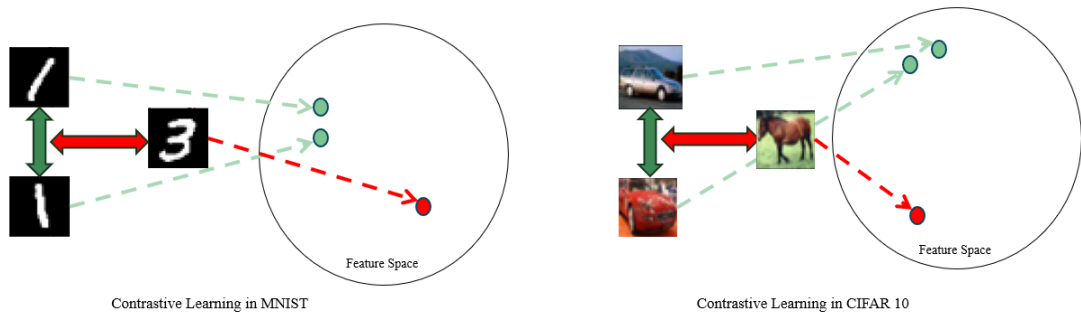


Figure 2.1: Contrastive Learning Intuition

cated patterns present in data under question. By facilitating a straight transmission of information across layers, Skip-GANomaly enhances the generator’s ability to generate realistic samples, both normal and anomalous. The extension to Skip-GANomaly exemplifies the continuous evolution in the field of semi-supervised anomaly detection, demonstrating the importance of refining architectures to achieve more nuanced and accurate anomaly discernment.

## 2.5 Contrastive Learning

Learning through contrast has come up as a powerful paradigm in machine learning [9, 10], aiming to learn meaningful representations by scrutinizing and detailing differences between unlikely (negative) and likely (positive) pairs. The underlying principle is trying to group positive instances closer in the embedding or feature space while pushing negative instances apart. Learning framework following this approach has exhibited remarkable success in various domains, spanning computer vision, natural language processing, and representation learning. Figure 1 showcases a simplified version of this learning methodology. The red and green arrows represent the negative and positive pairings respectively.

Success of contrastive learning can be attributed to its competence of leveraging huge amounts of unlabelled data effectively. In semi-supervised anomaly detection, where obtaining labeled anomalous instances can be challenging, contrastive learning becomes particularly valuable. By applying contrastive learning principles to anomaly detection models like GANomaly and Skip-GANomaly, we can enhance their capacity to learn meaningful embeddings for distinguishing between normal and anomalous instances. Contrastive learning complements the generative nature of these models by fostering a discriminative understanding of the data. The contrastive loss, applied strategically within the architecture, encourages the model to capture subtle differences in normal and anomalous instances, thereby improving the overall efficacy of semi-supervised anomaly detection.

## 2.6 SOTA Techniques for Anomaly Detection

This sub-section, ‘State-of-the-Art Techniques for Anomaly Detection’, serves to pro-

vide a comprehensive overview of the latest advancements in AD techniques. While we will cover traditional techniques in brief, we will focus in depth on deep learning based approaches, and emerging trends in graph-based anomaly detection. By examining these techniques, we aim to address the first part of the discussed problem statement, and at the same time set the stage for subsequent discussions on challenges, limitations, and future directions in the field. Traditional anomaly detection methods, including statistical techniques, distance-based methods, and clustering methods, form the foundation of anomaly detection and continue to play a crucial role in various domains. Deep learning approaches on the other hand were the catalyst that catapulted artificial intelligence into the limelight with AD techniques based on convolutional neural networks (CNNs), recurrent neural networks (RNNs), and especially, generative adversarial networks (GANs). Additionally, a potentially promising sub-field of deep learning is Graph Neural Networks (GNNs), which from an AD perspective, will majorly focus on anomalous node detection and anomalous edge detection.

### 2.6.1 Traditional Anomaly Detection Methods

Traditional AD methods have been the cornerstone of AD research, offering robust and interpretable approaches for identifying outliers in data. Throughout this paper, we have endeavored to select the most industry-prevalent techniques as the explainability and interpretability of AD techniques would only be relevant if they are industry-prevalent. Following in this stead, this subsection provides an overview of two state-of-the-art techniques within each subcategory of traditional anomaly detection methods: Statistical Techniques, Distance-based Methods, and Clustering Methods. A brief description about each of these techniques is given below.

#### Statistical Techniques

**Z-Score (ZS)**[24], also known as 'Standard Score,' is a widely used statistical technique in AD. It involves calculating deviations from the mean of individual data points based on the standard deviation and referencing these Z-scores against a predefined threshold value to determine anomalies. This method demonstrates simple yet effective AD capabilities, particularly in univariate datasets[25].

$$z_i = \frac{x_i - \mu}{\sigma} \quad (2.1)$$

**Grubb's Test (GT)**[26, 27], a classic method in AD, serves as an initial starting point for normally distributed datasets[28] [29]. It compares the max deviation of data points from the mean to the expected deviation based on the data distribution. Large deviations are flagged as anomalies, making it a reliable technique for detecting outliers.

$$G = \frac{|x_i - \bar{x}|}{s} \quad (2.2)$$

## Distance-Based Techniques

**K-Nearest Neighbours (KNN)**[30], is a widely used distance-based algorithm for AD. It identifies anomalies by calculating distances between each data point and its nearest neighbors in the feature space, with significantly distant points considered outliers. KNN is versatile and applicable to both numerical and categorical data[31].

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (2.3)$$

**Local Outlier Factor (LOF)**[32], assesses the local density of data points to their neighbors. Anomalies are identified by deviations in a data point's density from its neighbors' densities. Points with substantially lower densities compared to their neighbors are flagged as outliers. LOF is particularly effective in detecting anomalies in datasets with varying densities and non-linear distributions.

$$\frac{\sum_{o \in N(p)} \frac{LRD(o)}{LRD(p)}}{|N(p)|} \quad (2.4)$$

## Clustering Techniques

**K-means Clustering (KMC)**[33], is a popular unsupervised learning algorithm that groups data points based on similarity, with outliers considered as anomalies. It's one of the most widely used clustering-based AD techniques due to its efficiency and scalability, making it suitable for large-scale tasks[34].

$$\sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (2.5)$$

**DBSCAN**[35], Density-Based Spatial Clustering of Applications with Noise, identifies clusters based on regions of high data density separated by low-density regions. Anomalies are data points in low-density regions or not belonging to any cluster. Robust to noise and capable of detecting clusters of arbitrary shapes, it is suitable for detecting anomalies in complex datasets.

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (2.6)$$

While I have provided the objective functions of all the techniques, it is important to note that distance parameters and objective functions for KNN, KMC, and DBSCAN can change based on the specific implementation of algorithms.

### 2.6.2 Deep Learning Anomaly Detection Methods

Revolutionizing the AI world, deep learning approaches have become the go-to tool for AD, leveraging complex neural networks to capture intricate patterns and anomalies in data. Similar to traditional based approaches, let us go over some of the most

prevalent deep learning based approaches.

### Autoencoder-based Techniques

**Variational Autoencoder (VAE)**[18] [19] are widely used architectures for AD, featuring encoder and decoder sub-modules. They map input data to a latent vector and reconstruct it via the decoder. Anomalies are identified based on reconstruction error, with VAEs offering probabilistic interpretations and effectively capturing complex data distributions[20].

$$-\mathbb{E}_{z \sim q(z|x)} [\log p(x|z)] + \text{KL}(q(z|x)||p(z)) \quad (2.7)$$

**Sparse Autoencoder (SAE)**[19] are widely used in AD, imposing sparsity constraints on latent representations to learn compact, informative features. They identify anomalies based on reconstruction error, offering robustness to noise and effectively capturing subtle anomalies in high-dimensional data[36].

$$L_{recon} + L_{sparsityReg} \quad (2.8)$$

### CNN Based Techniques

**DeepConvLSTM (DCLSTM)**[37] is a deep learning architecture that combines CNNs with LSTM networks for AD in sequential data. It extracts spatial features and captures temporal dependencies using CNN and LSTM components, respectively. Anomalies are detected based on deviations from learned patterns in sequential data. Widely used in industry for detecting anomalies in time-series data, such as sensor readings and network traffic.

$$-\frac{1}{N} \sum_i y_i \log(\hat{y}_i) \quad (2.9)$$

**Capsule Network (CapsNet)**[38] addresses limitations of traditional CNNs by using capsules, groups of neurons representing object parts and poses. CapsNet captures spatial hierarchies and deformations, making it suitable for anomaly detection tasks requiring an understanding of object structure and relationships.

$$T_c \max(0, m^+ - \|v_c\|)^2 + \lambda (1 - T_c) \max(0, \|v_c\| - m^-)^2 \quad (2.10)$$

### RNN Based Techniques

**LSTM**[39] a type of RNN, are widely employed for anomaly detection in sequential data due to their ability to capture long-range dependencies and temporal dynamics. They excel at modeling complex sequential patterns, making them suitable for various industrial applications in time-series data, natural language processing, and other sequential data domains.

$$-\frac{1}{N} \sum_i y_i \log(\hat{y}_i) \quad (2.11)$$

**Gated Recurrent Unit (GRU)**[40] a variation of LSTM, is commonly used for AD tasks due to its simpler architecture, making it computationally efficient and easier to train. With a memory state capturing dependencies across time steps, GRUs detect anomalies in sequential data accurately and efficiently. Widely used in industry for applications like fraud detection and cybersecurity.

$$-\frac{1}{N} \sum_i y_i \log(\hat{y}_i) \quad (2.12)$$

### GAN Based Techniques

**GANomaly (GANM)**[7], combines GANs with autoencoder architectures for semi-supervised anomaly detection. A generator sub-network generates normal data samples, while an encoder-decoder network reconstructs input data. Anomalies are identified based on the network’s inability to accurately recreate them. GANomaly is versatile, detecting anomalies in diverse data types like images, time-series data, and tabular data.

$$W_1 * L_{adv} + W_2 * L_{con} + W_3 * L_{enc} \quad (2.13)$$

**Skip-GANomaly (SGANM)** [8] improves upon GANM with skip connections and enhanced adversarial training, enhancing overall AD performance. Skip connections between encoder and decoder layers aid information flow and gradient propagation during training. Skip-GANomaly is a semi-supervised AD framework capable of detecting anomalies in various data modalities, showing promising results in practical applications.

$$W_1 * L_{adv} + W_2 * L_{con} + W_3 * L_{lat} \quad (2.14)$$

While I have provided the objective functions of all the techniques, it is important to note that objective functions for DCLSTM, LSTM, and GRU can change based on the specific implementation of algorithms.

### 2.6.3 Graph Based Anomaly Detection Methods

Graph-based AD techniques leverage the inherent relationships and structures within data represented as graphs to identify anomalous patterns. This subsection provides an overview of two state-of-the-art techniques within each subcategory of graph-based anomaly detection methods: Anomalous Node Detection and Anomalous Edge Detection.

## Anomalous Node Detection Techniques

**NetWALK**[41] uses random walks to traverse the graph structure. Anomaly scores are assigned to nodes based on random walk frequencies, identifying anomalies as nodes with unusual visitation patterns. Effective in capturing local and global anomalies in graph-structured data, NetWALK finds applications in domains like social and biological networks.

$$\gamma L_{AE} + L_{Clique} + \lambda \|W\|_F^2 + \beta KL \quad (2.15)$$

**ResGCN**[42] extends traditional graph convolutional networks (GCNs) by introducing residual connections within the architecture. Learning node representations through graph convolution operations, residual connections aid gradient propagation and alleviate the vanishing gradient problem. ResGCN detects anomalies based on deviations from learned patterns, offering improved convergence and performance compared to traditional GCNs, showing promising results in various anomaly detection tasks.

$$(1 - \alpha)E_S + \alpha E_A \quad (2.16)$$

## Anomalous Edge Detection Techniques

**DeepSphere (DS)**[43] detects anomalous edges in graph-structured data using spherical convolutions to capture geometric relationships between nodes. Learning representations of edges, anomalies are identified based on deviations from learned edge representations. Robust to noise, DS effectively detects anomalies in large-scale graph datasets with high accuracy.

$$L_h + \lambda L_{res} \quad (2.17)$$

**DeepFD**[44] DeepFD uses a deep autoencoder architecture to learn compact representations of graph edges, detecting anomalies based on reconstruction errors where higher errors indicate anomalous edges. Capable of capturing complex patterns, it finds applications in network intrusion and fraud detection.

$$L_{recon} + \alpha L_{sim} + \gamma L_{reg} \quad (2.18)$$

## 2.7 Explainable AI for Anomaly Detection

Due to the critical roles AD plays across industries, the need for increased explainability in current AD models has been at an all time high. Explainable AI (XAI) may satiate this need by providing varied parameters for assessing AD models, through which, traditionally intangible objectives like enhancing stakeholder’s trust and acceptance, result validation, ethical compliance and many others may be accomplished. Through examining the below XAI parameters across our 18 approaches, we aim to bring forth valuable insights for enhancing understanding and trust in these AD systems.

- **Interpretability (INT):** A vital cornerstone for enhancing confidence among stakeholders and AI systems. Techniques with high INT can offer transparent explanations, while medium INT struggles to balance interpretability and complexity. Low INT, however, leads to skepticism of model decisions. For ensuring trust and acceptance by stakeholders in AD models interpretability is essential. GAN based models often lack proper interpretability [45].
- **Feature Importance (FI):** provides meaning to role of features in AD outcomes. High FI techniques provide clear measures of feature importance, aiding in stakeholder’s understanding of influential features and their impact on AD. Low FI hinders understanding of AD systems. Cybersecurity and fraud detection are domains where FI can be critical for effective risk mitigation.
- **Model Complexity (MC):** Stakeholder’s understanding and trust in model decisions are inversely proportional to MC. High MC methods, like deep learning, can be intricate and challenging to comprehend. Low MC techniques, such as traditional statistics, offer simplicity and transparency. Balancing sophistication and transparency is crucial for effective AD.
- **Visualization (VIZ):** Pivotal for XAI in AD [46, 47], VIZ provides stakeholders with intuitive representations of complex data patterns. VIZ capabilities give exploration capabilities and a better understanding of model outputs through interactive dashboards, heatmaps, or scatter plots.
- **Human-Readable Outputs (HRO):** Providing stakeholders with AD as readable outputs is a surefire way to improve explainability. Techniques with HRO outputs, such as textual descriptions or visualizations, enhance interpretability and facilitate collaboration. Conversely, methods lacking HRO may lead to ambiguity [48].
- **Robustness to Perturbations (RP):** AD models in the actual world have their dependability of applications directly correlated to RP. High RP demonstrates resilience against variations of input data and tends to maintain consistent performance across diverse datasets. On the other hand, low RP methods show sensitivity to variations, resulting in unreliable performance [49]. Stakeholders almost always prefer AD models with High RP.

## 2.8 SOTA Techniques Through Lens of Explainable AI

Having understood the 18 SOTA industry prevalent AD techniques and garnered the foundations of the 6 XAI parameters, this thesis will now conclusively tackle the first part of our problem statement. Using the XAI parameters, the SOTA techniques will be evaluated. This evaluation will help provide a go-to tabular representation through which one can decide which AD technique to select based on the explainability demands of the problem domain. Tables 2.1 contain the final results of our evaluation of 18 SOTA techniques across the previously discussed XAI parameters.



Method	Technique	Name	Objective Function	Metric	Output	INT	FI	MC	VIZ	HRO	RP	
Traditional AD Methods	Statistical	ZS	$z_i = \frac{x_i - \mu}{\sigma}$	ZSL	SSA	High	High	Low	No	Yes	Med	
		GT	$G = \frac{ x_i - \bar{x} }{s}$	GSL	SSA	Med	Med	Low	No	Yes	Med	
	Distance	KNN	$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$	DKNN	FA	High	Med	Low	Yes	Yes	Low	
		LOF	$\frac{LRD(o)}{\sum_{p \in N(p)} \frac{LRD(p)}{ N(p) }}$	LOFS	SSA	Med	High	Med	No	Yes	Med	
	Clustering	KMC	$\sum_{i=1}^k \sum_{x \in C_i} \ x - \mu_i\ ^2$	CM	FA	Med	Med	Low	Yes	Yes	Low	
		DBSCAN	$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$	DCM	CA	Med	High	Med	Yes	Yes	High	
Deep Learning AD Methods	AE	VAE	$-\mathbb{E}_{z \sim q(z x)} [\log p(x z)] + \text{KL}(q(z x) \  p(z))$	AS	FA	Med	Med	Med	Yes	Yes	Med	
		SAE	$L_{recon} + L_{sparsityReg}$	AS	FA	Med	Med	Med	Yes	Yes	Med	
	CNN	DCLSTM	$\frac{1}{N} \sum_i y_i \log(\hat{y}_i)$	PL	PRED	Low	High	High	Yes	Yes	Med	
		CapsNet	$T_c \max(0, m^+ - \ v_c\ )^2 + \lambda(1 - T_c) \max(0, \ v_c\  - m^-)^2$	ML	REC	Low	High	High	Yes	Yes	Med	
	RNN	LSTM	$-\frac{1}{N} \sum_i y_i \log(\hat{y}_i)$	PL	PRED	Med	High	High	Yes	Yes	Low	
		GRU	$-\frac{1}{N} \sum_i y_i \log(\hat{y}_i)$	PL	PRED	Med	High	High	Yes	Yes	Low	
	GAN	GANM	$W_1 * L_{adv} + W_2 * L_{con} + W_3 * L_{enc}$	AS	FA	Med	Med	High	High	Yes	Yes	High
		SGANM	$W_1 * L_{adv} + W_2 * L_{con} + W_3 * L_{lat}$	AS	FA	Med	Med	High	High	Yes	Yes	High
Graph AD Methods	Ano-Node	NetWALK	$\gamma L_{AE} + L_{Clique} + \lambda \ W\ _F^2 + \beta KL$	AS	NDCC	Med	High	High	No	Yes	High	
		ResGCN	$(1 - \alpha) E_S + \alpha E_A$	LES	AL	Low	High	High	Yes	Yes	High	
	Ano-Edge	DS	$L_h + \lambda L_{res}$	LES	AL	Med	High	High	Yes	Yes	High	
		DeepFD	$L_{recon} + \alpha L_{sim} + \gamma L_{reg}$	LES	AL	Low	High	High	Yes	Yes	High	

Table 2.1: AD Approaches through Explainable AI

It is important to note that KNN, KMC, DBSCAN Distance Parameters, DCLSTM, LSTM, and GRU Objective Functions can change based on implementation. To preserve table based spacing, the following abbreviations have been used: AE - Autoencoder, Ano-Node - Anomalous Node, Ano-Edge - Anomalous Edge, ZSL - Z Score List, GSL - G Score List, DKNN - Distance to KNN, LOFS - LOF Score, CM - Cluster Members, DCM - Density Cluster Measure, AS - Anomaly Score, PL - Prediction Loss, ML - Margin Loss, LES - Location in Embedding Space, SSA - Statistically Significant Anomalies, FA - Flags Anomalies, CA - Cluster Assignment, PRED - Predictions, REC - Reconstruction, NDCC - Nearest Distance to Cluster Centre, AL - Anomalous Label, Med - Medium.

## Chapter 3

### METHODOLOGY

In the realm of anomaly detection (AD), numerous state-of-the-art approaches have been developed, each offering distinct advantages and limitations - both, from the perspective of the ability as well as explainability of the approaches. After a comprehensive evaluation of these approaches, it became evident that a semi-supervised learning framework is particularly well-suited for practical applications in industry. One of the primary challenges in real-world scenarios is the scarcity of labeled abnormal instances. Large datasets with sufficient labeled anomalies are rare, making fully supervised learning impractical for many applications. Conversely, unsupervised methods, while useful, often struggle with high false-positive rates due to their reliance on assumptions about the underlying data distribution.

Semi-supervised learning provides a balanced solution by leveraging both labeled and unlabeled data, making it more adaptable to the constraints of industrial settings. This approach allows the model to learn from the limited labeled anomalies and a large pool of unlabeled data, enhancing its ability to detect rare and subtle anomalies without the need for an extensive labeled dataset.

#### 3.1 Why Choose GAN Based Model?

To tackle the second part of the problem statement, it is essential to first understand the reasons behind choosing GAN-based models. Within the semi-supervised AD framework, various models have been explored, with Skip-GANomaly emerging as a particularly effective choice. Skip-GANomaly, and similar models, offer several advantages over traditional deep learning models and Graph Neural Networks (GNNs) concerning explainability and performance. These advantages are as follows:

- **Interpretability of Outputs:** Models based on Generative Adversarial Networks (GANs), inherently generate outputs that can be compared with the original input data. This generation process facilitates a more intuitive understanding of where and how anomalies occur, as discrepancies between the input and generated output highlight the anomalous regions. This is also further strengthened by the 'Yes' rating of GANM and SGANM approaches under HRO XAI parameter, as shown in Table 2.1.

- **Robust Representation Learning:** The GAN-based structure of Skip-GANomaly, and similar models, ensures robust learning of normal data patterns, which is crucial for accurately identifying deviations. By capturing the complex distributions of normal data, these models can more effectively detect anomalies, even when they are subtle or rare. This is evident from the rating of 'High' being achieved by GAN based methods in XAI parameter RPO, as shown in Table 2.1.
- **Model Complexity and Training Efficiency:** While GNNs are powerful for structured data and relational learning, they often come with higher computational complexity and longer training times. GAN based models, in contrast, offer a more streamlined approach that balances complexity and efficiency, making them more practical for real-time industrial applications.
- **Applicability to Diverse Data Types:** GAN based models are highly versatile and can be applied to various data types, including time-series, image, and tabular data, whereas GNNs are primarily suited for graph-structured data.

Thus, based on the above compelling factors, paired with the superior compatibility of GAN based models with semi-supervised problem conditions, choosing GAN based models for research and improvements is the logical choice. By selecting the Skip-GANomaly model and enhancing it with contrastive learning, this research aims to develop a robust, explainable, and practical solution for semi-supervised anomaly detection. This proposed modified model has been termed ConGANomaly. The following sections will detail the implementation process and the novel modifications introduced to the model's loss function, demonstrating how these enhancements contribute to superior AD performance.

## 3.2 Proposed Approach

This and subsequent sections in the methodology chapter delineate the approach undertaken to integrate contrastive learning into the GANomaly and Skip-GANomaly based models, elucidating the steps to enhance their anomaly detection capabilities. Building upon the foundation laid by these base models, the methodology encompasses a comprehensive overview of the integration of contrastive learning principles. The model used for integration follows the design principles of the base models but introduces changes in various parts to incorporate contrastive learning. The primary components of the model are: a generator (G) and a DCGAN based classifier discriminator (C). The generator synthesizes normal samples, while the classifier discriminator attempts to distinguish between real and fake (generated) instances. Training in an adversarial environment of this model encourages the classifier generator to produce realistic samples, enhancing the overall anomaly detection capabilities. The model's objective loss function combines adversarial loss, contextual loss, latent loss, and contrastive loss providing a comprehensive loss metric for the overall model. Figure 3.1 showcases the proposed architecture with contrastive learning data pairings. The subsequent sections delve into the specifics of architecture used and foundational components. Equations

accompany these descriptions to provide a clear understanding of the model formulations.

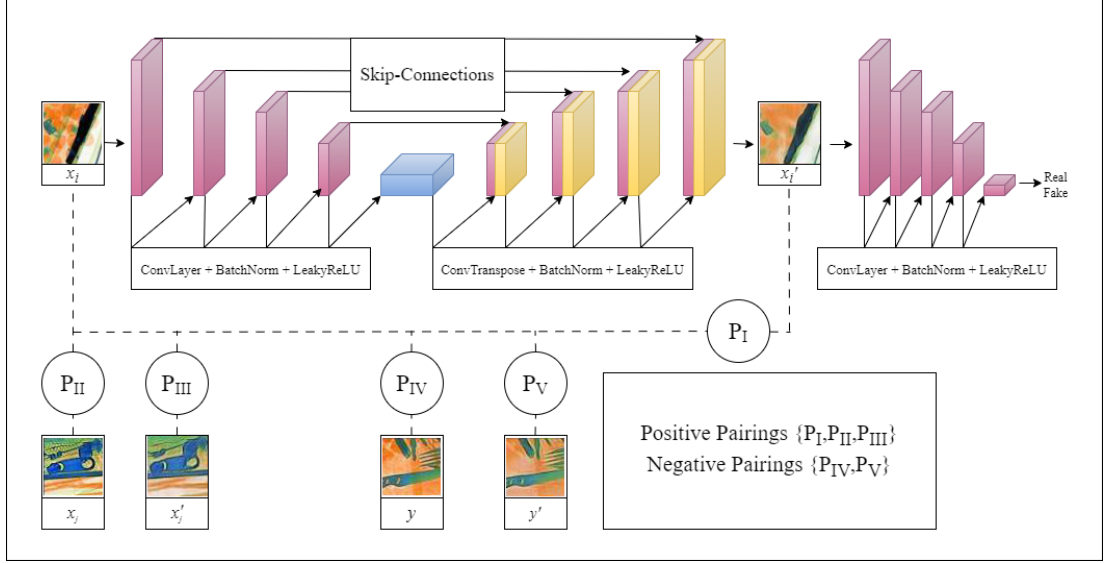


Figure 3.1: Proposed Architecture with Contrastive Learning Data Pairings

### 3.2.1 Generator Sub-Network

For an isolated input image ‘x’, the image first goes through a bow tie autoencoder network. This bow tie autoencoder is fundamentally a symmetric combination of an encoder and a decoder which will help the network identify two main things – the bottleneck features and the generated image  $x'$ . This specific sub-network is called the Generator and is denoted by ‘G’. G network comprises 2 sub-networks called the ‘Encoder’ (represented by  $G_E$ ) and ‘Decoder’ (represented by  $G_D$ ).  $G_E$  is designed to efficiently capture hierarchical features from input data. The encoder processes the input through convolutional layers, progressively reducing the spatial dimensions. The encoded features are then transmitted through the skip connections to the corresponding decoder layers, facilitating the preservation of detailed information. While  $G_D$  in Skip-GANomaly complements the encoder by reconstructing the input data from the encoded features. The main novel addition in this model is the ‘Skip-connections’ introduced in the  $G_D$ . The skip connections help in directing the passage of information, aiding in the recovery of intricate details. The decoder layers gradually up-sample the features, reconstructing the input with a high level of fidelity. This bow-tie architecture, with skip connections, enhances the model’s capacity to generate realistic samples.

### 3.2.2 Discriminator Sub-Network

The discriminator in network architecture performs a binary classification task, distinguishing between actual (input/real) and fake (generated) data. This network is represented as ‘C’. Trained adversarially, C guides the generator G to produce more

authentic samples. This network is responsible for vital contributions in the overall adversarial training process, contributing to the refinement of the generator’s ability to generate realistic data instances. Additionally, C has the task of extracting features so that the latent representations from  $x$  and  $x'$  can be computed (nomenclature is as per our discussion in previous sub-section). This particular extraction’s role will become clearer when we discuss the loss functions.

### 3.2.3 Individual Loss Functions

This model employs the weighted summation of multiple individual losses where the weights signify the relative importance of each loss and can be varied depending upon the use cases. During our experimentation with different combinations of weights, we found that this kind of customization allowed our model to be more robust and useful. Each of the individual losses in the loss function has a specific purpose which we will discuss in the following paragraphs. We have intentionally not discussed contrastive loss in this sub-section as it would require much more discussion on pairing strategies, embedding layer modifications, etc. Consequently, we will discuss contrastive loss in its own subsection in detail.

#### Adversarial Loss

The adversarial loss ( $L_{adv}$ ) [6] is fundamental to the adversarial training of the model. It is formulated to encourage G to create samples that are indistinguishable from actual input data. The adversarial loss is computed based on the standard GAN objective, promoting a competitive interplay between the generator and the discriminator. Equation for  $L_{adv}$  is represented as follows:

$$L_{adv} = E_{x \sim P_x} [\log(D(x))] + E_{x \sim P_x} [\log(1 - D(x'))] \quad (3.1)$$

#### Contextual Loss

The contextual loss  $L_{con}$  focuses on the preservation of spatial structures in the reconstructed samples. It measures the dissimilarity of given input data with obtained reconstructed output, ensuring G captures both global and local contextual information. This loss component contributes to the faithful reconstruction of the input. Equation for  $L_{con}$  is represented as follows:

$$L_{con} = E_{x \sim P_x} |x - x'|_1 \quad (3.2)$$

#### Latent Loss

The latent loss ( $L_{lat}$ ) plays a pivotal role in encoding a notion of anomaly within the architectural framework. It encourages the model in understanding of the discriminative representations in the latent space, aiding in the effective separation of normal and anomalous instances. The latent loss is instrumental in refining the anomaly detection

capabilities of the model. Equation for  $L_{lat}$  is represented as follows:

$$L_{lat} = E_{x \sim P_x} |f(x) - f(x')|_2 \quad (3.3)$$

### 3.2.4 Contrastive Loss

The targeted novelty of our research work is aimed towards incorporating the concept of contrastive learning [9, 10, 50, 51] into anomaly detection models. The incorporation of contrastive learning into the GANomaly and Skip-GANomaly models involves a strategic modification of the architecture, with a primary focus on the embedding layer and data pairs. The objective of Contrastive learning is to learn representations so optimize the embedding space such that there exists maximum similarity between positive pairs and minimum similarity for the negative ones. To achieve this, we introduce a contrastive loss component into the overall loss function. Since this requires multiple modifications to the architecture and how the data is processed, let us explore all each of the modifications one by one.

#### Embedding Layer Enhancement

The embedding layer in both GANomaly and Skip-GANomaly is crucial for encoding the input data into a latent space. To adapt the models for contrastive learning, we enhance the utilization of the embedding layer by giving more focus to the image obtained from the up-sampling of the latent representations –  $x'$ . The embedding layers will update their representations based on the contrastive loss in the backward propagation step. The modified embedding layer learns to map similar instances close together while pushing dissimilar instances apart. And in order to compute the contrastive losses, we would require 'data pairs'.

#### Data Pairing Enhancement

The research work being encapsulated by this thesis will be introducing the data pairing concept to facilitate the computation of the contrastive loss of the model. We have devised certain potential pairing options that we believe would help the model better adapt and learn the contrasting latent representations of different possible classes of the training dataset. Let us explore them in detail.

- **Positive Pairing Approach:** The positive pairing approach results in the positive pairs being formed which are used to guide the model to learn the features of the latent distribution representation by providing it two samples from the same or similar latent space. This helps the model compare the computed distribution from input  $x$  to another input  $x_{alt}$  through which it can learn to map similar instances closer in the latent space. We hope to achieve this in the following potential ways:

- Pair the input with the reconstructed generated image:

$$P_I = (x_i, x'_i); x'_i \in G(x_i) \quad (3.4)$$

where  $G(\cdot)$  is the model generator function

- Pair the input with another input from the same class:

$$P_{II} = (x_i, x_j); x_i, x_j \in C_k \quad (3.5)$$

where  $C_k$  is any  $k^{th}$  class of the dataset

- Pair the input with another reconstructed input from the same class:

$$P_{III} = (x_i, x'_j); x_i, x_j \in C_k, x'_j \in G(x_j) \quad (3.6)$$

where  $G(\cdot)$  is the model generator function

- **Negative Pairing Approach:** The negative pairing approach results in the negative pairs being formed which are used to guide the model to contrast and learn the distinction between features of the latent distribution representation by providing it two samples from two different latent spaces. This helps the model compare the computed distribution from input  $x$  to an input  $x_{alt}$  from another class through which it can learn to push dissimilar instances apart in the latent space. We hope to achieve this in the following potential ways:

- Pair the input with another input from a different class:

$$P_{IV} = (x, y); x \in C_i, y \in C_k \quad (3.7)$$

where  $C_i$  and  $C_k$  are any  $i^{th}$  and  $k^{th}$  classes of the dataset

- Pair the input with the reconstructed generated image of an input from a different class:

$$P_V = (x, y'); x \in C_i, y' \in G(y) \quad (3.8)$$

where  $G(\cdot)$  is the model generator function

## Contrastive Loss Function

The contrastive loss ( $L_{cont}$ ) is introduced to explicitly account for the similarity in relationships between embeddings. It encourages the model to pull positive pairs close ( $z_i, z_j$ ) and while pushing negative pairs apart ( $z_i, z_k$ ). Here, as discussed above, positive pairs  $(z_i, z_j) \in P_I, P_{II}, P_{III}$  and negative pairs  $(z_i, z_k) \in P_{IV}, P_V$ . The contrastive loss is defined as:

$$L_{cont} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N 1_{i \neq j} \log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^N 1_{i \neq j} \exp(\text{sim}(z_i, z_j)/\tau)} \quad (3.9)$$



Where the  $\text{sim}(\cdot)$  function is the function to compute the similarity measure between the positive pairs and 1 is the indicator function. This is one of the methods of implementation of contrastive loss which we have currently, this can be modified based on the experiences we have during implementation so as to better enhance and capture the concept of contrastive learning.

### Final Modified Loss Function

The modified architecture retains the key components of base models while incorporating the contrastive learning enhancements. Based on the model, the objective function becomes a combination of adversarial loss, contextual loss, latent loss, and the newly introduced contrastive loss. Thus the final loss function becomes:

$$L = W_{adv} * L_{adv} + W_{con} * L_{con} + W_{lat} * L_{lat} + W_{cont} * L_{cont} \quad (3.10)$$

The above loss function takes into consideration all the four different losses discussed previously. The incorporation of contrastive learning into the models aims to refine the learned representations, fostering improved discrimination between normal and anomalous instances in the embedding space.

## 3.3 Experimental Setup

### 3.3.1 Datasets

In our research, we leverage four datasets, for the training and evaluation of the proposed contrastive learning enhanced model. These datasets are selected based on a two-fold reasoning. The primary reason is that results on these datasets can be compared with the other existing state-of-the-art GAN-based anomaly detection models. Secondly, these datasets are widely and easily available which will enhance the reproducibility of our results and findings further facilitating validations and further future works built on top of our research.

#### CIFAR-10

For this dataset [52], we employed a similar leave-one-class-out approach [7, 8] which converts dataset into 10 unique anomaly cases. Consequently, there would be approximately 45,000 normal training samples. For testing, 3:2 ratio out of a total of 15,000 samples would be used for normal and abnormal respectively. Every time one anomaly case is selected, the rest all will act as normal cases. Figure 3 (b) depicts sample images from this dataset.

#### MNIST

Approach for this dataset [53] will be similar to CIFAR-10 dataset. Each digit will

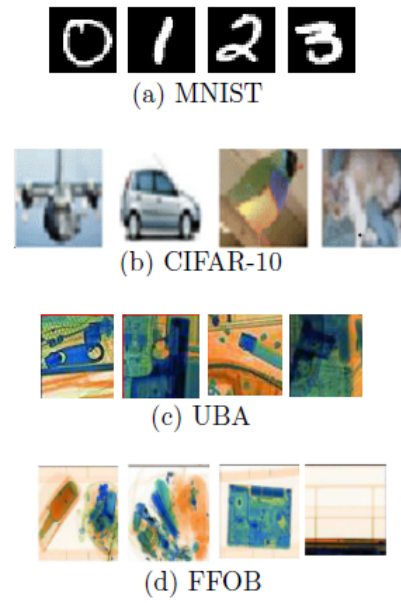


Figure 3.2: Dataset Sample Images - (a) MNIST (b) CIFAR-10 (c) UBA (d) FFOB

serve as an anomaly case. Here also we will get 10 unique anomaly cases. Figure 3.2 (a) depicts sample images from this dataset.

### Full Firearm vs. Operational Benign (FFOB)

Dataset [54] first introduced by UK government consisting of expertly concealed threats (firearms) and non-threat (operational benign) items. With a total of 72,325 data samples, 4,680 are anomaly class (threats) while rest will be considered as a normal class. Figure 3.2 (d) depicts sample images from this dataset.

### University Baggage Anomaly Dataset (UBA)

Dataset [55] used in [7, 8] consists of 230,275 total samples in a 64 x 64 individual sample size with 3 anomaly cases. Sample distribution of anomaly cases is 13,452, 45,855, and 63,496 which correspond to gun component, gun, and knife. Remaining instances are normal cases. Figure 3.2 (c) depicts sample images from this dataset.

### 3.3.2 Setup Configurations

Utilization of dual avenues were used for implementing the model and running the training and testing. The primary avenue was an AMD Ryzen 9 5900HX with Radeon Graphics - 3.30 GHz-based system with overclocking available. This system was equipped with an NVIDIA GeForce RTX 3060 GPU with 6 GB VRAM and 16GB system RAM. The software specifications included CUDA 12.1, CUDNN 12.x, Python 3.10.13, and PyTorch 2.1.1. The secondary avenue was Google colab based notebooks and infrastructure.

## Chapter 4

### RESULTS & ANALYSIS

#### 4.1 Results Obtained

The evaluation performance of our proposed contrastive learning enhanced model is evaluated by utilizing the concept of area under the curve (AUC) of the receiver operating characteristics (ROC) [56]. ROC is the graphical representation of the varying performance of the model on the basis of true positive rate (TPR) and false positive rate (FPR) across different threshold values. The decision of selecting this evaluation metric is based on previous work in this field [7, 8, 5, 57].

Table 4.1 showcases the performance of the proposed contrastive learning enhanced model in comparison to 4 other models - namely AnoGAN [5], EGBAD [57], GANomaly [7], and Skip-GANomaly [8]. As evident from the results, the proposed modifications have resulted in the model outperforming all of the other four models across all the datasets. Performance in CIFAR-10 shows the highest increase from 0.730 (of Skip-GANomaly) to 0.869. UBA shows the least improvement with an increase of 0.045 in the evaluation metric. Let us now dwell deeper into performance across each dataset.

Figure 4.1 shows the performance of each model on the MNIST dataset. As discussed in the previous section, there were 10 anomaly cases that could be possible using the leave-one-out strategy. As evident from analyzing the graph, the proposed modifications have resulted in superior performance across 9 of the 10 anomaly cases. Even in the case of Digit 2, for which the model under-performed, the drop in evaluation metric was extremely small (0.01). The highest improvement can be seen in the case where digit 9 was the anomaly - an improvement from the previous maximum of 0.8 to 0.89. The overall improved evaluation metric across the whole dataset, as reflected in Table 1 also, is 0.919 (from a previous maximum of 0.881).

Figure 4.2 showcases the performance of each model for the CIFAR-10 dataset. In this dataset also, there was a possibility of 10 anomaly cases. However, unlike the

Datasets	AnoGAN	EGBAD	GANomaly	Skip-GANomaly	ConGANomaly
<b>MNIST</b>	0.445	0.506	0.789	0.881	<b>0.919</b>
<b>CIFAR-10</b>	0.434	0.462	0.610	0.730	<b>0.869</b>
<b>UBA</b>	0.569	0.597	0.643	0.940	<b>0.945</b>
<b>FFOB</b>	0.703	0.712	0.882	0.903	<b>0.944</b>

Table 4.1: AUC Results Overview

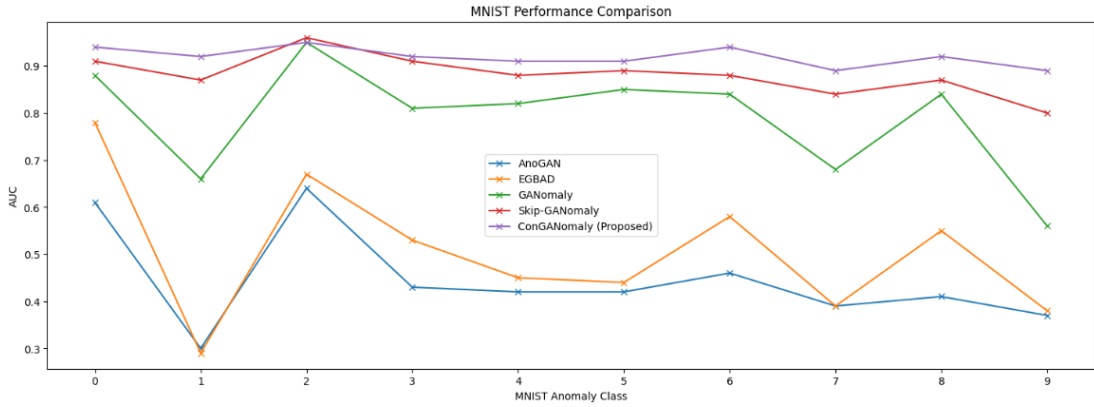


Figure 4.1: MNIST Performance Comparison

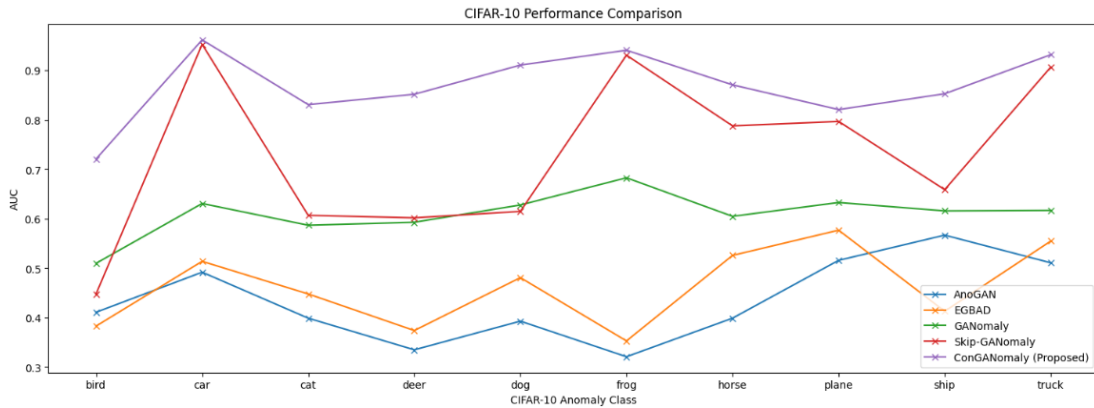


Figure 4.2: CIFAR-10 Performance Comparison

MNIST dataset where we saw a marginal increase in performance, here we can see a relatively larger improvement in performance across the dataset. All 10 anomaly cases have superior evaluation metric scores - with the highest improvement being seen in the cases where bird, cat, or dog class was the anomaly.

Table 4.2 depicts the performance of the model on the UBA and FFOB datasets. In UBA dataset, the Knife class case gained the highest improvement in the evaluation metric followed closely by the gun class case. While there was a slight fall in the evaluation metric for the gun parts class case from the previous maximum, the overall performance on the UBA dataset was still an improvement. In FFOB, having only 1 anomaly case, we saw an overall improvement in the evaluation metric.

Model	UBA				FFOB
	gun	gun-parts	knife	overall	full-weapon
<b>AnoGAN</b>	0.598	0.511	0.599	0.569	0.703
<b>EGBAD</b>	0.614	0.591	0.587	0.597	0.712
<b>GANomaly</b>	0.747	0.662	0.520	0.643	0.882
<b>Skip-GANomaly</b>	0.972	<b>0.945</b>	0.904	0.940	0.903
<b>ConGANomaly</b>	<b>0.981</b>	0.933	<b>0.923</b>	<b>0.945</b>	<b>0.944</b>

Table 4.2: AUC results for UBA and FFOB dataset

## 4.2 Analyzing Challenges Faced

While AD is vital across domains, deep learning based methods still face persistent challenges. Through this section, an exploration of these challenges and future directions for advancing anomaly detection research and applications will be undertaken.

- **Lack of Explainable Approaches:** Deep learning models have a reputation of working as black boxes, which makes comprehending their process of making decisions in AD hard. Moreover, as evident from Table 2, none of the prevalent approaches are a one-stop solution which is positive across all parameters. Addressing this challenge is crucial for enhancing anomaly detection models' transparency and explainability, aligning with our research question's objectives on the importance of explainable AI techniques.
- **Data Imbalance and Labeling Issues:** AD datasets often exhibit class imbalance, with normal instances outnumbering anomalies, as observed across the reviewed papers and models. This imbalance poses challenges for deep learning models, hindering their ability to learn from limited anomaly examples. While models like GANomaly and Skip-GANomaly partly address this issue, effectively managing data imbalance remains a critical need in the AD domain.
- **Anomaly Generalization:** Deep learning models, particularly supervised ones, often struggle to generalize to all possible anomalies, as they tend to overfit to the anomalies present in labeled datasets. Although semi-supervised models like GANomaly and Skip-GANomaly may offer partial solutions, they may sacrifice feature importance and struggle with complex anomalies. Developing techniques to mitigate overfitting and enhance generalization is essential for improving the robustness of deep learning-based anomaly detection systems, addressing our research question on the limitations of these approaches.
- **Contrastive Learning Pairings:** While contrastive learning has lots of benefits, implementing the data pairings can be sometimes quite complex. Having too many pairings can result in longer training times and an increase in computation requirements. Whereas, having too few pairings may not yield any significant performance improvement. Thus, a trade-off exists and needs to be carefully adjusted based on the features of the problem domain.

Addressing these challenges opens avenues for future research directions. This includes enhancing explainability in AD methods through improving existing approaches or developing novel, more transparent methods. Additionally, efforts can focus on mitigating data imbalances through modifications to existing approaches, proposing new methods, or introducing improved metrics.

## Chapter 5

### CONCLUSION, FUTURE SCOPE AND SOCIAL IMPACT

#### 5.1 Conclusion

The field of AD has seen significant advancements, driven by the increasing complexity and volume of data across various domains. This thesis has thoroughly explored and evaluated 18 SOTA AD techniques, focusing on their applicability and performance within the framework of XAI. Through this comprehensive evaluation, it has been demonstrated that while traditional and deep learning-based AD methods offer substantial capabilities, they often fall short in terms of explainability and interpretability — key factors for real-world deployment and trustworthiness.

In response to these challenges, this research introduced a modified version of SkipGANomaly based model called ConGANomaly, enhanced with novel loss functions and the incorporation of contrastive learning. The modifications were aimed at improving the semi-supervised AD performance by better distinguishing between normal and anomalous samples. The proposed model was rigorously tested and evaluated across multiple datasets, showcasing significant improvements in performance metrics. These improvements were illustrated through detailed graphs and tables, underscoring the model's efficacy and robustness.

The integration of XAI parameters in the evaluation of AD techniques provided critical insights into how these models can be made more interpretable and reliable. The results underscore the importance of balancing model complexity with explainability, ensuring that AD systems not only perform well but also offer transparency in their decision-making processes. This thesis contributes to the ongoing dialogue in the AI community about the necessity of explainable and interpretable models, particularly in high-stakes applications where understanding the reasoning behind anomaly detection is crucial.

In conclusion, this research not only advances the technical boundaries of AD through the proposed model enhancements but also emphasizes the critical role of explainability in developing trustworthy AI systems. The findings of this thesis pave the way for future research and practical applications, where both performance and interpretability are paramount.

## 5.2 Future Scope and Social Impact

The advancements and findings presented in this thesis open several avenues for future research and development in the field of anomaly detection (AD). Key areas for future work include:

- **Enhanced Explainability:** While this thesis has made strides in improving the explainability of AD models, further work is needed to develop more sophisticated techniques that can provide deeper insights into the model's decision-making processes. Future research could focus on integrating advanced XAI methods such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) with GAN-based AD models.
- **Real-Time Anomaly Detection:** Implementing the ConGANomaly model in real-time systems presents a significant challenge and opportunity. Future research should explore optimizing the model for real-time anomaly detection, including reducing computational overhead and improving processing speeds without compromising performance.
- **Hybrid Models:** Combining the strengths of various AD techniques, such as integrating graph-based methods with GAN-based models, could potentially yield even more powerful and flexible AD systems. Future studies could investigate hybrid models that leverage the unique advantages of multiple AD approaches.
- **Adaptive Learning:** Developing AD models that can adapt to evolving data patterns and contexts is another promising area. Future work could explore adaptive learning techniques that allow models to continuously learn and improve from new data without requiring extensive retraining.

The advancements in AD, particularly through the lens of explainable AI, have significant social implications across various sectors:

- **Healthcare:** Improved AD models can enhance the early detection of anomalies in medical data, leading to faster diagnosis and treatment of diseases. The explainability of these models ensures that healthcare professionals can trust and understand the AI's recommendations, ultimately improving patient outcomes.
- **Security and Surveillance:** Enhanced AD systems can play a crucial role in identifying security threats and suspicious activities in real-time, aiding in the prevention of crimes and enhancing public safety. The transparency provided by explainable models ensures that security personnel can make informed decisions based on AI insights.
- **Finance:** In the financial sector, effective and explainable AD models can detect fraudulent activities and irregular transactions, protecting consumers and institutions from financial losses. The interpretability of these models is critical for regulatory compliance and maintaining trust among stakeholders.

- **Industrial Applications:** AD models can be used to monitor the health and performance of industrial machinery, predicting failures and reducing downtime. Explainable AI ensures that maintenance teams understand the reasons behind predictions, enabling more effective and timely interventions.
- **Ethical AI Deployment:** By prioritizing explainability, this research contributes to the development of ethical AI systems. Transparent AD models promote accountability and fairness, ensuring that AI-driven decisions can be audited and understood by human operators.

Thus, the contributions of this thesis not only advance the technical capabilities of anomaly detection but also emphasize the importance of explainability in fostering trust and ethical deployment of AI systems. These advancements have the potential to bring about significant positive social changes across various critical sectors, improving safety, security, and quality of life.



## References

- [1] A. Abdallah, M. A. Maarof, and A. Zainal, “Fraud detection system: A survey,” *Journal of Network and Computer Applications*, vol. 68, pp. 90–113, 2016.
- [2] M. Ahmed, A. N. Mahmood, and J. Hu, “A survey of network anomaly detection techniques,” *Journal of Network and Computer Applications*, vol. 60, pp. 19–31, 2016.
- [3] M. Ahmed, A. N. Mahmood, and M. R. Islam, “A survey of anomaly detection techniques in financial domain,” *Future Generation Computer Systems*, vol. 55, pp. 278–288, 2016.
- [4] B. R. Kiran, D. M. Thomas, and R. Parakkal, “An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos,” *Journal of Imaging*, vol. 4, no. 2, p. 36, 2018.
- [5] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, “Unsupervised anomaly detection with generative adversarial networks to guide marker discovery,” in *International conference on information processing in medical imaging*. Springer, 2017, pp. 146–157.
- [6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” *Advances in neural information processing systems*, vol. 27, 2014.
- [7] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, “Ganomaly: Semi-supervised anomaly detection via adversarial training,” in *Computer Vision—ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part III 14*. Springer, 2019, pp. 622–637.
- [8] S. Akçay, A. Atapour-Abarghouei, and T. P. Breckon, “Skip-ganomaly: Skip connected and adversarially trained encoder-decoder anomaly detection,” in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–8.
- [9] W. Ågren, “The nt-xent loss upper bound,” *arXiv preprint arXiv:2205.03169*, 2022.
- [10] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” in *International conference on machine learning*. PMLR, 2020, pp. 1597–1607.
- [11] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A survey,” *ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, 2009.
- [12] C. C. Aggarwal and S. Sathe, “Theoretical foundations and algorithms for outlier ensembles,” *Acm sigkdd explorations newsletter*, vol. 17, no. 1, pp. 24–47, 2015.

- [13] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [14] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural networks*, vol. 61, pp. 85–117, 2015.
- [15] L. Akoglu, H. Tong, and D. Koutra, “Graph based anomaly detection and description: a survey,” *Data mining and knowledge discovery*, vol. 29, pp. 626–688, 2015.
- [16] W. Hamilton, Z. Ying, and J. Leskovec, “Inductive representation learning on large graphs,” *Advances in neural information processing systems*, vol. 30, 2017.
- [17] A. Z. Olivier Chapelle, Bernhard Schölkopf, *Semi-Supervised Learning*. MIT Press, 2006.
- [18] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” *arXiv preprint arXiv:1312.6114*, 2013.
- [19] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, “Adversarial autoencoders,” *arXiv preprint arXiv:1511.05644*, 2015.
- [20] J. An and S. Cho, “Variational autoencoder based anomaly detection using reconstruction probability,” *Special lecture on IE*, vol. 2, no. 1, pp. 1–18, 2015.
- [21] S. Zhao, J. Song, and S. Ermon, “Infovae: Balancing learning and inference in variational autoencoders,” in *Proceedings of the aaai conference on artificial intelligence*, vol. 33, 2019, pp. 5885–5892.
- [22] J. Donahue, P. Krähenbühl, and T. Darrell, “Adversarial feature learning,” *arXiv preprint arXiv:1605.09782*, 2016.
- [23] J. Donahue and K. Simonyan, “Large scale adversarial representation learning,” *Advances in neural information processing systems*, vol. 32, 2019.
- [24] P. J. Rousseeuw and M. Hubert, “Anomaly detection by robust statistics,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 2, p. e1236, 2018.
- [25] E. Eskin, A. Arnold, M. Prerau, L. Portnoy, and S. Stolfo, “A geometric framework for unsupervised anomaly detection: Detecting intrusions in unlabeled data,” *Applications of data mining in computer security*, pp. 77–101, 2002.
- [26] F. E. Grubbs, “Procedures for detecting outlying observations in samples,” *Technometrics*, vol. 11, no. 1, pp. 1–21, 1969.
- [27] ———, *Sample criteria for testing outlying observations*. University of Michigan, 1949.
- [28] D. Hawkins and D. Hawkins, “A single outlier in normal samples,” *Identification of Outliers*, pp. 27–41, 1980.
- [29] D. M. Hawkins, “Multivariate outlier detection,” in *Identification of outliers*. Springer, 1980, pp. 104–114.
- [30] L. E. Peterson, “K-nearest neighbor,” *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009.

- [31] S. Ramaswamy, R. Rastogi, and K. Shim, “Efficient algorithms for mining outliers from large data sets,” in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000, pp. 427–438.
- [32] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, “Lof: identifying density-based local outliers,” in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000, pp. 93–104.
- [33] J. A. Hartigan and M. A. Wong, “Algorithm as 136: A k-means clustering algorithm,” *Journal of the royal statistical society. series c (applied statistics)*, vol. 28, no. 1, pp. 100–108, 1979.
- [34] C. C. Aggarwal and K. Subbian, “Anomaly detection in large datasets: A survey,” *ACM Computing Surveys (CSUR)*, vol. 49, no. 1, p. 15, 2016.
- [35] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *kdd*, vol. 96, 1996, pp. 226–231.
- [36] I. J. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and harnessing adversarial examples,” *arXiv preprint arXiv:1412.6572*, 2014.
- [37] R. Murugesan, E. Mishra, and A. H. Krishnan, “Deep learning based models: Basic lstm, bi lstm, stacked lstm, cnn lstm and conv lstm to forecast agricultural commodities prices,” 2021.
- [38] S. Sabour, N. Frosst, and G. E. Hinton, “Dynamic routing between capsules,” *Advances in neural information processing systems*, vol. 30, 2017.
- [39] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [40] K. Cho and e. a. Van Merriënboer, “Learning phrase representations using rnn encoder-decoder for statistical machine translation,” *arXiv preprint arXiv:1406.1078*, 2014.
- [41] W. Yu, W. Cheng, C. C. Aggarwal, K. Zhang, H. Chen, and W. Wang, “Netwalk: A flexible deep embedding approach for anomaly detection in dynamic networks,” in *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 2018, pp. 2672–2681.
- [42] Y. Pei, T. Huang, W. van Ipenburg, and M. Pechenizkiy, “Resgcn: Attention-based deep residual modeling for anomaly detection on attributed networks,” *Machine Learning*, vol. 111, no. 2, pp. 519–541, 2022.
- [43] X. Teng, M. Yan, A. M. Ertugrul, and Y.-R. Lin, “Deep into hypersphere: Robust and unsupervised anomaly discovery in dynamic networks,” in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, 2018.
- [44] H. Wang and e. a. Zhou, “Deep structure learning for fraud detection,” in *2018 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2018, pp. 567–576.
- [45] I. Marin, S. Gotovac, and M. Russo, “Evaluation of generative adversarial network for human face image synthesis,” in *2020 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, 2020.

- [46] K. Wongsuphasawat and e. a. Smilkov, “Visualizing dataflow graphs of deep learning models in tensorflow,” *IEEE transactions on visualization and computer graphics*, vol. 24, no. 1, pp. 1–12, 2017.
- [47] D. T. Huff, A. J. Weisman, and R. Jeraj, “Interpretation and visualization techniques for deep learning models in medical imaging,” *Physics in Medicine & Biology*, vol. 66, no. 4, p. 04TR01, 2021.
- [48] L. H. Gilpin and e. a. Bau, “Explaining explanations: An overview of interpretability of machine learning,” in *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*. IEEE, 2018, pp. 80–89.
- [49] J. Xu, J. Chen, S. You, Z. Xiao, Y. Yang, and J. Lu, “Robustness of deep learning models on graphs: A survey,” *AI Open*, vol. 2, pp. 69–78, 2021.
- [50] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, “Momentum contrast for unsupervised visual representation learning,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 9729–9738.
- [51] O. Henaff, “Data-efficient image recognition with contrastive predictive coding,” in *International conference on machine learning*. PMLR, 2020, pp. 4182–4192.
- [52] A. Krizhevsky, V. Nair, and G. Hinton, “Cifar-10 (canadian institute for advanced research),” *URL <http://www.cs.toronto.edu/kriz/cifar.html>*, vol. 5, no. 4, p. 1, 2010.
- [53] Y. LeCun, C. Cortes, C. Burges *et al.*, “Mnist handwritten digit database,” 2010.
- [54] S. Akcay, M. E. Kundegorski, C. G. Willcocks, and T. P. Breckon, “Using deep convolutional neural network architectures for object classification and detection within x-ray baggage security imagery,” *IEEE transactions on information forensics and security*, vol. 13, no. 9, pp. 2203–2215, 2018.
- [55] T. Hassan, S. Akçay, M. Bennamoun, S. Khan, and N. Werghe, “Unsupervised anomaly instance segmentation for baggage threat recognition,” *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–12, 2021.
- [56] C. X. Ling, J. Huang, H. Zhang *et al.*, “Auc: a statistically consistent and more discriminating measure than accuracy,” in *Ijcai*, vol. 3, 2003, pp. 519–524.
- [57] H. Zenati, C. S. Foo, B. Lecouat, G. Manek, and V. R. Chandrasekhar, “Efficient gan-based anomaly detection,” *arXiv preprint arXiv:1802.06222*, 2018.

## Proof of Publishing

Proofs for the paper: Shikhar Asthana and Anurag Goel, “Unveiling Anomalies: A Review of Anomaly Detection Through Lens of Explainable AI” accepted to be published in “3rd International Conference on ‘Smart Technologies and Systems for Next Generation Computing’” to be held on July 18 - 19, 2024 at IFET College of Engineering, Tamil Nadu, India:

### 1. Acceptance Mail

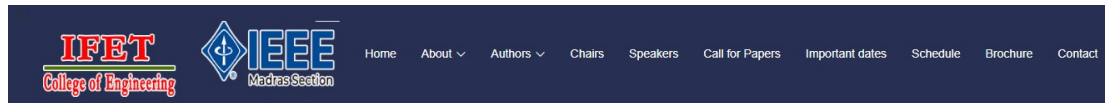
From: ICSTSN 2024 <icstsn2024@ifet.ac.in>  
Date: Thu, May 23, 2024 at 3:41 PM  
Subject: ICSTSN 2024  
To: Anurag Goel <anurag@dtu.ac.in>

Dear Author,

#### **Congratulations!!!**

The review and selection process for your paper ID ICSTSN 424 entitled “Unveiling Anomalies: A Review of Anomaly Detection Through Lens of Explainable AI” has been completed. Based on the recommendations from the reviewers assigned for your paper, I am pleased to inform you that your paper has been **ACCEPTED** by the Technical Program Committee (TPC) for ORAL PRESENTATION which is organized by IFET College of Engineering, Villupuram, Tamil Nadu, India during 18<sup>th</sup> - 19<sup>th</sup>, July 2024. I am also glad to inform you that the proceedings of ICSTSN 2024 will be submitted for inclusion in IEEE Xplore.

### 2. Scopus Indexed



#### **Publication**

All accepted and presented papers will be published in conference proceedings and will be eligible for submission to be included in IEEE Xplore Digital Library (Scopus indexed)

### 3. Registration Payment Receipt



#### **E-receipt**

Transaction Reference Number:	DES1756828
Transaction Type	NEFT Fund Transfer
Date of Transaction	24/05/2024
From ICICI Bank Account	628401578785-SHIKHAR ASTHANA
To payee	03780200000501-ICSTSN
Transaction Amount(Rs.)	INR 7,000.00
Remarks	ICSTSN 424

Proofs for the paper: Shikhar Asthana and Anurag Goel, “ConGANomaly: A Contrastive Learning Approach Of Anomaly Detection Using Generative Adversarial Networks” accepted to be published in “The 15th International IEEE Conference on Computing, Communication and Networking Technologies (ICCCNT)” to be held on June 24 - 28, 2024 at IIT - Mandi, Himachal Pradesh, India:

## 1. Acceptance Mail

From: 15th ICCCNT 2024 <[15thicccnt2024@easychair.org](mailto:15thicccnt2024@easychair.org)>  
Date: Sun, 26 May, 2024, 8:57 am  
Subject: 15th ICCCNT 2024 submission 2424  
To: Anurag Goel <[anurag@dtu.ac.in](mailto:anurag@dtu.ac.in)>

"Dear Authors,  
Paper ID:2424  
Title: ConGANomaly: A Contrastive Learning Approach Of Anomaly Detection Using Generative Adversarial Networks

Congratulations! Your paper got accepted.

Similarity/Plagiarism Index: 9.7%

1. Why leakyReLU is used as a activation function in the proposed architecture?
2. All the variables used in the equation should be explained in the content.
3. Provide formula for all the evaluation metrics used in the work.
4. Provide a table indicating total number of datasets and number taken for training, testing and validation.
5. How the generalizability of the model is achieved?
6. Work flow of the paper is good.

Author affiliation and paper should be in IEEE conference template  
(<https://www.ieee.org/conferences/publishing/templates.html>)

\*\*\*Complete the registration process immediately after receiving this email in order to prepare the presentation schedule (<https://15icccnt.com/register/index.php>)

For making payments (Indian Authors), using following bank account

## 2. Scopus Indexed



**THE 15th INTERNATIONAL IEEE CONFERENCE ON COMPUTING, COMMUNICATION AND NETWORKING TECHNOLOGIES (ICCCNT)**

June 24 - 28, 2024, IIT - Mandi, Himachal Pradesh, India.

A 5 Day Hybrid Technical Fest - participate online / offline

[Home](#)   [General Info](#) ▾   [About](#) ▾   [Authors](#) ▾   [Program](#) ▾

Call for papers

ICCCNT 2024 will be the next event in a series of highly successful International Conferences on Computing, Communication and Networking, previously held in a blended form in IIT Delhi 2023, in Kharagpur - Virtually (2021,2020), IIT Kanpur (2019), IISc Bengaluru (2018) and various premier locations. The ICCCNT Conference offers an opportunity to bring together researchers, academicians and industrial experts of different disciplines, discuss new issues, tackle complex problems and find advanced solutions breeding new trends in Computational Science.

### 3. Registration Payment Receipt

THE 15th INTERNATIONAL IEEE CONFERENCE ON COMPUTING, COMMUNICATION AND NETWORKING TECHNOLOGIES (ICCCNT)  
June 24 - 28, 2024, IIT - Mandi, Himachal Pradesh, India.

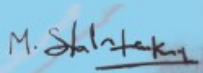
## Payment Receipt

Date: 2024-05-30

<b>Receipt No.</b>	16041		
<b>Author Name.</b>	Shikhar Asthana	<b>Author Address.</b>	B1/423, BLOCK-15, KAILASH DHAM SOCIETY, SECTOR 50, NOIDA, UTTAR PRADESH 201301
<b>Paper ID</b>	2424		
<b>Payment Reference ID</b>	DEU2131639		
<b>Payment Type</b>	neft		
<b>Description</b>	Conference Registration Fee		
		<b>Amount Paid</b>	<b>Rs. 8600.00</b>

Thank you for the payment.

Print



(Authorized Signatory)

