UNIVERSITY OF HULL

A machine learning-based approach for detection of vehicle insurance claim fraud

A dissertation submitted in partial satisfaction
of the requirements for the degree
Master of Science Artificial Intelligence

By

202314466

2024

ABSTRACT

Motor vehicle insurance plays a crucial role in today's economy and everyday life, especially with the growing number of vehicles on the road. It helps cover costs related to accidents, natural disasters, and other damages under comprehensive and third-party liability policies. However, one of the biggest challenges in this industry is insurance fraud, where false claims lead to massive financial losses. These fraudulent activities don't just cost insurance companies billions of dollars; they also hurt honest customers by driving up premiums and diverting resources that could improve customer services and benefits.

Traditionally, detecting fraudulent claims has been a manual process, but it's often slow and prone to errors. With the rise of Artificial Intelligence, insurance companies are now turning to machine learning algorithms, which offer more accuracy and efficiency in detecting fraud.

This study explores how machine learning methods, like Random Forest (RF) and Convolutional Neural Networks (CNN), can classify fraudulent and non-fraudulent images. The research uses a publicly available dataset from Kaggle, which includes training, validation, and test data. Since the training dataset is imbalanced, the study balances it using techniques like Generative Adversarial Networks (GAN), SMOTE, and advanced data augmentation methods, including CutMix, Mixup, and Attention CutMix (Attn_CutMix). For pre-processing, a Deep Convolutional Autoencoder (DCAE) is used to denoise the images before running the classification models. The findings reveal that CNN outperformed RF in terms of accuracy and precision, particularly with SMOTE-balanced data, while also demonstrating the importance of data augmentation techniques in enhancing model robustness.

Keywords: Insurance fraud, Machine Learning, GAN, SMOTE, Data Augmentation, Convolutional Neural Networks, CutMix, Mixup, Attn_CutMix

Table of Contents

ABSTRACT ..... **I**

TABLE OF CONTENTS ..... **II**

LIST OF FIGURES ..... **IV**

LIST OF TABLES ..... **V**

SECTION 1: INTRODUCTION **1**

1.1 BACKGROUND 1

1.2 PROBLEM STATEMENT 3

1.3 THE STUDY OBJECTIVES 5

1.4 SIGNIFICANCE OF THE STUDY 6

SECTION 2: LITERATURE REVIEW **8**

2.1 INSURANCE FRAUD DETECTION MARKET TRENDS 8

2.2 TRADITIONAL METHODS IN FRAUD DETECTION: A MANUAL APPROACH 9

2.3 ADVANCING FRAUD DETECTION: THE ROLE OF AUTOMATION 10

2.4 RELATED WORK 12

2.5 MACHINE LEARNING APPROACHES TO INSURANCE FRAUD DETECTION 14

2.5.1 Random Forests *14*

2.5.2 Convolutional Neural Networks (CNNs) *15*

2.5.3 Data Preparation and Balancing Techniques *16*

2.5.4 Generative Adversarial Networks (GANs) *17*

2.5.5 Data Augmentation *17*

2.5.6 Synthetic Minority Oversampling Technique (SMOTE) *19*

2.5.7 Deep Convolutional Autoencoders (DCAEs) *20*

SECTION 3: RESEARCH METHODOLOGY **22**

3.1 INTRODUCTION 22

3.2 DATA EXPLORATION 22

3.3 DATA PREPARATION 23

3.3.1 Resizing *23*

3.3.2 Denoising of the Images *24*

3.4 ADDRESSING DATA IMBALANCE 25
3.4.1 Generative Adversarial Network (GAN) 25
3.4.2 Data Augmentation 25
3.4.3 SMOTE 25
3.5 CLASSIFICATION MODELS 26
3.5.1 Convolution Neural Networks (CNN) 26
3.5.2 Random Forest 27
3.6 Evaluation 27
3.6.1 Confusion Matrix 27
3.6.2 Performance Metrics for Model Evaluation 28
SECTION 4: RESULTS AND DISCUSSIONS **30**
4.1 INTRODUCTION 30
4.2 MODEL PERFORMANCE EVALUATION 30
4.2.1 Model Performance on Balanced Datasets 30
4.2.1.1 CNN Model Evaluation Using GAN-Balanced Data 30
4.2.1.2 Random Forest Model Evaluation Using GAN-Balanced Data 36
4.2.1.3 Performance Evaluation of CNN on SMOTE Balanced Data 39
4.2.1.4 Performance Evaluation of Random Forest on SMOTE Balanced Data 42
4.2.1.5 Performance Evaluation of CNN on Augmented Data 44
4.2.1.6 Performance Evaluation of Random Forest on Augmented Data 47
4.3 COMPARATIVE ANALYSIS OF MODEL PERFORMANCES 49
SECTION 5: CONCLUSION AND RECOMMENDATIONS **52**
REFERENCES **54**

List of Figures

Figure 1: Benefits of Insurance Fraud Detection by using AI (Rout, 2023).....12
Figure 2: A pictorial representation of how SMOTE works (SWASTIK, 2024) 20
Figure 3: Summary of the CNN model with GAN usage for data balancing 31
Figure 4: Performance of CNN Model with GAN-Based Data Balancing: Confusion Matrix 36
Figure 5: Performance of Random Forest Model with GAN-Based Data Balancing: Confusion Matrix.....38
Figure 6: CNN model summary with SMOTE for data balancing. 39
Figure 7: Performance of CNN Model with SMOTE-Based Data Balancing: Confusion Matrix. 41
Figure 8: Confusion Matrix for Random Forest Model Performance on Test Set. 43
Figure 9: Confusion Matrix illustrating the performance of the model on the test data.....46
Figure 10: Confusion Matrix illustrating the performance of the model on the test data.....48

List of Tables

Table 1: Components of the Confusion Matrix 27
Table 2: CNN Architecture Using GAN-Balanced Data 32
Table 3: Confusion Matrix for CNN Model Evaluation Using GAN-Balanced Data 35
Table 4: Confusion Matrix for Random Forest Model Evaluation Using GAN-Balanced Data 37
Table 5: Performance Evaluation of CNN on SMOTE-Balanced Data 39
Table 6: Confusion Matrix for Performance Evaluation of CNN on SMOTE-Balanced Data. 41
Table 7: Architecture of Random Forest Model on SMOTE-Balanced Data.....42
Table 8: Confusion Matrix for Performance Evaluation of Random Forest on SMOTE-Balanced Data 43
Table 9: Performance Evaluation of CNN on Augmented Data.....44
Table 10: Confusion Matrix for Performance Evaluation of CNN on Augmented Data 46
Table 11: Confusion Matrix for Performance Evaluation of Random Forest on Augmented Data 48
Table 12: Comparative Analysis of Model Performances 49

Section 1: Introduction

1.1 Background

This section outlines the problem statement, establishes the primary aim of the study, details the specific objectives and presents the research questions. The section concludes by highlighting the significance of the study.

Insurance plays a critical role today by providing financial protection against unforeseen events. In exchange for a premium, an insurer agrees to compensate the insured party for losses related to an insurable interest, whether caused by accidents, political disturbances, or natural calamities like floods, typhoons, and tornadoes etc., relieving the insured party of risk related to insurmountable financial damage (Schrijver et al., 2024).

While insurance provides a critical safety net for individuals and businesses, the growing cases of fraudulent activities has revealed significant vulnerabilities within the industry. Fraudulent claims not only undermine the financial stability of insurers but also increase costs for honest policyholders. Moreover, the complexity of modern fraud schemes has outpaced many traditional detection methods, making it increasingly challenging to identify and mitigate fraudulent practices efficiently.

Traditional methods of detecting fraud have often proven to be slow and riddled with vulnerabilities that enable fraudulent behaviors. For example, external or internal investigators can sometimes be compromised. Imagine a scenario where a client files a motor vehicle insurance claim worth $100,000 and offers the investigator a 10% kickback upon claim approval. Such unethical incentives can tempt investigators, allowing fraudulent claims to proceed and causing significant financial losses for insurance companies.

Fraudulent behaviors like these lead to immense losses within the insurance industry. They drive up operational costs, making insurance more expensive for both insurers and policyholders. According to Terry from the Association of British Insurers (2024), fraudulent claims totaling £1.1 billion were detected in 2023, a 4% increase compared to the previous year. Of this, motor vehicle insurance fraud alone accounted for £501 million, representing 54% of all detected fraudulent claims and marking an 8% increase from 2022.

While insurance fraud remains a persistent challenge, considerable progress has been made in addressing it.The Insurance Fraud Detection market, valued at $4.2 billion in 2023, is projected to grow at a compound annual growth rate (CAGR) of over 25% between 2024 and 2032, according to Global Market Insights (2024). This rapid growth is being driven by advancements in Artificial Intelligence (AI), Machine Learning (ML), and big data analytics, which are revolutionizing how fraud is detected and prevented. Innovations like real-time monitoring and predictive analytics, coupled with collaborations between insurers and tech companies, are paving the way for more effective fraud detection solutions worldwide.

Europe has emerged as a leader in combating insurance fraud, accounting for over 35% of the global fraud detection market in 2023, according to Global Market Insights (2024). Countries like France, Italy, the UK, and Germany are making significant progress by adopting AI-driven solutions to detect and prevent fraud. Similarly, North America, particularly the U.S. and Canada, and Asia-Pacific nations such as China, India, Japan, and Australia are heavily investing in AI technologies to combat fraud. For example, in the U.S. alone, nearly 20% of auto insurance payouts annually are estimated to be lost to fraudulent activites (Faheem et al., 2022).

Closer to home, Kenya is also stepping up its efforts to address the growing threat of motor vehicle insurance fraud. As reported by Business Daily (Ashok, 2024), Kenyan insurers are beginning to adopt AI tools to reduce losses and improve their ability to detect fraudulent activities. This, however, is not easy to implement in Kenya because of the significant challenge posed by the availability and quality of data. While artificial intelligence (AI) thrives on large, accurate datasets, Kenya faces difficulties in collecting comprehensive and reliable data, especially in rural and underserved regions (Muiruri, 2023).

Globally, insurance companies and technology providers are working together to develop innovative and tailored solutions that can keep up with evolving fraud tactics. These collaborative efforts are paving the way for more effective fraud detection systems, helping to protect both insurers and their customers.With developing technology, insurance companies have resulted to looking for cutting edge solutions to improve fraud detection. This paper suggests machine learning methods could be used to improve fraud detection, with the primary

goal being to classify images as fraudulent or non-fraudulent images submitted to insurance companies for the need of compensation.

Insurance fraud is primarily committed for financial gain and involves acts of deception aimed at obtaining money illegally (Christopher & Aditi, 2020). This encompasses a range of dishonest behaviors, such as staging incidents, falsifying details, and exaggerating claims, all designed to unfairly extract benefits from insurance companies. Recognizing this, insurance companies have taken major steps to reduce losses by investing in technology to combat fraud. This has enhanced their ability to process information and improve the detection and management of fraud-related incidents.

Fraud in the insurance industry can be classified into four categories, according to Frimpong (2016). The first type is internal fraud, where insurance employees deceive the company either on their own or in partnership with others inside or outside the organization. The second type is false claim fraud, where the insured party submits misleading information to support their claim. The third type involves insurance brokers who conspire with policyholders or insurers to commit fraud. Lastly, there is fraud by the insurer, where policyholders are misled through unfair practices, such as hidden clauses, unfair premiums, or dishonest compensation schemes. Motor vehicle insurance fraud is widespread, and traditional fraud detection methods have often been slow, inaccurate, and prone to loopholes. In contrast, machine learning approaches have proven to be highly reliable for fraud detection. These methods offer greater accuracy, provided there is a careful balance between minimizing false positives and correctly identifying fraudulent cases (true negatives). Additionally, machine learning models can process vast amounts of data in real-time, adapt to emerging fraud patterns, and continuously improve over time, making them a more efficient and scalable solution compared to manual processes.

1.2 Problem Statement

Automobile insurance fraud detection is a critical concern for insurers, leading to substantial financial losses. Traditional methods, such as manual claims review and rule-based systems, often struggle to identify sophisticated fraudulent activities, including staged accidents and inflated damage costs. To address these challenges, machine learning has emerged as a promising solution, offering enhanced accuracy and efficiency in detecting complex fraud patterns.

For instance, Allianz UK has developed a machine-learning tool named 'Incognito' to combat fraudulent claims. This tool analyzes claims data to identify potentially fraudulent activities, which are then reviewed by fraud experts for further investigation. Since its implementation, Allianz has reported significant savings, with £1.7 million saved to date and an additional £3.4 million held in claim reserves pending investigation outcomes (Allianz, 2023). However, Allianz's tool faces limitations, including its reliance on expert validation, which increases processing times and operational costs, potentially limiting the system's scalability. Furthermore, its dependency on structured claims data makes it less effective in detecting fraud patterns embedded in unstructured data sources, such as customer communications or external reports. This gap highlights the need for methods that can adapt to diverse and incomplete data inputs. Additionally, the system's capability to handle emerging fraud schemes remains a concern, as it primarily focuses on known patterns.

Similarly, other insurers, like Wipro, have explored various machine learning techniques to identify potential insurance claim frauds with high accuracy. Wipro has applied supervised learning models, such as random forests and logistic regression, as well as unsupervised techniques like anomaly detection to address fraud detection challenges. Additionally, the use of natural language processing (NLP) allows them to analyze unstructured data from claims forms, communications, and reports (Guha et al., 2023). While these advancements are promising, Wipro's approaches also face notable limitations. Challenges such as managing false positives, ensuring model interpretability, and adapting to evolving fraud tactics persist. Moreover, scalability to handle growing data volumes efficiently remains an ongoing issue.

This study aims to tackle the limitations faced by existing solutions by developing a binary classification model for motor vehicle insurance claims using machine learning techniques, specifically Convolutional Neural Networks (CNN) and Random Forests (RF). Unlike the models employed by Allianz and Wipro, which rely heavily on structured data and manual validation, this study will focus on strategies to address class imbalance, which is a significant challenge in fraud detection. Class imbalance often create biaseness in model predictions, leading to either excessive false positives or undetected fraudulent claims.

To mitigate this, the study will investigate advanced data balancing methods such as:

- **SMOTE (Synthetic Minority Over-sampling Technique):** A technique to generate synthetic samples for the minority class, addressing the imbalance directly.

• **Generative Adversarial Networks (GANs):** To create realistic fraudulent claims, improving the diversity of the training set.
• **Data Augmentation Methods:** Techniques like CutMix, Attn_CutMix, and Mixup will be used to modify existing data and enhance training diversity, addressing challenges in adaptability and scalability.

By comparing the effectiveness of CNN and RF models trained with these balancing methods, this study seeks to improve on existing solutions in several key areas:

• **Accuracy and False Positives:** Optimizing models to reduce false positives while maintaining high fraud detection rates.
• **Scalability:** Ensuring the techniques can handle large and complex datasets efficiently.
• **Model Interpretability:** Utilizing simpler models like RF alongside CNNs to balance performance with transparency.
• **Adaptability:** Leveraging data augmentation and GANs to prepare models for evolving fraud patterns.

Through these improvements, the study aims to develop a more robust, efficient, and adaptable system for motor vehicle insurance fraud detection, addressing the limitations of current tools used by industry leaders such as Allianz and Wipro.

1.3 The Study Objectives

The primary objective of this study is to explore how machine learning techniques can be applied to detect vehicle insurance claim fraud. Specifically, the goal is to develop a model that predicts and performs binary classification on images in the dataset, categorizing them as either fraudulent or non-fraudulent. Below, specific objectives are what the study aims at achieving.

i. **Develop an Enhanced Model for Detecting Fraudulent Auto Insurance Claims:** Design a machine learning-based model specifically aimed at accurately identifying fraudulent auto insurance claims. This model will incorporate advanced methodologies, including deep learning techniques like Convolutional Neural Networks for image analysis and ensemble learning methods such as Random Forest. These approaches will

enhance the model's detection accuracy, reliability, and adaptability to varied data patterns.

ii. **Implement the Enhanced Detection Model:** Build and deploy the developed fraud detection model. Employ data balancing strategies such as Generative Adversarial Networks (GAN), Synthetic Minority Oversampling Technique (SMOTE), and Data Augmentation to address class imbalances, ensuring robust and equitable performance across the dataset.

iii. **Evaluate Model Performance Against Internal Benchmarks:** Evaluate the effectiveness of the implemented model by comparing its performance to other state-of-the-art models developed within this study. Metrics such as precision, recall, F1 score, accuracy, and area under the curve (AUC) will be used for a thorough and comprehensive assessment.

1.4 Significance of the study

This study plays a critical role in advancing the detection of fraudulent claims in motor vehicle insurance, offering significant benefits and impacts for the insurance industry. By integrating advanced machine learning models such as Convolutional Neural Networks (CNN) and Random Forest classifiers, the study enhances the accuracy and efficiency of fraud detection. These models allow insurers to move away from outdated, error-prone manual processes, reducing delays and improving fraud detection accuracy. Automating fraud detection contributes to cost savings for insurance companies, as faster and more precise identification of fraudulent claims leads to reduced operational expenses and prevents unnecessary fraudulent payouts.

Beyond financial savings, the study has broader societal impacts. By effectively identifying and mitigating fraudulent claims, machine learning helps reduce the financial burden passed on to consumers through higher insurance premiums. As fraud is minimized, premiums can be lowered for all policyholders. Additionally, the rapid and accurate identification of fraudulent claims accelerates the processing of legitimate claims, ensuring policyholders receive rightful compensation more quickly. This improves overall customer satisfaction, builds trust in the insurance system, and enhances the insurer's reputation. Insurers can also allocate more

resources to processing genuine claims, optimizing the claims process and improving operational efficiency.

Furthermore, this study lays the groundwork for the future of AI-driven fraud prevention systems. The scalability and flexibility of the machine learning models developed here make them applicable not only in the insurance industry but across various sectors facing fraud challenges, including banking, healthcare, and e-commerce. These models can be adapted to detect fraudulent activities in different domains, leading to more efficient, automated systems capable of preventing fraud in multiple industries.

As fraud tactics continue to evolve, machine learning models offer the flexibility to adapt continuously, ensuring that organizations can stay ahead of emerging threats and protect their operations. This adaptability ensures that the insurance industry can maintain the integrity of their operations in a changing fraud landscape. In essence, this study provides the insurance industry with a powerful tool to improve fraud detection, reduce costs, and enhance customer satisfaction, while also paving the way for future innovations in AI-powered fraud prevention.

Section 2: Literature Review

This section explores the global landscape of motor vehicle insurance and the diverse methods used by different countries to identify fraudulent claims. It will cover both manual and automated systems in place, highlighting how automated fraud detection systems leverage machine learning and deep learning techniques to combat fraudulent claims effectively.

2.1 Insurance Fraud Detection Market Trends

According to Preeti Wadhwani (2024), the global insurance fraud detection market was valued at USD 4.2 billion in 2023 and is projected to grow at a compound annual growth rate (CAGR) of over 25% from 2024 to 2032. This rapid expansion is driven by advancements in Artificial Intelligence (AI), Machine Learning (ML), and big data analytics. Key factors contributing to this growth include the adoption of real-time monitoring, predictive analytics, and increased collaboration between insurers and technology providers. The primary focus remains on proactive measures to tackle increasingly complex fraudulent activities across various insurance sectors.

For instance, the Équité Association (2023), a Canadian national organization dedicated to combating insurance fraud, launched EQ Insights in October 2023, a platform designed to detect fraud. This advanced insurance crime detection platform is designed to combat fraud and insurance-related crimes across Canada by leveraging cutting-edge technologies such as enhanced network link analysis, intelligence-driven fraud alerts, and comprehensive reporting. These innovations significantly improve the detection and prevention of fraudulent activities.
The rapid growth of digital transactions within the insurance sector has emerged as a major catalyst for the development of fraud detection solutions. With more processes like policy issuance, claims handling, and premium payments being conducted online, the sheer volume and complexity of transactions have increased, creating new avenues for fraudulent activities. This underscores the critical need for advanced fraud detection systems capable of processing large volumes of digital data in real-time, detecting anomalies, and identifying suspicious patterns.

Globally, several markets have embraced technology for fraud detection. European countries have led the way, with the region holding over 35% of the global insurance fraud detection market in 2023, according to Preeti Wadhwani (2024). Key markets such as Italy, France, the

UK, and Germany have seen significant growth due to the increasing adoption of digital insurance processes and sophisticated fraud detection technologies.

In addition, countries such as China, India, Japan, Australia, Kenya, Nigeria, and several other African nations have also made strides in fraud detection by implementing advanced technologies like AI, Machine Learning (ML), and blockchain. Collaborations between insurance companies and technology firms have fostered innovation and the creation of tailored solutions to address the unique fraud challenges present in each market.

2.2 Traditional Methods in Fraud Detection: A Manual Approach

Traditionally, fraud detection has relied on rule-based systems and manual reviews. This involves setting specific rules to identify suspicious behaviour and manually examining activities that fall outside these rules. It also includes auditing, where trained individuals review reports to detect fraud. While these methods have been effective in the past, they are often slow, costly, and prone to errors. With the large amounts of data today, manual fraud detection has become increasingly difficult and inefficient in keeping up with the complexity of modern transactions (Jarrod et al.,).

Traditionally, fraud detection in Kenya has involved a manual process, heavily reliant on human expertise. According to the Association of Kenya Insurers (2020), insurance companies use trained agents to analyse and assess each claim to determine if it is genuine or fraudulent. These agents base their judgments on data collected at the scene and investigation reports. The Insurance Regulatory Authority (2021) outlines that policyholders are required to collect key details at the scene of an accident, such as the driver's name, license information, insurance status (whether valid or expired), and vehicle details like registration number, make, and year of manufacture.

Once a claim is submitted, a supervisor at the insurance company manually reviews it, scoring the claim based on a checklist related to the specifics of the damaged components. If a claim receives a high score, the case is forwarded to an investigator, who conducts a thorough investigation and generates a report. The supervisor then makes the final decision on whether the claim is fraudulent or genuine based on the findings (Association of Kenya Insurers, 2020).

However, this manual approach has several loopholes that can easily facilitate fraud:

i. Compromise and Corruption: Supervisors or investigators can be bribed to alter their findings in favour of the claimant, leading to fraudulent claims being approved.
ii. Human Limitations: The process relies heavily on human knowledge and judgment, which can be flawed, inconsistent, or limited by experience.
iii. Inconsistency and Subjectivity: The manual evaluation process can lead to inconsistencies in decision-making, as different supervisors may interpret the same case differently, increasing the chances of errors.

2.3 Advancing Fraud Detection: The Role of Automation

Automated Fraud Detection in motor vehicle insurance is a technology-driven solution designed to identify and prevent fraudulent claims. It uses advanced algorithms and artificial intelligence to analyse claim patterns, detect unusual activities, and flag potential fraud in real-time. This system helps insurance companies protect themselves from financial losses and reputational damage, while also making the fraud detection process faster and more accurate. By reducing the need for manual reviews, automated fraud detection improves the overall efficiency of handling motor vehicle insurance claims (Tookitaki, 2023).

The automated system uses advanced algorithms to interpret patterns in large datasets. Machine learning models, a key part of these systems, are designed to learn and adapt over time, much like detectives who continuously refine their skills. These systems don't just perform surface-level checks; they dig deep into vast amounts of data, similar to searching for a needle in a digital haystack. By analysing the data and other key features of the dataset, they identify patterns that deviate from the norm, effectively detecting potential fraud. And all these happen in real-time.

One of the main strengths of automated fraud detection is its ability to learn from experience. Machine learning algorithms, similar to adaptive learners, constantly evolve by gathering insights from each case they analyse. This ongoing learning process improves their accuracy over time, helping the system adjust to new fraud techniques. As a result, the chance of false positives is reduced, allowing for more precise and reliable identification of potential threats (Tookitaki, 2023).

According to Nicholas Martino (2023), the impact of automating fraud detection is significant, offering protection not only just from financial losses but also helping maintain a company's reputation and customer trust. As illustrated in Figure 2 below, some of the key advantages include:

i. Proactive Fraud Detection

Predictive analytics offers the chance to detect fraud before it happens. With today's digital fraud schemes, advanced data mining and behaviour-based algorithms are essential for identifying scams early. This technology helps uncover the root causes of fraudulent activity, allowing insurers to predict and prevent fraud in a timely manner.

ii. Faster Fraud Detection

AI not only speeds up the fraud detection process but also identifies patterns that allow for early warnings and quick responses to suspicious activities. As client numbers grow, claims adjusters face pressure to balance speed and accuracy. However, the more data ML receives, the faster and more accurately it can detect fraud.

iii. Accurate Fraud Detection

Predictive analytics provides far more accurate results than manual processes. By processing large amounts of data, ML tools can make decisions with unprecedented precision, surpassing the abilities of human agents.

iv. Reduced Human Interventions

By maximizing technology and data analytics, insurers can reduce the need for manual reviews in claims processing. This not only speeds up the process but also allows insurance agents to focus on more critical tasks.

v. Cost Savings

AI's improved accuracy and reduction of false positives help insurers minimize financial losses. Automation also reduces the need to hire more staff as businesses scale, resulting in cost savings.

vi. Enhanced Customer Experience

With AI-powered fraud detection reducing costs, insurers can offer more competitive insurance plans and an overall improved experience for customers.

Figure 1: Benefits of Insurance Fraud Detection by using AI (Rout, 2023).

2.4 Related Work

Fraud detection has been a major focus in machine learning research, especially in financial systems like automobile insurance. Stolfo et al. (1997a, 1997b) laid important groundwork by introducing meta-learning techniques, where multiple classifiers were combined to improve the detection of fraudulent transactions. In 2000, Stolfo and his team took this further by using cost-based metrics to train and evaluate fraud detection systems, tailoring them to financial data. Artis et al. (2002) used a binary choice model for detecting fraud in automobile insurance claims, incorporating assumptions about misclassification into logistic regression frameworks. Phua et al. (2004) improved on ensemble methods by combining bagging and stacking techniques, showing how this hybrid approach outperformed simpler models. Pathak et al. (2005) applied fuzzy logic systems to flag fraudulent claims, while Pinquet et al. (2007) used a statistical two-equation model to analyze fraud risk with data from a Spanish insurer.

Bermúdez et al. (2008) also worked with Spanish insurance data, using a modified version of a Naive Bayes model to detect fraud.

In the 2010s, neural networks and other advanced methods became more common. Xu et al. (2011) introduced a neural network ensemble that used random rough subspace techniques to improve consistency in datasets. Tao et al. (2012) tackled issues with overlapping samples in fraud detection by using a fuzzy support vector machine with dual membership functionality. Around the same time, Benard and Vanduffel (2014) used portfolio optimization techniques, like the Sharpe ratio, to make fraud detection strategies more effective.

As the effort to detect fraud using machine learning progressed, addressing the challenge of imbalanced datasets became increasingly important. Sundarkumar and Ravi (2015) developed a method combining k-Reverse Nearest Neighborhood with one-class support vector machines, which enhanced performance in fraud detection for datasets with a low proportion of fraudulent claims. Nian et al. (2016) introduced an anomaly detection technique called spectral ranking (SRA), which outperformed traditional outlier detection methods for auto insurance claims. Similarly, Subudhi and Panigrahi (2017) applied genetic algorithms and fuzzy clustering to improve the performance of classifiers in detecting automobile insurance fraud.

In 2018, Yaqi et al. improved Random Forest for automobile insurance fraud identification by integrating Principal Component Analysis (PCA) and Potential Nearest Neighbor (PNN) methods, boosting classifier accuracy and diversity. In his 2022 paper, Krishna Kanth Kondapaka explores contemporary deep learning architectures, specifically Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), in the context of auto insurance. CNNs are highlighted for their effectiveness in recognizing patterns in image data, making them suitable for applications like fraud detection in auto insurance. RNNs, on the other hand, are well-suited for analyzing sequential data, such as customer behavior patterns. The paper discusses the core functionalities and strengths of each approach, emphasizing their applicability to various auto insurance-related tasks (Kondapaka, 2022).

Building on this rich legacy, we seek to enhance fraud detection methodologies by incorporating both traditional and advanced machine learning techniques. Inspired by Kondapaka's (2022) exploration of deep learning methods, we leverage Random Forests and Convolutional Neural Networks (CNNs) in our approach. Random Forests serve as a reliable

ensemble method for structured data, while CNNs are employed to capture intricate patterns in image data. These methodologies address critical challenges, such as handling imbalanced datasets and ensuring scalability to real-world applications, with the goal of achieving higher accuracy and robustness in detecting fraudulent insurance claims.

2.5 Machine Learning Approaches to Insurance Fraud Detection

Machine learning (ML) has become a valuable tool for improving fraud detection in the insurance industry. ML algorithms are highly effective at analysing large amounts of data to detect patterns and anomalies that suggest fraudulent activity. Techniques like logistic regression, decision trees, random forests, and gradient boosting are especially useful in this area. These methods learn from past labelled data to differentiate between legitimate and fraudulent claims with a high degree of accuracy. By examining historical claims data, ML models can spot suspicious patterns that traditional methods might miss. This not only makes fraud detection more accurate but also increases efficiency by automating the process of flagging and prioritizing suspicious claims (Chandravanshi et al., 2024).

This section will explain the key machine learning models the study used for classifying insurance claims, particularly for image-based fraud detection tasks. The focus will be on Random Forests, CNNs how to prepare and balance data using methods like GAN, SMOTE, and augmentation.

2.5.1 Random Forests

Random Forest is a widely-used ensemble learning method that builds upon the strengths of decision trees, providing higher accuracy and robustness in classification tasks. By constructing multiple decision trees on random subsets of the data, the Random Forest algorithm averages the results to produce a final prediction, reducing the risk of overfitting. Each tree in the forest contributes to the overall prediction by voting, and the class (fraud or non-fraud) with the majority of votes becomes the final output.

In the context of image-based fraud detection, Random Forest can be applied by first extracting relevant features from the images (such as damage patterns, texture, or colour) and then using these features as input for the model. One of the key advantages of Random Forest is its ability

to handle high-dimensional data with multiple features, making it well-suited for image classification tasks where the visual characteristics may vary significantly between cases.

Advantages of Random Forest for Binary Classification of Images:

- **Handling of large feature sets:** Random Forest is capable of processing large numbers of input features, which is ideal when analysing complex image datasets where multiple characteristics (e.g., pixel values, colour histograms, or manually extracted features) are important for classification.
- **Robustness against overfitting:** By averaging the predictions of multiple trees, Random Forest reduces overfitting, leading to more reliable and generalizable models.
- **Interpretability:** Despite its complexity, Random Forest models can provide insights into feature importance, helping us understand which image characteristics contribute most to fraud detection.
- **Flexibility:** Random Forest can be used for both structured data and images, making it a versatile tool in insurance fraud detection.

In our study, Random Forests were used to classify images of car accidents as fraudulent or non-fraudulent by analysing key visual features such as damage patterns, textures, and colour distributions. The model utilized multiple decision trees, each trained on different subsets of image data, to identify patterns associated with fraudulent claims. By aggregating the outputs of these decision trees, Random Forests provided a robust classification method that effectively distinguished between legitimate and suspicious claims. This ensemble approach allowed the model to consider a wide range of features extracted from the images, ensuring a comprehensive analysis for fraud detection.

2.5.2 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) represent one of the most advanced and specialized deep learning architectures for image classification tasks. Unlike traditional models like Random Forest that require explicit feature extraction, CNNs learn to identify important features directly from raw image data through a series of convolutional and pooling layers. These layers detect local patterns, such as edges, shapes, and textures, which are then combined to form a high-level understanding of the image content.

CNNs are particularly effective for binary classification of images in the insurance sector, where claims often involve photographic evidence of damage. The model can be trained to classify images as either fraudulent or non-fraudulent by learning to recognize specific visual patterns that distinguish between legitimate and suspicious claims.

Advantages of CNNs for Binary Classification of Images:

- **Automated feature extraction:** CNNs learn hierarchical representations of the image data, eliminating the need for manual feature extraction. This makes them highly effective for complex image datasets where patterns are difficult to quantify.
- **High accuracy with large datasets:** CNNs performs exceptionally well when trained on large datasets, as they can capture fine-grained details and patterns that may not be visible through manual analysis.
- **Spatial awareness:** The convolutional layers in CNNs preserve the spatial relationships between pixels, allowing the model to understand the structure and arrangement of objects within an image, which is critical when analysing car accident claims.
- **End-to-end learning:** CNNs are trained in an end-to-end fashion, meaning that the model learns to classify images directly from the pixel data, without requiring intermediate steps like manual feature extraction.

In our study, CNNs were employed to classify images of car accidents as fraudulent or non-fraudulent based on damage patterns. Through multiple layers of convolution and pooling, the CNN was able to detect and capture subtle visual differences between legitimate and suspicious claims, leading to highly accurate predictions.

2.5.3 Data Preparation and Balancing Techniques

One of the major challenges in fraud detection, especially with image-based data, is the imbalance between legitimate and fraudulent claims, with legitimate claims often outnumbering fraudulent ones. This imbalance can lead to models being biased toward predicting non-fraudulent claims. To handle this issue and improve model robustness, we employed several data balancing techniques, including Generative Adversarial Networks (GANs), SMOTE (Synthetic Minority Over-sampling Technique), data augmentation, and

Deep Convolutional Autoencoders (DCAEs). These methods help create a more diverse and balanced dataset, enabling the models to learn more effectively and perform better.

GANs, SMOTE, and data augmentation were specifically used to address the class imbalance by generating synthetic fraudulent samples. While GANs and SMOTE created new instances of the minority class, data augmentation techniques were applied to multiply the existing images of the minority class through transformations such as rotations, shifts, and zooms. This increased the diversity of the dataset, allowing the model to generalize better and reduce bias toward the majority class. Additionally, DCAEs were employed to pre-process and denoise the images, enhancing their quality before they were fed into the model. By combining these methods, we were able to create a well-rounded dataset that significantly improves the model's capacity to distinguish between fraudulent and non-fraudulent claims.

2.5.4 Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) are a deep learning technique that helps create synthetic data through a unique two-part system: the generator and the discriminator. The generator is responsible for creating new, realistic data samples, while the discriminator works to distinguish between real and fake data. This back-and-forth competition between the two networks helps both improve over time, resulting in the generation of high-quality synthetic data.

In fraud detection, particularly when working with images, GANs are especially useful for tackling the problem of class imbalance. Fraudulent claims are often underrepresented in real-world datasets, which can cause models to become biased toward predicting legitimate claims. GANs help by generating synthetic fraudulent images that closely resemble real ones, thus increasing the number of fraud examples and creating a more balanced dataset.

2.5.5 Data Augmentation

Data augmentation is a technique used to artificially expand the size of a dataset by creating new variations of the existing data. This is done by applying various transformations to the original data, such as rotations, flipping, shifts and zooms, which results in modified data points that still retain the core characteristics of the original data. In the context of image augmentation, it can include flipping an image horizontally, rotating it, adjusting brightness,

or zooming in and out (Haseeb et al. 2022). Advanced augmentation techniques, such as MixUp, CutMix, and Attention CutMix, can further enhance the diversity of training data.

• **MixUp** blends two images and their corresponding labels to create new samples. For two images x_i and x_j with labels y_i and y_j, the new augmented image x_{mix} and label y_{mix} are defined as:

$$x_{mix} = \lambda x_i + (1-\lambda)x_j, \quad y_{mix} = \lambda y_i + (1-\lambda)y_j$$

where $\lambda \sim \text{Beta}(\alpha, \alpha) and \alpha > 0$. This smooth interpolation between samples encourages the model to generalize better by exposing it to mixed representations of the data (Zhang et al., 2018).

• **CutMix** replaces a rectangular region of one image with the corresponding region from another image while proportionally mixing their labels. If x_i and x_j are the original images, and $y_j$and $y_j$are their labels, the augmented image x_{cut} and label y_{cut} are given as:

$$x_{\text{cut}} = M \odot x_i + (1-M) \odot x_j, \quad y_{\text{cut}} = \lambda y_i + (1-\lambda)y_j$$

where M is a binary mask that defines the cut region, $\lambda = \frac{\text{area of cut region}}{\text{total area of the image}}$, and $\odot$ represents element-wise multiplication. This technique allows the model to combine spatially distinct features from different images effectively (Yun et al., 2019).

• **Attention CutMix** builds upon CutMix by focusing on semantically important regions of the image. The binary mask M is determined using an "attention map" that highlights regions with high significance (e.g., areas containing objects of interest). The mathematical formulation remains similar to CutMix but integrates an additional step to calculate M based on the attention map. This ensures that the regions being mixed contribute more meaningfully to the model's learning process (Huang et al., 2023).

In imbalanced datasets, where one class (e.g., non-fraudulent images) significantly outweighs another (e.g., fraudulent images), data augmentation can help address this imbalance by artificially increasing the representation of the minority class. By generating diverse variations of the minority class, such as fraud cases in the context of fraud detection, augmentation helps the model better learn the features specific to the minority class without overfitting.

For example, applying augmentation techniques to fraud images can produce a more balanced dataset where the model receives more examples of fraud, allowing it to learn more effectively and make more accurate predictions for both classes. This helps reduce the model's bias toward the majority class (e.g., non-fraud cases) and improves its generalization ability, particularly for detecting rare or unseen fraud instances.

Data augmentation is important in balancing datasets by:

i. Increasing Representation of the Minority Class: By augmenting the minority class, the model is exposed to more diverse examples, helping to reduce bias toward the majority class.

ii. Improving Model Robustness: Augmenting the data adds controlled variety, which helps the model generalize better by exposing it to a broader range of examples.

iii. Preventing Overfitting: Instead of simply duplicating existing data, data augmentation creates new, unique examples, which reduces the risk of overfitting, particularly in smaller or imbalanced datasets.

2.5.6 Synthetic Minority Oversampling Technique (SMOTE)

Another effective approach to handling imbalanced datasets is oversampling the minority class using the Synthetic Minority Oversampling Technique (SMOTE). SMOTE is designed to create synthetic samples for the underrepresented class, helping to balance the dataset and improve model performance. By doing so, it reduces bias and ensures that the model captures the key features of the minority class. Unlike random oversampling, which can lead to overfitting by simply duplicating existing samples, SMOTE generates new instances by interpolating between similar examples in the feature space.

The implementation of SMOTE begins by determining the total number of synthetic observations, N, that need to be generated. This is often set to achieve a 1:1 balance between the two classes, though it can be adjusted depending on the specific requirements of the task. The process starts by randomly selecting an instance from the minority class as illustrated in Figure 3 below. Next, the algorithm identifies the K nearest neighbours (typically set to 5 by default) of that instance. N of these neighbours are selected to generate new synthetic samples. The differences in the feature vectors between the selected instance and its neighbours are calculated using a distance metric. Then, a random value between 0 and 1 is multiplied by these differences and added to the original feature vector to create new synthetic samples. This

process helps generate more diverse data points for the minority class, improving the model's ability to make accurate predictions (SWASTIK, 2024).

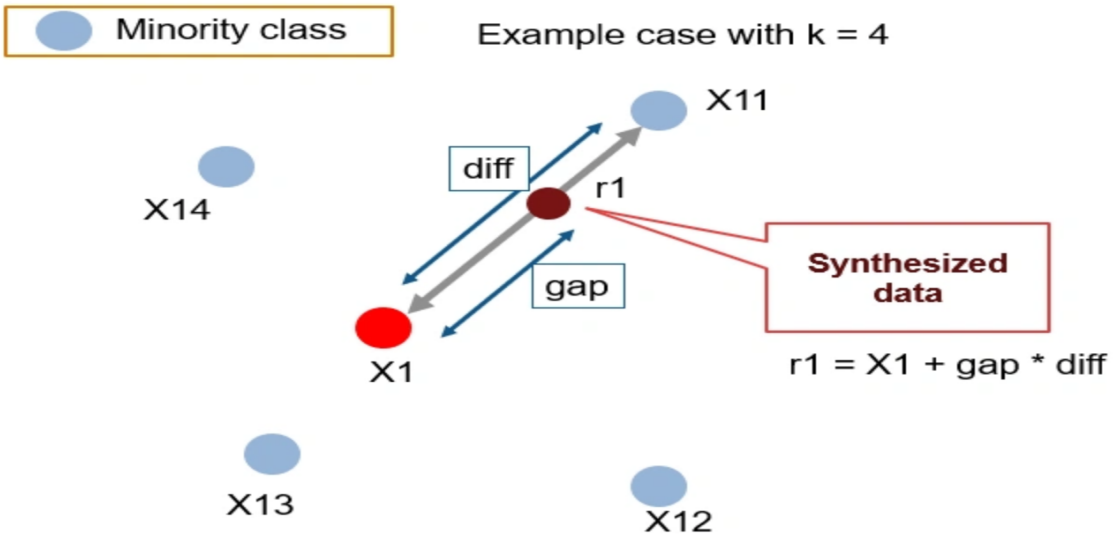

Figure 2: A pictorial representation of how SMOTE works (SWASTIK, 2024)

2.5.7 Deep Convolutional Autoencoders (DCAEs)

Denoising of images in an image-based fraud detection project plays a critical role in the performance of machine learning models. Insurance claims often involve images taken under varying conditions, such as poor lighting, different camera pixel quality, and other environmental factors that can introduce noise to the images. Noise in images can obscure important details, like damage patterns on vehicles, making it difficult to accurately classify the images as either genuine or fraudulent. This, therefore, creates the need for image denoising to retain only the relevant information.

Denoising techniques aim to reduce noise while preserving important image features (Laine, 2021). By smoothing out noisy pixels and enhancing the clarity of images, denoising helps machine learning models focus on key details, such as cracks, dents, scratches, or paint wear, that are often essential in determining whether claims are genuine.

In this study, we employed the approach of using autoencoders, particularly Deep Convolutional Autoencoders (DCAEs). Autoencoders work by learning to compress images into a lower-dimensional representation and then reconstructing them, minimizing noise in the process (Bergmann & Stryker, 2023). The autoencoder is trained on pairs of clean and noisy

images, learning to differentiate between key image features and noise. This process allows the model to denoise new, unseen images effectively before they are input into a classifier model, such as our CNN or Random Forest.

In our study, denoising was applied to images of car accidents before the classification process. The deep convolutional autoencoder processed the images, removing background noise while preserving essential features like damage textures and patterns. This preprocessing step improved the clarity of the images, ensuring that subsequent classification models could focus on the meaningful aspects of the images. The result was cleaner input data, reducing the risk of misclassification due to noise-related artifacts.

Section 3: Research Methodology

3.1 Introduction

This section explains the process of developing and testing machine learning models to detect fraudulent insurance claims. The goal of the study is to create a strong model that can accurately classify claim images as fraudulent or non-fraudulent using a motor vehicle insurance dataset from Kaggle. Key steps in this process included balancing the dataset, enhancing image quality through techniques like denoising, and choosing the right performance metrics, with a focus on accuracy. The study also explored how synthetic data could enhance fraud detection while reducing potential biases and investigated methods to balance the dataset for fair and accurate predictions. Throughout, a systematic approach was followed, involving data exploration, preprocessing, model development, and evaluation, which allowed for continuous improvements and ensured the final model was reliable and effective in identifying fraudulent claims.

3.2 Data Exploration

The dataset used for this study was sourced from Kaggle, specifically the Vehicle Insurance Fraud Classification dataset provided by Gaurav Dutta. This dataset consists of labeled images of vehicle insurance claims, categorized as either fraudulent or non-fraudulent.

The dataset is organized into three main directories:

i.Training Data Directory: Contains two subdirectories, one for fraudulent images and another for non-fraudulent images.

ii.Test Data Directory: Also contains two subdirectories for fraudulent and non-fraudulent images.

iii.Validation Data Directory: A standalone directory for validation purposes.

The training set includes 6,463 images, with 372 labeled as fraudulent and 6,091 as non-fraudulent. The test set consists of 1,616 images, including 93 fraudulent and 1,523 non-fraudulent images. The validation set contains 3,462 images. It is important to note that the images across these directories vary in dimensions, which may require preprocessing to standardize their size before training and evaluation.

3.3 Data Preparation

Based on the data exploration phase, it is clear that the dataset is highly imbalanced, particularly in the training data, where there are only 372 fraud images compared to 6,091 non-fraud images. This imbalance poses a significant challenge for model training, as it is likely to result in a model biased toward predicting non-fraud cases, thus limiting its effectiveness in detecting fraud. Simply reducing the number of non-fraud images to match the fraud count would risk losing valuable information critical for accurate classification. Consequently, addressing this imbalance without sacrificing key data is essential for building a reliable model.

Additionally, the images in the dataset have varying dimensions, requiring resizing to a consistent size for effective model training. It is crucial to ensure that the resizing process does not discard important features in the images, as losing key visual details could negatively impact the model's accuracy.

3.3.1 Resizing

The original dataset contains images with varying dimensions, ranging from (870, 652) to (1376, 1032), (834, 625), (1372, 1029), and (1258, 943), leading to inconsistencies in image resolution. Such variations posed a challenge, as machine learning models require uniform input dimensions to effectively process images.

To address this, all images were resized to a standardized 128x128 pixels. This resizing step not only made the images consistent but also ensured the model could process each image without being affected by size differences. After resizing, both fraud and non-fraud datasets were reshaped to the following dimensions:

i. Fraud Dataset Shape: (6091, 128, 128, 3)

ii. Non-Fraud Dataset Shape: (6091, 128, 128, 3)

Choosing 128x128 as the target resolution was a deliberate decision. This size struck a balance between preserving enough image detail for the model to detect fraud-related patterns and maintaining manageable computational requirements. Additionally, resizing the images to 128x128 helped optimize memory usage and training time, making the model training process more efficient and resource-friendly. This resizing step was essential for ensuring that the data was both consistent and computationally feasible for training the model.

3.3.2 Denoising of the Images

Image denoising was applied to improve the quality of images before they were fed into classification models. Insurance claim images, often taken under various conditions, can contain significant noise due to factors like poor lighting, resolution differences, and environmental conditions. This noise can obscure critical features, such as damage patterns on vehicles, making it challenging to classify the images accurately. By removing this noise, the model is able to focus on the relevant details, leading to improved classification accuracy.

In our experiments, we trained a Deep Convolutional Autoencoder (DCAE) model on both fraud and non-fraud images over 20 epochs, allowing it to learn how to handle diverse noise patterns and image types. The DCAE model compresses the images while retaining important details, such as dents, cracks, or paint damage—key features for assessing insurance claims. By removing noise, the DCAE enhanced the clarity of the images before they were fed into models like CNN and Random Forest, providing a cleaner dataset that supported more accurate classification.

Here is a summary of the DCAE architecture used in this study:

i. **Input:** The input to the model was an image of shape (128, 128, 3), where 128x128 represents the height and width, and 3 corresponds to the RGB color channels.

ii. **Encoder:** A series of convolutional layers with ReLU activation functions were used to extract features from the noisy images. MaxPooling layers were applied to downsample the feature maps, reducing the spatial dimensions while preserving important patterns. The encoded representation captured key characteristics of the image while filtering out noise.

iii. **Decoder:** The decoder mirrored the encoder, using convolutional layers and upsampling to reconstruct the denoised image from the encoded representation. The final layer used a tanh activation function to produce pixel values in the range of [-1, 1], matching the original image's range.

During training, we added Gaussian noise with a factor of 0.2 to the images. This allowed the DCAE to learn how to effectively remove varying levels of noise. The model was trained over 20 epochs with a batch size of 32, and the loss function used to measure the difference between the denoised output and the clean target image was Mean Squared Error (MSE). After training, the model achieved a test MSE of 0.0027117403224110603. This low reconstruction loss indicates that the model effectively denoised the images while preserving the key features necessary for classification.

3.4 Addressing Data Imbalance

3.4.1 Generative Adversarial Network (GAN)

To address the challenge of class imbalance, we utilized Generative Adversarial Networks (GANs) to generate synthetic fraud images, thereby augmenting the dataset with additional fraud cases. This approach helped balance the dataset by increasing the number of fraud samples, which were underrepresented. After training the GAN model, we assessed its performance by comparing the synthetic images to real fraud cases.

The GAN achieved a Mean Squared Error (MSE) of 0.29, which indicates that the generated images were highly similar to the actual fraud images. This low MSE score suggests that the GAN successfully created realistic and diverse fraud images, enhancing the training dataset. By introducing more variety in fraud samples, the model was able to learn critical features associated with fraudulent claims, improving its ability to generalize across different fraud scenarios. This not only improved the model's overall performance but also reduced its bias toward the non-fraud class, making the fraud detection process more accurate and balanced.

3.4.2 Data Augmentation

Another technique used to address the data imbalance was aggressive data augmentation. Transformations such as rotations, shifts, shearing, and zooming were applied to the fraud images to increase the diversity of the dataset. These augmentations generated multiple variations of the fraud images until the dataset was balanced with the non-fraud images. This approach helped reduce overfitting and improved the model's ability to generalize across various fraud scenarios.

The Mean Squared Error (MSE) between the original and augmented images averaged 0.04, indicating that the augmented images retained the essential structures of the originals while introducing helpful variations. This moderate MSE value suggests that the augmentation added sufficient diversity without compromising the core features important for fraud detection. This approach allowed the model to generalize more effectively by exposing it to a wider array of fraud-related patterns.

3.4.3 SMOTE

To further address the data imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to generate synthetic samples for the minority (fraud) class by interpolating between existing fraud cases. Unlike random oversampling, SMOTE helps

prevent overfitting by creating realistic variations of the fraud images rather than simply duplicating them. This method provides the model with a more diverse set of fraud examples, enabling it to learn distinguishing features more effectively and avoid biases toward the majority (non-fraud) class.

The effectiveness of SMOTE was evaluated by calculating the Mean Squared Error (MSE) between the original and synthetic samples, which resulted in an average MSE of 0.108. This low MSE indicates that the synthetic samples preserved the critical characteristics of real fraud instances while introducing useful variations. By balancing the dataset, SMOTE helped improve model generalization, leading to more accurate predictions across both classes and reducing the tendency to favor the majority class.

3.5 Classification Models

In this part of section 3, the study applied Convolution Neural Networks (CNN) and Random Forest to classify car accident images in the validation dataset as either fraudulent or non-fraudulent. After data preparation, the models were trained on the prepared training dataset to identify patterns associated with fraudulent claims. The key focus during this stage was for the models to learn the fraudulent patterns as well as get acquainted to the challenges encountered during training like class weights since the fraud and non-fraud dataset had a huge imbalance.

3.5.1 Convolution Neural Networks (CNN)

For the CNN model, we used pre-processed images from the DCAE to allow the network to learn critical features directly from the data. The CNN architecture comprised three convolutional layers with ReLU activation for feature extraction, followed by max-pooling layers for dimensionality reduction. L2 regularization was applied to the convolutional layers to prevent overfitting. After flattening the feature maps, fully connected layers processed the extracted features, with a dropout layer to mitigate overfitting further. The final layer employed a sigmoid activation function for binary classification (fraudulent or non-fraudulent). To enhance model performance, we utilized a learning rate scheduler to dynamically adjust the learning rate, and early stopping was implemented to prevent overfitting by halting training when validation performance ceased to improve. Additionally, a model checkpoint was used to save the best-performing model based on validation accuracy.

3.5.2 Random Forest

The Random Forest model comprised multiple decision trees, each trained on different subsets of the data. These trees were optimized using grid search to explore hyperparameters such as the number of trees (n_estimators), tree depth (max_depth), and the minimum number of samples required to split nodes (min_samples_split, min_samples_leaf). This hyperparameter tuning helped ensure the selection of the most effective parameter combinations while considering available computational resources. The use of decision trees trained on different subsets of the data contributed to the model's ability to reduce overfitting and improve generalization.

3.6 Evaluation

The performance of the models was evaluated using the test data to assess their ability to accurately categorize insurance claims. Key performance metrics used in this evaluation included the confusion matrix, classification accuracy, and a comprehensive classification report, which contained precision, recall, and F1-score. These metrics provided valuable insights into the models' performance and the impact of different data balancing techniques on their effectiveness.

3.6.1 Confusion Matrix

The confusion matrix breaks down the model's predictions into four categories: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). It helps evaluate how well the model distinguishes between fraudulent and non-fraudulent claims, highlighting both correct predictions and errors, such as misclassifying fraud or non-fraud.

- True Positives (TP): Correctly identified fraud cases.
- True Negatives (TN): Correctly identified non-fraud cases.
- False Positives (FP): Non-fraud cases incorrectly labelled as fraud (false alarms).
- False Negatives (FN): Fraud cases incorrectly labelled as non-fraud (missed fraud).

Table 1: Components of the Confusion Matrix

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

3.6.2 Performance Metrics for Model Evaluation

Classification Accuracy: Measures the overall correctness by calculating the ratio of correctly classified instances TP + TN) to total instances (TP + TN + FP + FN).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: Indicates the model's ability to correctly identify positive instances (fraudulent claims). It's the ratio of true positives (TP) to total predicted positives (TP + FP).

$$Precision = \frac{TP}{TP + FP}$$

Recall: Measures how well the model detects actual positive instances (fraudulent claims). It's the ratio of true positives (TP) to the total number of actual positives (TP + FN).

$$Recall = \frac{TP}{TP + FN}$$

F1-Score: The harmonic mean of precision and recall, balancing the two metrics. It's useful for imbalanced datasets, where it combines both detection and prediction accuracy. The score ranges from 0 to 1, with 1 being the best.

$$F1\ Score = 2\ x\ \frac{Precision\ x\ Recall}{Precsion + Recall}$$

Specificity (True Negative Rate): Focuses on how well the model identifies negative instances (non-fraudulent claims). It's the ratio of true negatives (TN) to the total actual negatives (TN + FP).

$$Specificity = \frac{TN}{TN + FP}$$

AUC-ROC (Area Under the Curve - Receiver Operating Characteristic): The area under the Receiver Operating Characteristic curve, which plots the true positive rate (recall) against the false positive rate (FPR) at various thresholds. A higher AUC score (closer to 1) indicates better model performance. FPR is the proportion of actual negative cases incorrectly classified as positive.

$$FPR = \frac{FP}{FP + TN}$$

Section 4: Results and Discussions

4.1 Introduction

In this section, we shall present the findings of the study and examine how the different machine learning algorithms performed at classifying the images. The analysis includes an evaluation of each model's effectiveness, along with an assessment of the impact of data balancing techniques such as GAN, SMOTE and data augmentation on the overall predictive accuracy. This will provide insights into how these approaches contribute to the models' abilities to accurately identify fraudulent cases, especially in the scenarios where the dataset is imbalanced.

4.2 Model Performance Evaluation

In this subsection, we evaluate and compare the performance of the two binary classification models used in this study: Convolutional Neural Network (CNN) and Random Forest. The comparison will focus on key performance metrics, including accuracy, precision, recall, F1-score, and AUC (Area Under the Curve), to assess each model's effectiveness in distinguishing between fraudulent and non-fraudulent images.

4.2.1 Model Performance on Balanced Datasets

4.2.1.1 CNN Model Evaluation Using GAN-Balanced Data

The classification model CNN was trained on the fraud detection dataset, which was balanced using images generated by GAN. The classification model achieved an accuracy of 99.71% across 30 epochs as shown in figure 8. The model architecture is as follows.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 126, 126, 64)	1,792
batch_normalization (BatchNormalization)	(None, 126, 126, 64)	256
activation (Activation)	(None, 126, 126, 64)	0
max_pooling2d (MaxPooling2D)	(None, 63, 63, 64)	0
dropout (Dropout)	(None, 63, 63, 64)	0
conv2d_1 (Conv2D)	(None, 61, 61, 128)	73,856
batch_normalization_1 (BatchNormalization)	(None, 61, 61, 128)	512
activation_1 (Activation)	(None, 61, 61, 128)	0
max_pooling2d_1 (MaxPooling2D)	(None, 30, 30, 128)	0
dropout_1 (Dropout)	(None, 30, 30, 128)	0
conv2d_2 (Conv2D)	(None, 28, 28, 256)	295,168
batch_normalization_2 (BatchNormalization)	(None, 28, 28, 256)	1,024
activation_2 (Activation)	(None, 28, 28, 256)	0
max_pooling2d_2 (MaxPooling2D)	(None, 14, 14, 256)	0
dropout_2 (Dropout)	(None, 14, 14, 256)	0
conv2d_3 (Conv2D)	(None, 12, 12, 512)	1,180,160
batch_normalization_3 (BatchNormalization)	(None, 12, 12, 512)	2,048
activation_3 (Activation)	(None, 12, 12, 512)	0
max_pooling2d_3 (MaxPooling2D)	(None, 6, 6, 512)	0
dropout_3 (Dropout)	(None, 6, 6, 512)	0
flatten (Flatten)	(None, 18432)	0
dense (Dense)	(None, 512)	9,437,696
batch_normalization_4 (BatchNormalization)	(None, 512)	2,048
activation_4 (Activation)	(None, 512)	0
dropout_4 (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 1)	513

Total params: 32,979,333 (125.81 MB)
Trainable params: 10,992,129 (41.93 MB)
Non-trainable params: 2,944 (11.50 KB)
Optimizer params: 21,984,260 (83.86 MB)

Figure 3: Summary of the CNN model with GAN usage for data balancing

Table 2, below summarizes the CNN model evaluation using GAN-balanced data, covering the architecture from the initial convolutional block to the output layer.

Table 2: CNN Architecture Using GAN-Balanced Data

i. Initial Convolutional Block:

Layer	**Purpose**	**Filters**	**Output Shape**	**Parameters**	**Additional Info**
Initial Convolutional Layer	Extract features from input images	64 (3x3)	(126, 126, 64)	1,792	None
Batch Normalization	Normalize data for stability	-	(126, 126, 64)	256	-
MaxPooling2D	Reduce spatial dimensions by half	-	(63, 63, 64)	0	None
Dropout	Prevent overfitting	-	(63, 63, 64)	0	Randomly disables connections

ii. Second Convolutional Block:

Layer	**Purpose**	**Filters**	**Output Shape**	**Parameters**	**Additional Info**
Second Convolutional Layer	Extract deeper features	128 (3x3)	(61, 61, 128)	73,856	None
Batch Normalization	Normalize data for stability	-	(61, 61, 128)	512	-
MaxPooling2D	Reduce spatial dimensions by half	-	(30, 30, 128)	0	None
Dropout	Prevent overfitting	-	(30, 30, 128)	0	Randomly disables connections

iii. Third Convolutional Block:

Layer	Purpose	Filters	Output Shape	Parameters	Additional Info
Third Convolutional Layer	Extract deeper features	256 (3x3)	(28, 28, 256)	295,168	None
Batch Normalization	Normalize data for stability	-	(28, 28, 256)	1,024	-
MaxPooling2D	Reduce spatial dimensions by half	-	(14, 14, 256)	0	None
Dropout	Prevent overfitting	-	(14, 14, 256)	0	Randomly disables connections

iv. Fourth Convolutional Block:

Layer	Purpose	Filters	Output Shape	Parameters	Additional Info
Fourth Convolutional Layer	Extract even deeper features	512 (3x3)	(12, 12, 512)	1,180,160	None
Batch Normalization	Normalize data for stability	-	(12, 12, 512)	2,048	-
MaxPooling2D	Reduce spatial dimensions by half	-	(6, 6, 512)	0	None
Dropout	Prevent overfitting	-	(6, 6, 512)	0	Randomly disables connections

v. Flattening and Fully Connected Layers:

Layer	Purpose	Filters	Output Shape	Parameters	Additional Info
Flatten	Flatten the multi-dimensional output into a single vector	-	(18,432)	0	-
Dense (Fully Connected)	Map features to output	512	(512)	9,437,696	Batch Normalization, Dropout

vi. Output Layer:

Layer	Purpose	Filters	Output Shape	Parameters	Additional Info
Output Layer	Output prediction for binary classification	1	(1)	513	-

This CNN model processes input images by extracting hierarchical features through multiple convolutional layers. Pooling layers reduce data complexity by down-sampling, while dense layers make the final classification decision. Regularization techniques, such as dropout and batch normalization, enhance robustness and minimize overfitting.

Testing the model on the test data yielded the following metrics,

CNN Binary Classification Model Metrics:

• **Threshold:** 0.5

• **Accuracy:** 96.41%

• **Precision:** 0.76

• **Recall:** 0.55

• **F1 Score:** 0.64

Table 3: Confusion Matrix for CNN Model Evaluation Using GAN-Balanced Data

	Predicted: Non-Fraud	**Predicted: Fraud**
True: Non-Fraud	1507	16
True: Fraud	42	51

Model Performance Analysis

The model achieved an overall accuracy of 96.41%, meaning it correctly classified the majority of test samples. However, the performance on the fraud class reveals key limitations. Specifically:

- **Precision for the fraud class:** 0.76, indicating that only 76% of the predicted fraudulent images were correctly classified as fraud.
- **Recall for the fraud class:** 0.55, meaning only 55% of the actual fraudulent images were identified by the model.
- **F1 Score:** 0.64, reflecting a moderate balance between precision and recall, but also highlighting room for improvement in detecting fraud cases.

Insights and Potential Causes

While the model performs exceptionally well at identifying non-fraudulent images, it struggles to consistently detect fraudulent cases. Despite being trained on balanced data generated by a GAN, the precision and recall for fraud remain relatively low. This discrepancy points to several potential causes:

- **Quality of GAN-generated images:** If the synthetic fraud images lack sufficient realism or diversity, the model may struggle to generalize effectively to real fraudulent cases.
- **Residual class imbalance bias:** Even with balanced data, the model might still favor the majority class (non-fraud), leading to reduced sensitivity in detecting the minority fraud class.
- **Complexity of fraud patterns:** Fraudulent cases often exhibit subtle or complex features that are difficult to capture, requiring more sophisticated model tuning or additional data augmentation.

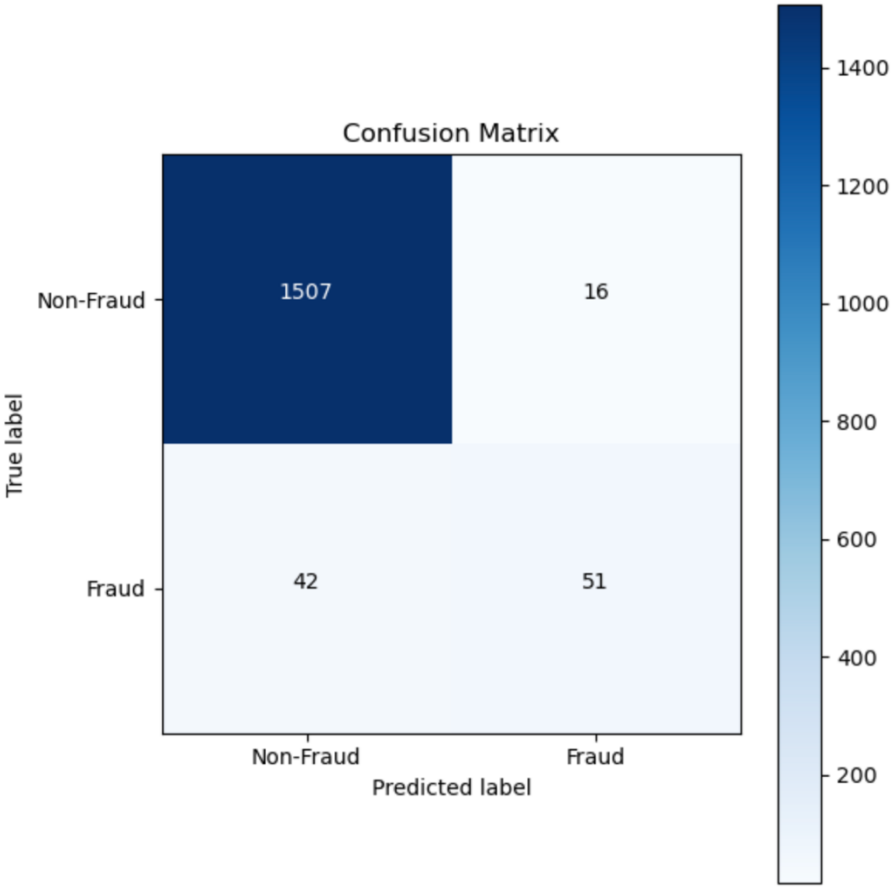

Figure 4: Performance of CNN Model with GAN-Based Data Balancing: Confusion Matrix

4.2.1.2 Random Forest Model Evaluation Using GAN-Balanced Data

The Random Forest model was tested on the dataset balanced using a GAN, and the results offered some valuable insights. To improve efficiency, Principal Component Analysis (PCA) was applied to reduce the feature dimensions to 200, ensuring we retained most of the data's variance while simplifying its structure. This step was critical in speeding up the training process and reducing the risk of overfitting. The Random Forest model was optimized using RandomizedSearchCV with 5-fold cross-validation to determine the best hyperparameters. The search space included parameters such as the number of trees (n_estimators), tree depth (max_depth), minimum samples for splits (min_samples_split), and leaves (min_samples_leaf), as well as the number of features considered at each split (max_features). Although the dataset was balanced using GAN, the model was further configured with class_weight = {0: 1.0, 1: 15} to fine-tune the balance between the classes. This configuration assigned more weight to the fraudulent cases (class 1) to improve their detection, compensating for any residual imbalance.

To optimize the Random Forest model's performance, hyperparameter tuning using RandomizedSearchCV with 5-fold cross-validation was applied. This process allowed us to test various configurations while keeping computational resource constraints in mind. The following ranges of hyperparameters were explored during tuning:

- **Number of Trees (n_estimators):** [25, 50, 75, 100, 150]
- **Maximum Tree Depth (max_depth):** [10, 20, 30, None]
- **Minimum Samples to Split (min_samples_split):** [2, 5, 10]
- **Minimum Samples per Leaf (min_samples_leaf):** [1, 2, 4]
- **Features per Split (max_features):** ['sqrt', 'log2', None]
- **Bootstrap:** [True, False]
- **Criterion:** ['gini', 'entropy']

After evaluating all combinations, the model achieved its best performance using a configuration that struck a balance between predictive accuracy and computational efficiency. The cross-validation results showed that even relatively shallow trees and fewer estimators could yield competitive results without significantly increasing training time.

The best cross-validation F1 score achieved during training was 0.04%, which reflects a significant imbalance between precision and recall, particularly for the fraudulent cases. This score suggests that the model is struggling to correctly identify fraudulent instances, leading to a high number of false negatives. While the model shows promising accuracy in identifying non-fraudulent cases, there is a need for further improvement in capturing the fraudulent cases more effectively.

When evaluated on the on the test dataset, the Random Forest model achieved an overall accuracy of 94.06%. The confusion matrix for this evaluation was as follows:

Confusion Matrix (Validation):

[[1518 5]

[91 2]]

Table 4: Confusion Matrix for Random Forest Model Evaluation Using GAN-Balanced Data

	Predicted: Non-Fraud	**Predicted: Fraud**
True: Non-Fraud	1518	5
True: Fraud	91	2

From the confusion matrix, it is evident that the model performed better in identifying non-fraudulent cases but struggled with fraudulent ones. Below are the key metrics derived from this evaluation:

- **Precision:** 0.29
- **Recall:** 0.02
- **F1-Score:** 0.04

This indicates that while the model had a high accuracy in identifying non-fraudulent cases, it was ineffective in detecting fraudulent ones, as reflected in the zero values for precision, recall, and F1-score for the fraud class.

In conclusion, the combination of GAN-balanced data and PCA contributed to building a fairly effective model, with accuracy serving as a promising starting point. However, the model's relatively high number of false negatives (91 cases) highlights a weakness in its ability to correctly identify fraudulent cases. This limitation, reflected in the lower recall, suggests room for improvement. The challenge might stem from the Random Forest algorithm's difficulty in detecting subtle patterns within complex, high-dimensional data.

.

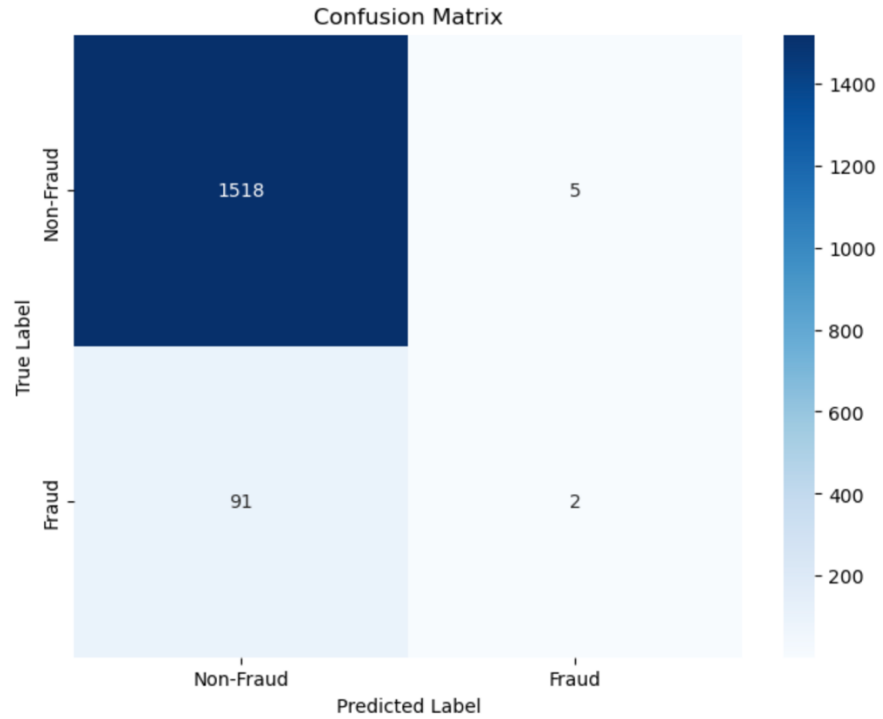

Figure 5: Performance of Random Forest Model with GAN-Based Data Balancing: Confusion Matrix

4.2.1.3 Performance Evaluation of CNN on SMOTE Balanced Data

The CNN model with SMOTE model for balancing data achived an accuracy of 99.96% across 25 epochs out of 30 set epochs. Early stopping, was implemented to prevent overfitting, ensuring that the model's performance remained optimal. The architecture of the model is as follows:

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 126, 126, 32)	896
max_pooling2d (MaxPooling2D)	(None, 63, 63, 32)	0
conv2d_1 (Conv2D)	(None, 61, 61, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 30, 30, 64)	0
conv2d_2 (Conv2D)	(None, 28, 28, 128)	73,856
max_pooling2d_2 (MaxPooling2D)	(None, 14, 14, 128)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 128)	3,211,392
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 1)	129

Total params: 9,914,309 (37.82 MB)
Trainable params: 3,304,769 (12.61 MB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 6,609,540 (25.21 MB)

Figure 6: CNN model summary with SMOTE for data balancing.

Table 5 below summarizes the performance evaluation of the CNN model on SMOTE-balanced data, covering the evaluation results from the confusion matrix.

Table 5: Performance Evaluation of CNN on SMOTE-Balanced Data

Layer	Description	Output Shape	Filters/Neurons	Trainable Parameters
Initial Convolutional Layer	Convolutional layer with 32 filters of size 3x3 to extract features from the input images.	(126, 126, 32)	32 filters (3x3)	896
Max Pooling Layer	MaxPooling2D layer that reduces spatial	(63, 63, 32)	-	0

	dimensions by half, focusing on important features.			
Second Convolutional Block	Convolutional layer with 64 filters of size 3x3 followed by MaxPooling2D and dropout for overfitting.	(30, 30, 64)	64 filters (3x3)	-
Third Convolutional Block	Convolutional layer with 128 filters of size 3x3 followed by MaxPooling2D and dropout for generalization.	(14, 14, 128)	128 filters (3x3)	-
Flattening and Dense Layer	Flatten the output to a vector, followed by a fully connected layer with 128 neurons for classification.	(25,088)	128 neurons	3,211,392
Output Layer	Dense layer with a single neuron to output binary classification (fraud or non-fraud).	(1)	1 neuron	129

In summary, this architecture effectively extracts features through convolutional layers and uses dense layers for classification, with dropout applied to help prevent overfitting and improve generalization.

On the test data, the model performance was as below:

CNN Binary Classification Model Metrics:

Threshold: 0.5

Accuracy: 96.47%

Precision: 0.74

Recall: 0.59

F1 Score: 0.66

Table 6: Confusion Matrix for Performance Evaluation of CNN on SMOTE-Balanced Data

Actual/Predicted	**Non-Fraud**	**Fraud**
Non-Fraud	1504	19
Fraud	38	55

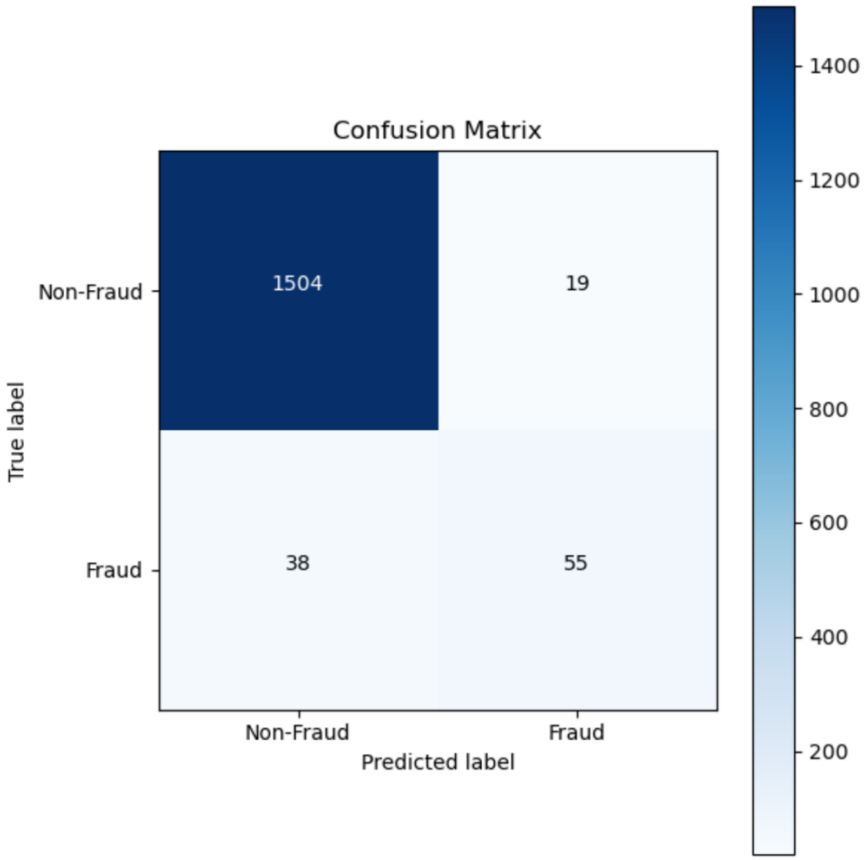

Figure 7: Performance of CNN Model with SMOTE-Based Data Balancing: Confusion Matrix.

Analysis of Model Performance

The classification model achieved an overall accuracy of 96.47%, showcasing its ability to correctly classify the majority of test samples. When analyzing performance specific to the fraud detection task, several key observations emerge:

• **Precision (0.74):** The model correctly identified fraud 74% of the time when it predicted a sample as fraudulent. This indicates a reasonable level of confidence in its positive predictions.

• **Recall (0.59):** However, the model was able to identify only 59% of the actual fraudulent cases, indicating that a significant portion of fraudulent samples went undetected.

• **F1 Score (0.66):** The F1 score, which balances precision and recall, reflects moderate effectiveness in identifying fraudulent cases.

Despite the high overall accuracy, the model's performance in detecting fraud is less satisfactory. The lower recall and F1 score for the fraud class highlight the model's struggle to correctly identify fraudulent samples. This could stem from limitations in the quality or representativeness of the SMOTE-generated synthetic samples used during training. Alternatively, it may indicate a persistent bias towards classifying samples as non-fraud, even with a balanced dataset. Improving the training data or adopting more advanced techniques could help enhance the model's fraud detection capabilities.

4.2.1.4 Performance Evaluation of Random Forest on SMOTE Balanced Data

The Random Forest model, trained on SMOTE-balanced data, was evaluated to assess its ability to detect fraudulent images. SMOTE generated synthetic samples for the minority class (fraudulent images), helping the model avoid bias toward the majority class (non-fraudulent images). The model architecture is outlined below in Table 7:

Table 7: Architecture of Random Forest Model on SMOTE-Balanced Data

Step	**Details**
Dimensionality Reduction	PCA was applied to reduce the number of features to 50 components, helping to manage the high-dimensional data.
Classifier	RandomForestClassifier with the following hyperparameters:
n_estimators	25, 50, 75 - Number of trees in the forest.
max_depth	10, 20, 30, None - Maximum depth of the trees.
min_samples_split	2, 5, 10 - Minimum samples required to split a node.
min_samples_leaf	1, 2, 4 - Minimum samples required at a leaf node.
max_features	'sqrt', 'log2' - Number of features considered for splitting each node.

GridSearchCV	Used for 5-fold cross-validation with F1-score as the evaluation metric.
Best Hyperparameters	n_estimators: 50, max_depth: 20, min_samples_split: 5, min_samples_leaf: 2, max_features: 'sqrt'
Best Cross-Validation F1	0.91

The training process involved key steps to optimize the model's performance. First, Principal Component Analysis (PCA) reduced the feature space to 50 components, simplifying the model and reducing the risk of overfitting. Hyperparameter tuning was then performed using GridSearchCV, which explored different configurations to identify the best settings for the model. The optimal hyperparameters were selected based on the highest F1-score achieved through 5-fold cross-validation. As a result, the model achieved a strong test accuracy of 95.17%, reflecting its overall performance. However, the confusion matrix (Figure 9) provides deeper insights:

Table 8: Confusion Matrix for Performance Evaluation of Random Forest on SMOTE-Balanced Data

Actual/Predicted	**Non-Fraud**	**Fraud**
Non-Fraud	1488	35
Fraud	43	50

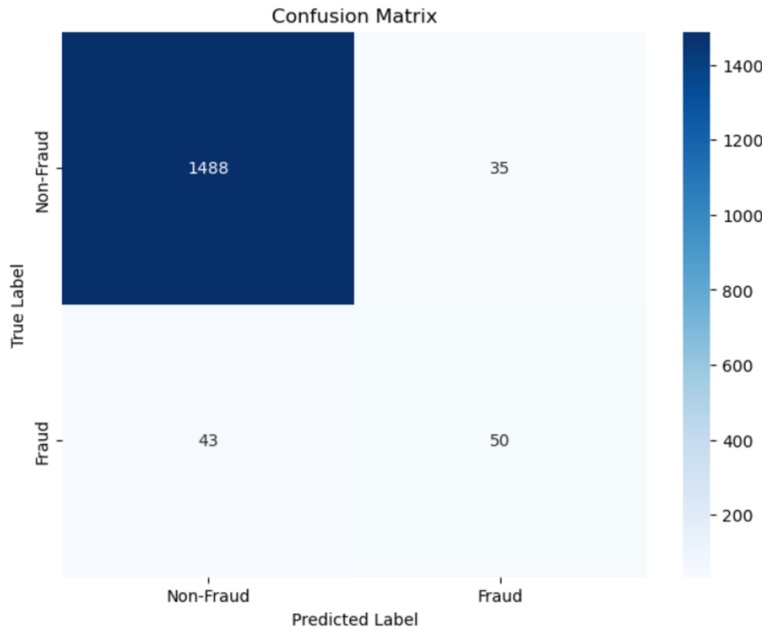

Figure 8: Confusion Matrix for Random Forest Model Performance on Test Set.

The confusion matrix reveals that the model successfully identified 1488 non-fraudulent images and 50 fraudulent ones. However, it misclassified 35 non-fraudulent images as fraudulent and 43 fraudulent images as non-fraudulent. While the model performed well in detecting non-fraudulent images, the false negatives highlight an area where improvements are needed for better fraud detection.

Despite achieving an overall accuracy of 95.17%, the model's recall and F1-score for fraudulent images were notably lower, at 53.8% and 56.2%, respectively. The precision for fraud detection was higher, at 58.8%, but the model still misclassified 43 fraudulent images as non-fraudulent. This suggests that while the model can distinguish non-fraudulent images effectively, it struggles to detect fraud with the same level of confidence. On a positive note, the model excelled at classifying non-fraudulent images, as evidenced by the high number of true negatives.

Although the Random Forest model demonstrated solid performance overall, its ability to accurately identify fraudulent images still requires improvement. The false negatives point to a potential bias towards classifying images as non-fraudulent. To enhance fraud detection, applying techniques like class weighting or adjusting the classification threshold could help improve the model's sensitivity to fraudulent images, without significantly impacting its accuracy in detecting non-fraudulent cases.

4.2.1.5 Performance Evaluation of CNN on Augmented Data

In this section, we evaluate the performance of the CNN model trained on augmented data. The results indicate that while the model performs well in terms of accuracy, there is still room for improvement, especially in detecting fraud cases. Here, we detail the architecture, performance metrics, and key observations derived from the evaluation.

The architecture of the CNN model is designed for binary classification, with layers configured to learn relevant features from image data. Below, Table 10 summarizes the model architecture:

Table 9: Performance Evaluation of CNN on Augmented Data

Layer (type)	**Output Shape**	**Param #**
Conv2D	(None, 126, 126, 32)	896
MaxPooling2D	(None, 63, 63, 32)	0
Conv2D	(None, 61, 61, 64)	18,496
MaxPooling2D	(None, 30, 30, 64)	0

Conv2D	(None, 28, 28, 128)	73,856
MaxPooling2D	(None, 14, 14, 128)	0
Flatten	(None, 25088)	0
Dense	(None, 128)	3,211,392
Dropout	(None, 128)	0
Dense	(None, 1)	129

This architecture consists of three convolutional layers with corresponding max pooling layers for downsampling, followed by a fully connected dense layer with dropout regularization to reduce overfitting. The output layer consists of a single unit with a sigmoid activation function for binary classification.

The model demonstrated strong performance during training, achieving an accuracy of 95.67% and a loss of 0.3185. The validation accuracy of 95.49% closely mirrors the training accuracy, suggesting that the model generalizes well to unseen data. This indicates that the model is not overfitting and is capable of making accurate predictions on new, unseen examples.

Upon testing the model on the test data, the following key metrics were achieved, as illustrated in Figure 9, which presents the confusion matrix:

• **Threshold:** The model uses a threshold of 0.5 for classification, meaning that if the predicted probability for the fraud class exceeds 0.5, the image is classified as fraud. If the probability is lower than 0.5, the image is classified as non-fraud.

• **Accuracy:** The model achieved an accuracy of 91.71%. While this is relatively high, it is important to note that accuracy can be misleading in imbalanced datasets, as it does not account for the distribution of the classes.

• **Precision:** The precision of the model is 0.35, indicating that only 35% of the cases predicted as fraud are indeed actual fraud cases. This highlights a significant number of false positives, where non-fraud images are mistakenly classified as fraud.

• **Recall:** With a recall of 0.49, the model correctly identifies nearly half of the fraud cases but still misses more than half. This is a concern for fraud detection tasks, where missing fraud cases can have serious consequences.

• **F1 Score:** The F1 score, which balances precision and recall, is 0.41. This suggests that while there is some balance between precision and recall, both metrics need significant improvement for better fraud detection performance.

Table 10: Confusion Matrix for Performance Evaluation of CNN on Augmented Data

Actual/Predicted	**Non-Fraud**	**Fraud**
Non-Fraud	1436	87
Fraud	47	46

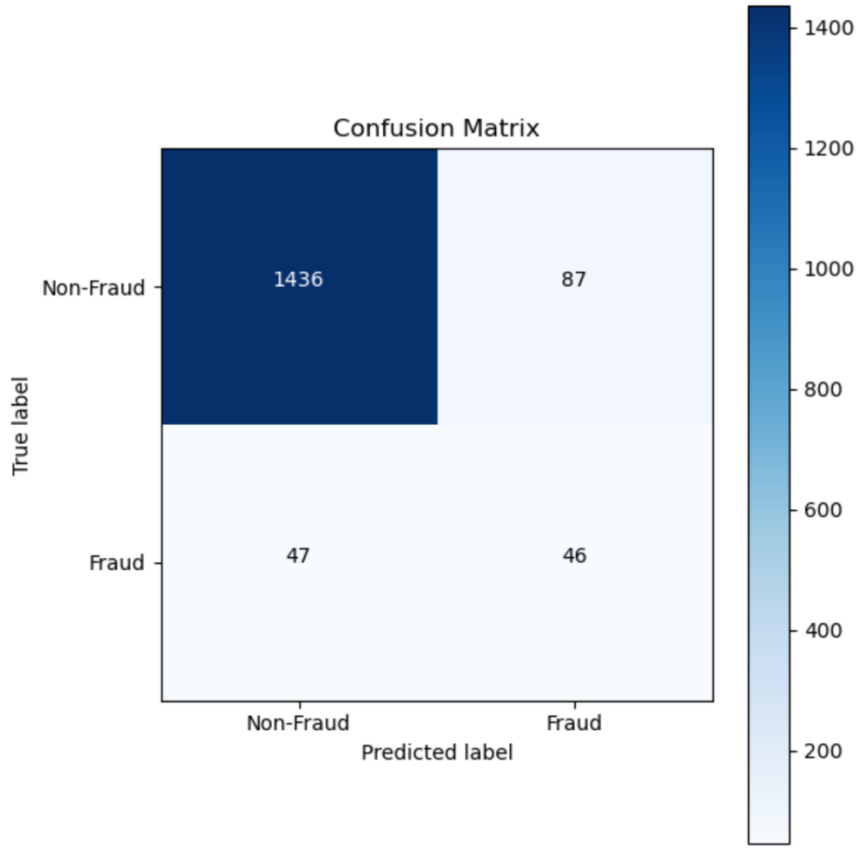

Figure 9: Confusion Matrix illustrating the performance of the model on the test data.

• **True Negatives (1436):** Non-fraud images correctly classified as non-fraud.

• **False Positives (87):** Non-fraud images incorrectly classified as fraud, contributing to the low precision.

• **False Negatives (47):** Fraud images incorrectly classified as non-fraud, affecting the recall.

• **True Positives (46):** Fraud images correctly identified as fraud.

Observations:

• While the model's overall accuracy is high, the precision and recall metrics reveal significant issues with fraud detection. The low precision suggests that many non-fraud cases are being incorrectly labeled as fraud. The moderate recall indicates that the model is missing a considerable number of fraud cases.

• Class imbalance remains a challenge, as there are far fewer fraud images than non-fraud images in the dataset. To improve the model's performance in detecting fraud, strategies such as further adjustments to class weights, threshold tuning, or exploring more advanced data augmentation techniques could be beneficial.
• **Threshold Adjustment:** Lowering the classification threshold from 0.5 to a value such as 0.3 could potentially increase recall, although this would likely decrease precision. This trade-off should be explored further, considering the specific needs of the application.
• **Class Weights:** Fine-tuning the class weights could also improve the model's ability to correctly classify fraud cases, as the current imbalance between fraud and non-fraud images is impacting performance.

In conclusion, while this model shows promise with good accuracy, there is still room for improvement in detecting fraud, particularly in terms of precision and recall. Further refinements in the training process and adjustments to the model's settings could help address these issues.

4.2.1.6 Performance Evaluation of Random Forest on Augmented Data

This section evaluates the performance of the Random Forest model trained on augmented data, with PCA applied for dimensionality reduction. The augmentation technique was used to balance the dataset by generating synthetic data, which was then utilized for model training.
To enhance efficiency, Principal Component Analysis (PCA) was applied to reduce the feature dimensions to 50 components, retaining most of the data's variance while simplifying the dataset. This was essential in speeding up the training process and reducing the risk of overfitting. Following this, the Random Forest model was fine-tuned using GridSearchCV with 5-fold cross-validation to identify the optimal hyperparameters. The hyperparameters tested included:

- **Number of Trees (n_estimators):** [25, 50, 75]
- **Maximum Tree Depth (max_depth):** [10, 20, 30, None]
- **Minimum Samples to Split (min_samples_split):** [2, 5, 10]
- **Minimum Samples per Leaf (min_samples_leaf):** [1, 2, 4]
- **Features per Split (max_features):** ['sqrt', 'log2']

Additionally, the model was configured with class_weight='balanced' to handle any remaining class imbalance. The best combination of hyperparameters from this search resulted in an F1 score of 84.36% during cross-validation, reflecting strong performance when trained on the augmented dataset.

Upon evaluating the model on the validation set, we obtained the following performance metrics:

- **Accuracy:** 93.19%
- **Precision:** 0.33
- **Recall:** 0.17
- **F1-Score:** 0.23

The confusion matrix for the validation set is as follows:

Table 11: Confusion Matrix for Performance Evaluation of Random Forest on Augmented Data

Actual/Predicted	**Non-Fraud**	**Fraud**
Non-Fraud	1490	33
Fraud	77	16

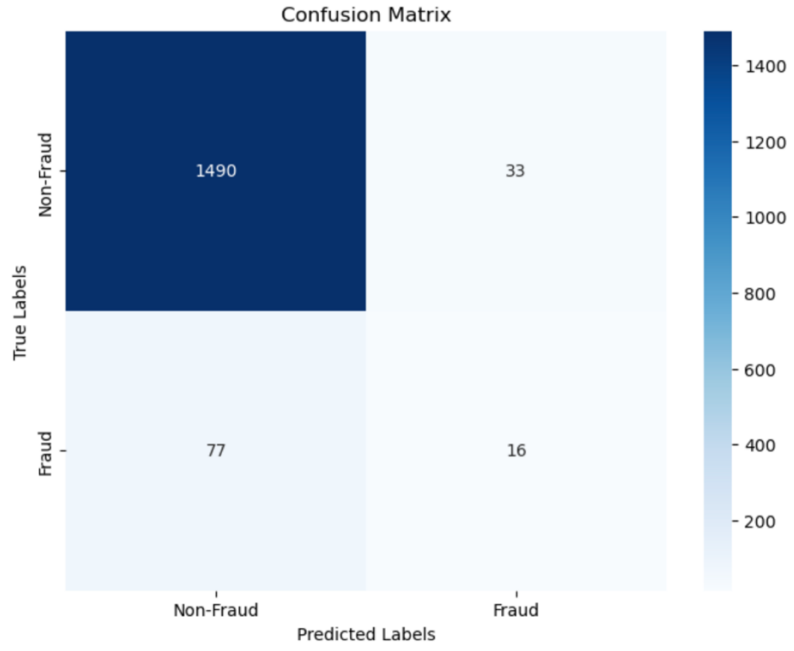

Figure 10: Confusion Matrix illustrating the performance of the model on the test data.

While the model achieved a high accuracy (93.19%), it struggled with identifying fraudulent cases. The low precision, recall, and F1 score indicate that, although the model was successful at identifying non-fraudulent cases (TN), it missed many fraudulent cases (FN). This suggests that while the model is good at distinguishing the majority class, it requires further tuning or additional strategies to improve its performance on the minority class (fraudulent cases).

4.3 Comparative Analysis of Model Performances

This section, as summarized in Table 12, compares the performance analysis of the Convolutional Neural Networks (CNN) and Random Forest models on test data, employing three data balancing techniques: GAN-balanced, SMOTE-balanced, and augmented datasets. By evaluating the impact of each balancing strategy and model architecture on fraud detection, we aim to identify their respective strengths and limitations. The findings will offer valuable insights into the most effective approaches for achieving accurate and efficient fraud classification, particularly in the context of insurance claims or comparable applications.

Table 12: Comparative Analysis of Model Performances

Metric	**CNN (GAN)**	**RF (GAN)**	**CNN (SMOTE)**	**RF (SMOTE)**	**CNN (Augmented)**	**RF (Augmented)**
Accuracy	94.99%	94.06%	96.47%	95.20%	91.71%	93.19%
Threshold	0.5	0.5	0.5	0.5	0.5	0.5
Precision	0.76	0.29	0.74	0.59	0.35	0.33
Recall	0.55	0.02	0.59	0.54	0.49	0.17
F1-Score	0.64	0.04	0.66	0.56	0.41	0.23
Predicted Unlabelled Dataset Count						
Fraud	1728	492	141	64	1319	933
Non-Fraud	1734	2970	3321	3398	2143	2529

Accuracy: CNN consistently outperformed RF, with its highest accuracy of 96.47% achieved using SMOTE-balanced data. This result highlights SMOTE's effectiveness in creating a balanced representation of fraud and non-fraud cases, which significantly boosted CNN's performance. Similarly, RF performed best with SMOTE, achieving 95.20% accuracy,

while RF's performance dropped slightly to 94.06%. When using the augmentation technique, both models experienced a slight decline in accuracy, with CNN at 91.71% and RF at 93.19%. Interestingly, the gap between the models was smallest with augmented data, suggesting that this technique leveled the playing field, making their performance more comparable.

Precision: CNN demonstrated a clear advantage in precision, particularly with GAN-balanced data, where it achieved a precision score of 0.76. This result underscores its ability to minimize false positives, a critical factor in fraud detection. RF, in contrast, struggled with precision across all techniques, with its worst performance recorded on GAN-balanced data (0.29). This indicates that RF frequently misclassified non-fraud cases as fraud, particularly under this balancing strategy.

Recall: When it came to recall, which measures the ability to correctly identify fraudulent cases, SMOTE again emerged as the most effective technique for both models. CNN achieved a recall of 0.59, closely followed by RF at 0.54. However, with GAN-balanced and augmented data, CNN maintained relatively stable recall scores of 0.43 and 0.49, while RF showed a significant decline, particularly with augmented data, where its recall dropped to just 0.17. This highlights RF's difficulty in generalizing well to fraudulent cases under these conditions.
F1-Score: The F1-score, which balances precision and recall, further reinforced SMOTE's superiority. CNN achieved its highest F1-score of 0.66 with SMOTE, reflecting its strong overall performance. RF, though less effective overall, also performed best with SMOTE, achieving an F1-score of 0.56. However, GAN-balanced data resulted in RF's lowest F1-score of 0.04, revealing its struggles with this particular balancing strategy.

Fraud and Non-Fraud Predictions: Examining the number of fraud and non-fraud cases identified by each model adds further context. CNN trained on GAN data predicted the highest number of fraud cases (1728), showing its sensitivity to fraudulent patterns. RF, however, consistently predicted fewer fraud cases across all techniques, with its lowest count of only 64 fraud cases occurring with SMOTE-balanced data. This disparity suggests that CNN is more adept at detecting fraud, whereas RF may under-detect fraudulent cases, potentially missing key instances.

In conclusion, CNN emerged as the stronger performer across all data balancing techniques, with SMOTE-balanced data consistently enhancing its metrics. RF also benefited from

SMOTE but struggled with precision and recall, particularly with GAN-balanced and augmented data. While augmentation proved to be a viable alternative, GAN-balanced data posed significant challenges, especially for RF. These findings highlight the importance of selecting the right data balancing strategy for the specific model architecture to optimize fraud detection outcomes.

Section 5: Conclusion and Recommendations

In this study, we successfully developed predictive models for identifying potential automobile insurance fraud claims using machine learning techniques. The dataset, sourced from Kaggle, presented significant challenges due to its highly imbalanced nature, with a training distribution of 6,091 non-fraud cases to only 372 fraud cases. This imbalance posed difficulties in balancing the dataset and training effective classification models.

A balanced dataset would enable a more accurate evaluation of the model's performance, ensuring its applicability in real-world scenarios. This balance ensures that the model can learn effectively from both fraud and non-fraud cases, reducing the risk of bias toward the majority class. Moreover, balanced data provides a solid foundation for assessing key performance metrics such as precision, recall, and F1-score, which are critical for understanding the model's capability to detect fraudulent claims accurately.

To further enhance the model's robustness, it is essential to test its generalization capabilities by evaluating it on datasets from diverse sources. This approach allows the model to encounter varied patterns and anomalies present in different insurance environments, thereby providing a realistic measure of its effectiveness in real-world applications. Cross-validation across these datasets can reveal areas where the model excels and identify weaknesses that need to be addressed. Additionally, fine-tuning the model through hyperparameter optimization is recommended. Exploring various configurations, such as adjusting the learning rate, regularization parameters, thresholds, and network architecture (for neural networks), along with tuning parameters specific to ensemble methods like the number of estimators, maximum depth, and feature selection, can significantly enhance the model's performance. By evaluating and experimenting with these settings, the model can be tailored to effectively address the complexities of insurance fraud detection, resulting in a more robust and adaptive solution.

To address computational costs and optimize model efficiency, reducing the number of features used in the prediction process is essential. In our study, we applied Principal Component Analysis (PCA) to the Random Forest model for dimensionality reduction, retaining only the most relevant components. This technique helped eliminate redundant or less informative features, thereby reducing computation time and memory requirements. By doing so, we made the model more efficient and suitable for practical implementation in insurance fraud detection systems. Additionally, we explored advanced data augmentation techniques like MixUp, CutMix, and Attn_CutMix to further improve model performance, particularly in handling the

highly imbalanced dataset. These steps contributed to enhancing the overall effectiveness of our fraud detection approach.

In conclusion, this study makes a valuable contribution to the field of insurance fraud detection by demonstrating the effectiveness of machine learning techniques in identifying potential automobile insurance fraud claims. However, it is crucial to recognize that the reliability and generalizability of the predictive model can be further improved by utilizing balanced datasets, cross-validating on diverse environments, and experimenting with different hyperparameter settings. Additionally, optimizing feature selection and exploring advanced techniques such as PCA for dimensionality reduction and data augmentation methods like MixUp, CutMix, and Attn_CutMix can enhance the model's performance. By continually refining the model with updated data, insurance companies can better combat fraudulent activities, reduce financial losses, and offer more competitive premiums to honest policyholders.

References

Aslam, F., Hunjra, A.I., Ftiti, Z., Louhichi, W. & Shams, T., 2022. Insurance fraud detection: Evidence from artificial intelligence and machine learning. *ScienceDirect*, [online]. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0275531922001325 [Accessed 8 January 2025].

Ben-Hutta, A., 2023. Allianz develops machine-learning tool to support growing fraud claims. *Coverager*, [online]. Available at: https://coverager.com/allianz-develops-machine-learning-tool-to-support-growing-fraud-claims/ [Accessed 8 January 2025].

Christopher, A. & Dubey, A., 2020. The exigency for an insurance frauds control act in India: Challenges to be addressed. *SSRN*, [online]. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3870318 [Accessed 8 January 2025].

Gangcuangco, T., 2024. ABI outlines extent of insurance fraud. *Insurance Business*, [online]. Available at: https://www.insurancebusinessmag.com/uk/news/claims/abi-outlines-extent-of-insurance-fraud-507601.aspx [Accessed 8 January 2025].

Global Market Insights, 2024. Insurance fraud detection market. [online]. Available at: https://www.gminsights.com/industry-analysis/insurance-fraud-detection-market [Accessed 8 January 2025].

Huang, H., Zhou, X. & He, R., 2023. Attention-guided CutMix data augmentation network for fine-grained visual classification. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1–10. Available at: https://doi.org/10.1109/ICCV.2023.00001 [Accessed 8 January 2025].

Laine, R.F., Jacquemet, G. & Krull, A., 2021. Imaging in focus: An introduction to denoising bioimages in the era of deep learning. *Biomedical Signal Processing and Control*, 67, p.102588. [online] Available at: https://www.sciencedirect.com/science/article/pii/S1357272521001588 [Accessed 8 January 2025].

Martino, N., 2023. The power of automation: How AI is transforming fraud detection in financial institutions. *LinkedIn*, [online]. Available at: https://www.linkedin.com/pulse/power-automation-how-ai-transforming-fraud-detection-nicolas-martino [Accessed 8 January 2025].

Muiruri, D., 2023. AI a game changer for insurance. *Business Daily Africa*, [online]. Available at: https://www.businessdailyafrica.com/bd/opinion-analysis/columnists/ai-a-game-changer-for-insurance--4352032 [Accessed 8 January 2025].

Schrijver, G., Sarmah, D.K. & El-hajj, M., 2024. Automobile insurance fraud detection using data mining: A systematic literature review. *Journal of Insurance Research*, [online]. Available at: https://www.sciencedirect.com/science/article/pii/S2667305324000164 [Accessed 8 January 2025].

Shah, A., 2024. The transformative impact of AI on Kenya's insurance industry. *Business Daily Africa*, [online]. Available at: https://www.businessdailyafrica.com/bd/opinion-analysis/columnists/the-transformative-impact-of-ai-on-kenya-s-insurance-industry--4545286 [Accessed 8 January 2025].

Tookitaki, 2024. Unveiling the dark reality of insurance fraud: A comprehensive guide. [online]. Available at: https://www.tookitaki.com/glossary/insurance-fraud [Accessed 8 January 2025].

Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J. & Yoo, Y., 2019. CutMix: Regularization strategy to train strong classifiers with localizable features. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6023–6032. Available at: https://doi.org/10.1109/ICCV.2019.00614 [Accessed 8 January 2025].

Zhang, H., Cisse, M., Dauphin, Y.N. & Lopez-Paz, D., 2018. mixup: Beyond empirical risk minimization. *International Conference on Learning Representations (ICLR)*, [online]. Available at: https://arxiv.org/abs/1710.09412 [Accessed 8 January 2025].

Note: All access dates are as of 8 January 2025, the date of finalizing the reference list.