Journal of Global Research in Electronics and Communication
Volume 1, No. 12, December 2025
Available Online at: www.jgrec.info
RESEARCH PAPER

AI-Driven Fraud Detection in Insurance Claims: A Deep Learning Framework for FinTech Risk Intelligence

Dr. Hemant N. Patel
Assistant Professor
Sankalchand Patel College of Engineering
Sankalchand Patel University
Visnagar
hp15284@gmail.com

***Abstract*—The growth rate of fraudulent insurance claims is a significant challenge to financial institutions, and it is causing huge losses to the economy and inefficiency in operations. The purpose of this research is to find a solution to this growing problem by creating an automated system that can reliably forecast the number of fraudulent property insurance claims. To combat insurance data misclassifications and improve fraud detection, an Artificial Neural Network (ANN) machine learning model is trained in this study. The model has been trained and evaluated using a publicly available Kaggle dataset, which has 38 customer profiles, claim history, and policy details as features. The use of extensive preprocessing operations such as cleaning of data, label coding, working with missing values and balancing with the help of SMOTE, and normalization with the help of Standard Scaler provide quality of data and resilience. The proposed ANN model is tested with the help of common performance measures, including accuracy (ACC), precision (PRE), recall (REC), F1-score (F1), confusion matrix, ROC curve, and AUC. As the experimental outcome shows, with an accuracy of 96.67, the experiment is more accurate than the current baseline models like XGBoost, Decision Trees and Bi LSTM networks. The results show that ANN is effective in identifying unusual patterns of claims and this can be applied in enhancing the fraud prevention strategies and decision-making in the finance and insurance industry.**

Keywords*—Insurance Fraud Detection, Finance, Artificial Intelligence (AI), Machine Learning, Deep Learning (DL), Risk Intelligence.

I. INTRODUCTION

Insurance is typically defined as the transmission of the risk of loss from the insured to the insurer in consideration of a premium. In order to mitigate the effects of life's inevitable risks, people turn to insurance markets [1]. Insuring people has been around since ancient times, but the industry really took off in the twentieth century, thanks to increased international trade and investments from companies all over the globe. Neither Europe nor North America is home to more than half of the world's insurance companies [2]. A major portion of the insurance needs of the rest of the world are served by these companies. As a result, reasonable effort is necessary. Pursuing monetary gain by dishonest and unlawful means is known as financial fraud [3]. Certain industries are more vulnerable to financial fraud than others, including the insurance, banking, tax, and corporate spheres [4]. Companies and sectors have been facing a growing problem with financial fraud, money laundering, and other forms of financial transaction fraud at recent times. Huge sums of money are lost daily due to fraud, and this happens despite multiple attempts to curb financial crime. As a result, society and the economy suffer.

Traditional rule-based systems struggle to identify complicated fraud, despite the vital nature of detecting such actions in the financial industry [5]. Improving the accuracy and efficacy of fraud detection procedures through the use of AI techniques, particularly deep learning and machine learning, has shown encouraging results. As a result of the substantial monetary losses caused annually by fraudulent claims [6]Insurance fraud detection is an important issue in the sector. There is a tendency for the statements to be overstated, fabricated, or deliberately twisted. Business operations are disrupted and policyholder premiums are increased as a result of this [7]. It could be challenging to detect fraud due to the large volume of claims data or the number of possible permutations used in an attempt to defraud. This requires the ability to identify unusual patterns or abnormalities. The sheer volume and complexity of contemporary fraud cases have rendered conventional methods and strategies obsolete.

The field of fraud detection and prevention has been revolutionized by the rise of artificial intelligence (AI) [8]. Advanced methods for real-time fraud detection and mitigation are available through AI's use of machine learning, data analytics, and predictive modelling. Within this framework, AI has become a game-changer that can radically alter the way fraud is detected [9]. The integration of cognitive capacities with deep learning models has brought about a fresh age of intelligent data analysis in the field of finance technology (fintech) [10]. AI-based systems utilize machine learning (ML) and deep learning (DL). Strong tools capable of extracting valuable insights are in high demand due to the ever-increasing complexity and volume of data [11]. The capacity of deep learning models to learn hierarchical representations from unstructured data has made them a revolutionary tool. These models are now being used more and more to spot suspicious patterns in massive datasets, identify fraudulent activity as it happens [12], and cut down on financial losses. Research and debate have recently focused on the efficacy of these systems in many industries, including as healthcare, banking, and insurance [13].

A. Structure of the Paper

The outline of the paper is as follows: Section II provides a literature survey on the topic of insurance fraud detection. Methods for data preparation, implementing models, and enhancing interpretability are detailed in Section III. Section IV delves into the outcomes of the experiments and the analysis of comparisons. Section V concludes the study while also outlining possible avenues for further investigation.

B. Significance and Contribution of This Study

The study is also important as fraudulent insurance claims have kept financial organizations incurring heavy financial losses and straining them with administrative workloads. This research enables the insurers to identify suspicious trends in a timely manner and implement better risk responses and operational protection by creating an effective and scalable machine-learning model. The findings provide the data to confirm that ANN-based systems are more effective in fraud detection compared to a number of classical ML methods.

- Applied numerous amounts of preprocessing such as cleaning, encoding, scaling and missing-value treatment.
- Used SMOTE to equalize the unequal classes and enhance generalization of the model.
- Created a binary fraud detection ANN architecture.
- Carried out model assessment based on ACC, PRE, REC, F1, ROC, and confusion matrix.
- Undergone comparative analysis with XGBoost, Decision Trees and Bi-LSTM.

C. Justification and Novelty

The challenges that insurance companies are currently encountering are becoming hard to overcome because of the advanced cases of fraudulent activities that cannot be easily identified via the traditional rules base or manual system of investigations. Such traditional methods are not always capable of finding the concealed patterns in large and complicated data, which leads to tardy recognition and increased losses. Thus, AI-based fraud detection models necessary to detect risks at an early stage. The necessity of having a solid, automated, and data-driven framework that can enhance the level of accuracy of the decisions and the lowest rate of false positives justifies this research. Through ANN, ANN overcomes the main data issues, namely, missing values, skewed classes and noisy features, which make the study methodologically robust and significantly improve the performance and operational effectiveness of fraud detection among insurers.

The originality of the current research is the incorporation of a full and optimized ANN-based pipeline with advanced preprocessing like the statistic operations that are unique to insurance fraud detection. In contrast to earlier literature which uses single models or small preprocessing, the current study uses SMOTE-based balancing, Standard Scaler normalization, and an optimized ANN architecture to demonstrate better predictive performance. The relative comparison with several ML and DL models also confirms the originality and usefulness of the suggested strategy in the financial fraud detection context.

II. Literature Review

The literature review indicates the recent progress in the field of insurance fraud detection based on such models as Cat Boost, Random Forest, Logistic Regression, and hybrid FL-GA-PSO. Accuracy has been enhanced with effective feature selection and model fusion, as well as, providing better tools with respect to the detection and prevention of fraudulent insurance claims.

Xie (2025) CatBoost algorithm with Bayesian optimization was finally used to achieve a prediction accuracy of 0.95. At the same time, the constructed random forest and multi-attention mechanism model can output the prediction probability and average attention weight map of each case, which shows that the model pays more attention to the severity of the accident. The results of the study provide effective technical support for the identification and prevention of insurance fraud, and propose the construction of a risk scoring mechanism to assist in the development of personalized programs [14].

Feng et al. (2024) four baseline algorithms were selected with the best performance: XGBoost, LightGBM, CatBoost, and Random Forest. A feature selection algorithm has been designed based on the concept of greed, which selects three of the most essential features from 80 features while still ensuring high accuracy. Finally, the baseline model was trained using the three selected features, and grid search was used for parameter tuning. After model fusion, an accuracy of 92.3% was achieved on the test set [15]

Al-Ghazi et al. (2024) used a dataset provided in Kaggle titled Healthcare Provider Fraud Detection Analysis using RF Classifier and LR. The best-performing model in this test, the Logistic Regression, is then used to which features are the most important for the classification. Research shows that the most important feature in detecting health insurance fraud is the amount of money reimbursed associated with a provider. The LR model achieved an ACC of 0.90, PRE of 0.93, REC of 0.91, and an F1 of 0.90, outperforming the RF model in comparative analysis [16].

Preetham et al. (2024) proposed model recorded an accuracy of 78 %. The results reveal that by successfully modelling the links between observed and hidden factors, Hidden Naive Bayes shows potential in identifying phoney insurance claims. To improve the algorithm and investigate its sturdiness in managing massive and intricate insurance datasets, additional study is necessary. This study advances insurance fraud detection by highlighting Hidden Naive Bayes's promising future [17].

Y et al. (2023) suggested research takes use of the best features of three different technologies—Federated Learning (FL), Genetic Algorithm (GA), and Particle Swarm Optimization (PSO). For optimal feature subset extraction, the proposed model employs GA. Federated learning with particle swarm optimization (FPSO) is subsequently fed the optimized feature subset. The results demonstrate that the proposed hybrid model achieves an accuracy of 94.47% and that it has room for improvement by incorporating additional nature-based algorithms specifically designed for detecting fraud [18].

Rath and Panigrahi (2023) showcase an innovative approach to improving overall performance by combining correlation-based and forward feature selection strategies based on feature relevance. A number of supervised learning techniques are compared in this study. These algorithms include LR, DT Classifier, RF, SVM, and ANN. A remarkable

mean accuracy rate of 93.83% was achieved by the RF technique, which produced the highest accuracy [19].

Recent studies in the detection of insurance fraud have been emphasized in Table I and depict great accuracy levels in models such as Cat Boost, Random Forest, and hybrid FL-GA-PSO frameworks. Primary constraints include constraints of the datasets and scalability, whereas the future goal is larger datasets, enhanced robustness, and enhanced interpretability.

TABLE I. SUMMARY OF PREVIOUS WORK ON INSURANCE FRAUD DETECTION STUDIES

Authors (Year)	Dataset	Methods Used	Key Findings	Limitations & Future Work
Xie (2025)	Not specified	CatBoost with Bayesian Optimization; Random Forest; Multi-Attention Mechanism	CatBoost achieved 0.95 accuracy; Attention model highlights accident severity; Supports fraud identification and prevention; Proposes risk-scoring mechanism.	Further validation on diverse datasets; Expand risk-scoring mechanism.
Feng et al. (2024)	Not specified	XGBoost, LightGBM, CatBoost, Random Forest; Greedy Feature Selection (3 features from 80); Grid Search; Model Fusion	Final fused model achieved 92.3% accuracy using only 3 essential features.	Extend feature selection approach; Test on larger datasets; Explore advanced fusion methods.
Al-Ghazi et al. (2024)	Kaggle: Healthcare Provider Fraud Detection Analysis	Logistic Regression; Random Forest	Logistic Regression performed best with 0.90 accuracy, 0.93 precision, 0.91 recall, 0.90 F1; Most important feature: reimbursed amount.	Improve RF performance; Explore more ML models; Use balanced datasets.
Preetham et al. (2024)	Not specified	Hidden Naive Bayes	Achieved 78% accuracy; Effective at modelling hidden and observed factors in fraud cases.	Needs improvement for large/intricate datasets; Further robustness testing required.
Y et al. (2023)	Not specified	Federated Learning + GA + PSO (Hybrid FPSO model)	Hybrid GA-FPSO-FL approach achieved 94.47% accuracy; Shows potential for fraud detection.	Explore other nature-inspired algorithms; Scalability studies.
Rath & Panigrahi (2023)	Not specified	Correlation-based FS + Forward Selection; LR, DT, RF, SVM, ANN	Random Forest achieved the best result: 93.83% mean accuracy.	Extend evaluation with more algorithms; Improve FS automation.

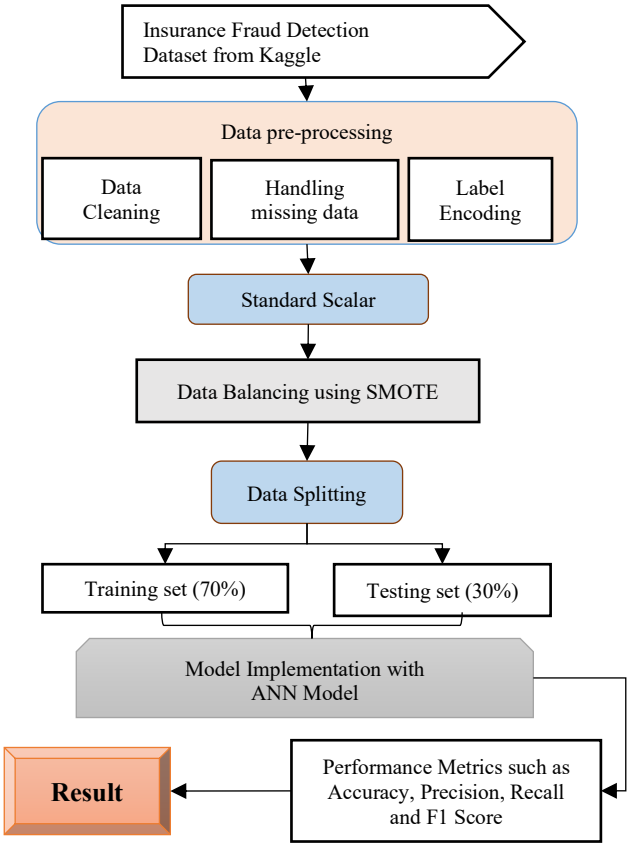

Fig. 1. Proposed Flowchart for Insurance Fraud Detection

III. METHODOLOGY

The first stage in the methodology is to collect and learn the Kaggle insurance fraud dataset that has 38 features. Preprocessing Data cleaning Data cleaning involves the removal of invalid entries, missing values, encoding categorical data with labels, and Standard Scaler to normalize numerical data and SMOTE to deal with class imbalance. The data that has been processed is divided into a 70:30 training and testing ratio. A model of an Artificial Neural Network is then generated that has optimized layers and activation functions and is used to do binary classification. ACC, REC, PRE, F1, confusion matrix, and ROC curve are used to evaluate model performance, and the model is then compared with other machine learners, as shown in Figure 1.

A. Data Collection

This publicly available Kaggle dataset provides detailed information for insurance fraud detection, containing 38 customer and claim-related features along with a target variable indicating whether fraud was reported. The data includes demographics, policy details, claim history, and other relevant attributes that help in understanding patterns associated with fraudulent behavior. A lot of people use it to create and test machine-learning models that try to guess how likely fraud is to happen, which helps insurance companies measure risk and make decisions.

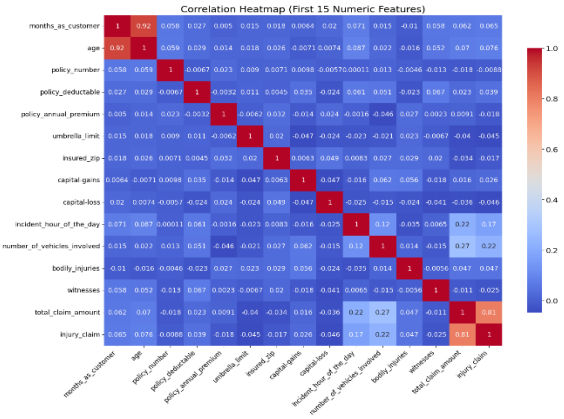

Fig. 2. Correlation Heatmap of Top 15 Features

Figure 2 shows the correlation between the first 15 numerical features, where red indicates strong positive relationships, blue indicates negative relationships, and most features display only weak correlations with each other.

B. Dataset Preprocessing

To have consistency and compatibility of the models, Data Cleaning, missing data, scaling the data and Class imbalance were used to preprocess the datasets. These procedures were critical in preventing data leaks and improving the models' performance. There is a list of these in the following:

- **Data Cleaning:** Data cleaning involves the removal of corrupt data and erroneous entries in a record set or a database table[20]. The primary application of cleaning step is founded on the identification of incomplete, erroneous, inconsistent and irrelevant data and the methods of modifying or erasing this useless data.
- **Handling Missing Data:** There is a stage of preparing missing values in a dataset, which is one of the most essential stages. There are various reasons why missing values may happen, among them are wrong data entry, system malfunction, or inadequate data collection [21]. Unfinished or damaged records are corrected or deleted in order to ensure the data is of quality and to avoid bias in training.
- **Categorical Encoding using Label Encoding:** One of the most involved parts of data preprocessing is categorical data encoding, which involves changing data without a definite numerical value to an exact one [22]. Every piece of category data now has a numerical value thanks to this update. Each distinct textual value is converted to an integer based on the sequence in the label encoding data encoding method.

C. Scale Data Using Standard Scalar

One application of Z-score normalization is the Standard Scaler method. To standardize attributes, it is necessary to subtract the mean from the value of each attribute and divide the result by the standard deviation S. This yields an attribute with a zero mean and unit variance [23]. A value x_i is translated into x_i' using Equation (1), where $\bar{x}$ is the mean of the x variable.

$$x_i' = \frac{x_i - \bar{x}}{s} \quad (1)$$

D. Handling Class Imbalance Using SMOTE

The production of synthetic samples is the meat and potatoes of the SMOTE algorithm. By producing fresh, synthetic samples of the underrepresented class, the algorithm proves its mettle in this stage of tackling class imbalance [24]. To get N synthetic samples for every minority sample xi, SMOTE takes the required oversampling rate as N. Here is the procedure outlined in Equation (2) for creating each new synthetic sample x.

$$x_{new} = x_i + \lambda(x_j - x_i) \quad (2)$$

The distribution of classes both before and after SMOTE application is illustrated in Figure 3. The dataset is skewed at the outset since the count of the majority class (0) is substantially more than that of the minority class (1). Once SMOTE is applied, the distribution of the minority class is balanced because it is synthetically oversampled to match the majority class. The model's performance is enhanced by preventing bias towards the majority class, thanks to this balancing.

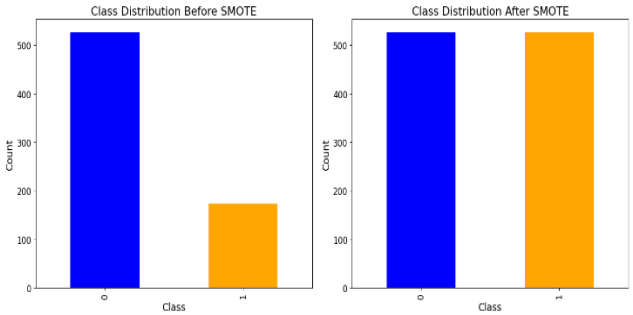

Fig. 3. Class Distribution before and After Balancing

E. Data Splitting

A training data set and a testing data set were created with the data divided 70:30. These are the model parameters that are learnt from the training set and evaluated on the testing set. They help with the model's prediction power on unknown data and generalizability.

F. Artificial Neural Network (ANN)

ANNs can be used to solve problems in areas like game design and pattern recognition by using current statistics and biological ideas. The principle idea of ANNs is that of neuron mimics connected in the following manner [25], Multiply the input value x_i by the weights w_i and add the multiplied values of the inputs. The weight of the connections between the neurons of a neuron determines its output, w_1 more influential than w_2 due to the weight of the connection between the neurons and even when they are equal in weight.

After multiplying each input value x_i by its associated weights w_i Add the resulting results. The strength of the weighted connections between neurones influences the output of neurons. Even when both w_1 and w_2 have the same weight, w_1's bigger weight makes it more influential, as in Equation (3).

$$\sum = (x_1 * w_1) + (x_2 * w_2) + \cdots + (x_n * w_n) \quad (3)$$

Each row vector represents an input, and the weights are supplied by w = [w_1, w_2, …, w_n] and x = [x_1, x_2, …, x_n] accordingly. Equation (4) gives the dot product.

$$x.w = (x_1 * w_1) + (x_2 * w_2) + \cdots + (x_n * w_n) \quad (4)$$

Hence, Equation (4) is equal to Equation (5).

$$\sum = x.w \quad (5)$$

This product of multiplied numbers again with the added bias b, and call z. Applying a bias or offset to the entire activation function is necessary to achieve the desired output values, as in Equation (6).

$$z = x.w + b \quad (6)$$

A non-linear activation function is used to transform z according to the provided value. A neuron would only generate a straight line if activation functions weren't used to make the output of the neuron non-linear. The functions also have a significant impact in the neural network's learning rate. Use the sigmoid function (or logistic function) as the activation function, as shown in Equation (7), even though the activation function is typically a step function, which is its binary form (7).

$$\hat{y} = \sigma(z) = \frac{1}{1+e^{-z}} \quad (7)$$

The projected value is given in equation (7) after the forward prorogation and it is the sigmoid activation function.

G. Model Evaluation

ACC, specificity, REC, PRE, and F1 were the performance metrics used to assess the ML models' capacity to foretell fraudulent transactions. The confusion matrix was used to evaluate the model performance by showing the True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) [26]. This assists in gauging both the accuracy and the level of effectiveness of the models in detecting fraudulent transactions, reducing false alarms, and reducing the issue of missed fraud. The performance measure calculating equations are as follows:

- **True Positive (TP):** True Positive (TP) data is the amount of positive data that is accurately predicted as True.
- **True Negative (TN):** True Negative (TN) refers to the total number of negative data points that were accurately anticipated to be negative.
- **False Positive (FP):** The count of the data points that are forecasted in the positive class yet belong to the negative class is known as False Positive (FP).
- **False Negative (FN):** The number of negative predictions that turn out to be positive False Negatives (FN).

1) Accuracy

ACC, the most straightforward performance metric, is defined as the proportion of samples correctly classified to the total samples. It comes in handy when the target classes are balanced out. Accuracy can be defined as follows with Equation (8):

$$Accuracy = \frac{TP + TN}{TP+TN+FP+FN} \quad (8)$$

2) Precision

Precision may be regarded as a metric of the accuracy of a classifier. This ratio of positive samples appropriately classified to total samples classified as positive; precision is presented with Equation (9):

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

3) Recall

REC, sensitivity, or the true-positive rate is the percentage of positively predicted observations out of an actual class that are accurate. To find out how well the model predicts the positive cases, it uses this metric. Equation (10) finds the definition of REC:

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

4) F1 Score

PRE and REC means are weighted to give F1. This score, therefore, considers FP and FN. It is superior to ACC in the case of unbalanced classes. Equation (11) defines F1:

$$F1\ Score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (11)$$

IV. Results, Discussion & Comparative Analysis

The studies were carried out using a Windows 11 laptop from Dell Inspiron, powered by an Intel® Core TM i5-1135G7 CPU running at 2.40 GHz. Used Python notebooks on Google Colab to implement and test the model. The results of an artificial neural network (ANN) model used to detect insurance fraud are shown in Table 2. The model's ACC is 96.67%, with a REC of 95.76%, an F1 of 96.03%, and a PRE of 98.14%. The model does decently in terms of overall accuracy, but it might do better when it comes to detecting fake insurance, as seen by its middling REC and F1.

TABLE II. Results of Insurance Fraud Detection

Measures	ANN
Accuracy	96.67
Precision	98.14
Recall	95.76
F1 Score	96.93

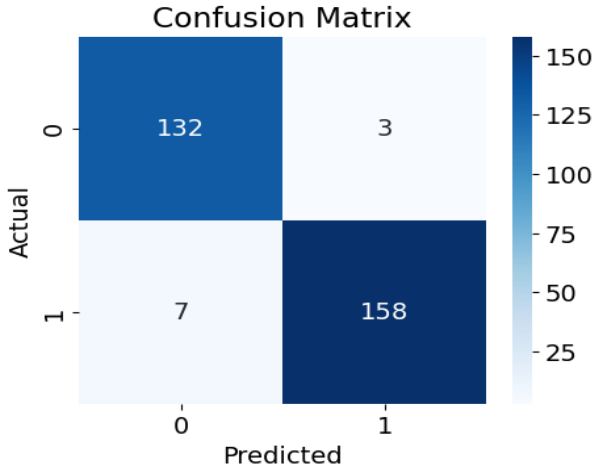

Fig. 4. Confusion Matrix of ANN Model

Figure 4 indicates that the confusion matrix indicates that the ANN model is incredibly accurate and correctly recognizes most of the samples of the two classes, class 0 and class 1. The model records very low misclassifications, and thus it has a good balance between sensitivity and specificity. This distribution indicates the strength and accuracy of the ANN with regard to the classification exercise.

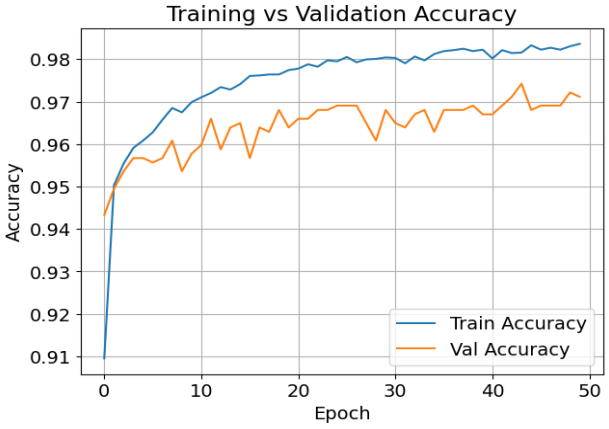

Fig. 5. Training vs Validation Accuracy Curve

Figure 5 the accuracy graph indicates that the ANN model is progressively enhancing the performance in 50 epochs, and training accuracy is increasing gradually, and the validation accuracy is increasing immediately. The two curves settle at around 97-98%, which shows that the model does not overfit too. The trend establishes the fact that the DNN model is able to extrapolate to unknown data.

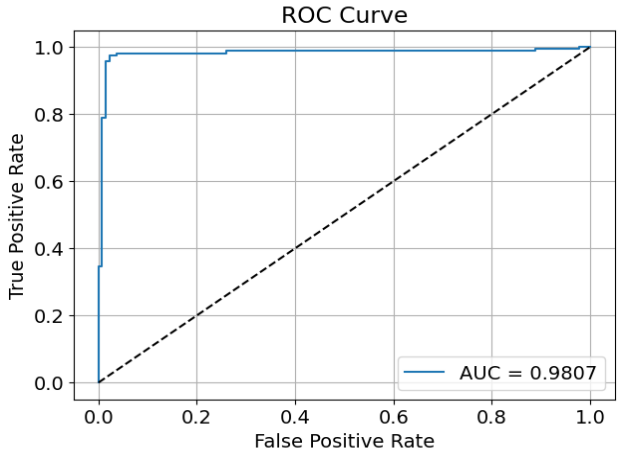

Fig. 6. ROC Curve

Figure 6 indicates that the ROC curve of the ANN model exhibits a high level of discriminative power, and the line is sharply increasing towards the top-left corner, indicating that it has a good separation between classes. The model has a high AUC of 0.9807, which means that it is able to correctly identify fraudulent and non-fraudulent cases with a high degree of consistency. In the whole, the ROC plot indicates an extremely trustworthy ANN classifier.

A. Comparative Analysis

Table III gives a comparative analysis of the various models that are applied in Fraud Insurance Prediction, and a definite difference in model performance is evident. ANN model has the best ACC with 96.67% and has a high PRE, REC, and F1, which means that it is effective in fraud case detection. The next model is the DT with 93.32% ACC, which has equal performance, and the XGBoost is 89% with moderate reliability. Conversely, the Bi-LSTM model has a much lower ACC of 80.80% and almost negligible PRE and F1, implying that it is extremely hard to make the correct decision on the claims of fraud.

TABLE III. COMPARATIVE ANALYSIS OF DIFFERENT MODELS ON FRAUD INSURANCE PREDICTION

Models	Accuracy	Precision	Recall	F1 Score
ANN	96.67	98.14	95.76	96.93
XGB[27]	89.00	86.00	88.00	87.00
DT[28]	93.32	93.46	93.17	93.31
Bi-LSTM[29]	80.80	5.90	14.30	8.5

The suggested ANN model does better than all the others that were tested, getting the best ACC, PRE, memory, and F1. This demonstrates its superior capability in detecting both fraud insurance and non-fraudulent insurances on the Dataset.

V. CONCLUSION AND FUTURE WORK

The increasing numbers of fraudulent insurance claims has posed a huge burden on the insurance industry in terms of financial burden and at the same time, early and proper detection of such cases is of utmost importance in terms of operational stability, as well as economical safety. To tackle this dilemma, a good, solid, data-based strategy was necessary that could help in detecting sophisticated fraud patterns that the conventional rule-based frameworks tend to ignore. This paper was able to devise a high-performance machine-learning structure based on an ANN to fulfill this requirement. The proposed ANN model showed a high degree of accuracy of 96.67%, which is excellent predictive accuracy over other models like XGBoost, DT and Bi-LSTM. The study pre-processed the data, cleaning it, encoding labels, addressing missing values, SMOTE balancing, and Standard Scaler normalizing the data, to quality-check the data and enhance the generalization of the model. PRE, REC, F1, confusion matrix, and ROC-AUC were used to identify the good results of the model to differentiate between fraudulent and legitimate claims with minimum misclassification.

In the future, better models can help represent more behavioral regularities in the framework of this study. Generalization can also be enhanced using larger and more heterogeneous datasets across various insurance fields. The use of Explainable AI (XAI), such as SHAP and LIME, will assist in enabling the insurers to understand the decisions made by the model, as well as improve transparency. Moreover, real-time monitoring of frauds, federated learning with privacy-preserving training with blockchain to generate claim records, are all promising future research and industry applications.

REFERENCES

[1] M. Demirkaya, “Insurance is generally defined as the transfer of risk of loss from the assured to the insurer in exchange for a premium (2),” 2019.

[2] S. R. P. Madugula and N. Malali, “AI-powered life insurance claims adjudication using LLMs and RAG Architectures,” *Int. J. Sci. Res. Arch.*, vol. 15, no. 1, pp. 460–470, Apr. 2025, doi: 10.30574/ijsra.2025.15.1.0867.

[3] S. K. Das, “Transforming Loan and Distribution Processing in Retirement Systems: A QA Automation Approach,” *Int. J. Bus. Manag. Sci.*, vol. 05, no. 05, pp. 09–26, May 2025, doi: 10.55640/ijbms-05-05-02.

[4] A. Ali *et al.*, “Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review,” *Appl. Sci.*, vol. 12, no. 19, Sep. 2022, doi: 10.3390/app12199637.

[5] A. H. Ali and A. A. Hagag, “An enhanced AI-based model for financial fraud detection,” *Int. J. Adv. Appl. Sci.*, vol. 11, no. 10, pp. 114–121, Oct. 2024, doi: 10.21833/ijaas.2024.10.013.

[6] S. Gupta, “An Advanced Machine Learning Method Used for Identification of Banking Fraud in Financial Transactions,” *J. Glob. Trends Artif. Intell.*, vol. 1, no. 1, pp. 28–35, 2025.

[7] P. A. Kumar and S. Sountharrajan, “Insurance claims estimation and fraud detection with optimized deep learning techniques,” *Sci. Rep.*, vol. 15, no. 1, p. 27296, Jul. 2025, doi: 10.1038/s41598-025-12848-0.

[8] O. A. Bello and K. Olufemi, “Artificial intelligence in fraud prevention: Exploring techniques and applications challenges and opportunities,” *Comput. Sci. IT Res. J.*, vol. 5, no. 6, pp. 1505–1520, Jun. 2024, doi: 10.51594/csitrj.v5i6.1252.

[9] P. Adhikari, P. Hamal, and F. B. Jnr, “Artificial Intelligence in fraud detection: Revolutionizing financial security,” *Int. J. Sci. Res. Arch.*, vol. 13, no. 1, pp. 1457–1472, Sep. 2024, doi: 10.30574/ijsra.2024.13.1.1860.

[10] S. B. Shah, “Evaluating the Effectiveness of Machine Learning in Forecasting Financial Market Trends: A Fintech Perspective,” in *2025 3rd International Conference on Integrated Circuits and Communication Systems (ICICACS)*, IEEE, Feb. 2025, pp. 1–6. doi: 10.1109/ICICACS65178.2025.10968297.

[11] B. Sailesh and S. Lingesh, “Transforming Fintech through Deep Learning: Insights, Innovations, and Ethical Considerations,” *Int. J. Res. Publ. Rev.*, vol. 5, no. 9, pp. 2818–2827, 2024.

[12] D. Patel, “Enhancing Banking Security: A Blockchain and Machine Learning- Based Fraud Prevention Model,” *Int. J. Curr. Eng. Technol.*, vol. 13, no. 06, pp. 576–583, Dec. 2023, doi: 10.14741/ijcet/v.13.6.10.

[13] Y. Macha and S. K. Pulichikkunnu, “A Data-Driven Framework for Medical Insurance Cost Prediction Using Efficient AI Approaches,” *IJRAR- Int. J. Res. Anal. Rev.*, vol. 11, no. 4, pp. 887–893, 2024.

[14] H. Xie, "CatBoost and Bayesian Optimization Algorithm-Based Classification of Fraudulently Insured Persons and Random Forest-Based Fraud Prediction Combining Multiple Attention Mechanisms," in *2025 5th International Conference on Machine Learning and Intelligent Systems Engineering (MLISE)*, IEEE, Jun. 2025, pp. 103–110. doi: 10.1109/MLISE66443.2025.11100175.

[15] J. Feng *et al.*, "Medical Insurance Fraud Risk Monitoring and Identification Model Based on Feature Selection and Machine Learning," in *2024 3rd International Conference on Cloud Computing, Big Data Application and Software Engineering (CBASE)*, IEEE, Oct. 2024, pp. 571–574. doi: 10.1109/CBASE64041.2024.10824405.

[16] M. K. Al-Ghazi, R. Bertrand, M. D. Q. Destra, A. A. S. Gunawan, and K. E. Setiawan, "Classification of Health Insurance Fraud Risk with Machine Learning," in *2024 International Conference on Information Technology Research and Innovation (ICITRI)*, IEEE, Sep. 2024, pp. 24–29. doi: 10.1109/ICITRI62858.2024.10699052.

[17] G. Preetham, K. Siddu, B. Ramesh, M. A. Jabbar, and S. Sucharita, "Insurance Claim Fraud Detection Using Hidden Naive Bayes," in *2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT)*, IEEE, Mar. 2024, pp. 1–6. doi: 10.1109/ICDCOT61034.2024.10516207.

[18] S. Y, N. Victor, G. Srivastava, and T. R. Gadekallu, "A Hybrid Federated Learning Model for Insurance Fraud Detection," in *2023 IEEE International Conference on Communications Workshops (ICC Workshops)*, IEEE, May 2023, pp. 1516–1522. doi: 10.1109/ICCWorkshops57953.2023.10283682.

[19] S. Rath and S. Panigrahi, "Healthcare Insurance Fraud Prediction with Correlation based Forward Feature Selection," in *2023 2nd International Conference on Ambient Intelligence in Health Care (ICAIHC)*, IEEE, Nov. 2023, pp. 1–5. doi: 10.1109/ICAIHC59020.2023.10431444.

[20] V. Agarwal, "Research on Data Preprocessing and Categorization Technique for Smartphone Review Analysis," *Int. J. Comput. Appl.*, vol. 131, no. 4, pp. 30–36, Dec. 2015, doi: 10.5120/ijca2015907309.

[21] S. A. Farooq, O. Konda, A. Kunwar, and N. Rajeev, "Anxiety Prediction and Analysis- A Machine Learning Based Approach," in *2023 4th International Conference for Emerging Technology (INCET)*, IEEE, May 2023, pp. 1–7. doi: 10.1109/INCET57972.2023.10170115.

[22] F. Bolikulov, R. Nasimov, A. Rashidov, F. Akhmedov, and Y.-I. Cho, "Effective Methods of Categorical Data Encoding for Artificial Intelligence Algorithms," *Mathematics*, vol. 12, no. 16, Aug. 2024, doi: 10.3390/math12162553.

[23] L. B. V. de Amorim, G. D. C. Cavalcanti, and R. M. O. Cruz, "The choice of scaling technique matters for classification performance," *Appl. Soft Comput.*, vol. 133, Jan. 2023, doi: 10.1016/j.asoc.2022.109924.

[24] G. Husain *et al.*, "SMOTE vs. SMOTEENN: A Study on the Performance of Resampling Algorithms for Addressing Class Imbalance in Regression Models," *Algorithms*, vol. 18, no. 1, Jan. 2025, doi: 10.3390/a18010037.

[25] S. K. Saini, "Mathematical Model of ANN," *J. Emerg. Technol. Innov. Res.*, vol. 8, no. 3, pp. 89–93, 2021.

[26] R. Qasrawi, S. P. V. Polo, D. A. Al-Halawa, S. Hallaq, and Z. Abdeen, "Assessment and Prediction of Depression and Anxiety Risk Factors in Schoolchildren: Machine Learning Techniques Performance Analysis," *JMIR Form. Res.*, vol. 6, no. 8, Aug. 2022, doi: 10.2196/32736.

[27] P. R. Jagdale and M. M. Sukale, "Fraud Detection in Auto Insurance Claims Using Machine Learning Algorithms," *Int. J. Latest Technol. Eng. Manag. Appl. Sci.*, vol. 14, no. 13, pp. 245–256, Oct. 2025, doi: 10.51583/IJLTEMAS.2025.1413SP050.

[28] R. Q. Majumder, "Designing an Intelligent Fraud Detection System for Healthcare Insurance Claims Using a Machine Learning Approach," in *2025 Global Conference in Emerging Technology (GINOTECH)*, IEEE, May 2025, pp. 1–6. doi: 10.1109/GINOTECH63460.2025.11076870.

[29] M. J. D. Ebinezer and B. C. Krishna, "Life Insurance Fraud Detection: A Data-Driven Approach Utilizing Ensemble Learning, CVAE, and Bi-LSTM," *Appl. Sci.*, vol. 15, no. 16, Aug. 2025, doi: 10.3390/app15168869.

