IRJEMS International Research Journal of Economics and Management Studies
Published by Eternal Scientific Publications
ISSN: 2583 – 5238 / Volume 1 Issue 2 December 2022 / Pg. No: 45-53
Paper Id: IRJEMS-V1I2P106, Doi: 10. 56472/25835238/IRJEMS-V1I2P107

Original Article

Efficient Framework for Forecasting Auto Insurance Claims Utilizing Machine Learning Based Data-Driven Methodologies

Venkataswamy Naidu Gangineni[1], Sriram Pabbineedi[2], Sriram Pabbineedi, Mitra Penmetsa[3], Jayakeshav Reddy Bhumireddy[4], Rajiv Chalasani[5], Mukund Sai Vikram Tyagadurgam[6]

[1]*University of Madras, Chennai.*
[2]*University of Central Missouri.*
[3]*University of Illinois at Springfield.*
[4]*University of Houston.*
[5]*Sacred Heart University.*
[6]*University of Illinois at Springfield.*

Received Date: 29 November 2022 Revised Date: 05 December 2022 Accepted Date: 16 December 2022

Abstract: *The industry has changed a lot because of digitization and big data and now it's crucial to forecast claims for efficiency, less fraud and better customer experiences. Traditional models are not well-suited to process complex and varied data that come from features like driver information, insurance plans and vehicle tracking. To overcome this, machine learning (ML) can use large amounts of information, adapt to any situation and spot complex patterns to enhance the decision-making process. This work presents a framework that uses InceptionV3, a deep learning (DL) model that was tailored for efficient feature learning and classification accuracy. The data went through several treatments before the training step, including filling in the missing data, encoding the labels and removing unnecessary information. The InceptionV3 model demonstrated superior performance with 97% accuracy, 96% precision, 95% recall, and 95% F1-score, outperforming both Logistic Regression (LR) and Extreme Gradient Boosting (XGBoost). This proves that the model can help insurers predict claims with better accuracy, swiftly decide on matters based on data, prevent more losses and offer more relevant services to clients.*

Keywords: *Auto Insurance, Machine Learning, Deep Learning, Inceptionv3, Claim Prediction, Data-Driven, Insurance.*

I. INTRODUCTION

Digital technologies, the collecting of big data and the need for quick and precise decisions have led to major changes in the global auto insurance industry [1]. Today, insurers have to process both structured details like who the customer is and what vehicle they use, as well as unstructured data collected from sensors, telematics records and how drivers behave.

In this environment where data is key, forecasting auto insurance claims rightly becomes a key concern. Traditional methods used by actuaries in setting rates and understanding risks are not well-equipped to handle the new features of data in insurance [2]. Because they rely on guidelines and thought processes from experts, these old systems usually lag behind changing trends and lead to problems such as misuse of capital, increased reserves, higher risk of fraud and subpar prices.

Therefore, insurance companies are starting to use ML more often to tackle these issues. Thanks to ML, we can now use data in a powerful way to discover difficult patterns, make decisions automatically and greatly improve our ability to predict the future [3]. By working with large-scale data, ML algorithms can reveal hidden relationships and are thus well-suited to tasks including claim forecasting, detecting fraud and segmenting customers.

As expenses for claims keep increasing and regulations become stricter, insurers are required to implement sophisticated and flexible methods for estimating losses. In this setup, the proposed study uses a ML-based approach that is designed to forecast auto insurance claims [4][5]. This framework's primary objective is to increase forecasting systems' accuracy and efficiency by using preprocessing, carefully choosing, and assessing the appropriate elements [6]. By adopting this approach, companies will likely save money, improve their daily operations and satisfy customers with faster claims and fair prices.

A) Motivation and Contribution of study

The auto insurance claim data is becoming more complex and overwhelming, coupled with the rise in demand for fast, accurate and automated decision-making, calls for advanced forecasting techniques that are better than traditional ML methods. Since many models fail to notice complex interactions and nonlinear aspects present in insurance data, their results are often not as good as they could be. This study is driven by the belief that the InceptionV3 architecture in DL can lead to more accurate forecasts, better generalization and higher efficiency in processing claims. The integration of deep learning is meant to boost not only the identification of risks and fraud, but also to simplify the claim review process and raise the bar for customer service in insurance.

This research offers important advancements:
- Uses a complete dataset of auto insurance claims for training and validation of the model.
- Ensures all the data used for analysis is clean by addressing problems with missing values and coding categorized elements.
- Relies on the InceptionV3 DL model to extract important information and make good forecasts.
- Models are assessed using accuracy, precision, recall and F1-score which offer a solid measure of their effectiveness.
- Provides useful information to help speed up claims and save money while also improving insurance services.

B) Novelty of paper

A novel aspect of this study is that a deep learning algorithm, InceptionV3, was successfully applied to auto insurance data which is usually handled by ML models like LR and XGBoost. Rather than relying on the models described before, InceptionV3 makes use of advanced hierarchical feature extraction to find complex patterns in the data and achieve better predictions. The proposed solution performs accurately, surpasses standard models and keeps precise recall, high F1-score and does not show excessive overfitting. This approach is not only accurate and effective, but it also makes it easier to spot risks, identify fraud and improve the standard for processing claims for auto insurance companies.

C) Structure of paper

The paper is structured as follows: Section I introduces the study's motivation. Section II reviews related work. Section III details the methodology, including data preprocessing and model implementation. Section IV presents results and comparative analysis. Section V concludes with insights and future research directions.

II. LITERATURE REVIEW

This section reviews and emphasizes predictive analytics in the auto insurance claims forecasting. The focus is on utilizing ML-based, data-driven methodologies for efficient claim prediction. Key works reviewed in this field include:

Singh et al. (2019) Provide a comprehensive system to automate the registration, processing, and decision-making procedures for auto insurance claims. Using images of damaged automobiles, the system calculates the extent of damage and the cost of repairs. It uses instance segmentation models as Mask R-CNN, PANet, and a transfer learning-based VGG16 network to find and identify fragments and damages. For components and damage localization, the system receives good map ratings (0.38 and 0.40, respectively) [7]

S Limma (2019) looks at the speed with which Awash Insurance Company S.C. processes auto insurance claims, the skill of claims staff, consumers' understanding of auto insurance, and service providers' proficiency. Data was gathered from 134 customers, with 87.3% returning questionnaires from top 10 branches in 2017/18. [8] Nowak et al. (2019) uses ML models to reduce repairs in motor insurance workshops. The models predict customer decisions based on data from the company's database and additional features. The models were created using DT, RF, GBoost, ad boost, naive Bayesian, LR , and NN. They produced an area under the curve (ROC) of above 0.8, which permitted a pilot study to be implemented in production [9].

Duan et al. (2018) propose a model that considers auto types in auto insurance claims. They divide vehicle insurance prices into three risk categories based on the indicator of auto burden. The LR model is used to fit claim frequency, and a clustering approach is used to classify rates. According to the report, more than 80% of vehicles with an auto load score of 20 or above are classified as being at the greatest risk. Whereas the Gamma distribution is more suited for fitting claim expenses, The Poisson distribution fits claim frequency more accurately. The results may help inform GLM based insurance policy recommendations [10].

Y. B. Gessese (2018) aimed to investigate the impact on customer satisfaction of EIC's car insurance claims processing. The process's five main components were characterized as follows: filing a claim, responding to one, towing a damaged car, assessing damage, handling repairs, and resolving complaints. 'Complaint or dispute settlement' was a sixth procedure for anybody bringing any claims. The research involved 102 customers and found a significant correlation between customer satisfaction and these processes. This was supported by the regression model, which had an R2 of 0.843 and a significant F ((6,95) = 38.914, $p < 0.001$) [11].

Anggraini et al. (2018) used a SWOT analysis approach to ascertain the insurance claims and premium strategy at PT. Tangerang's Assurance Sonar Mas Branch from 2005 to 2015. The study included both qualitative and quantitative approaches. The findings demonstrated that the business may raise premiums by focusing on premiums, collaborating with leasing companies, offering no-claim money back car insurance products, and increasing premiums due to tariff changes from OJK.

Additionally, the launch of new car models encourages consumers to switch to new models and get insurance. Claim suppression tactics used to stifle claims from 2005-2010 decreased from 9.51% to 8.35% from 2011-2015, aiming to increase underwriting income [12].

The comparative analysis of background study based on their Methodology, Key Findings, Limitations, and Future gap are provided in Table I.

Table 1: Summary of Reviewed Works on Machine Learning-Driven Auto Insurance Claims Forecasting

Author	Methodology	Data	Key Findings	Limitation	Future gap
Singh et al. (2019)	Deep learning (Mask R-CNN, PANet, VGG16 with transfer learning) for instance segmentation	Damaged car images	System can automatically identify damaged parts and assess damage level with mAP of 0.38 (parts) and 0.40 (damage)	Limited to image-based analysis; may not handle all damage types or real-world variations	Integrate cost estimation directly and expand model generalizability across diverse damage scenarios
Lemma (2019)	Quantitative analysis using structured questionnaires	117 returned questionnaires from AIC customers	Competency of staff, service providers, and customer awareness influence claims processing time	Results limited to a single company (Awash Insurance) and period	Expand study across multiple firms and incorporate automated systems in processing pipeline
Nowak et al. (2019)	Machine learning models (DT, RF, GB, AdaBoost, NB, LR, NN) to predict customer decisions	Real motor insurance data from a company	Achieved over 0.8 AUC for predicting customer decision, reducing external repairs	Lack of interpretability in complex models like NN; limited to decision prediction	Real-time deployment with interpretability features and dynamic policy optimization
Duan et al. (2018)	Generalized Linear Model (GLM), clustering, Poisson and Gamma distributions	Data from Chinese insurance company + Auto Burden Index	Claim frequency best fits Poisson; claim cost fits Gamma; auto burden index improves model accuracy	Auto burden index may not be generalizable outside dataset/geography	Extend to international datasets; incorporate deep learning-based risk modeling
Belay (2018)	Correlation and regression analysis	Survey data from 102 customers at EIC	Strong correlation between satisfaction and repair handling, damage assessment, complaint settlement achieved R2 of 0.843	Only examines customer perspective; lacks integration with technical claim metrics	Develop an integrated system linking customer feedback and digital claims management
Anggraini (2018)	SWOT Analysis, qualitative and quantitative mixed methods	Internal company data of motor insurance claims and premium data	Marketing mix and claim control reduced claim ratio and improved underwriting income	Descriptive insights only; lacks predictive analytics or tech-based solutions	Apply predictive models to claims control and simulate policy effects on customer retention and costs

III. METHODOLOGY

The methodology for forecasting auto insurance claims follows a structured, data-driven approach as illustrated in the flowchart in Figure 1. Leveraging advanced machine learning techniques has become increasingly vital for accurate and timely predicting claims for vehicle insurance. The first step is gathering data on vehicle insurance claims, which undergoes a data preprocessing stage. During preprocessing, missing values are handled to maintain data integrity, and irrelevant columns. After cleaning, the dataset moves to label encoding, where every categorical variable is assigned a related number for use by ML models. Encoding is pursued with the data and then to effectively execute model learning and evaluation, the data is divided into divisions for training and assessment. InceptionV3 was trained using the given data because of its known strengths in extracting features and accurately forecasting outcomes. Following training, in order to assess the model's accuracy, precision, recall, and F1-score, it applies what it has learnt to new data. Businesses can be certain that their estimates of insurance claim rates are correct in this way.

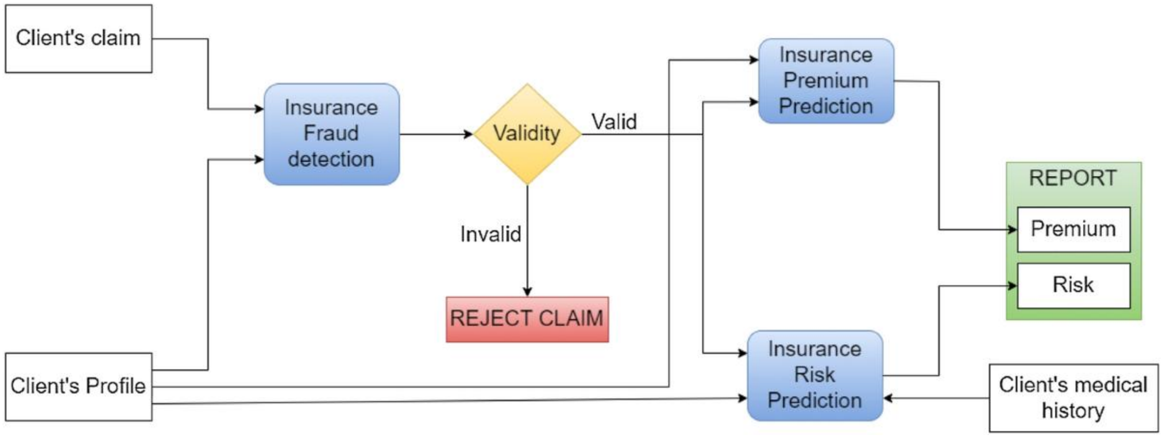

Fig 1. Proposed Methodology for Auto Insurance Claim Using Machine Learning

A) Data Collection

This Auto Insurance Claim Dataset includes information about policyholders such as their vehicles and the corresponding claims. Usually, it includes insurance details including coverage type and demographic data like age, gender, and driving history, annual premium and information about vehicle specifications. Related to the claim fields include number of claim(s) filed, claim amount, claim reason and claim status. The data set is also likely to have a fraud indicator which will be used in identifying suspicious activity. This data has broad application in the predictive modeling of claim likelihood, cost estimation and fraud detection which is crucial to the goal of enhanced risk assessment capability, pricing strategy and operational efficiency in the insurance industry. Figure 2 illustrates the relationship between the characteristics.

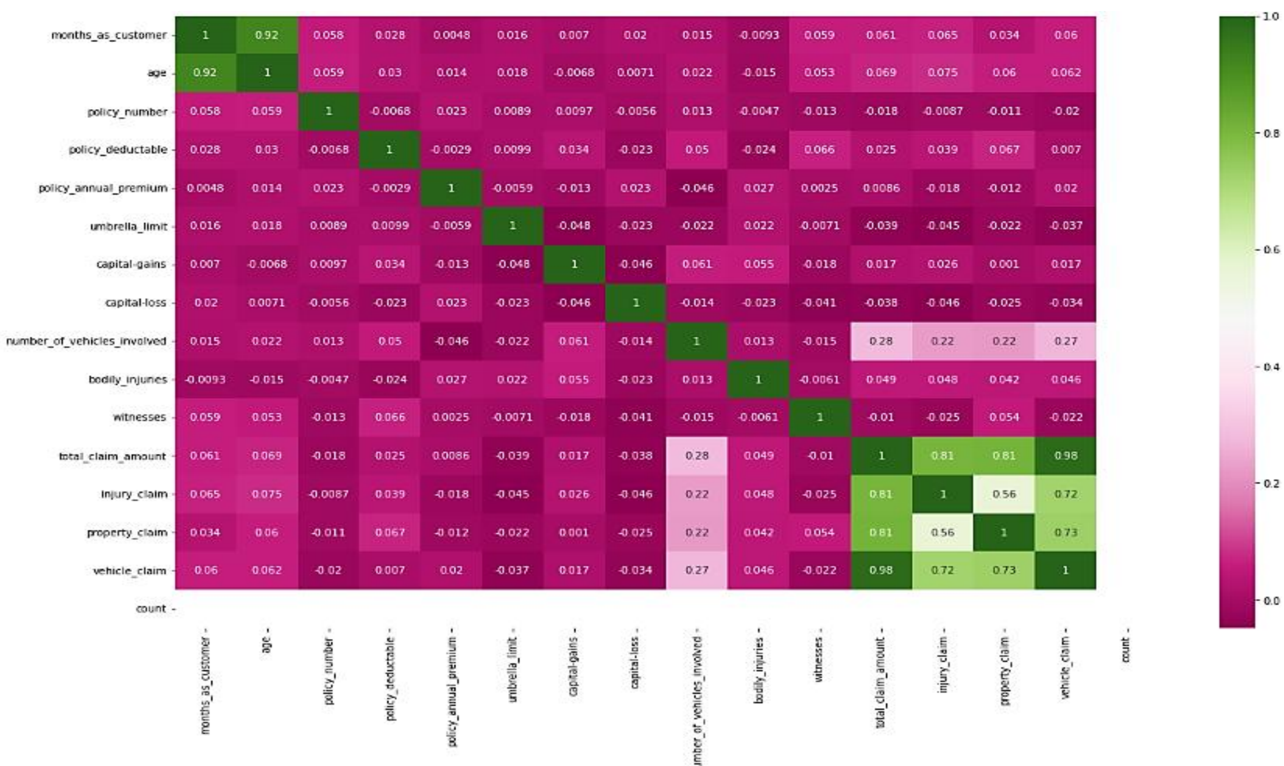

Fig 2. Correlation Matrix for Auto Insurance Claim Data

Figure 2 is a correlation heatmap showing relationships between numerical features in a dataset. Green indicates strong positive correlations, while pink shows weak or negative ones. Notably, Total_claim_amount has a strong relationship with injury_claim, property_claim, and vehicle_claim.

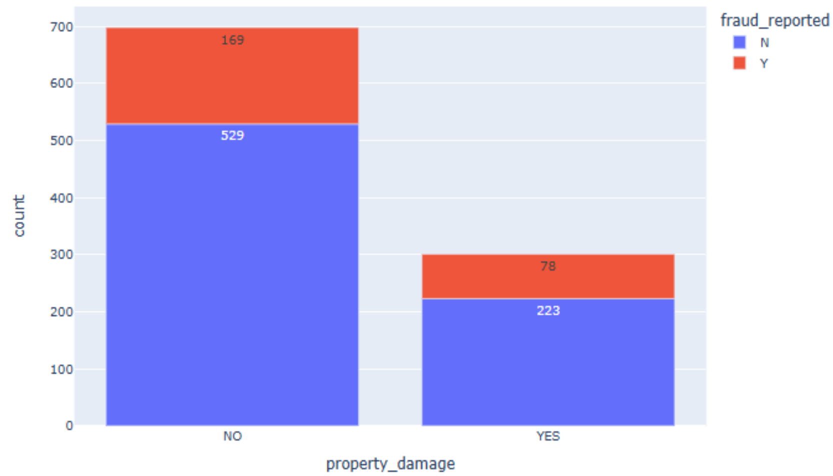

Fig 3. Distribution of fault category of the Auto Insurance Claim Dataset

The Figure 3 is a stacked bar chart showing the relationship between property damage and fraud reports. The chart compares counts of fraud reported ("Y") and not reported ("N") across two categories: property damage ("YES" and "NO"). It reveals that more fraud cases were reported when there was property damage, while fraud was less frequently reported without property damage.

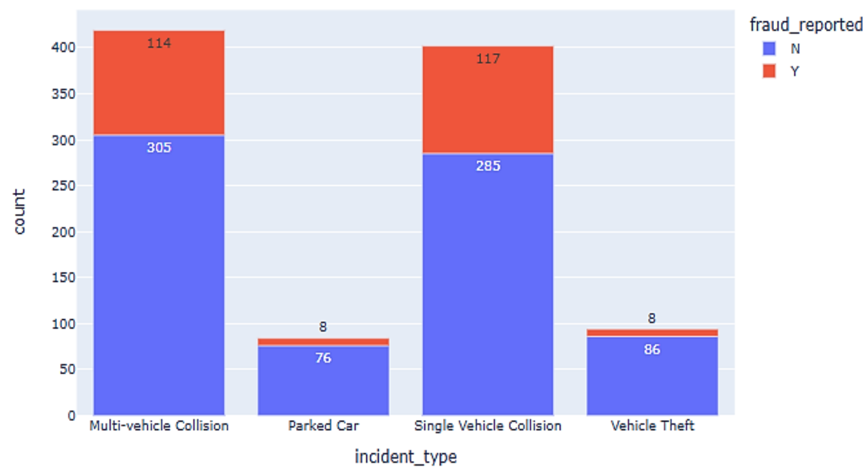

Fig 4. Distribution of Fraud Reports by Incident Type

The bar chart in Figure 4 shows the distribution of auto insurance claims the kind of event, such as vehicle theft, parked cars, single vehicle collisions, and multi-vehicle collisions. The y-axis represents claim count, with blue indicating non-fraud claims ("N") and red representing fraud claims ("Y"). Fraud is most prevalent in Multi-vehicle Collisions and Vehicle Theft, while Parked Car incidents have the least fraud.

B) Data Pre-processing

Data preprocessing is a foundational component of any ML-based framework, especially when forecasting complex outcomes such as auto insurance claims. Several essential Preprocessing techniques were used to get the dataset ready for efficient learning.:

- **Handling Missing Values:** Missing or inconsistent data is common and must be addressed before modeling. This standardization helps in identifying and handling missing data effectively, which is essential for accurate model training.
- **Dropping Unnecessary Columns:** The column _c39 was dropped as it provided no useful information. Removing irrelevant features simplifies the dataset and improves model performance by reducing noise.

C) Label Encoding

For features with many unique string categories, such as 'insured hobbies' and 'insured occupation', was employed. This encoder transforms each unique category into a corresponding integer value, preserving the feature as a single column rather than expanding it into multiple dummy variables. Label encoding is memory-efficient and suitable when the categorical variable does not have a meaningful order. It simplifies the dataset and prepares it for ML models that expect numerical inputs.

D) Data Splitting

Data splitting is a popular method for evaluating model performance in ML. To ensure that the model is tested on unknown data for a more accurate estimation of its prediction power, the dataset in this study was divided by 20% for testing and 80% for instruction.

E) Implementation of InceptionV3 Model

The third version of Google's founding the name of CNN is Inception-v3. InceptionNet-v3 included several new procedures, including RMSProp Optimiser, Factorised 7 × 7 convolutions, BatchNorm in the Auxiliary Classifiers, and Label Smoothing. Reducing the number of parameters in convolutions without compromising network performance is possible through factorization [13]. The process of label smoothing prevents overfitting. The model uses its inception modules to extract hierarchical features from the input image, taking into consideration different receptive field widths. The model uses a sequence of convolutional, pooling, and inception modules to extract high-level feature representations from an input picture x. The output prediction Þ for a picture x is provided by Equation (1), and the final classification is carried out using a fully linked softmax layer:

$$\hat{y} = softmax(W.f(x) + b)$$

Where:

- f(x) is the feature vector extracted from the input image by the convolutional base of InceptionV3,
- W is the weight matrix of the dense layer,
- b is the bias vector, and
- $softmax$ converts the raw output scores into a probability distribution over predefined class.

F) Performance Matrix

Performance metrics are crucial for assessing the effectiveness of ML models. In this work, a confusion matrix comprising the following elements was used to assess model performance:

The TP is the number of times a claim was correctly predicted by the classifier. If FP is the number of instances when the classifier incorrectly anticipated that a claim would occur, but in fact [14], The number of instances where the classifier correctly predicted that a claim would not materialize was not captured by TN. FN represents the frequency with which the classifier incorrectly anticipated that a claim would not come to fruition. The Calculations are shown below:

Accuracy: Its definition is the ratio of the sum of TP and TN to the sum of TP, TN, FP, and FN. [15]. It displays the overall frequency of correct categorization algorithms. as shown in Equation (2).

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

Recall: The proportion of a positive class that is properly categorized is called recall, and it basically indicates how effectively the model can identify the class type given in Equation (3).

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \qquad (3)$$

Precision: The precision metric is employed to assess the class's classification reliability and if it is in the correct class, as indicated by Equation (4).

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

F1 Score: A model performance statistic known as the F-measure combines recall and accuracy into a single number known as the F1 score. The formula in Equation (5) is as follows:

$$\text{F1} = \frac{2 * (\text{precision} * \text{recall})}{\text{precision} + \text{recall}}$$

These metrics are utilized for comparative analysis across various models to evaluate their effectiveness in predicting insurance claim events accurately.

IV. RESULTS AND DISCUSSION

It reviews the outcome of applying a proposed model on auto insurance claims using methods-based ML and data analysis. I performed experiments using Python 3.9, TensorFlow 2.x and Scikit-Learn 1.2 on a 64-bit Ubuntu 20.04 system equipped with an eight-core AMD Ryzen 7 processor and 32 GB RAM. Table II outlines the results of the InceptionV3 model's prediction on auto insurance claims. According to the results, the model performs very well in predicting outcomes consistently.

This suggests that InceptionV3 recognizes elaborate patterns in the data, supporting dependable and precise predictions for claims. Being able to perform well in many aspects demonstrates that the model is fit for use in data-driven insurance applications.

Table 2: Performance Metrics of InceptionV3 Model for auto insurance claims

Performance Metric	InceptionV3 model
Accuracy	97
Precision	96
Recall	95
F1-score	95

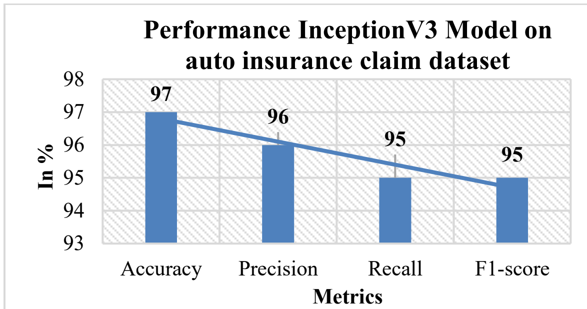

Fig 5. Performance InceptionV3 Model for forecasting auto insurance claims

Figure 5 illustrates the various performance measures of the InceptionV3 model when it was used for forecasting auto insurance claims. Its performance accuracy is 97%, As for the F1-score, it is 95%, 95%, and 96%, respectively. This reliability and balance are essential for boosting operational productivity in the auto insurance industry.

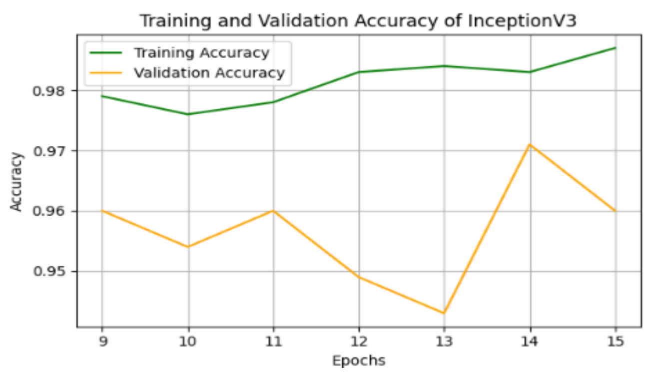

Fig 6. Training and Validation Accuracy of InceptionV3 model

As seen in Figure 6, InceptionV3's accuracy during training stays above 97% for epochs 9 to 15. The model's validation accuracy is between 94% and 97% which suggests it has low risk of overfitting. This points to the accuracy of the model in making auto insurance claim forecasts.

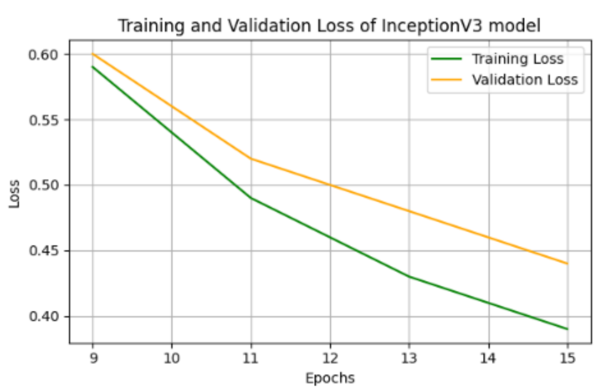

Fig 7. Training and Validation loss of InceptionV3 Model

Figure 7 illustrates how the InceptionV3 model's loss in training and validation The drops between epochs 9 and 15. This suggests strong generalization and efficient learning, demonstrating the model’s suitability for accurately forecasting auto insurance claims.

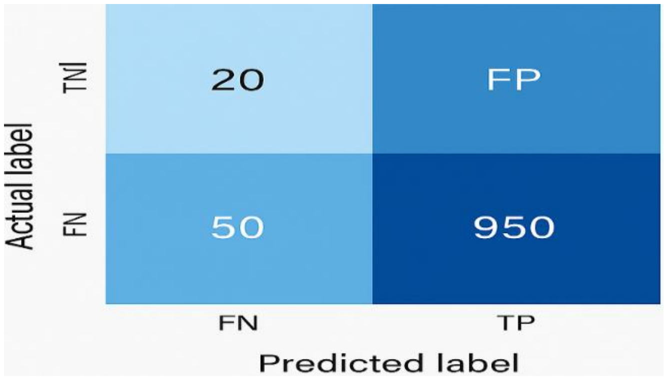

Fig 8. Confusion matrix of InceptionV3

A confusion matrix for the InceptionV3 model used to anticipate vehicle insurance claims is shown in Figure 8. The model's ability to forecast claim outcomes is seen in the matrix. Twenty TN, fifty FN, 950 TP, and twenty FP are among the values.

A) Comparative Analysis

A comparison of ML and DL models used for car insurance claim forecasting is provided in this section. Table III compares how accurate several algorithms are in making predictions. InceptionV3 came out on top, showing accuracy of 97%, proving it can extract important elements needed for claim prediction. Using LR, the model was correct 87.1% of the time, indicating that it performs reasonably well using a simple structure. While XGBoost is strong in various uses, in this instance it showed a weaker performance, with an accuracy of 83.97% which indicates that it may improve if more work is done on it or the features are changed.

Table 3: Comparison of Auto Insurance Claim Dataset Performance Using ML and DL Models

Models	**Accuracy**
InceptionV3	97
LR [16]	87.1
XGBoost [17]	83.97

The framework suggests several strengths in effectively forecasting auto insurance claims with the help of ML and data. InceptionV3, the model achieves 97% accuracy, surpassing the results of LR and XGBoost. Better data quality is achieved by using initial steps that deal with missing values and remove unnecessary elements. Converting categorical variables using label encoding is vital for this framework which makes it suitable for adoption by insurance companies for better early risk detection, fraud control, processed claims and more satisfied customers.

V. CONCLUSION AND FUTURE SCOPE

It is necessary to have models that are designed to handle different and complicated situations for accurate auto insurance claim predictions. The study designed and applied a solid DL model, InceptionV3, to predict auto insurance claims. Detailed preprocessing of data, encoding labels correctly and dividing the model's generalization was enhanced by dividing the data into training and validation groups. From the experiments, it was clear that InceptionV3 performed much better than LR and XGBoost, offering a predictive accuracy of 97%, along with precision, recall and F1-scores at 96%, 95% and 95% respectively. Both results demonstrate that the model can learn about complex and nonlinear links in insurance data, making it an appropriate fit for use in real auto insurance systems. Thanks to the model, identifying elements of fraud and estimating the chances of a specific claim provide valuable support for understanding risks, dealing with fraud and making key decisions for operations.

In the future, we plan to enhance the understanding of models by using explainable AI techniques such as SHAP and LIME, helping insurance practitioners to trust them. Instead of training from scratch, it is also possible to transfer the model to other datasets from various insurance providers. Changes will be applied in real time using edge computing or cloud-based APIs to ensure that insurance services can support both growth and fast processing.

VI. REFERENCES

[1] B. Itri, Y. Mohamed, Q. Mohammed, and B. Omar, "Performance comparative study of machine learning algorithms for automobile insurance fraud detection," in 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS), 2019, pp. 1–4. doi: 10.1109/ICDS47004.2019.8942277.

[2] M. Ayuso, M. Guillen, and J. P. Nielsen, "Improving automobile insurance ratemaking using telematics: incorporating mileage and driver behaviour data," Transportation (Amst)., vol. 46, pp. 735–752, 2019.

[3] C. Oham, R. Jurdak, S. S. Kanhere, A. Dorri, and S. Jha, "B-FICA: BlockChain based Framework for Auto-Insurance Claim and Adjudication," in 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), IEEE, Jul. 2018, pp. 1171–1180. doi: 10.1109/Cybermatics_2018.2018.00210.

[4] L. Weerasinghe, "A comparative study of data mining algorithms in the prediction of auto insurance claims," 2016.

[5] S. Singamsetty, "Fuzzy-Optimized Lightweight Cyber-Attack Detection For Secure Edge-Based Iot," J. Crit. Rev., vol. 6, no. 07, pp. 1028–1033, 2019, doi: 10.53555/jcr.v6:i7.13156.

[6] P. Li, B. Shen, and W. Dong, "An anti-fraud system for car insurance claim based on visual evidence," arXiv Prepr. arXiv1804.11207, 2018.

[7] R. Singh, M. P. Ayyar, T. V. Sri Pavan, S. Gosain, and R. R. Shah, "Automating Car Insurance Claims Using Deep Learning Techniques," in 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM), 2019, pp. 199–207. doi: 10.1109/BigMM.2019.00-25.

[8] S. Lemma, "Factors Affecting Motor Insurance Claim Processing Time: The Case Of Awash Insurance Company Sc," st. mary's University, 2019.

[9] R. M. Nowak, Ł. Neumann, W. Franus, M. Dąmbski, A. Smółkowski, and P. Zawistowski, "Machine learning models for predicting customer decision in motor claims settlement," in Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2019, R. S. Romaniuk and M. Linczuk, Eds., SPIE, Nov. 2019, p. 48. doi: 10.1117/12.2536523.

[10] Z. Duan, Y. Chang, Q. Wang, T. Chen, and Q. Zhao, "A Logistic Regression Based Auto Insurance Rate-Making Model Designed for the Insurance Rate Reform," Int. J. Financ. Stud., vol. 6, no. 1, p. 18, Feb. 2018, doi: 10.3390/ijfs6010018.

[11] Y. B. Gessese, "The effect of motor insurance claim management on customer satisfaction at Ethiopian insurance corporation," St. Mary's University, 2018.

[12] A. Anggraini, "The Analysis of Premium Strategy and Vehicle Claims in Increasing Revenues of PT Asuransi Sinar Mas Tangerang," Sci. J. Reflect. Econ. Accounting, Manag. Bus., vol. 1, no. 3, pp. 251–260, 2018.

[13] V. Vijayan and E. Sherly, "Real time detection system of driver drowsiness based on representation learning using deep neural networks," J. Intell. Fuzzy Syst., vol. 36, no. 3, pp. 1977–1985, Mar. 2019, doi: 10.3233/JIFS-169909.

[14] R. Bhowmik, "Detecting auto insurance fraud by data mining techniques," J. Emerg. Trends Comput. Inf. Sci., vol. 2, no. 4, pp. 156–162, 2011.

[15] M. Eling and M. Luhnen, "Frontier efficiency methodologies to measure performance in the insurance industry: Overview, systematization, and recent developments," Geneva Pap. Risk Insur. Pract., vol. 35, pp. 217–265, 2010.

[16] H. Moon, Y. Pu, and C. Ceglia, "A predictive modeling for detecting fraudulent automobile insurance claims," Theor. Econ. Lett., vol. 9, no. 06, p. 1886, 2019.

[17] J. Pesantez-Narvaez, M. Guillen, and M. Alcañiz, "Predicting Motor Insurance Claims Using Telematics Data XGBoost versus Logistic Regression," Risks, vol. 7, no. 2, p. 70, Jun. 2019, doi: 10.3390/risks7020070.

[18] J. Zhang, Y. Zhu, X. Zhang, M. Ye, and J. Yang, "Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas," J. Hydrol., vol. 561, pp. 918–929, Jun. 2018, doi: 10.1016/j.jhydrol.2018.04.065.

[19] L. J. Cao and F. E. H. Tay, "Feature selection for support vector machines in financial time series forecasting," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 1983, pp. 268–273, 2000, doi: 10.1007/3-540-44491-2_38.

[20] M. Khashei and M. Bijari, "An artificial neural network (p, d, q) model for timeseries forecasting," Expert Syst. Appl., vol. 37, no. 1, pp. 479–489, 2010, doi: 10.1016/j.eswa.2009.05.044.

[21] Chinta, P. C. R., & Karaka, L. M. (2020). Agentic AI and Reinforcement Learning: Towards More Autonomous and Adaptive AI Systems.

[22] Kalla, D., Kuraku, D. S., & Samaah, F. (2021). Enhancing cyber security by predicting malwares using supervised machine learning models. International Journal of Computing and Artificial Intelligence, 2(2), 55-62.

[23] Katari, A., & Kalla, D. (2021). Cost Optimization in Cloud-Based Financial Data Lakes: Techniques and Case Studies. ESP Journal of Engineering & Technology Advancements (ESP-JETA), 1(1), 150-157.

[24] Kalla, D., Smith, N., Samaah, F., & Polimetla, K. (2021). Facial Emotion and Sentiment Detection Using Convolutional Neural Network. Indian Journal of Artificial Intelligence Research (INDJAIR), 1(1), 1-13.

[25] Routhu, K., Bodepudi, V., Jha, K. M., & Chinta, P. C. R. (2020). A Deep Learning Architectures for Enhancing Cyber Security Protocols in Big Data Integrated ERP Systems. Available at SSRN 5102662.

[26] Kuraku, S., & Kalla, D. (2020). Emotet malware a banking credentials stealer. Iosr J. Comput. Eng, 22, 31-41.

[27] Kalla, D., & Samiuddin, V. (2020). Chatbot for medical treatment using NLTK Lib. IOSR J. Comput. Eng, 22, 12.

[28] Routhu, K., & Jha, K. M. (2021). Leveraging Big Data Analytics and Machine Learning Techniques for Sentiment Analysis of Amazon Product Reviews in Business Insights. Available at SSRN 5106490.

[29] Chinta, P. C. R., & Katnapally, N. (2021). Neural Network-Based Risk Assessment for Cybersecurity in Big Data-Oriented ERP Infrastructures. Neural Network-Based Risk Assessment for Cybersecurity in Big Data-Oriented ERP Infrastructures.

[30] Karaka, L. M. (2021). Optimising Product Enhancements Strategic Approaches to Managing Complexity. Available at SSRN 5147875.

[31] Boppana, S. B., Moore, C. S., Bodepudi, V., Jha, K. M., Maka, S. R., & Sadaram, G. AI And ML Applications In Big Data Analytics: Transforming ERP Security Models For Modern Enterprises