IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 3, July 2023

A Machine and Deep Learning Framework for Robust Health Insurance Fraud Detection and Prevention

Suhag Pandya
Independent Researcher
spandya5886@ucumberlands.edu

Abstract*: Healthcare fraud is the deliberate submission of false information or the fabrication of facts in order to get entitlement payments. As a result, it wastes healthcare funds and raises healthcare expenses. For both insurance firms and consumers, predicting health insurance prices is an essential undertaking. The purpose of this research is to examine the feasibility of using ML models for accurate identification of medical insurance fraud. Using a dataset with more than 1300 entries and important characteristics such charges, smoking status, geography, BMI, age, sex, and children, this research investigates the use of ANN for strong health insurance fraud detection. Traditional models like Ridge, Lasso, and XGBoost fared poorly when compared to the ANN, which achieved a R² of 92.72 and a low RMSE of 0.27, according to error measures utilised to assess a model's performance. A findings show that an ANN is good at identifying fraudulent health insurance claims, which bodes well for its future use in better fraud prevention systems. Limitations include the dataset's small size and limited features, suggesting that future studies should expand the dataset and explore more advanced ML techniques for further optimisation*

Keywords: Healthcare, Insurance Fraud, Machine Learning, Fraud Detection, Predictive Modeling, Claims Processing

I. INTRODUCTION

Individual and societal health are fundamentally dependent on healthcare systems. However, as these systems grow more complex, they also become vulnerable to fraud and misuse, posing significant challenges to their sustainability[1]. Fraud in healthcare systems not only undermines the quality of care but also leads to substantial financial losses. Among the various types of fraud, health insurance fraud is particularly concerning due to its widespread impact on both private and public funds[2].

Health insurance, a contractual agreement between insurers and subscribers, is designed to provide financial coverage for healthcare expenses[3]. It includes compensation for losses stemming from accidents, medical treatments, and other health-related events[4]. While health insurance serves as a crucial safety net, its complexity and vast operational scale make it susceptible to fraudulent activities. Health insurance fraud typically falls into two categories: consumer-related fraud, such as submitting false claims or misusing medical benefits, and provider-related fraud, including incorrect billing for unrendered services and unbundling of medical procedures[1][5].

The detection and prevention of health insurance fraud have become paramount for safeguarding financial resources and ensuring the equitable delivery of healthcare services[6][6]. The ever-increasing complexity and amount of healthcare data makes it difficult for traditional fraud detection methods to keep up. This has led to the adoption of AI [7][8]and ML techniques, which offer transformative capabilities in analysing large datasets and uncovering hidden patterns indicative of fraudulent activities [9].

Machine learning, with its ability to learn and improve from experience, enables systems to predict fraudulent behaviors with high precision[10][11]. Techniques fraud detection in health insurance. Furthermore, deep learning, an advanced subset of machine learning, leverages hierarchical data representations to enhance detection accuracy and scalability [12][13]. These technologies not only improve the efficiency of fraud detection systems but also contribute to a more secure and reliable healthcare ecosystem [14][15].

DOI: 10.48175/IJARSCT-14000T

A. Aim and Contribution of Study

This study's main goal is to create a robust framework for detecting and preventing health insurance fraud using advanced machine learning techniques. By leveraging predictive models. This study contributes to the field of health insurance fraud detection are listed in below:

- Collect and use health insurance datasets to predict insurance fraud.
- High-quality analysis data is guaranteed by advanced preparation procedures, which include managing missing values and feature scaling.
- Feature extraction, focusing on influential variables like age, BMI, smoking status, and charges to enhance model accuracy.
- Explores various regression models, including Ridge, Lasso, XGBoost, and ANN, for predicting health insurance premiums fraudulent claims.
- Employs multiple error metrics—RMSE, MAE, MSE, and Adjusted R-Squared—to assess the performance of the models.

B. Structure of paper

What follows is an outline of the rest of the paper**. Section II** requires a survey of the research on the topic of health insurance fraud prediction software**. Section III** explains the process and procedures, while **Section IV** discusses and analyses the findings. The findings and future directions of the study are detailed in **Section V**.

II. LITERATURE REVIEW

ML algorithms for health insurance premium prediction remain an unexplored and unimproved area in the healthcare industry. This section offers some earlier research on machine learning-based health insurance.

This study, Ataabadi et al. (2022) suggests using a method based on ML to forecast the expense of claims by looking at past data from similar patients and identifying claims that are out of the ordinary or fraudulent. The RASA online platform, used for supplemental insurance by well-known firms like as Day, uses a real-world private dataset to assess 700,000 claims. Absolute error for deduction rate dropped from 35 to 23 with the use of the suggested data sampling method in rare instances. According to the assessment, there are around 0.5 percent of the dataset's instances that have an absolute inaccuracy of more than 20%. A lower or higher range might be used to alter the anomalous rates [16].

In this work, Amponsah, Adekoya and Weyori (2022) propose a system that detects and prevents healthcare fraud, especially in the claims processing sector, by using blockchain technology and ML approaches. The data is then entered into a smart contract on the Ethereum blockchain with the purpose of detecting and preventing healthcare fraud. Data from comparative tests shows that the best tool achieves a sensitivity level of 98.09% and a classification accuracy of 97.96%. Consequently, the proposed fix enhances the blockchain smart contract's fraud detection accuracy rate to 97.96% [17].

In this study, Saldamli et al. (2020) the purpose of providing health insurance companies with confusing or incorrect information is to deceive them into paying out for policyholders' bogus claims. Additionally, a single policyholder may seek compensation from more than one insurance company. The NHCAA estimates that there is a yearly financial loss of billions of dollars. It is critical to construct a system that can safely administer and oversee insurance operations by combining data from all insurance providers if they are to stop health insurance fraud. Since data stored and distributed on blockchain cannot be undone, they propose using it to solve the problem of health insurance fraud[18].

This study, Rayan, (2019) offers a combination of domain knowledge (Rule Engine), supervised learning (Decision Trees & Averaged Perceptron), and unsupervised learning (outlier analysis, k-means Clustering) to identify false claims in a dataset. A weighted priority queue of pending claims is sent to the investigation team, outlining a claims most likely to be fraudulent along with notes for both proactive and retrospective review. A 209.4 percent improvement in hit rate is seen in our first case study with a single insurance [19].

This study, Akbar et al. (2020) aims to examine, with the use of cutting-edge methods, statistical modelling approaches to the evaluation of fabricated health benefits. Following data collection and exploratory data analysis, the most effective models may be determined using algorithms like XGB for tree classification and random forest regression. Comparatively, the RF method achieved 81% accuracy with class 1 recall, but the XGB Tree method of random sub-

DOI: 10.48175/IJARSCT-14000T

sampling achieved 86% overall accuracy and 87% with fraudulent providers. Based on the findings, the XGB approach outperforms the others when dealing with clean, fine-tuned data[20].

Table I lists the methods, datasets, main results, and limitations of a number of research that have been conducted on the topic of health insurance fraud detection. It emphasises the advancements in machine learning and blockchain technologies while outlining the gaps for future research.

Table: Comparative Analysis of Previous Studies on Health Insurance Fraud Detection using ML methods

Author	Techniques Used	Data	Findings	Limitation/Gap
Ataabadi et al. (2022)	Data Sampling, Machine Learning	700,000 claims from RASA web portal	Reduced absolute error in exceptional cases from 35 to 23; identified ~0.5% abnormal cases.	Limited to supplementary insurance dataset; broader datasets may reveal more comprehensive insights.
Amponsah, Adekoya, Weyori (2022)	Decision Tree, Blockchain Smart Contracts	Claims dataset	Achieved 97.96% classification accuracy; 98.09% sensitivity for fraud detection.	Focuses on integration with blockchain; lacks comparison with non-blockchain approaches.
Saldamli et al. (2020)	Blockchain	Data integrated from multiple insurers	Enhanced data security and monitoring; prevention of financial losses.	Limited to data integration; does not include advanced ML techniques for fraud detection.
Rayan (2019)	Rule Engine, Decision Trees, Averaged Perceptron, k-Means Clustering	Outstanding claims dataset	Increased fraud detection hit rate by 209.4%.	Requires further validation across diverse insurers and datasets.
Akbar et al. (2020)	Random Forest, Extreme Gradient Boost (XGBoost)	Clean and tuned dataset	XGBoost achieved 86% accuracy; 87% for detecting illegitimate providers.	Relies on clean and preprocessed data; performance on raw data needs evaluation.

III. METHODOLOGY

The methodology for Robust Health Insurance Fraud Detection and Prevention involves several key stages. An first step in data collecting is amassing a dataset that comprises more than 1300 entries and seven important characteristics that are used to forecast health insurance premiums: age, sex, region, children, smoking, and charges. Data preprocessing follows, focusing on cleaning the dataset by removing duplicates, handling missing data through imputation, and ensuring accurate and reliable information for analysis.

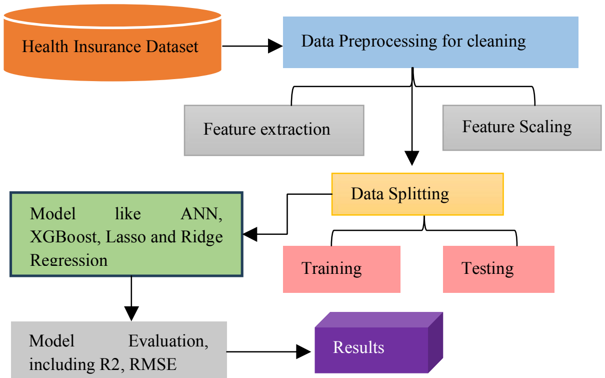
Health Insurance Dataset
Data Preprocessing for cleaning
Feature extraction
Feature Scaling
Data Splitting
Model like ANN, XGBoost, Lasso and Ridge Regression
Training
Testing
Model Evaluation, including R2, RMSE
Results

Fig. 1. Proposed Flowchart for health insurance

Feature extraction is performed to enhance machine learning model performance by selecting influential attributes such as age, BMI, and smoking status. Feature scaling is then applied using interval scaling techniques to normalise the features within a defined range, ensuring consistency across the dataset. To ensure the model is performing as expected, the data is divided into two sets: one for training and one for testing, with a ratio of 80:20. Health insurance cost prediction and fraud detection make use of a number of ML models, including XGBoost, Ridge, and ANN. The models' efficacy is assessed using error metrics like RMSE, MAE, MSE, and Adjusted R-squared. This enables the detection and prevention of fraudulent activities within the health insurance domain.
The steps of the flowchart briefly explained in below:

A. Data Collection

There were more than 1300 records in the dataset organised into 7 columns: charges, smoking, region, children, body mass index (BMI), sex, and age. This dataset was used to forecast the health insurance premium. The analysis and visualisation representation are provide below:

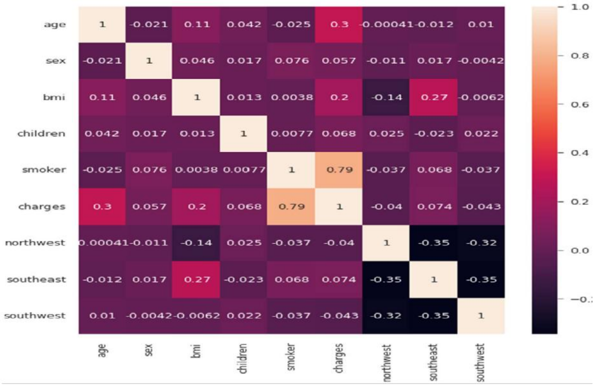

Fig. 2.Correlation Matrix of data

Figure 2 shows correlation matrix heatmap showing the relationship between various variables. With a value between -1 and 1, each heatmap cell represents a different correlation coefficient among two variables. The colour gradient runs from purple (negative correlation) to red (positive correlation). Notable correlations include a strong positive correlation (0.79) between smoker status and charges, suggesting smokers have higher charges, and a moderate positive correlation (0.2) between BMI and charges, indicating higher BMI is associated with higher charges. This heatmap is useful for identifying potential relationships for further statistical analysis.

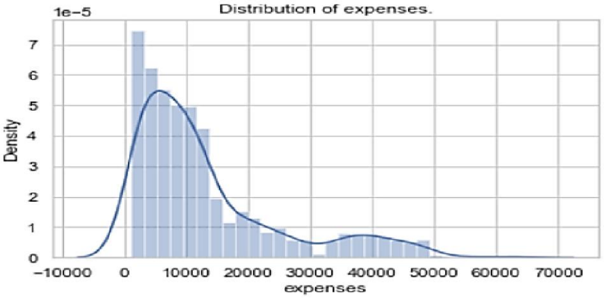

Fig. 3.Density distribution of expenses

The density distribution of expenses in Figure 3 is right-skewed, displaying that most expenses are concentrated towards a lower end, with a few significantly higher values extending the tail to the right. The median likely falls between 10,000 and 20,000, while the mode appears near 5,000, reflecting the most frequent expense range. The distribution spans a broad range, from approximately -10,000 to 70,000, highlighting considerable variability and potential outliers on the higher end, representing unusually large expenses. This pattern suggests a predominance of lower expenses, with occasional high-value transactions or category-specific higher costs contributing to the skewness.

B. Data Preprocessing

Data preparation is the process of preparing raw data for use in data science activities like data mining and ML [21]. Data cleaning and preparation include removing duplicates to make sure the data is correct and dependable for analysis or modelling [22][23]. There can be a lot of missing or unnecessary information in the data [24]. This aspect is handled by cleansing the data. Filling in blanks with the attribute mean or most likely value is part of dealing with missing data.

C. Feature Extraction

The purpose of feature engineering in ML is to enhance the efficiency of ML algorithms by the extraction of valuable features from unstructured data using domain expertise [25][26]. The medical insurance cost dataset is dominated by factors such as age, BMI, and smoking status [27].

D. Feature Scaling

To convert a range of characteristics into a range of intervals, the interval technique makes use of boundary information [28][29]. Scaling using two extreme values, the maximum and lowest, is employed by popular interval scaling techniques like [0, 1] [30]. This equation may be written as (1):

$$x_i' = \frac{x_i - \min\,(x_i)}{\max(x_i) - \min\,(x_i)} \qquad (1)$$

Max and min are functions that discover the maximum and lowest values, respectively, and xi is the ith converted eigenvector [31].

E. Data Splitting

A frequent method for validating models is data splitting, which is separating a dataset into two independent sets: training and testing. A ratio of 80:20 divides the dataset.

F. Detection with Artificial Neural Network (ANN) Model

The patterns for processing information are called ANNs, which were created by McCulloch and Pitts by imitating the neural network seen in the human brain. There are three layers in an ANN: input, hidden, and output [32]. As a result of the interconnections between the neurones in each layer, the output of one layer is used as an input by the following layer [33]. A layer's output signal is coupled to the layer below it via signal-strengthening or -weakening interfaces according to the weighting parameters [34]. Hidden and output layer neurones' outputs will be computed with the help of an activation function, like a sigmoid or linear function [35][36]. Input and output layer neurone densities are proportional to the number of input and output variables, respectively [37]. The number of neurones in the subsurface layer is not considered in any predetermined manner. Still, the problem's complexity and the iterative technique decide how many hidden layers there are, along with their neurone counts[38]. A transfer function, sometimes known as an activation function, is used to calculate the output [39][40]. Optimal behaviour for the activation function would be a step-like pattern [41]. Also, common activation functions that meet the criterion of being continuous and derivable at all places are (2), which is needed by most optimisation algorithms nowadays[42][43]:

Hyperbolic tangent sigmoid(tansig): (a)

$$(a) = \frac{e^{a} - e^{-a}}{e^{a} + e^{-a}} \qquad (2)$$

logarithmic sigmoid(logsig):

$$f(a) = \frac{1}{1 + e^{-a}} \qquad (3)$$

pure linear(purelin):

$$f(a) = a \qquad (4)$$

exponential(exponential):

$$f(a) = e^{-a} \qquad (5)$$

sin transfer function(sine):

$$f(a) = \sin(a) \qquad (6)$$

G. Performance metrics

The accuracy or mistakes of ML models may be used to assess their predictive capacity. It is difficult to determine the models' accuracy since regressions are utilised in this study. Consequently, the prediction error is used to measure the performance of a regression model. The intriguing thing about errors is that they reveal the degree to which the predicted values were in line with the anticipated ones [44][45]. This paper use the RMSE, MAE, MSE and Adjusted R Squared that explained below:

R2-Square

As a statistical metric, R-squared score is used to assess the efficacy of regression models [46]. It provides a handy 0–100 scale for gauging the strength of the association between the dependent variable and regression models. It can be found as (7):

$$R^2 = \frac{\text{Variance explained by model}}{\text{Total variance}} \qquad (7)$$

Mean absolute error (MAE)

The MAE takes all the absolute mistakes and finds their average. When calculating the MAE, the absolute value of the difference among the anticipated and actual values is used, as opposed to the RMSE, which results in negative numbers [47]. By avoiding the square root of the errors, the MAE gives an error measure that is unit-independent of the output variable. Here is the formula for calculating the MAE (8):

$$MAE = \frac{\sum_{j=1}^{N} |y_j - \hat{y}_j|}{N} \qquad (8)$$

In this case, N is included in the numerator, which displays the absolute value of the sum of the residuals. A lower MAE, similar to a lower RMSE, suggests a better match.

Root Mean Square Error (RMSE)

RMSE have widely been utilised in evaluating an accuracy of a recommender system, RMSE measures the magnitude of error produced by a model, making it very useful for quantitative assessments in forecasting and regression analysis, given by (9):

$$RMSE = \sqrt{\frac{\sum_{j=1}^{N} (y_i - \hat{y}_i)^2}{N}} \qquad (9)$$

Mean Squared Error (MSE)

In regression assignments in particular, the MSE is a popular statistic for assessing the precision of a prediction model. As a whole, it's a measure of how far off the actual dataset values are from the forecasted values. A lower MSE value indicates better predictive performance, as it reflects smaller deviations from the true values. The mathematical equation for MSE is expressed as (10):

$$MSE = \frac{1}{N} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad (10)$$

Where:

- The data points are represented by n.
- For the i-th data point, the actual value is represented by yi.
- The expected value for the i-th data point is denoted as $\hat{y}_i$.

IV. RESULT AND DISCUSSION

This section covers a simulated outcome for health insurance based on machine learning. The experiment utilised Python programming language, Windows 10, CPU, operating system, etc. A performance of a model is evaluated with error matrices like RMSE and R-square. The following ANN models are compared (see Table III) with existing models such as Ridge[48], Lasso[49], and XGBoost[50], where ANN models outperformance and predicts health insurance are provided in Table II.

Table 2: Findings of the ANN Model for Health Insurance Fraud Detection

Performance Metrics	Artificial Neural Network (ANN)
R2	92.72
RMSE	0.27
MSE	0.07
MAE	0.14

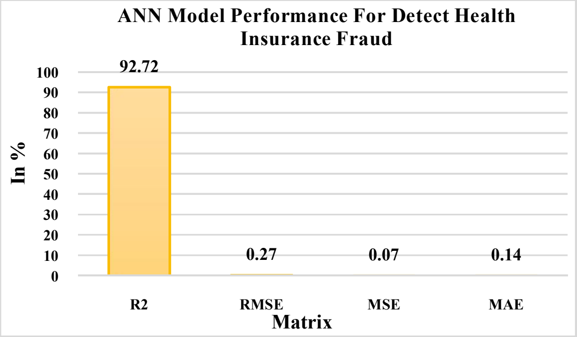

Fig. 4. ANN Model Performance on Insurance Dataset

Table II and Figure 4 highlight the performance metrics of the ANN model for health insurance fraud detection. The model demonstrates a robust predictive capability, achieving an R² score of 92.72, indicating a high proportion of variance in the data being accurately captured. Additionally, the model exhibits lower error rates, with an RMSE0.27, an MSE0.07, and an MAE0.14, reflecting its precision in minimising prediction errors. These results demonstrate that the ANN model is capable of detecting health insurance claim fraud.

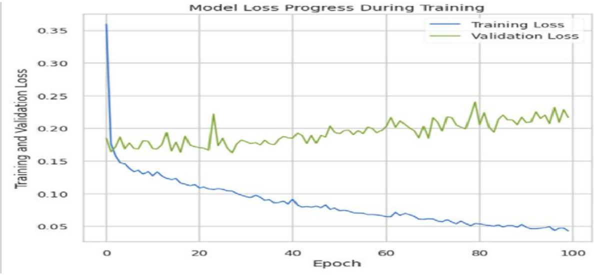

Fig. 5. Training and validation loss of ANN model

Figure 5 shows the training and validation loss of a model over 100 epochs. The graph features two lines: a blue line displays the training loss, which steadily decreases, starting from a high point near 0.35 and reducing to approximately 0.05. The green line represents the validation loss, which exhibits some fluctuations but generally trends around 0.20. The loss is measured from 0 to 0.35 on the y-axis, while epochs are labelled on the x-axis from 0 to 100.

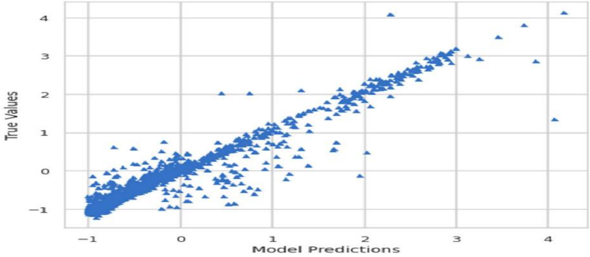

Fig. 6. ANN Model predictions vs. true values.

Figure 6 illustrates a relationship between model predictions and true values for a regression model. The x-axis is labelled "Model Predictions" and ranges from approximately -1 to 4, while the y-axis is labelled "True Values" and also ranges from about -1 to 4. The data points, represented by blue triangles, are scattered across the graph but form a generally diagonal line, shows a positive correlation among the predicted and true values. This suggests that model predictions align reasonably well with the true values, with increasing alignment as the values increase. The distribution of data points suggests the model performs consistently across the range of values but may have some variance or error, particularly noticeable where points stray from the diagonal line.

Table 3: Comparison of Model Performance for Health Insurance Fraud Detection

Model	R2	RMSE
Ridge	78.38	46.52
Lasso	79.78	56.71
XGBoost	86.81	44.50
ANN	92.72	0.27

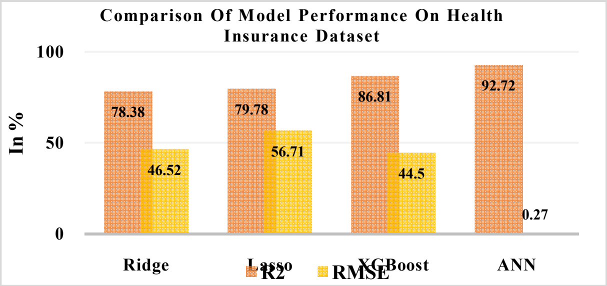

Fig. 7. Comparison of Model Performance

Table III presents a comparative analysis of different models used for health insurance fraud detection, showcasing their performance based on R² and RMSE metrics. Among the models, the ANN outperforms the others with an exceptional R² score of 92.72 and a significantly low RMSE of 0.27, indicating superior predictive accuracy and minimal error. In contrast, XGBoost achieves an R² of 86.81 with a higher RMSE of 44.50, making it the second-best performer. Lasso and Ridge regression models exhibit comparatively lower performance, with R² scores of 79.78 and 78.38 and RMSE values of 56.71 and 46.52, respectively. These results highlight the ANN model's dominance in accurately detecting health insurance fraud.

V. CONCLUSION & FUTURE WORK

A person's health insurance premiums are among their most substantial yearly outlays. A third of GDP goes into health insurance, and everyone has different levels of medical care requirements. Healthcare expenses fluctuate every year due to a multitude of causes, including but not limited to changes in medical, pharmaceutical trends, and political considerations. Predicting future health insurance rates using regression methods is the focus of this paper. The ANN model demonstrated superior performance in detecting health insurance fraud, achieving an R² score of 92.72 and low error metrics, such as RMSE (0.27), MSE (0.07), and MAE (0.14), outperforming models like Ridge, Lasso, and XGBoost. The research may not have captured the intricacy of health insurance fraud across varied demographics due to the very small dataset of 1300 items. The limited set of features, such as charges, smoking, and BMI, may also restrict the model's ability to identify more subtle fraudulent behaviors. To improve forecast accuracy and generalisability across various insurance scenarios, future research should centre on growing the dataset, adding more features, and investigating sophisticated ML approaches like DL models or ensemble methods.

REFERENCES

[1] B. Patel, V. K. Yarlagadda, N. Dhameliya, K. Mullangi, and S. C. R. Vennapusa, "Advancements in 5G Technology: Enhancing Connectivity and Performance in Communication Engineering," *Eng. Int.*, vol. 10, no. 2, pp. 117–130, 2022, doi: 10.18034/ei.v10i2.715.

[2] I. Matloob, S. A. Khan, and H. U. Rahman, "Sequence mining and prediction-based healthcare fraud detection methodology," *IEEE Access*, vol. 8, pp. 143256–143273, 2020.

[3] R. Arora, S. Gera, and M. Saxena, "Mitigating Security Risks on Privacy of Sensitive Data used in Cloud-based ERP Applications," in *2021 8th International Conference on Computing for Sustainable Global Development (INDIACom)*, 2021, pp. 458–463.

[4] V. K. Yarlagadda, “Harnessing Biomedical Signals: A Modern Fusion of Hadoop Infrastructure, AI, and Fuzzy Logic in Healthcare,” *Malaysian J. Med. Biol. Res.*, vol. 8, no. 2, 2021.

[5] R. Arora, S. Gera, and M. Saxena, “Impact of Cloud Computing Services and Application in Healthcare Sector and to provide improved quality patient care,” *IEEE Int. Conf. Cloud Comput. Emerg. Mark. (CCEM), NJ, USA, 2021*, pp. 45–47, 2021.

[6] K. V. V. and S. G. Jubin Thomas , Piyush Patidar, “An analysis of predictive maintenance strategies in supply chain management,” *Int. J. Sci. Res. Arch.*, vol. 06, no. 01, pp. 308–317, 2022, doi: DOI: https://doi.org/10.30574/ijsra.2022.6.1.0144.

[7] P. Khare and S. Srivastava, “The Impact of AI on Product Management : A Systematic Review and Future Trends,” vol. 9, no. 4, 2022.

[8] R. Bishukarma, “Adaptive AI-Based Anomaly Detection Framework for SaaS Platform Security,” *Int. J. Curr. Eng. Technol.*, vol. 12, no. 07, pp. 541–548, 2022, doi: https://doi.org/10.14741/ijcet/v.12.6.8.

[9] J. Thomas, H. Volikatla, V. V. R. Indugu, K. Gondi, and D. S. Gondi, “Machine Learning Approaches for Fraud Detection in E-commerce Supply Chains,” *Innov. Comput. Sci. J.*, vol. 8, no. 1, 2022.

[10] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” *SN Comput. Sci.*, 2021, doi: 10.1007/s42979-021-00592-x.

[11] S. K. R. Anumandla, V. K. Yarlagadda, S. C. R. Vennapusa, and K. R. V Kothapalli, “Unveiling the Influence of Artificial Intelligence on Resource Management and Sustainable Development: A Comprehensive Investigation,” *Technol. \& Manag. Rev.*, vol. 5, no. 1, pp. 45–65, 2020.

[12] R. Goyal, “The Role Of Business Analysts In Information Management Projects,” *Int. J. Core Eng. Manag.*, vol. 6, no. 9, pp. 76–86, 2020.

[13] A. Goyal, “Scaling Agile Practices with Quantum Computing for Multi-Vendor Engineering Solutions in Global Markets,” *Int. J. Curr. Eng. Technol.*, vol. 12, no. 06, 2022, doi: : https://doi.org/10.14741/ijcet/v.12.6.10.

[14] K. Kapadiya *et al.*, “Blockchain and AI-empowered healthcare insurance fraud detection: an analysis, architecture, and future prospects,” *IEEE Access*, vol. 10, pp. 79606–79627, 2022.

[15] V. K. Yarlagadda and R. Pydipalli, “Secure Programming with SAS: Mitigating Risks and Protecting Data Integrity,” *Eng. Int.*, vol. 6, no. 2, pp. 211–222, Dec. 2018, doi: 10.18034/ei.v6i2.709.

[16] P. E. Ataabadi, B. S. Neysiani, M. Z. Nogorani, and N. Mehraby, “Semi-Supervised Medical Insurance Fraud Detection by Predicting Indirect Reductions Rate using Machine Learning Generalization Capability,” in *2022 8th International Conference on Web Research, ICWR 2022*, 2022. doi: 10.1109/ICWR54782.2022.9786251.

[17] A. A. Amponsah, A. F. Adekoya, and B. A. Weyori, “A novel fraud detection and prevention method for healthcare claim processing using machine learning and blockchain technology,” *Decis. Anal. J.*, 2022, doi: 10.1016/j.dajour.2022.100122.

[18] G. Saldamli, V. Reddy, K. S. Bojja, M. K. Gururaja, Y. Doddaveerappa, and L. Tawalbeh, “Health Care Insurance Fraud Detection Using Blockchain,” in *2020 Seventh International Conference on Software Defined Systems (SDS)*, 2020, pp. 145–152. doi: 10.1109/SDS49854.2020.9143900.

[19] N. Rayan, “Framework for analysis and detection of fraud in health insurance,” in *Proceedings of 2019 6th IEEE International Conference on Cloud Computing and Intelligence Systems, CCIS 2019*, 2019. doi: 10.1109/CCIS48116.2019.9073700.

[20] N. A. Akbar, A. Sunyoto, M. Rudyanto Arief, and W. Caesarendra, “Improvement of decision tree classifier accuracy for healthcare insurance fraud prediction by using Extreme Gradient Boosting algorithm,” in *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, 2020, pp. 110–114. doi: 10.1109/ICIMCIS51567.2020.9354286.

[21] R. K. Arora, A. Soni, R. Garine, and A. Kumar, “Impact of Cloud-based Mobile Application during Pandemic (Covid-19),” *SSRN*, 2022.

[22] S. A. and A. Tewari, “AI-Driven Resilience: Enhancing Critical Infrastructure with Edge Computing,” *Int. J. Curr. Eng. Technol.*, vol. 12, no. 02, pp. 151–157, 2022, doi: https://doi.org/10.14741/ijcet/v.12.2.9.

[23] Mani Gopalsamy, “An Optimal Artificial Intelligence (AI) technique for cybersecurity threat detection in IoT

Networks," *Int. J. Sci. Res. Arch.*, vol. 7, no. 2, pp. 661–671, Dec. 2022, doi: 10.30574/ijsra.2022.7.2.0235.

[24] B. Boddu, "Ensuring Data Integrity and Privacy: A Guide for Database Administrators," *https://www.ijfmr.com/research-paper.php?id=10880*, vol. 4, no. 6, p. 6, 2022.

[25] A. P. A. Singh, "STRATEGIC APPROACHES TO MATERIALS DATA COLLECTION AND INVENTORY MANAGEMENT," *Int. J. Bus. Quant. Econ. Appl. Manag. Res.*, vol. 7, no. 5, 2022.

[26] R. Goyal, "Software Development Life Cycle Models: A Review Of Their Impact On Project Management," *Int. J. Core Eng. Manag.*, vol. 7, no. 2, pp. 78–87, 2022.

[27] B. Boddu, "Serverless Databases Are the Future of Database Management," *https://jsaer.com/download/vol-6-iss-1-2019/JSAER2019-6-1-277-282.pdf*, vol. 6, no. 1, p. 5, 2020.

[28] A. P. A. Singh, "Streamlining Purchase Requisitions and Orders : A Guide to Effective Goods Receipt Management," *J. Emerg. Technol. Innov. Res.*, vol. 8, no. 5, pp. g179–g184, 2021.

[29] K. Patel, "Quality Assurance In The Age Of Data Analytics: Innovations And Challenges," *Int. J. Creat. Res. Thoughts*, vol. 9, no. 12, pp. f573–f578, 2021.

[30] S. Bauskar, "Predictive Analytics For Sales Forecasting In Enterprise Resource," *Int. Res. J. Mod. Eng. Technol. Sci.*, vol. 04, no. 06, pp. 4607–4618, 2022, doi: https://www.doi.org/10.56726/IRJMETS26271.

[31] M. Gopalsamy, "Advanced Cybersecurity in Cloud Via Employing AI Techniques for Effective Intrusion Detection," *Int. J. Res. Anal. Rev.*, vol. 8, no. 01, pp. 187–193, 2021.

[32] J. Thomas, K. V. Vedi, and S. Gupta, "Enhancing Supply Chain Resilience Through Cloud-Based SCM and Advanced Machine Learning: A Case Study of Logistics," *J. Emerg. Technol. Innov. Res.*, vol. 8, no. 9, 2021.

[33] S. R. Bauskar and S. Clarita, "Evaluation of Deep Learning for the Diagnosis of Leukemia Blood Cancer," *Int. J. Adv. Res. Eng. Technol.*, vol. 11, no. 3, pp. 661–672, 2020, doi: https://iaeme.com/Home/issue/IJARET?Volume=11&Issue=3.

[34] V. V. Kumar, A. Sahoo, and F. W. Liou, "Cyber-enabled product lifecycle management: A multi-agent framework," in *Procedia Manufacturing*, 2019. doi: 10.1016/j.promfg.2020.01.247.

[35] Mani Gopalsamy, "Enhanced Cybersecurity for Network Intrusion Detection System Based Artificial Intelligence (AI) Techniques," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 12, no. 01, pp. 671–681, Dec. 2021, doi: 10.48175/IJARSCT-2269M.

[36] A. Goyal, "Enhancing Engineering Project Efficiency through Cross-Functional Collaboration and IoT Integration," *Int. J. Res. Anal. Rev.*, vol. 8, no. 4, pp. 396–402, 2021.

[37] N. Abid, "A Climbing Artificial Intelligence for Threat Identification in Critical Infrastructure Cyber Security," *Int. J. Res. Anal. Rev.*, vol. 9, no. 4, 2022.

[38] E. Mohamad, R. Shirani Faradonbeh, D. Jahed Armaghani, M. Monjezi, and M. Z. Abd Majid, "An optimized ANN model based on genetic algorithm for predicting ripping production," *Neural Comput. Appl.*, vol. 28, 2017, doi: 10.1007/s00521-016-2359-8.

[39] M. A. Shajahan, N. Richardson, N. Dhameliya, B. Patel, S. K. R. Anumandla, and V. K. Yarlagadda, "AUTOSAR Classic vs. AUTOSAR Adaptive: A Comparative Analysis in Stack Development," *Eng. Int.*, vol. 7, no. 2, pp. 161–178, Dec. 2019, doi: 10.18034/ei.v7i2.711.

[40] S. A. and A. Tewari, "Security Vulnerabilities in Edge Computing: A Comprehensive Review," *Int. J. Res. Anal. Rev.*, vol. 9, no. 4, pp. 936–941, 2022.

[41] R. Bishukarma, "The Role of AI in Automated Testing and Monitoring in SaaS Environments," *Int. J. Res. Anal. Rev.*, vol. 8, no. 2, pp. 846–852, 2021.

[42] H. Maouz, L. Khaouane, S. Hanini, Y. Ammi, M. Laidi, and H. Benimam, "The prediction of carbonyl groups during photo-thermal and thermal aging of polymers using artificial neural networks," *Alger. J. Environ. Sci. Technol.*, vol. 6, no. 3, 2020.

[43] S. Bauskar, "BUSINESS ANALYTICS IN ENTERPRISE SYSTEM BASED ON APPLICATION OF ARTIFICIAL INTELLIGENCE," *Int. Res. J. Mod. Eng. Technol. Sci.*, vol. 04, no. 01, pp. 1861–1870, 2022, doi: DOI : https://www.doi.org/10.56726/IRJMETS18127.

[44] V. V. Kumar, F. W. Liou, S. N. Balakrishnan, and V. Kumar, "Economical impact of RFID implementation in remanufacturing: a Chaos-based Interactive Artificial Bee Colony approach," *J. Intell. Manuf.*, 2015, doi:

10.1007/s10845-013-0836-9.

[45] M. Z. Hasan, R. Fink, M. R. Suyambu, M. K. Baskaran, D. James, and J. Gamboa, "Performance evaluation of energy efficient intelligent elevator controllers," in *IEEE International Conference on Electro Information Technology*, 2015. doi: 10.1109/EIT.2015.7293320.

[46] V. Kumar and F. T. S. Chan, "A superiority search and optimisation algorithm to solve RFID and an environmental factor embedded closed loop logistics model," *Int. J. Prod. Res.*, vol. 49, no. 16, 2011, doi: 10.1080/00207543.2010.503201.

[47] S. Bauskar and S. Clarita, "AN ANALYSIS: EARLY DIAGNOSIS AND CLASSIFICATION OF PARKINSON'S DISEASE USING MACHINE LEARNING TECHNIQUES," *Int. J. Comput. Eng. Technol.*, vol. 12, no. 01, pp. 54-66., 2021, doi: 10.5281/zenodo.13836264.

[48] C. A. ul Hassan, J. Iqbal, S. Hussain, H. AlSalman, M. A. A. Mosleh, and S. Sajid Ullah, "A Computational Intelligence Approach for Predicting Medical Insurance Cost," *Math. Probl. Eng.*, 2021, doi: 10.1155/2021/1162553.

[49] Y. A. Christobel and S. Subramanian, "An Empirical Study Of Machine Learning Regression Models to Predict Health Insurance Cost," *Webology*, 2022.

[50] G. K. Patra, C. Kuraku, S. Konkimalla, V. N. Boddapati, and M. Sarisa, "An Analysis and Prediction of Health Insurance Costs Using Machine Learning-Based Regressor Techniques," *J. Comput. Eng. Technol.*, vol. 12, no. 3, pp. 102–113, 2021.

DOI: 10.48175/IJARSCT-14000T