Universal Library of Engineering Technology

Research Article

ISSN: 3064-996X

Open Access | PP: 27-34

DOI: https://doi.org/10.70315/uloap.ulete.2022.004

Leveraging Artificial Intelligence Algorithms for Risk Prediction in Life Insurance Service Industry

Srikanth Reddy Vangala[1], Ram Mohan Polam[2], Bhavana Kamarthapu[3], Ajay Babu Kakani[4], Sri Krishna Kireeti Nandiraju[5], Sandeep Kumar Chundru[6]

[1]University of Bridgeport.
[2]University of Illinois at Springfield.
[3]Fairleigh Dickinson University.
[4]Wright State University.
[5]University of Illinois at Springfield.
[6]University of Central Missouri.

Abstract

Accurate risk prediction is essential in the life insurance industry to enhance underwriting processes and detect fraudulent activities effectively. This paper suggests a model for healthcare insurance fraud detection and risk assessment using an Artificial Neural Network (ANN) applied to the Prudential Insurance dataset. Data cleaning, categorical variable encoding, and resolving class imbalance using the Synthetic Minority Oversampling Technique (SMOTE) are all part of the extensive preparation that the dataset goes through in order to optimize model learning. Our ANN model has been trained and tested using industry-standard regression metrics, and it has achieved impressive results: R^2= 92.7%, MAE= 14.3%, MSE= 72.7, and RMSE= 27. These outcomes prove that the model is strong and can correctly identify complicated nonlinear relationships in the data, which greatly improves the accuracy of risk predictions. The findings suggest that the proposed ANN model is a powerful tool for life insurance providers to optimize underwriting, improve fraud detection accuracy, and deliver personalized insurance products.

Keywords: *Healthcare Insurance, Artificial Neural Network (ANN), Prudential Insurance Dataset, Risk Prediction, Life Insurance.*

INTRODUCTION

There has been very little consumer participation in the insurance market. Insurance companies had the lowest customer interaction rates across all research industries1. Brokers, for example, get an astounding $45 billion in yearly pay from insurers worldwide2, however many insurers have minimal contact with a large section of the end-consumers due to the fact that a lot of their business is intermediary. Also, insurers aren't getting as much data about their customers' wants and how to tailor their goods to those demands since the sector isn't moving quickly enough to digitize. [1]. As a whole, the insurance industry has never had very engaged customers. Customers interact with insurance companies the least of any of the other sectors surveyed.

In this context, AI and related technologies are spreading throughout society and industry [2], as well as starting to be implementedinhealthcare.Provider,payer,andpharmaceutical organizations' administrative operations, along with several facets of patient care, might be revolutionized by these technologies. Numerous studies have demonstrated that AI is equally as capable as humans in performing vital healthcare tasks like diagnosing illnesses. Algorithms can still identify malignant tumors with more accuracy than radiologists, and they can even teach researchers how to create cohorts for costly clinical research. Still, we anticipate a long time before AI can fully supplant humans in general medical procedure sectors, and there are a number of reasons why. Both the opportunities and the challenges for the quick adoption of AI in healthcare are discussed in this article. AI has the ability to automate many parts of the treatment process [3].

In order to safeguard their financial stability, millions of households rely on life insurance. In the US, life insurance firms are responsible for overseeing benefits of trillions of dollars. There are many different kinds of insurance policies, but they all share the underwriting process's calculation of mortality risk at the individual level [4]. Human discretion and point-based systems that treat risk variables separately have traditionally been used for this purpose. Despite their imprecision and lack of consistency, these approaches are adequate for industrial use. Therefore, insurers may only use traditional underwriting to a certain extent when trying to evaluate risk from data and provide products at competitive prices [5].

In the insurance industry, particularly in areas requiring high-dimensional data analysis and predictive modelling. ML algorithms can automatically learn from historical data to identify complex patterns and relationships among variables without being explicitly programmed, making them highly effective for risk classification and fraud detection [6]. DL, a subset of ML that uses ANN with multiple layers, excels at processing unstructured data such as medical records, sensor signals, and textual claim descriptions. These tools make it possible to increase the correctness of risk evaluations, as they deal with the non-linear effects of different types of risk. As opposed to traditional statistics, ML and DL models are flexible [7], can handle large volumes of data, and evolve on their own with more data, giving insurers an advantage in all areas of insurance policy making

Motivation and Contribution of Paper

The motivation behind this study comes from the fact that healthcare insurance fraud is happening more often and is costing the insurance industry a lot of money, which could hurt the system and make it less reliable. Traditional fraud detection methods often can't pick up the complicated and twisting patterns that come with fraud, which often causes too many mistakes on top of some missed detections. With the rapid advancement of ML and DL technologies, there is a good reason to start looking at smarter and data-based ways to spot fraud. Specifically, ANN can quickly pick up patterns and connections in big and varied sets of data by themselves, which makes them useful for helping improve how accurately we spot fraud. This research tries to use the strong points of artificial neural networks to make it easier for life insurance companies to predict risks and find fraud so they can keep customers safe and make the process work better. Contributions of the study are: (1) There are three main things you should think about to improve relationships in a family; (2) the best way to improve family relationships is to be open and honest when you talk, trust each other, and spend time together; and (3) including more members of the family, like children or older relatives, can help relationships get better:

- Developed an ANN model that is able to efficiently and accurately identify healthcare fraud cases.
- Processed the data in multiple ways, like cleaning missing values, transforming categorical variables, and using SMOTE to resolve the imbalance issue.
- Performed trials comparing to standard approaches (PLS-SEM and LR) and it was determined that ANN fare better on measures of performance.
- Applied the model to the real-world data of Prudential Insurance, confirming it can be used to detect fraud in life insurance.

Justification and Novelty

The main contribution of this study is bringing together an ANN model with advanced ways to fix the data, such as SMOTE for fairness and encoding for categories, to create a powerful framework for finding fraud in healthcare insurance. Compared to PLS-SEM and LR, the ANN model presented here works well with difficult nonlinear patterns found in big data sets used in insurance industries. The urgent requirement for more precise and extensible fraud detection systems in the insurance industry, where traditional approaches frequently fall short of keeping up with changing fraud strategies, justifies this research. By achieving a significantly higher R^2 score and lower prediction errors, the study not only validates the effectiveness of ANN but also presents a comprehensive and replicable methodology that can be adopted by insurers to enhance fraud prevention and ensure data-driven decision-making.

Structure of the Paper

The following paper structured: Section II provides a summary of the research on credit scoring. Section III explains the suggested process, Section IV gives a discussion and comparison of the experimental outcomes. Finally, Section V outlines the paper's weaknesses and prospective research goals in its conclusion.

LITERATURE REVIEW

This section reviews key literature and provides an overview of the milestones, methodologies, with life insurance service industry, the methods and datasets used, key findings and challenges.

Shih, William and Chang (2019) stated, they found other possible ailments in addition to those that have already been discussed in the literature, such diabetes and hypertension. Our suggested approaches achieve an average first analysis accuracy of 86%, compared to a standard Neural Network algorithm's 49.5% using the same training data [8].

Mishra and Reddy (2018) the telecom sector used ensemble-based classifiers for churn prediction, specifically Bagging, Boosting, and RF. The well-known classifiers, DT, NB Classifier, and SVM, were pitted against the classifiers based on ensembles. Experimentally, RF works better than other methods with a greater accuracy of 91.66 percent, a lower error rate, a lower specificity, and a higher sensitivity [9].

Kaewkiriya (2017) The objective is to introduce a framework for predicting life insurance policyholders that is based on many algorithms. Three components make up the framework. The first one is the module for preparing data. The second one is a module for cleansing data. The third one is a module for extracting data. There are three distinct phases of data extraction. Choose features as a starting point. The best characteristic for data analysis is chosen in this stage. In the second phase, the K-mean approach is used to cluster the data. Finally, a recommendation model is built by extracting data using a neural network technique. Using a combination of algorithms yielded the best prediction accuracy (92.83%), according to the comparative findings [10].

Almasi et al. (2017) reveals obstacles that must be overcome for test generating tools to enhance fault detection. Classification of the undiscovered errors reveals that nearly all of them (97.62%) are the development of "complex state configuration of objects" (47.62%) or "specific primitive values" (50.00%) are both necessary [11].

Peters and Maxemchuk (2017) Compare the distributed application's performance to that of a centralized one by calculating its Shannon entropy. This metric measures the amount of information kept in the system as well as the amount of information revealed when components are compromised. With this distributed strategy, the clearing house's entropy may be reduced by about 79% in the case of an assault [12].

Table I summarizes the literature review on risk prediction in life insurance, including the methodology, dataset, main conclusions, limitations

Table I. Literature review summary of risk prediction in life insurance and healthcare

Study	Methodology	Dataset	Key Findings	Limitations	Future Directions
Shih, William and Chang (2019)	Proposed method vs. traditional Neural Network	Medical datasets with diseases like hypertension, diabetes	Proposed method achieved 86% accuracy vs. 49.5% by traditional NN; discovered additional potential diseases	Lack of detail on model architecture and disease types	Apply method to broader disease categories and larger datasets
Mishra and Reddy (2018)	SVM, Decision Tree, and Naïve Bayes classifiers against ensemble-based classifiers (Bagging, Boosting, Random Forest)	Telecom industry customer churn dataset	Random Forest outperformed others with 91.66% accuracy, low error, high sensitivity, and low specificity	No insight into feature importance or scalability	Explore model scalability and integrate real-time churn prediction
Kaewkiriya (2017)	Multi-algorithm framework (Feature selection, K-means clustering, Neural Network)	Life insurance customer data	Multi-algorithm model achieved highest predictive accuracy (92.83%)	Framework may be complex to implement; real-time capability not assessed	Extend framework for real-time recommendation systems; test on cross-industry datasets
Almasi et al. (2017)	Fault classification in test generation tools	Software testing data	Almost all (97.62%) of the errors either have complicated state configurations (47.62%) or are associated with particular values (50%)	No proposed solutions for fault detection improvement	Develop tools that better handle specific values and complex object states
Peters and Maxemchuk (2017)	Entropy-based performance comparison of distributed vs. centralized applications using Shannon entropy	Application-level security datasets	Distributed model reduced entropy loss by ~79% in case of attack	Limited to entropy-based evaluation, ignoring latency or throughput	Combine entropy analysis with performance metrics; test under diverse cyberattack scenarios

METHODOLOGY

The proposed methodology involves a structured and systematic approach to healthcare insurance fraud detection using an ANN model illustrated in Figure 1. It starts with the collection of the Prudential Insurance dataset, which then goes through a careful cleaning and checking process to make sure the data is ready to use. This phase deals with filling in any missing values, turning categories into something the machine can read, and using a method called SMOTE to make sure the model gets enough training examples of both types in the data. The balanced dataset is pre-processed and then divided into distinct training and testing portions to ensure the model is evaluated fairly and properly. The training set aids in the development of the ANN model, and then this model is used to see how it works with the testing set of data. The performance of the model is checked using regular ability scores from linear regression like the R^2 value, MAE, MSE, and RSME. These metrics help show how good and reliable the model is at making predictions and using what it has learned in new situations. The final output shows that the ANN model can do a good job at detecting fake insurance cases, giving us results we can trust. Figure 1 displays the suggested life insurance methodology's flow diagram.

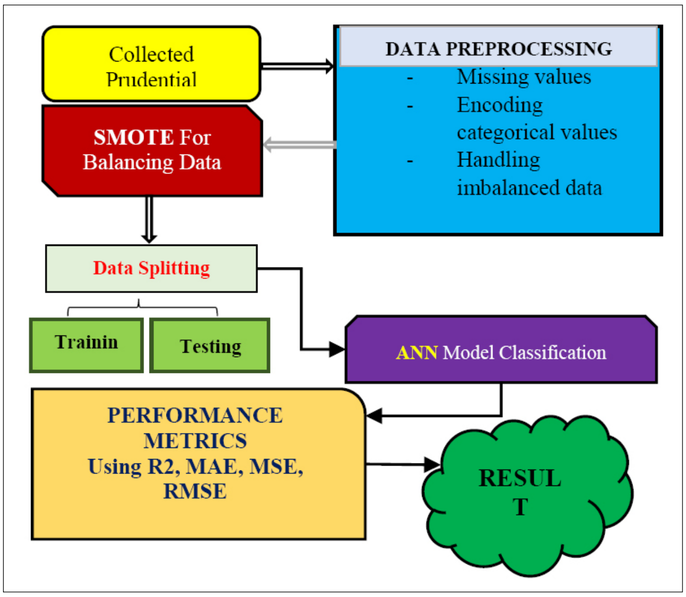

Fig 1. Flowchart for Healthcare Insurance

A quick explanation of the subsequent phases in the flowchart based on risk prediction in life insurance is provided below:

Data Collection

The research is based on a Prudential Insurance Dataset. The information is derived from test data that has been posted to www.kaggle.com. One hundred twenty-eight variables make up the offered dataset. A categorical variable (ordinal data), the response variable (risk score) has eight categories of risk. Customers may be concerned about the disclosure of personal information, such as health records or financial details, in the data supplied, which might include sensitive information about them. EDA involves examining and visually representing the information to improve comprehension of the dataset as a whole by revealing trends, identifying outliers, and summarizing important features.

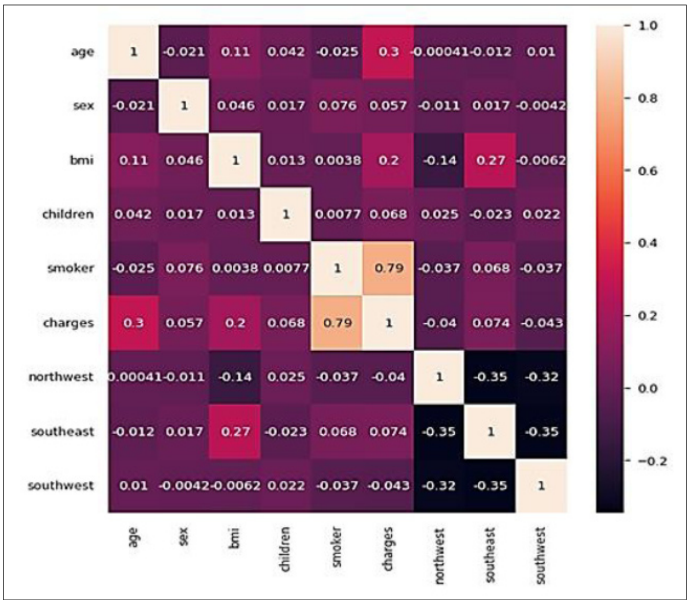

Fig 2. Heatmap of the Correlated Features

The Pearson correlation coefficients among the different characteristics in a health insurance dataset are shown in Figure 2. The strongest positive correlation is observed between **smoker** and **charges** (r = 0.79), indicating that smoking status significantly increases insurance charges. A moderate correlation is also present between **age** and **charges** (r = 0.30), suggesting that charges tend to rise with age. Meanwhile, **bmi** shows a weak positive correlation with **charges** (r = 0.20), and a moderate correlation with **southeast** region (r = 0.27). Negative correlations are observed among some regional variables (e.g., **southeast** with **northwest** and **southwest**), likely due to mutual exclusivity of geographic categories. Overall, most variables exhibit weak or no correlation, indicating a diverse influence of features on insurance charges.

Data Preprocessing

In data analysis, cleaning and converting raw data into a useable format is the initial stage known as pre-processing. Tasks like as categorical variable encoding, missing value handling, and data cleaning, and unbalanced data handling (sometimes using methods like SMOTE to rectify class distributions and enhance model learning) are covered in this research. We will now proceed with the following pre-processing steps:

- **Data Cleaning:** There were some difficulties in first acquiring the data and then selecting the patients with high-quality waveforms appropriate for machine learning because the entire database is somewhat big [13]. Data cleaning is a process of fixing or removing the incorrect data.
- **Missing Values:** It means that there is no value in any one variable or attribute. Missing values show up when there is a problem with the data source [14].
- **Encoding Categorical Variables:** This step counts the bits required to represent each category variable and then finds the values of the numerical variables [15].

Data Balancing Using SMOTE:

An effective solution to the class imbalance issue in land cover classification has been implemented using SMOTE, The most popular method of informed oversampling [16]. This technique creates synthetic instances along the line segment from one randomly selected instance to one of its neighbours in the minority class, hence oversampling the minority class. By utilizing SMOTE to equalize the data, it is possible to learn effectively from both labelled and unlabelled samples. The categorization results were enhanced by implementing SMOTE. [17].

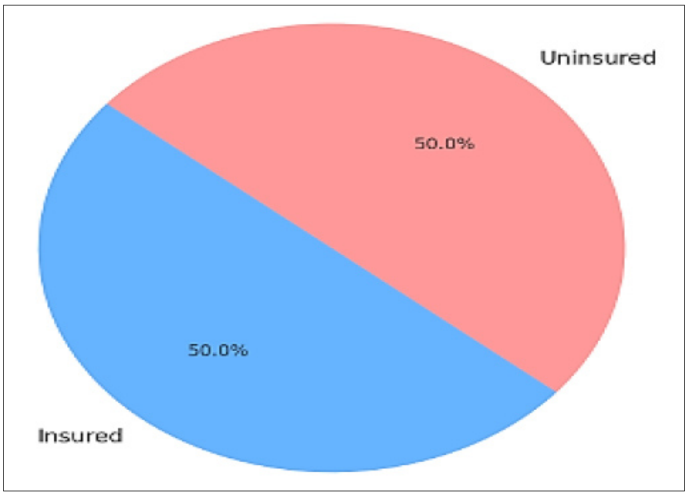

Fig 3. Balanced Healthcare Insurance Data

The Figure 3 shows an equal distribution between insured and uninsured individuals, each comprising 50% of the dataset. This balanced representation helps eliminate bias during model training and supports fair analysis in healthcare insurance prediction.

Data Splitting:

The dataset is divided into two sets: a training set and a testing set. 80% is typically allocated to instruction and 20%

to testing, which provides a balanced approach—ensuring sufficient data for learning while retaining enough unseen data to effectively evaluate the model's performance.

Classification of Proposed ANN Model

The ANN's foundational method is a machine learning strategy that, given historical and current training data, can make predictions, cluster data, and identify patterns in a way that is reminiscent of a human neural network. Recent advances in computing power have reignited interest in ANN, and solutions to The ReLu function has been offered as a solution to the information blurring problem. The problem with the ANN approach is then resolved, and additional hidden layers are added between the input and output layers of the ANN, better results may be achieved. ANNs enable the incorporation of several factors, with the ability to apply varying weights to each variable, resulting in outputs that nearly mimic measurements [18].

Equation (1) describes an ANN, which has the basic architecture illustrated

$$n_k^h = \sum_{j=1}^{R} \omega_{kj}^h p_j + b_{h'}^k,\ k = 1\ to\ S \quad (1)$$

Where R is the quantity of input variables and S is The quantity of neurons that are concealed. Further, p is the input variable, b^h is the bias of the hidden layer, ω^h is the weight. An activation function takes the computed value as input. The activation function is used to adjust the weight total of the previous layer's data to produce the ANN's output from its input [19].

Performance Metrics

The key performance indicators like MAE, which calculates the average absolute differences between expected and actual stock prices; MSE, which highlights larger discrepancies by calculating the average of the squares of the errors; and RMSE, which is the square root of MSE and provides a measure of prediction error, are used to assess the effectiveness of the ANN classifier models, R2, which calculates goodness of fit for the regression model in the same units as Risk Prediction in Life Insurance.

R-Squared (R2) – The R2 value is the most common name for the coefficient of determination. With R-Squared, one may determine what percentage of the variance in the dependent variable can be accounted for by the independent variables. It's may evaluate the fit of the regression model to the data (the degree to which observed values correspond to predicted ones) by computing R-squared. The following equation shows the formula for assessing R2 Equation (2):

$$R^2 = 1 - \frac{SSR}{TSS} \quad (2)$$

Mean Absolute Error (MAE) – The Mean Absolute Error, sometimes called the Mean Absolute Deviation, is the average of the absolute values that deviate from the expected and actual values. The MAE uses the absolute difference between the expected and actual values, rather than the squared difference. This is the main distinction between the two metrics. It is mathematically calculated in Equation (3):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i| \quad (3)$$

Mean Squared Error (MSE): The squared discrepancy between the expected and observed values is known as the mean square error or mean square deviation. So, the MSE shows us how near the line of best fit is to the data set. In every case, the MSE is positive. To remove negative signs, the square is taken. Prediction accuracy improves when the MSE approaches zero. It is numerically presented in Equation (4).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2 \quad (4)$$

Root Mean Squared Error (RMSE) The RMSE is calculated by taking the square root of the sum of all the errors [4]. That is to say, the RMSE is just the standard deviation of the errors. One other thing we can learn from RMSE is how near the line of best fit is to the collection of points. In Equation (5), we may find the RMSE formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2} \quad (5)$$

The model's accuracy in predicting life insurance risks and in comparing results are both measured by these performance matrices.

EXPERIMENTAL RESULTS AND ANALYSIS

The Prudential Insurance dataset, which primarily aims to detect healthcare insurance fraud, was used to validate the suggested ANN mod. Empirical experiments were carried out using Python 2.7 on a system equipped with a 3.6 GHz Intel i7 processor, 8GB of RAM, and running Microsoft Windows 10. As presented in Table II the ANN model delivered robust performance across key regression metrics. It achieved a high R^2 score of 92.7, indicating a solid capacity to measure and clarify the variation of the dependent variable of interest. In addition, the model's MAE was 14.3, MSE was 72.7, and RMSE was 27. This indicates that the model reliably predicts outcomes. These results confirm the model's effectiveness in accurately identifying fraudulent activities within prudential insurance data.

Table II. Performance Evaluation of Proposed ANN Model for Insurance Fraud

Measures	**ANN**
R2	92.7
MAE	14.3
MSE	72.7
RMSE	27

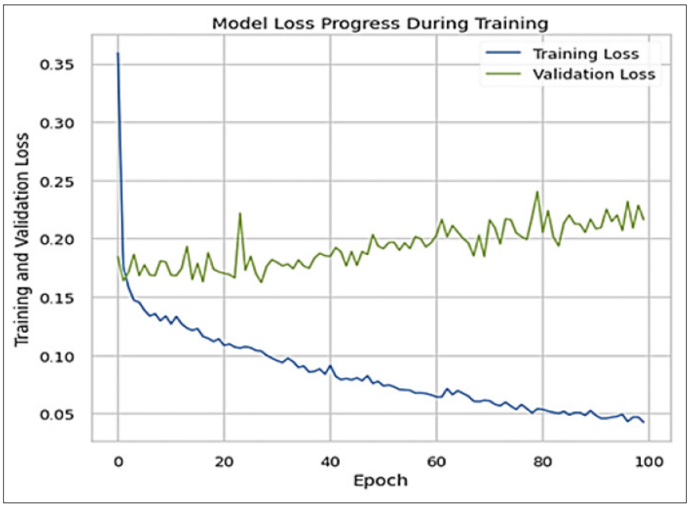

Fig 4. Training Loss vs. validation loss

The training and validation loss progression over 100 epochs is shown in Figure 4. Indicative of the model's successful learning from the training data, the training loss (blue line) steadily declines. When compared, the validation loss (green line) is stable at the outset but starts to vary and eventually increases after 20 to 30 epochs. Because the model's performance on training data is higher than its generalization on unknown data, the fact that the two sets of losses are so different indicates that overfitting may be occurring. In this case, the rising trend of the validation loss indicates that the model's performance is declining on validation data, which calls for action to prevent overfitting by measures such as early halting, regularization, or dropout.

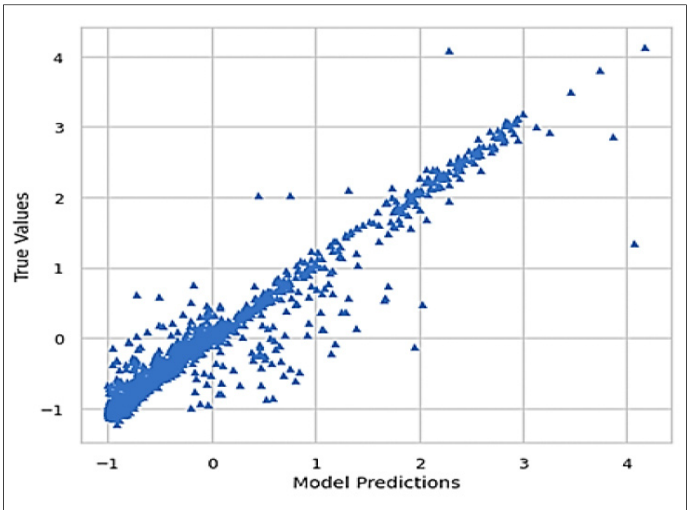

Fig 5. Model Predictions Vs. True Values

Analyzing the relationship between the x- and y-axes of model predictions and actual values is a common way to measure the effectiveness of regression models (Figure 5). A data point is represented by each blue triangle. A high positive correlation between predicted and actual values, as seen by the dense clustering of dots along the diagonal line, implies that the model is generally functioning effectively with minimum error. However, some dispersion at higher values may indicate increasing prediction variance or outliers.

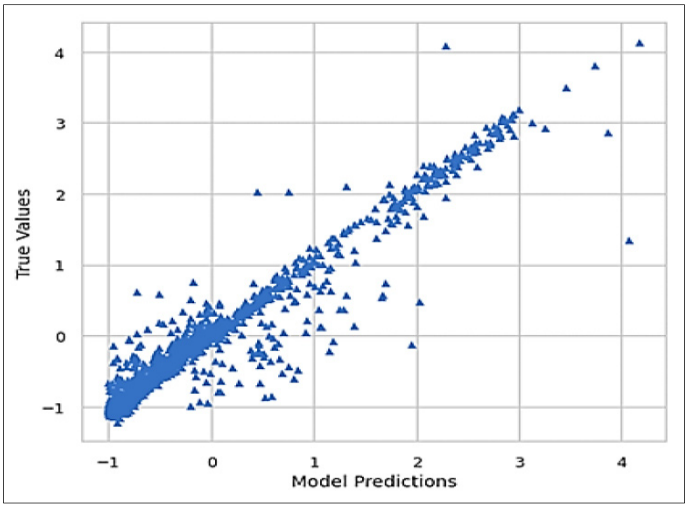

Fig 6. Inverse Transform of Model Predictions Vs. True Values

Figure 6 displays the outcomes of the testing of the proposed ANN model. The y-axis displays the actual values, while the x-axis displays the projected values. The data points, represented as blue triangles, are tightly clustered along the diagonal line, indicating a strong linear relationship and high predictive accuracy. The excellent performance of the model is reflected in this visual alignment. The plot confirms the ANN model's effectiveness in closely approximating true outcomes in healthcare insurance fraud detection.

Comparison and Discussion

The experiments evaluate the proposed model's predictive power in comparison to other basic learner models. Table III provides a comprehensive overview of the models' functionality. It is obvious from the analysis that the proposed ANN performed much better than the baseline models—PLS-SEM and LR. The ANN model scored an R^2 of 92.7, which is much better than the R^2 scores of 64.1 from PLS-SEM and 62 from LR. The significant improvement means that the ANN model can much better recognize and explain the different values in the target variable, which makes it more reliable and practical for healthcare insurance fraud detection. It is evident from the study that, mainly when working with complex data, ANN-style deep learning models offer a stronger solution for life insurance risk prediction.

Table III. Comparison between proposed and baseline Models' performance for risk prediction in life Insurance

Performance Metrics	**PLS-SEM** [20]	**LR** [21]	**ANN**
R2	64.1	62	92.7

The proposed ANN model offers several advantages that make it highly suitable for healthcare insurance fraud detection. It is better able to predict the future because it catches complex patterns in data that many other models struggle to see. Automatic learning in ANN lessens the burden of feature engineering, and its reliability in real-life makes it work well with messy or incomplete data. Moreover, the model can be easily modified and updated to stay ahead of fraud as new trends emerge. Because it can deal with a high number of input variables, it is more effective when processing complex data from the insurance sector.

CONCLUSION AND FUTURE DIRECTION

It has been established in the health insurance literature that how much something costs (premiums, deductible, out-of-pocket max) plays a key role in consumers' preferences. Yet, these financial reasons are only a small part of what consumers balance when making a choice. The results of this study demonstrate that the ANN model is effective in detecting life insurance fraud and making risk predictions. The high value of R^2 (92.7%) and the low error metrics verified that the ANN was able to handle complicated, nonlinear patterns and make trustworthy predictions using the SMOTE technique for class balancing among the data. The findings demonstrate that using this model can lead to better underwriting, improved pricing, and expanded efforts to catch fraud within the industry. Since the model uses a single set of data, it may not apply well to all types of insurance portfolios. Moreover, noise in the data and values that are missing can influence the ANN's behaviour. Further research could look into CNNs and RNNs, as these advanced deep learning architectures are well suited for handling images and sequences in data, respectively. The use of XAI would increase how transparent the model is and how stakeholders trust it. Including data of 24-hour sources such as wearable devices or social media could help to improve how risks are assessed. Moreover, using online learning systems allows for ongoing changes to fight against fraud and adapt to changing forces in the insurance sector.

REFERENCES

1. N. Kumar, J. D. Srivastava, and H. Bisht, "Artificial intelligence in insurance sector," *J. Gujarat Res. Soc.*, vol. 21, no. 7, 2019.
2. V. Kolluri, "A Pioneering Approach To Forensic Insights: Utilization Ai for Cybersecurity Incident Investigations," *Int. J. Res. Anal. Rev. (IJRAR*, vol. 3, no. 3, 2016.
3. T. Davenport and R. Kalakota, "The potential for artificial intelligence in healthcare," *Futur. Healthc. J.*, vol. 6, no. 2, pp. 94–98, Jun. 2019, doi: 10.7861/futurehosp.6-2-94.
4. V. Kolluri, "An Innovative Study Exploring Revolutionizing Healthcare With AI: Personalized Medicine: Predictive Diagnostic Techniques and Individualized Treatment," *JETIR - Int. J. Emerg. Technol. Innov. Res. (www. jetir. org| UGC issn Approv. ISSN*, vol. 3, no. 11, pp. 2349–5162, 2016.
5. M. Maier, H. Carlotto, F. Sanchez, S. Balogun, and S. Merritt, "Transforming Underwriting in the Life Insurance Industry," *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 01, pp. 9373–9380, Jul. 2019, doi: 10.1609/aaai.v33i01.33019373.
6. D. D. Rao, "Multimedia Based Intelligent Content Networking for Future Internet," in *2009 Third UKSim European Symposium on Computer Modeling and Simulation*, 2009, pp. 55–59. doi: 10.1109/EMS.2009.108.
7. V. Kolluri, "Machine Learning in Managing Healthcare Supply Chains: How Machine Learning," *J. Emerg. Technol. Innov. Res.*, vol. 3, no. 6, pp. 2349–5162, 2016.
8. C. Shih, C. C. C. William, and Y. W. Chang, "The causes analysis of ischemic stroke transformation into hemorrhagic stroke using PLS (partial least square)-GA and swarm algorithm," in *Proceedings - International Computer Software and Applications Conference*, 2019. doi: 10.1109/COMPSAC.2019.00108.
9. A. Mishra and U. S. Reddy, "A comparative study of customer churn prediction in telecom industry using ensemble based classifiers," *Proc. Int. Conf. Inven. Comput. Informatics, ICICI 2017*, no. January, pp. 721–725, 2018, doi: 10.1109/ICICI.2017.8365230.
10. T. Kaewkiriya, "Framework for prediction of life insurance customers based on multi-algorithms," in *7th International Conference on Innovative Computing Technology, INTECH 2017*, 2017. doi: 10.1109/INTECH.2017.8102444.
11. M. M. Almasi, H. Hemmati, G. Fraser, A. Arcuri, and J. Benefelds, "An industrial evaluation of unit test generation: Finding real faults in a financial application," in *Proceedings - 2017 IEEE/ACM 39th International Conference on Software Engineering: Software Engineering in Practice Track, ICSE-SEIP 2017*, 2017. doi: 10.1109/ICSE-SEIP.2017.27.
12. E. Peters and N. Maxemchuk, "A Privacy-Preserving Distributed Medical Insurance Claim Clearinghouse & EHR Application," in *Proceedings - 2017 IEEE 2nd International Conference on Connected Health: Applications, Systems and Engineering Technologies, CHASE 2017*, 2017. doi: 10.1109/CHASE.2017.62.
13. G. Slapničar, N. Mlakar, and M. Luštrek, "Blood Pressure Estimation from Photoplethysmogram Using a Spectro-Temporal Deep Neural Network," *Sensors*, vol. 19, no. 15, p. 3420, Aug. 2019, doi: 10.3390/s19153420.
14. I. B. Aydilek and A. Arslan, "A hybrid method for imputation of missing values using optimized fuzzy c-means with support vector regression and a genetic algorithm," *Inf. Sci. (Ny).*, 2013, doi: 10.1016/j.ins.2013.01.021.
15. J. Salvador Meneses, Z. R. Chavez, and J. G. Rodriguez, "Compressed kNN: K-Nearest Neighbors with Data Compression," *Entropy*, vol. 21, no. 3, p. 234, Feb. 2019, doi: 10.3390/e21030234.
16. T. W. Cenggoro, S. M. Isa, G. P. Kusuma, and B. Pardamean, "Classification of imbalanced land-use/land-cover data

using variational semi-supervised learning," in *2017 International Conference on Innovative and Creative Information Technology (ICITech)*, IEEE, Nov. 2017, pp. 1–6. doi: 10.1109/INNOCIT.2017.8319149.

17. G. Douzas, F. Bacao, J. Fonseca, and M. Khudinyan, "Imbalanced learning in land cover classification: Improving minority classes' prediction accuracy using the geometric SMOTE algorithm," *Remote Sens.*, 2019, doi: 10.3390/rs11243040.
18. R. Salakhutdinov, A. Mnih, and G. Hinton, "Restricted Boltzmann machines for collaborative filtering," in *ACM International Conference Proceeding Series*, 2007. doi: 10.1145/1273496.1273596.
19. S. Lee, S. Jung, and J. Lee, "Prediction Model Based on an Artificial Neural Network for User-Based Building Energy Consumption in South Korea," *Energies*, vol. 12, no. 4, p. 608, Feb. 2019, doi: 10.3390/en12040608.
20. S. S. Weedige, H. Ouyang, Y. Gao, and Y. Liu, "Decision making in personal insurance: Impact of insurance literacy," *Sustain.*, vol. 11, no. 23, pp. 1–24, 2019, doi: 10.3390/su11236795.
21. A. R. Sarker, M. Sultana, S. Ahmed, R. A. Mahumud, A. Morton, and J. A. M. Khan, "Clients' experience and satisfaction of utilizing healthcare services in a community based health insurance program in bangladesh," *Int. J. Environ. Res. Public Health*, vol. 15, no. 8, 2018, doi: 10.3390/ijerph15081637.
22. Chinta, P. C. R., & Karaka, L. M. (2020). Agentic AI and Reinforcement Learning: Towards More Autonomous and Adaptive AI Systems.
23. 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.
24. 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.
25. 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.
26. 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*.
27. Kuraku, S., & Kalla, D. (2020). Emotet malware—a banking credentials stealer. *Iosr J. Comput. Eng*, *22*, 31-41.
28. Kalla, D., & Samiuddin, V. (2020). Chatbot for medical treatment using NLTK Lib. *IOSR J. Comput. Eng*, *22*, 12.
29. 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*.
30. 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*.
31. Karaka, L. M. (2021). Optimising Product Enhancements Strategic Approaches to Managing Complexity. *Available at SSRN 5147875*.
32. 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.

Citation: Srikanth Reddy Vangala, Ram Mohan Polam, et al., "Leveraging Artificial Intelligence Algorithms for Risk Prediction in Life Insurance Service Industry", Universal Library of Engineering Technology, 2022; 27-34. DOI: https://doi.org/10.70315/uloap.ulete.2022.004.