



ARTICLE

A Deep Learning Model for Insurance Claims Predictions

Umar Isa Abdulkadir* and Anil Fernando*

Department of Computer and Information Science, University of Strathclyde, Glasgow, UK

*Corresponding Authors: Umar Isa Abdulkadir. Email: umar.isa@strath.ac.uk; Anil Fernando. Email: anil.fernando@strath.ac.uk

Received: 24 August 2023 Accepted: 11 March 2024 Published: 22 April 2024

ABSTRACT

One of the significant issues the insurance industry faces is its ability to predict future claims related to individual policyholders. As risk varies from one policyholder to another, the industry has faced the challenge of using various risk factors to accurately predict the likelihood of claims by policyholders using historical data. Traditional machine-learning models that use neural networks are recognized as exceptional algorithms with predictive capabilities. This study aims to develop a deep learning model using sequential deep regression techniques for insurance claim prediction using historical data obtained from Kaggle with 1339 cases and eight variables. This study adopted a sequential model in Keras and compared the model with ReLU and Swish as activation functions. The performance metrics used during the training to evaluate the model performance are R2 score, mean square error and mean percentage error with values of 0.5%, 1.17% and 23.5%, respectively obtained using ReLU, while 0.7%, 0.82%, and 21.3%, obtained using Swish function. Although the results of both models using ReLU and Swish were fairly tolerable, the performance metrics obtained of the model using Swish, as the activation function tends to perform more satisfactorily than that of ReLU. However, to investigate the model performance deeply, therefore, this study recommends that more interest be channeled to the interpretability and explainability of the proposed model and the provisions of AI technologies by insurance industries to enhance accurate claim prediction and minimize losses.

KEYWORDS

Claim predictions; insurance; deep learning; ReLU; swish; artificial intelligence

1 Introduction

Deep learning in this present age has gathered significant momentum about artificial intelligence as a research direction with inherent potential for the present and future. Although deep learning has recently attracted attention, its significance as a vital aspect of artificial intelligence has promoted its necessity as a study area in modern society. There is an outburst of demand for deep learning models that have spread through various fields, both as a profession and as a study area. Reference [1] opined that the increasing demand for deep learning in various fields can be attributed to its ability to perform tasks swiftly at a summative frequency. This was acknowledged to have attracted ubiquitous relevance to their models and concepts; a significant among them is the convolutional neural network [2]. Most of these models have been utilized in everyday activities, such as speech recognition and



picture classification, and in other more suffocated areas, such as predictions. The classification layer employed by most of these models is the conventional SoftMax function. However, some studies [3,4] have used other classification functions, such as SoftMax, as this study adopted the use of rectified linear units (ReLU) as the model classification layer. Conventionally, a deep neural network (ReLU) is adopted to serve as an activation function for hidden layers. This new approach is novel about insurance claims prediction, which is the main objective of this study. Although players in the insurance business did not openly adopt deep learning until recently, confirmation of the increase in the use of machine learning methods in insurance is the growing number of jobs offered to data scientists and machine learning engineers in insurance industries [5].

Previously, big players in the insurance industry were reluctant to adopt new technologies such as machine learning with preference over traditional approaches because of the many regulations associated with the insurance business and the uncertainty surrounding the new technologies [5,6]. However, recent developments suggest that the insurance industry embraces the dynamic involvement of machine learning and its applications in day-to-day insurance processes.

As reference [7] opined, these new usages include but are not limited to fraud detection, risk profiling, virtual assistance targeted at customer advisory, and claims prediction. This study focuses on claim prediction, a vital operation associated with the insurance industry. As stated in [8], claims processing and prediction are crucial in maintaining continued market penetration for any insurance business, whether for life or non-life insurance. Accurate claims prediction is vital for insurance companies to make adequate recommendations on the best type of insurance package for every potential policyholder, as risk differs from one client to another [5].

The importance of machine learning in the insurance industry is evident, with the need to automate numerous claims and client complaints coming in hundreds and thousands of hours [3]. The industry's discrete regulatory and compliance requirements have provided indisputable relevance to the role of deep learning as financial technology in this modern era [9]. Automating claims processes with deep learning algorithms enables swift claims processing from preliminary reports to the informed analytical information necessary for client interaction. However, deep learning algorithms are not limited to hastening claims processing and also play a vital role in claims prediction by enabling a more detailed understanding of claims demand, which invariably aids in cost minimization, as insurers minimize risk and maximize profit [10]. Another notable usage of deep learning in the insurance business is the deep learning-generated chat box that services customers' inquiries, freeing staff to concentrate on the more complex work demands of the industry.

In insurance claims prediction, the most used machine learning algorithms are primarily categorized into supervised [7], unsupervised, and semi-supervised models [11]. To highlight the significance of the categorization, it is inherent to analyze the rudimentary essentials of the first two initial algorithm classes. Supervised deep learning operates by deducing a function that maps a given data input to a given data output. Training data collection was employed as an input system during the model training phase.

A label that indicates a desired output value is attached for every input value, with the essential role of supervising the model; hence, the model already has knowledge of the possible output as the input is submitted to the system. Unsupervised learning involves or is referred to as, an analytical technique that operates by inferring a function with the best characterization of the inputted data without supervision. Reference [12] unsupervised learning is more subjective than supervised learning because it has no specified goal such as response prediction.

To evaluate the significance of deep learning in enhanced claims prediction and other aspects of insurance claims processing and management, it is necessary to consider this aspect of the study area to unify and advance new findings. Claim prediction is a delicate aspect of the insurance process that significantly enhances the insurance sector penetration and profit maximization [13]. However, the acceptance of deep learning models in the insurance industry is yet to achieve optimal performance and maximum success, although there is a persistent clamor for more utilization of these new technologies in the sector. Therefore, this study intends to present a holistic perception of deep learning model utilization for insurance claims prediction and thereby propose a more reliable argument for that effect.

2 Brief Review of Empirical Literature

As stated in [14], different methods such as time series, simulation models, statistical models, and machine learning have been considered to address issues focused on predictions. A review of the empirical literature as shown in Table 1, indicates that various studies have been conducted on artificial intelligence models and insurance sector processes. Few of these deep learning studies have focused on comparing several ReLU activation functions.

These studies aimed to consider their efficiency and accuracy [14–16], with no direct link to their utilization in the insurance process. Taking a closer dimension to artificial intelligence and insurance processes, reference [5] indicated that a significant number of AI applications in the insurance industry for claim prediction and fraud detection, especially in motor insurance, focused mainly on the imbalanced dataset, which is a significant problem of AI in the industry.

Reference [17] applied XGBoost accuracy to predict claims and compared its performance with other techniques, such as neural networks, random forests, and AdaBoost. Although the results of this study indicated that XGBoost produced a more accurate prediction in terms of normalized Gini, other studies [18] indicated that this simple technique is prevalent with some form of bias in their results.

According to [19], comparing the performance of XGBoost and logistic regression as a predictive technique for accident insurance claims, combined with a small amount of training data, indicated that logistic regression is a more appropriate model considering its good predictive capability and interpretability. Reference [20] utilized data mining techniques to predict possible fraud in claims applications and ascertain the premium weight of various customers regarding their details and financial records using three classifiers.

The adopted algorithms were Naïve Bayes and Random Forest, and the results indicated that random forest showed a higher predictive ability than other techniques. This implies that the study's predictive ability is preferable to detecting fraud than predicting claims.

Other studies [2] applied deep learning to a different concept that predicts the terminal call rate. Although this study had a slight deviation for the insurance sector, it utilized a model similar to that adopted for this study, which is a deep neural network that has six layers combined with a convolutional neural network.

3 Data Set and Model Architecture

3.1 Variables and Data Set

The variables adopted for the study included age, sex, body mass index (BMI), steps, children, smokers, region, charges, and insurance claims. The variable descriptions are presented in Table 2.

Table 1: Tabular form of literature review

Reference	Methodology/Model	Focus	Key findings
[14]	Time series, simulation models, statistical models, machine learning	Predictions in the insurance sector	Various methods considered; focus on efficiency and accuracy of ReLU activation functions.
[5]	Artificial Intelligence applications in the insurance industry	Claim prediction, fraud detection, imbalanced datasets	AI applications in motor insurance, addressing imbalanced datasets; [5] emphasizes the prevalence of imbalanced datasets in AI in the industry.
[17]	XGBoost, neural networks, random forests, AdaBoost	Claim prediction	XGBoost outperformed in terms of normalized Gini; some bias noted in the results of this study.
[19]	XGBoost, logistic regression	Accident insurance claim prediction	Logistic regression deemed more appropriate with small training data due to good predictive capability and interpretability.
[20]	Data mining techniques (Naïve Bayes, Random Forest)	Fraud prediction in claims applications	Random Forest showed higher predictive ability than other techniques in detecting fraud.
[2]	Deep neural network with convolutional neural network	Predicting terminal call rate	Utilized a deep neural network similar to the study; six layers combined with a convolutional neural network.

Table 2: Dataset descriptions [21]

Name	Description
Age	Policy holder's age
Sex	policyholder's gender (male = 1, female = 0)
BMI	Body mass index
Steps	Number of walking step of policy holders per day in average
Children	Number of policyholder's children and dependents
Smokers	Policyholder smoking status (smoker = 1, non-smoke = 0)
Region	Area of resident of policy holder
Charges	costs medicals billed per individual by health insurance
Insurance claim	Number of insurance claims received

The dataset used to build the claim predictor was sourced from Kaggle [21]. The dataset obtained consisted of eight variables and 1339 cases. The dataset was divided into 80/20 trains/tests and split using Sklearn’s train_test_split function with 30 random states. The structure of the dataset is shown in Fig. 1.

	age	sex	bmi	children	smoker	region	charges	insuranceclaim
0	19	0	27.900	0	1	3	16884.92400	1
1	18	1	33.770	1	0	2	1725.55230	1
2	28	1	33.000	3	0	2	4449.46200	0
3	33	1	22.705	0	0	1	21984.47061	0
4	32	1	28.880	0	0	1	3866.85520	1

Figure 1: Dataset used for training the proposed model [21]

3.2 Proposed Model

This study developed a deep learning model to predict insurance claims based on insurance policyholder data, as indicated in Table 2. The design model was structured to capture different phases, as adopted from the sequential model. A structural overview is shown in Fig. 2.

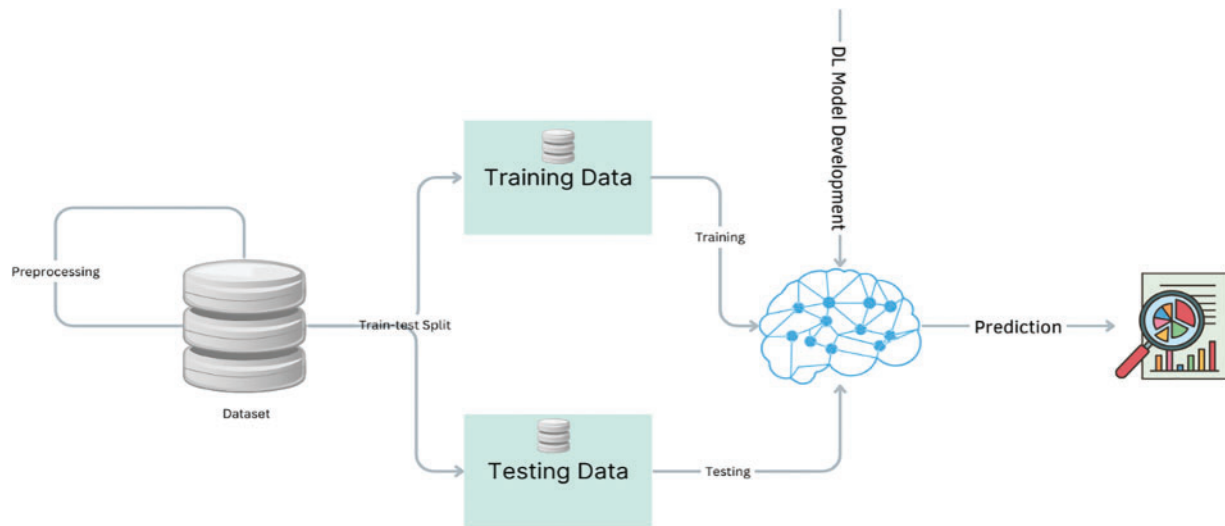


Figure 2: Semantic overview of the model

3.3 Model Architecture

The model used in this study was structured using a neural network. The neural network accepts inputs and processes them in hidden layers, whose weights are attuned during training, which enables the model to split predictions as an output [22]. The study adopted Keras using is sequential deep regression model, a neural network library, and the model type employed was the sequential model, which is the easiest method of building a model using Keras, as it allows the model to be structured layer by layer.

Each layer has weights relative to the following one. To design a neural network, an activation function must be available to do the transformation from summed weighted input directly into node activation [23]. The activation function chosen for this study is the rectified linear activation functions (ReLU), see Fig. 3, also known as short control, short model. The ReLU is a piecewise-linear function. If the input is positive, it simply outputs its value; otherwise zero. As a result, the model that uses it has become many different kinds of neural networks' default activation function. It is also easier to train and usually accomplishes better results.

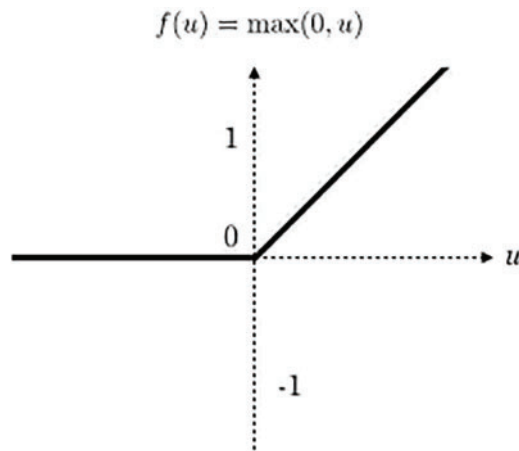


Figure 3: ReLU activation function [22]

Alternatively, swish is a smooth continuous function (see Fig. 4), that has attributes that allow a minimal number of negative weights to be circulated through it. When used in deep neural networks, this is an essential property and is crucial for the feat of nonmonotonic smooth activation functions. In addition, the trainable parameter provides a provision that allows the activation function to propagate maximum information and provides smoother gradients for easy optimization and generalization.

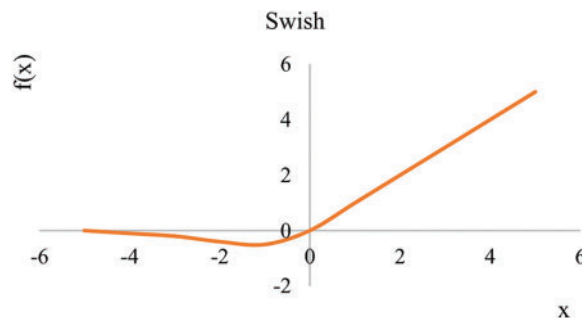


Figure 4: Swish activation function [24]

As an activation function, Swish (see Fig. 4) is self-gating; this implies that it modifies the input by utilizing it as a gate multiplier together with its sigmoid. This concept was first adopted in Long Short-Term Memory (LSTMs) [24].

The study used 'adam' as the optimizer, as it is generally considered a perfect optimizer in various instances. Theoretically, the learning rate was adjusted using the Adam optimizer throughout training. The study also used the mean squared error as its loss function, which involves calculating the average

of the square difference between predicted and actual values. In theory, it is adopted as the usual loss function for almost all regression problems and your model performs better if you are closer to zero [22].

4 Results and Discussion

Considering that the study did not adopt an existing baseline system, the model architecture comprised three models. In the first model, which is an ordinary model, there is only one set of layers, that is, the output layer. An epoch equivalent of 1000, Adam optimizer of 0.001, and validation split of 0.1 was used.

Considering the above performance output shown in Fig. 5, it is evident that the ordinary model does not indicate significant training that can inform predictions. Hence, the study adopted a deep learning model that has 5 layers and an output layer using the ReLU activation function and Adam optimizer of 0.003 with a validation loss function as part of the evaluation. The model used epochs of 4000 with a validation split of 0.2. The illustration of the model performance indicates the following.

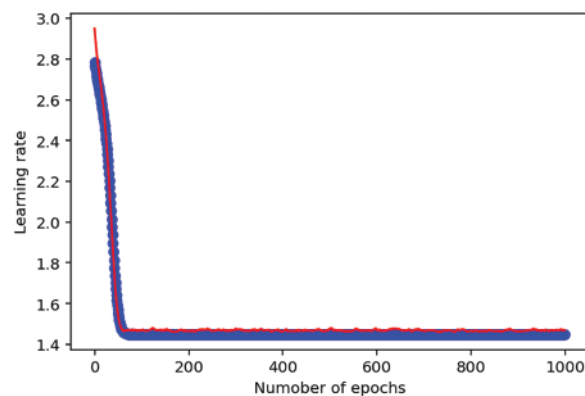


Figure 5: Ordinary model training output

In Fig. 6, the blue line indicates the training loss, which is constantly decreasing from the illustration. However, the red line, which indicates the trading loss validation, is constantly increasing. Based on this, the performance of the training loss is consistent, whereas the validation is insufficient and requires adjustment in the model. The same model was redesigned using the Adam optimizer with a learning rate of 0.003, an epoch of 5000, a validation split of 0.2, a patient parameter of 15, and an early stop parameter. The second illustration of the model performance with the new adjustments indicates the following.

From the plots shown in Fig. 7, it is evident that both the training loss and training validation are constantly moving downward. However, the training validation plot is imperfect, as the training lines do not seem to be moving smoothly. The performance of this model can be considered to be pretty good. Still, with the study's objective centered on developing a near-perfect model for predicting insurance claims, there is a need to compare the performance of this model with a similar model using the Swish activation function.

To achieve this, the study maintained the structural arrangement of the previous model, which has five layers and an output layer. The study replaced the ReLU activation function with the Swish activation function, maintaining an Adam optimizer of 0.003 and 4000 epochs with a validation split of 0.2. The model performance is shown in the figure below.

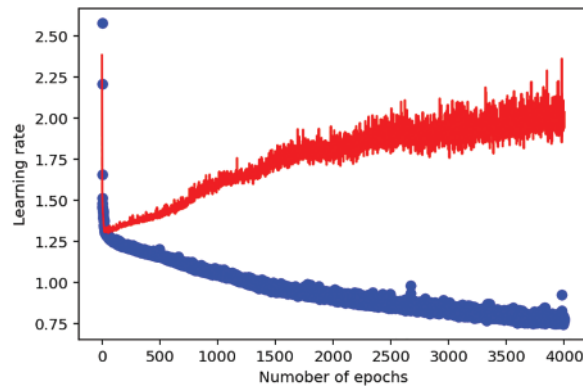


Figure 6: First deep learning model output with ReLU activation function

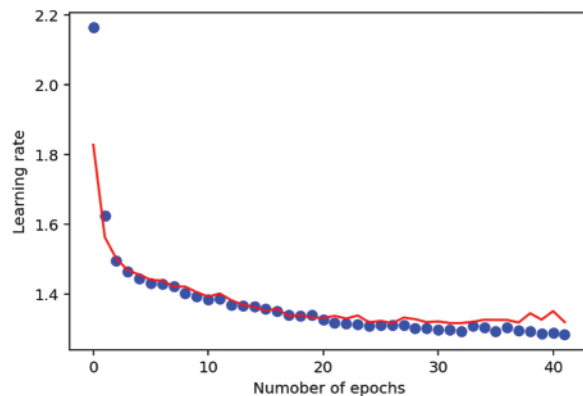


Figure 7: Second deep learning model output with ReLU activation function

From the training results, as indicated in the above [Fig. 8](#), the training loss, as indicated in the previous model with a similar structure, moved downward, while the training validation was moving upward. This implies that this model using attributes similar to the formal one is imperfect, even with adopting the Swish activation function. Therefore, this study applied some adjustments by redesigning the same model using an Adam optimizer of 0.003, an epoch of 5000, a validation split of 0.2, a patient parameter of 15, and an early stop parameter. The performance illustration of the redesigned model with the new adjustments indicates the following.

From the training performance results above shown in [Fig. 9](#), the plot of the training loss and training loss validation followed a perfect downward trend. This implies that this model with Swish as the activation function performed better than the previous models with the ReLU activation function. This result confirms the study of [25], which indicated that, although the ReLU activation function is the most widely used, the Swish activation function, in most cases, performs better than ReLU because of its nit-monotonic attributes. This implies that it allows the value of the function to decrease regardless of the increase in the values of the inputs.

The choice of activation functions has a large impact on the performance of neural network models. In this case, during training two well-known activation functions were used: Rectified Linear Unit (ReLU) and Swish. The effect on model performance of these different types was measured in terms of several important metrics.

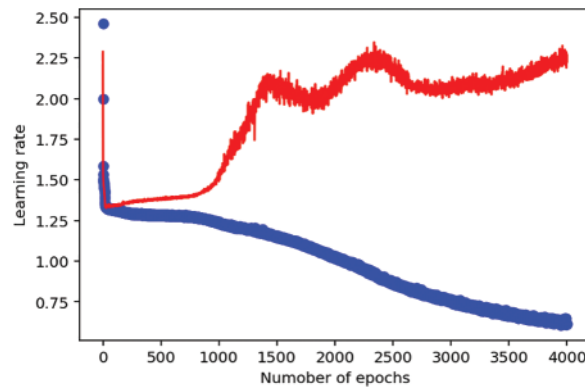


Figure 8: Adjusted deep learning model with Swish activation function

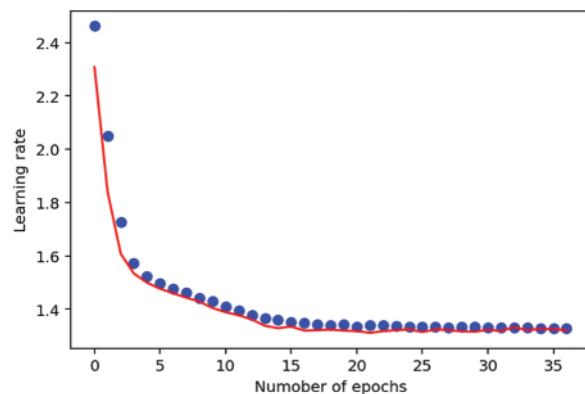


Figure 9: Third deep learning model with Swish activation function

R2 scores (a measure of how well the model explains variance in data) for calculations using ReLU and Swish were 0.5, respectively. Thus, the higher R2 score indicates better prediction abilities. This implies that Swish captures more of the variation in data than ReLU.

With ReLU, the mean square error (average squared difference between predicted and observed values) was 1.17, but only with Swish it is dropped down to 0.82. Therefore, swish yielded a smaller MSE. That is to say this activation produces more accurate predictions and lower overall error for the model. In addition, the mean percent error (MPE) decreased from 23.5% with ReLU to 21.3% for Swish. That is to say, applying the Swish function obtains a model with smaller average percentage deviation from true values (higher precision).

So to sum up: The Swish activation function gave better results on almost all measures—higher R2 scores, lower MSEs and smaller MPE values. These findings also emphasize the significance of choosing suitable activation functions for machine learning applications in order to optimize model accuracy and ability to predict. As a result, the current work achieves better results with non-monotonic and smoother Swish than ReLU. Unlike ReLU, Swish is distinguished in that it allows the value of a function to decrease based on particular values which are entered. This way the model can distinguish finer relationships between data points and better visualize them. For healthcare claims forecasting, the more complex product is Swish rather than the simple ReLU stimulation. It excels in reflecting fine details of connections between parameters that may be messy and nonlinear.

These results have significant practical implications for the insurance industry. The choice of stimulation functions directly affects the reliability and efficiency of prediction models. Given Swish's improved efficiency, insurance firms that use models developed using deep learning to anticipate claims may increase the reliability of their projections by using Swish as the activation function. This enhancement may result in improved risk assessment, better-informed purchasing procedures, and eventually, more effective utilization of assets throughout the risk management sector.

The study's modest novelty is investigating activation processes in insurance claim prediction, contrasting the widely used ReLU with the relatively newer Swish activation function. Although activation functions are essential in neural network models, no new idea or revolutionary method is presented in this study. The result supports previous findings that the selection of the activation mechanism can significantly impact model performance.

5 Conclusion

The study provides an overview of deep learning models with the objective of developing a deep learning model for insurance claims prediction. This study adopted a neural network model using the data from Kaggle [21]. The model used a sequential model in keras, and its performance was compared using the ReLU activation function and Swish. After observation, the results of training showed that applying the swish activation function in our model occurred with a far better prediction (the R2 score on both training set and test set are 0.98 and 0.71, respectively). It says that the model fits perfectly and makes a perfect job of an unseen data set.

Choice of activation functions, such as ReLU and Swish affect performance metrics greatly. The R2 value for a model trained with ReLU is 0.5, and the mean percentage error rate 23.7 percent and mean square error score are both at around 0.8. These measurements show how well the Swish function works in this particular training environment. They highlight just how much it increases model accuracy and reduces errors compared to ReLU.

The outcome of the experiment indicates that the deep learning model has a powerful advantage in insurance claims prediction. The tentative prediction capacity illustrated with the models tends to improve and enhance the accuracy of claims prediction in the insurance industry, which is vital for business sustainability and profit maximization. Therefore, it is imperative that the insurance industry leverage AI technologies by adopting deep learning models to enhance claim predictions. In addition, after conducting a systematic literature review and several experiments, it was observed that there are several areas that need to be investigated, especially the performance of the model and its interpretability and explainability. Hence, it is recommended that the interpretability, explainability, and optimization of the proposed deep learning model be investigated further.

6 Future Recommendations

Suggestion for Future Research: Although the analysis rightly notes that comprehension, along with clearness in the framework of insurance claim forecasting, requires further investigation, offering specific research inquiries or guidelines can improve the direction that investigators in this field can take. The following are some recommended areas of inquiry:

Interpretable Features of the Model

Look at ways to improve the readability of characteristics that go into predicting insurance claims. How may deep learning models be enhanced or changed to give more precise insights into the variables affecting a claim?

Explain Ability in the Making of Model Decisions

Examine methods for increasing the transparency of the deep learning model's decision-making process. Does the model explain why projections are generated, particularly regarding insurance claims where openness is essential?

A Comparative Analysis of Explain Ability Methods

Compare and ascertain the efficacy of different explain ability methodologies (e.g., LIME and SHAP values) in the insurance claim prediction arena. Which techniques offer a comprehensible understanding of the behavior of the model?

Collaboration between Humans and AI to Comprehend Predictions

To increase comprehension, look for methods to make it easier for human specialists and the AI model to collaborate. To create a mutually beneficial relationship and enable better-informed decision-making, how can AI systems effectively convey forecasts to domain professionals?

Feature Engineering's Effect on Interpretation

Examine how feature engineering affects the ability to interpret of the model. In insurance claims, why do various methods of feature engineering impact the capacity to decipher and comprehend the reinforcement learning model's forecasts?

Easy-to-Use Justifications

Look at ways to produce understandable answers for stakeholders who are not engineers, including policyholders or brokers who sell insurance. How might findings from intricate models be comprehensible to a wide range of audiences?

Ethics in Explain Ability: A Perspective

Discuss the moral issues around the comprehensible nature of the model. How may the ethical ramifications of employing interpretable models in the insurance sector be lessened to maintain transparency and equal treatment?

Long-Term the Ability to Explain the Model

Examine how model interpretations relate to time. How can a deep learning model retain its comprehension as time passes, particularly in dynamic contexts like the insurance industry, where risk variables are subject to change?

Comparing Explain Ability Methodologies

Provide benchmark datasets and assessment metrics to evaluate clearness algorithms' effectiveness in the insurance claim prediction arena. This could aid in the field's establishment of standardized procedures.

Combining Domain Knowledge

Look at ways to incorporate domain-specific expertise into the accessibility model. How can the comprehension of the model's predictions be improved by utilizing past knowledge of insurance laws, rules, and industry practices?

Future research may significantly advance the topic of interpretability and explain ability in insurance claim prediction by exploring these research avenues, ultimately resulting in the development of more reliable and efficient predictive models for the insurance industry.

Acknowledgement: We sincerely acknowledge all contributors for their support and wisdom. Expertise and dedication greatly enrich this paper. Thank you, reviewers for your helpful suggestions which improved the quality of our work. Thank you, colleagues and institution for your encouragement and assistance in making this publication possible.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: Conceptualization: Umar Abdulkadir Isa and Anil Fernando; Methodology: Umar Abdulkadir Isa and Anil Fernando; Software: Umar Abdulkadir Isa; Validation: Umar Abdulkadir Isa; Formal analysis: Umar Abdulkadir Isa; Investigation: Umar Abdulkadir Isa; Writing–original draft: Umar Abdulkadir Isa; Writing–review and editing: Umar Abdulkadir Isa and Anil Fernando; Supervision: Anil Fernando.

Availability of Data and Materials: Sample Insurance Claim Prediction Dataset: <https://www.kaggle.com/datasets/easonlai/sample-insurance-claim-prediction-dataset> (accessed on 08 February 2023) (Available at Kaggle for Research Purposes).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] N. Almufadi, A. M. Qamar, R. U. Khan, and M. T. Ben Othman, “Deep learning-based churn prediction of telecom subscribers,” *Int. J. Eng. Res. Technol.*, vol. 12, no. 12, pp. 2743–2748, 2019.
- [2] A. Ferencek, D. Kofjač, A. Škraba, B. Sašek, and M. K. Borštnar, “Deep learning predictive models for terminal call rate prediction during the warranty period,” *Bus Syst. Res.*, vol. 11, no. 2, pp. 36–50, 2020. doi: [10.2478/bsrj-2020-0014](https://doi.org/10.2478/bsrj-2020-0014).
- [3] E. Alamir, T. Urgessa, A. Hunegnaw, and T. Gopikrishna, “Motor insurance claim status prediction using machine learning techniques,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 3, pp. 457–463, 2021. doi: [10.14569/issn.2156-5570](https://doi.org/10.14569/issn.2156-5570).
- [4] M. N. Favorskaya and V. V. Andreev, “The study of activation functions in deep learning for pedestrian detection and tracking,” *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.-ISPRS Arch*, vol. 42, no. 2, pp. 53–59, 2019. doi: [10.5194/isprs-archives-XLII-2-W12-53-2019](https://doi.org/10.5194/isprs-archives-XLII-2-W12-53-2019).
- [5] S. Baran and P. Rola, “Prediction of motor insurance claims occurrence as an imbalanced machine learning problem,” *Articulo Investig.*, vol. 2, no. 3, pp. 77–98, 2022.
- [6] T. Pijl and P. Groenen, “A framework to forecast insurance claims,” *Econ. Res.*, vol. 7, no. 28, pp. 334–347, 2017.
- [7] C. Gomes, Z. Jin, and H. Yang, “Insurance fraud detection with unsupervised deep learning,” *J. Risk Insur.*, vol. 88, no. 3, pp. 591–624, 2021. doi: [10.1111/jori.12359](https://doi.org/10.1111/jori.12359).
- [8] K. Kaushik, A. Bhardwaj, A. D. Dwivedi, and R. Singh, “Article machine learning-based regression framework to predict health insurance premiums,” *Int. J. Environ. Res. Public Health*, vol. 19, no. 13, pp. 7898, 2022. doi: [10.3390/ijerph19137898](https://doi.org/10.3390/ijerph19137898).
- [9] V. Selvakumar, D. K. Satpathi, P. T. V. Praveen Kumar, and V. V. Haragopal, “Predictive modeling of insurance claims using machine learning approach for different types of motor vehicles,” *Univers. J. Account. Financ.*, vol. 9, no. 1, pp. 1–14, 2021. doi: [10.13189/ujaf.2021.090101](https://doi.org/10.13189/ujaf.2021.090101).
- [10] S. Wang, S. Jin, D. Bai, Y. Fan, H. Shi and C. Fernandez, “A critical review of improved deep learning methods for the remaining useful life prediction of lithium-ion batteries,” *Energy Rep.*, vol. 7, no. 9, pp. 5562–5574, 2021. doi: [10.1016/j.egyr.2021.08.182](https://doi.org/10.1016/j.egyr.2021.08.182).
- [11] J. E. van Engelen and H. H. Hoos, “A survey on semi-supervised learning,” *Mach. Learn.*, vol. 109, no. 2, pp. 373–440, 2020. doi: [10.1007/s10994-019-05855-6](https://doi.org/10.1007/s10994-019-05855-6).

- [12] T. Hastie, R. Tibshirani, and J. Friedman, “The elements of statistical learning data mining, inference, and prediction,” in *Overview of Supervised Learning, Springer Series in Statistics*, 2nd ed. California: Stanford, 2009, vol. 26, pp. 9–11.
- [13] S. Abdelhadi, K. Elbahnasy, and M. Abdelsalam, “A proposed model to predict auto insurance claims using machine learning techniques,” *J. Theor. Appl. Inf. Technol.*, vol. 98, no. 22, pp. 3428–3437, 2020.
- [14] N. Vinod and G. Hinton, “Rectified linear units improve restricted boltzmann machines vinod,” *J. Appl. Biomech.*, vol. 33, no. 5, pp. 384–387, 2017. doi: [10.1123/jab.2016-0355](https://doi.org/10.1123/jab.2016-0355).
- [15] D. A. Clevert, T. Unterthiner, and S. Hochreiter, “Fast and accurate deep network learning by exponential linear units (ELUs),” in *4th Int. Conf. Learning Represent. (ICLR), 2016-Conf. Track Proc.*, 2016, pp. 1–14.
- [16] Y. Bai, “RELU-function and derived function review,” in *SHS Web Conf.*, 2022, vol. 144, no. 7, pp. 02006. doi: [10.1051/shsconf/202214402006](https://doi.org/10.1051/shsconf/202214402006).
- [17] M. A. Fauzan and H. Murfi, “The accuracy of XGBoost for insurance claim prediction,” *Int. J. Adv. Soft Comput. Appl.*, vol. 10, no. 2, pp. 159–171, 2018.
- [18] R. D. Burri, R. Burri, R. R. Bojja, and S. R. Buruga, “Insurance claim analysis using machine learning algorithms,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 22, no. 6, pp. 577–582, 2019. doi: [10.35940/ijitee.F1118.0486S419](https://doi.org/10.35940/ijitee.F1118.0486S419).
- [19] J. Pesantez-Narvaez, M. Guillen, and M. Alcañiz, “Predicting motor insurance claims using telematics data—XGboost versus logistic regression,” *Risks*, vol. 7, no. 2, pp. 70, 2019. doi: [10.3390/risks7020070](https://doi.org/10.3390/risks7020070).
- [20] G. Kowshalya and M. U. Nandhini, “Predicting fraudulent claims in automobile insurance,” in *Second Int. Conf. Inventive Commun. Comput. Technol. (ICICCT)*, Coimbatore, India, 2018, pp. 1338–1343. doi: [10.1109/ICICCT.2018.8473034](https://doi.org/10.1109/ICICCT.2018.8473034).
- [21] L. Eason, “Sample insurance claim prediction dataset,” Accessed: Jul. 10, 2023. [Online]. Available: www.kaggle.com/datasets/easonlai/sample-insurance-claim-prediction-dataset
- [22] E. Allibhai, “Building a deep learning model using Keras,” *Towards Data Science*, 2018. Accessed: Jul. 11, 2023. [Online]. Available: towardsdatascience.com/building-a-deep-learning-model-using-keras-1548ca149d37
- [23] J. Brownlee, “A Gentle Introduction to the Rectified Linear Unit (ReLU),” *Optoelectronics and Fiber Optic Technology*, 2019. Accessed: Jul. 13, 2023. [Online]. Available: machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/
- [24] D. Misra, “The swish activation function,” *Paperspace Blog*, 2016. Accessed: Jul. 14, 2023. [Online]. Available: blog.paperspace.com/swish-activation-function/
- [25] S. Sharma, S. Sharma, and A. Athaiya, “Activation functions in neural networks,” *Int. J. Eng. Appl. Sci. Technol.*, vol. 4, no. 12, pp. 310–316, 2020. doi: [10.33564/IJEAST.2020.v04i12.054](https://doi.org/10.33564/IJEAST.2020.v04i12.054).