Explainable AI in Insurance: A Drop in the Ocean?

Bart Custers
University of Hull
Hull, United Kingdom

December 2024

ABSTRACT

The use of Artificial Intelligence (AI) in the insurance industry has significantly evolved over the past two decades, driven by the search for predictive performance and automation. With an increasing availability of data, opaque AI models enable insurance companies to differentiate more between customers, raising concerns about fairness and trust. Explainable AI (XAI) is a key measure to maintain trust in insurance and mitigate unfair practices, as XAI techniques can enhance transparency and interpretability around AI model decisions. This paper presents a systematic literature review of (X)AI applications in insurance and attempts to show to which degree XAI is utilized. A common taxonomy of XAI techniques is used to classify and evaluate 71 reviewed articles, focusing on four key domains of the insurance value chain: pricing and underwriting, risk assessment, claims handling, and fraud detection. The literature review shows widespread implementation of ensemble methods and neural networks, with feature relevance explanations being the predominant XAI technique. A positive development are hybrid models, which constrain the black-box component and retain intrinsic interpretability. To conclude, this paper identifies the trends in (X)AI use cases in insurance and suggests potential further research opportunities.

1 INTRODUCTION

Predictive modelling techniques have been broadly accepted in the insurance industry for many years, where risk prediction and differentiation are essential for maintaining profitability. Traditionally, insurers relied on univariate analysis and linear or logistic regression [47]. Over the past twenty years, the use of artificial intelligence (AI) for prediction tasks has grown significantly [62]. A recent survey amongst more than 400 insurance executives showed that 70% of insurers aim to implement AI models within two years [23].

Contrary, insurance customers might not always accept AI systems. Lawsuits against health insurance companies for example, for using AI tools to wrongly reject healthcare claims [60], could erode trust in AI and the insurance industry. Combined with an increasing availability of data, opaque AI tools enable insurance companies to differentiate more between low and high risks [25]. This raises concerns about fairness, as the use of AI could potentially lead to unfair or discriminatory outcomes [26], particularly for vulnerable customers.

Therefore, explainable AI (XAI) is a key measure to maintain trust in insurance and mitigate unfair practices, as XAI techniques can advance transparency and interpretability around AI model decisions. In a report on AI governance, the European Insurance and Occupational Pensions Authority (EIOPA) identified model explainability as a key governance principle [24]. This paper examines the application of XAI in insurance through a systematic literature review, addressing the following research questions:

- What are the current use cases of AI in insurance?
- To which extent are XAI techniques applied in these use cases?
- What are the main XAI techniques utilized?

This paper contributes to the literature in three ways. First, it provides an updated review of XAI applications. Although [62] conducted a thorough review of XAI applications in insurance, the focus was on articles between 2000 and 2021. As shown in the remainder of this study, many XAI applications in insurance were done in the last two years. Second, this paper focuses exclusively on black-box AI models. Third, this paper investigates the extent to which XAI is applied, by benchmarking XAI techniques across the insurance value chain.

This paper is structured as follows: Section 2 describes the methodology and key definitions, and shows exploratory data analysis on the search results. Section 3 present the findings of the literature review across insurance subdomains. Section 4 presents a discussion of the review results, after which section 5 concludes this paper.

2 RESEARCH METHODOLOGY

2.1 The insurance value chain

The literature review per subdomain is an important contribution of this research, for which several domains along the insurance value chain are explored. The insurance value chain (IVC) has been defined in [25] for example, though this approach is modified in this research for two reasons. Firstly, this research aims to focus on insurance processes that directly affect customers or could potentially have a critical impact on customers. Hence, processes such as marketing, sales and asset management have been ignored. Secondly, to create more focus, pricing and risk prediction were separated (by separating when the keyword 'pricing' was used in the title or abstract), and fraud detection was treated as a separate subdomain. As a result, the following four domains are considered: pricing and underwriting, risk assessment, claims handling, and fraud detection. AI use cases in these subdomains generally aim for increased predictive performance or automation of manual tasks. Further explanation on insurance subdomains and examples of use cases are outlined in [25].

2.2 XAI taxonomy

For reviewing and classifying XAI methodologies this research aligns with the XAI taxonomy as developed by [5]. Focusing on post-hoc explainability, this literature review identifies 'explanation by simplification' methods such as LIME or surrogate models, 'feature relevance explanations' such as SHAP and feature importance (FI), and 'visualization' methods such as partial dependence plots (PDP) and individual conditional expectation (ICE) plots. Furthermore, as highlighted by EIOPA [24], hybrid methods are also recognized as a transparent approach, as they typically contain some transparent elements.

Table 1: List of search strings

Search strings combined with insurance and/or subdomains:
Explainable artificial intelligence (XAI)
Artificial intelligence (AI)
Machine Learning
Neural networks (NN)
Deep learning
Genetic algorithm

2.3 Extraction process and screening

To perform a structured literature search, the databases Google Scholar, IEEE Digital Library and

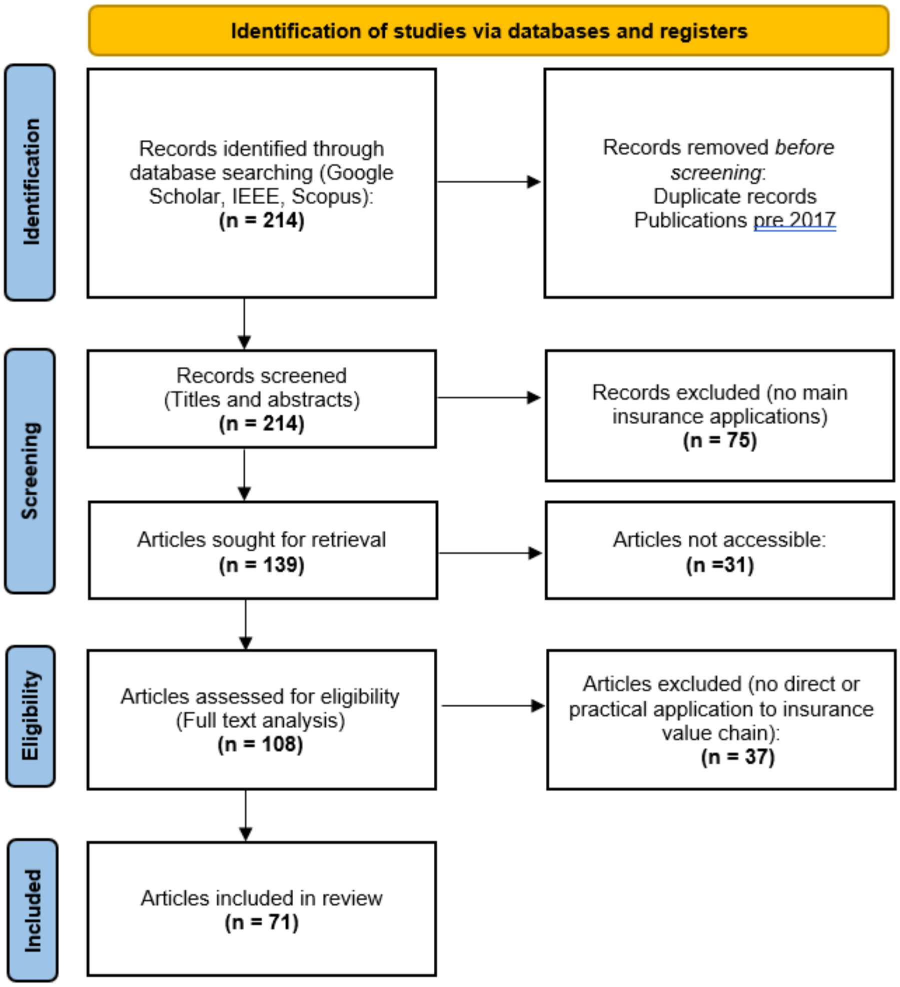
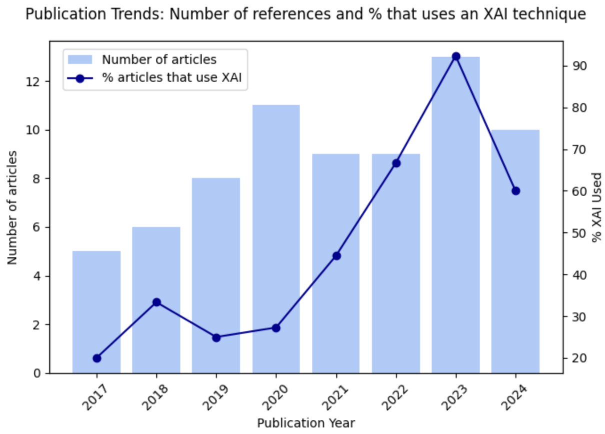


Figure 1: PRISMA Diagram for literature selection process

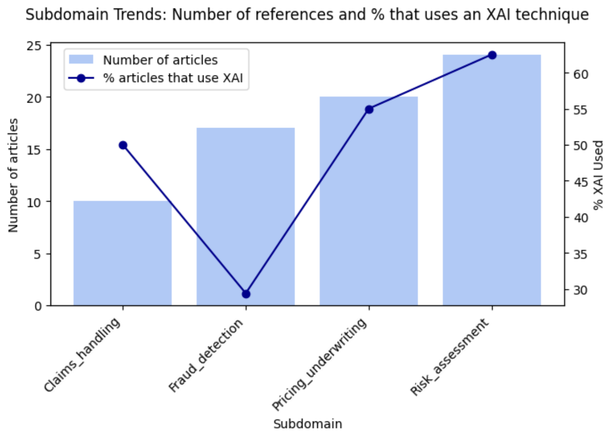
Scopus were explored. Table 1 outlines the search terms used alongside general words such as 'insurance' and the subdomain terms. Intrinsically interpretable models such as linear or logistic regression were not included in the search terms, as this research aims to focus on the use of opaque models such as ensemble learning and neural networks. The structure of the selection process is shown in Figure 1. After deduplicating the search results, articles published before 2017 were removed as XAI only gained increasing attention in the scientific community around 2017 [5]. The resulting 214 articles were screened by titles and abstracts and checked for accessibility. After this process, 108 references remained on which a full text analysis was conducted. After the full text review 37 articles were removed, mainly because they were general reviews that lacked a practical application to the IVC. The resulting 71 articles were selected for the literature review in section 3.

2.4 Exploratory data analysis

The selected articles were reviewed to check whether XAI techniques were applied. An exploratory data analysis was conducted to summarize overall trends. Figure 2a shows the number of references per publication year and the degree to which XAI techniques were used. The plot shows the increased attention to XAI in the last four years. In addition, Figure 2b depicts a similar view across subdomains. Especially among pricing, underwriting and risk assessment more than half of the reviewed articles used XAI. Fraud detection remains slightly behind with around 30% that used XAI.



(a) XAI application over time



(b) XAI application per subdomain

Figure 2: Trends in (X)AI applications in insurance

To obtain overall insight into the main (X)AI methods used, Figure 3a provides a heatmap between the main AI techniques and XAI methods. Dominant techniques are artificial neural networks (ANN) and ensemble learning methods such as Random Forest (RF), Gradient Boosting Machines (GBM) and Extreme Gradient Boosting (XGB). Feature importance is the main XAI approach utilized. Furthermore, Figure 3b illustrates which XAI techniques are used in each subdomain. Here, XAI techniques are summarized (based on [5]) on a higher level for clarity. The next section will discuss the applications in more detail.

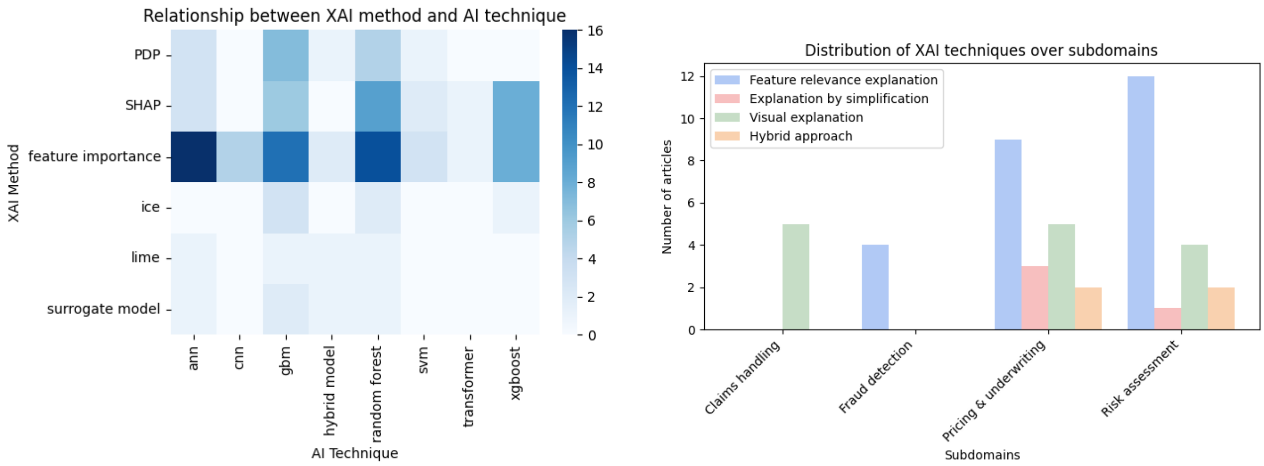
3 LITERATURE REVIEW

3.1 Pricing and underwriting

From the twenty articles examined for pricing and underwriting a majority focuses on applications in the auto insurance domain (11) or life/health domain (6). The studies found in these subdomains all use tabular datasets on which various ensemble learning techniques or neural network approaches are tested, either for price predictions or underwriting decisions. Frequently a range of different models is applied and compared, as done in [19, 74, 22]. In [70, 85, 43, 12], neural network architectures were used to investigate their potential in outperforming traditional models. More uncommon approaches to insurance pricing are [84] and [46]. In [46] reinforcement learning was used for repricing of contracts at renewal. In [84] a genetic algorithm was used for price setting of occupational accident insurance.

Out of the twenty articles reviewed, eleven studies included some form of XAI. As can be seen in Table 2, feature relevance explanations (feature importance and SHAP) and visual explanations (partial dependence plots) are the most used techniques. In [10] only feature importance was used in a heatmap to detect common feature importances for a range of models that predicted underwriting decisions in life insurance. An extensive analysis on model interpretability was performed by [34], where a traditional generalized linear model (GLM) model was compared against RF and XGB for modelling car claims frequency and severity. Feature importance graphs, ICE plots and PDPs were used to compare the workings of different models. Car insurance pricing was also studied in [82] in which the explainability of ANNs was enhanced with feature importance and PDPs.

The SHAP technique is applied in many studies. An interesting approach is [14], in which SHAP was used to examine Irish car insurance premiums and discover the main contributors. SHAP



(a) Heatmap of AI method and XAI technique **(b) Distribution of XAI techniques per subdomain**

Figure 3: Exploration of AI techniques and explanation methods used in insurance

Table 2: Summary of XAI methods in pricing and underwriting studies

Ref.	Model	XAI Tech	Scope	Origin
[34]	RF, GBM	FI, PDP, ICE	Global	Agnostic
[13]	ANN, Transformer, Hybrid	Hybrid	Global	Specific
[39]	RF, GBM	FI, LIME, SHAP	Global, Local	Agnostic
[82]	ANN	FI, PDP	Global	Agnostic
[61]	RF, GBM, XGB	SHAP, ICE	Global	Agnostic
[35]	GBM, ANN, Hybrid	FI, PDP, Surrogate	Global, Local	Specific
[55]	RF, GBM	FI, PDP, Surrogate	Global	Agnostic
[32]	XGB	SHAP	Global	Agnostic
[10]	XGB	FI	Global	Agnostic
[54]	RF	FI, SHAP	Global, Local	Agnostic

was also utilized in [32] to study the predictions of an XGBoost model that analysed the probability that someone would buy the insurance. Explainability for customers was an important concept in [54], in which SHAP was applied to generate feature contributions for translation into a transparent customer dashboard for underwriting decisions. In [61] multiple models were assessed for predicting medical costs for health insurance premiums. SHAP and ICE plots were used for feature importance. Most of the studied articles focus on global explanations, but [39] used both SHAP and LIME to generate local explanations for car insurance premiums.

An XAI technique rarely encountered in this research is the use of surrogate models. In [55], the author used telematics (driving behaviour) data for predicting car claim frequency. Ensemble methods RF and GBM were used, after which an interpretable decision tree was fitted on the outcomes for global explanations. A surrogate model was also used in [35], in which GLM, GBM and ANN approaches were used on car claim datasets from several countries. After modelling the data was grouped based on partial dependence effects and a GLM was fitted to obtain interpretable results. The authors also employed a hybrid approach, the Combined Actuarial Neural Network (CANN), which combines a traditional GLM with a neural network [81]. The approach was to leverage the transparency and interpretability of the GLM while using the neural network's ability to detect complex interactions. By fitting the neural network on top of the GLM predictions, transparency was enhanced as it showed the portion of the prediction explained by the GLM and the additional adjustment by the neural network. The CANN model was also studied in [13] alongside a hybrid approach termed LocalGLMNet. LocalGLMNet also aims to integrate a GLM with a neural network, by allowing the neural network to dynamically fit the GLM coefficients. Additionally, the author included a transformer mechanism, revealing that the transformer further enhances these models by improving the feature interaction mapping.

3.2 Risk assessment

There were twenty-four references found on risk prediction and assessment. As seen in the pricing domain, there are quite some references that examine neural network frameworks [1, 15, 45, 31, 71] without explicitly using methodologies for interpretation and explanation. In [1, 31] the prediction of healthcare claims was investigated. In [15] a neural network optimized with Particle Swarm Optimization (PSO) was used on car insurance claims. An often seen practice is to employ multiple models and compare performances. This was done for car insurance claim predictions in [9, 28], using amongst others RF, GBM, XGB and neural networks, for life insurance risk assessment [40]. In [29], a deep learning framework for agricultural insurance was utilized and benchmarked against traditional methods (GLM) and ensemble methods (RF, GBM). As agricultural insurance is heavily dependent on complex weather patterns, the deep learning structures outperformed other methods, but no further model explanations were provided.

From the twenty-four reviewed articles, fifteen studies used an XAI technique. As can be seen in Table 3, feature importance and SHAP are typically employed for model interpretability, for example in [8, 69, 56] for explanation of RF and XGB predictions. Feature importance and SHAP were also utilized in [52], where a range of models (a.o. RF, GBM, ANN) was tested on different datasets (life, health and occupational risk). In [76], several models (GBM, XGB, CatBoost) were combined with a transformer technique similar to [13] in order to improve the interaction mapping. Models were analysed with feature importance and SHAP.

SHAP was also used in [49], in which car telematics data was used to identify low-risk and risky drivers, employing model techniques such as Support Vector Machine (SVM) and XGBoost. Furthermore, SHAP values were then used to generate a counterfactual model that could advise drivers on how to improve their driving score, hence providing a practical and useful example of how XAI methods could benefit customers. Another combination frequently observed is feature importance and partial dependence plots (PDP). In [77] several tree-based models were utilized on claim size prediction of car insurance. In [66] a range of models was tested on multiple property insurance coverages from a US local government insurance fund. Next to RF and GBM, multivariate decision trees were tested.

Claim frequency prediction for car insurance was the main topic of [38, 17] and explained by feature importance, PDPs and ICE plots [17]. In [7], claim frequency of car insurance claims was also the main topic, but only explained with feature importance. Both [33, 78] employed the CANN model similar to some references in the pricing domain. This hybrid modelling of GLM and neural network aims to retain the transparency of the GLM and the interaction detection enhancement of

Table 3: Summary of XAI methods in risk assessment studies

Ref.	Model	XAI Tech	Scope	Origin
[52]	RF, GBM, ANN	FI, SHAP	Global	Agnostic
[38]	RF, SVM, GBM, ANN	FI, PDP	Global	Agnostic
[57]	XGB, TabNet	FI, Decision Masks	Global	Specific
[33]	ANN, Hybrid	Hybrid	Global	Agnostic
[3]	RF, GBM	FI (LOCO, VNS), SHAP (TE)	Global	Agnostic
[49]	SVM, XGB	FI, SHAP, Counterfact.	Global, Local	Agnostic
[76]	GBM, XGB, ANN, Transformer	FI, SHAP	Global	Agnostic
[78]	ANN, Hybrid	FI, LIME	Local	Agnostic
[8, 56]	RF, XGB	FI, SHAP	Global, Local	Agnostic
[69]	RF, XGB	FI, SHAP	Global	Agnostic
[17]	GBM	FI, PDP, ICE	Global	Agnostic
[7]	RF, XGB, ANN	FI	Global	Agnostic
[66, 77]	RF, GBM	FI, PDP	Global	Agnostic

the neural network. Specifically, in [33] the CANN was only exploited to detect and list the most promising (pairwise) interactions, which could then be added to the GLM. Although this approach is very transparent, as the black-box model is only used in assisting the intrinsically interpretable method, the manual character of adding the interactions in the GLM could be considered a disadvantage.

Another approach to improve the explainability of neural network methods was shown in [57], in which the authors applied the TabNet model on car telematics data. The TabNet model [4] was specifically developed for tabular data and thus especially suitable for insurance modelling. It aims to approximate the explainability of simpler tree-based methods and combines this with the effectiveness of (deep) neural networks. A specific technique used within the TabNet structure is sequential attention, which produces interior decision masks that show which features were used for model decisions.

In [3], the authors investigated prediction of car claims with severe injuries, utilizing RF and GBM. To detect the most important features, several methods were applied: Leave-One-Covariate-Out (LOCO), TreeExplainer (TE) and Variable Neighborhood Search (VNS). The LOCO method iteratively removes a feature from the model after which the impact on the model is assessed. The TE method basically calculates SHAP values. Originally, the VNS approach is an optimization approach and employed here to optimize for the set of features with the highest impact.

3.3 Claims handling

Of the ten articles found on claims handling, all aimed to automate (part of) this process and all focused on vehicle damage detection for car insurance claims. Typically a Convolutional Neural Network (CNN) is employed for this task, as shown in [44, 48, 16, 41]. In [50], the authors proposed a two-step approach, in which step one involved identifying the vehicle and cropping the image, while step two aimed to detect the damage. For five out of ten references, one could label them as using XAI, as they included some form of masking that depicts the prediction in the image. Transfer learning with existing models (e.g. Resnet, EfficientNet) was used in [63, 80]. Model explanations were provided by means of damage localization, i.e. masking the damaged areas as detected by the model. In [64] multiple models were applied with a PSO optimization to obtain the optimal number of models and find confidence thresholds: predictions that did not reach the threshold were excluded. With the predictions, 'bounding boxes' were created that identified the damaged part of the vehicle. Regarding explainability, [65, 73] demonstrated how model predictions can be integrated into a transparent dashboard that can be used by claim managers.

3.4 Fraud detection

The seventeen articles examined for fraud detection applications in insurance show the potential for partly or fully automated systems. A common approach is the use of ensemble methods (RF, GBM, XGB) on tabular claims data [51, 21, 20, 68]. Furthermore, SVM techniques potentially outperformed traditional models such as logistic regression in [59, 67, 6]. The predictive power of neural networks was demonstrated in [83], in which a neural network was fitted on auto insurance claims and tuned with a genetic algorithm.

A combination of structured and unstructured data was used in a deep learning framework in [79], in which Latent Dirichlet Allocation (LDA) topic modelling was used on text data from claim files, to enrich already available structured data. In [42], the authors focused on the imbalanced data problem and optimal threshold setting of several deep learning methods. Usually, fraudulent claims are a minimal proportion of total claims, leading to imbalanced datasets. In [2], a deep learning method was combined with Monte Carlo to quantify the uncertainty in point predictions from the neural network. Although no specific XAI techniques were applied, this framework provides insight when a model can be trusted (i.e. when a desired confidence level is reached). A completely different approach would be an unsupervised approach such as Fuzzy C-Means clustering, as examined in [75].

Explainability techniques used in fraud detection are summarized in Table 4. An extensive XAI analysis was performed in [72], in which a range of predictive modelling techniques, from logistic regression to neural networks, was used for fraud detection in Brazilian property insurance claims. In the evaluation, this study first showed global model explanations using feature importance graphs. Secondly, the authors showed local explanations with SHAP for false positive and false negative predictions. In [18] both unsupervised learning methods (isolation forests) and supervised methods were used (XGBoost), evaluated with SHAP feature importance.

In [27], an attention network was used to detect health insurance fraud. As health case data can vary in number of variables for example, the authors proposed the attention network approach as it could potentially utilize the variety in variables better as compared to methods such as gradient boosting. The attention weights were used as a means of model explanation, as they depict the most relevant features. An unsupervised deep learning approach was the use of an autoencoder [30]. Without labeled data, this mechanism was expected to detect anomalies in the data that could point at fraud. Variable importance graphs were then used to find the most important drivers of fraud.

Finally, the only article that used image data was [53]. Here, the approach was to detect fraud by predicting whether a vehicle damage was already claimed before. The deep learning framework, based on person identification approaches, aimed to recognize the damaged vehicle and classify whether a claim was already processed in the past. For interpretability, the relevant features for decision making were projected on the images.

Table 4: **Summary of XAI methods in fraud detection studies**

Ref.	Model	XAI Tech	Scope	Origin
[53, 30]	ANN	FI	Local	Agnostic
[27]	ANN	Attention weights	Global	Specific
[18]	XGB	SHAP	Global	Agnostic
[72]	RF, SVM, GBM, ANN	FI, SHAP	Global, Local	Agnostic

4 DISCUSSION

XAI in the insurance domain cannot be portrayed as a 'drop in the ocean'. As shown in Figure 1, academic studies in the insurance domain are increasingly incorporating XAI techniques. A positive development is the introduction of hybrid approaches, which are slowly being adopted in pricing and risk assessment. The CANN model is particularly valuable as it combines a transparent GLM with predictive enhancements from an ANN, ensuring transparent rating structures. Moreover, this hybrid approach constrains the black-box component, reducing the risk of unfair practices. This could maintain trustworthiness while still benefiting insurers. From an XAI perspective, other approaches such as TabNet and LocalGLMNet can also be regarded as advances, as they intent to address the black-box nature of some AI models.

However, many of the XAI applications reviewed in this paper are still largely limited to global feature importance techniques. In these cases, black-box models are only shallowly explained through feature contributions, with the focus of XAI techniques leaning more toward model interpretability rather than model explainability. Moreover, research [36, 37] has shown that feature importance techniques such as SHAP could produce misleading results and might ignore interaction effects. Frequently used partial dependence plots also have limitations as correlations with other features are ignored for these plots [58].

Finally, the fraud detection references in this paper are lagging in XAI applications. Both fraud detection and claims handling are also applying solely visual techniques or feature relevance techniques. Here, hybrid approaches such as neuro-symbolic AI [11] could further enhance explainability and trustworthiness. In summary, the following recommendations for XAI in insurance are suggested:

- Enhance the adoption of hybrid model frameworks, thus constraining the influence of black-box models on predictions.
- Move beyond the shallow use of feature relevance techniques, by considering their shortcomings and implementing local explanations to improve explainability.
- Explore hybrid approaches such as neuro-symbolic AI, specifically in the claims handling and fraud detection domain. These domains lacked adoption of hybrid approaches and especially in these domains customers might be more likely to demand explanations for model decisions.

This literature review has certain limitations that future research could address. A significant number of articles (31) were inaccessible, thus limiting the scope of this review and influencing the reported percentage of papers utilizing XAI. Future research could broaden the scope by extending the list of search strings to identify additional XAI applications. Furthermore, future research could extend this paper's framework to other insurance subdomains such as marketing, claim reserving or asset management.

5 CONCLUSION

This research aims to present the main (X)AI use cases in insurance and evaluate the extent to which XAI techniques are utilized. In the insurance industry, it is imperative to deploy AI carefully to mitigate unfair practices, maintain trust, and increase acceptance of AI tools. While explainable AI alone might not guarantee careful deployment, it could serve as a first line of defense. By providing interpretability and explainability for black-box models, insurance companies can obtain more confidence in AI models through a deeper understanding of model decisions. From a regulatory perspective, EIOPA has recognized this importance by defining XAI as a key governance principle [24].

The literature review encompassed 71 articles published between 2017 and 2024, focusing on key insurance processes: pricing and underwriting, risk assessment, claims handling, and fraud detection. General statistics have been derived from the literature review, indicating a growing trend in XAI applications over time. Furthermore, the exploratory analysis shows the main modelling approaches (ensemble models and neural networks) versus the dominant XAI techniques. Feature relevance explanations, such as feature importance and SHAP, are dominant in the reviewed AI models. Visual explanations such as PDP and ICE are also frequently applied. In addition, hybrid approaches such as CANN appear promising in enhancing explainability, as they blend transparent models with the predictive power of neural networks. Hybrid approaches have not been detected in claims handling and fraud detection, but could be explored in future research.

Although this review indicates that many of the articles published in the last two years use some XAI technique, one should be careful not to be overoptimistic. Firstly, some of the employed XAI techniques such as feature importance may provide shallow and misleading insights, raising questions whether they provide sufficiently meaningful explanations for model decisions. Secondly, the percentage of articles that apply XAI may be biased as the literature search was limited to the search terms in Table 1. Moreover, the inaccessibility of several articles means that important perspectives might be ignored. Future research could broaden the scope of search terms and explore additional subdomains to present an even broader perspective. Furthermore, continuous monitoring of XAI developments in the insurance industry is essential to ensure the development of trustworthy and explainable AI solutions.

ABBREVIATIONS

ANN	Artificial Neural Network
CANN	Combined Actuarial Neural Network
CNN	Convolutional Neural Network
EIOPA	European Insurance and Occupational Pensions Authority
FI	Feature Importance
ICE	Individual Conditional Expectation
GBM	Gradient Boosting Machine
GLM	Generalized Linear Model
LIME	Local Interpretable Model-Agnostic Explanations
PDP	Partial Dependence Plot
PSO	Particle Swarm Optimization
RF	Random Forest
SHAP	SHapley Additive exPlanations
SVM	Support Vector Machine
XGB	Extreme Gradient Boosting

REFERENCES

[1] Umar Isa Abdulkadir and Anil Fernando. 2024. A Deep Learning Model for Insurance Claims Predictions. *Journal on Artificial Intelligence* 6, 1 (2024), 71–83. https://doi.org/10.32604/jai.2024.045332

[2] Mouad Ablad, Bouchra Frikh, and Brahim Ouhbi. 2020. Uncertainty quantification in deep learning context: Application to insurance. In *6th IEEE Congress on Information Science and Technology (CiSt)* (Agadir - Essaouira, Morocco), Vol. 2020-June. IEEE, 110–115. https://doi.org/10.1109/CiSt49399.2021.9357201

[3] Mostafa Amini, Ali Bagheri, and Dursun Delen. 2022. Discovering injury severity risk factors in automobile crashes: A hybrid explainable AI framework for decision support. *Reliability Engineering and System Safety* 226 (10 2022), 108720. https://doi.org/10.1016/j.ress.2022.108720

[4] Sercan Arik and Tomas Pfister. 2021. TabNet: Attentive Interpretable Tabular Learning. In *The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21)*. 6679–6687. https://doi.org/10.1609/aaai.v35i8.16826

[5] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58 (6 2020), 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

[6] Faheem Aslam, Ahmed Imran Hunjra, Zied Ftiti, Wael Louhichi, and Tahira Shams. 2022. Insurance Fraud Detection: Evidence from Artificial Intelligence and Machine Learning. *Research in International Business and Finance* 62 (2022), 101744. https://doi.org/10.1016/j.ribaf.2022.101744

[7] Sebastian Baran and Przemysław Rola. 2022. Prediction of motor insurance claims occurrence as an imbalanced machine learning problem. (4 2022), 1–12. arXiv:abs/2204.0 http://arxiv.org/abs/2204.06109

[8] Lluís Bermúdez, David Anaya, and Jaume Belles-Sampera. 2023. Explainable AI for paid-up risk management in life insurance products. *Finance Research Letters* 57 (11 2023), 104242. https://doi.org/10.1016/j.frl.2023.104242

[9] Spoorthi Bhoji, Trilok Nath Pandey, Bibhuti Bhusan Dash, Rabinarayan Satpathy, Utpal Chandra De, and Sudhansu Shekhar Patra. 2024. Performance Analysis of Classifiers in Predicting Car Insurance Claim. In *2024 8th International Conference on Inventive Systems and Control, ICISC*. IEEE, 33–38. https://doi.org/10.1109/ICISC62624.2024.00013

[10] Rhys Biddle, Shaowu Liu, Peter Tilocca, and Guandong Xu. 2018. Automated underwriting in life insurance: Predictions and optimisation. In *Proceedings of the Australasian Database Conference* (Brisbane, Australia), Vol. 10837 LNCS. 135–146. https://doi.org/10.1007/978-3-319-92013-9_11

[11] Tim Bohne, Anne Kathrin Patricia Windler, and Martin Atzmueller. 2023. A Neuro-Symbolic Approach for Anomaly Detection and Complex Fault Diagnosis Exemplified in the Automotive Domain. In *K-CAP 2023 - Proceedings of the 12th Knowledge Capture Conference*. ACM, 35–43. https://doi.org/10.1145/3587259.3627546

[12] Noorhannah Boodhun and Manoj Jayabalan. 2018. Risk prediction in life insurance industry using supervised learning algorithms. *Complex Intelligent Systems* 4, 2 (6 2018), 145–154. https://doi.org/10.1007/s40747-018-0072-1

[13] Alexej Brauer. 2024. Enhancing actuarial non-life pricing models via transformers. *European Actuarial Journal* 14 (6 2024), 991–1012. https://doi.org/10.1007/s13385-024-00388-2

[14] Adrian Byrne. 2024. *Pricing Risk: An XAI Analysis of Irish Car Insurance Premiums*. Vol. 2156 CCIS. Springer, 315–330. https://doi.org/10.1007/978-3-031-63803-9_17

[15] Xi Chen, Liang Xu, Yalong Wang, Xiang Zhai, and Xiang Guo. 2020. Application of Improved PSO-ELM in Auto Insurance Customer Risk Level Prediction. In *2020 IEEE International Conference on Mechatronics and Automation (ICMA)* (Beijing, China). IEEE, 979–983. https://doi.org/10.1109/ICMA49215.2020.9233591

[16] Abraham C. Chua, Christian Rei B. Mercado, John Phillip R. Pin, Angelo Kyle T. Tan, Jose Benito L. Tinhay, Elmer P. Dadios, and Robert Kerwin C. Billones. 2021.

Damage Identification of Selected Car Parts Using Image Classification and Deep Learning. In *2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)* (Manila, Philippines). IEEE, 1–5. https://doi.org/10.1109/HNICEM54116.2021.9731806

[17] Carina Clemente, Gracinda R. Guerreiro, and Jorge M. Bravo. 2023. Modelling Motor Insurance Claim Frequency and Severity Using Gradient Boosting. *Risks* 11, 9 (9 2023), 163. https://doi.org/10.3390/risks11090163

[18] Jörn Debener, Volker Heinke, and Johannes Kriebel. 2023. Detecting insurance fraud using supervised and unsupervised machine learning. *Journal of Risk and Insurance* 90, 3 (9 2023), 743–768. https://doi.org/10.1111/jori.12427

[19] Michel Denuit, Arthur Charpentier, and Julien Trufin. 2021. Autocalibration and Tweedie-dominance for insurance pricing with machine learning. *Insurance: Mathematics and Economics* 101 (11 2021), 485–497. https://doi.org/10.1016/j.insmatheco.2021.09.001

[20] Najmeddine Dhieb, Hakim Ghazzai, Hichem Besbes, and Yehia Massoud. 2019. Extreme gradient boosting machine learning algorithm for safe auto insurance operations. In *2019 IEEE International Conference on Vehicular Electronics and Safety (ICVES)* (Cairo, Egypt). IEEE, 1–5. https://doi.org/10.1109/ICVES.2019.8906396

[21] Najmeddine Dhieb, Hakim Ghazzai, Hichem Besbes, and Yehia Massoud. 2020. A Secure AI-Driven Architecture for Automated Insurance Systems: Fraud Detection and Risk Measurement. *IEEE Access* 8 (2020), 58546–58558. https://doi.org/10.1109/ACCESS.2020.2983300

[22] Aman Dubey, Tejisman Parida, Akshay Birajdar, Ajay Kumar Prajapati, and Sagar Rane. 2018. Smart Underwriting System: An Intelligent Decision Support System for Insurance Approval Risk Assessment. In *2018 3rd International Conference for Convergence in Technology (I2CT)* (Pune, India). IEEE, 1–6. https://doi.org/10.1109/I2CT.2018.8529792

[23] Earnix. 2024. Earnix Industry Trends Report 2024. https://earnix.com/insights-center/brochures/earnix-2024-industry-trends-report/

[24] EIOPA. 2021. Artificial Intelligence Governance Principles: Towards Ethical and Trustworthy Artificial Intelligence in the European Insurance Sector. https://doi.org/10.2854/49874

[25] Martin Eling, Davide Nuessle, and Julian Staubli. 2022. The impact of artificial intelligence along the insurance value chain and on the insurability of risks. *Geneva Papers on Risk and Insurance: Issues and Practice* 47 (4 2022), 205–241. https://doi.org/10.1057/s41288-020-00201-7

[26] Alessandro Fabris, Alan Mishler, Stefano Gottardi, Mattia Carletti, Matteo Daicampi, Gian Antonio Susto, and Gianmaria Silvello. 2021. Algorithmic Audit of Italian Car Insurance: Evidence of Unfairness in Access and Pricing. (7 2021). https://doi.org/10.1145/3461702.3462569 arXiv:2105.10174

[27] Helmut Farbmacher, Leander Löw, and Martin Spindler. 2022. An explainable attention network for fraud detection in claims management. *Journal of Econometrics* 228, 2 (6 2022), 244–258. https://doi.org/10.1016/j.jeconom.2020.05.021

[28] Muhammad Arief Fauzan and Hendri Murfi. 2018. The Accuracy of XGBoost for Insurance Claim Prediction. *International Journal of Advances in Soft Computing and Its Applications* 10, 2 (2018), 159–171. https://api.semanticscholar.org/CorpusID:202775972

[29] Azar Ghahari, Nathaniel K. Newlands, Vyacheslav Lyubchich, and Yulia R. Gel. 2019. Deep Learning at the Interface of Agricultural Insurance Risk and Spatio-Temporal Uncertainty in Weather Extremes. *North American Actuarial Journal* 23, 4 (10 2019), 535–550. https://doi.org/10.1080/10920277.2019.1633928

[30] Chamal Gomes, Zhuo Jin, and Hailiang Yang. 2021. Insurance fraud detection with unsupervised deep learning. *Journal of Risk and Insurance* 88, 3 (9 2021), 591–624. https://doi.org/10.1111/jori.12359

[31] Sam Goundar, Suneet Prakash, Pranil Sadal, and Akashdeep Bhardwaj. 2020. Health Insurance Claim Prediction Using Artificial Neural Networks. *International Journal of System Dynamics Applications* 9, 3 (6 2020), 40–57. https://doi.org/10.4018/ijsda.2020070103

[32] Alex Gramegna and Paolo Giudici. 2020. Why to buy insurance? An explainable artificial intelligence approach.

Risks 8, 4 (12 2020), 137. https://doi.org/10.3390/risks8040137

[33] Yevhen Havrylenko and Julia Heger. 2024. Detection of interacting variables for generalized linear models via neural networks. *European Actuarial Journal* 14 (8 2024), 551–580. https://doi.org/10.1007/s13385-023-00362-4

[34] Roel Henckaerts, Marie-Pier Côté, Katrien Antonio, and Roel Verbelen. 2020. Boosting insights in insurance tariff plans with tree-based machine learning methods. (4 2020). arXiv:1904.10890 http://arxiv.org/abs/1904.10890

[35] Freek Holvoet, Katrien Antonio, and Roel Henckaerts. 2023. Neural networks for insurance pricing with frequency and severity data: a benchmark study from data preprocessing to technical tariff. (10 2023). arXiv:2310.12671 http://arxiv.org/abs/2310.12671

[36] Danial Hooshyar and Yeongwook Yang. 2024. Problems with SHAP and LIME in interpretable AI for education: A comparative study of post-hoc explanations and neural-symbolic rule extraction. *IEEE Access* 12 (2024), 137472–137490. https://doi.org/10.1109/ACCESS.2024.3463948

[37] Xuanxiang Huang and Joao Marques-Silva. 2023. The Inadequacy of Shapley Values for Explainability. (2 2023). arXiv:2302.08160 http://arxiv.org/abs/2302.08160

[38] Yifan Huang and Shengwang Meng. 2019. Automobile insurance classification ratemaking based on telematics driving data. *Decision Support Systems* 127 (12 2019), 113156. https://doi.org/10.1016/j.dss.2019.113156

[39] Umar Abdulkadir Isa and Anil Fernando. 2023. Explainable AI and Interpretable Model for Insurance Premium Prediction. *Journal on Artificial Intelligence* 5 (5 2023), 31–42. https://doi.org/10.21203/rs.3.rs-2364785/v1

[40] Rachna Jain, Jafar A. Alzubi, Nikita Jain, and Pawan Joshi. 2019. Assessing risk in life insurance using ensemble learning. *Journal of Intelligent and Fuzzy Systems* 37, 2 (2019), 2969–2980. https://doi.org/10.3233/JIFS-190078

[41] Amirreza Jalili, Hamed Tabrizchi, and Bagher Babaali. 2024. EfficientNet-based vehicle damage insurance verification. In *2024 20th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP)* (Babol, Iran). IEEE. https://doi.org/10.1109/AISP61396.2024.10475312

[42] Justin M. Johnson and Taghi M. Khoshgoftaar. 2019. Medicare fraud detection using neural networks. *Journal of Big Data* 6 (12 2019), 63. https://doi.org/10.1186/s40537-019-0225-0

[43] Keshav Kaushik, Akashdeep Bhardwaj, Ashutosh Dhar Dwivedi, and Rajani Singh. 2022. Machine Learning-Based Regression Framework to Predict Health Insurance Premiums. *International Journal of Environmental Research and Public Health* 19, 13 (7 2022), 7898. https://doi.org/10.3390/ijerph19137898

[44] Maleika Heenaye Mamode Khan, Mohammad Zafir Hussein Sk Heerah, and Zuhairah Basgeeth. 2021. Automated detection of multi-class vehicle exterior damages using deep learning. In *International Conference on Electrical, Computer, Communications and Mechatronics Engineering, ICECCME*. IEEE, 1. https://doi.org/10.1109/ICECCME52200.2021.9590927

[45] Mark Kiermayer and Christian Weiß. 2021. Grouping of Contracts in Insurance using Neural Networks. *Scandinavian Actuarial Journal* 4 (11 2021), 295–322. http://arxiv.org/abs/1912.09964

[46] Elena Krasheninnikova, Javier García, Roberto Maestre, and Fernando Fernández. 2019. Reinforcement learning for pricing strategy optimization in the insurance industry. *Engineering Applications of Artificial Intelligence* 80 (4 2019), 8–19. https://doi.org/10.1016/j.engappai.2019.01.010

[47] Kevin Kuo and Daniel Lupton. 2023. Towards Explainability of Machine Learning Models in Insurance Pricing. *Variance* 16, 1 (3 2023). http://arxiv.org/abs/2003.10674

[48] Donggeun Lee, Juyeob Lee, and Eunil Park. 2024. Automated vehicle damage classification using the three-quarter view car damage dataset and deep learning approaches. *Heliyon* 10, 14 (7 2024). https://doi.org/10.1016/j.heliyon.2024.e34016

[49] Hong Jie Li, Xing Gang Luo, Zhong Liang Zhang, Wei Jiang, and Shen Wei Huang. 2023. Driving risk prevention in usage-based insurance services based on interpretable machine learning and telematics data. *Decision Support Systems* 172 (9 2023), 113985. https://doi.org/10.1016/j.dss.2023.113985

[50] Lin Li, Koshin Ono, and Chun Kit Ngan. 2021. A Deep Learning and Transfer Learning Approach for Vehicle Damage Detection. In *Proceedings of the International Florida Artificial Intelligence Research Society Conference, FLAIRS*, Vol. 34. Florida Online Journals, University of Florida. https://doi.org/10.32473/flairs.v34i1.128473

[51] Yaqi Li, Chun Yan, Wei Liu, and Maozhen Li. 2018. A principle component analysis-based random forest with the potential nearest neighbor method for automobile insurance fraud identification. *Applied Soft Computing Journal* 70 (9 2018), 1000–1009. https://doi.org/10.1016/j.asoc.2017.07.027

[52] Catalina Lozano-Murcia, Francisco P. Romero, Jesus Serrano-Guerrero, Arturo Peralta, and Jose A. Olivas. 2024. Potential Applications of Explainable Artificial Intelligence to Actuarial Problems. *Mathematics* 12 (3 2024), 635. https://doi.org/10.3390/math12050635

[53] Luca Maiano, Antonio Montuschi, Marta Caserio, Egon Ferri, Federico Kieffer, Chiara Germanò, Lorenzo Baiocco, Lorenzo Ricciardi Celsi, Irene Amerini, and Aris Anagnostopoulos. 2023. A deep-learning–based antifraud system for car-insurance claims. *Expert Systems with Applications* 231 (11 2023), 120644. https://doi.org/10.1016/j.eswa.2023.120644

[54] Marc Maier, Hayley Carlotto, Sara Saperstein, Freddie Sanchez, Sherriff Balogun, and Sears Merritt. 2020. Improving the Accuracy and Transparency of Underwriting with Artificial Intelligence to Transform the Life-Insurance Industry. *AI Magazine* 41(3) (2020), 78–93. https://doi.org/10.1609/aimag.v41i3.5320

[55] Arthur Maillart. 2021. Toward an explainable machine learning model for claim frequency: a use case in car insurance pricing with telematics data. *European Actuarial Journal* 11, 2 (12 2021), 579–617. https://doi.org/10.1007/s13385-021-00270-5

[56] Leandro Masello, German Castignani, Barry Sheehan, Montserrat Guillen, and Finbarr Murphy. 2023. Using contextual data to predict risky driving events: A novel methodology from explainable artificial intelligence. *Accident Analysis and Prevention* 184 (5 2023). https://doi.org/10.1016/j.aap.2023.106997

[57] Kevin McDonnell, Finbarr Murphy, Barry Sheehan, Leandro Masello, and German Castignani. 2023. Deep learning in insurance: Accuracy and model interpretability using TabNet. *Expert Systems with Applications* 217 (5 2023), 119543. https://doi.org/10.1016/j.eswa.2023.119543

[58] Christoph Molnar. 2022. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* (2nd ed.). https://christophm.github.io/interpretable-ml-book

[59] Brian Ndirangu Muthura and Abraham Matheka. 2023. A Hybrid Model for Detecting Insurance Fraud Using K-Means and Support Vector Machine Algorithms. *Open Journal for Information Technology* 6, 2 (11 2023), 143–156. https://doi.org/10.32591/coas.ojit.0602.05143m

[60] Elizabeth Napolitano. 2023. *MoneyWatch Lawsuits take aim at use of AI tool by health insurance companies to process claims.* Retrieved December 23, 2024 from https://www.cbsnews.com/news/health-insurance-humana-united-health-ai-algor

[61] Ugochukwu Orji and Elochukwu Ukwandu. 2024. Machine learning for an explainable cost prediction of medical insurance. *Machine Learning with Applications* 15 (3 2024), 100516. https://doi.org/10.1016/j.mlwa.2023.100516

[62] Emer Owens, Barry Sheehan, Martin Mullins, Martin Cunneen, Juliane Ressel, and German Castignani. 2022. Explainable Artificial Intelligence (XAI) in Insurance. *Risks* 10 (12 2022), 230. https://doi.org/10.3390/risks10120230

[63] Kalpesh Patil, Mandar Kulkarni, Anand Sriraman, and Shirish Karande. 2017. Deep learning based car damage classification. In *16th IEEE International Conference on Machine Learning and Applications (ICMLA)* (Cancun, Mexico), Vol. 2017-December. IEEE, 50–54. https://doi.org/10.1109/ICMLA.2017.0-179

[64] Sergio A. Pérez-Zarate, Daniel Corzo-García, Jose L. Pro-Martín, Juan A. Álvarez García, Miguel A. Martínez del Amor, and David Fernández-Cabrera. 2024. Automated Car Damage Assessment Using Computer Vision: Insurance Company Use Case. *Applied Sciences* 14, 20 (10 2024). https://doi.org/10.3390/app14209560

[65] Jihad Qaddour and Syeda Ayesha Siddiqa. 2023. Automatic damaged vehicle estimator using enhanced deep learning algorithm. *Intelligent Systems with Applications*

18 (5 2023). https://doi.org/10.1016/j.iswa.2023.200192

[66] Zhiyu Quan and Emiliano A. Valdez. 2018. Predictive analytics of insurance claims using multivariate decision trees. *Dependence Modeling* 6, 1 (12 2018), 377–407. https://doi.org/10.1515/demo-2018-0022

[67] Seema Rawat, Aakankshu Rawat, Deepak Kumar, and A. Sai Sabitha. 2021. Application of machine learning and data visualization techniques for decision support in the insurance sector. *International Journal of Information Management Data Insights* 1, 2 (11 2021), 100012. https://doi.org/10.1016/j.jjimei.2021.100012

[68] Laiqa Rukhsar, Waqas Haider Bangyal, Kashif Nisar, and Sana Nisar. 2022. Prediction of Insurance Fraud Detection using Machine Learning Algorithms. *Mehran University Research Journal of Engineering and Technology* 41 (1 2022), 33–40. https://doi.org/10.22581/muet1982.2201.04

[69] Rahul Sahai, Ali Al-Ataby, Sulaf Assi, Manoj Jayabalan, Panagiotis Liatsis, Chong Kim Loy, Abdullah Al-Hamid, Sahar Al-Sudani, Maitham Alamran, and Hoshang Kolivand. 2023. *Insurance Risk Prediction Using Machine Learning*. Vol. 165. Springer, 419–433. https://doi.org/10.1007/978-981-99-0741-0_30

[70] K. M. Sakthivel and Ch. Rajitha. 2017. Artificial Intelligence for Estimation of Future Claim Frequency in Non-Life Insurance. *Global Journal of Pure and Applied Mathematics* 13, 6 (2017), 1701–1710. https://api.semanticscholar.org/CorpusID:44183213

[71] Vigneshwaran Selvakumar, Dipak Kumar, Satpathi, and P. T. V. Praveen Kumar. 2020. Modeling and Predicting of Motor Insurance Claim Amount using Artificial Neural Network. *International Journal of Recent Technology and Engineering* 8, 6 (2020). https://doi.org/10.35940/ijrte.F9873.038620

[72] Matheus Kempa Severino and Yaohao Peng. 2021. Machine learning algorithms for fraud prediction in property insurance: Empirical evidence using real-world microdata. *Machine Learning with Applications* 5 (9 2021), 100074. https://doi.org/10.1016/j.mlwa.2021.100074

[73] Ranjodh Singh, Meghna P. Ayyar, Tata Venkata Sri Pavan, Sandeep Gosain, and Rajiv Ratn Shah. 2019. Automating Car Insurance Claims Using Deep Learning Techniques. In *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)* (Singapore). IEEE, 199–207. https://doi.org/10.1109/BigMM.2019.00-25

[74] Giorgio Alfredo Spedicato, Christophe Dutang, and Leonardo Petrini. 2018. Machine Learning Methods to Perform Pricing Optimization. A Comparison with Standard GLMs. *Variance* 12 (2018), 69–89. https://www.researchgate.net/publication/323202093

[75] Sharmila Subudhi and Suvasini Panigrahi. 2020. Use of optimized Fuzzy C-Means clustering and supervised classifiers for automobile insurance fraud detection. *Journal of King Saud University - Computer and Information Sciences* 32, 5 (6 2020), 568–575. https://doi.org/10.1016/j.jksuci.2017.09.010

[76] Tiejiang Sun, Jingyun Yang, Jiale Li, Jiaying Chen, Mingyue Liu, Li Fan, and Xukang Wang. 2024. Enhancing Auto Insurance Risk Evaluation With Transformer and SHAP. *IEEE Access* 12 (2024), 116546–116557. https://doi.org/10.1109/ACCESS.2024.3446179

[77] Edossa Merga Terefe. 2023. Tree-Based Machine Learning Methods For Vehicle Insurance Claims Size Prediction. (2 2023). arXiv:2302.10612 http://arxiv.org/abs/2302.10612

[78] George Tzougas and Konstantin Kutzkov. 2023. Enhancing Logistic Regression Using Neural Networks for Classification in Actuarial Learning. *Algorithms* 16, 2 (2 2023), 99. https://doi.org/10.3390/a16020099

[79] Yibo Wang and Wei Xu. 2018. Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. *Decision Support Systems* 105 (1 2018), 87–95. https://doi.org/10.1016/j.dss.2017.11.001

[80] Daniel Widjojo, Endang Setyati, and Yosi Kristian. 2022. Integrated Deep Learning System for Car Damage Detection and Classification Using Deep Transfer Learning. In *IEEE 8th Information Technology International Seminar (ITIS)* (Surabaya, Indonesia). IEEE, 21–26. https://doi.org/10.1109/ITIS57155.2022.10010292

[81] Mario V. Wüthrich and Michael Merz. 2018. EDITORIAL: YES, WE CANN! *ASTIN Bulletin* 49 (2018), 1 – 3. https://api.semanticscholar.org/CorpusID:159006725

[82] Shengkun Xie. 2021. Improving explainability of major risk factors in artificial neural networks for auto insurance rate regulation. *Risks* 9, 7 (7 2021), 126. https://doi.org/10.3390/risks9070126

[83] Chun Yan, Meixuan Li, Wei Liu, and Man Qi. 2020. Improved adaptive genetic algorithm for the vehicle Insurance Fraud Identification Model based on a BP Neural Network. *Theoretical Computer Science* 817 (5 2020), 12–23. https://doi.org/10.1016/j.tcs.2019.06.025

[84] Josh Jia Ching Ying, Chi Kai Chang, and Yen Ting Chang. 2020. Applying a genetic algorithm to determine premium rate of occupational accident insurance. In *2020 International Conference on Machine Learning and Cybernetics (ICMLC)* (Adelaide, Australia), Vol. 2020-December. IEEE, 94–98. https://doi.org/10.1109/ICMLC51923.2020.9469570

[85] Wenguang Yu, Guofeng Guan, Jingchao Li, Qi Wang, Xiaohan Xie, Yu Zhang, Yujuan Huang, Xinliang Yu, and Chaoran Cui. 2021. Claim Amount Forecasting and Pricing of Automobile Insurance Based on the BP Neural Network. *Complexity* (2021), 1–17. https://doi.org/10.1155/2021/6616121