Towards a Hierarchical-structured Decision Support System for Detecting Vehicle Insurance Fraud

[1]Judit Lukács, [2]Péter Váradi, and [1*] Richárd Horváth

[1]Bánki Donát Faculty of Mechanical and Safety Engineering, Obuda University, József körút 6, H-1088 Budapest, Hungary; lukacs.judit@bgk.uni-obuda.hu; horvath.richard@bgk.uni-obuda.hu

[2]Doctoral School on Safety and Security Sciences, Obuda University, József körút 6, H-1088 Budapest, Hungary; varadi.peter@bgk.uni-obuda.hu

[*]Corresponding author

Abstract: *Currently, insurance fraud is a global issue, that has become increasingly widespread. The issue concerns several fields: life insurance, real estate and the vehicle insurance market. This paper focuses on the motor insurance sector, which is a competitive field in Hungary and aims to provide a decision support system, to detect fraudulent claims at an early stage. Based on actual insurance data, a hierarchical fuzzy tree structure was chosen to determine the total risk level of the investigated claims. To achieve this, simple-determinable parameters were divided into subsystems: in the first level, the person-related and the vehicle-related risk levels were investigated. Then, the results were separated into the policyholder and the claimant side. Finally, on the third level, the total risk level was calculated, which is proper for screening the claims. The system can be a useful tool to help focus expertise and human capacity on more critical cases and can also improve the efficiency of the compensation processes.*

Keywords: insurance fraud, artificial intelligence, fuzzy logic

1 Introduction

The high incidence of insurance fraud and the financial losses it causes are a serious problem for insurance companies and their customers [1]. This problem can threaten not only the profitability of the companies, but also society as a whole, as fraud, among other factors, has a direct impact on the development of insurance premiums and can create distrust among customers.

According to the Hungarian Central Statistical Office, approximately 15,000–18,000 road accidents occur in Hungary every year [1]. A certain proportion of the related insurance claims may involve suspected or actual fraud. However, the exact share of fraudulent cases is difficult to determine from publicly available accident statistics alone. Therefore, the early identification of suspicious claims is an important practical task for insurance companies, especially in the motor insurance sector. [2][3]. In addition, with the advent of e-claim reporting systems, the submission of claims has become simpler and faster, but digital channels also open up new opportunities for fraudsters. The rapid development of artificial intelligence offers insurers the opportunity to combat fraudulent cases with more effective methods.

AI-based detection can provide better solutions and improve risk assessment processes (see Figure 1), reducing the need for human resources for clear-cut loss events, thereby increasing accuracy and efficiency for less clear-cut fraudulent cases.

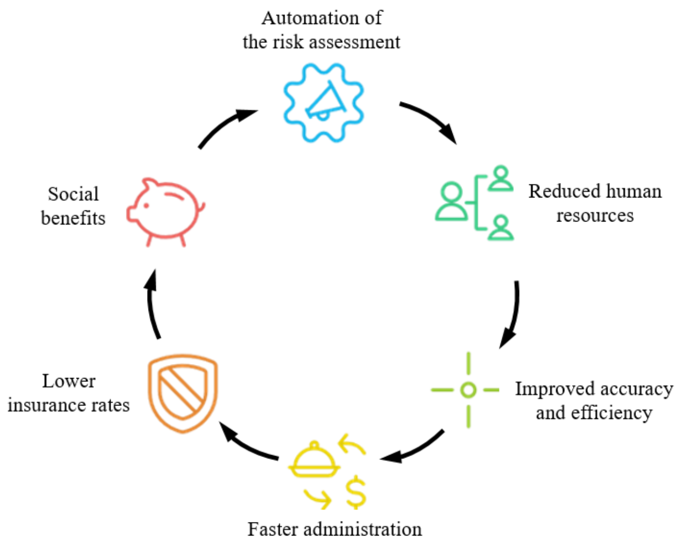

Figure 1
Ideal cycle of efficient insurance claims settlement

The acceleration of the process also means that those, entitled to payments, will receive their money sooner, with insurance premiums increasing less, as a result of reduced expenses for the insurers.

In total, it can be said that the topic is increasingly important today, as the fight against insurance fraud is not only in the common interest of insurers, but also for customers and society. The motivation of this research is to demonstrate how artificial intelligence technologies can be applied, thereby facilitating the smooth operation of the insurance system.

In Hungary, the claim settlement process related to the mandatory vehicle insurance market consists of the steps shown in Figure 2.

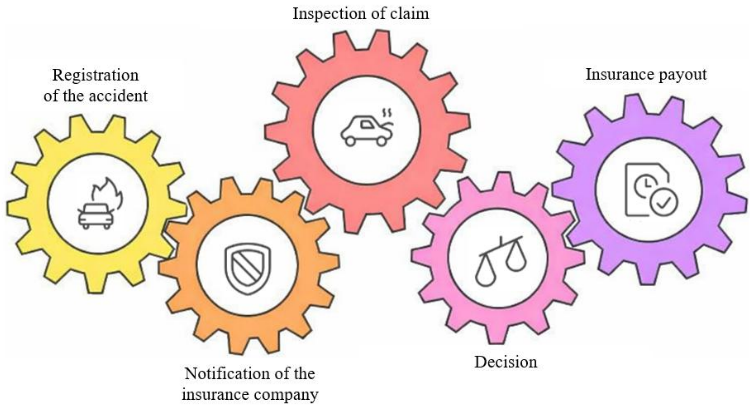

Figure 2
Claims settlement process in Hungary

Faster administration and online reporting greatly accelerate the process for most insurance companies because data is automatically transmitted to the system, reducing the chance of human error.

However, the use of digital channels also raises new security issues, such as the need for data protection and protection against digital fraud. Mishandling of data in e-claims systems or false reports can easily lead to fraud, which is why insurers are devoting more resources to developing artificial intelligence-based fraud detection methods.

The official processes concerning insurance claims operate similarly in other countries as well. However, there might be some differences based on the specific countries and legal systems. These can be influenced by the degree of digitalization, the automation of the claim reporting systems, the organization of the insurance systems, and the methods of detecting fraud.

The detection of insurance fraud is a classification problem, which can be considered either a binary or a multi-classification type. Several studies showed that machine learning methods can provide high effectiveness. Viaene et al. [4] examined the explicative capabilities of Bayesian learning neural networks, particularly Multi-Layer Perceptrons with Automatic Relevance Determination (MLP-ARD), for detecting auto insurance claim fraud. The classification methodology of MacKay [5] was used to train these networks, allowing for the determination of which input features are most informative. The study used a dataset of closed personal injury protection claims from Massachusetts, USA, providing a real-world empirical evaluation. The results showed that MLP-ARD consistently performed well and provided valuable insights into the relative importance of various inputs compared to traditional methods, such as logistic regression and decision trees. Li et al. [6] proposed a novel Random Forest (RF) variant to enhance automobile insurance fraud detection by improving classifier diversity and addressing information loss. Principal Component Analysis was integrated at each decision tree node to transform data and increase the diversity of individual weak learners within the forest. Traditional majority voting concepts were replaced with a Potential Nearest Neighbor to mitigate information loss from out-of-bag samples. Evaluated against other ensemble methods on 12 benchmark datasets and real-world insurance fraud data, the complex technique demonstrated high classification accuracy, reduced variance, and enhanced stability and proper effectiveness in identifying fraudulent claims. David et al. [7] focused on the policyholder risk. To determine appropriate premiums, claim frequency was studied, in the case of non-life insurance. For that purpose, regression analysis, standard Poisson, and Negative Binomial models were used to identify relevant risk factors and predict expected claim frequency. Finally, models were evaluated on the French vehicle insurance portfolio.

Even more studies propose deep-learning concepts for specific categorization problems. A Neural Factorization Machine was introduced by He et al. [8] to handle prediction and classification problems, which combined the linearity in modelling second-order feature interactions with the non-linearity of neural networks in modelling higher-order feature interactions. Guo et al. [9] designed a model that could effectively distinguish fraudulent from normal behaviors and offered interpretable insights into human behavioral patterns. A self-historical attention mechanism is incorporated to capture behavioral periodicity and long-term dependencies, while an interaction module integrates source-target relationship information. The efficiency of the model was evaluated with the help of real-world telecommunication data. Wang et al. [10] introduced a novel deep learning model for automobile insurance fraud detection by leveraging Latent Dirichlet Allocation (LDA) for text analytics. The main goal was to create a complex system that can also analyze textual data from accident descriptions, which extends the amount of information provided by numerical data. It was revealed that deep-learning algorithms can outperform conventional machine learning models (support vector machine, random forests, etc.) in the case of identifying fraudulent cases. The need for more complex solutions is also demonstrated.

Hybrid methods are also preferred in insurance fraud detection. Subudhi et al. [11] presented a novel hybrid approach for automobile insurance fraud detection by combining Genetic Algorithm (GA) with Fuzzy C-Means (FCM) clustering with various supervised classifiers. It was revealed that under-sampling can reduce class imbalance and remove noisy data, thereby enhancing the effectiveness of subsequent supervised models. The proposed system first extracted a test set, then applied GA-FCM to the training data to generate meaningful clusters and identify outliers, which were discarded. Finally, the pre-processed data were used to train supervised classifiers (decision tree, support vector machine, multi-level perceptron, etc.), and suspicious cases were further analyzed. The empirical evaluation using a real-world automobile insurance dataset demonstrated that this hybrid approach effectively improved fraud detection accuracy and minimized misclassification. In addition, Majhi et al. [12] also turned to FCM, which was combined with a modified Whale Optimization Algorithm. With this solution, a better quality was achieved in global optima search, and a more robust model was introduced. Zheng [13] introduced a novel CTGAN-Stacking model designed for robust automobile insurance fraud detection, specifically addressing challenges posed by imbalanced datasets. A key novelty was the usage of CTGAN (Conditional Tabular Generative Adversarial Network) to generate synthetic fraud samples, thereby balancing the dataset for more effective training. The model integrated various algorithms like Random Forest, XGBoost, and ANOVA to create refined input features for a Stacking ensemble. Random Forest, SVM, and neural networks were used as base learners, with logistic regression serving as the meta-learner for final prediction. The proposed CTGAN-Stacking approach was evaluated on real-world data with a high accuracy in identifying fraud compared to single models, offering a robust tool for the insurance industry.

In the past few decades, fraudulent cases also involved the Chinese automobile insurance market. Several concepts were analyzed to minimize the financial loss caused by these incidents [14][15]. Game theory has been proven to be a potential solution [16][17]. However, soft techniques gained more attention thanks to their advantageous behavior. Several methods were investigated, such as data mining [18], Random forests and ant colony algorithm [19], genetic algorithm [20], machine learning [21][22], etc.

It should also be emphasized, that in the case of fraud detection systems, it is not enough to examine performance, but it is also necessary to analyze cost-effectiveness for long-term sustainability. This approach is also a prominent research area in many other scientific fields, especially in healthcare [22][23]. In addition, Benedek and Nagy [24] presented a cost-based comparison, concluding that the better performance of AI cannot be clearly demonstrated between the traditional and artificial intelligence-based systems examined.

These days, in Hungary, there is no effective, artificial intelligence-based “good practice” for screening insurance fraud [24]. It can be stated that fuzzy-based decision support systems can provide encouraging results in the field of fraud detection. The investigation of Váradi et al. [25][26] showed promising potential; the creation of a theoretical decision support system is also presented [27]. Based on all this, it can be stated that the detection and screening of motor insurance fraud is an important task both at the domestic and international levels.

In this paper, a fuzzy-based decision support system is introduced, to highlight risky insurance claims in the motor insurance market. Based on real-life insurance data containing claims from the past ten years, a hierarchical model was created. Simple independent variables were used to describe the participants of the accident. These data were separated and divided into sub-inferences: the personal and the material risks were analyzed from the policyholder and the claimant side as well. Finally, total risk levels were analyzed. The provided system can be a potential tool for screening claims, highlighting critical cases, can help in the distribution of human capacity and improve the efficiency of the compensation processes.

2 Materials and methods

Within the framework of the research, our primary goal was to create a system that is able to follow real trends, which can provide effective assistance in filtering potential fraudulent cases based on easily identifiable factors.

Within the framework of this study, a hierarchical, multi-level fuzzy tree is presented, which is suitable for predicting the risk of fraudulent claims. The system was created based on real cases. It uses easily defined variables to determine the risk experienced in relation to the persons and vehicles involved in the accident, which is examined by both the policyholder and the claimant party. Finally, the overall risk level of the claim process is determined, which also indicates the risk of the case. The created fuzzy tree is particularly suitable for the optimal allocation of human resources and expertise required by the insurer, thus facilitating the acceleration of the claim settlement process.

2.1 Database

To achieve the above goal, real-life insurance data were used, containing accidents that occurred in the past ten years, and an insurance claim was related to the incident. The database contained, in accordance with the GDPR requirements, the data of the persons and vehicles involved in the case (no data that can be used for exact identification), as well as the decision of the insurance company regarding insurance fraud.

The data source was provided by an insurance company in the vehicle insurance market. The available database contained cases with diverse properties. Thus, it was necessary to clarify the examination criteria. During the research, the boundary conditions were defined as follows: those cases were examined in which two passenger vehicles were involved, their owners are natural people, and no personal injury occurred.

After the filtering process, nearly 4,000 cases were gained, in which the current Eurotax value of the vehicles involved (causing and damaged) at the time of the accident was determined. The modeling process was carried out in the MATLAB environment.

2.2 Applied method

In terms of the computational paradigm, the methods can be classified into two classes that operate in fundamentally different ways: hard computing and soft computing methods. Traditional computing is fundamentally based on hard methods; they are based on precision, certainty, and rigidity. These methods completely ignore inflexibility and uncertainty. Precise mathematical models are used to describe problems, which are further classified into symbolic logic and reasoning, and traditional numerical modeling and search algorithms. However, the description of real-life problems that are difficult to characterize exactly and (at least partially) imprecise is difficult to implement and takes a disproportionate amount of time with these methods [28].

In contrast, soft computing is the world of approximate models. In this case, there are further categories: approximate inference, functional optimization, and random search. The great advantage of soft methods is that these techniques are properly suited to real-world problems; imprecision and uncertainty can be handled. This makes them well-suited for use in all industries, business, and many areas of science, whether it is engineering [29]or economics [30], weather [31], or decision/behavior prediction [32].

The introduction of fuzzy logic is attributed to L. A. Zadeh [33]. Fuzzy inference itself is particularly suitable for characterizing human thinking, because instead of binary logic (yes-no/black-white), it monitors the degree of conformity with the help of the membership function. The membership function is suitable for characterizing goodness on a scale between 0 (completely bad) and 1 (perfectly good). This makes it possible to take partial truth into account and to use linguistic variables (e.g., "small", "medium", "large", etc.).

The fuzzy inference system is built from four basic elements, as illustrated in Figure 3.

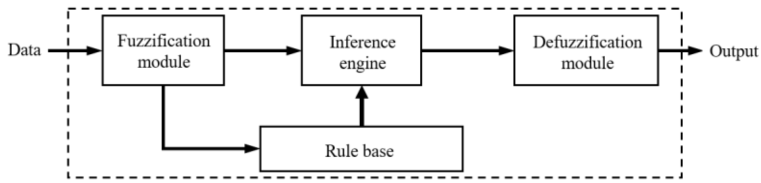

Figure 3
Fuzzy inference system architecture (based on [34][35])

The membership functions (fuzzy sets) are defined in a unit that influences the degree of membership. The rule base contains the accumulated expert knowledge in the form of IF…THEN… rules. The inference engine performs the merging of the currently active, firing rules. Finally, with the help of defuzzification, a specific value from the output set is gained [36]. The central part of the system is the inference engine. Mamdani - type inference [37] was chosen due to its ability to easily describe expert knowledge works. The system works as follows:

The basic idea is that the greater the adequacy of a condition on the input side (the "truer"), the more significantly it should be taken into account in the output on the result side. For example, if we take a simple system with two inputs and one output, where the inputs have 3 possible levels, while the output has 4 (see Figure

4), then a possible partition of the fuzzy sets could be the following:

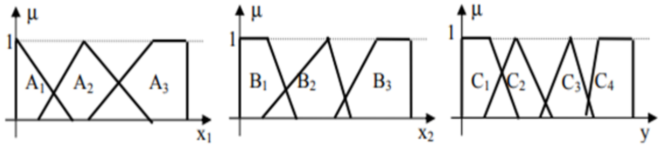

Figure 4
Example of the partition [36]

For an arbitrary pair of measured values, x_{1M} and x_{2M}, two rules are activated, since these values can belong to multiple subsets, to different degrees:

$$R_1 : \text{IF } x_1 = A_2 \text{ AND } x_2 = B_3 \text{ THEN } y = C_3$$

$$R_2 : \text{IF } x_1 = A_3 \text{ AND } x_2 = B_3 \text{ THEN } y = C_2$$

In the first step, rules R_1 and R_2 are activated based on Figures 5 and 6. To calculate the firing strength, for both cases, the minimal degree is chosen from the antecedents.

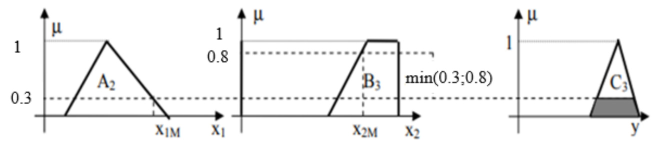

Figure 5
Operation of rule R_1 [36]

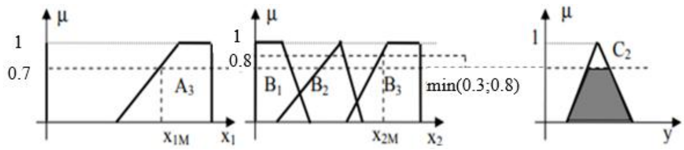

Figure 6
Operation of rule R_2 [36]

The two rules are firing together, so the partial results must be combined. This is how the aggregate result set is gained (see Figure 7).

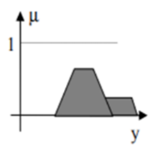

Figure 7
Aggregated fuzzy output [36]

Since this result is hard to understand, defuzzification is needed to obtain a clear, exact, crisp replacement value. Different defuzzification techniques exist; in this case, the effectiveness of the centroid and Largest of Maxima methods was compared.

3 Adaptation

During the model-building procedure, the main goal was to maintain the flexibility and efficiency of the system. This issue also affected the structural design. In general, many input parameters would have to be taken into account, which would also significantly increase the number of potential parameter combinations. An extreme increase in the size of the rule base would practically result in a significant increase in computing capacity and reduced operation speed.

As a result, a hierarchical fuzzy tree structure was chosen that can be created by hierarchically connecting simpler inference systems, which allowed the size of the rule base to be kept under control even with many independent input variables.

3.1 Hierarchical tree structure

The first level of the model (see Figure 8) contained the independent input parameters, while the subsequent levels grouped them in more detail and determined and summarized the associated risk levels.

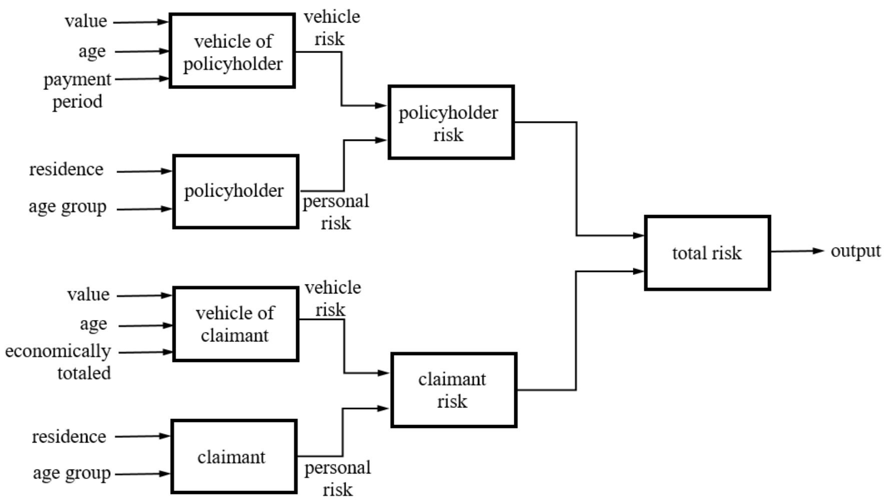

Figure 8
Structure of the model

The detailed explanation of the structure, containing the independent input and output parameters, the rules, the blocks and also their operation is described in Sections 3.2-3.4.

3.2 Input parameters

The first level contained the characteristics of the customers and vehicles involved in the incident. These were divided into two main groups: material (vehicle-related) and personal factors. Table 1 contains the characteristic values and linguistic expressions of the levels of the quantitative variables used in the model.

In addition, two more qualitative variables were introduced in the model: for the material risk on the causing side, the *payment* period of the insurance contract, which can be quarterly, semi-annually, or annually, and on the claimant side, the fact of whether the vehicle is *economically totaled* (yes, no).

Table 1
Independent variables and their linguistic description

Vehicle-related input parameters		**Human-related input parameters**	
***Value*, EUR**	**Linguistic variable**	***Residence*, people (pop.)**	**Linguistic variable**
1-1000	very low	1-5 000	village
1000-3000	low	5 000-20 000	small town
3000-7000	medium	20 000-100 000	medium town
7000-10000	medium expensive	100 000-1 000 000	big town
9000+	expensive	1 000 000+	metropolis
***Age*, years**	**Linguistic variable**	***Age group***	**Linguistic variable**
0-3	new	-1964	Baby Boomer
2-8	young	1965-1980	Generation X
7-15	medium old	1981-1995	Generation Y
14+	old	1996-	Generation Z

The partitions of the independent variables are shown in Figure 9. Due to scaling problems, the partition of *Residence* is not shown graphically.

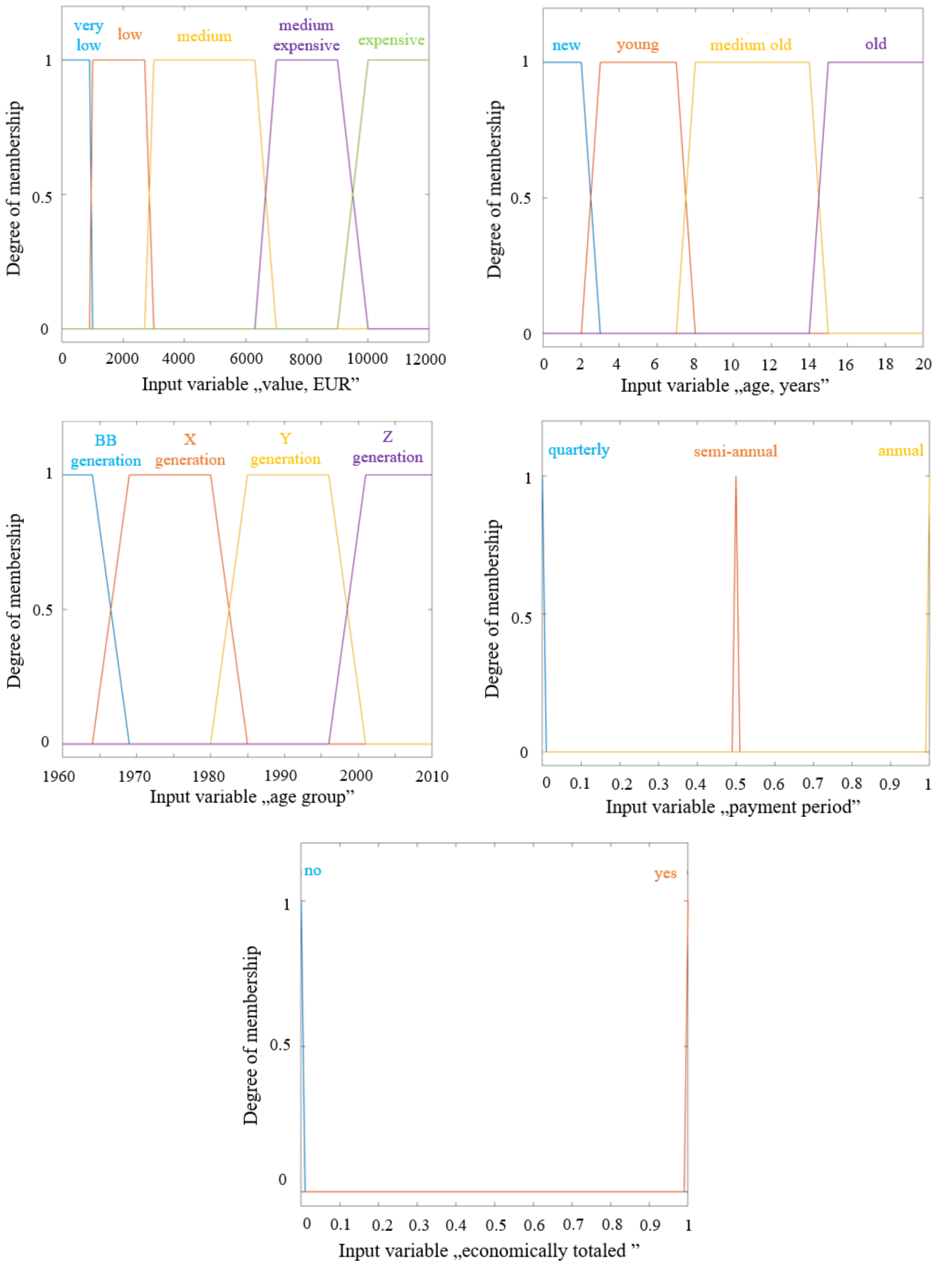

Figure 9
Partitions of the independent variables and the output

3.3 Rules

The decision basis of the fuzzy inference system is provided by the rule base, which is responsible for classifying risks into different levels.

3.3.1 1st Level

On the output side of the first level of the fuzzy tree, the vehicle-related (V_{risk}, %) and personal risks (P_{risk}, %) were determined on the policyholder and claimant sides. Their partition contained 6 different levels: negligible, low, medium, hypothetical, reasonable, and suspicious (see Figure 10). Based on the available data, the probability of being a fraud was calculated for the given parameter combination. These results were converted into rules.

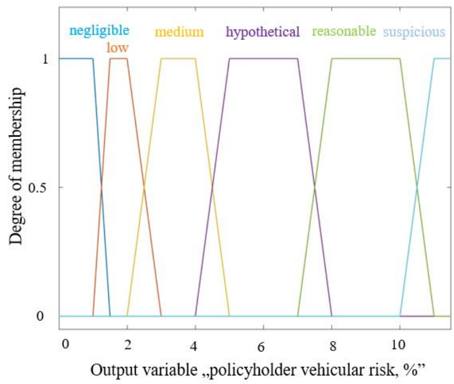

Figure 10
Partition V_{risk}, %

The vehicle-related and the personal risks were estimated with the same concept on the first level of the model.

3.3.2 2nd Level

At the next level of the fuzzy tree, the values of the material and personal risks were merged at each side. The relationships merged in the rule base are given in Table 2, where the individual colors refer to membership functions in Figure 11.

Table 2
Rule base for the second level

PH_{risk} / C_{risk}		V_{risk}, %					
		negligible	low	medium	hypothetical	reasonable	suspicious
P_{risk}, %	negligible						
	low						
	medium						

	hypothetical						
	reasonable						
	suspicious						

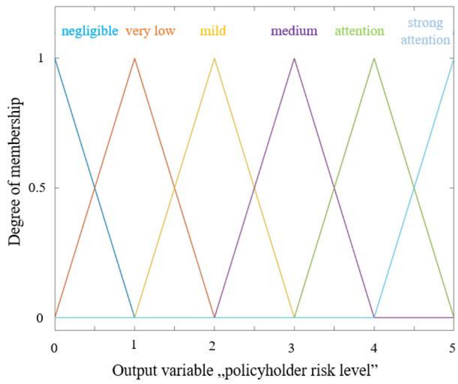

Figure 11
Partition of the output variables on the second level

In the second level, the policyholder risk (PH_{risk}, -) and the claimant risk (C_{risk}, -) are determined similarly. The output is a risk level value between 0 and 5.

3.3.3 3rd Level

Finally, at the third level of the system, risk levels of the two sides were combined, which gave the total risk level for the given claim, also on a scale of 0…5. A higher risk level means a stronger suspicion of being a fraud. In this step, the rule base is shown in Table 3. The output membership functions were the same as at the second level.

Table 3
Rule base at the third level

T_{risk}		PH_{risk}, %					
		-	very low	mild	medium	attention	strong att.
C_{risk}, %	-						
	very low						
	mild						
	medium						
	attention						
	strong att.						

3.4 Defuzzification method

In fuzzy logic systems, defuzzification is an essential step in which the fuzzy sets are transformed into exact, crisp values [35]. The results of two defuzzification methods, the LOM (Last of Maxima) and the centroid (Center of Gravity, COG) methods, were investigated, since there is no generally applicable method.

4 Results

The accuracy and adequacy of the mathematical model were tested using the available database. For all 4,000 cases, the total risk levels were calculated with the help of the fuzzy tree using the centroid and LOM defuzzification techniques. These values were compared with the decisions of the insurance company, indicating fraud.

It was concluded that the centroid method provides better accuracy, so further analyses were fulfilled with the centroid. The total risk level is shown in Figure 12 plotted against the policyholder and the claimant risk, where the lower risk levels are marked with blue and the higher ones with yellow.

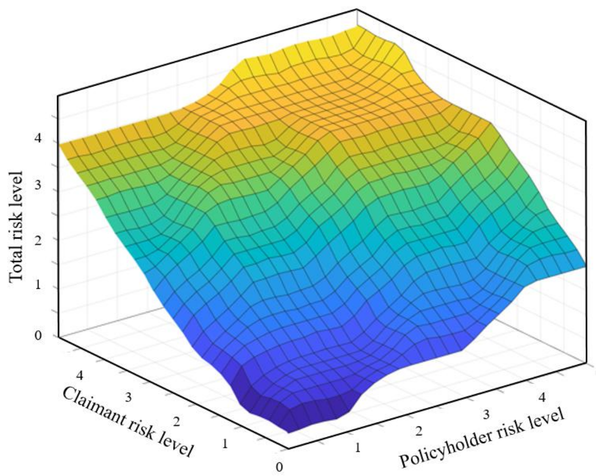

Figure 12
Total risk level

The threshold of the total risk level was set to 2. This means that having a total risk level of less than 2, under the fraud suspicion, is considered sufficiently low; there is no need for a more detailed overview of the case. However, the risk level of 2… 5 means potentially higher risk. In these cases, a comprehensive, more detailed and deeper investigation procedure is needed to make the final decision. As a result, the model can separate risky and less risky incidents to help focus human expertise and optimize resources.

In total, the model matched the decision of the insurance company in 72%. It effectively succeeded in nearly 2,900 cases with low risk levels. This opportunity provides the possibility for regrouping the expertise and resources necessary for the remaining claims with higher risk levels. The method can help to make it more effective for fraud detection, and the administration can become more accelerated, so the run-through time can be reduced.

5 Conclusions

Within the framework of this study, a three-level fuzzy tree was presented, which aimed to filter out motor insurance fraud. During the creation of the model, real-life insurance claims were used for the last 10 years. The impact of the vehicles and the people involved was analyzed separately on both the policyholder and the claimant parties. For the combined analysis, Mamdani-type inference systems were applied in a hierarchical structure. Finally, we determined the total risk level for each case, which was interpreted on a scale of 0…5. The results were compared with the original database.

The conclusions are as follows:

- A ten-input fuzzy tree was presented for determining the risk level of motor insurance claims.
- The total risk level was introduced and used to evaluate the claims: a threshold risk level of 2 was considered as a separation level between risky (potential fraud) and non-risky cases.
- It was found that the best performance is provided with the centroid defuzzification technique, which resulted in ~72% convergence with the examined real-life insurance data Regarding the aim of the study – providing a potential tool for early detection of fraudulent claims – it is a reasonable ratio.
- It can be stated that the system can help in the efficient grouping of human resources during the claims settlement process.

Overall, it can be concluded that the presented model can be an effective tool for detecting motor insurance fraud, which saves money and time for insurance companies and can provide faster, more accurate administration for retail customers.

Acknowledgements

This work was prepared with the professional support of the University Research Scholarship Program code number 2024-2.1.1 of the Ministry of Culture and Innovation, financed by the National Research, Development and Innovation Fund.

References

[1] https://www.insurancefraud.org (Accessed: October 14, 2025)

[2] https://www.ksh.hu/docs/hun/xstadat/xstadat_evkozi/e_feb002.html (Accessed: October 12, 2025)

[3] https://www.autoklub.hu/klub/hirek/baleseti-statisztikak/ (Accessed: October 12, 2025)

[4] Viaene, S., Dedene , G., Derrig, RA (2005). Auto claim fraud detection using Bayesian learning neural networks. *Expert Systems with Applications 29*(3), 653-666. https://doi.org/10.1016/j.eswa.2005.04.030

[5] MacKay, D. J. (1992). The evidence framework applied to classification networks. *Neural computation*, *4*(5), 720-736. https://doi.org/10.1162/neco.1992.4.5.720

[6] Li, Y., Yan, C., Liu, W., & Li, M. (2018). A principle component analysis-based random forest with the potential nearest neighbor method for automobile insurance fraud identification. Applied Soft Computing, 70, 1000-1009. https://doi.org/10.1016/j.asoc.2017.07.027

[7] David, M., & Jemna, DV (2015). Modeling the frequency of auto insurance claims by means of poisson and negative binomial models. Anal stiintifice ale Universitatii "Al. I. Cuza" din Iasi. Stinte economic /Scientific Annals of the "Alexandru Ioan Cuza" University of Iaşi Economic Sciences 62 (2), 2015, 151-168 https://doi.org/10.1515/aicue -2015-0011

[8] He, X., & Chua, TS (2017, August). Neural factorization machines for sparse predictive analytics. In Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval (pp. 355-364). https://doi.org/10.1145/3077136.3080777

[9] Guo, J., Liu, G., Zuo, Y., Wu, J. (2018, November). In 2018 IEEE International Conference on Data Mining (ICDM), pp. 127-136. https://doi.org/10.1109/ICDM.2018.00028

[10] Wang, Y., & Xu, W. (2018). Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. Decision Support Systems, 105, 87-95. https://doi.org/10.1016/j.dss.2017.11.001

[11] Subudhi, S., & Panigrahi, S. (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), 568-575. https://doi.org/10.1016/j.jksuci.2017.09.010

[12] Majhi, SK (2021). Fuzzy clustering algorithm based on modified whale optimization algorithm for automobile insurance fraud detection. Evolutionary Intelligence, 14(1), 35-46. https://doi.org/10.1007/s12065-019-00260-3

[13] Zheng, Y. (2025). Analysis of Auto Insurance Fraud Prevention Model

Based on CTGAN—Stacking Model. *Journal of Computers*, *36*(5), 89-101. https://doi.org/10.63367/199115992025103605006

[14] Wang, H. (2016). A Research on Chinese Insurers' Moral Hazard Screening in Operation: From the Big Data Hadoop Clustering Analysis Technology Perspective. Insurance Studies, 2, 59-67.

[15] Zhao, G., & Wu, H. (2010). Is There Moral Hazard in Chinese Automobile Insurance Market?— Evidence from Dynamic Renewal Policies. Journal of Financial Research, 6, 175-188.

[16] Tuo, G., Duan, J. (1999). Game Theory Analysis of Insurance Fraud. Journal of Capital University of Economics and Business (3), 51-54.

[17] Liu, X., & Jin, J. (2004). The Insurance Fraud Game and Insurance Contract Based on Optimal Game Strategies. Systems Engineering—Theory & Practice, 24(2), 19-24.

[18] Tang, J., & Mo, Y. (2013). Construction of auto insurance anti-fraud system based on data mining technology. Shanghai insurance, 11, 39-42.

[19] Yan, C., Li, Y., & Sun, H. (2017). A Research on Automobile Insurance Fraud Identification Based on Random Forest Model and Ant Colony Optimization Algorithm. Insurance Studies, 6, 114-127.

[20] Yan, C., Li, M., & Zhou, X. (2019). Improved genetic algorithm for vehicle insurance fraud identification model based on BP neural network. Journal of Shandong University of Science and Technology (Natural Science), 38(5), 72-80.

[21] Yang, J., Chen, K., Ding, K., Na, C., & Wang, M. (2023). Auto insurance fraud detection with multimodal learning. Data Intelligence, 5(2), 388-412. https://doi.org/10.1162/dint_a_00191

[22] Lee, A., Taylor, P., Kalpathy-Cramer, J., & Tufail, A. (2017). Machine learning has arrived! Ophthalmology, 124(12), 1726-1728. https://doi.org/10.1016/j.ophtha.2017.08.046

[23] Hill, NR, Sandler, B., Mokgokong, R., Lister, S., Ward, T., Boyce, R., ... & Gordon, J. (2020). Cost-effectiveness of targeted screening for the identification of patients with atrial fibrillation: evaluation of a machine learning risk prediction algorithm. Journal of medical economics, 23(4), 386-393. https://doi.org/10.1080/13696998.2019.1706543

[24] Benedek, B., Nagy, B. Traditional versus AI- based fraud detection: Cost-effectiveness of car insurance in the field of. CREDIT INSTITUTIONS REVIEW, 77. https://doi.org/10.25201/HSZ.22.2.77

[25] Váradi, P., Lukács, J., & Horváth, R. (2023, May). Examination of vehicle fraud detection possibilities with the help of Fuzzy inference system. In 2023 IEEE 17th International Symposium on Applied Computational Intelligence and Informatics (SACI) (pp. 000353-000358). IEEE. https://doi.org/10.1109/SACI58269.2023.10158631

[26] Váradi, P., Lukács, J., & Horváth, R. (2024, April). Response Surface Methodology to Characterize the Indicators for the Early Detection of

Fraudulent Cases in the Motor Insurance Market. In 2024 IEEE 11th International Conference on Computational Cybernetics and Cyber-Medical Systems (ICCC) (pp. 000201-000204). IEEE. https://doi.org/10.1109/ICCC62278.2024.10582971

[27] Váradi, P., Lukács, J., & Horváth, R. (2024). Theoretical fuzzy inference system for suspected fraud for early detection of cases in the motor insurance in the sector. Journal of Transportation Sciences, 74(5), 31-43. https://doi.org/10.24228/KTSZ.2024.5.3

[28] Chaturvedi, DK (2008). Soft computing. Studies in Computational Intelligence, 103, 509-612.

[29] Tóth-Laufer, E., Horváth, R. (2017). Fuzzy model based surface roughness prediction of fine turning. FME Transactions, 45(1), 181-188. https://doi.org/10.5937/fmet1701181T

[30] Singh, AK, & Parida, P. (2013). Soft computing in financial decision making. Global Journal of Management and Business Studies, 3(2), 103-110.

[31] Cadenas, JM, Garrido, MC, Martínez-España, R., & Guillén-Navarro, MA (2020). Making decisions for frost prediction in agricultural crops in a soft computing framework. Computers and Electronics in Agriculture, 175, 105587. https://doi.org/10.1016/j.compag.2020.105587

[32] Forgács, A., Lukács, J., & Horváth, R. (2021). The Investigation of the Applicability of Fuzzy Rule-based Systems to Predict Economic Decision-Making. Acta Polytechnica Hungarica, 18(11). https://doi.org/10.12700/APH.18.11.2021.11.6

[33] Zadeh, LA (1965). Fuzzy sets, Information and control, 8(3), 338-353.

[34] Kóczy, LT, Tikk, D. (2000). Fuzzy systems . TypoTEX, Budapest.

[35] Yager, RR, Filev , DP (1994). Essentials of fuzzy modeling and control. New York, 388.

[36] Johanyák, Cs. Zs. (2004) Fuzzy inference methods. Colleges Mathematics, Physics and Computer Science XXVIII National Conference of Educators

[37] Mamdani, EH (1974, December). Application of fuzzy algorithms for control of simple dynamic plants. In Proceedings of the institution of electrical engineers (Vol. 121, No. 12, pp. 1585-1588). IET https://doi.org/10.1049/piee.1974.0328