

Vehicle-Related Risk Level in the Case of Claims on the Motor Insurance Market in Hungary [†]

Judit Lukács ¹, Péter Váradi ² and Richárd Horváth ^{1,*} ¹ Bánki Donát Faculty of Mechanical and Safety Engineering, Obuda University, 1088 Budapest, Hungary; lukacs.judit@bgk.uni-obuda.hu² Doctoral School on Safety and Security Sciences, Obuda University, 1034 Budapest, Hungary; varadi.peter@phd.uni-obuda.hu

* Correspondence: horvath.richard@bgk.uni-obuda.hu

[†] Presented at the Sustainable Mobility and Transportation Symposium 2025, Győr, Hungary, 16–18 October 2025.

Abstract

Insurance fraud, characterized by false or exaggerated claims, is a major economic crime worldwide, undermining trust between insurance companies and their customers. Detecting these cases is a priority issue nowadays. This paper presents a fuzzy inference system for the early identification of suspicious claims in the compulsory motor liability insurance market. The study focuses exclusively on cases involving two privately owned passenger cars where no personal injury, but only property damage, occurred. A Mamdani-type inference system was created, using simple independent input parameters: the *value* (in EUR) and the *age* of the vehicle (in years) and the payment *period* of the insurance contract. The last parameter was introduced as a qualitative factor. These were linked to the risk level resulting from the characteristics of the vehicles involved in the incident. For this purpose, real insurance data were used.

Keywords: insurance fraud detection; economic risk; risk level; Mamdani-type fuzzy inference; motor insurance



Academic Editors: Szabolcs Kocsis-Szürke, Dániel Csikor and Boglárka Eisinger Balassa

Published: 7 November 2025

Citation: Lukács, J.; Váradi, P.; Horváth, R. Vehicle-Related Risk Level in the Case of Claims on the Motor Insurance Market in Hungary. *Eng. Proc.* **2025**, *113*, 37. <https://doi.org/10.3390/engproc2025113037>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Economic crime occurs in many areas of our lives [1]. The transport and transportation sectors are no exception. But here, as critical infrastructure, it is vital to detect, handle, and reveal these incidents and minimize their occurrence all around the world.

Schrijver et al. [2] presented a detailed review on potential methods for automobile insurance fraud detection. Machine learning and data mining are proposed as useful tools. Benedek et al. [3] compared traditional and AI-based methods. Several challenges were introduced: imbalanced datasets, the lack of standardized fraud indicators, the limited availability of proper datasets, etc. It was highlighted that cost-sensitive models have to be developed.

Real-life insurance data can provide more information about the situations in the vehicle insurance market. Lincoln et al. [4] highlighted the challenges of vehicle insurance fraud with the investigation of the Australian market. The research contained recommendations from a practical point of view: potential “red flags” like missing data, behavioral indicators, vehicles, and financial factors. It was concluded that a multi-layered approach and combination of different techniques might be required to detect fraudulent cases. Duwadi and Ghimiri [5] also used a real insurance claims dataset, preprocessed using oversampling (SMOTE) and undersampling to address class imbalance. In addition, multiple-based

classifiers were applied with hyperparameter tuning. As a result, a robust machine learning framework was demonstrated to enhance insurance fraud detection, minimize financial loss, and improve prevention strategies. Bodaghi and Teimourpour [6] proposed a network-based approach to identify organized fraud groups in insurance claims. In this study, a collision network was constructed using real-life insurance data in Iran. Cycle detection algorithms were used to reveal recurring patterns of suspicious claims. In addition, a social network analysis was carried out to emphasize structural fraud patterns over individual anomalies to improve fraud prevention strategies. In total, it can be said that, based on the literature, there are several challenging issues and also potential methodologies in detecting and preventing insurance fraud. However, general solutions are not present yet.

Modern methods, like fuzzy logic and soft computing, are widely used in safety studies. Pinto et al. [7] proposed the application of fuzzy severity functions to objectively evaluate occupational accident severity in the construction industry. Multiple severity predictors helped to analyze consequences. It was demonstrated that fuzzy set theory provided a practical and adaptable possibility to improve the accuracy of decision-making. Ali et al. [8] applied a hybrid modeling approach: decision tree analysis was combined with a panel mixed logit model to analyze stop/go decisions in a driving simulator. The aim of the study was to justify the need for personalized driving assistance systems to improve safety. Regarding the insurance market, Shapiro [9] said that fuzzy logic can be promising in all areas of the insurance market: it was demonstrated that the application of fuzzy systems is not limited to fraud detection; it can also provide a useful tool in understanding insurance decision-making. Furthermore, Kalra et al. [10] proposed the use of fuzzy expert systems in predicting and detecting fraud in healthcare insurance. By assigning fraud detection scores, the faster and more effective identification of fraudulent claims is enabled.

This study focuses on the vehicle insurance market. The vehicle-related risk is analyzed and modeled in detail. For that purpose, real-life insurance data were used from Hungary with the following limitations: only those reported claims were taken into consideration where two passenger cars were involved in the incidents, and these were owned by natural persons, and only property damage occurred. A Mamdani-type inference system was introduced, using three independent input parameters describing the vehicle of the policyholder: the *value* (in EUR) and the *age* of the vehicle (in years) and the *payment period* of the insurance contract. The last two parameters were introduced as qualitative factors. These were linked to the *vehicular risk*, in %, resulting from the characteristics of the vehicles involved in the incident. The results can be used to highlight suspicious insurance claims at a very early stage and might help insurance companies to detect fraudulent cases.

2. Materials and Methods

In Hungary, a compulsory liability insurance is needed to drive a vehicle on the roads. It is intended to help owners to compensate the at-fault party for any damage caused in an accident. However, some believe that compensation payments are also an alternative means of generating income.

Previous studies have revealed that the detection of insurance fraud in the vehicle insurance market causes significant economic damage and some further problems for policyholders, insurance companies, and society as well. As a result, investigating fraudulent cases is important, where several problems appear: a lack of information, incorrect datasets, etc. [3] The theoretical application of fuzzy rule-based systems was already determined [11]. So, in this case, the extension was aimed at helping with realistic data.

2.1. Dataset Applied

For the investigations, real-life insurance data were used over the past ten years. Since the claims varied over a very wide range of features, several limitations were introduced at the beginning. Insurance claims were selected with the following criteria:

- Double-vehicle-related incidents where two passenger cars were involved,
- Natural people owned the vehicles,
- No personal injury was reported.

This concept made it possible to analyze the features of the policyholder and the third party as well. In addition, company cars with additional issues were eliminated from the set. In total, around 4000 cases were investigated. The parameters of the people and vehicles were available, as well as the decision of the insurance company on whether it was fraud or not.

2.2. Methods

Fuzzy set theory was introduced by Zadeh [12], aiming to handle imprecision, uncertainty, and a lack of information. This concept fits into the circumstances that are to be faced when analyzing insurance claims [9–11].

The Mamdani-type fuzzy inference system [13] is a useful tool when the goal is to model human decision-making. The basic structure contains four blocks: (1) fuzzification is important in the definition of the membership functions; (2) the rule base contains the available information sourced from previous knowledge; (3) the inference engine combines the activated rules and determines the firing strength; and (4) the defuzzification module helps to transform the output fuzzy set into a crisp value [14].

2.3. Adaptation

This study focuses on vehicle-related risk, so simple parameters were taken into consideration which are easy to determine and do not raise any GDPR issues. For that purpose, a Mamdani-type triple-input–single-output fuzzy inference system was created to estimate the vehicular risk from the perspective of the policyholder.

Two quantitative independent input variables were defined: the *value* (in EUR) and the *age* of the vehicle (in years). The *value* varied at five levels: very low, low, medium, medium expensive, and expensive; the *age* was at four levels: new, young, medium old, and old. The payment *period* of the insurance contract was added as a qualitative feature: on the Hungarian market, policyholders can choose from quarterly, semi-annual, or annual payment periods.

On the output side, the *vehicular risk* was introduced with the help of the database: for each possible combination, the relative incidence rate of fraud occurrence was calculated in %. The linguistic expressions of the levels were as follows: negligible, low, medium, hypothetical, reasonable, and suspicious. This parameter was expected to provide information on how often a given case is concluded as fraud.

The partitions of the variables are shown in Figure 1a–c for the independent input variables and Figure 1d for the single output.

In Figure 1, two types of membership functions were used: trapezoid-shaped for the *value*, the *age*, and the *vehicular risk* (Figure 1a,b,d), where the transition can provide a better description for quantitative data, and, for the *payment period*, the so-called fuzzy singleton was applied. This specific membership function works like the characteristic function in crisp set theory: at the one peak, truthness appears only at a single value. This function can be used to describe qualitative data, where transitions cannot be defined between different levels.

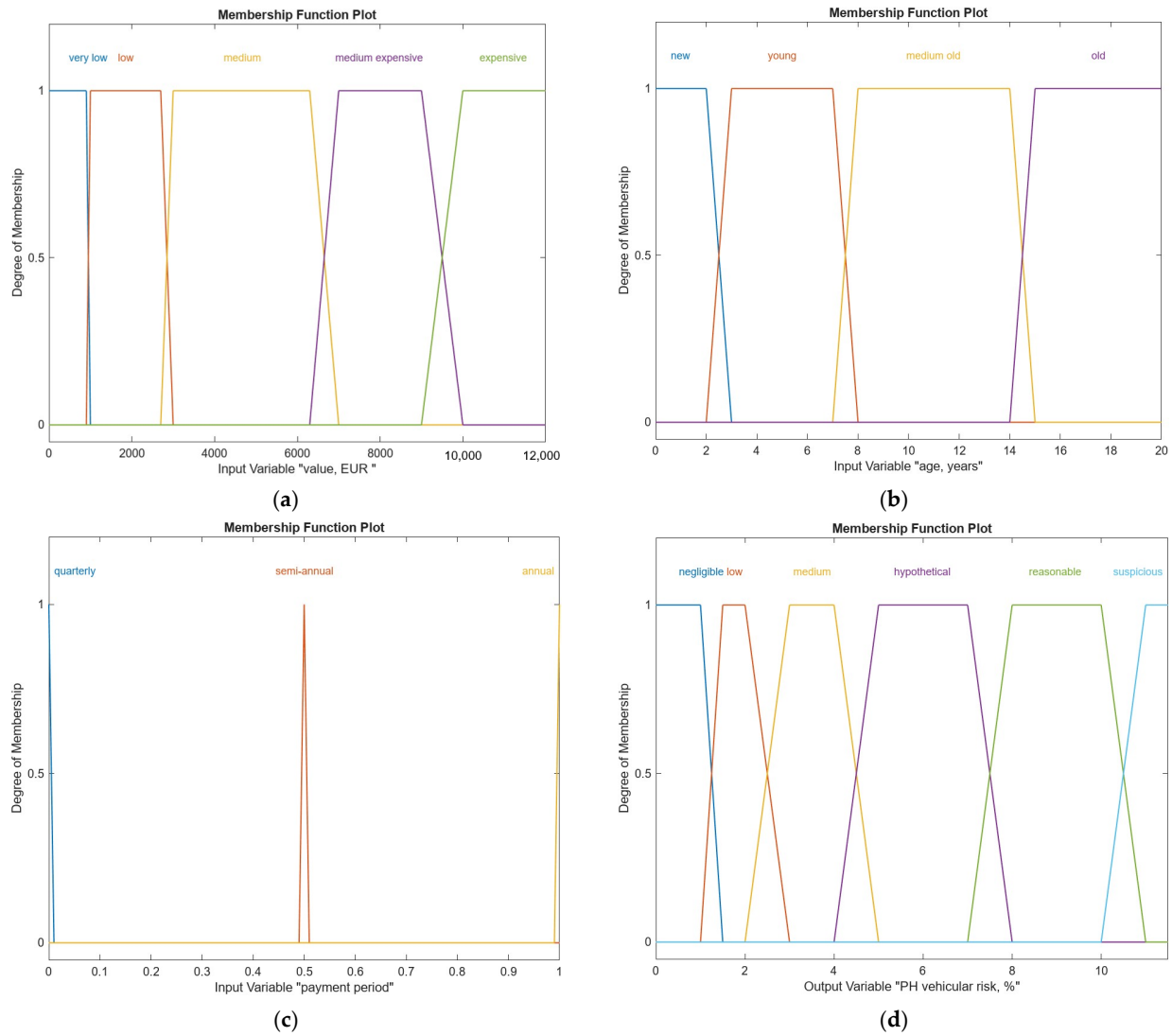


Figure 1. Partition of the input and output variables: (a) *value* of the vehicle, EUR; (b) *age* of the vehicle, years; (c) *payment period*, qualitative parameter; (d) *vehicular risk*, %.

The ranges of the quantitative independent input and the output variables are shown in Table 1.

Table 1. Ranges of the quantitative parameters used in the model.

Value	Range in EUR	Age	Range in Years	Veh. Risk	Range in %
Very low	0–1000	New	0–3	Negligible	0–1.5
Low	900–3000	Young	2–8	Low	1–3
Medium	2700–7000	Medium old	7–15	Medium	2–5
Med. exp.	6300–10,000	Old	14–	Hypothetical	4–8
Expensive	9000–			Reasonable	7–11
				Suspicious	10–

The rule base contained all the possible combinations of the independent variables. In total, 60 rules were defined to introduce previous expert knowledge into the system.

Finally, the Largest of Maxima method was used in defuzzification to highlight all the potential fraudulent claims.

3. Results

As a result, *vehicular risk*, in %, was calculated based on the available data. Figure 2 shows the response surface for the relative incident rate of fraud occurrence. Since the *payment period* is a qualitative parameter varied at three levels (quarterly, semi-annual, and annual), for a better graphical representation, these cases are shown separately.

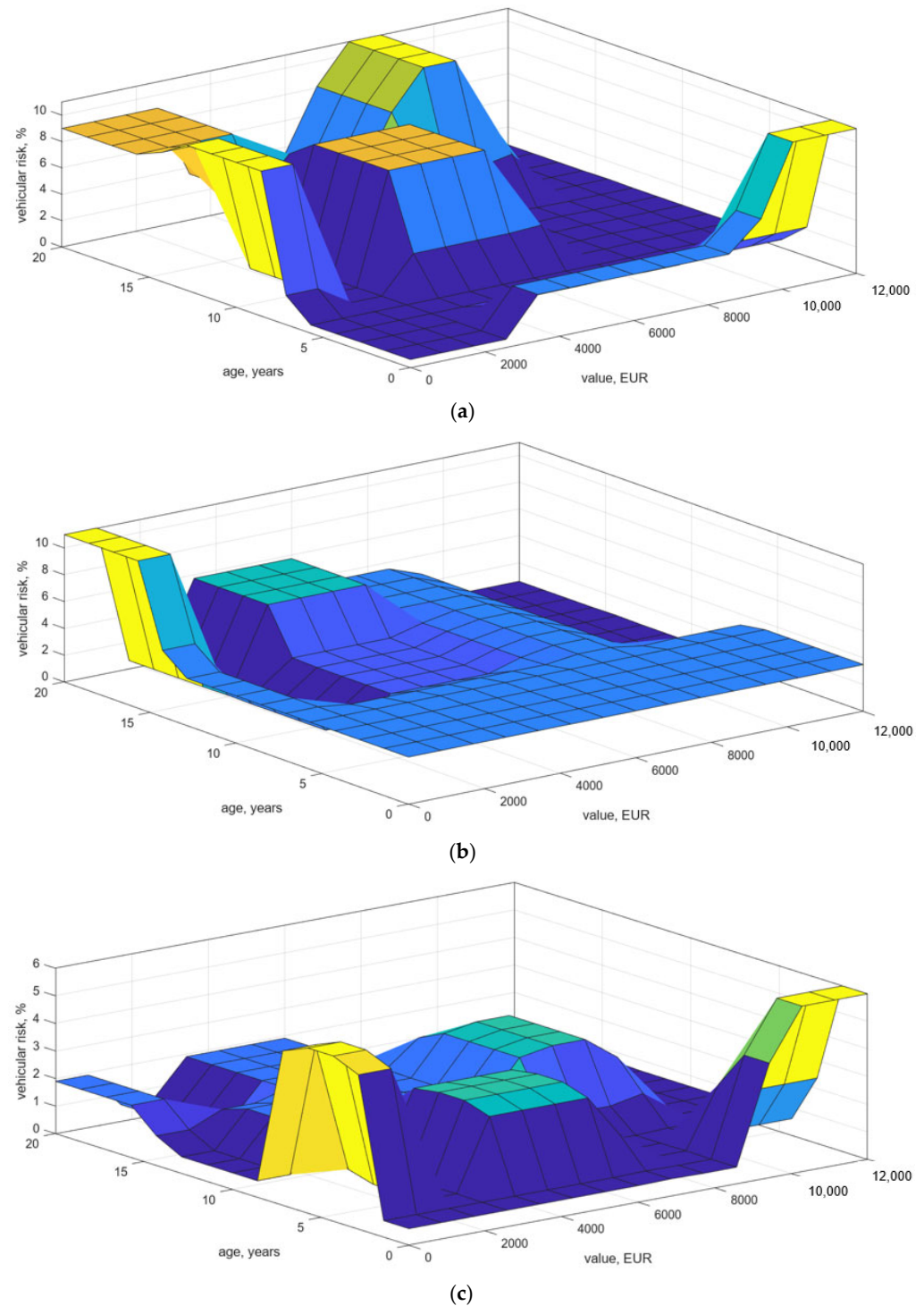


Figure 2. Output surface for the *vehicular risk* in %: (a) for quarterly; (b) for semi-annual; (c) for annual payment period.

Based on Figure 2, it can be said that the trends in a classic sense cannot be revealed, since the parameters have a combined effect. However, critical areas can be found. Figure 2a

shows the *vehicular risk* for the quarterly, Figure 2b for the semi-annual, and Figure 2c for the annual *payment period*. The higher risk is marked with yellow, and the lower with blue. It can be said that the most fraudulent area is in the case of quarterly payments. In addition, the highest *vehicular risk* can also be observed here with 11–12%, compared to semi-annual with 10–11% and the annual period with 6–7%.

What is more, the highest possible risk is not only connected to the payment period, but also to the magnitude of high-risk fraud cases. As the level of the payment period is increased, the amount of high-risk claims is reduced.

Although the introduced model can be useful to estimate the vehicular risk on the policyholder, as well as person-related risk, the claimant side needs to be analyzed to obtain more precise information about the claims.

4. Summary

Insurance frauds cause huge economic damage. As a result, the detection and identification of fraudulent claims is a key but challenging issue nowadays.

This study focuses on predicting the risk, in %, related to the vehicle of the policyholder. For that purpose, a Mamdani-type inference system was introduced with three simply determinable independent input variables: the *value* (in EUR) and the *age* of the vehicle (in years) and the *payment period* of the insurance contract (defined as a qualitative parameter). The single output of the system was the *vehicular risk* to the policyholder, calculated as a percentage. To implement previous expert knowledge into the systems, a real-life insurance dataset was used, containing claims from the past ten years.

The conclusions are as follows:

- It can be said that the introduced fuzzy inference system can predict the vehicular risk from the perspective of the policyholder.
- On the one hand, exact trends and simple relations in terms of the independent variables cannot be defined. However, risky areas can be determined.
- It was revealed that the increased level of the *payment period* resulted in reduced maximal *vehicular risk* values. In addition, with the enhancement of the payment period, the magnitude of the high-risk area also decreased.

In total, it can be concluded that the model can help detect vehicle-related risk at an early stage of the claims process. However, to obtain more detailed information on the risk, several further parameters have to be investigated: the human factor on the policyholder side and the claimant side as well.

Author Contributions: Conceptualization, J.L., P.V. and R.H.; methodology, J.L.; software, J.L.; formal analysis, J.L., P.V. and R.H.; investigation, J.L. and P.V.; resources, J.L. and P.V.; writing—original draft preparation, J.L.; writing—review and editing, J.L., P.V. and R.H.; visualization, J.L. and R.H.; supervision, J.L. and R.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the 2024-2.1.1 University Research Scholarship Program of the Ministry for Culture and Innovation from the Source of the National Research, Development and Innovation Fund.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Passas, N. Globalization, Criminogenic Asymmetries and Economic Crime. In *International Crimes*; Routledge: London, UK, 2017; pp. 17–42.
2. Schrijver, G.; Sarmah, D.K.; El-Hajj, M. Automobile Insurance Fraud Detection Using Data Mining: A Systematic Literature Review. *Intell. Syst. Appl.* **2024**, *21*, 200340. [[CrossRef](#)]
3. Benedek, B.; Ciumas, C.; Nagy, B.Z. Automobile Insurance Fraud Detection in the Age of Big Data—A Systematic and Comprehensive Literature Review. *J. Financ. Regul. Compliance* **2022**, *30*, 503–523. [[CrossRef](#)]
4. Lincoln, R.; Wells, H.; Petherick, W. An Exploration of Automobile Insurance Fraud. *Humanit. Soc. Sci. Pap.* **2003**, *64*.
5. Duwadi, N.; Ghimire, B.R. Automobile Insurance Fraud Detection Using Ensemble Learning Models. *Int. J. Eng. Appl. Sci. Technol.* **2024**, *9*, 187–199. [[CrossRef](#)]
6. Bodaghi, A.; Teimourpour, B. The Detection of Professional Fraud in Automobile Insurance Using Social Network Analysis. *arXiv* **2018**, arXiv:1805.09741. [[CrossRef](#)]
7. Pinto, A.; Ribeiro, R.A.; Nunes, I.L. Fuzzy Approach for Reducing Subjectivity in Estimating Occupational Accident Severity. *Accid. Anal. Prev.* **2012**, *45*, 281–290. [[CrossRef](#)] [[PubMed](#)]
8. Ali, Y.; Haque, M.M.; Zheng, Z.; Bliemer, M.C. Stop or Go Decisions at the Onset of Yellow Light in a Connected Environment: A Hybrid Approach of Decision Tree and Panel Mixed Logit Model. *Anal. Methods Accid. Res.* **2021**, *31*, 100165. [[CrossRef](#)]
9. Shapiro, A.F. Fuzzy Logic in Insurance. *Insur. Math. Econ.* **2004**, *35*, 399–424. [[CrossRef](#)]
10. Kalra, G.; Rajoria, Y.K.; Boadh, R.; Rajendra, P.; Pandey, P.; Khatak, N.; Kumar, A. Study of Fuzzy Expert Systems towards Prediction and Detection of Fraud Case in Health Care Insurance. *Mater. Today Proc.* **2022**, *56*, 477–480. [[CrossRef](#)]
11. Váradi, P.; Lukács, J.; Horváth, R. Examination of Vehicle Fraud Detection Possibilities with the Help of Fuzzy Inference System. In Proceedings of the 2023 IEEE 17th International Symposium on Applied Computational Intelligence and Informatics (SACI), Timisoara, Romania, 18–20 May 2023; pp. 353–358. [[CrossRef](#)]
12. Zadeh, L.A. Fuzzy Set Theory. *Inf. Control* **1965**, *8*, 338–353. [[CrossRef](#)]
13. Mamdani, E.H. Application of Fuzzy Algorithms for Control of Simple Dynamic Plant. In Proceedings of the Institution of Electrical Engineers; IET: London, UK, 1974; Volume 121, pp. 1585–1588.
14. Wang, P.P. *Computing with Words*; John Wiley & Sons: Hoboken, NJ, USA, 2001.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.