CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 10 (2) (2025), Pages 998-1014
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Transformation of Traditional Models to AI: SLR on the Application of Machine Learning in Mortality Prediction

Vita Nuraini and Nina Fitriyati*

Department of Mathematics, Faculty of Sciences and Technology, UIN Syarif Hidayatullah Jakarta, Indonesia

Abstract

The application of machine learning (ML) in actuarial science and life insurance has driven the digital transformation in mortality risk prediction. This article conducts research using the systematic review methodology (SLR) with the PRISMA approach to evaluate the performance comparison between ML methods and traditional actuarial models in predicting mortality risk. This study analyzed publication trends, geographic and institutional distribution, and methodologies used in the literature published between 2019 and 2025. The SLR results show that ML methods, especially Random Forest and XGBoost, have superior predictive accuracy compared to traditional actuarial models such as Traditional Logistic Regression and Cox Proportional Hazards. However, despite the obvious accuracy advantage, issues of interpretability and long-term stability remain a major challenge in implementing ML in the actuarial industry. This study also highlights the need for a hybrid approach that combines the strengths of both methodologies to enhance prediction accuracy while maintaining high interpretability. This study suggests the need for further development in the application of ML through the regulation and compliance of the insurance industry. The findings provide insights for actuarial practitioners, regulators, and academics on the potential and challenges of using ML in the prediction of mortality risk.

Keywords: Machine learning, traditional actuarial model, hybrid model, mortality risk.

1 Introduction

Machine learning (ML) in the current digital era encourages digital transformation in various areas of life, including actuarial science and life insurance [1]. In life insurance, predicting the risk of death is the primary factor in calculating premiums, which are traditionally modeled using mortality tables and the Lee-Carter model [2], but this model has shortcomings. To address these shortcomings, the machine learning (ML) method offers the advantage of improved prediction accuracy in dealing with the complexity of modern mortality, which is influenced by several factors, including the risk of the COVID-19 pandemic [3]. In addition, machine learning (ML) methods, such as deep learning, random forest, and neural networks, can also overcome the challenges of capturing complex interactions and non-linear relationships between risk variables in traditional models with a strong mathematical foundation and good interpretability [4].

*Corresponding author. E-mail: nina.fitriyati@gmail.com

Submitted: August 07, 2025 Reviewed: October 10, 2025 Accepted: October 19, 2025
DOI: https://doi.org/10.18860/cauchy.v10i2.35972

Machine learning (ML) methods offer several advantages in terms of predictions required by the actuarial industry; however, applying these methods still presents significant challenges. These challenges are interpretability in the context of actuarial and insurance, which is difficult to use because many algorithms in machine learning (ML) methods function as "*black boxes*" [5]. Not only that, but limitations in long-term validation of insurance products, regulatory compliance with government regulations, and ethical considerations toward customers are also challenges that need to be resolved before machine learning (ML) methods are fully implemented [6]. Several studies have compared the performance of various machine learning (ML) methods with simple actuarial models in certain contexts in actuarial science. For example, the Lee-Carter model and recurrent neural networks (RNN) methods can be compared to predict mortality rates. However, few studies integrate the findings of various comparison studies related to the performance of machine learning (ML) methods with traditional techniques to predict mortality risk.

This study was conducted to fill the gap using a systematic review (SLR) research method, which employed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology for the review. The objectives of this SLR research are to analyze the performance trends of ML methods compared to the performance of traditional actuarial models in predicting mortality risk, identify patterns present in publications from year to year at specific publication year boundaries, analyze the geographic and institutional distribution of relevant publications, identify key journals that publish on this topic, explore the most frequently used methodologies for comparison, and identify any research gaps for future studies. Thus, through the results of this SLR, actuarial practitioners, regulators, and researchers or academics gain important insights into the development of the application of machine learning (ML) methods in the actuarial industry to reduce the gaps that traditional actuarial models have in mortality risk prediction—not forgetting while maintaining a strong mathematical basis and implementing good interpretability for mortality risk management in the current era.

2 Methods

2.1 Research Design

This research employs a Systematic Literature Review (SLR), which utilises the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology [7] to conduct a comparative analysis of the performance of machine learning (ML) methods with traditional actuarial models in predicting mortality risk. The purpose of choosing this method is to obtain a comprehensive overview of the research status on this topic at a specified time interval, as SLR can systematically organize and synthesize research evidence.

2.2 Search Strategy

Specifically, this study used the Scopus database as a platform in the literature search conducted in February 2025. The search string used was (("machine learning" OR "artificial intelligence" OR "deep learning" OR "predictive modelling") AND ("mortality prediction" OR "life insurance" OR "insurance risk" OR "actuarial science" OR "actuarial models" OR "mortality tables")). The search string was developed based on the PICOS (Population, Intervention, Comparison, Outcomes, and Study Design) framework to identify literature relevant to this study's research question.

Scopus was deliberately chosen for this study due to its broad multidisciplinary coverage, encompassing journals and conference proceedings in healthcare, actuarial science, finance, and computational modeling. Scopus integrates indexed content from major repositories, including IEEE, Springer, and Elsevier, thereby covering the vast majority of relevant literature on this topic. We used only Scopus to avoid overlap with other databases. Furthermore, using a single, high-quality database ensures methodological consistency, reproducibility, and transparent data management throughout the review process.

2.3 Inclusion and Exclusion Criteria

Articles selected for further review were determined based on inclusion and exclusion criteria. The inclusion criteria included publication years between 2019 and 2025, peer-reviewed articles in English journals, and studies that directly compared machine learning (ML) methods with traditional actuarial models in the context of mortality risk prediction. Only open access articles were selected to ensure complete transparency, reproducibility, and accessibility of the reviewed literature. This approach allows other researchers to verify independently included studies and replicate the systematic review process without institutional access restrictions. Furthermore, initial screening revealed substantial thematic overlap between open-access and subscription-based publications in Scopus, confirming that this restriction did not materially impact the representativeness or validity of the findings. Inclusion and exclusion criteria were chosen to define the boundaries of this study, although they may introduce bias in selecting articles for the study. The specific range of years, from 2019 to 2025, was also chosen because recent bibliometric analyses have shown a marked increase in machine learning publications and applications since 2019. For example, research from Alzoubi, et al. [8] shows that in the last five to six years, since 2019, there has been a significant growth in the number of articles using ML in cloud security, and research from Ayanwale, et.al. [9] which states the same thing , making it a suitable range to look at trends. Language restriction was also imposed, and English was chosen because it is an international language that can unify understanding and avoid terminology inequality. Other criteria were determined in accordance with this research topic to enhance the homogeneity of the sample to be studied.

2.4 Study Selection Process

The study selection is conducted in three stages, following the recommendations of the PRISMA methodology. The first stage involves selection, which is performed by applying several filters to the Scopus database using a predetermined search string. The filters applied in this study selection are the year of publication (2019–2025), the type of document (article), the stage of publication that has been finalized, and the type of source (journals, English-language studies, and open-access). Based on the entire application of the filter, 634 articles were identified from 1,404 studies. From these 634 articles, the second stage of selection was carried out. Namely, the selection was carried out by the predetermined inclusion and exclusion criteria, resulting in 25 articles that directly discussed the comparison of the performance of the ML method with traditional actuarial models in predicting mortality risk.

2.5 Data Extraction

The 25 selected articles from the study selection exercise were extracted using a structured template that included publication information (author, year, title, journal, citations), methodology (dataset, sample size, observation period), methods compared (types of ML methods and traditional actuarial models), comparison results (performance metrics and values), application context (medical, insurance), interpretability and implementation aspects, and main conclusions. This structured template was implemented to make it easier to obtain a comprehensive overview of the research status of this topic.

2.6 Quality Assessment

This study employed the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) criteria [10] to evaluate the quality of 25 articles that had passed the study selection process. With a total of 22 STROBE criteria, namely title and abstract, introduction (background/rationale and objectives), methods (study design, setting, participants, variables, data sources/measurement, bias, study size, quantitative variables, and statistical methods), results (participants, descriptive data, outcome data, main results, and other analyses), discussion (key results, limitations, interpretation, generalisability), and other information (funding), the study results were divided into three categories. Categorisation is performed using a specific value standard, namely high-quality category studies ($\geq$ 80%), medium (60%–79%), or low ($<$ 60%), based on 22 STROBE criteria relevant to research in this topic. Table 1 shows examples of 5 articles assessed based on the STROBE criteria, and Figure 1 shows the STROBE score distribution.

Table 1: Example of Article Assessment based on STROBE Criteria

No	STROBE Criteria	Article 1	Article 5	Article 15	Article 20	Article 25
			Title and abstract			
1.	Title and abstract	✓	✓	✓	✓	✓
			Introduction			
2.	Background/rationale	✓	✓	✓	✓	✓
3.	Objectives	✓	✓	✓	✓	✓
			Methods			
4.	Study Design	✓	✓	✓	✓	✓
5.	Setting	✓	-	✓	✓	✓
6.	Participants	✓	-	✓	✓	✓
7.	Variables	✓	✓	✓	✓	✓
8.	Data sources/ measurement	✓	✓	✓	✓	✓
9.	Bias	-	-	-	-	-
10.	Study size	✓	-	✓	-	✓
11.	Quantitative variables	✓	✓	✓	✓	✓
12.	Statistical Methods	✓	✓	✓	✓	✓
			Results			
13.	Participants	✓	-	✓	-	✓
14.	Descriptive Data	✓	-	✓	✓	✓
15.	Outcome data	✓	✓	✓	✓	✓
16.	Main results	✓	✓	✓	✓	✓
17.	Other analyses	-	-	✓	✓	✓
			Discussion			
18.	Key results	✓	✓	✓	✓	✓
19.	Limitations	✓	✓	✓	✓	-
20.	Interpretation	✓	✓	✓	✓	✓
21.	Generalizability	✓	✓	✓	✓	-
			Other Information			
22.	Funding	✓	✓	✓	✓	-
	Total	91%	68%	95%	86%	82%

2.7 Data Analysis and Synthesis

Data analysis was conducted with several approaches. The first approach was to analyze the performance comparison between ML methods and traditional actuarial models based on various metrics, including AUC, accuracy, and C-Index. The second approach is bibliometric analysis, which includes publication trends, geographical and institutional distribution of research, citation patterns, and major journals based on the collected study data. The third analysis is a methodological analysis to determine the distribution of the number of ML methods and traditional actuarial models often used in comparison, as well as their application context. Based on these three analyses, research gaps will be identified, and the existing gaps in the literature on this topic will be analyzed to provide a clearer direction for future research. All analysis results are visualized in graphs and diagrams to facilitate the interpretation and communication of the findings, ranging from PRISMA diagrams and publication trend charts to other visual representations relevant to this research.

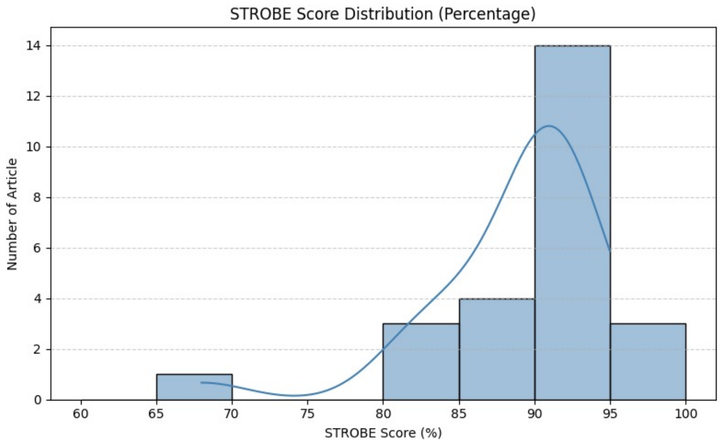

Figure 1: STROBE Score Distribution of 25 Selected Articles

2.8 Data Visualization

Data visualization is applied in this research to facilitate the more effective interpretation and communication of findings. The data visualization carried out in this study is a PRISMA diagram in the form of a flow chart describing the entire study selection process, a temporary trend graph related to the publication of the comparison of the ML method with the actuarial model in the form of a bar graph, heat maps that visualize the geographical distribution of the most productive countries contributing articles, data tables that present a comparison of the methods used for each article, bar graphs as visualizations comparing ML methods with traditional actuarial models, line graphs showing the results of methodological evolution analysis, performance metrics between methods in the form of dot plots, spider charts visualizing the strengths and weaknesses of ML methods and traditional actuarial models on various dimensions, and plots of identified research gaps in the form of bar graph visualizations.

2.9 Limitations of the Methodology

Applying the SLR research method, along with the PRISMA methodology, to this study aimed to employ a comprehensive and systematic approach. However, there are still some limitations that need to be recognized. These limitations are that the studies used for research only come from the Scopus database, examining English-language articles, methodological heterogeneity between studies that limits the ability to make direct comparisons, potential selection bias in the inclusion or exclusion criteria set, temporal limitations on publishing articles from 2019 – 2025, and the possibility of dominance of certain authors or institutions. To overcome these limitations, the researchers ensured maximum transparency in the methodology documentation and carefully interpreted the results.

3 Results and Discussion

3.1 Research Results

3.1.1 Selection Process and Study Characteristics

According to the Scopus database, a search string determined using the PICOS framework identified 1,404 relevant studies. From the 1,404 studies, screening was then carried out by applying several filters, namely year of publication (2019 – 2025), the type of document, namely the article, the final publication stage, the type of source in the form of journals, English-language studies, and open access, which resulted in 634 articles. Next, a selection was made based on the established inclusion and exclusion criteria. This selection process resulted in 25 articles comparing the ML method with traditional actuarial models. A flowchart illustrating the application of the PRISMA methodology for study selection is presented in Figure 2.

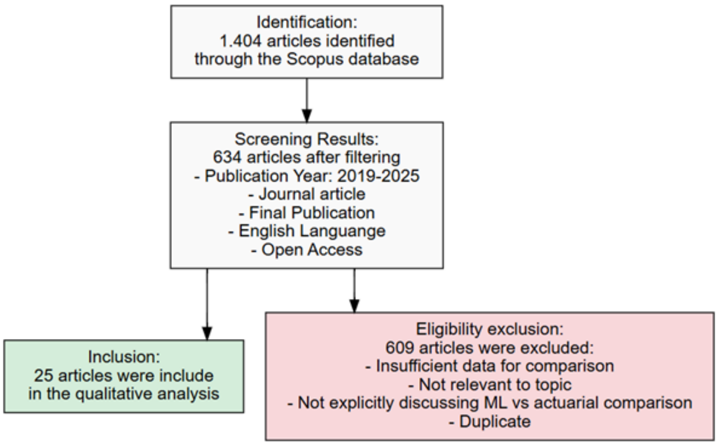

Figure 2: PRISMA Diagram of the Study Selection Process

The publication time trend shows a steady increase in research on this topic. From no explicit comparison of ML methods with traditional actuarial models in 2019, it increased to four articles in both medical and insurance contexts in 2020. And again experienced an up-and-down pattern in the interval 2021 – 2024 with a total of six articles in 2021, then decreased in 2022 to a total of five articles, and increased again in 2023 with six articles, followed by a decrease in the number to three articles in 2024 and no relevant research in 2025 when data collection was carried out in this study for both medical and insurance contexts. This trend is visualized as a line chart according to context (medical or insurance), presented in Figure 3. It can be seen that the medical context excels with 19 articles compared to the insurance context, which has published only six articles. Nonetheless, this trend still indicated an increasing interest in comparative analysis of this topic over the past five years.

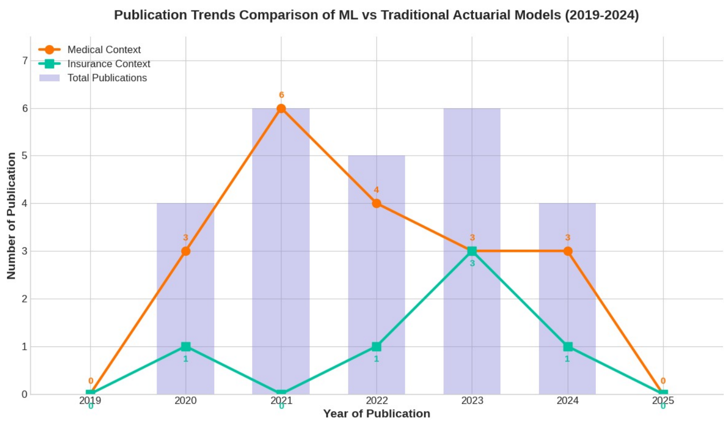

Figure 3: Publication Trend of ML Method Comparison with Actuarial Model

For the STROBE quality assessment, several important findings emerged. The first important finding was the variation in the studies' quality. There are 96% of articles used in this study that showed good methodological quality according to the 22 STROBE criteria set, this quality assessment was not only decriptive but also served to guide the interpretation of findings. In synthesizing the results, greater attention was given to studies with higher methodological rigor. Conversely, findings from studies with weaker reporting of bias or limited generalizability were interpreted more cautiously. In addition, temporal trends were also a finding that influenced this STROBE assessment, as recent studies have shown a polarization in article strength, balancing an increasing number of high-quality studies with a decreasing number of low-quality studies. Finally, there are findings related to methodological limitations in addressing bias and generalizability that influence the selection of models to apply in practice. Based on the findings and assessment of these various aspects, three categories can be concluded, with each category comprising a specific number of articles: 24 high-quality articles, one medium-quality article, and no low-quality articles.

There is a dominance of European (44%) and international collaboration (36%), with a high percentage for geographical distribution. Followed by Asia, North America, and Africa with a fairly low rate of 12%, 8%, and 4%, respectively. The most productive institution contributing articles on this topic is Seoul National University, with two published articles. The other articles came from various leading universities, including the University of Oxford, the University of Cambridge, and the National University of Singapore, each contributing 1 article. This geographical distribution is presented in Figure 4. The journal that most frequently publishes on this topic is BioMed Central (BMC), with four articles published, although these articles span different branches of health science. Based on the data, 68% of the journals that published this topic focused on healthcare, followed by 12% in insurance and 20% in other fields.

Figure 4: Geographic Distribution of the Most Productive Institutions Contributing Articles

3.1.2 Method Comparison

Table 2 presents the distribution of methods compared to the 25 articles that have passed the selection process. The most frequently used ML methods for comparison are Random Forest (20%) and Extreme Gradient Boosting (XGBoost) (13%). Other methods that can be considered to be used as a comparison are the Decision Tree method (7%), Recurrent Neural Networks (RNN) (7%), Gradient Boosting Model (GBM) (7%), Support Vector Machine (SVM) (6%), and developed methods such as Elastic Net, TabNet, GT-A Model, and AutoML (2%). As for traditional actuarial models, the most frequently used model is Traditional Logistic Regression (31%). Several other models can be considered for use in comparing with ML methods, namely the Cox Proportional Hazards (19%), Generalized Linear Model (GLM) (16%), Lee-Carter Model (6%), and other prediction models specifically used in the medical field, including Lung Allocation Score (LAS) and Clinical Risk Index for Babies II (CRIB-11) (3%). The most frequently used methods for comparison, depending on the context, produce different results. In the medical context, Random Forest (58%) and Traditional Logistic Regression (47%) were the most dominant, whereas in the insurance context, Gradient Boosting Machine (GBM) (33%) and Generalized Linear Models (GLM) (50%) were the most used. The results of this context-based method comparison are presented in Figure 5.

Table 2: Distribution of Methods in the 25 Articles Examined

No	Author (Year)	Citation Number	Machine Learning Method	Traditional Actuarial Model
1	Boo, Y & Choi, Y (2020) [11]	17	Multilayer Perceptron (MLP), Decision Tree	Traditional Logistic Regression
2	Bithew et al. (2020) [12]	22	Random Forest, Logistic Regression, K-Nearest Neighbors (KNN)	Traditional Logistic Regression
3	Maier et al. (2020) [13]	15	Random Survival Forest (RSF)	Cox Proportional Hazards Model
4	Lee et al. (2021) [14]	30	LASSO, Ridge Regression, Elastic Net, Random Forest, SVM, XGBoost	TIMI, GRACE, ACTION
5	Li et al. (2021) [15]	18	Random Survival Forest (RSF)	Cox Proportional Hazards
6	Lee et al. (2021) [16]	16	Random Forest	CRIB-II, Traditional Logistic Regression
7	Brahmbhatt et al. (2022) [17]	15	LASSO, Random Forest	LAS Model (clinical model)
8	Garcia-Montemayor et al. (2020) [18]	12	Random Forest	Traditional Logistic Regression
9	Sinha et al. (2023) [19]	10	Random Forest, Neural Networks, XGBoost, Weighted SVM	EuroSCORE II, Traditional Logistic Regression
10	McDonnell et al. (2023) [20]	35	TabNet, XGBoost	Traditional Logistic Regression
11	Shen et al. (2024) [21]	3	GT-A (GNN + Transformer)	Lee-Carter Model
12	Chen & Khaliq (2022) [22]	7	RNN (LSTM, BiLSTM, GRU)	Lee-Carter Model
13	Clemente et al. (2023) [23]	6	Gradient Boosting Model (GBM)	Generalized Linear Model (GLM)
14	Vagliano et al. (2022) [24]	5	AutoML	Traditional Logistic Regression
15	Nakamura et al. (2022) [25]	4	RNNSurv, DeepSurv	Cox Proportional Hazard
16	Nistal-Nuño (2022) [26]	11	TE, Random Forest, XGBoost, Naïve Bayes, Bayesian Network	Traditional Logistic Regression
17	Chia et al. (2021) [27]	9	Decision Tree	Cox Proportional Hazards, Traditional Logistic Regression
18	Chou & Ghimire (2021) [28]	8	Random Forest	Linear Regression
19	Lopes et al. (2023) [29]	4	XGBoost	Traditional Logistic Regression
20	Andrade & Valencia (2023) [30]	4	Fuzzy Random Survival Forest (FRSF)	Cox Proportional Hazards
21	Kovacs et al. (2021) [31]	3	Decision Tree	Generalized Linear Model (GLM)
22	Penny-Dimri et al. (2023) [32]	3	Decision Tree, Random Forest, GBM	Cox Proportional Hazards

No	Author (Year)	Citation Number	Machine Learning Method	Traditional Actuarial Model
23	Huber et al. (2023) [33]	2	Random Forest, SVM, XGBoost	Traditional Logistic Regression
24	Wilson et al. (2024) [34]	1	GBM, Artificial Neural Networks (ANN)	Generalized Linear Model (GLM)
25	Kagerbauer et al. (2024) [35]	1	Random Forest, GBM, XGBoost, Deep Learning, Stacked Ensembles	Generalized Linear Model (GLM)

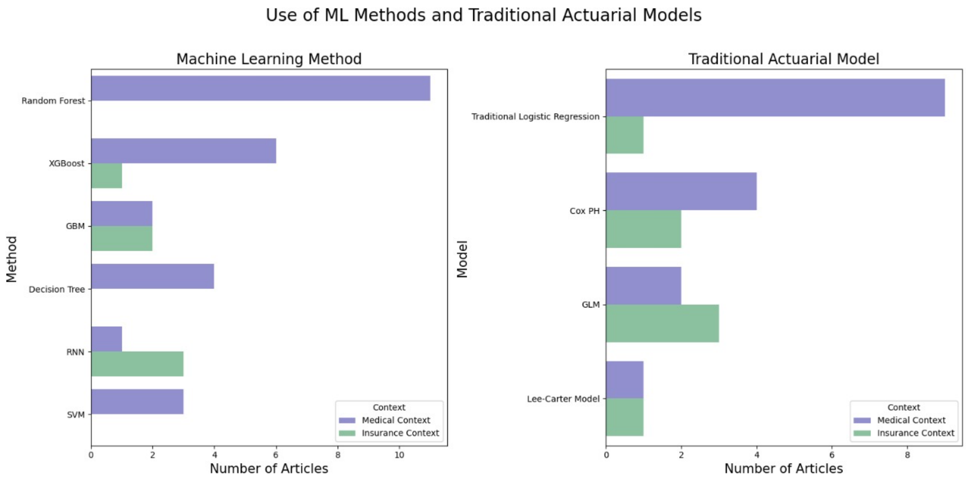

Figure 5: Comparison of Methods Based on Context

The methodological evolution analysis in this study identified three phases of development. The initial research, conducted between 2019 and 2021, involved more comparisons between the Random Forest and traditional logistic regression. Then, there was a change in the second phase, from 2022 to 2023, which showed an increase in the Traditional Logistic Regression and the Cox Proportional Hazards used as a comparison. The last phase, spanning 2024–2025, marks the emergence of the Generalized Linear Model (GLM), which is beginning to be used and compared with the Gradient Boosting Machine (GBM), particularly in the insurance context. In addition, a hybrid approach uses two methods at once by integrating the strengths of ML methods and traditional actuarial models starting in 2022. A visualization of the results of the methodological evolution analysis in this study is presented in Figure 6.

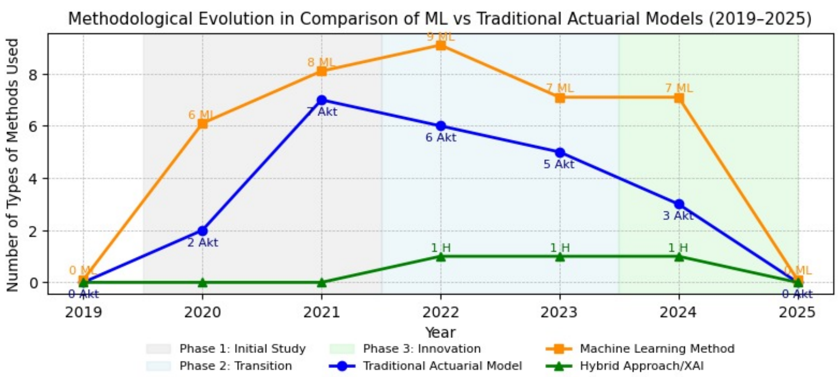

Figure 6: Methodological Evolution Analysis Results

3.1.3 Performance Comparison

Of the 25 articles that have gone through the selection process, 60% of the total articles state that the machine learning method is far superior. While 16% of the 25 articles show the opposite result, namely, the traditional actuarial model is far superior, the remaining 4% show that there is no difference in mortality prediction results between ML methods and conventional actuarial models. Based on the visualization in Figure 7, it can be seen that, in general, the ML method is far superior based on the accuracy (59.2%), sensitivity (35.8%), mean absolute error (MAE) (22.2%), Area Under the Curve (AUC) (17%) values, which are more significant. However, if focused on the context, the results will vary. In the medical context, the ML method excels due to the increase in accuracy (59.2%) in the death rate scenario. Another aspect of the insurance context that yields varied results is that the ML method is far superior for short-term mortality projections, which is inversely proportional to the traditional actuarial model. In contrast, the latter is superior for long-term mortality projections.

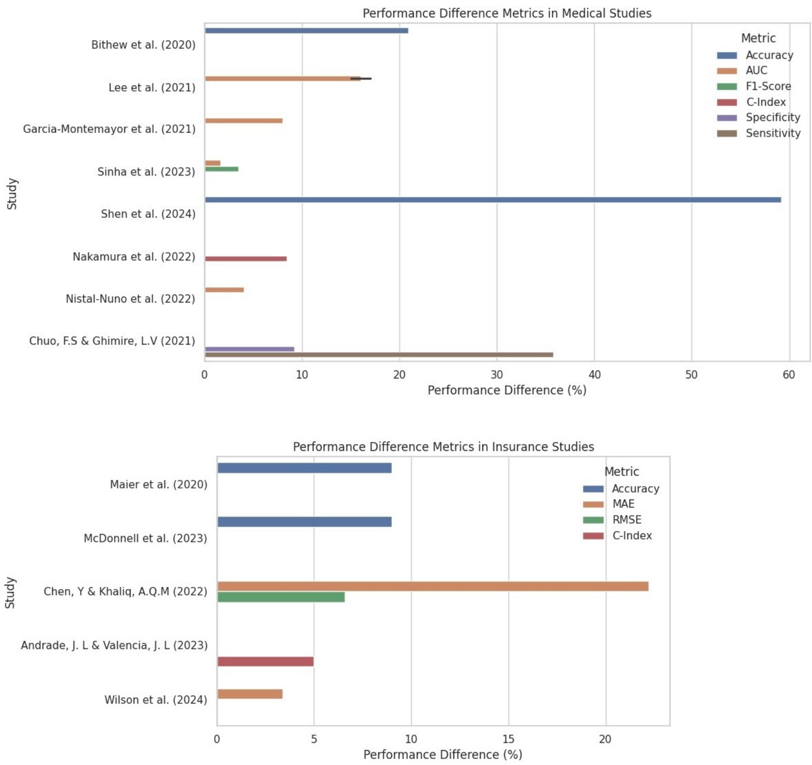

Figure 7: Performance Metrics in Medical Studies (top) and Insurance Studies (bottom)

3.1.4 Relative Strengths and Limitations

To review this section, several aspects are compared between the ML method and traditional actuarial models: data requirements, interpretability, predictive accuracy, adaptability, regulatory compliance, long-term stability, capturing non-linearities, and computational complexity. ML methods and traditional actuarial models have different relative strengths and limitations. The ML method has advantages in terms of predictive accuracy (85/100), computational complexity (60/100), capturing non-linearity (90/100), data requirements (70/100), and adaptability (80/100). These advantages are offset by weaknesses in interpretability (35/100), regulatory compliance (45/100), and long-term stability (60/100)—the advantages and disadvantages of the traditional actuarial model. The strengths of the traditional actuarial model were long-term stability (85/100), regulatory compliance (85/100), and interpretability (90/100), while the weaknesses were computational complexity (15/100), data requirements (30/100), capturing non-linearity

(20/100), adaptability (50/100), and precision accuracy (65/100). The relative strengths and limitations were visualized using spider charts presented in Figure 8.

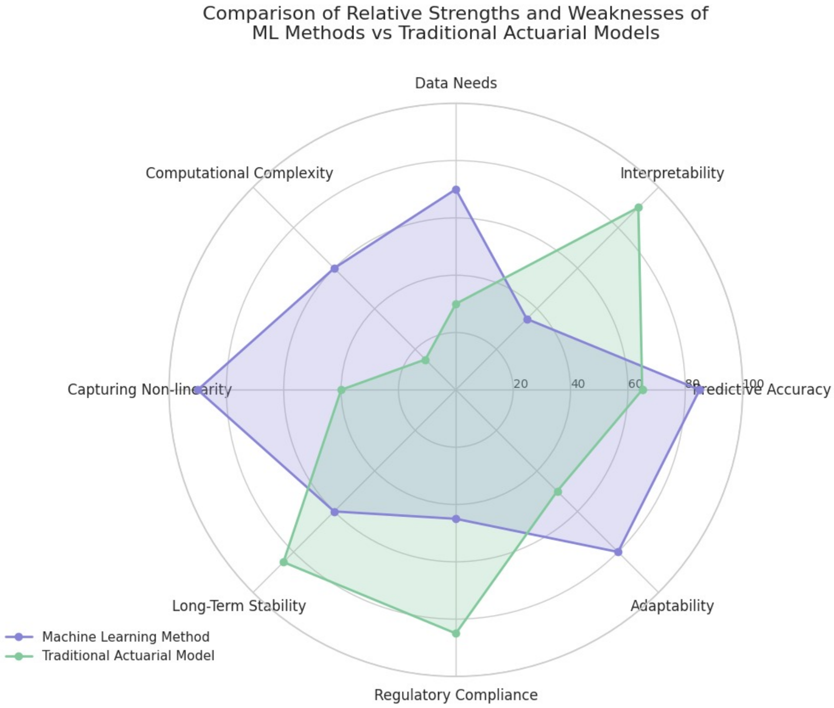

Figure 8: Comparison of Relative Strengths and Weaknesses

3.1.5 Research Gaps

Of the 25 articles examined, some areas have underdeveloped research analyses, commonly referred to as gaps. Methodological gaps refer to limited comprehensive external validation in 68% of the articles, limited predictive evaluation in homogeneous populations in 72% of the articles, and limited testing in homogeneous populations in 64% of the articles. The technical side of the gap refers to the need for a systematic hybrid approach (88%), improved bias reduction methods (20%), and model calibration to mitigate the risk of tail events in predictions (32%). The implementation gap refers to an insufficient regulatory framework for ML methods in actuarial science (32%) and limited cost-benefit analysis due to specific factors (40%). The interpretability gap refers only to the methods used to interpret complex ML models (60%). All of these gaps are presented in Figure 9, which concludes that there is a significant need for further research on this topic.

3.2 Discussion

3.2.1 Interpretation of Main Findings

Based on the identification using the SLR method, several consistent patterns were found in comparing the performance of the ML method with traditional actuarial models in predicting mortality risk. The most significant finding in this study is that the ML method is significantly superior to the conventional actuarial model in 60% of cases. This result suggests good potential for this methodology to improve the accuracy of predicting mortality risk. However, when narrowed down by context (medical or actuarial), the results suggest that the superiority of ML methods is uneven for each context. In particular, ML methods consistently excel in the medical context, especially for Random Forest and Extreme Gradient Boosting (XGBoost) methods, which can effectively capture the complexity and non-linearity of medical data well. In contrast, the more moderate and variable superiority in the insurance context, especially for long-term projections, makes the traditional actuarial model the winner, with the Cox Proportional Hazards model as the representative in predicting mortality risk.

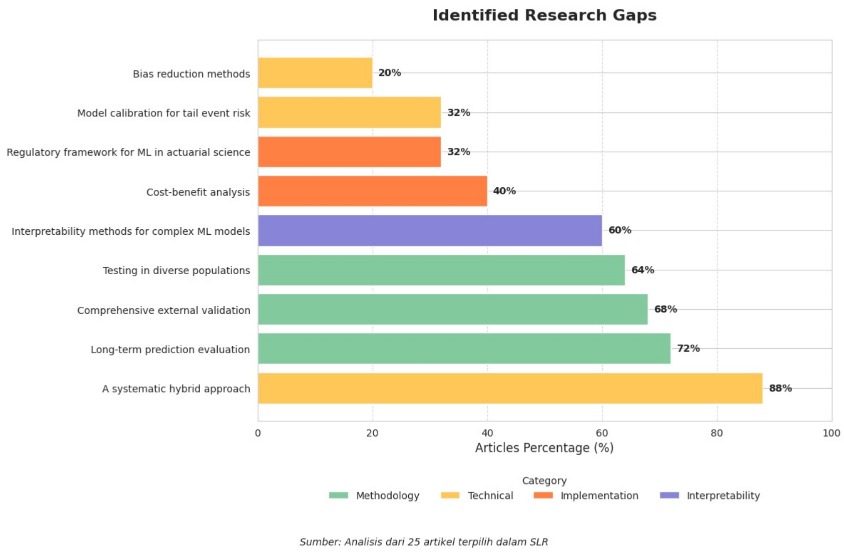

Figure 9: Plot of Identified Research Gaps

An interesting result of this comparative study is the dominance of Random Forest over Traditional Logistic Regression and Gradient Boosting Machine (GBM) over Generalized Linear Model (GLM), especially in 2024. The advantage of the Random Forest lies in its inherent balance between high prediction accuracy and reasonably good interpretability. These advantages are compared with those of Traditional Logistic Regression, which remains the standard in survival analysis because it involves fairly complex computations, making the comparison between the two interesting.

3.2.2 Accuracy and Interpretability Challenges

When selecting a methodological approach, an actuarial practitioner considers both predictive accuracy and interpretability. Based on this research, ML methods excel in accuracy but have poor interpretability scores, especially for Deep Learning methods, due to their "black box" nature. Interpretability is an important aspect of actuarial analysis. That is why only a few ML methods can be used in the actuarial industry. It cannot be denied that in the context of the insurance industry, transparent decisions that can be explained and defended objectively are needed. Thus, traditional actuarial models are still maintained today. This is due to the relatively higher level of interpretability in predicting mortality risk. Additionally, conventional actuarial models continue to ensure regulatory compliance. However, based on the methodological evolution results identified, many researchers are trying to resolve this gap and realize that in the future, predictions can be made not only with one methodology but by combining the strengths of both.

3.2.3 Implications for Practice and Research

The results of this study have important implications for various stakeholders in the actuarial and insurance world in predicting mortality risk. For actuarial practitioners, the results indicate that ML methods significantly enhance prediction accuracy, although they must still consider specific needs, such as short-term predictions or contexts involving complex data. For long-term applications that require high interpretability, traditional actuarial models are a better choice. However, it is interesting that 88% of the articles identified the need for a systematic hybrid approach to improve the quality of mortality risk prediction. Thus, actuarial practitioners must master programming and AI to use ML methods that align with the times.

For regulators, it is essential to consider that technological developments will continue to advance. Therefore, it is necessary to develop new standards for the interpretability and stability policies of machine learning methods in the actuarial industry, given the strength of these methods. This process should not overlook the importance of conducting a thorough review of the method to be applied. For researchers or academics, significant research gaps are identified through the SLR. Starting from the need for more comprehensive external validation, to better long-term evaluation. Additionally, there are limitations to the interpretation of ML methods in the insurance industry, which is a significant highlight for the development of future innovations.

3.2.4 Strengths and Limitations of the Review

The main strength of this SLR is that it is a comprehensive search strategy with an in-depth analysis of 25 articles specifically comparing ML methods with traditional actuarial models. Other strengths include quality assessment using the STROBE instrument and systematic identification of research gaps. Behind these advantages, some limitations need to be recognized, namely the use of databases only from Scopus, the focus on English-language articles, the potential bias of publications that support the results of ML methods positively, and the relatively recent emergence of this research domain, which limits the long-term perspective.

3.2.5 Future Research Directions

Based on the research conducted using the SLR method, with several gaps identified, several promising research directions will be explored in the future. These research directions are the development of a hybrid approach that integrates the theoretical basis of traditional actuarial models with the predictive accuracy capabilities of the ML method in predicting mortality risk. Additionally, a multi-institutional external validation study assesses the generalizability of performance comparisons for each article published on this topic, ensuring that they adhere to the same standards. Another research direction is the development of XAI (Explainable Artificial Intelligence) techniques in actuarial science to overcome the gap between the advantages of ML methods and traditional actuarial models, thereby enabling the use of ML methods in the actuarial industry. Finally, an important research direction is to comprehensively analyze the adoption of ML methods in actuarial practice to take advantage of technological developments and improve accuracy in mortality risk prediction.

4 Conclusion

This study compares machine learning (ML) methods and traditional actuarial models in predicting mortality risk through a Systematic Literature Review (SLR) approach with the PRISMA methodology. Based on the analysis of 25 selected articles, it was found that ML methods, especially Random Forest and XGBoost, have superior prediction accuracy compared to traditional actuarial models such as Traditional Logistic Regression and Cox Proportional Hazards. However, the main challenges in applying machine learning (ML) methods are the low level of interpretability, difficulty in long-term validation, and compliance with insurance regulations, which are still obstacles to fully adopting this technology in the actuarial industry.

This study also found that while machine learning (ML) methods have shown advantages in the medical context, their application in the insurance industry still has limitations, especially in more accurate long-term mortality projections using traditional actuarial models. Therefore, a hybrid approach is needed that combines the strengths of machine learning (ML) methods in improving prediction accuracy with the advantages of traditional actuarial models in interpretability and regulatory compliance.

In addition, this study identified various research gaps that need to be further explored, such as the need for more comprehensive external validation, the development of bias reduction methods in prediction, and the use of Explainable Artificial Intelligence (XAI) techniques to improve the transparency of machine learning (ML) methods in life insurance. The implications of these findings highlight the need for clearer regulatory standards and the development of more adaptive prediction models to meet the needs of the actuarial industry in the digital era.

As a recommendation, future research could explore hybrid approaches that combine the interpretability of traditional actuarial models and regulatory elements with the higher predictive accuracy of machine learning (ML) methods. Furthermore, additional studies are required to evaluate the effectiveness of the XAI approach in enhancing transparency and trust in ML models within the actuarial industry.

CRediT Authorship Contribution Statement

Vita Nuraini contributed to the methodology, formal analysis, writing - original draft, and visualization. Nina Fitriyati contributed to conceptualization, data curation, validation, writing – review & editing, supervision, and project administration.

Declaration of Generative AI and AI-assisted technologies

The author utilized the generative AI tool ChatGPT (version 4, OpenAI) to assist in developing ideas, generating code, and formatting in LaTeX. Additionally, Grammarly was employed to support grammar correction and the refinement of English language usage. All outputs produced by these tools were carefully reviewed and manually revised to ensure precision, originality, and adherence to academic standards.

Declaration of Competing Interest

The authors declare no competing interests.

Funding and Acknowledgments

This research received no external funding. We thank the reviewers for their insightful feedback and helpful suggestions, which have enhanced the quality of this work.

Data and Code Availability

The dataset was retrieved from Scopus.com and is publicly accessible. The visualizations were generated and executed using Google Colab.

References

[1] S. Levantesi, A. Nigri, and G. Piscopo, "Longevity risk management through machine learning: State of the art," *Insur. Mark. Co.*, vol. 11, no. 1, pp. 11–20, 2020. DOI: 10.21511/ins.11(1).2020.02.

[2] R. Lee and L. R. Carter, "Modeling and forecasting u.s. mortality," *J. Am. Stat. Assoc.*, vol. 87, pp. 659–671, 1992.

[3] R. Nopour, L. Erfannia, N. Mehrabi, M. Mashoufi, A. Mahdavi, and M. Shanbehzadeh, "Comparison of two statistical models for predicting mortality in covid-19 patients in iran," *Shiraz E Med. J.*, vol. 23, no. 6, pp. 659–671, 2022. DOI: 10.5812/semj.119172.

[4] B. Avanzi, G. Taylor, M. Wang, and B. Wong, "Machine learning with high-cardinality categorical features in actuarial applications," *ASTIN Bull.*, vol. 54, no. 2, pp. 213–238, 2024. DOI: 10.1017/asb.2024.7.

[5] N. Bishop, "Application of machine learning techniques in insurance underwriting," *J. Actuar. Res.*, vol. 2, no. 1, pp. 1–3, 2024. DOI: 10.47941/jar.1756.

[6] B. Mahohoho, C. Chimedza, F. Matarise, and S. Munyira, "Artificial intelligence-based automated actuarial pricing and underwriting model for the general insurance sector," *Open J. Stat.*, vol. 14, no. 03, pp. 294–340, 2024. DOI: 10.4236/ojs.2024.143014.

[7] M. J. Page et al., "The prisma 2020 statement: An updated guideline for reporting systematic reviews," *BMJ*, vol. 372, 2021. DOI: 10.1136/bmj.n71.

[8] Y. I. Alzoubi, A. Mishra, and A. E. Topcu, "Research trends in deep learning and machine learning for cloud computing security," *Artif Intell Rev*, vol. 57, no. 5, p. 132, 2024. DOI: 10.1007/s10462-024-10776-5.

[9] M. A. Ayanwale, R. R. Molefi, and S. Oyeniran, "Analyzing the evolution of machine learning integration in educational research: A bibliometric perspective," *Discover Education*, vol. 3, no. 1, p. 47, 2024. DOI: 10.1007/s44217-024-00119-5.

[10] E. von Elm, D. G. Altman, M. Egger, S. J. Pocock, P. C. Gøtzsche, and J. P. Vandenbroucke, "Strengthening the reporting of observational studies in epidemiology (strobe) statement: Guidelines for reporting observational studies," *BMJ*, vol. 335, no. 7624, pp. 806–808, Oct. 2007. DOI: 10.1136/bmj.39335.541782.AD.

[11] Y. Boo and Y. Choi, "Comparing logistic regression models with alternative machine learning methods to predict the risk of drug intoxication mortality," *International Journal of Environmental Research and Public Health*, vol. 17, no. 3, 2020. DOI: 10.3390/ijerph17030897.

[12] F. H. Bitew, S. H. Nyarko, L. Potter, and C. S. Sparks, "Machine learning approach for predicting under-five mortality determinants in ethiopia: Evidence from the 2016 ethiopian demographic and health survey," *Genus*, vol. 76, no. 1, 2020. DOI: 10.1186/s41118-020-00106-2.

[13] M. Maier, H. Carlotto, S. Saperstein, F. Sanchez, S. Balogun, and S. Merritt, "Improving the accuracy and transparency of underwriting with artificial intelligence to transform the life-insurance industry," *AI Magazine*, vol. 41, no. 3, pp. 78–93, 2020. DOI: 10.1609/AIMAG.V41I3.5320.

[14] W. Lee et al., "Machine learning enhances the performance of short and long-term mortality prediction model in non-st-segment elevation myocardial infarction," *Scientific Reports*, vol. 11, no. 1, pp. 1–14, 2021. DOI: 10.1038/s41598-021-92362-1.

[15] Y. Li, M. Chen, H. Lv, P. Yin, L. Zhang, and P. Tang, "A novel machine-learning algorithm for predicting mortality risk after hip fracture surgery," *Injury*, vol. 52, no. 6, pp. 1487–1493, 2021. DOI: 10.1016/j.injury.2020.12.008.

[16] J. Lee, J. Cai, F. Li, and Z. A. Vesoulis, "Predicting mortality risk for preterm infants using random forest," *Scientific Reports*, vol. 11, no. 1, pp. 1–9, 2021. DOI: 10.1038/s41598-021-86748-4.

[17] J. M. Brahmbhatt et al., "The lung allocation score and other available models lack predictive accuracy for post-lung transplant survival," *Journal of Heart and Lung Transplantation*, vol. 41, no. 8, pp. 1063–1074, 2022. DOI: 10.1016/j.healun.2022.05.008.

[18] V. Garcia-Montemayor et al., "Predicting mortality in hemodialysis patients using machine learning analysis," *Clinical Kidney Journal*, vol. 14, no. 5, pp. 1388–1395, 2020. DOI: 10.1093/ckj/sfaa126.

[19] S. Sinha et al., "Comparison of machine learning techniques in prediction of mortality following cardiac surgery: Analysis of over 220 000 patients from a large national database," *European Journal of Cardio-Thoracic Surgery*, vol. 63, no. 6, ezad183, 2023. DOI: 10.1093/ejcts/ezad183.

[20] K. McDonnell, F. Murphy, B. Sheehan, L. Masello, and G. Castignani, "Deep learning in insurance: Accuracy and model interpretability using tabnet," *Expert Systems with Applications*, vol. 217, p. 119 543, 2023. DOI: 10.1016/j.eswa.2023.119543.

[21] Y. Shen, X. Yang, H. Liu, and Z. Li, "Advancing mortality rate prediction in european population clusters: Integrating deep learning and multiscale analysis," *Scientific Reports*, vol. 14, no. 1, pp. 1–16, 2024. DOI: 10.1038/s41598-024-56390-x.

[22] Y. Chen and A. Q. M. Khaliq, "Comparative study of mortality rate prediction using data-driven recurrent neural networks and the lee–carter model," *Big Data and Cognitive Computing*, vol. 6, no. 4, 2022. DOI: 10.3390/bdcc6040134.

[23] C. Clemente, G. R. Guerreiro, and J. M. Bravo, "Modelling motor insurance claim frequency and severity using gradient boosting †," *Risks*, vol. 11, no. 9, pp. 1–20, 2023. DOI: 10.3390/risks11090163.

[24] I. Vagliano et al., "Can we reliably automate clinical prognostic modelling? a retrospective cohort study for icu triage prediction of in-hospital mortality of covid-19 patients in the netherlands," *International Journal of Medical Informatics*, vol. 160, p. 104 688, Apr. 2022. DOI: 10.1016/j.ijmedinf.2022.104688.

[25] K. Nakamura et al., "Risk of mortality prediction involving time-varying covariates for patients with heart failure using deep learning," *Diagnostics*, vol. 12, no. 12, pp. 1–14, 2022. DOI: 10.3390/diagnostics12122947.

[26] B. Nistal-Nuño, "Machine learning applied to a cardiac surgery recovery unit and to a coronary care unit for mortality prediction," *Journal of Clinical Monitoring and Computing*, vol. 36, no. 3, pp. 751–763, 2022. DOI: 10.1007/s10877-021-00703-2.

[27] A. H. T. Chia et al., "Explainable machine learning prediction of icu mortality," *Informatics in Medicine Unlocked*, vol. 25, p. 100 674, 2021. DOI: 10.1016/j.imu.2021.100674.

[28] F. S. Chou and L. V. Ghimire, "Machine learning for mortality prediction in pediatric myocarditis," *Frontiers in Pediatrics*, vol. 9, no. April, pp. 1–8, 2021. DOI: 10.3389/fped.2021.644922.

[29] R. R. Lopes et al., "Temporal validation of 30-day mortality prediction models for transcatheter aortic valve implantation using statistical process control – an observational study in a national population," *Heliyon*, vol. 9, no. 6, e17139, 2023. DOI: 10.1016/j.heliyon.2023.e17139.

[30] J. L. Andrade and J. L. Valencia, "A fuzzy random survival forest for predicting lapses in insurance portfolios containing imprecise data," *Mathematics*, vol. 11, no. 1, 2023. DOI: 10.3390/math11010198.

[31] D. Kovacs, D. R. Msanga, S. E. Mshana, M. Bilal, K. Oravcova, and L. Matthews, "Developing practical clinical tools for predicting neonatal mortality at a neonatal intensive care unit in tanzania," *BMC Pediatrics*, vol. 21, no. 1, pp. 1–10, 2021. DOI: 10.1186/s12887-021-03012-4.

[32] J. C. Penny-Dimri, C. Bergmeir, C. M. Reid, J. Williams-Spence, L. A. Perry, and J. A. Smith, "Tree-based survival analysis improves mortality prediction in cardiac surgery," *Frontiers in Cardiovascular Medicine*, vol. 10, no. July, pp. 1–8, 2023. DOI: 10.3389/fcvm.2023.1211600.

[33] M. Huber, P. Schober, S. Petersen, and M. M. Luedi, "Decision curve analysis confirms higher clinical utility of multi-domain versus single-domain prediction models in patients with open abdomen treatment for peritonitis," *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, pp. 1–12, 2023. DOI: 10.1186/s12911-023-02156-w.

[34] A. A. Wilson, A. Nehme, A. Dhyani, and K. Mahbub, "A comparison of generalised linear modelling with machine learning approaches for predicting loss cost in motor insurance," *Risks*, vol. 12, no. 4, 2024. DOI: 10.3390/risks12040062.

[35] S. M. Kagerbauer et al., "Susceptibility of automl mortality prediction algorithms to model drift caused by the covid pandemic," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, pp. 1–13, 2024. DOI: 10.1186/s12911-024-02428-z.