International Journal of Artificial Intelligence, Data Science, and Machine Learning
Grace Horizon Publication | Volume 7, Issue 3, 64-70, 2026
ISSN: 3050-9262 | https://doi.org/10.63282/3050-9262.IJAIDSML-V7I3P109

Original Article

Predictive Underwriting: How Machine Learning Is Reducing Loss Ratios in Modern Insurers

Jalees Ahmad
Independent Researcher, USA.

Received On: 15/06/2026 ***Revised On: 18/07/2026*** ***Accepted On: 25/07/2026*** ***Published On: 02/08/2026***

Abstract - *Insurance underwriting is shifting from generalised linear models toward ensemble learning and text-derived features, and the shift is usually justified by an appeal to loss-ratio improvement. This review examines how well that justification is actually supported. It surveys the peer-reviewed evidence on model performance in insurance pricing, the use of natural language processing to recover signal from unstructured records, and the fairness and explainability constraints that govern deployment, and it separates what the academic literature establishes from what is asserted in industry reporting. Three findings emerge. First, the performance advantage of gradient boosting and related ensemble methods over generalised linear models is well established in the pricing literature, with consistent gains in discriminatory power across motor, health and property lines. Second, the link from that statistical gain to a realised loss-ratio improvement is far more weakly evidenced: the frequently cited figures for loss-ratio reduction originate in consultancy reporting rather than peer-reviewed study, and this review could not locate independent replication. Third, a 2025 theoretical result establishes an analytical relationship between model validation performance and loss-ratio degradation, and shows that model improvement carries diminishing marginal returns, which reframes model investment as a prioritisation problem rather than a uniform good. The review also argues that fairness in algorithmic underwriting is a prudential concern and not only a conduct one, since proxy-driven mispricing raises loss-ratio volatility. It concludes that the defensible position is augmented underwriting, in which models triage and humans adjudicate, and identifies the absence of independent loss-ratio evidence as the field's most consequential gap.*

Keywords - *Predictive Underwriting, Machine Learning, Loss Ratio, Insurance Pricing, Gradient Boosting, Natural Language Processing, Explainable AI, Algorithmic Fairness, Actuarial Science.*

1. Introduction

The insurance enterprise exists to quantify uncertainty and price it. For most of the twentieth century that quantification rested on generalised linear models and static actuarial tables, an approach whose virtue was interpretability and whose limitation was rigidity.

A structural feature of the traditional arrangement compounds that limitation. Underwriting prices risk from historical assumptions and demographic averages, while claims handling observes the realised risk only after a loss has occurred. The two functions have historically operated with limited feedback between them, so information generated at the point of loss does not systematically return to the point of pricing. The economic consequence is the classical adverse selection problem [1]: where the insurer's information is poorer than the applicant's, the portfolio drifts toward risks whose premiums do not reflect their exposure. Figure 1 sets out the divide and what closing it is supposed to achieve.

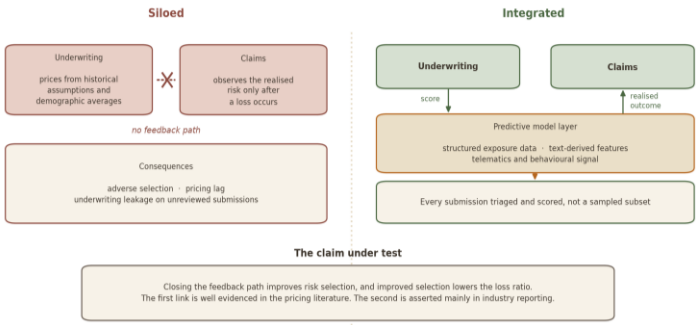

Fig 1: The Knowledge Gap between Underwriting and Claims, and the Integrated Arrangement Intended to Close it. The Claim under Test in this Review is the Second Link, From Improved Risk Selection to a Realized Loss-Ratio Reductio

Machine learning is now widely proposed as the mechanism for closing that loop, and the proposal is usually accompanied by quantified promises about loss ratios. This review takes those promises as its object of study rather than its premise.

1.1. Contribution

This paper contributes:

- A synthesis of the peer-reviewed evidence on model performance in insurance pricing and underwriting, distinguishing what is established from what is asserted.
- An explicit separation of two claims that are routinely conflated: that machine learning improves discriminatory power, which is well supported, and that it reduces loss ratios, which is much less so.
- A treatment of the loss-ratio error framework and its diminishing-returns implication for where model investment should be directed.
- An argument that algorithmic fairness in underwriting is a prudential risk, not solely a conduct risk.

1.2. What this review does not claim

It reports no new empirical work. Where quantified industry figures are reproduced, they are attributed to their source and identified as industry-reported. Section 8 states the limitations, of which the most significant is that the central commercial claim in this field rests on evidence this review could not independently verify.

2. Review Methodology

Sources were identified through Crossref and arXiv searches across the intersecting domains of insurance pricing, actuarial machine learning, explainable AI, algorithmic fairness, and natural language processing applied to clinical and claims text. Search terms combined method terms (generalised linear model, gradient boosting, XGBoost, random forest, neural network, natural language processing) with domain terms (insurance pricing, underwriting, claim frequency, claim severity, loss ratio, reserving).

Inclusion required that a source be either peer-reviewed, a recognised preprint with verifiable provenance, or an official regulatory or standards document. Industry and consultancy publications were admitted only where a claim originates with them and no peer-reviewed equivalent exists, and every such source is identified as industry-reported at the point of use.

Every reference in this paper was verified against Crossref or arXiv for title, venue, year and authorship before inclusion. Sources published by houses appearing on recognised predatory-publisher lists were excluded.

The distinction that structures the review. Claims are sorted into two categories throughout: those supported by peer-reviewed empirical study, and those originating in industry reporting without independent replication. This distinction is maintained deliberately, because the field's most-repeated numbers fall into the second category and the conflation of the two is, in this author's assessment, the principal weakness of the existing literature on this subject.

3. From Generalised Linear Models to Ensemble Learning

3.1. The methodological shift

Generalised linear models were designed for inference. They require the modeller to specify functional form and interaction terms explicitly, which makes their output interpretable and their assumptions auditable, and which also limits their capacity to represent high-dimensional non-linear structure.

Gradient boosting takes the opposite approach. Friedman's formulation [2] constructs an additive model in a stagewise manner, fitting each successive learner to the residual of the ensemble so far, and thereby discovers interaction structure without it being specified. The XGBoost implementation [3] made the approach computationally practical at scale through regularisation, sparsity handling and parallelised tree construction, and it is now the default choice for tabular insurance data.

3.2. What the pricing literature establishes

The evidence for improved discriminatory power in insurance applications is consistent.

Guelman [4] applied gradient boosting trees to auto insurance loss cost modelling and reported superior predictive performance against the conventional benchmark. Clemente et al. [5] modelled motor claim frequency and severity with gradient boosting and reported the same direction of result. Blier-Wong et al. [6] survey machine learning across property and casualty pricing and reserving and find the pattern general rather than dataset-specific. Wüthrich [7] extends the case into individual claims reserving. Krúpová et al. [8] report on explainable boosting machines for claim severity and frequency, indicating that the interpretability penalty is itself becoming negotiable.

Verbelen et al. [9] address a different lever, showing that telematics data carries predictive power beyond conventional rating factors in motor pricing. That result matters here because it establishes that the gains attributed to machine learning are partly gains from new data rather than from new algorithms, a distinction the industry literature rarely draws.

Figure 2 summarises the reported discriminatory power ranges against the interpretability each model class affords.

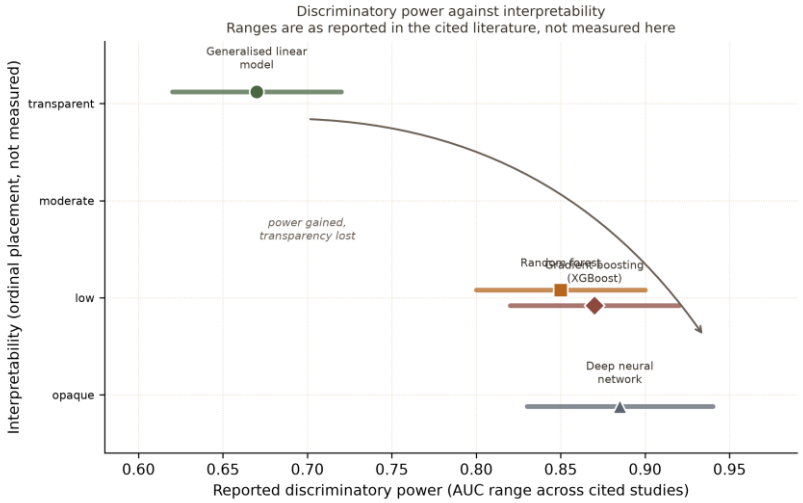

Fig 2: Discriminatory Power against Interpretability across Model Classes. Ranges are as Reported across the Cited Literature, Not Measured in this Review, and Vary Substantially by Line of Business and Dataset. Interpretability is an Ordinal Placement rather than a Measurement

Table 1 gives the same comparison with sources attached.

Table 1: Model Classes, Reported Performance, and Evidentiary Basis

Model Class	Reported Discriminatory Power	Principal Strength	Evidentiary Basis
Generalised linear model	Lower, and the standing benchmark	Interpretability, regulatory familiarity, stable extrapolation	Established practice; benchmark in [4][5][6]
Random forest	Consistently above GLM	Robustness to noise, handles mixed data types	[6]
Gradient boosting / XGBoost	Highest among tabular methods in most reported comparisons	Strong performance with tractable stability	[2][3][4][5][8]
Deep neural network	Comparable to or above boosting on some tasks	Representation learning on unstructured inputs	[6]; limited insurance-specific evidence

The consistent finding across [4][5][6] is that gradient boosting outperforms the generalised linear benchmark on tabular insurance data. What none of these studies report is a realised loss-ratio outcome? They report statistical performance on held-out data, which is a different quantity, and Section 5 takes up the distance between the two.

3.3. Automation of complex underwriting

A frequently cited operational benefit is triage automation in life and health underwriting, where a substantial share of applications historically required manual review. Figures circulating in industry reporting for accuracy of automated underwriting decisions, cycle-time reduction and cost reduction are specific and repeated, but this review was unable to locate peer-reviewed studies reporting them under stated methodology. They are therefore reproduced here only with that qualification attached, and Section 8 records the gap.

4. Natural Language Processing and Unstructured Records

A large share of an insurer's record is narrative: adjuster notes, clinical documentation, correspondence, incident reports. Conventional pricing models cannot consume it, so the signal it carries is lost by construction rather than by choice.

Natural language processing addresses this by converting narrative into structured features. Figure 3 shows the pipeline and its principal constraint.

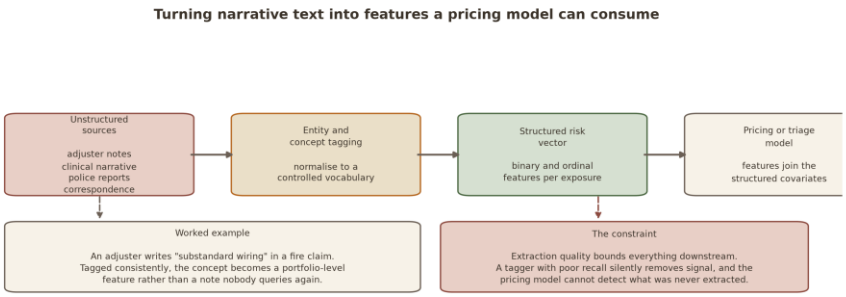

Fig 3: Extraction Pipeline Converting Narrative Sources into Features Consumable by a Pricing or Triage Model, with the Extraction-Quality Constraint that Bounds Everything Downstream

The mechanism is worth stating concretely. An adjuster recording "substandard wiring" in a fire claim narrative is generating portfolio-level intelligence that a structured database will not retain. Tagged consistently against a controlled vocabulary, that concept becomes a queryable feature, and a rising frequency of the tag within a segment becomes an underwriting signal rather than an observation nobody revisits.

In health underwriting the analogous target is behavioural and social risk factors documented in narrative clinical text but absent from structured fields. Related work on unsupervised discovery of patient subgroups from electronic health records [10] demonstrates that clinically meaningful structure is recoverable from such records, which is the enabling result for the underwriting application even though it was not conducted for underwriting purposes.

A caution the industry literature omits. Extraction quality bounds every downstream claim. A tagger with poor recall removes signal silently, and no pricing model can detect what was never extracted. Reported performance improvements from text-derived features are therefore conditional on extraction quality that is seldom reported alongside them.

5. Loss Ratio: The Claim and Its Evidence

This section addresses the paper's title claim directly.

5.1. The mechanism

The argued causal chain runs: better discrimination produces better risk selection; better selection produces closer alignment between premium and expected loss; closer alignment reduces the loss ratio. Each link is plausible and the first is well evidenced. Underwriting leakage, meaning policies bound at prices that do not reflect their risk because relevant factors were not assessed, is the specific inefficiency the chain is supposed to remove, and comprehensive scoring of submissions rather than sampled manual review is the specific mechanism.

5.2. The evidentiary position

Widely repeated figures for loss-ratio reduction attributable to predictive underwriting originate in consultancy and industry reporting [17][18]. This review could not locate peer-reviewed studies that independently establish them.

That is a material finding rather than a bibliographic inconvenience, and it warrants stating plainly. The performance literature reports discriminatory power, not loss ratios. The loss-ratio literature is industry-produced, does not disclose methodology, portfolio composition, or period, and is generated by parties with a commercial interest in the conclusion. Between the statistical result and the financial result there is a gap that the published evidence does not close.

Table 2 reproduces the commonly cited figures with their provenance and status attached, which is the responsible way to present them.

Table 2: Commonly Cited Operational and Financial Impacts, With Provenance

Claimed impact	Typical figure cited	Source type	Status in this review
Improvement in risk assessment precision	~20%	Industry reporting [17][18]	Not independently verified
Reduction in aggregate loss ratio	2% to 5%	Industry reporting [17][18]	Not independently verified
Reduction in policy issuance cycle time	Up to 90%	Industry reporting [17]	Not independently verified
Reduction in underwriting operating cost	~40%	Industry reporting [17]	Not independently verified
Share of insurer data held in unstructured form	~80%	Industry reporting [18]	Not independently verified

Improvement in discriminatory power over GLM	Consistent, magnitude varies by line	Peer-reviewed [4][5][6][8]	Established
Predictive value of telematics beyond conventional factors	Material	Peer-reviewed [9]	Established

5.3. The loss ratio error framework

A 2025 theoretical contribution by Hedges [11] supplies what the empirical literature does not: an analytical relationship between a model's validation performance and its financial consequence. The framework introduces a loss-ratio error quantity capturing the expected degradation in loss ratio attributable to imperfect model performance.

Its most useful result is structural. Model improvement exhibits diminishing marginal returns: raising a weak model's discriminatory power yields a substantially larger loss-ratio benefit than an equivalent gain applied to an already strong model. Figure 4 illustrates the shape.

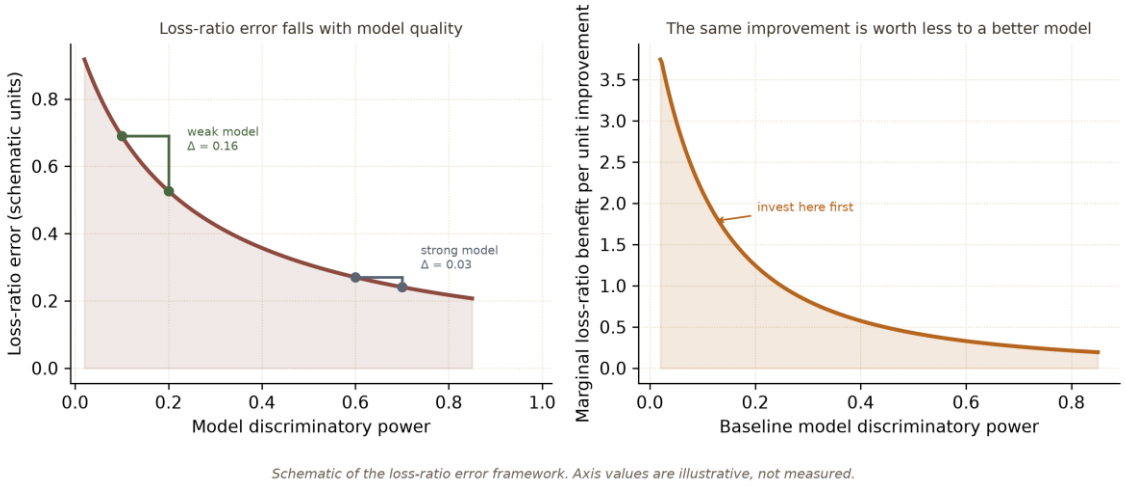

Fig 4: Schematic of the Loss-Ratio Error Framework, Showing Loss-Ratio Error Falling with Model Quality and the Marginal Benefit of Improvement Falling with Baseline Quality. Axis Values are Illustrative of the Framework's Structure, not Measured Quantities

The practical consequence is a reordering of investment priorities. Where model development has conventionally been directed at the lines of business that are already best modelled, because that is where teams and data are strongest, the framework implies the opposite: the largest financial return sits with the worst-performing models. This converts a qualitative judgement into a quantitative allocation problem, and it is the most actionable result in the current literature on this subject.

The framework is theoretical and recent. It has not, at the time of writing, been empirically validated against portfolio outcomes, and that validation is the natural next contribution.

6. Governance, Explainability, and Fairness

6.1. The explainability requirement

Ensemble and deep models are not natively interpretable, which creates friction with regulatory expectations and consumer-facing justification. The dominant response has been post-hoc explanation, principally LIME [12], which fits a locally faithful interpretable surrogate around an individual prediction, and Shapley-value attribution, which allocates a prediction across its input features.

Rudin 13] argues forcefully against relying on post-hoc explanation for high-stakes decisions, on the ground that an explanation [of a black box is not the black box, and that in consequential domains inherently interpretable models should be preferred. Underwriting is squarely a high-stakes decision domain in Rudin's sense, and the emergence of explainable boosting machines in insurance applications [8] suggests the field is moving toward that position rather than away from it.

6.2. Fairness is not solved by removing the attribute

Excluding a protected characteristic from a model does not remove its influence. Correlated variables reconstruct it: geography, occupation and deprivation measures all carry residual information about protected characteristics, and a model denied the attribute will route around it through the proxies. Mehrabi et al. [14] survey the mechanisms comprehensively.

Lindholm et al. [15] address the insurance case directly, formalising discrimination-free pricing and showing that naive exclusion is insufficient while a construction exists that removes both direct and indirect discrimination. Israni et al. [16] evaluate performance and fairness jointly for non-life pricing, which is the practically important framing because it treats the two as a trade-off surface rather than as sequential constraints.

Figure 5 contrasts the two approaches.

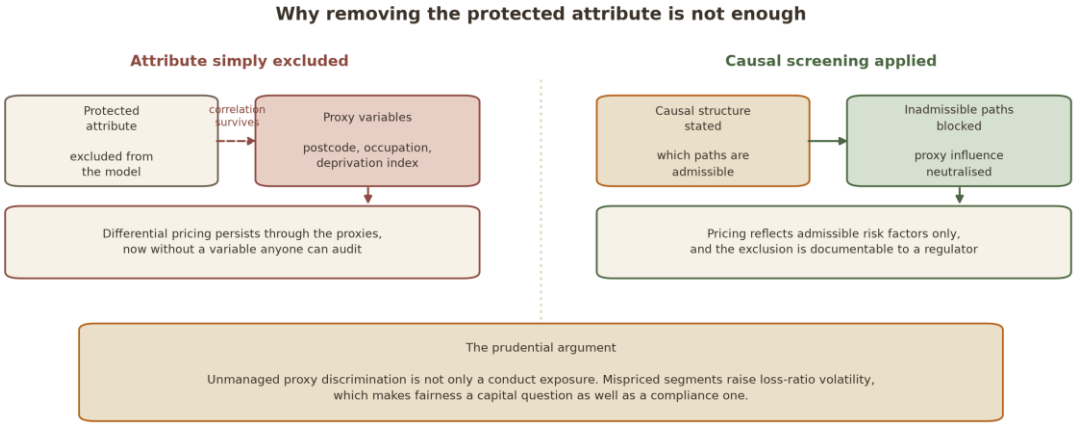

Fig 5: Naive Exclusion of a Protected Attribute against Causal Screening. Under Exclusion the Correlation Survives through Proxies and becomes Harder to Audit, because there is no Longer a Variable to Inspect

6.3. Fairness as a prudential matter

The framing this review advances is that proxy-driven differential pricing is not only a conduct exposure. A segment priced on a proxy is a segment priced on something other than its risk, and mispricing raises the variance of realised losses against expectation. Unmanaged algorithmic bias therefore raises loss-ratio volatility, which makes it a capital question as well as a compliance one.

This reframing has a practical consequence: it moves fairness assessment out of the conduct function alone and into the model risk and capital functions, where it acquires quantitative treatment and budget.

7. Discussion: Augmented Rather Than Automated

The evidence supports a narrower conclusion than the industry framing.

What is established: ensemble methods outperform generalised linear models on discriminatory power across insurance lines [4][5][6][8]; new data sources carry genuine incremental signal [9]; text-derived features are technically recoverable [10]; naive exclusion of protected attributes does not produce fair pricing [14][15]; and model improvement carries diminishing marginal returns [11].

What is not established: that any of this has produced the loss-ratio reductions routinely claimed. The gap is not evidence of absence, and the mechanism is plausible. But a field that has produced this volume of methodological work without producing a single independently published portfolio-level loss-ratio study has a gap where its commercial justification should be.

The operational conclusion follows from the explainability and fairness constraints rather than from performance. Models can triage comprehensively where manual review reaches only a sample, and that is a real capability. Adjudication of the complex, the ambiguous and the contestable remains with human underwriters, both because those cases are where model confidence is lowest and because accountability for a declined or loaded policy must rest with a person who can explain it. Augmented underwriting is not a compromise position pending better models. It is what the governance constraints require.

8. Limitations

The central commercial claim is unverified. The loss-ratio figures this paper reproduces are industry-reported and were not independently confirmed. They are presented with provenance in Table 2 and should not be read as findings.

This is a narrative review, not a systematic one. Section 2 states the search approach but does not follow a formal protocol with screening counts and dual review. Selection bias in source identification cannot be excluded.

Coverage is uneven by line of business. The pricing evidence is weighted toward motor, which is the best-served line in the published literature. Health, property and specialty lines are less well represented, and generalisation across lines is assumed rather than demonstrated.

The loss-ratio error framework is recent and unvalidated. Hedges [11] is a December 2025 preprint. Its structural result is compelling and its empirical validation is outstanding.

Regulatory treatment is described, not analysed. Fairness and explainability obligations vary substantially by jurisdiction, and this review does not attempt a comparative regulatory analysis. No primary data. Every quantity in this paper originates in a cited source. Nothing was measured here.

9. Conclusion

Machine learning has demonstrably improved the discriminatory power of insurance pricing models. That much the peer-reviewed literature establishes consistently,

across methods, datasets and lines of business. Whether it has reduced loss ratios is a different question, and the honest answer is that the published evidence does not currently support the confident figures in circulation. The improvement in statistical performance is real; its translation into financial outcome is asserted rather than demonstrated. The loss-ratio error framework [11] is valuable precisely because it makes that translation analytically tractable for the first time, and its empirical validation against portfolio outcomes is the single most useful contribution the field could now make.

Two further conclusions follow. Fairness in algorithmic underwriting should be treated as a prudential exposure rather than only a conduct one, because proxy-driven mispricing raises loss-ratio volatility and therefore has capital consequences. And the defensible operating model is augmented underwriting, in which models triage at a scale manual review cannot reach while humans adjudicate the cases where accountability and explanation matter most. The field does not need better claims. It needs a published portfolio-level study, with disclosed methodology and a properly established pre-deployment baseline that measures what predictive underwriting actually does to a loss ratio.

References

[1] M. Rothschild and J. Stiglitz, "Equilibrium in competitive insurance markets: An essay on the economics of imperfect information," *The Quarterly Journal of Economics*, vol. 90, no. 4, pp. 629-649, 1976. doi: 10.2307/1885326

[2] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189-1232, 2001. doi: 10.1214/aos/1013203451

[3] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785-794. doi: 10.1145/2939672.2939785

[4] L. Guelman, "Gradient boosting trees for auto insurance loss cost modeling and prediction," *Expert Systems with Applications*, vol. 39, no. 3, pp. 3659-3667, 2012. doi: 10.1016/j.eswa.2011.09.058

[5] C. Clemente, G. R. Guerreiro and J. M. Bravo, "Modelling motor insurance claim frequency and severity using gradient boosting," *Risks*, vol. 11, no. 9, art. 163, 2023. doi: 10.3390/risks11090163

[6] C. Blier-Wong, H. Cossette, L. Lamontagne and E. Marceau, "Machine learning in P&C insurance: A review for pricing and reserving," *Risks*, vol. 9, no. 1, art. 4, 2020. doi: 10.3390/risks9010004

[7] M. V. Wüthrich, "Machine learning in individual claims reserving," *Scandinavian Actuarial Journal*, vol. 2018, no. 6, pp. 465-480, 2018. doi: 10.1080/03461238.2018.1428681

[8] J. Krúpová, A. Rachdi and Q. Guibert, "Explainable boosting machine for predicting claim severity and frequency in car insurance," *European Actuarial Journal*, 2026. doi: 10.1007/s13385-026-00461-y

[9] R. Verbelen, K. Antonio and G. Claeskens, "Unravelling the predictive power of telematics data in car insurance pricing," *Journal of the Royal Statistical Society Series C: Applied Statistics*, vol. 67, no. 5, pp. 1275-1304, 2018. doi: 10.1111/rssc.12283

[10] Y. Wang, Y. Zhao, T. M. Therneau, E. J. Atkinson, A. P. Tafti, N. Zhang, S. Amin, A. H. Limper, S. Khosla and H. Liu, "Unsupervised machine learning for the discovery of latent disease clusters and patient subgroups using electronic health records," *Journal of Biomedical Informatics*, vol. 102, art. 103364, 2020. doi: 10.1016/j.jbi.2019.103364

[11] C. E. Hedges, "A theoretical framework bridging model validation and loss ratio in insurance," arXiv preprint arXiv:2512.03242, December 2025.

[12] M. T. Ribeiro, S. Singh and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1135-1144. doi: 10.1145/2939672.2939778

[13] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, pp. 206-215, 2019. doi: 10.1038/s42256-019-0048-x

[14] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM Computing Surveys*, vol. 54, no. 6, pp. 1-35, 2021. doi: 10.1145/3457607

[15] M. Lindholm, R. Richman, A. Tsanakas and M. V. Wüthrich, "Discrimination-free insurance pricing," *ASTIN Bulletin*, vol. 52, no. 1, pp. 55-89, 2021. doi: 10.1017/asb.2021.23

[16] T. Israni, K. Daly, C. Condon and E. Wolsztynski, "Dual evaluation of performance and fairness from machine learning models for non-life insurance pricing," *British Actuarial Journal*, 2026. doi: 10.1017/s1357321725100317

[17] Capgemini Research Institute, *World Property and Casualty Insurance Report: The Underwriter of the Future*, 2024. Industry report.

[18] Accenture, *The Future of Insurance Underwriting: Balancing Speed and Precision with AI*, 2024. Industry report.

[19] National Association of Insurance Commissioners, *Model Bulletin on the Use of Artificial Intelligence Systems by Insurers*, adopted December 2023.

[20] Society of Actuaries Research Institute, *Avoiding Unfair Bias in Insurance Applications of AI Models*, 2022.

[21] European Parliament and Council, *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*, 2024.

[22] European Parliament and Council, *Directive 2009/138/EC on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II)*, 2009