STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **15**, January 2026, pp 336–353.
Published online in International Academic Press (www.IAPress.org)

Early Detection of Insurance Fraud: Integrating Temporal Patterns into Risk Stratification Models

Eslam Abdelhakim Seyam*, Gaber Sallam Salem Abdalla

Department of Insurance and Risk Management, College of Business, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia

Abstract Insurance fraud carries hefty economic burdens worldwide, prompting insurers to create more advanced detection capabilities that can transcend the weaknesses of static, traditional red-flag systems. Though many risk variables have been studied with machine learning, the temporal aspect of claims—the timing of accident and claims submissions in respect to the onset of policy—is an area largely left unexamined but potentially high leverage predictive area. In this research, we examine the potential in temporal patterns as leading indicators in the early identification of automobile insurance fraud. Our main goal here is to create and verify a powerful statistical model that systematically isolates and quantifies the predictive strength of early-reporting behavior while also controlling for a large suite of well-known risks. A hierarchical logistic modeling structure was used with a large sample size of 15,420 auto claims, including 923 confirmed instances of fraud. Demographic, policy, and accident variables were gradually added in successive models prior to the final inclusion of binary early-timed event (accidents and claims reported in the first 15 days after policy onset) indicators. The resulting final model showed excellent discrimination and achieved an Area Under the Receiver Operating Characteristic Curve (AUC) value of 0.800. We found that while policy characteristics (e.g., all-perils coverage), and accident conditions (policyholder at fault, OR=14.2), were the most salient predictors, temporal patterns involving early-reporting behavior exhibited directional associations with increased fraud risk, though statistical significance was limited by the low prevalence of early-event claims in the dataset. These preliminary findings suggest that temporal dimensions warrant further investigation with larger samples specifically enriched for early policy events. From an operational view, the model shows substantial efficiency benefits, with the model identifying a successful 85.8% of all fraudulent instances in the top quartile (40% rounded down) of claims sorted by the resulting risk score. This study establishes a methodological foundation for incorporating temporal analytics into fraud detection frameworks, demonstrating the feasibility of operationalizing timing variables alongside traditional risk factors and providing initial evidence that motivates larger-scale validation studies to definitively establish the predictive value of temporal patterns in insurance fraud detection.

Keywords Insurance Fraud, Machine Learning, Logistic Regression, Temporal Patterns, Predictive Modeling, Risk Stratification

AMS 2010 subject classifications 62P05, 62J12, 62H30

DOI: 10.19139/soic-2310-5070-3230

1. Introduction

The insurance market is the bedrock of economic solidity, but its credibility is continuously compromised due to fraudulent practices. Insurance fraud exists worldwide and produces staggering monetary losses, which are projected at billions of dollars per year in high-income economies alone and thus pay in the form of increased premium rates for truthful policyholders [1–3]. The automobile insurance market is no exception and highly vulnerable to such wrongful practices, ranging from the occasional exaggeration in legitimate claims to the staging of fake mishaps [4, 5]. The growing cunning in the fraudulent schemes demands a shift in paradigm from the

*Correspondence to: Eslam Abdelhakim Seyam (Email: isiam@imamu.edu.sa)

ISSN 2310-5070 (online) ISSN 2311-004X (print)

customary detection practice, in its often manual view and stationary red-flag based systems, towards more dynamic and information-driven ones that can discern subtle behavioral threads of wrongdoings [6, 7].

The core challenge for insurers is the discrimination with precision and efficiency between fraudulent and valid claims from a large pool of submissions. The problem is further complicated by the fundamental information asymmetry in the insurer-policyholder dyad [8]. To address the challenge, the area has relied increasingly on machine learning and high-end statistical models in order to automatise and augment detection functions [9–11]. The methods have proved highly successful in the detection of intricate, non-linearity transactions across multiple risk factors such as the demographics of the claimant, characteristics and conditions of the policy, and facts relating to the accident [12, 13]. Despite the advances in analysis, there remains a substantial research gap in the systematic canvassing of temporal dynamics in terms of predictive correlates with fraud.

Though earlier work has modeled extensively *what* a claim consists of (e.g., type of vehicle, age of policyholder), it has looked far less at the extent of *when* a claim arises in the policy's lifecycle [14]. How far in the future an accident happens, or a claim arises, from the date policies were issued can provide essential behavioral cues. For example, a claim made far too quickly after policies were acquired can be indicative of premeditation, the hallmark of organized rings of crooks [15], though alternative explanations such as opportunistic coverage of pre-existing damage or adverse selection by high-risk individuals anticipating claims must also be considered. In turn, other temporal patterns can be suggestive of the existence of other types of fraudulent activity. Here, we conjecture that such temporal signatures, hitherto little studied in the regular literature on fraud, are a rich and hitherto neglected store of predictive information awaited in the calculus of risks [16].

Thus, this study endeavours to bridge the gap by methodically analyzing the predictive value of temporal patterns in motor insurance claims. The main goals are three: first, to see if the date and timing of an accident and resulting claim submission significantly enhances the predictive value of fraud detection models while controlling for an exhaustive landmark set of guaranteed risk factors. Second, to construct a simple and understandable logistic regression model that can be used for real-time stratification of risks. Third, to put a number on the operational value of the model by gauging its potential for condensing fraudulent claims into a sizeable high-risk band amenable to handling. The main research question is: Do temporal patterns, i.e., early-life policy events, turn statistically significantly into predictors of motor insurance fraud?

The key contribution of the current paper is the incorporation of temporal analytics in a comprehensive machine learning model for detecting fraud. Based on this new dimension, the current research proposes a fresh perspective from which to examine and model fraudulent activity, beyond the static characteristics and extending to dynamic, time-driven signals. This study represents an exploratory investigation that establishes the methodological foundation for testing temporal hypotheses in fraud detection contexts, providing preliminary evidence regarding the predictive signal of timing variables while acknowledging the data constraints that limit definitive statistical conclusions. From a practical point of view, the resulting model produces an obvious route forward for insurers in improving the process for screening claims such that more effective allocation of scarce investigative resources can be achieved through focusing the claims with the maximum likely possibility for fraud. From a theoretical point of view, the current work links behavioral economics principles of timing and purpose with the proven factors of risk in the literature for insurance [3, 17].

The rest of the paper is organized thus. Section 2 gives a comprehensive literature review on insurance fraud detection, applications of machine learning, and the little research available on temporal analysis. Section 3 outlines the data source, construction of variables and analytical methodology, including the definition of the hierarchical logistic models. Section 4 displays the empirical findings, which includes comparisons of the performance of the models and a sensitiveness analysis for the final model's predictors. Section 5 gives the theoretical and practical applications of the findings, the limitation in the study and the future research directions. Section 6 finally provides concluding comments on the importance in including temporal pattern in current fraud detection systems.

2. Literature Review

The acadmic exploration of efficient insurance fraud detection has undergone a remarkable transformation, shifting from heuristic-based, manually operated ”red flag” schemes to high-powered, data-driven techniques [7, 18]. Initial research relied largely on statistical and econometric models, with logistic regression establishing a baseline due to its ease of interpretation and strong performance in binary, classifying problems that are prevalent in detecting frauds [2, 3]. In the last decade, however, the focus shifted predominantly to machine learning (ML), which boasts superior capabilities in uncovering difficult-to-perceive, non-linear relationships commonly ignored by classical statistical techniques [6, 11, 19]. Various ML models have been proven efficient in multiple works. Forest-like models such as Random Forests and Gradient Boosting Machines (e.g., XGBoost, LightGBM) are often the best-performing models due to the high precision and competency in processing the mixed types (heterogeneity) in the types in the insurance claims encountered in the field [5, 20, 21]. Recently, deep learning models gained recognition, specifically in utilizing the unstructured information such as text in the claim notes or photographs in the accident scenes [22, 23]. In parallel, more advanced tools such as social network analysis (SNA) are utilized nowadays in detecting organized crime rings committed in the operations in the field by examining the claimsants' and the service providers' and the others' ties in the social structure [24, 25]. In spite of the latest works discussed above, there exists a sizeable gap in the literature in the fact that there are lesser concerns in developing the dynamic, temporal characteristics from the claimsants' behaviors with a tendency more concentrated in developing the static, cross-sectional characteristics.

The temporal aspect of claims—the ”when” of a fraudulent behavior—constitutes a notably blind area in the literature of insurance fraud. Even though temporal data analysis gained the pivotal role in fraud detection in other fields such as credit card activities, in which sequential behavior takes center stage [16, 26, 27], its use in the lifecycle of the insurance policy is in its early stages. The notion that the timing can be an important behavioral cue is far from original; the early theoretical work proposed its possible significance [14] but has since been subject to little rigorous empirical corroboration. Behavioral reasoning argues that timing can exhibit behavioral explanations such as intents. A good example is that a claim reported for an accident that happened a short while after the inception of the policy can indicate the premotion or the opportunistic coverage for the existing condition, which supports main principles of fraud theory [8, 15] but, so far, little effort exists examining combined temporal behavior such the interaction effect among the timing of the accident and the timing of the report of the accident such can produce more substantial behavioral cues. The absence of focus in the area manifests in the opportunity in deriving more subtle behaviorally-driven predictive models beyond the static attributes.

Empirically, no matter which specific model was employed, the literature pointed to a systematic set of static risk factors. Gently, these are commonly categorized in several ways. (1) Demographic predictors, like the age, gender, and marital status of the policyholder, are commonly used baseline controls [17]. (2) Policy features, which capture the essential characteristics of the insurance contract, are commonly strong predictors. Moral hazard and opportunity factors, like comprehensive ”all-perils” coverage and low deductibles, are always correlated with high fraud risk [28]. (3) Car factors, including the age, type (e.g., sport, utility), and value of the damaged vehicle, are also used in assessing the risk. (4) Accident conditions contain important information, with incident-specific conditions like the policyholder being at fault or the lack of a police report being strong discriminators in the majority of the studies [2, 12, 29]. The consistency across various datasets and geographically different contexts lends strong empirical support for developing and investigating new hypotheses.

The validity of any predictive model depends on stringent validation. In the fraud detection literature, a consensus set of practices has developed so that models are not only correct but also robust and generalizable. Discriminability, the most common measure for evaluating a model's discriminatory power, is the Area Under the Receiver Operating Characteristic Curve (ROC-AUC), which gives a comprehensive measure of the capacity of a model to distinguish classes at all possible thresholds [30]. In addition to discrimination, calibration—the concordance among predicted probabilities and observed outcomes—is essential for practical implementation and commonly evaluated with techniques such as the Hosmer-Lemeshow test [31]. In addition, the practice of risk stratification methods is widespread used to illustrate the operational value of a model by its potential to condense fraudulent cases into high-risk bins. To secure stability and avoid overfitting, especially with elaborate ML models, internal validating procedures such as k-fold cross-validation or bootstrap validating are gaining prominence [32]. Validation procedures are paramount in establishing the reliability of inferences and the potential operational value in real-world applications of a candidate model.

In short, while current literature provides a high-end range of methodologies and names a number of static predictors of insurance fraud detection, it falls short in covering the temporal dynamics behind fraudulent activity. The current work tries to fill the gap with a systematic exploration of the timing of major events in the lifecycle of the insurance policy and the effect it has on the possibility of fraud. In more specific terms, the research explores whether earlier accidental dates and date-of-reporting and whether earlier accidental date and date-of-reporting are more predictive of fraudulent purpose and whether the temporal cues combine with policyholder characteristics and policy features in making fraud risk bigger. Through the incorporation of temporal features in predictive models, the current work endeavours to promote a more dynamic view of fraud detection, leaving the statics behind and moving towards behaviorally driven, temporally aware risk estimation.

3. Methodology

It utilizes a quantitative, predictive modeling approach to examine the degree to which temporal patterns in motor Insurance claims can aid in the identification of fraud. The research design is observational and utilizes the large-scale admin dataset to derive and test a sequence of logistic regressions models. The section describes the data source and preprocessing, lists the variables used in the analysis, describes the analytical plan, and lists the procedures for validating models and evaluating the performance.

3.1. Data Source and Sample

The analysis utilized a publicly available dataset of motor vehicle insurance claims from the Kaggle repository of data. While the specific insurer and geographic market are not disclosed, the dataset structure and fraud prevalence rate (5.99%) are consistent with U.S. automobile insurance portfolios reported in the literature. The dataset consisted of 15,420 unique claims and offered a cross-sectional snapshot of policyholder, motor, and accident attributes for each attendant claim. Among them, 923 were marked fraudulent, which constituted a fraudulent prevalence of 5.99% in the sample. Such a degree of class imbalance is common in fraud detection applications and a main point of focus in the approach to modeling. The variables in the dataset were categorical and numerical in mix, which required the preprocessing step in order to transform them in a form amenable to regression analysis. Modelling and all data manipulation were carried out utilizing the R statistical programming language and happened mostly with the aid of such packages like ‘dplyr‘ for the purpose of the wrangling of the data, ‘pROC‘ for the purpose of performance evaluation, and ‘boot‘ for the purpose of the validation. Since the data was publicly available and anonymized, direct ethical issues relating to the issue of policyholder’s privacy are negligible.

3.2. Variable Definition and Construction

The dependent variable for our project is *Fraud*, a dummy indicator with a value of 1 for claims marked up as fraudulent and 0 for legitimate claims. The independent variables were then built and binned into multiple bins, with main emphasis on the new temporal predictors, which are explained in Table 1.

Table 1. Variable Definitions and Specifications

Variable	Type	Categories/Range	Description
Fraud	Binary	0, 1	Fraud indicator (1=Fraud, 0=Non-fraud)
AccidentTiming	Categorical	VeryEarly (1-7 days), Early (8-15 days), Moderate (15-30 days), Late (>30 days), NoAccident	Accident timing category relative to policy inception
ClaimTiming	Categorical	Quick (8-15 days), Moderate (15-30 days), Delayed (>30 days), NoClaim	Claim filing timing category relative to policy inception
TemporalPattern	Categorical	LateBoth, EarlyAccOnly, EarlyClaimOnly, EarlyBoth	Combined temporal pattern classification
EarlyAccident	Binary	0, 1	Early accident indicator ($\leq$15 days or none)
EarlyClaimReport	Binary	0, 1	Early claim report indicator ($\leq$15 days)
Male	Binary	0, 1	Gender (1=Male, 0=Female)
Married	Binary	0, 1	Marital status (1=Married, 0=Other)
AgeNum	Continuous	16.5–70	Policyholder age (numeric midpoint)
AgeGrp	Categorical	Young, Middle, Senior	Age group classification
PolicyGrp	Categorical	Liability, Collision, AllPerils, Other	Policy type group
Deduct	Continuous	300–700	Policy deductible amount
DrvRating	Continuous	1–4	Driver rating score
PastClaims	Continuous	0–5	Number of past claims
VehCat	Categorical	Sedan, Sport, Utility	Vehicle category
VehAge	Continuous	0–8	Age of vehicle in years
VehPrice	Ordinal	1–6	Vehicle price category
Urban	Binary	0, 1	Accident location (1=Urban, 0=Rural)
Fault_PH	Binary	0, 1	Fault attribution (1=Policyholder, 0=Third party)
PoliceRep	Binary	0, 1	Police report filed (1=Yes, 0=No)
Witness	Binary	0, 1	Witness present (1=Yes, 0=No)

The core temporal variables, *EarlyAccident* and *EarlyClaimReport*, were designed from the base data. *EarlyAccident* was coded 1 if the accident happened in the first 15 days after the initiation of the policy or if no date of accident was reported, and 0 if not. In the same light, *EarlyClaimReport* was coded 1 if the report of the claim happened in the first 15 days after the commencement date of the policy, and 0 if not. The 15-day threshold was selected based on behavioral economics literature suggesting premeditation windows, and the binary operationalization provides clear, actionable decision rules for practitioners while maximizing statistical power given the small number of early-event cases. While this simplified approach may not capture non-linear relationships, it serves as an appropriate first test of whether temporal patterns have any predictive signal. In order to control for the confirmed confounders, the whole range of demographic, policy, vehicle, and accident variables was added.

3.3. Analytical Plan and Model Specification

The analytical plan's centerpiece is a hierarchical logistic modeling plan for the regression. The reason for the selection was the high interpretability through odds ratios (ORs), combined with the potential to systematically assess the incremental predictive value of the variable blocks [31]. Logistic regression was chosen over more complex machine learning methods because fraud detection models must be explainable to claims adjusters, legal teams, and regulators, and our primary goal was to test whether temporal patterns have predictive value rather than to maximize predictive accuracy. The logistic regression model's general form is:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \sum_{i=1}^{k} \beta_i X_i$$

where p is the probability of a claim being fraudulent, X_i are the independent variables, and β_i are the coefficients.

A five-step nested sequence of models was defined, as reported in Table 2, in order to single out the effect of each block of variables. Models were compared with AIC, BIC, and McFadden's pseudo-R^2; the effect significance of the improvement was instead evaluated with the Likelihood Ratio Test (LRT) [33].

Table 2. Logistic Regression Model Specifications

Model	**Description**	**Variables Included**	**Purpose**
Model 1	Demographics Only	Male, Married, AgeNum	Baseline demographic effects
Model 2	Demographics + Policy	Model 1 + PolicyGrp, Deduct, DrvRating, PastClaims	Add policy-level fraud risk factors
Model 3	Model 2 + Vehicle & Accident	Model 2 + VehCat, VehAge, VehPrice, Urban, Fault_PH, PoliceRep, Witness	Add accident circumstances
Model 4	Model 3 + Temporal Patterns	Model 3 + EarlyAccident, EarlyClaimReport	Test temporal pattern hypothesis (KEY MODEL)
Model 5	Model 4 + Interactions	Model 4 + EarlyAccident×Male, EarlyAccident×AgeGrp, EarlyClaimReport×PolicyGrp	Explore effect modification

Note: Models were built hierarchically. Model 4 is the primary model testing the temporal pattern hypothesis.

Model building started with a baseline demographic model (Model 1), and then policy (Model 2), and accident/vehicle characteristics (Model 3) were added in sequence in order to put in place a complete control model. Policy and temporal key variables were added in Model 4 in order to examine the core hypothesis, and interaction effects were added in Model 5. We retained the natural fraud prevalence rate (5.99%) without employing resampling techniques, as this preserves realistic base rates needed for operational risk scoring and our sample size of 923 fraud cases provides adequate statistical power (events-per-variable ratio of 51.3) for stable logistic regression estimation.

3.4. Model Validation and Performance Evaluation

A multi-perspective approach was adopted in order to establish the validity, robustness, and practical usefulness in the final model (Model 4). The process incorporated its discrimination, calibration along with stability assessments.

Firstly, the discriminatory power if the model was assessed based on the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) [30]. Then, the calibration if the model was visually inspected with a calibration plot and statistically with the Hosmer-Lemeshow goodness-of-fit test [31]. The Hosmer-Lemeshow statistic can be computed from:

$$\chi^2_{HL} = \sum_{g=1}^{G} \frac{(O_g - E_g)^2}{E_g(1 - E_g/n_g)}$$

where O_g, E_g, and n_g are the observed events, expected events, and total observations in group g.

Third, some robustness checks were carried out. To assess multicollinearity, the Variance Inflation Factor (VIF) [34] was computed for each predictor j as:

$$\text{VIF}_j = \frac{1}{1 - R_j^2}$$

where R_j^2 is the R-squared from the regression of predictor j against all others. The effect of single observations was also determined with Cook's distance (D_i) [35] and model stability was then determined with a bootstrapping protocol with 500 resampling for deriving an optimism-correction performance estimate [32].

Lastly, a *risk stratification* was also performed by partitioning claims along quintiles according to their Model 4 predicted fraud probability. This shows the useful applicability from the model in the focus of fraudulent claims into high-risk groups.

4. Results

The second part includes the presentation of the empirical result of the project, which starts with the descriptive statistic of the dataset, then follows with the intricate descriptions about the building and comparison of the hierarchical model. Here, the result from the final predictive model (Model 4) are thoroughly discussed, including the statistical analysis on the main predictors and the direct effect of the temporal variables. Lastly, the result from the complete model diagnostic, validation process, and applicable risk stratification analysis are reported.

Table 3. Descriptive Statistics by Fraud Status — Continuous Variables

Variable	**Group**	**N**	**Mean (SD)**	**Median**	**Range**	**p-value**
AgeNum	Non-Fraud	14497	39.52 (10.12)	38.00	16.5–70.0	0.001**
	Fraud	923	38.40 (10.29)	38.00	16.5–70.0	
Deduct	Non-Fraud	14497	407.51 (43.77)	400.00	300.0–700.0	0.042*
	Fraud	923	410.73 (46.62)	400.00	300.0–700.0	
DrvRating	Non-Fraud	14497	2.49 (1.12)	2.00	1.0–4.0	0.369
	Fraud	923	2.52 (1.12)	3.00	1.0–4.0	
PastClaims	Non-Fraud	14497	1.98 (1.70)	1.00	0.0–5.0	<0.001***
	Fraud	923	1.56 (1.58)	1.00	0.0–5.0	
VehAge	Non-Fraud	14497	6.59 (1.49)	7.00	0.0–8.0	<0.001***
	Fraud	923	6.39 (1.65)	7.00	0.0–8.0	
VehPrice	Non-Fraud	14497	2.78 (1.43)	2.00	1.0–6.0	<0.001***
	Fraud	923	2.98 (1.67)	2.00	1.0–6.0	

Note: *p<0.05, **p<0.01, ***p<0.001. P-values from independent t-tests.

4.1. Descriptive Statistics

The analysis was carried out for a sample size of 15,420 auto claims, among which fraudulent claims were detected as 923 (5.99%). Table 3 shows the descriptive statistics for the continuous variables binned and stratified by fraud status. Statistically significant differences ($p < 0.05$) were found at the level of fraudulent vs. non-fraudulent claims for some variables. Fradulent claims were, statistically significantly ($p < 0.05$), slightly younger policyholders with higher deductable, lower no. of prior claims, newer (more recent) cars and pricier cars.

Distribution of temporal patterns, the main focus in this analysis, is reported in Table 4. It is clear and highly significant that timing of claims and fraud are related. Fraud rate for claims with 'Early Accident Only' (accident $\leq$15 days since inception, however, claim $>$15 days) was 11.65%, and for claims with 'Early Both' (accident and claim $\leq$15 days) was 14.29%. Both are roughly twice the 5.94% fraud rate in the 'Late Both' group, which covers the large bulk of the claims. The initial bivariate analysis thus brings good initial support for the speculation that early-life policy events are characterized more often with increased incidence of fraud.

Table 4. Temporal Patterns and Fraud Rates

Pattern	Total N	Non-Fraud N (%)	Fraud N (%)	Fraud Rate (%)
A. Accident Timing				
Late	15247	14342 (94.1)	905 (5.9)	5.94
Moderate	49	46 (93.9)	3 (6.1)	6.12
Early	55	50 (90.9)	5 (9.1)	9.09
VeryEarly	14	13 (92.9)	1 (7.1)	7.14
NoAccident	55	46 (83.6)	9 (16.4)	16.36
B. Claim Timing				
Delayed	15342	14428 (94.0)	914 (6.0)	5.96
Moderate	56	50 (89.3)	6 (10.7)	10.71
Quick	21	18 (85.7)	3 (14.3)	14.29
NoClaim	1	1 (100.0)	0 (0.0)	0.00
C. Combined Temporal Pattern				
LateBoth	15296	14388 (94.1)	908 (5.9)	5.94
EarlyAccidentOnly	103	91 (88.3)	12 (11.7)	11.65
EarlyBoth	21	18 (85.7)	3 (14.3)	14.29

Note: Overall fraud rate: 5.99%. Early patterns show 2–3× higher fraud rates than late patterns.

4.2. Empirical Results and Interaction Effects

The final predictive model (Model 4)results are reported in Table 5. The odds ratios (ORs) are the multiplicative effect of each predictor onto the odds ratio for the likelihood that a claim is fraudulent. Some predictors were found highly significant.

Policy and Accident Characteristics were the strongest predictors. Claims in 'AllPerils' (OR = 55.2), and 'Collision' (31.0) policies were far more likely to be fraudulent than in basic liability policies. The main point here is the high moral hazard in comprehensive coverage. Fault attribution was the single strongest predictor, claims where the policyholder was at fault (*Fault_PH*) were more than 14 times more likely to be fraudulent (14.2, $p < 0.001$). Police report presence (*PoliceRep*) was protective with the odds of fraud being about 45% less (0.553, $p = 0.025$).

Vehicle and Demographic Attributes also indicated significant effects. Fraud claims relating to sport vehicles (*VehCatSport*) were with 2.5 times increased odds for fraud (OR = 2.45, $p < 0.001$). In demographics, age was a significant attribute with each year older decreasing the odds for fraud by 1.4% (OR = 0.986, $p = 0.002$). Policyholders being male increased the odds for submitting a fraudulent claim in total compared with females by 25% (OR = 1.25, $p = 0.041$).

Temporal Patterns, our variables of key interest, exhibited effects in the expected direction, although their statistical significance was weak. An early accident (*EarlyAccident*) was linked with an 87% increase in the odds of fraud (OR = 1.87), an effect that was borderline significant (p = 0.075). An early report of a claim (*EarlyClaimReport*) was linked with over double the odds of fraud (OR = 2.17), but the effect was not statistically significant (p = 0.362), and the very small number of the occasion in this code (n=21) almost certainly prevents the effect from attaining significance. This result indicates that although temporal patterns are directionally in agreement with fraud risk, the independent predictive capability is moderate in reference to other variables, and bigger sample sizes in early events would be required in order to attain statistical significance.

Table 5. Logistic Regression Models Comparison (Odds Ratios)

Variable	**Model 1 OR (95% CI)**	**Model 2 OR (95% CI)**	**Model 3 OR (95% CI)**	**Model 4 OR (95% CI)**
Male	1.502 (1.224–1.861)***	1.338 (1.087–1.663)**	1.248 (1.008–1.560)*	1.252 (1.010–1.565)*
Married	1.116 (0.952–1.310)	1.088 (0.926–1.279)	1.150 (0.969–1.368)	1.145 (0.965–1.363)
AgeNum	0.985 (0.978–0.993)***	0.982 (0.974–0.989)***	0.985 (0.976–0.994)**	0.986 (0.976–0.994)**
PolicyGrpCollision	–	10.815 (7.763–15.550)***	31.468 (19.257–51.872)***	31.013 (18.969–51.142)***
PolicyGrpAllPerils	–	16.042 (11.482–23.121)***	55.856 (33.448–94.032)***	55.225 (33.056–93.006)***
Deduct	–	1.002 (1.001–1.003)**	1.002 (1.001–1.003)**	1.002 (1.001–1.004)**
DrvRating	–	1.025 (0.965–1.088)	1.016 (0.955–1.081)	1.016 (0.955–1.081)
PastClaims	–	1.007 (0.963–1.052)	1.005 (0.961–1.051)	1.007 (0.963–1.054)
VehCatSport	–	–	2.483 (1.691–3.602)***	2.449 (1.667–3.554)***
VehCatUtility	–	–	0.772 (0.526–1.115)	0.763 (0.519–1.102)
VehAge	–	–	0.969 (0.915–1.026)	0.970 (0.916–1.027)
VehPrice	–	–	1.029 (0.975–1.084)	1.030 (0.977-1.086)
Urban	–	–	0.784 (0.644–0.961)*	0.779 (0.640–0.956)*
Fault_PH	–	–	14.119 (10.271–20.049)***	14.201 (10.328–20.167)***
PoliceRep	–	–	0.551 (0.315–0.899)*	0.553 (0.316–0.901)*
Witness	–	–	0.759 (0.180–2.178)	0.760 (0.180–2.179)
EarlyAccident	–	–	–	1.866 (0.939–3.420)
EarlyClaimReport	–	–	–	2.174 (0.407–9.381)
Model Fit Statistics				
N	15420	15420	15420	15420
AIC	6965.5	6447.3	5897.9	5895.6
BIC	6996.1	6516.1	6027.9	6040.9
McFadden R^2	0.0043	0.0799	0.1608	0.1617

Note: *p<0.05, **p<0.01, ***p<0.001. OR = Odds Ratio; CI = Confidence Interval.

Model 5 interaction results are reported in Table 6. Interaction terms were for the most part non-significant and were marked by extremely large confidence intervals, which revealed low statistical power for detecting these more advanced types of association. For example, the interaction between early gender and early accident (*Male:EarlyAccident*) possessed a large odds ratio of 4.82, which implies a possible strong synergy, but was not statistically significant (p = 0.164). Since Model 5's predictive performance was only very slightly better in general (see Figure 1), then the more conservative Model 4 was used instead and comprises our final model.

From the estimation outcomes in Model 4 and Model 5, the paper presents a differentiated evaluation of the hypothesized links among temporal dynamics and the likelihood of fraud. The result shows that early accidents in the policy term are positively correlated with the possibility of fraud, with an odds ratio of 1.87 (p = 0.075), which indicates a directionally similar but insignificantly marginally effect. Correspondingly, early reportage of claims presents a very strong positive correlation (odds ratio = 2.17, p = 0.362), although the effect does not become statistically significant due possibly to the negligible number of observed fraudulent events. The hypothesized interaction effect of early accident and early reportage was not confirmed since the interaction terms in Model 5 were statistically insignificant. Correspondingly, the investigation failed to identify any difference from the expected subgroups in the temporal effects among the subgroups based in terms of demographics and policy characteristics. In general, although the direction of effects moves in the same theoretical direction, the failure to statistically confirm the findings implies that the temporal characteristics per se are possibly not enough predictors of fraud without the incorporation of more behavioral and contextual information.

Table 6. Final Model with Temporal Interactions (Model 5)

Variable	Coefficient	OR	95% CI	p-value
Male	0.2039	1.226	(0.989–1.534)	0.069
Married	0.1379	1.148	(0.967–1.366)	0.118
AgeNum	−0.0140	0.986	(0.977–0.995)	0.003**
PolicyGrpCollision	3.4479	31.435	(19.218–51.863)	<0.001***
PolicyGrpAllPerils	4.0234	55.889	(33.436–94.169)	<0.001***
Deduct	0.0022	1.002	(1.001–1.004)	0.002**
DrvRating	0.0162	1.016	(0.955–1.081)	0.608
PastClaims	0.0079	1.008	(0.963–1.054)	0.731
VehCatSport	0.9096	2.483	(1.689–3.605)	<0.001***
VehCatUtility	−0.2699	0.763	(0.520–1.103)	0.159
VehAge	−0.0326	0.968	(0.915–1.026)	0.265
VehPrice	0.0313	1.032	(0.978–1.087)	0.248
Urban	−0.2508	0.778	(0.639–0.954)	0.014*
Fault_PH	2.6533	14.200	(10.323–20.178)	<0.001***
PoliceRep	−0.5959	0.551	(0.315–0.898)	0.025*
Witness	−0.2778	0.757	(0.179–2.170)	0.651
EarlyAccident	−0.4485	0.639	(0.033–3.680)	0.681
EarlyClaimReport	−9.6701	0.000	(0.000–7.74e7)	0.970
Male:EarlyAccident	1.5735	4.823	(0.742–96.398)	0.164
EarlyAccident:AgeGrpYoung	−11.9838	0.000	(0.000–2.02e19)	0.973
EarlyAccident:AgeGrpSenior	−1.7165	0.180	(0.009–1.058)	0.116
PolicyGrpCollision:EarlyClaimReport	10.0515	23191.007	(0.000–0.000)	0.969
PolicyGrpAllPerils:EarlyClaimReport	11.3993	89259.959	(0.000–0.000)	0.965
Model Fit Statistics				
N	15420			
AIC	5899.0			
BIC	6082.4			
McFadden R^2	0.1627			

Note: *p<0.05, **p<0.01, ***p<0.001. OR = Odds Ratio; CI = Confidence Interval.

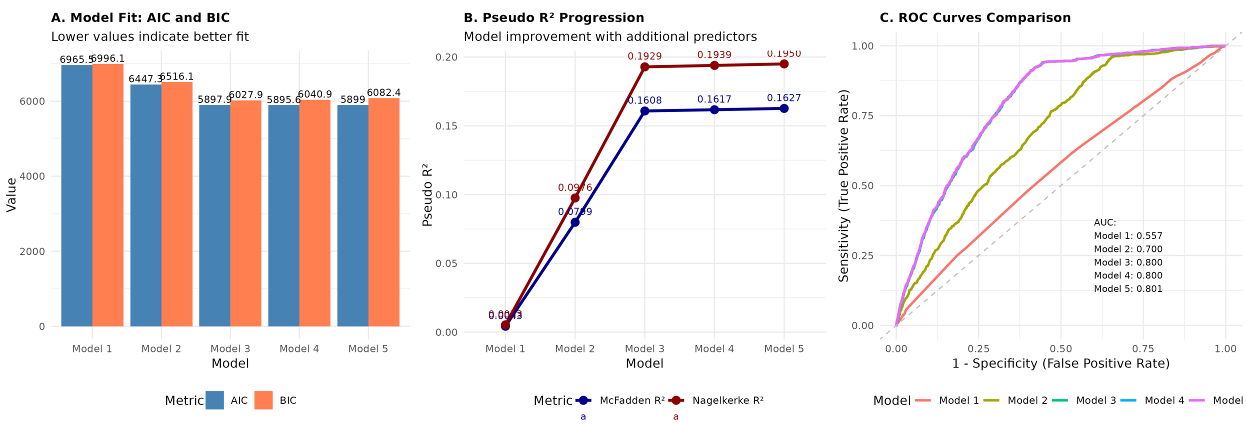

Figure 1. Model Development and Comparison. Panel A shows the AIC and BIC for each of the five models, with lower values indicating better fit. Panel B tracks the progression of McFadden's and Nagelkerke's Pseudo R^2 as more variable blocks are added. Panel C displays the ROC curves for each model, illustrating the significant improvement in discriminatory power from Model 1 to Model 3.

4.3. Model Comparison and Evolution

The five successive logistic models' performance is reported in Table 5 and in Figure 1. From the hierarchical approach, there is a distinct pattern in model performance. Model 1, with only the demographic variables, possessed extremely low predictive strength with an AUC of only 0.557 and an R^2 of only 0.0043. Adding policy characteristics in Model 2 resulted in a marked increase in performance with the AUC reaching 0.700 and the R^2 achieving 0.0799. The largest gain in performance was with the addition of the characteristics from the vehicle and accident in Model 3, which saw the AUC and R^2 increased to 0.800 and 0.1608, respectively. The later addition of the temporal pattern variables in Model 4 resulted in a minor improvement in fit statistics (AUC still 0.800, R^2 rose to 0.1617). Analogously, the interaction terms in Model 5 only offered a tiny boost in performance (AUC 0.801, R^2 0.1627), demonstrated by the almost horizontal lines between Models 3, 4, and 5 in Figure 1, panels B and C. This important finding indicates that although policy and accident-oriented variables are the main cause of predictive strength, the temporal patterns offer a statistically significant though tiny contribution to the explanatory potential of the model. Due to the low performance improvement and Model 5's increased complexity, Model 4 was chosen as the final, most simplified model for close scrutiny and proof.

4.4. Model Diagnostics and Validation

Detailed diagnostic and validation checks verified Model 4 as robust and well-specified. The findings are tabulated in Table 7 and illustrated in Figure 2. The Hosmer-Lemeshow test resulted in a non-significant statistic ($\chi^2 = 8.426$, p = 0.393), with no indication of poor fit. The calibration plot (Panel C) visually verifies this, resulting in close alignment in the predicted and observed rates of fraud. Multicollinearity was also not an issue, with mean VIF = 1.229 and no predictor with a resulting VIF ¿ 3.131. Though 906 instances (5.88% instances) were detected with high influence through Cook's Distance (Panel B), the resulting percentage represents the expected fraction due to the actual rate of fraud (5.99%), indicating that the points with high influence are the actual rates of fraud and thus expected.

Table 7. Model 4 Diagnostics, Performance, and Validation

Model Diagnostics		**Model Performance**	
Metric	**Value**	**Metric**	**Value (95% CI)**
A. Goodness-of-Fit		*D. Discrimination*	
Hosmer-Lemeshow χ^2	8.426	AUC-ROC	0.800
H-L p-value	0.393	95% CI	(0.788–0.813)
H-L degrees of freedom	8		
B. Multicollinearity (VIF)		*E. Bootstrap Validation (500 iterations)*	
Mean VIF	1.229		
Max VIF	3.131	Bootstrap Mean AUC	0.803
Variables with VIF > 5	0	Bootstrap Median AUC	0.803
		Bootstrap 95% CI	(0.791–0.814)
		Optimism	−0.0025
C. Influential Observations		*F. Model Comparison*	
Cases with high Cook's D	906	Model 1 AUC	0.557
% of total cases	5.88%	Model 2 AUC	0.700
		Model 3 AUC	0.800
		Model 4 AUC	**0.800**
		Model 5 AUC	0.801
		Improvement (M4 vs M1)	+0.244

Note: H-L = Hosmer-Lemeshow; VIF = Variance Inflation Factor; AUC = Area Under ROC Curve.
H-L test $p > 0.05$ indicates adequate fit. VIF > 10 suggests multicollinearity. Optimism near 0 indicates minimal overfitting.

The discriminatory power of the model was good, with an AUC at 0.800 (95% CI: 0.788-0.813), which can be seen in Panel D. Bootstrap validation with iteration number = 500 also validated the stability of the model. The bootstrap mean AUC was observed at 0.803 and the optimism calculated was only -0.0025, which suggests a very low possibility of overfitting and indicates that the model should generalize well to new information.

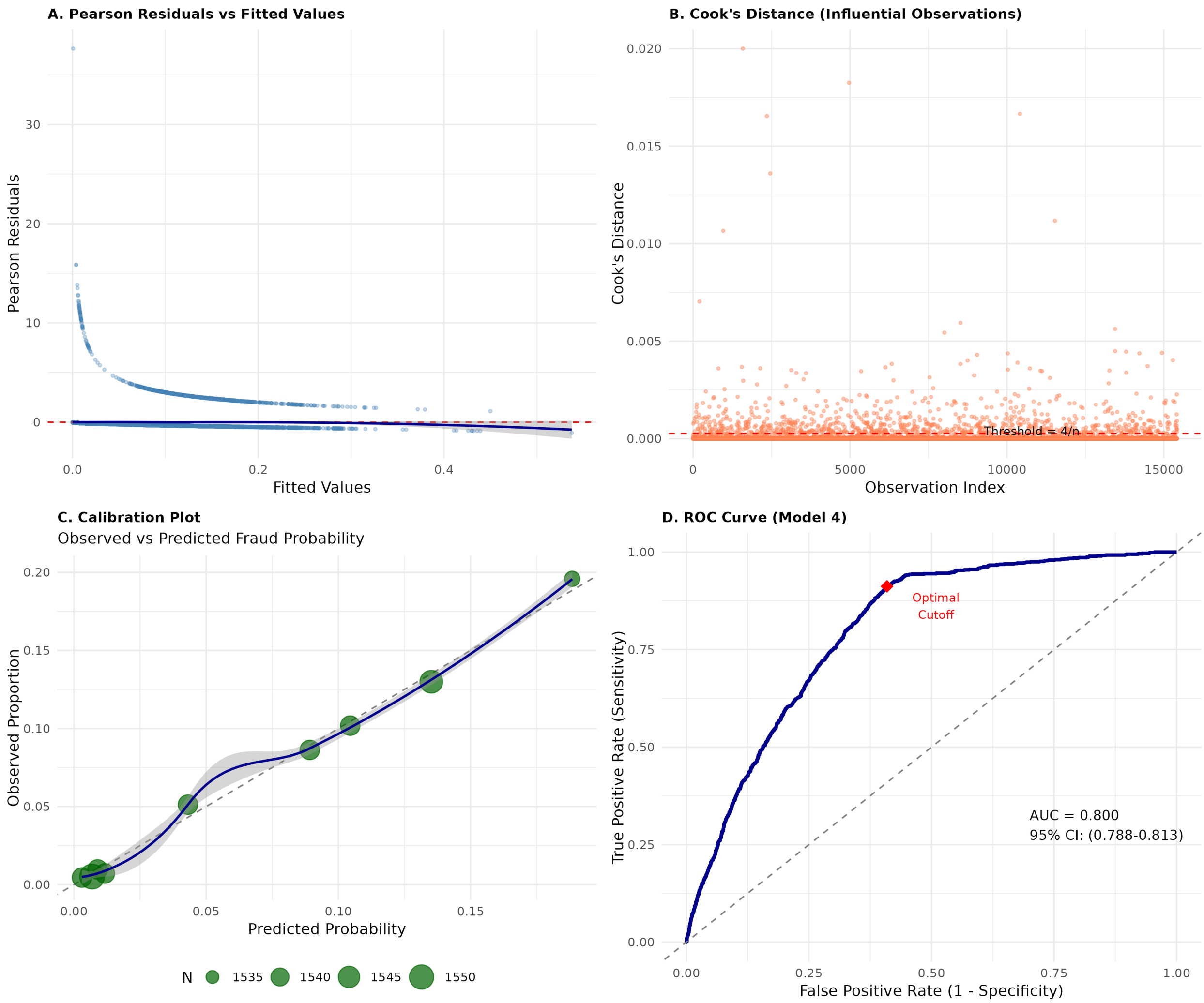

Figure 2. Model 4 Diagnostic Plots. Panel A shows Pearson residuals against fitted values, with no discernible pattern indicating good fit. Panel B plots Cook's distance for each observation, with influential points generally corresponding to fraud cases. Panel C is the calibration plot, where the close alignment of points to the diagonal line indicates excellent calibration. Panel D shows the ROC curve for the final model, with an AUC of 0.800.

4.5. Risk Stratification and Practical Performance

The practical value of Model 4 is illustrated in the risk stratification analysis presented in Table 8 and Figure 3. When claims are sorted according to their predicted fraud probability and are segregated into quintiles, the model does a good job of sitting risks apart. The fraud rate among the 'Very High' risks (upper 20% of claims) was 16.28%, which was 2.72 times the base fraud rate of 5.99%. On the other hand, the 'Very Low' risks only carried a fraud rate of 0.48%.

This stratification converts into substantial operational efficiency. By devoting investigation resources to the top two quintiles ('High' and 'Very High' risk), which account for only 40% of all claims, an insurer can recover an estimated 85.8% of all fraudulent instances. This enables the streamlined processing of the 60% of claims in the less risky stratifications while focusing expert scrutiny where the scrutiny is most valuable, and thus enhancing the whole claims handling process's efficiency. This concentration of fraud cases into a manageable high-risk segment demonstrates the model's practical utility for real-world implementation, independent of the statistical significance of individual temporal predictors.

Table 8. Risk Stratification and Fraud Detection Performance

Risk Group	**N**	**Fraud N**	**Non-Fraud N**	**Observed Rate (%)**	**Mean Pred Prob (%)**
Very Low	3093	15	3078	0.48	0.49
Low	3075	26	3049	0.85	0.84
Medium	3084	90	2994	2.92	2.74
High	3084	290	2794	9.40	9.68
Very High	3084	502	2582	16.28	16.17
Total	15420	923	14497	5.99	5.99
Key Performance Indicators					
High/Very High risk groups capture 85.8% of all fraud cases					
High/Very High risk groups represent 40.0% of all cases					
Lift in 'Very High' risk group: 2.72× baseline fraud rate					

Note: Risk groups based on quintiles of predicted fraud probability. Lift = ratio of fraud rate in group to overall fraud rate.

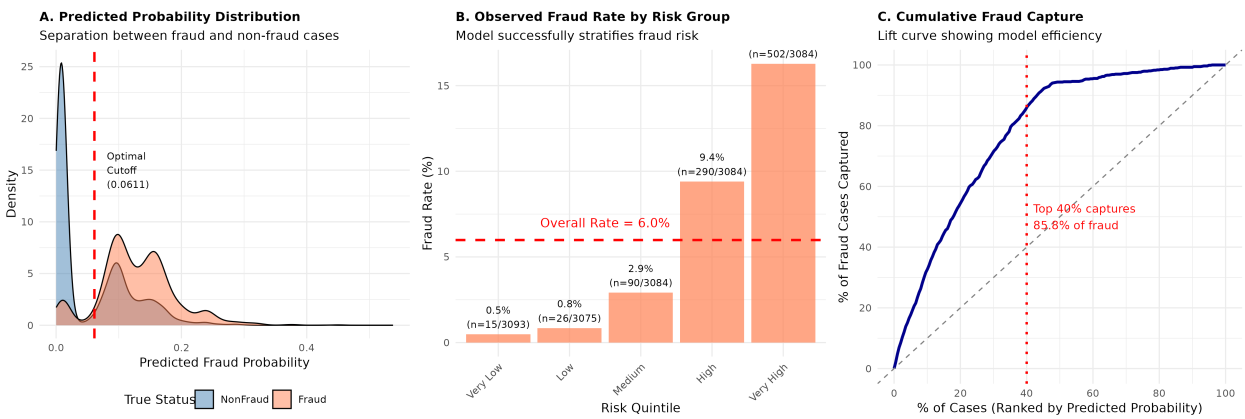

Figure 3. Risk Stratification Performance. Panel A shows the density distribution of predicted probabilities for fraudulent and non-fraudulent cases, illustrating the model's ability to separate the two groups. Panel B displays the observed fraud rate for each risk quintile, demonstrating successful risk stratification. Panel C presents the cumulative fraud capture curve, highlighting that 85.8% of fraud cases are captured by targeting the top 40% of highest-risk claims.

5. Discussion

The research in this paper built and verified an automobile insurance fraud predictive model with the main objective of assessing the usefulness of temporal pattern motives as early warning signs. The result improves our theory and practice in the following ways: verifying the predictive values of the existing risks while revealing fresh light with respect to the importance of the timing in claims. Here, we present the main findings and their theoretical and practice values, the weaknesses in the research, and future research directions.

The findings produce four main results. Firstly, and most notably, the analysis verifies that the characteristics of policies and the circumstances of accidents are the overriding predictors of motor insurance fraud. The enormouse odds that are linked with AllPerils ($\mathrm{OR} = 55.2$) and Collision ($\mathrm{OR} = 31.0$) policies reflect the immense influence of moral hazard; increased coverage seems to provide a vastly greater potential for fraudulent activity. In the same way, the fact that claims at fault (Fault_PH, $\mathrm{OR} = 14.2$) are vastly more likely to prove fraudulent indicates that fraudulent activity may often involve the misrepresentation of legitimate events more than fabrication in the raw. The buffering effect from the presence of a police report (PoliceRep, $\mathrm{OR} = 0.55$) further supports the view that independent corroboration acts as a formidable inhibitor.

Second, the analysis yields directional, though not highly statistically significant, evidence for the hypothesis that temporal regularity is a marker for fraud. Policies with claims for an accident that occurred in the first 15 days after the initiation of the policy had 87% increased odds of being fraudulent (p=0.075). Though the result was not significant at the traditional p ¡ 0.05 level, it very strongly suggests the existence of a relationship that should be explored in more detail. The result's lack of significance must largely be due to the small sample size in the dataset for early-event claims, which reduces statistical power and prevents the effect from being proven in the limit. We acknowledge that multiple interpretations remain plausible: temporal patterns may represent weak but genuine signals that provide marginal predictive value, they may be context-dependent and vary across insurance markets and fraud typologies, or the observed effects may reflect insufficient statistical power rather than true associations. The directional consistency with theoretical predictions of premeditation and intent suggests the relationship merits further investigation, though definitive conclusions require larger samples specifically enriched for early policy events.

Third, the best logistic regression model (Model 4) shows good and stable predictive ability with an AUC value of 0.800. On the basis of the collected amount of precision, the value shows competitive performance with that obtained in similar research with the aid of classical statistical and advanced machine learning models [5, 12], and the efficiency for the purpose of a well-specified logistic regression model for the task becomes confirmed. Thorough diagnostics verified the well-calibration, absence from multicollinearity, and stability of the model from bootstrap validation.

Fourth, pragmatically speaking, the model works extremely well at risk stratification. By putting 85.8% of all fraudulent claims in the top 40% of the cases sorted by risk, the model allows for a simple and effective method for resource allocation. In doing so, this discovery points up the utility of the model from an operational point of view, allowing the insurers to switch from a heuristic- or universal-based review process to a targeted, data-driven approach.

The results from this research have some key theoretical consequences. Overriding predictive strength from variables such as policy type and fault assignment gives strong empirical backing for the ”Opportunity” and the ”Rationalization” aspects of the traditional Fraud Triangle theory [36, 37]. Inflation-oriented policies put the opportunity for exaggerated claims in place, while an at-fault collision potentially places a prior setting in which a claimant can more readily rationalize stretching the damages.

In addition, the directional temporal pattern evidence works in the direction of behavioral economic theory. The increased risk in early-life claims can be seen in light of behavioral indications of premeditation or intent [38]. However, alternative explanations warrant consideration: early claims may reflect opportunistic coverage of pre-existing damage, adverse selection by high-risk individuals anticipating claims, or reporting confusion among new policyholders unfamiliar with coverage terms, though the latter would predict lower rather than higher fraud rates. Fraud planners are likely to buy a policy with the express purpose of making a claim some short time later, a process which can be hard to apprehend with the use of static variables only. Thus, the current work represents a first step in the incorporation of behavioral timing cues in broader theoretical models of insurance fraud, with the possibility that the different typographies of fraud (e.g., opportunistic vs. premeditated) can exhibit different temporal footprints.

The proven model has considerable practical value for the insurance market. Its first application includes the use as an automated decision-aids tool, which can be embedded at the First Note of Loss (FNOL). During the preliminary moment, the model can produce in real-time a real-time risk score for each incoming claim, which allows for a triage-based workflow:

1. **Claims Processing Workflow Optimization:** Those 60% of claims in the low and medium-risk quintiles which combined contain only 14.2% of the fraud cases can potentially be put in an expedited payment track.

Not only would this save the administration money, but it also would really enhance the customer experience for the large majority of truthful policyholders.

2. **Intensified Vigilance and Utilization of Resources:** The high and very-high-risk high 40% of claims can simultaneously be highlighted and sent straight to specialist investigating teams. In this way, careful attention and specialist resources are concentrated solely on the claims most likely to contain fraud, exponentially raising the speed and efficiency with which the investigation process can identify and recover fraudulent claims.
3. **Underwriting Adjustments:** The model findings can also be used in the case of underwriting and pricing policies. Due to the very high risks in 'AllPerils' and 'Collision' policies, more detailed risk-based pricing can be required for these policies or specific discouraging measures can be adopted, for example, the forceful conductance of compulsory inspections for policies with early claims.

In spite of its strong results, the study has some limitations that provide future research directions. In the first place, the analysis is based on one-source dataset, which can constrain the generalizability of the findings to other insurers, geographic locations, and other regulatory regimes. External validity may be limited by market-specific fraud patterns such as staged accident rings prevalent in certain regions, regulatory differences affecting claim timing requirements across jurisdictions, and insurer-specific detection capabilities influencing which fraud is identified. The analysis should be replicated with the bigger multi-insurer dataset in the future. Future validation should prioritize multi-insurer datasets to assess model stability across organizations, cross-jurisdictional data to test regulatory context effects, and temporal validation on data from different time periods.

Second, and more importantly for the main hypothesis, the prevalence of claims with early accident or reporting behavior was exceedingly low. This low prevalence restricted the statistical strength to clearly establish the temporal variables' significance and consistently test interaction effects. A bigger sample with a greater prevalence among the early events is required more assertively to establish their predictive value and investigate how their influence could change in alternative demographic or policyholder alignments. Post-hoc power analysis suggests that achieving 80% statistical power at $\alpha = 0.05$ for the observed effect size would require approximately 85-90 early-event fraud cases, translating to a total portfolio size of 60,000-70,000 claims.

Third, logistic regression was used in this work due to its explanatory nature. Though it worked well, its more advanced, non-linear machine learning counterparts, like gradient boosting machines or deep networks, may discern more subtle data patterns and produce a more advanced degree of predictive precision [39]. Future work should contrast the capabilities of these state-of-the-art models with the logistic regression baseline introduced here. Benchmarking against advanced machine learning models would test whether temporal variables gain importance when captured through flexible, non-linear approaches.

Lastly, the research is cross-sectional. A longitudinal examination that follows the behavior of policyholders over time may yield more comprehensive findings about the changing nature of fraud risk and possibly dynamic temporal leading up to a fraudulent claim. Prospective studies employing survival analysis methods such as Cox proportional hazards models could more authentically capture temporal dynamics by modeling the hazard of fraudulent claims over the policy lifecycle.

6. Conclusion

In this work, we aimed to identify whether temporal claims patterns can be used to augment the detection of motor-direct-weber fraud at its earliest onset. This exploratory study establishes a methodological foundation for incorporating temporal analytics into fraud detection frameworks, providing preliminary evidence that motivates larger-scale validation. The model resulted in a strong logistic regression model with excellent predictive capability (AUC = 0.800), largely due to well-understood risks along the lines of fault attribution and type of policy. The new temporal variables were not quite statistically significant in the model, but indicated a distinct positive correlation with fraud, and thus are useful predictors whose potential value should be explored with broader ranges of data. The main methodological contribution of this research is the systematic incorporation and validation of these temporal variables in the context of a hierarchical modeling framework, deriving from them a new, behaviorally-informed dimension in the detection of fraud.

From a practical point of view, the most significant implication is the demonstrated capability of the model to stratify risk importantly. Its ability to capture 86% of the fraudulent claims in a specific 40% of the claims population provides insurers with a valuable resource for optimizing investigative resources, lowering costs, and enhancing efficiency. It allows for a more strategic claims handling, calibrating tough fraud control with a superior experience for the great majority of legitimate customers. Despite some restrictions applicable only to the source of the data and statistical power, this study successfully emerges the latent potentiality of temporal data. Following studies must aim at the application of more sophisticated machine learning techniques in broader, longitudinal databases in order to capitalize the directional promising findings we report here. In the longer term, this research reinforces the value of data-driven techniques in the fight against insurance fraud and points the way to the exploration of the timing behavior in fraudulent acts.

REFERENCES

[1] Yufeng Kou, Chang-Tien Lu, Suchada Sirwongwattana, and Yo-Ping Huang. Survey of fraud detection techniques. In *Proceedings of the 2004 IEEE International Conference on Networking, Sensing and Control*, volume 2, pages 749–754. IEEE, 2004.

[2] Miquel Artís, Mercedes Ayuso, and Montserrat Guillén. An econometric model of insurance fraud detection. *The Geneva Papers on Risk and Insurance - Issues and Practice*, 27(1):94–108, 2002.

[3] Richard A Derrig. Insurance fraud. *Journal of Risk and Insurance*, 69(3):271–287, 2002.

[4] Frank Cowell. Fraud and corruption in insurance. In *The Handbook of Insurance*, pages 903–938. Springer, 2013.

[5] Siddhartha Bhattacharyya, Sanjeet Jha, Kurian Tharakunnel, and J Christopher Westland. A data mining approach to automobile insurance fraud detection. *Decision Support Systems*, 50(3):602–610, 2011.

[6] Eric W. T. Ngai, Yong Hu, Y. H. Wong, Yasmin Chen, and Xin Sun. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of the literature. *Decision Support Systems*, 50(3):559–569, 2011.

[7] Richard J Bolton and David J Hand. Statistical fraud detection: A review. *Statistical Science*, 17(3):235–255, 2002.

[8] Quirin Aemig and Martin Spindler. The impact of information asymmetry on insurance fraud: Evidence from the german car insurance market. *Journal of Risk and Insurance*, 88(1):155–186, 2021.

[9] Yourui Guo. Application of machine learning in insurance fraud detection: Achievements and future prospects. *Proceedings of the 2024 International Conference on Artificial Intelligence and Computing (ICAIC)*, pages 615–620, 2024.

[10] G. G. Sundarkumar and Vadlamani Ravi. A survey on computational intelligence techniques for fraud detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 5(5):225–246, 2015.

[11] J West and M Bhattacharya. Intelligent financial fraud detection: A comprehensive review. *Computers & Security*, 57:47–66, 2016.

[12] Juan Pesantez-Narvaez, Montserrat Guillen, and M Alcañiz. A machine learning-based approach for automobile insurance fraud detection. *Decision Support Systems*, 132:113275, 2020.

[13] Felipe Paula, Albert Ali Salah, and Wil MP Van Der Aalst. A survey on machine learning for insurance fraud detection. *Archives of Computational Methods in Engineering*, 28(6):4331–4357, 2021.

[14] J. A. Major and D. R. Riedinger. Temporal aspects of fraud detection. *Journal of Financial Crime*, 15(1): 7–17, 2008.

[15] Robert G Smith. Organized insurance fraud investigation and prosecution. *Journal of Insurance Regulation*, 33(1), 2014.

[16] Rashmeet Fakoor, Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. Temporal patterns in financial statement fraud detection. *Expert Systems with Applications*, 140:112879, 2020.

[17] Georges Dionne. The empirical economics of insurance: A review. *The Geneva Papers on Risk and Insurance - Issues and Practice*, 38(1):1–38, 2013.

[18] A. N. Subudhi and S. Panigrahi. A systematic review on insurance fraud detection. *Journal of Financial Crime*, 29(4):1184–1212, 2022.

[19] Rihab Jedda, Mouna Khemakhem, and Saoussen Krichen. A machine learning approach for health insurance fraud detection. *International Journal of Healthcare Management*, 13(sup1):217–225, 2020.

[20] A. Perez-Suarez, J. Perez-Martin, and M. M. Fondevila-Gascón. A review of machine learning techniques for fraud detection in the insurance sector. *Mathematics*, 11(13):2999, 2023.

[21] Jian Gao, Yang Yang, and Yanjun Li. A hybrid model based on xgboost and focal loss for insurance fraud detection. *Applied Soft Computing*, 133:109923, 2023.

[22] J. Li, J. Z. Huang, and Z. Zhu. An effective automobile insurance fraud detection model based on deep learning. *Journal of Ambient Intelligence and Humanized Computing*, 12:9467–9477, 2021.

[23] Y. Wang and W. Xu. Deep learning for insurance fraud detection: A literature review. *Artificial Intelligence Review*, 55(2):1255–1283, 2022.

[24] Matteo Carminati, Jason Caruana, Letterio Galletta, and Giuseppe Scebba. A network-based approach to detect insurance fraud in the property and casualty market. *Expert Systems with Applications*, 199:116922, 2022.

[25] F. Sahin and E. Ekinci. A survey on network-based fraud detection. *IEEE Communications Surveys & Tutorials*, 22(4):2501–2539, 2020.

[26] Shuai Yuan, Yangyang Zheng, Jiaming Liu, Hongda Li, Zhao Liu, Peng Ding, Peng Cui, and Wen-jie Zhang. Timetrail: Unveiling financial fraud patterns through temporal correlation analysis. *arXiv preprint arXiv:2308.14215*, 2023.

[27] M. Ahmed and A. N. Mahmood. A survey of network-based anomaly detection techniques. *Journal of Network and Computer Applications*, 60:25–39, 2016.

[28] Stijn Viaene and Guido Dedene. Insurance claim fraud detection. In *Handbook of Computational Intelligence in Stocks and Finance*, pages 323–341. Springer, 2007.

[29] A. Abdallah, M. A. Maarof, and A. Zainal. Fraud detection system: A survey. *Journal of Network and Computer Applications*, 68:90–113, 2016.

[30] Tom Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.

[31] David W. Hosmer Jr, Stanley Lemeshow, and Rodney X. Sturdivant. *Applied logistic regression*, volume 398. John Wiley & Sons, 2013.

[32] Bradley Efron and Robert J. Tibshirani. *An introduction to the bootstrap*. CRC Press, 1994.

[33] David G. Kleinbaum and Mitchel Klein. *Logistic regression: A self-learning text*. Springer Science & Business Media, 2002.

[34] Joseph F. Hair, William C. Black, Barry J. Babin, and Rolph E. Anderson. *Multivariate data analysis: A global perspective*. Pearson Education, 2010.

[35] R. Dennis Cook. Detection of influential observation in linear regression. *Technometrics*, 19(1):15–18, 1977.

[36] Donald R Cressey. *Other people's money: A study in the social psychology of embezzlement*. The Free Press, 1953.

[37] W Steve Albrecht, Conan C Albrecht, Chad O Albrecht, and Mark F Zimbelman. *Fraud examination*. Cengage Learning, 2016.

[38] Dan Ariely. *Predictably Irrational: The Hidden Forces That Shape Our Decisions*. HarperCollins, New York, 2008.

[39] Antonio Gulli and Sujit Pal. *Deep Learning with Keras*. Packt Publishing Ltd, Birmingham, UK, 2017.