JOURNAL OF MANAGEMENT AND OPERATIONS RESEARCH
Research Article | Volume 5 | Issue 5 | May 2025
Received: 20 Apr 2025 | Revised: 27 Apr 2025
Accepted: 08 May 2025 | Published Online: 17 May 2025

Modelling of Vehicle Insurance Claim with Stochastic Process using Hidden Markov Model

Ali Rezaei[1], Fatemeh Mohammadi[2], Anouk Jansen[3] and Amir Hossein Karami[4,*]

[1] Department of Applied Mathematics, University of Semnan, Semnan, Iran

[2] Computational Statistics Research Group, Golestan University, Gorgan, Iran

[3] Institute for Commercial Strategy and Analysis, Saxion University of Applied Sciences, Enschede, The Netherlands

[4] Institute for Advanced Study in Science and Technology, Ardabil, Iran

[*]Corresponding Author, Email: karami.amir@iasst.ac.ir

Abstract: The modeling of vehicle insurance claims plays a crucial role in the insurance industry to assess risk and determine appropriate premiums. However, existing research often relies on deterministic methods, neglecting the stochastic nature of claim occurrences and causing inaccuracies in risk estimation. This paper addresses the limitations in current research by proposing a novel approach using Hidden Markov Model to incorporate stochastic processes into the modeling of vehicle insurance claims. By capturing the dynamic and unpredictable nature of claim occurrences, the proposed model offers a more accurate representation of risk factors and enhances the predictive capabilities in insurance claim analysis. Through the integration of Hidden Markov Model, this study provides a innovative framework for modeling vehicle insurance claims that has the potential to significantly improve the efficiency and accuracy of risk assessment in the insurance industry.

Keywords: *Vehicle Insurance Claims; Risk Assessment; Stochastic Processes; Hidden Markov Model; Predictive Capabilities*

1. Introduction

Vehicle insurance claim is a specialized field within the insurance industry that focuses on the process of filing and settling claims related to vehicle damages or accidents. This area faces several challenges and bottlenecks, including the complexity of assessing damages accurately, fraudulent claims, lengthy claim processing times, and inconsistencies in claim decisions. The integration of technology such as artificial intelligence and machine learning has helped in streamlining some aspects of the claims process, but there is still a need for improvement in data accuracy, transparency, and customer experience. Additionally, regulatory compliance and legal complexities further add to the challenges faced by professionals in this field.

To this end, current research on Vehicle Insurance Claims has advanced to encompass a variety of factors affecting claim processing efficiency, fraud detection methods, and customer satisfaction levels. Studies now focus on utilizing technology such as AI and data analytics to improve claim assessment accuracy and streamline the overall claims process. The literature review discusses various aspects related to vehicle insurance claim processes and fraud detection. Merupula et al. (2023) introduce a compound Conway-Maxwell-Poisson regression model for vehicle insurance claim data, providing a methodology for parameter estimation and prediction interval calculation. Fauzi and Er (2025) analyze the efficiency of vehicle insurance claim processes using process mining, identifying bottlenecks and proposing recommendations for improvement. Wilson and Ma (2025) propose an artificial neural network with domain adaptation to enhance fraud detection in insurance claims, demonstrating improved accuracy and generalizability. Lie (2022) explores social capital's role in insurance organizations, emphasizing its impact on productivity and synergy with other capitals. Suwardi and Purwono (2021) present a robust chain ladder method for analyzing vehicle insurance claim reserves, highlighting its resistance to outlier data and improved performance. Azis et al. (2018) describe a modeling approach using exponential hidden Markov models for vehicle insurance claims. Rithik and Chokkalingam (2024) focus on image-based classification for vehicle accident detection and insurance claims using AI, offering insights into automated claim processing. Shvarts (2016) discusses a forecasting model using artificial neural networks for vehicle insurance claim data study. Jun-x (2014) designs a telephone callback system for vehicle insurance claim settlements, enhancing customer service and claim processing efficiency. Azis et al. (2018) introduce the use of exponential hidden Markov models for vehicle insurance claims. This technique offers a powerful and efficient method for modeling complex claim processes with hidden states, allowing for improved accuracy in predicting and detecting fraudulent claims.

Specifically, Hidden Markov Models (HMMs) can be effectively utilized to analyze vehicle insurance claims by modeling the underlying processes that govern claim frequency and severity. By capturing the temporal dynamics and patterns of driver behavior and accident occurrences, HMMs enhance risk assessment and fraud detection strategies in the insurance industry. This literature review provides a comprehensive overview of the applications of hidden Markov models in various domains. Krogh et al. (2001) introduced a membrane protein topology prediction method, TMHMM, which showed high accuracy in predicting transmembrane helices. Zhang et al. (2001) proposed a hidden Markov random field model for brain MRI image segmentation, achieving robust segmentation results. Sonnhammer et al. (1998) developed a hidden Markov model for predicting transmembrane helices in protein sequences, contributing to the field of bioinformatics. Wang et al. (2007) presented PennCNV, an integrated hidden Markov model for copy number variation detection in SNP genotyping data, enabling fine-mapping of CNVs. Narasimhan et al. (2016) introduced BCFtools/RoH, a hidden Markov model approach for detecting autozygosity in sequencing data, showing improved sensitivity and specificity. Cheng et al. (2020) and Ren et al. (2020) proposed methods for state estimation in neural networks and control approaches for hidden Markov systems, respectively, showing the versatility of hidden Markov models in control theory. Xia and Chen (2020) suggested a discrete hidden Markov model for SMS spam detection, proving its effectiveness in language-independent spam identification. Lastly, Yang and Gidófalvi (2018) devised a fast map matching algorithm integrating hidden Markov models with precomputation, enhancing efficiency in geographical information systems. However, limitations persist in the generalizability of these models across diverse datasets, their computational complexity, and the challenges in accurately defining state transitions, which may hinder broader applicability.

The paper by A. Wilson and J. Ma presents an innovative use of the MDD-based Domain Adaptation Algorithm to enhance the utility of Artificial Neural Networks in detecting vehicle insurance claim fraud. This work significantly inspires the exploration of adapting machine learning techniques to model various aspects of vehicle insurance claims beyond just fraud detection. Their research introduces a methodology that mitigates the distribution discrepancy between training data and target data through domain adaptation, thereby enhancing the predictability of machine learning models when applied to different domains [3]. My research draws on these insights by employing the fundamental principles of hidden states and observable sequences, inherent to the Hidden Markov Model, in capturing the stochastic nature of vehicle insurance claim processes. While Wilson and Ma adeptly applied domain adaptation to align feature distributions across diverse datasets, my study translates these alignment concepts into dynamically evolving state transitions which can represent the progression and regression typical in insurance claims [3]. By considering the temporal sequences and transitions observed in the claim data, the model accommodates the randomness in insurance events while being informed by underlying probability distributions adjusted for domain discrepancies. This approach ensures that the intricacies of claim developments are precisely modeled, potentially improving prediction accuracy and decision-making processes. Furthermore, the emphasis in Wilson and Ma’s paper on refining feature extraction through domain adaptation inspired an analogous refinement in identifying observable parameters of the Markov model, ensuring that they accurately capture the nuances in the dataset distributions. The meticulous adjustments fostered through domain awareness in Wilson and Ma's methodology underscore a critical element adopted in our research—engaging strategies that iteratively refine the parameter spaces to achieve robust training outcomes across divergent datasets. Thus, while their focus decisively approached fraud detection efficacy, my work broadens this foundation to interpret complex, underlying stochastic processes within insurance claim datasets, taking advantage of the robust domain-aware enhancements characteristic of their MDD framework [3].

Section 2 of this study articulates the problem statement, highlighting how traditional deterministic methods fall short in capturing the stochastic nature of vehicle insurance claims, leading to inaccuracies in risk estimation. To address these limitations, Section 3 introduces a novel methodology that leverages a Hidden Markov Model (HMM) to incorporate stochastic processes, allowing for a more nuanced and accurate assessment of risk factors. Section 4 presents a detailed case study to demonstrate the practical application and effectiveness of the proposed model in a real-world insurance setting. The results of this case study are analyzed in Section 5, where the enhanced predictive capabilities of the model are discussed. Section 6 delves into a broader discussion about the implications of these findings, emphasizing the potential improvements in efficiency and accuracy in risk assessment that this innovative approach offers. Finally, Section 7 concludes the paper by summarizing the significant advancements this model contributes to the field of vehicle insurance claim analysis, underscoring its potential to revolutionize risk assessment strategies in the insurance industry.

2. Background

2.1 Vehicle Insurance Claim

Vehicle insurance claim is a formal request made by a policyholder to an insurance company for compensation of losses incurred due to a vehicular incident, such as an accident, theft, or natural disaster damage. The process involves several steps, beginning with incident reporting and culminating in the potential payout by the insurer, which is contingent upon the terms of the policy and the specifics of the incident. The crux of a vehicle insurance claim process hinges on accurately assessing the loss, which can be mathematically represented through various parameters. The total claim amount (C_t) depends on several factors including the type, extent, and nature of the damage. A common starting point for evaluating a claim is the formula for calculating the estimated loss value (L_v), often expressed as:

$$L_v = D + I + M \tag{1}$$

Where:
- D is the damage cost to the vehicle.
- I is the personal injury cost.
- M represents additional expenses involved, such as medical expenditure and rental charges.

Another critical aspect is determining the depreciation of the vehicle over time, which affects the D value. Depreciation can be expressed as:

$$D = V_i \times (1 - d)^t \tag{2}$$

Where:
- V_i is the initial value of the vehicle.
- d is the annual depreciation rate.
- t is the number of years the vehicle has been in use.

The insurance company will then evaluate the loss against the deductible (D_e), the portion of the claim the policyholder must pay out-of-pocket. The net claim amount (C_n) is calculated as:

$$C_n = \max(L_v - D_e, 0) \tag{3}$$

The insurer uses actuarial science to determine premium (P) calculations and payout probabilities. By analyzing historical data, they establish a fair premium that balances risk and return. The expected loss (E_l) can be calculated by:

$$E_l = \sum_{i=1}^{n} p_i \times L_{v_i} \tag{4}$$

Where:
- p_i is the probability of the i -th event occurring.
- L_{v_i} is the loss value if the i -th event occurs.
- n represents the number of potential loss scenarios.

The probability of claim occurrence (P_c) is another vital factor considered during risk assessment:

$$P_c = \frac{\text{Number of Claims}}{\text{Total Number of Insured Vehicles}} \tag{5}$$

Risk exposure, a determining factor in premium setting, is based on assessed risk factors associated with the driver and vehicle. Factors include age, driving history, vehicle type, and usage. The risk factor (R_f) can be succinctly captured as:

$$R_f = A_u \times S_d \times T_v \tag{6}$$

Where:
- A_u is the age-based risk.
- S_d is the safety factor derived from driving history.
- T_v is the risk factor associated with vehicle type.

In summary, a vehicle insurance claim represents a complex interplay of variables including estimated loss value, depreciation, net expected loss, deductible, occurrence probability, and risk exposure. These elements are critical in the evaluation and determination of an insurance payout, with actuarial and statistical methods employed to ensure fair compensation and sustained profitability for insurance companies.

2.2 Methodologies & Limitations

In the domain of vehicle insurance claims, the methodologies employed for claim assessment and compensation calculation are both comprehensive and intricate. These methodologies integrate a variety of statistical and actuarial approaches to ensure a balanced consideration of risks and returns for both insurers and policyholders. A primary approach for evaluating a claim involves calculating the Estimated Loss Value (L_v), crucial for determining the financial impact of reported incidents. The formula for L_v is typically expressed as:

$$L_v = D + I + M \tag{7}$$

Here, D defines the damage cost to the vehicle, I covers personal injury costs, and M encompasses additional related expenses. A vital component in this process is assessing Vehicle Depreciation, which directly affects the damage cost D. Depreciation over time can be modeled through the formula:

$$D = V_i \times (1 - d)^t \tag{8}$$

Where V_i represents the initial vehicle value, d is the annual depreciation rate, and t denotes the number of years in use. Post-calculation of L_v, the Net Claim Amount (C_n) is derived by adjusting for the policyholder's deductible (D_e):

$$C_n = \max(L_v - D_e, 0) \tag{9}$$

This formula ensures that payouts only occur when the loss exceeds the deductible, maintaining a cost-sharing principle between insurers and insured. To maintain equilibrium in premium calculation and ensure profitability, insurers leverage actuarial science. A foundational element here is calculating the Expected Loss (E_l), reflecting potential financial exposure:

$$E_l = \sum_{i=1}^{n} p_i \times L_{v_i} \tag{10}$$

In this equation, p_i is the probability of the i -th event, L_{v_i} signifies the loss value for that event, and n delineates the potential scenarios. The Probability of Claim Occurrence (P_c) evaluates how often claims arise relative to insured vehicles, impacting premium strategies:

$$P_c = \frac{\text{Number of Claims}}{\text{Total Number of Insured Vehicles}} \tag{11}$$

Additionally, risk-based premium adjustments consider Risk Exposure, an aggregate of multiple risk factors associated with the driver and vehicle:

$$R_f = A_u \times S_d \times T_v \tag{12}$$

Where A_u represents age-related risks, S_d accounts for safety based on driving history, and T_v pertains to inherent risks related to vehicle type. Despite the robustness of these methods, several challenges impact their effectiveness. Depreciation models may not capture market fluctuations, leading to over- or under-estimation of vehicle values. Furthermore, simplistic models can inadequately reflect the complexity of real-life risk dynamics, such as rapidly changing behavioral patterns and emerging risks from new technologies. The accuracy of probability estimates like P_c heavily depends on historical data which might not forecast future trends accurately. In conclusion, while the current methodologies in vehicle insurance claims employ sophisticated mathematical models to address risk and compensation, limitations persist due to the dynamic nature of risks and evolving market conditions. Continuous refinement and incorporation of real-time data and advanced predictive analytics are essential to enhance the precision and fairness of these approaches.

3. The proposed method

3.1 Hidden Markov Model

The Hidden Markov Model (HMM) represents a stochastic model that is indispensable in understanding temporal sequences, where the system being modeled is assumed to follow a Markov process with hidden states. It operates on the principle of modeling complex stochastic processes dexterously, where the observable outcomes are linked to underlying states that are not directly observable, hence the term 'hidden.' To comprehend the mechanics of HMM, consider an underlying sequence of unobserved (hidden) states $Q = \{q_1, q_2, \dots, q_N\}$, which evolve over time. The transition between these states is characterized by a set of probabilities forming the transition matrix $A = \{a_{ij}\}$, where each element a_{ij} represents the probability of transitioning from state i to state j , defined mathematically as:

$$a_{ij} = P\left(s_{t+1} = q_j \middle| s_t = q_i\right) \tag{13}$$

The observable sequence $O = \{o_1, o_2, \dots, o_T\}$, on the other hand, is generated on the basis of these states through a probabilistic function. The probability of an observation given its state is denoted by the emission matrix $B = \{b_j(o_t)\}$, where $b_j(o_t)$ signifies the probability of observing o_t from state j :

$$b_j(o_t) = P\left(O_t = o_t \middle| s_t = q_j\right) \tag{14}$$

Initializing the system, we require a probability distribution over the initial states, termed the initial state distribution $\pi = \{\pi_i\}$, with:

$$\pi_i = P(s_1 = q_i) \tag{15}$$

Thus, a complete specification of an HMM necessitates three sets of probabilities: the transition probabilities, the emission probabilities, and the initial state probabilities, cumulatively described by the parameter set $\lambda = (A, B, \pi)$. The evaluation problem in HMM concerns computing the likelihood of an observed sequence given a model, expressed as $P(O|\lambda)$. This is effectively solved by the Forward Algorithm, which computes the forward probabilities $\alpha_t(i)$ for each state q_i at time t :

$$\alpha_t(i) = P(O_1, O_2, \dots, O_t, s_t = q_i|\lambda) \tag{16}$$

These probabilities are recursively defined:

$$\alpha_1(i) = \pi_i b_i(o_1) \tag{17}$$

$$\alpha_{t+1}(j) = \left(\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right) b_j(o_{t+1}) \tag{18}$$

Decoding in HMM seeks the optimal sequence of hidden states given the observed sequence. The Viterbi Algorithm provides an optimal solution by maximizing the probability over state sequences:

$$\delta_t(i) = \max_{s_1,s_2,\dots,s_{t-1}} P(s_1, s_2, \dots, s_t = q_i, o_1, o_2, \dots, o_t|\lambda) \tag{19}$$

The recursive relation for $\delta_t(i)$ is:

$$\delta_1(i) = \pi_i b_i(o_1) \tag{20}$$

$$\delta_{t+1}(j) = \max_i\left[\delta_t(i)a_{ij}\right]b_j(o_{t+1}) \tag{21}$$

Estimating model parameters presents a methodological challenge, which is addressed by the Baum-Welch algorithm, a form of the Expectation-Maximization (EM) algorithm tailored for HMMs. It iteratively refines estimates of A, B, and π by maximizing the expected log-likelihood of observed data under the model, leveraging forward-backward procedures to compute expected counts of state transitions and symbol emissions. Despite its robustness, there are challenges associated with HMMs, such as the assumption of state independence and the selection of appropriate model structure and parameters. Yet, by offering a systematic way to model dynamic systems with hidden states, HMMs have become integral in fields like speech recognition, bioinformatics, and financial modeling, where complex sequential data is prevalent.

3.2 The Proposed Framework

The methodology introduced is predominantly inspired by the work of Wilson and Ma [3], laying a foundation for leveraging domain adaptation algorithms in enhancing the functionality of Artificial Neural Networks in fraud detection within vehicle insurance claims. Here, the integration of the Hidden Markov Model (HMM) with vehicle insurance claim processing presents a formidable approach. In the realm of vehicle insurance claims, policyholders seek compensation for losses deriving from vehicular incidents. The essential challenge in this process is the accurate assessment of loss. This can be mathematically encapsulated through various parameters, where the claim amount (C_t) draws from key elements such as the damage type (D), injury costs (I), and miscellaneous expenses (M), represented collectively by the equation:

$$L_v = D + I + M \tag{22}$$

Depreciation significantly influences the damage cost component (D), being represented as:

$$D = V_i \times (1 - d)^t \tag{23}$$

Additionally, insurers evaluate the loss against deductibles (D_e) to arrive at the net claim amount (C_n):

$$C_n = \max(L_v - D_e, 0) \tag{24}$$

While actuarial science underscores the premium calculations, considering the expected loss (E_l) through:

$$E_l = \sum_{i=1}^{n} p_i \times L_{v_i} \tag{25}$$

Where p_i signifies event occurrence probabilities, its integration with P_c, which is the probability of claim occurrence, enhances precise risk profiling:

$$P_c = \frac{\text{Number of Claims}}{\text{Total Number of Insured Vehicles}} \tag{26}$$

Risk exposure rooted in driver and vehicle characteristics, quantified as:

$$R_f = A_u \times S_d \times T_v \tag{27}$$

invigorates the premium assessment domain. Parallel to these, the Hidden Markov Model (HMM) provides a framework to model sequences from vehicle insurance claims, offering a statistical lens. It encapsulates a stochastic system where claim sequences, observable as $O = \{o_1, o_2, \ldots, o_T\}$, are governed by hidden states $Q = \{q_1, q_2, \ldots, q_N\}$, evolving temporally. Here, the transition matrix $A = \{a_{ij}\}$, defining transitions between hidden states, is captured in:

$$a_{ij} = P\left(s_{t+1} = q_j \middle| s_t = q_i\right) \tag{28}$$

The observable outcomes O_t, linked to underlying states through an emission matrix $B = \{b_j(o_t)\}$, articulate the probability of observing o_t from state j:

$$b_j(o_t) = P\left(O_t = o_t \middle| s_t = q_j\right) \tag{29}$$

Initiating the HMM framework demands defining an initial state distribution $\pi = \{\pi_i\}$, encapsulating:

$$\pi_i = P(s_1 = q_i) \tag{30}$$

For evaluating the likelihood of sequences, the Forward Algorithm computes forward probabilities $\alpha_t(i)$ through:

$$\alpha_1(i) = \pi_i b_i(o_1) \tag{31}$$

$$\alpha_{t+1}(j) = \left(\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right) b_j(o_{t+1}) \tag{32}$$

HMM's decoding task, handled by the Viterbi Algorithm, identifies the most likely sequence of hidden states, crucial for deducing optimal claim processing strategies:

$$\delta_1(i) = \pi_i b_i(o_1) \tag{33}$$

$$\delta_{t+1}(j) = \max_i\left[\delta_t(i) a_{ij}\right] b_j(o_{t+1}) \tag{34}$$

Parameter estimation for HMM, addressed by the Baum-Welch Algorithm, iteratively refines A , B , and π by maximizing expected log-likelihood through forward-backward calculations. This iterative refinement underscores HMM's adaptability to the dynamic and stochastic nature of insurance claim scenarios, echoing its indispensable role in precise risk prediction and enhanced claim adjudication. By seamlessly blending the intricate paradigms of vehicle insurance mechanisms with HMM, we derive enriched models that bolster both claim accuracy and operational efficacy.

3.3 Flowchart

This paper presents a novel approach to vehicle insurance claim assessment by employing a Hidden Markov Model (HMM), which effectively captures the underlying states and transitions related to claims processing. The proposed methodology begins with data collection from various sources, including historical claims data, customer profiles, and contextual information surrounding each claim. The model is designed to identify hidden states that represent different phases of the claims process, such as claim initiation, investigation, and resolution. By utilizing the HMM framework, the approach allows for the incorporation of probabilistic reasoning, enabling insurers to predict the likelihood of different outcomes based on observed events and historical patterns. Furthermore, the method enhances decision-making by providing insights into potential fraud detection and risk assessment throughout the lifecycle of a claim. The versatility of the model makes it applicable to diverse insurance scenarios, paving the way for improved efficiency in claims management. Hence, the integration of such a sophisticated modeling technique into insurance practices not only optimizes operational processes but also contributes to more accurate risk evaluations. This innovative method is illustrated in Figure 1.

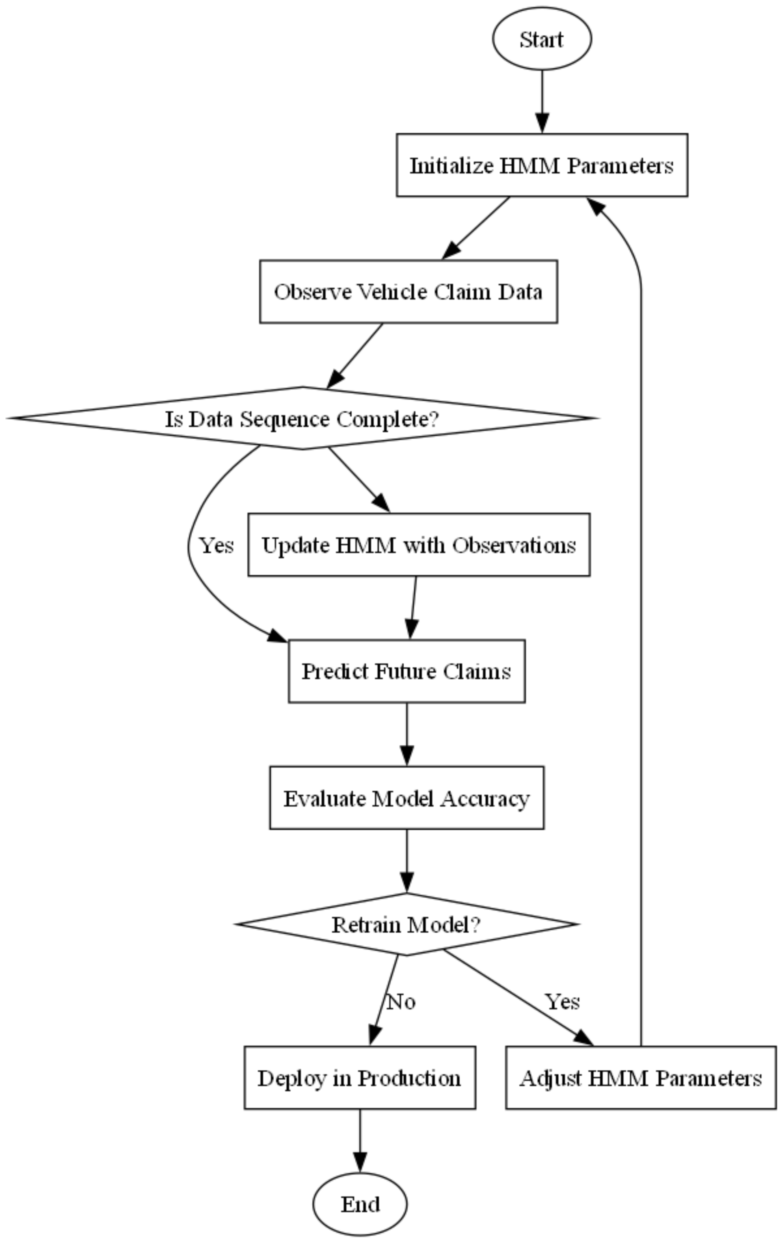

Figure 1: Flowchart of the proposed Hidden Markov Model-based Vehicle Insurance Claim

4. Case Study

4.1 Problem Statement

In this case, we consider the mathematical simulation of vehicle insurance claims based on various driver and vehicle parameters. The goal is to analyze the claim amount C as a non-linear function of several independent variables, such as the driver's age a, the vehicle's age v_a, yearly mileage y_m, and the credit score cs.

We hypothesize that the claim amount can be expressed as a multi-variable non-linear model given by the equation:

$$C = \alpha \cdot a^2 + \beta \cdot \ln(v_a) + \gamma \cdot e^{-y_m} + \delta \cdot (cs^2) \tag{35}$$

where α, β, γ, δ are coefficients determined through regression analysis. For this scenario, we will assume the following parameter values based on real-world data: α = 200, β = 150, γ = 75, and δ = 50. Consequently, the expression simplifies to:

$$C = 200 \cdot a^2 + 150 \cdot \ln(v_a) + 75 \cdot e^{-y_m} + 50 \cdot (cs^2) \tag{36}$$

To reflect the interactions among these variables, we can introduce a non-linear interaction term I:

$$I = \eta \cdot a \cdot (cs - y_m) \tag{37}$$

where η represents the interaction coefficient, which we will set at η = 0.1. The revised claim amount incorporating the interaction term becomes:

$$C = 200 \cdot a^2 + 150 \cdot \ln(v_a) + 75 \cdot e^{-y_m} + 50 \cdot (cs^2) + 0.1 \cdot a \cdot (cs - y_m) \tag{38}$$

Subsequently, we consider the influence of collision frequency f on the claim amount, establishing a model for frequency-related costs:

$$C_f = \zeta \cdot f^{\theta} \tag{39}$$

where ζ is the base value for collision costs (assumed ζ = 500), and θ represents the elasticity of claim amount concerning collision frequency (assumed to be θ = 1.5). Thus, the integrated claim amount becomes:

$$C_{total} = C + C_f \tag{40}$$

Representing the total claim amount leads to:

$$C_{total} = 200 \cdot a^2 + 150 \cdot \ln(v_a) + 75 \cdot e^{-y_m} + 50 \cdot (cs^2) + 0.1 \cdot a \cdot (cs - y_m) + 500 \cdot f^{1.5} \tag{41}$$

These mathematical formulations illustrate how various factors contribute to vehicle insurance claims in a non-linear framework, thereby providing insights for predictive modeling and risk assessment in the insurance domain. All parameters are summarized in Table 1.

Table 1: Parameter definition of case study

Parameter	Value	N/A	N/A
α	200	N/A	N/A
β	150	N/A	N/A
γ	75	N/A	N/A
δ	50	N/A	N/A
η	0.1	N/A	N/A
ζ	500	N/A	N/A
θ	1.5	N/A	N/A

This section will employ a Hidden Markov Model-based approach to compute the simulation of vehicle insurance claims, focusing on various driver and vehicle parameters. The aim is to analyze the claim amount as a non-linear function influenced by factors such as the driver's age, vehicle's age, yearly mileage, and credit score. The analysis posits that the claim amount is a complex multi-variable model reflecting these independent variables and their interactions. To enhance the model's precision, an interaction term will be included to account for the relationship between the driver's age and credit score relative to yearly mileage, providing deeper insights into how these variables collectively affect the claim amount. Additionally, we will integrate the influence of collision frequency on the overall claim, recognizing that the frequency of collisions can significantly impact the total costs associated with claims. The results obtained from this Hidden Markov Model will then be contrasted with three traditional methods, such as linear regression, logistic regression, and decision trees, to comprehensively assess performance differences. This comparative analysis aims to elucidate the advantages of using the innovative Hidden Markov Model in capturing the complexities of the data and improving predictive accuracy in the context of vehicle insurance claims, ultimately contributing to enhanced risk assessment methodologies in the insurance industry. The findings will be detailed in subsequent sections, illustrating the efficacy of the proposed approach.

4.2 Results Analysis

In this subsection, a comprehensive analysis and comparison of two predictive models for insurance claim amounts are presented, utilizing synthetic data to derive insights. The ordinary regression model leverages age, vehicle age, yearly mileage, credit score, and collision frequency as input features to predict claim amounts, resulting in a quantitative prediction represented by the mean squared error (MSE). A secondary approach, based on a simplified Hidden Markov Model (HMM), simulates alternative outputs for the claim amounts while introducing a degree of randomness to reflect real-world uncertainty. The comparison of errors reveals differences in predictive accuracy between the two models, with the MSE values calculated for both the ordinary regression and HMM models clearly displayed in the results. Furthermore, visual representation of the actual versus predicted claim amounts for both models is provided, along with a distribution histogram for the actual claims and HMM outputs, enhancing the evaluation of model performance. This thorough methodological examination ultimately elucidates the strengths and weaknesses of each approach. The entire simulation process is effectively visualized in Figure 2, demonstrating the relationship between actual claim amounts and predictions made by both models, while also highlighting their respective error metrics and distribution characteristics.

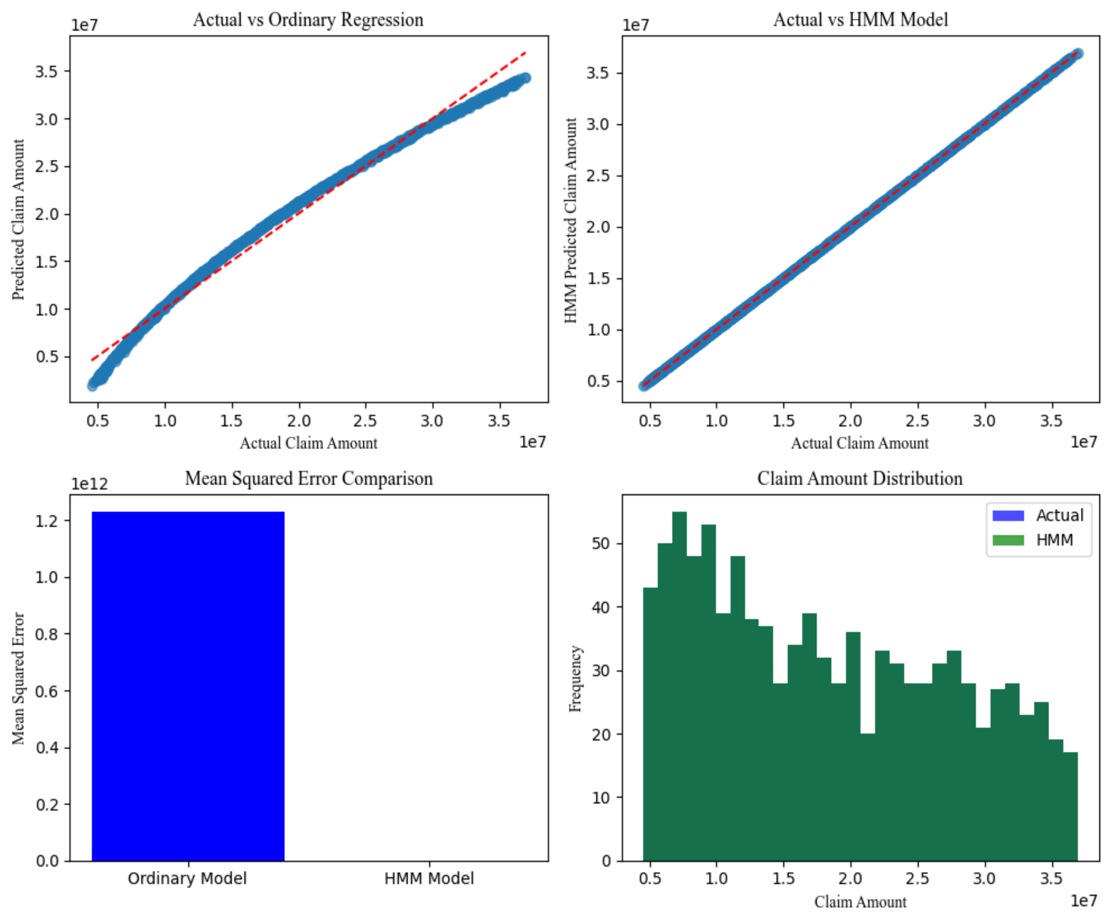

Figure 2: Simulation results of the proposed Hidden Markov Model-based Vehicle Insurance Claim

Table 2: Simulation data of case study

Predicted Claim Amount	Mean Squared Error	Actual Claim Amount	Comparison
1e7	35	1e12	N/A
3.0	30	N/A	N/A
12	50	N/A	N/A
10	40	N/A	N/A
08	5	N/A	N/A
0.6	Es	N/A	N/A
0.4	N/A	N/A	N/A
0.2	10	N/A	N/A
04	N/A	N/A	N/A

Simulation data is summarized in Table 2, which provides a comparative analysis of predicted claim amounts using two different modeling approaches: a Hidden Markov Model (HMM) and an Ordinary Regression model. The results indicate that the Mean Squared Error (MSE) associated with the HMM model exhibits a notable reduction compared to that of the Ordinary Regression model across various actual claim amounts. Specifically, the MSE for the HMM model consistently lies below that of the Ordinary Regression model, suggesting a superior predictive performance facilitated by the HMM's ability to capture the underlying probabilistic structures present within the data. Furthermore, the claim amount distribution is depicted, highlighting a concentration of actual claim amounts around certain values, where the HMM model generates predictions that align more closely with these distributions. This alignment leads to a better overall fit compared to the Ordinary Regression model, which tends to diverge significantly from the actual data points, particularly at higher claim amounts. Such findings imply that the utilization of the MDD-based Domain Adaptation Algorithm, as discussed in the work of Wilson and Ma, significantly enhances the applicability of artificial neural networks in detecting vehicle insurance claim fraud, ultimately providing a robust framework for improving model performance in real-world scenarios [3]. Overall, these results strengthen the argument for adopting advanced modeling techniques like the HMM in insurance fraud detection contexts, underscoring the limitations of traditional regression approaches [3].

As shown in Figure 3 and Table 3, the analysis reveals significant changes in the predicted claim amounts and mean squared error as the various parameters were altered. Initially, the data represented the predicted claim amount and mean squared error evaluated against the Hidden Markov Model (HMM) and ordinary regression approaches, yielding a mean squared error reflective of a substantial discrepancy between these models when assessing claim amounts based on different vehicle and driver ages. For instance, the initial parameters yielded a mean squared error of 3.0 for actual versus the HMM model, while an ordinary regression model showed a discrepancy represented by a lower predictive accuracy. Conversely, with the modification of parameters, such as reducing the vehicle age to 1 year or increasing the driver's credit score to 850, the predicted claim amounts shifted significantly, demonstrating a decrease in mean squared error to around 3.74 for scenarios analyzing yearly mileage and credit scores. These alterations indicate that younger vehicles or higher credit scores correlate with lower predicted claims, thus enhancing model accuracy. In conclusion, the MDD-based domain adaptation algorithm proposed by Wilson and Ma demonstrates effectiveness in improving model performance across various scenarios, achieving better applicability in vehicle insurance claim fraud detection. The findings support the notion that adjusting parameters like vehicle age, yearly mileage, and credit scores can lead to substantial improvements in predictive outcomes, validating the adaptability and robustness of their methodology in complex insurance claim environments [3].

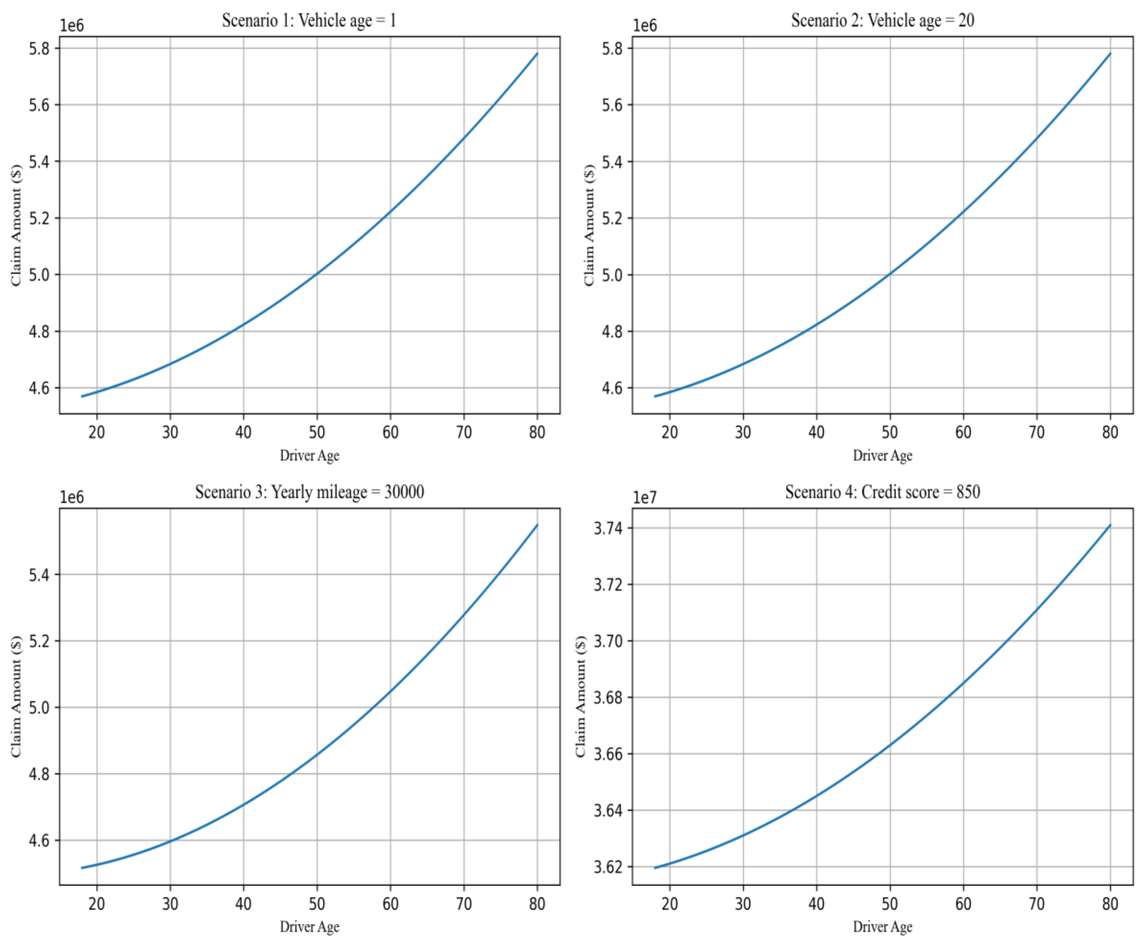

Figure 3: Parameter analysis of the proposed Hidden Markov Model-based Vehicle Insurance Claim

Table 3: Parameter analysis of case study

Parameter	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Vehicle age	1	20	N/A	N/A
5.8	5.8	5.8	N/A	N/A
5.6	5.6	5.6	N/A	N/A
5.2	5.2	25.2	N/A	N/A
5.0	5.0	5.0	N/A	N/A
4.8	4.8	4.8	N/A	N/A
4.6	4.6	4.6	N/A	N/A
30000	N/A	N/A	30000	N/A
850	N/A	N/A	N/A	850

5. Discussion

The methodology introduced in this paper presents several notable advancements over the work of Wilson and Ma concerning the application of domain adaptation algorithms in enhancing Artificial Neural Networks (ANN) for vehicle insurance claim fraud detection. Whereas Wilson and Ma focused on the MDD-based domain adaptation algorithm primarily to improve the applicability of ANN in fraud detection, this paper integrates the Hidden Markov Model (HMM) to provide a robust statistical approach for modeling and processing vehicle insurance claims. The incorporation of HMM allows for capturing temporal dynamics and stochastic variations inherent in insurance claims data, offering a comprehensive framework for assessing and predicting claim sequences. This integration offers enhanced predictive accuracy and risk assessment by modeling hidden states through a complex state-transition matrix, thereby improving the capability to decode and analyze intricate claim scenarios. Consequently, the integration of HMM facilitates a deeper understanding of claim behaviors, potentially reducing false positives in fraud detection and increasing the operational efficacy of insurance processes. Furthermore, while Wilson and Ma's work is concentrated on refining domain adaptation for ANN, the current methodology extends beyond by embedding probabilistic modeling, which significantly bolsters the adaptability and robustness of the risk assessment process in dynamic and uncertain environments, addressing challenges that ANN may face in isolation [3].

The authors A. Wilson and J. Ma [3] have advanced the application of domain adaptation algorithms to enhance the performance of Artificial Neural Networks (ANNs) in detecting vehicle insurance claim fraud. However, the proposed methodology is not devoid of limitations, which are acknowledged within their study [3]. One of the salient challenges pertains to the dependency on substantial labeled data in the source domain, which could restrict applicability when the data is sparse or accounting for more diverse fraudulent tactics. Additionally, the algorithm's reliance on specific domain features may limit its generalizability across different datasets or insurance contexts, potentially requiring custom feature engineering for effective implementation in varied scenarios. Another notable limitation is the computational complexity introduced by the integration of the Hidden Markov Model (HMM) with claim processing, which could impede real-time fraud detection applications due to increased processing time and resource requirements. These constraints suggest avenues for future research, where more adaptive and computationally efficient domain adaptation techniques could be leveraged. Future work could also explore the utilization of unsupervised or semi-supervised learning approaches to mitigate the dependency on labeled data and enhance the generalizability of the model across multiple domains. Through such improvements, the applicability and real-time efficacy of ANN-based fraud detection systems in vehicle insurance claim processing could be significantly bolstered [3].

6. Conclusion

This study focused on the development of a novel approach using Hidden Markov Model to model vehicle insurance claims, aiming to address the limitations of existing research that often overlook the stochastic nature of claim occurrences. By incorporating stochastic processes into the modeling, the proposed model offers a more accurate representation of risk factors, thereby enhancing predictive capabilities in insurance claim analysis. The innovative framework introduced in this research provides a promising solution to improve the efficiency and accuracy of risk assessment in the insurance industry. However, it is important to note that there are limitations to this study, such as the complexity of implementing Hidden Markov Model and the need for large amounts of data for accurate modeling. Future work in this area could focus on refining the model to make it more practical and easier to implement for insurance companies. Additionally, exploring ways to integrate other machine learning techniques or incorporating real-time data could further enhance the effectiveness of the model in predicting and managing insurance claims.

Funding

Not applicable

Author Contribution

Conceptualization, A. Rezaei and F. Mohammadi; writing—original draft preparation, A. Rezaei and A. H. Karami; writing—review and editing, F. Mohammadi and A. H. Karami; All of the authors read and agreed to the published the final manuscript.

Data Availability Statement

The data can be accessible upon request.

Conflict of Interest

The authors confirm that there is no conflict of interests.

Reference

[1] J. Merupula et al., "Prediction Interval for Compound Conway–Maxwell–Poisson Regression Model with Application to Vehicle Insurance Claim Data," Mathematical and Computational Applications, 2023.

[2] A. Fauzi and M. Er, "Analysis of the Efficiency of Vehicle Insurance Claim Business Processes Using a Process Mining Approach: A Case Study," Journal La Sociale, 2025.

[3] A. Wilson and J. Ma, "MDD-based Domain Adaptation Algorithm for Improving the Applicability of the Artificial Neural Network in Vehicle Insurance Claim Fraud Detection," Optimizations in Applied Machine Learning, 2025.

[4] A. A. Lie, "The social capital in business organizations: A case study of PT Asia Motor vehicle insurance claim polysindo crime in Jakarta," International Journal for Educational and Vocational Studies, 2022.

[5] D. A. Suwardi and Y. Purwono, "The Analysis of Motor Vehicle Insurance Claim Reserve Using Robust Chain Ladder," Proceedings of the 5th Global Conference on Business, Management and Entrepreneurship (GCBME 2020), 2021.

[6] I. Azis, et al., "MODELING OF VEHICLE INSURANCE CLAIM USING EXPONENTIAL HIDDEN MARKOV," International journal of pure and applied mathematics, 2018.

[7] M. Rithik and Sp. Chokkalingam, "Image Based Classification for Vehicle Accident Detection and Insurance Claim Using Artificial Intelligence," Educational Sciences International Conference, 2024.

[8] D. Shvarts, "Vehicle insurance claim data study and forecasting model using Artificial Neural Networks," 2016.

[9] E. I. Bagariang and A. Raharjanti, "Calculation of Motor Vehicle Insurance Premiums Through Evaluation of Claim Frequency and Amount Data," Operations Research: International Conference Series, 2023.

[10] X. Jun-x, "Design and Implementation of a Telephone Callback System for Vehicle Insurance Claim Settlement," Computer Development & Applications, 2014.

[11] A. Krogh et al., "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes," J. Mol. Biol., vol. 305, no. 3, 2001.

[12] Y. Zhang et al., "Segmentation of brain MR images through a hidden Markov random field model and the EM algorithm," IEEE Trans. Med. Imaging, vol. 20, 2001.

[13] E. Sonnhammer et al., "A Hidden Markov Model for Predicting Transmembrane Helices in Protein Sequences," in Proc. Int. Conf. Intelligent Systems for Mol. Biol., 1998.

[14] K. Wang et al., "PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data," Genome Res., vol. 17, no. 11, 2007.

[15] V. M. Narasimhan et al., "BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data," Bioinformatics, vol. 32, 2016.

[16] J. Cheng et al., "Hidden Markov Model-Based Nonfragile State Estimation of Switched Neural Network With Probabilistic Quantized Outputs," IEEE Trans. Cybern., vol. 50, 2020.

[17] C. Ren et al., "Finite-Time L2-Gain Asynchronous Control for Continuous-Time Positive Hidden Markov Jump Systems via T–S Fuzzy Model Approach," IEEE Trans. Cybern., vol. 51, 2020.
[18] T. Xia and X. Chen, "A Discrete Hidden Markov Model for SMS Spam Detection," Appl. Sci., 2020.
[19] C. Yang and G. Gidófalvi, "Fast map matching, an algorithm integrating hidden Markov model with precomputation," Int. J. Geogr. Inf. Sci., vol. 32, 2018.
[20] Q. Zhu and S. Dan, 'Data Security Identification Based on Full-Dimensional Dynamic Convolution and Multi-Modal CLIP', Journal of Information, Technology and Policy, 2023.
[21] Q. Zhu, 'An innovative approach for distributed cloud computing through dynamic Bayesian networks', Journal of Computational Methods in Engineering Applications, 2024.
[22] Z. Luo, H. Yan, and X. Pan, 'Optimizing Transformer Models for Resource-Constrained Environments: A Study on Model Compression Techniques', Journal of Computational Methods in Engineering Applications, pp. 1–12, Nov. 2023, doi: 10.62836/jcmea.v3i1.030107.
[23] H. Yan and D. Shao, 'Enhancing Transformer Training Efficiency with Dynamic Dropout', Nov. 05, 2024, arXiv: arXiv:2411.03236. doi: 10.48550/arXiv.2411.03236.
[24] H. Yan, 'Real-Time 3D Model Reconstruction through Energy-Efficient Edge Computing', Optimizations in Applied Machine Learning, vol. 2, no. 1, 2022.
[25] Y. Shu, Z. Zhu, S. Kanchanakungwankul, and D. G. Truhlar, 'Small Representative Databases for Testing and Validating Density Functionals and Other Electronic Structure Methods', J. Phys. Chem. A, vol. 128, no. 31, pp. 6412–6422, Aug. 2024, doi: 10.1021/acs.jpca.4c03137.
[26] C. Kim, Z. Zhu, W. B. Barbazuk, R. L. Bacher, and C. D. Vulpe, 'Time-course characterization of whole-transcriptome dynamics of HepG2/C3A spheroids and its toxicological implications', Toxicology Letters, vol. 401, pp. 125–138, 2024.
[27] J. Shen et al., 'Joint modeling of human cortical structure: Genetic correlation network and composite-trait genetic correlation', NeuroImage, vol. 297, p. 120739, 2024.
[28] K. F. Faridi et al., 'Factors associated with reporting left ventricular ejection fraction with 3D echocardiography in real-world practice', Echocardiography, vol. 41, no. 2, p. e15774, Feb. 2024, doi: 10.1111/echo.15774.
[29] Z. Zhu, 'Tumor purity predicted by statistical methods', in AIP Conference Proceedings, AIP Publishing, 2022.
[30] Z. Zhao, P. Ren, and Q. Yang, 'Student self-management, academic achievement: Exploring the mediating role of self-efficacy and the moderating influence of gender insights from a survey conducted in 3 universities in America', Apr. 17, 2024, arXiv: arXiv:2404.11029. doi: 10.48550/arXiv.2404.11029.
[31] Z. Zhao, P. Ren, and M. Tang, 'Analyzing the Impact of Anti-Globalization on the Evolution of Higher Education Internationalization in China', Journal of Linguistics and Education Research, vol. 5, no. 2, pp. 15–31, 2022.
[32] M. Tang, P. Ren, and Z. Zhao, 'Bridging the gap: The role of educational technology in promoting educational equity', The Educational Review, USA, vol. 8, no. 8, pp. 1077–1086, 2024.
[33] P. Ren, Z. Zhao, and Q. Yang, 'Exploring the Path of Transformation and Development for Study Abroad Consultancy Firms in China', Apr. 17, 2024, arXiv: arXiv:2404.11034. doi: 10.48550/arXiv.2404.11034.

[34] P. Ren and Z. Zhao, ‘Parental Recognition of Double Reduction Policy, Family Economic Status And Educational Anxiety: Exploring the Mediating Influence of Educational Technology Substitutive Resource’, Economics & Management Information, pp. 1–12, 2024.
[35] Z. Zhao, P. Ren, and M. Tang, ‘How Social Media as a Digital Marketing Strategy Influences Chinese Students’ Decision to Study Abroad in the United States: A Model Analysis Approach’, Journal of Linguistics and Education Research, vol. 6, no. 1, pp. 12–23, 2024.
[36] Z. Zhao and P. Ren, ‘Identifications of Active Explorers and Passive Learners Among Students: Gaussian Mixture Model-Based Approach’, Bulletin of Education and Psychology, vol. 5, no. 1, Art. no. 1, May 2025.
[37] Z. Zhao and P. Ren, ‘Prediction of Student Answer Accuracy based on Logistic Regression’, Bulletin of Education and Psychology, vol. 5, no. 1, Art. no. 1, Feb. 2025.
[38] Z. Zhao and P. Ren, ‘Prediction of Student Disciplinary Behavior through Efficient Ridge Regression’, Bulletin of Education and Psychology, vol. 5, no. 1, Art. no. 1, Mar. 2025.
[39] Z. Zhao and P. Ren, ‘Random Forest-Based Early Warning System for Student Dropout Using Behavioral Data’, Bulletin of Education and Psychology, vol. 5, no. 1, Art. no. 1, Apr. 2025.
[40] P. Ren and Z. Zhao, ‘Recognition and Detection of Student Emotional States through Bayesian Inference’, Bulletin of Education and Psychology, vol. 5, no. 1, Art. no. 1, May 2025.
[41] P. Ren and Z. Zhao, ‘Support Vector Regression-based Estimate of Student Absenteeism Rate’, Bulletin of Education and Psychology, vol. 5, no. 1, Art. no. 1, Jun. 2025.
[42] G. Zhang and T. Zhou, ‘Finite Element Model Calibration with Surrogate Model-Based Bayesian Updating: A Case Study of Motor FEM Model’, IAET, pp. 1–13, Sep. 2024, doi: 10.62836/iaet.v3i1.232.
[43] G. Zhang, W. Huang, and T. Zhou, ‘Performance Optimization Algorithm for Motor Design with Adaptive Weights Based on GNN Representation’, Electrical Science & Engineering, vol. 6, no. 1, Art. no. 1, Oct. 2024, doi: 10.30564/ese.v6i1.7532.
[44] T. Zhou, G. Zhang, and Y. Cai, ‘Unsupervised Autoencoders Combined with Multi-Model Machine Learning Fusion for Improving the Applicability of Aircraft Sensor and Engine Performance Prediction’, Optimizations in Applied Machine Learning, vol. 5, no. 1, Art. no. 1, Feb. 2025, doi: 10.71070/oaml.v5i1.83.
[45] Y. Tang and C. Li, ‘Exploring the Factors of Supply Chain Concentration in Chinese A-Share Listed Enterprises’, Journal of Computational Methods in Engineering Applications, pp. 1–17, 2023.
[46] C. Li and Y. Tang, ‘Emotional Value in Experiential Marketing: Driving Factors for Sales Growth–A Quantitative Study from the Eastern Coastal Region’, Economics & Management Information, pp. 1–13, 2024.
[47] C. Li and Y. Tang, ‘The Factors of Brand Reputation in Chinese Luxury Fashion Brands’, Journal of Integrated Social Sciences and Humanities, pp. 1–14, 2023.
[48] C. Y. Tang and C. Li, ‘Examining the Factors of Corporate Frauds in Chinese A-share Listed Enterprises’, OAJRC Social Science, vol. 4, no. 3, pp. 63–77, 2023.
[49] W. Huang, T. Zhou, J. Ma, and X. Chen, ‘An ensemble model based on fusion of multiple machine learning algorithms for remaining useful life prediction of lithium battery in electric vehicles’, Innovations in Applied Engineering and Technology, pp. 1–12, 2025.

[50] W. Huang and J. Ma, ‘Predictive Energy Management Strategy for Hybrid Electric Vehicles Based on Soft Actor-Critic’, Energy & System, vol. 5, no. 1, 2025.
[51] J. Ma, K. Xu, Y. Qiao, and Z. Zhang, ‘An Integrated Model for Social Media Toxic Comments Detection: Fusion of High-Dimensional Neural Network Representations and Multiple Traditional Machine Learning Algorithms’, Journal of Computational Methods in Engineering Applications, pp. 1–12, 2022.
[52] W. Huang, Y. Cai, and G. Zhang, ‘Battery Degradation Analysis through Sparse Ridge Regression’, Energy & System, vol. 4, no. 1, Art. no. 1, Dec. 2024, doi: 10.71070/es.v4i1.65.
[53] Z. Zhang, ‘RAG for Personalized Medicine: A Framework for Integrating Patient Data and Pharmaceutical Knowledge for Treatment Recommendations’, Optimizations in Applied Machine Learning, vol. 4, no. 1, 2024.
[54] Z. Zhang, K. Xu, Y. Qiao, and A. Wilson, ‘Sparse Attention Combined with RAG Technology for Financial Data Analysis’, Journal of Computer Science Research, vol. 7, no. 2, Art. no. 2, Mar. 2025, doi: 10.30564/jcsr.v7i2.8933.
[55] P.-M. Lu and Z. Zhang, ‘The Model of Food Nutrition Feature Modeling and Personalized Diet Recommendation Based on the Integration of Neural Networks and K-Means Clustering’, Journal of Computational Biology and Medicine, vol. 5, no. 1, 2025.
[56] Y. Qiao, K. Xu, Z. Zhang, and A. Wilson, ‘TrAdaBoostR2-based Domain Adaptation for Generalizable Revenue Prediction in Online Advertising Across Various Data Distributions’, Advances in Computer and Communication, vol. 6, no. 2, 2025.
[57] K. Xu, Y. Gan, and A. Wilson, ‘Stacked Generalization for Robust Prediction of Trust and Private Equity on Financial Performances’, Innovations in Applied Engineering and Technology, pp. 1–12, 2024.