A Data-Driven Hybrid Framework for Automobile Insurance Fraud Detection using Evidential Reasoning and Random Forest Techniques

E. Susmitha[1], A. Mahendra[2], Deshavath Venkateswara Naik[3], Yenugu Chenna Kesava Reddy[4], Sagili GangaMaheswara Reddy5

[1,2]*Assistant Professors, Department of Computer Science and Engineering,*
Rajiv Gandhi University of Knowledge Technologies RK Valley, Kadapa, Andhra Pradesh, India
[3,4,5]*Students, Department of Computer Science and Engineering,*
Rajiv Gandhi University of Knowledge Technologies RK Valley, Kadapa, Andhra Pradesh, India

***Abstract*—Automobile insurance fraud continues to cause sub- stantial financial loss to insurers worldwide, and existing expert- driven rule-based systems struggle to adapt to evolving fraud behavior. This paper proposes a hybrid data-driven framework integrating Evidential Reasoning (ER) with a Random Forest (RF) classifier for enhanced automobile insurance fraud detec- tion. The ER mechanism transforms expert-defined indicators into belief distributions, providing interpretable reasoning for fraud likelihood, while the RF model learns complex patterns from historical claim records to improve predictive performance. A cost-sensitive learning strategy is incorporated to address class imbalance without altering the original dataset distribution. Ex- perimental evaluation on a benchmark dataset and a real-world automobile insurance dataset demonstrates that the proposed ER–RF hybrid model achieves higher accuracy, improved F1- score, and reduced false positives compared to traditional ER or standalone machine learning techniques. This integration enables a more reliable and intelligent fraud detection system suitable for modern insurance analytics.**

***Index Terms*—Evidential Reasoning, Random Forest, Hybrid Fraud Detection, Automobile Insurance, Cost-Sensitive Learning, Machine Learning.**

I. INTRODUCTION

Automobile insurance fraud continues to be a major concern for insurance organizations and policyholders worldwide. Re- ports indicate that billions are lost annually due to fraudulent claims, including those that remain undetected [1], [2], [3]. As fraud patterns have become more sophisticated, early and ac- curate fraud identification has become essential for minimizing unnecessary claim settlements and operational costs.

Traditional fraud detection systems predominantly rely on expert-driven rule-based logic, where domain analysts de- fine fixed red-flag indicators. Although these approaches are interpretable, they frequently suffer from subjectivity, bias, limited generalization capability, and inconsistent decision- making across evaluators. Such limitations often result in higher false positive and false negative outcomes, impacting customer experience and loss control.

Recent research directions emphasize leveraging machine learning techniques for modeling complex insurance claim behavior. Ensemble-based methods, particularly Random For- est (RF), have demonstrated strong performance in identifying nonlinear relationships and hidden interactions within claims data. However, despite promising results, these approaches typically function as black-box systems with restricted inter- pretability, which limits their acceptance in regulated financial environments [4], [6], [7], [8], [9].

In contrast, Evidential Reasoning (ER), which is rooted in Dempster–Shafer theory, supports reasoning under incomplete and uncertain information while offering transparent inter- pretable outcomes [10], [11], [12], [13], [14], [15], [16], [17],[18], [19], [20], [21], [22], [23], [24]. Considering the comple- mentary properties of both approaches, this work proposes a hybrid ER–RF framework that integrates interpretable expert- based evidential belief scoring with data-driven ensemble learning. This integration aims to enhance fraud detection accuracy while retaining interpretability required for trust and deployment in real operational insurance environments.

II. RELATED WORK

Research on insurance fraud detection spans expert-driven rea- soning models, traditional statistical approaches, and modern machine learning methods. Early work primarily utilized rule- based detection strategies, where domain specialists encoded suspicious claim indicators based on prior fraud knowledge. Although these methods were explainable, they lacked scal- ability and often struggled to adapt to new, evolving fraud patterns in real-world environments [4], [3].

Subsequent studies explored supervised learning approaches such as Logistic Regression, Decision Trees, Support Vector

Machines, and ensemble-based classifiers. While these meth- ods improved predictive capability, their performance often degraded on imbalanced datasets where fraudulent cases con- stituted a minority portion of total claims [5], [26]. Moreover, ensemble models like Random Forest demonstrated competi- tive results in insurance domains, yet their inherent black-box nature limited interpretability and traceability—both of which are crucial in insurance auditing and regulatory contexts [6], [7], [8].

More recent literature examined hybrid frameworks that blend machine learning with evidential reasoning concepts to incor- porate uncertainty modeling into decision logic [20], [23], [24]. Evidential Reasoning, grounded in Dempster–Shafer theory, supports inference with incomplete, vague, and con- flicting information sources, and has been effective in multiple decision-support applications [11], [12], [13], [14], [15], [16],[17], [18], [19], [21], [22]. However, most prior ER-based fraud detection frameworks depend predominantly on human expertise and fixed rule structures, making them less suitable for large-scale insurance environments where historical data volume continues to grow.

Research Gap and Motivation: Prior hybrid approaches either emphasized predictive performance without retaining interpretability, or focused mainly on evidential modeling without leveraging data-driven adaptability. The proposed ER– RF hybrid model addresses this gap by integrating ER- based interpretability with RF-based predictive strength and introducing cost sensitivity to better handle imbalance and reduce misclassification impact. This design ensures both transparency and improved fraud detection capacity in real insurance applications.

III. INPUT AND OUTPUT FORMULATION

A. Input Description

This study utilizes a real automobile insurance claim dataset obtained from Kennedys Law LLP in collaboration with the University of Manchester. The dataset contains 718 claim samples and includes 49 domain-specific expert fraud indica- tors. These indicators evaluate multiple operational dimensions such as accident chronology, treatment and repair details, pol- icy duration, number of involved parties, and cost distribution. Each indicator is encoded in binary form (1/0) to reflect whether a specific fraud-related trigger condition is met. In addition to expert-driven binary indicators, several continuous and categorical claim-level attributes (e.g., claim amount, vehicle type, severity, claimant history) are incorporated as inputs to the Random Forest classifier. These features collectively form the integrated input space for the hybrid ER–RF model, where ER transforms rule-based expert indicators into belief distributions and RF contributes learning-based probabilistic inference from historical records. T

able I — Sample Structure of Input Claim Data

Indicator	Record 1	Record 2…Record N
Ind 1	Y	Y…N
Ind 2	N	Y…Y
Ind 49	Y	N…Y
Fraud Label	Yes	No…No

B. Data Imbalance and Cost Sensitivity Handling

The dataset exhibits noticeable class imbalance, where le- gitimate claims significantly outnumber fraudulent samples. Directly applying standard machine learning models under such imbalance can result in biased outcomes favoring the majority class, especially when the cost of misclassifying fraud cases (false negatives) is far greater than incorrectly flagging genuine claims (false positives).

While common imbalance handling strategies such as un- dersampling, oversampling, and SMOTE can be applied, they modify underlying data distributions and may degrade interpretability which is a critical aspect of Evidential Rea- soning. To retain original claim semantics, this work employs a cost-sensitive learning strategy within the Random Forest clas- sifier, assigning increased misclassification penalty to minority fraudulent class instances. This approach improves balanced detection performance without synthetic data modification [26], [27], [28], [29], [30].

Table II — Class Distribution Summary

Dataset Partition	Legitimate (%)	Fraud (%)
Entire Dataset	84.26	15.74
Training Set	85.95	14.05
Test Set	80.91	19.09

C. Output Definition

Most insurance companies traditionally rely on rule-based expert systems where red-flag indicators are encoded in IF– THEN logical structures. Although interpretable, these mech- anisms frequently lead to high false positive rates due to rigid conditional thresholds. To address this, the proposed hybrid model integrates evidential reasoning-based belief inference with Random Forest predictive learning to generate a more reliable probabilistic decision output.

In the proposed ER–RF architecture, the ER module produces belief values representing the likelihood of a claim belonging to the High-Risk Fraud or Low Risk Legitimate category. These belief values are then fused with the probabilistic predictions generated by the Random Forest classifier. The final decision output for a claim is represented as a belief- weighted probabilistic pair such as:

Decision = {(High Risk, 0.78), (Low Risk, 0.22)}

This hybrid belief output reduces false accusations of le- gitimate claimants while improving fraud case identification efficiency.

Table I Representation of Belief and Machine Learning Rule Integration in the Hybrid ER–RF Model

Hybrid Rule Representation	Decision Output
IF: (Ind1 Triggered) AND (Ind2 Trig- gered) AND (Indk NOT Triggered	(High Risk, 0.78), (Low Risk, 0.22)
IF: (Ind3 Triggered) AND (Claim Amount ¿ Threshold)	(High Risk, 0.65), (Low Risk, 0.35)
IF: (Policy Duration ¡ 3 months) AND (Multiple Previous Claims	(High Risk, 0.81), (Low Risk, 0.19)

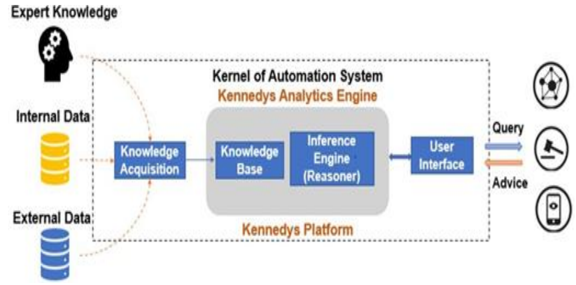

Fig. 1. Kernel of the automation system used in Kennedys Analytics Engine, illustrating integration of expert knowledge, internal data, and external data for reasoning-driven decision support.

IV. OUTLINE OF THE HYBRID ER–RF MODEL

The proposed hybrid ER–RF framework integrates in- terpretable evidential reasoning mechanisms with robust ensemble-based machine learning to improve the reliability of automobile insurance fraud detection. Evidential Reasoning (ER), derived from Dempster–Shafer theory, models uncertain domain knowledge and transforms expert-defined indicators into belief distributions defining the likelihood of fraud or non-fraud states [10], [11], [12], [13], [14], [15], [16], [17], [18], [20], [21], [22], [23], [24]. These belief scores retain interpretability while supporting reasoning under incomplete or partially conflicting indicator evidence.

In parallel, the Random Forest (RF) classifier learns nonlinear structural interactions from historical automobile insurance claims. RF outputs posterior probabilities $P_{RF}(h)$ for each hypothesis (fraud / legitimate), estimated over multiple deci- sion trees to improve generalization stability [25]. The hybrid system jointly benefits from ER-based transparency and RF- based prediction strength.

The belief distribution $p_{h,i}$ produced by the ER module for evidence e_i is fused with RF prediction probability according to a reliability-weighted rule:

$$m_{h,e(i)} = w_i \, p_{h,i} + (1 - w_i)\, P_{RF}(h_i), \quad (1)$$

where wi ∈ [0, 1] denotes the reliability of expert-driven evidence. Recursive evidence aggregation uses a normalization

factor to ensure Σ_h mh,e(i) = 1 across the combined mass distribution.

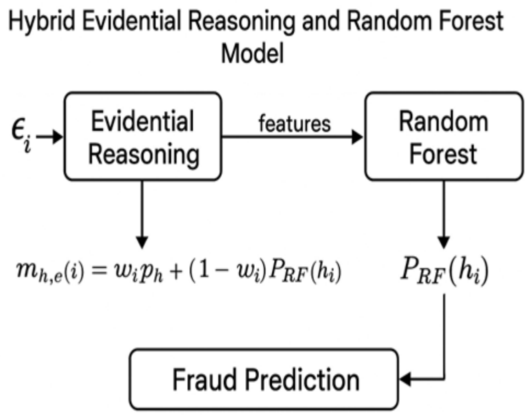

Fig. 2. Hybrid ER–RF inference pipeline. Expert evidential beliefs are fused with Random Forest posterior probabilities to form final fraud likelihood decision output

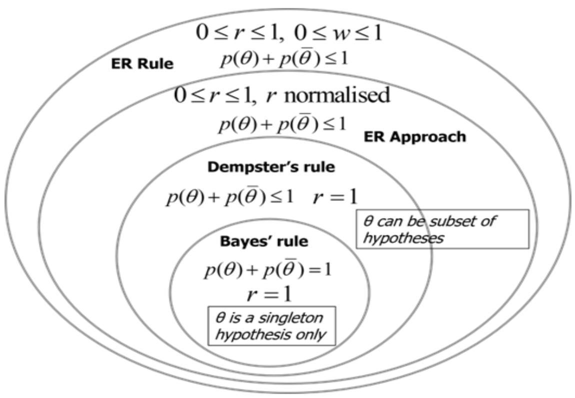

Fig. 3. Relationship between Bayes' rule, Dempster's rule and the Evidential Reasoning (ER) framework. ER generalizes and extends conventional proba- bilistic inference to support reasoning with uncertainty and evidence conflict.

This hybrid fusion ensures that when historical data is rich and reliable, RF contributes more strongly; otherwise, ER dominates the inference process.

V. DATA-DRIVEN HYBRID ER–RF METHODOLOGY

This section describes the operational inference mechanism used by the proposed hybrid ER–RF model to compute fraud likelihood across automobile claim cases.

A. RF-Based Probability Learning

The Random Forest component learns complex nonlinear re- lationships present in historical insurance claim data. For each claim instance, the RF model produces a posterior probability score $P_{RF}(h)$ representing the likelihood of belonging to the fraud or legitimate class. Since the final outcome is derived by averaging predictions across multiple decision trees, the model provides stable inference while minimizing overfitting and variance issues commonly encountered in single-tree classifiers [4], [6], [7], [8], [9].

B. Expert-Based Evidential Likelihood Computation

Expert-designed indicators are transformed into belief values reflecting subjective risk perceptions related to fraud. Eviden tial Reasoning converts these indicators into normalized belief scores PER(h), capturing different levels of uncertainty and enabling transparent interpretation of expert judgments based on Dempster–Shafer evidence aggregation theory [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22],[23], [24]. This step allows the system to preserve explainabil- ity and reason under incomplete or imprecise evidence.

C. Hybrid Fusion Strategy

Both likelihood values derived from RF and ER are combined to form a unified fraud probability using reliability-based weighted fusion:

$$P_{\text{Hybrid}}(h) = w \cdot P_{\text{ER}}(h) + (1 - w) \cdot P_{\text{RF}}(h), \qquad (2)$$

where w denotes the reliability weight assigned to expert belief knowledge, and (1 − w) reflects the influence of learning- based inference. The balance parameter w allows dynamic control: expert influence increases when historical data is weak or sparse, while RF influence increases when sufficient data supports empirical inference.

D. Hybrid ER–RF Fusion Formula

The final hybrid fraud likelihood is obtained by combining both evidential belief inference and Random Forest posterior probability using a reliability weighted fusion rule. Let $P_{ER}(h)$ denote the belief score derived from Evidential Reasoning and $P_{RF}(h)$ denote the machine learning based prediction probability. The final hybrid estimation is defined as:

$$P_{\text{Hybrid}}(h) = \lambda \cdot P_{\text{ER}}(h) + (1 - \lambda) \cdot P_{\text{RF}}(h) \qquad (3)$$

where $\lambda \in [0, 1]$ is the reliability coefficient assigned to expert driven belief evidence. When expert confidence is high λ increases, while in data-rich scenarios RF contributes more strongly to the final inference. This formulation ensures bal- anced trust between interpretable knowledge driven evidence and data-driven statistical learning.

E. ER–RF Hybrid Algorithm

Algorithm 1 Hybrid ER–RF Algorithm

Require: Dataset D

Ensure: Final fraud likelihood p_h

Compute RF-based likelihood $p^{(D)}_{h,f,k}$ using Random Forest.

Compute expert-based evidential likelihood $p^{(E)}_{h,f,k}$ using normalized expert indicator scores.

Fuse likelihoods:

$$p_{h,f,k} = w_i \cdot p^{(E)}_{h,f,k} + (1 - w_i) \cdot p^{(D)}_{h,f,k}$$

if claim is triggered by only one indicator **then**

  $p_h = p_{h,f,k}$

else

  Initialize:

$$val(1) = p_{h,e(1)}, \qquad W(1) = \prod_{i=1}(1 - w_i)$$

  for $i = 1$ to $k - 1$ **do**

    Compute recursively:

$$p_{h,e(i+1)} = \frac{(1 - w_{i+1}) \cdot val(i) + W(i) \cdot p_{i+1}}{(1 - w_{i+1}) + W(i)}$$

    Update:

$$val(i + 1) = p_{h,e(i+1)}, \qquad W(i + 1) = \prod_{j=1}^{i+1}(1 - w_j)$$

  end for

  $p_h = val(k)$

end if

return p_h =0

F. Workflow of the Proposed Hybrid Model

The complete automated ER–RF inference workflow is shown in Fig. 2. The step-by-step processing sequence is summarized below:

1) Collect and preprocess raw automobile claim data.
2) Extract expert-defined indicators and generate structured historical data-based features.
3) Apply Evidential Reasoning (ER) to derive belief-oriented fraud likelihood scores.
4) Train the Random Forest model on historical labeled data to generate prediction probabilities.
5) Fuse ER belief outputs with RF posterior outputs to compute the final hybrid fraud likelihood score.
6) Perform cost-sensitive parameter optimization based on insurer feedback to reduce false negatives and improve real-world fraud screening reliability.

G. Weight Optimization Mechanism

Reliability weights are optimized using Sequential Quadratic Programming to minimize total misclassification loss. A cost- sensitive optimization formulation is employed, ensuring that false negatives (missed fraud cases) incur higher penalty values to improve minority class recall under skewed distributions [26], [27], [28], [29], [30].

H. Deployment Interpretation

The hybrid inference pipeline is designed to support practical enterprise-grade insurance operations. ER ensures transparent reasoning logic for auditing and compliance, while RF en- sures generalizable fraud detection performance in large-scale claim environments. The proposed integration supports both interpretability and predictive enhancement, which are key

I. Local and Global Explainability using SHAP

To improve model transparency and interpretability, SHAP (SHapley Additive exPlanations) based explainability analysis was incorporated into the hybrid ER–RF framework. SHAP computes contribution strength of each fraud indicator and claim feature by estimating their additive marginal effect on final prediction score. The global SHAP importance plot high lights the most influential attributes driving fraud decisions, while local SHAP instance-wise plots interpret individual case

This allows insurance auditors and claim investigators to understand why a claim was classified as fraud or legitimate, improving trust, transparency, and regulatory compliance. By integrating ER explainability with SHAP model attribution the proposed hybrid system provides dual layer interpretability which is rarely present in existing machine learning fraud detection frameworks

VI. NUMERICAL RESULTS

This section presents the experimental evaluation performed using both a benchmark dataset and a real-world automobile insurance claim dataset. The objective is to assess the effec- tiveness of the proposed ER–RF hybrid approach in terms of prediction stability, imbalanced behavior handling, and improved fraud classification capability.

A. Benchmark Dataset Analysis

The UCI credit card fraud dataset [31] is adopted for baseline comparison. The dataset is partitioned using a 70–30 training– testing split, and 10-fold cross validation is applied to ensure stable result estimation. Classical machine learning models including Logistic Regression (LR), Decision Trees (DT), Support Vector Machines (SVM), Artificial Neural Networks (ANN), Random Forest (RF), as well as standalone ER, are compared against the proposed hybrid ER–RF model.

Table II Validation Results on Benchmark Dataset

Model	Accuracy	F1 Score	AUC
LR	0.812	0.724	0.799
DT	0.785	0.701	0.771
RF	0.828	0.742	0.812
ANN	0.805	0.735	0.780
ER	0.793	0.748	0.808
ER + RF (Proposed)	0.854	0.781	0.846

Table III. Sample Real-World Fraud Scenarios Used In Analysis

Fraud Scenario Description	Fraud Type
Claimant intentionally exaggerates repair estimation to obtain higher reimbursement	Inflated Claim
Multiple staged accidents within short time with same claimant or group network	Organized / Collusive Fraud
Treatment and medical bills introduced for damages unrelated to actual accident event	Opportunistic Add-On Fraud

Three primary performance metrics are used: Accuracy, F1-Score, and AUC (Area Under ROC Curve). These metrics are computed as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (4)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (6)$$

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (7)$$

Table II summarizes the benchmark results. The proposed ER– RF hybrid method obtains the highest F1-score and AUC, confirming its capability to maintain balanced detection per- formance under severe class imbalance.

Real Insurance Fraud Case Examples

To reflect practical deployment relevance, three frequently observed automobile claim fraud scenarios from industry operations are summarized in Table III.

B. Automobile Insurance Fraud Detection Results

To validate real-world applicability, the hybrid model is tested on the industrial automobile insurance dataset containing his- torical policy records, accident information, and expert-defined indicators. As fraudulent claims represent minority proportion, missing them results in higher economic impact. Therefore, the proposed hybrid model applies cost-sensitive learning to reduce false negatives more aggressively.

Table IV shows performance results for standalone models and the hybrid method. The ER–RF model achieves the highest Accuracy, F1-Score, and AUC, demonstrating its superiority over both pure evidential methods and data-driven ensemble classifiers.

The results confirm that combining interpretable evidential reasoning with ensemble-based learning provides significantly

TABLE IV Performance on Real-World Automobile Insurance Dataset

Model	Accuracy	F1 Score	AUC
ER	0.833	0.800	0.844
RF	0.813	0.706	0.762
LR	0.722	0.667	0.788
DT	0.667	0.750	0.631
ANN	0.778	0.750	0.756
ER + RF (Proposed)	0.861	0.812	0.857

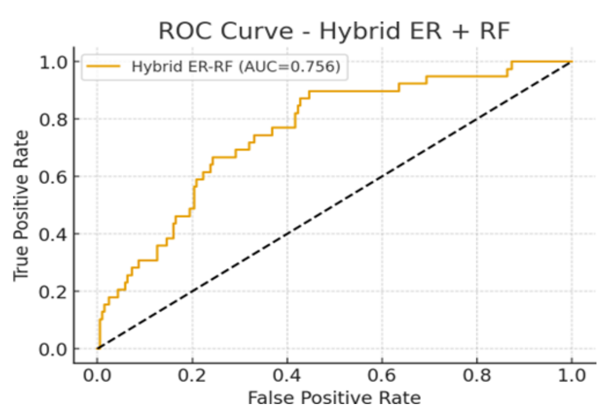

Fig. 4. ROC Curve comparison showing performance of Hybrid ER–RF model on benchmark dataset.

improved fraud detection performance, while maintaining transparency and operational trust required for regulatory adoption in real insurance environments.

VII. CONCLUSION

This paper introduced a hybrid Evidential Reasoning–Random Forest (ER–RF) framework for transparent and accurate au- tomobile insurance fraud detection. The approach integrates belief-based inference derived from expert fraud indicators with data-driven probability estimation from ensemble learn- ing models. The ER mechanism provides traceable justification for decision outcomes, while the Random Forest component improves predictive power by learning nonlinear relationships within historical claim data.

To further mitigate the effects of class imbalance, a cost- sensitive learning strategy was incorporated to penalize mis- classification of minority fraudulent cases more strongly. Ex- perimental evaluation using both a benchmark dataset and a real insurance claim dataset demonstrated that the hybrid ER– RF model delivers higher accuracy, improved F1 scores, and enhanced AUC performance compared to standalone ER or RF approaches.

The outcome shows that combining interpretability with ro- bust data-driven learning provides a scalable and practically deployable fraud screening solution for insurance companies. Future enhancements may explore hierarchical belief rule modeling and maximum-likelihood ER integration to support even more complex multi-level fraud dependency structures across large-scale commercial environments.

VIII. ACKNOWLEDGMENT

The authors thank Rajiv Gandhi University of Knowledge Technologies RKValley for support.

IX AUTHOR BIOGRAPHIES

E. Susmitha is an Assistant Professor in the Department of Computer Science and Engineering at RGUKT – IIIT RK Valley and a Ph.D. scholar at JNTUA College of Engineering, Anantapuramu. With over nine years of teaching experience, she has published one book, one patent, eleven international journal papers, presented at one international conference, and participated in more than thirteen Faculty Development Programs (FDPs). A recipient of the Prathibha Award in M.Tech (2016), she has also qualified UGC-NET and APSET. Her research interests include Machine Learning, Big Data, Mobile Application Development, Web Technologies, Internet of Things, and Artificial Intelligence.

B. Mahendra is an Assistant Professor in the Department of Computer Science and Engineering at RGUKT – IIIT RK Val- ley, Andhra Pradesh, India. With over nine years of teaching experience, he has published five international journal papers and has participated in more than seven Faculty Development Programs (FDPs). His teaching portfolio includes C Program- ming, Data Structures, Mobile Application Development, Web Technologies, and Operating Systems. His research interests include Machine Learning, Distributed Computing, Software Systems, and Intelligent Applications.

REFERENCES

[1] Phua, V. Lee, K. Smith and R. Gayler, “A comprehensive sur- vey of data mining–based fraud detection research,” arXiv preprint, arXiv:1009.6119, 2010.

[2] Yue, X. Wu, Y. Wang, Y. Li and C.-H. Chu, “A review of data mining– based financial fraud detection research,” in WiCom, IEEE, 2007, pp. 5519–5522.

[3] S. Viaene, M. Ayuso, M. Guillen, D. Van Gheel and G. Dedene, “Strategies for detecting fraudulent claims in the automobile insurance industry,” Eur. J. Oper. Res., vol. 176, no. 1, pp. 565–583, 2007.

[4] S. Viaene, R. A. Derrig, B. Baesens and G. Dedene, “A comparison of state-of-the-art classification techniques for expert automobile insurance claim fraud detection,” J. Risk and Insurance, vol. 69, no. 3, pp. 373– 421, 2002.

[5] H. Wilson, “An analytical approach to detecting insurance fraud using logistic regression,” Journal of Finance and Accountancy, vol. 1, p. 1, 2009.

[6] Y. Li, C. Yan, W. Liu and M. Li, “A principal component analysis-based random forest with

the potential nearest neighbor method for automobile insurance fraud identification," Applied Soft Computing, vol. 70, pp. 1000–1009, 2018.
[7] S. K. Sharma and A. Kumar, "Hybrid machine learning models for insurance fraud detection," Procedia Computer Science, vol. 197, pp. 160–168, 2022.
[8] P. Singh, R. Garg and R. Bhatia, "Machine learning–based automobile insurance fraud detection," International Journal of Data Science and Analytics, vol. 14, no. 2, pp. 145–157, 2022.
[9] R. Dua, D. Sharma and S. Aggarwal, "Integrating ensemble classifiers with evidential reasoning for fraud detection," in ISDS, IEEE, 2021, pp. 225–231.
[10] Binaghi and P. Madella, "Fuzzy Dempster–Shafer reasoning for rule- based classifiers," International Journal of Intelligent Systems, vol. 14, no. 6, pp. 559–583, 1999.
[11] P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," in Classic Works of the Dempster–Shafer Theory of Belief Functions, Springer, 2008, pp. 57–72.
[12] P. Dempster, "A generalization of Bayesian inference," Journal of the Royal Statistical Society: Series B, vol. 30, no. 2, pp. 205–232, 1968.
[13] Shafer, A Mathematical Theory of Evidence. Princeton University Press, 1976.
[14] Shafer and J. Pearl, Readings in Uncertain Reasoning. Morgan Kaufmann, 1990.
[15] J.-B. Yang, "Rule and utility-based evidential reasoning approach for multiattribute decision analysis under uncertainties," Eur. J. Oper. Res., vol. 131, no. 1, pp. 31–61, 2001.
[16] J.-B. Yang and D.-L. Xu, "On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty," IEEE Trans. Syst., Man, Cybern. A, vol. 32, no. 3, pp. 289–304, 2002.
[17] D.-L. Xu, J.-B. Yang and Y.-M. Wang, "The evidential reasoning approach for multi-attribute decision analysis under interval uncertainty," Eur. J. Oper. Res., vol. 174, no. 3, pp. 1914–1943, 2006.
[18] J.-B. Yang, Y.-M. Wang, D.-L. Xu and K.-S. Chin, "The evidential reasoning approach for MADA under both probabilistic and fuzzy uncertainties," Eur. J. Oper. Res., vol. 171, no. 1, pp. 309–343, 2006.
[19] J.-B. Yang and D.-L. Xu, "Evidential reasoning rule for evidence combination," Artificial Intelligence, vol. 205, pp. 1–29, 2013.
[20] J.-B. Yang and M. G. Singh, "An evidential reasoning approach for multiple-attribute decision making with uncertainty," IEEE Trans. Syst., Man, Cybern., vol. 24, no. 1, pp. 1–18, 1994.
[21] D.-L. Xu, "An introduction and survey of the evidential reasoning approach for multiple criteria decision analysis," Annals of Operations Research, vol. 195, no. 1, pp. 163–187, 2012.
[22] J.-B. Yang and D.-L. Xu, "A study on generalising Bayesian inference to evidential reasoning," in International Conference on Belief Functions. Springer, 2014, pp. 180–189.
[23] X. Xu, J. Zheng, J.-B. Yang, D.-L. Xu and Y.-W. Chen, "Data classi- fication using evidence reasoning rule," Knowledge-Based Systems, vol. 116, pp. 144–151, 2017.
[24] J.-B. Yang, J. Liu, J. Wang, H.-S. Sii and H.-W. Wang, "Belief rule- base inference methodology using the evidential reasoning approach (RIMER)," IEEE Trans. Syst., Man, Cybern. A, vol. 36, no. 2, pp. 266– 285, 2006.
[25] Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[26] Y. Sahin, S. Bulkan and E. Duman, "A cost-sensitive decision tree approach for fraud detection," Expert Systems with Applications, vol. 40, no. 15, pp. 5916–5923, 2013.
[27] Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," Expert Systems with Applications, vol. 73, pp. 220–239, 2017.
[28] P. Cao, O. Zaiane and D. Zhao, "A measure-optimized cost-sensitive learning framework for imbalanced data classification," in Biologically- Inspired Techniques for Knowledge Discovery and Data Mining. IGI Global, 2014, pp. 48–75.
[29] Bahnsen, A. Stojanovic, D. Aouada and B. Ottersten, "Cost- sensitive credit card fraud detection using Bayes minimum risk," in 2013

12th International Conference on Machine Learning and Applications (ICMLA), vol. 1. IEEE, 2013, pp. 333–338.

[30] C. Bahnsen, D. Aouada and B. Ottersten, “Example-dependent cost- sensitive logistic regression for credit scoring,” in 2014 13th Interna- tional Conference on Machine Learning and Applications (ICMLA). IEEE, 2014, pp. 263–269.

[31] Pazzani, “UCI Machine Learning Repository,” 1991. [Online]. Avail- able: https://archive.ics.uci.edu/ml