A Research on Enterprise Technical Risk Threshold Activation Model Construction in ICV Industry

ZHANG Yue[1](✉), CAO Yue[1] and BAI Chen[1]

[1] Institute of Scientific and Technical Information of China, Fuxing Rd. 15, 100038 Beijing, China
zhy@istic.ac.com

Abstract. Addressing the critical need for enhanced industrial risk monitoring, this research advances the analytical capabilities of management entities and policy advisors in scrutinizing enterprise technological risks in specific sectors. It introduces a machine learning-assisted approach to systematically comprehend the triggers and mitigators of technological risks. The research develops a Machine Learning-based Enterprise Technology Risk Threshold Activation (ETRTA) Model. The model, grounded in a multi-dimensional classification of enterprise risks, is adept at delving into the nuances of these risks in industry-specific contexts. Employing a suite of eight machine learning techniques, including Random Forest, XGBoost, etc. the model trains on various parameters to discern the characteristics of enterprise technological risks. Additionally, automated processes are employed to uncover consistent patterns in the activation of these risks. The efficacy of the model is highlighted by the classification prediction accuracy of three gradient boosting ensemble models, which stands at 82.59%. The accuracy facilitates the identification of enterprises at potential technological risk using extensive datasets. The future scope includes enhancing the prediction precision and robustness of the models and broadening their applicability in assessing enterprise technological risks in diverse industries.

Keywords: Enterprise Technology Risk, Threshold Activation, Classification Prediction, Data Mining, Intelligent and Connected Vehicle (ICV)

1 Introduction

In the current global landscape, the evolving international dynamics have heightened industrial risks for China, especially in critical emerging sectors. These industries confront a range of challenges, encompassing uncertainties in technological research and development, governance of emerging technologies, resource limitations, technological biases and monopolistic behaviors. It's critical to adhere to directives on scientific and technological security and enhance industrial risk monitoring. It requires enhancing the analytical capabilities of administrative organizations and policy researchers, focusing on an exhaustive analysis of industry-specific challenges, vulnerabilities, and pivotal points. Utilizing advanced modeling techniques is crucial for identifying and

C. Bai et al. (eds.), *Proceedings of 2023 China Science and Technology Information Resource Management and Service Annual Conference (COINFO 2023)*, Advances in Economics, Business and Management Research 293,
https://doi.org/10.2991/978-94-6463-498-3_6

preempting risks unique to each industry, thus improving industrial risk foresight and ensuring industrial security.

The concept of risk threshold activation is key within the risk management framework, playing a significant role in strategic foresight, effective risk management and proactive approaches. The research introduces a machine learning framework designed to systematically identify patterns of technology risk activation or mitigation within specific industries. The aim is to expand research perspectives, equip industries to proactively recognize potential threats, strengthen risk resilience, and provide insights for policymaking in industrial regulation and strategic resource allocation in the dynamic industrial environment.

2 Related Work

2.1 Literature review

The existing body of literature on industrial technological innovation risk, as explored by both domestic and international scholars, can be broadly categorized into three distinct types.

The first category comprises studies focusing on the interpretative analysis of specific risks and the application of corresponding risk response strategies. This category primarily deals with a detailed classification and understanding of various risks associated with enterprise technological innovation, including market risk, technology implementation risk, intellectual property infringement risk, and risks in collaborative projects. For instance, Song et al.[1] emphasized that managing these risks is critical for the success of a firm's innovation process, encompassing steps like environmental information gathering, risk identification, assessment, and decision-making. Effective management in this context is essential for mitigating uncertainty and fostering successful innovation implementation. Hao [2] delved into Motorola's Iridium system project to discuss innovation challenges and factors leading to innovation failure, re-evaluating technological innovation risks from a technological philosophy standpoint. Zhu et al.[3] conducted an analysis of the cotton textile industry's internal and external environments, identifying key sustainable development constraints and suggesting strategies for optimizing industry layout, enhancing crisis risk measures, and establishing industry risk early warning systems.

The second category of literature, rooted in risk management framework principles, includes research employing various indicator systems, wind control models, and scenario analysis techniques to explore risk propagation and diffusion concepts. Lei [4] focused on the economic risk diffusion model within the complex network of the industrial economy, elucidating the dynamic nature and pathways of risk diffusion across industries. Yang [5] constructed a risk propagation model for industrial symbiotic networks, incorporating both anticipatory and reactive strategies. The model examines how organizational behavior influences risk propagation in these networks, both before and after risk materialization in neighboring entities. Lin [6] utilized a risk scenario identification approach to dissect the risks in China's burgeoning new energy automotive industry, viewing these risks from their source and employing the FIR (factor-interaction behavior-risk) model. This model aids in understanding the interplay among various risk factors and constructing a multi-dimensional risk landscape for the industry. Additionally, Wang et al. [7] employed VAR and CD models, grounded in extreme value theory, to assess risk levels in banking and real estate sectors. Their research extended to studying how risks propagate within and beyond these sectors. Collectively, these studies offer valuable methodological insights for managing risks in diverse industries.

The third category centers on risk threshold research in risk management, utilizing algorithmic models and other methodologies to define risk thresholds and their determination techniques. The inception in the domestic sphere dates back to around 2010. Xia [8] underscored the importance of risk thresholds, emphasizing their impact on risk transmission, extending beyond the risk source, carrier, flow, characterization, and transmission path. Yang [9] developed the risk degree vector method, which quantitatively evaluates the potential risk associated with each role and task, enhancing the management and control of access rights. The method also includes a comprehensive assessment, using fuzzy logic to evaluate aggregated risks of user groups performing multiple tasks, taking into account various contributing factors. Yang [10] conducted an in-depth analysis of China's financial risk early warning indicators, employing a signaling methodology. The approach involved scrutinizing monthly and quarterly indicators to establish their respective thresholds. Lastly, Hu [11] provided a systematic review of the methods used to determine indicator thresholds in financial risk early warnings, including comparative, fluctuation, and expert solicitation techniques.

2.2 Summary

The prevailing research on gauging enterprise technology risk predominantly anchors on industrial economic theories and risk management principles. Yet, there's a marked deficiency in harnessing algorithmic models to scrutinize industrial growth through a risk management prism. This void underscores a nascent comprehension of the mechanisms triggering risk occurrences and the prerequisites for risk threshold activation.

Primarily, the nature of industrial risks, characterized by their latent development and gradual emergence, is not thoroughly grasped in the existing frameworks of enterprise technology risk identification and monitoring within China. Particularly in the area of early risk warning, which often zeroes in on specific sectors, the understanding of the intrinsic properties and developmental patterns of industrial risks remains superficial. As a result, the methodological models designed for these investigations may only offer limited applicability across diverse industrial domains.

Furthermore, domestic studies on "risk identification" predominantly focus on macro-level descriptive analyses and quantitative assessments. These investigations typically involve constructing a risk factor indicator system, assigning weights through expert judgment or predetermined criteria, and computing aggregate scores. However, this approach has not yet evolved into a standardized framework and grapples with challenges like the high subjectivity in quantifying index weights, data acquisition hurdles, and a deficit in anticipatory risk forecasting.

With the escalating imperative for individuals and organizations to proactively and systematically comprehend risks and their dynamic shifts, the urgency to refine and intensify research into enterprise technology risk thresholds is becoming more pronounced. This refinement should pivot towards devising more objective, data-centric, and predictive methodologies that can more accurately capture the subtleties and dynamics of enterprise technology risks.

3 ETRTA model design

To augment the efficacy, expertise, and proficiency in managing multifaceted, decentralized big data in enterprise technology risk identification, the research redefines the conventional risk identification task as a predictive challenge employing machine learning for binary classification. The applied methodology encompasses several pivotal phases.

Feature Engineering. This foundational stage entails the selection and transformation of input features in a dataset with labeled training samples. The goal is to reorganize the data to enhance its interpretability and utility for the model.

Supervised Learning Algorithms. Employed for the classification task, these algorithms use the training data to learn the model's parameters or decision rules. This learning phase is essential for the model's precision in classifying and forecasting based on the provided data.

Optimizing the Objective Function. The model's training involves refining an objective function. This step is critical to align the model closely with the training data, enabling comprehensive learning from established input-output correlations.

Category Determination and Prediction. Post-training, the model gains the ability to categorize input data or predict target values based on the acquired rules. It facilitates the prediction of whether an unexamined enterprise carries a technology risk.

By amalgamating these stages, this methodology provides enterprises with advanced, intelligent tools for risk detection and forecasting. It supports efficient, scientific, and adaptable decision-making in risk management, empowering enterprises to proactively confront and manage technological risks.

3.1 Research process design

The goal is to establish an Enterprise Technical Risk Threshold Activation Model (ETRTA), which involves clearly defining and categorizing enterprise risks in multiple dimensions through a machine learning approach. The method is employed to delve deeply into the characteristics of enterprise risks. To construct the ETRTA, we will use eight different algorithms for training parameter variables. The training is pivotal for learning the nuances of enterprise technical risk and identifying the ideal risk activation threshold. Additionally, the model's effectiveness will be thoroughly assessed to ensure its accuracy and applicability in practical scenarios.

The process for activating the enterprise technology risk threshold using machine learning is illustrated in Figure 1. This figure outlines the key steps and methodologies involved in the machine learning-based approach for identifying and managing enterprise technology risks.

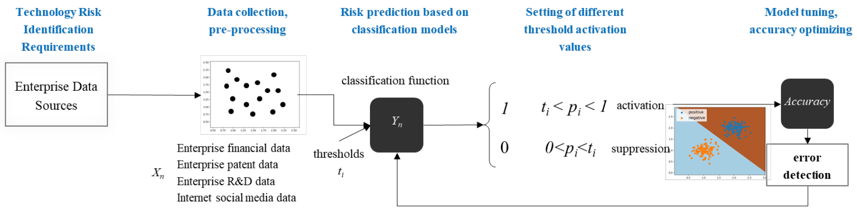

Fig. 1 Research process and main steps

To optimize the model's efficiency in evaluating technology risks, it is crucial to define the specific risk type under investigation and extract relevant data from corporate sources. This process involves using advanced network collection tools to compile comprehensive open-source data. Ensuring data quality and relevance involves several preprocessing steps, such as data cleaning, feature selection, standardization, normalization, dataset segmentation, and addressing imbalances, thus preparing the data for model input.

The modeling and iterative training phase involves a thorough exploration of integrated learning methods like Random Forest and Gradient Boosting Tree, as well as other machine learning models. This exploration is geared towards enhancing the model's performance and robustness by combining multiple 'weak learners'. A meticulous division of the dataset and repeated training and testing phases are employed to evaluate the model's generalization capabilities and minimize the risk of overfitting. Model tuning, especially adjusting the threshold activation parameters, forms a crucial component of this phase. It includes fine-tuning model configuration parameters such as risk thresholds to optimize overall performance.

The final stage of the study focuses on model evaluation and optimization. After training, the model's effectiveness is rigorously assessed using a range of evaluation metrics, ensuring its high quality and practical applicability.

3.2 Data Dimensions

Enterprise technology risk is characterized as the potential threats and uncertainties encountered in activities like technological innovation, new technology application, research and development, impacting an enterprise's technological advancement, assets, and market position. The risk stems from various sources: internal ones include the technology's evolution, complex nonlinearities, and dependencies, while external factors comprise market uncertainties, competitive pressures, regulatory changes, intellectual property issues, supply chain disruptions, and security concerns. In the technological innovation process, challenges like the immaturity of advanced technologies, insufficient reliability verification, complex risk causes, and lack of alternatives can lead to innovation failure. Additionally, gauging a technology's market adaptability and sophistication is challenging, making it tough to gather and effectively measure relevant risk data. Hence, this research focuses on external risk factors, excluding risks inherent to the technology itself.

Table 1 presents descriptors of firm data related to external technology risk, some of which directly indicate the intensity of a firm's financial investment in technology development. These include metrics such as R&D budgets, staffing levels, and progress in R&D projects. Intellectual property metrics, like patents, trademarks and copyrights, serve as common indicators of a firm's capacity for technological innovation and its ability to protect intellectual property. However, acquiring data on skilled personnel is more challenging. This involves gathering information about their technical expertise, competencies, and involvement in technological innovation projects. Furthermore, data indicators like project success rates, product development cycles, and technology demand tend to depend more on subjective evaluations and expert insights. This reliance on subjective assessments can make it difficult to accurately reflect overall trends in the data.

Table 1 Descriptors related to technology risk in common enterprise data

Description item classification	Data indicators
Based on in-house panel statistics on technology, production, finance, markets, etc.[12]	financial statements, R&D inputs, labor productivity, patent outputs[13][14] , product sales [1] (profit tax rate, market share, etc.), net present value ratio, internal rate of return, proportion of R&D expenditures to corporate sales, proportion of technological developers, amount of venture capital obtained [15] , etc.
Unstructured information based on heterogeneous text from multiple sources[1]	enterprise annual reports16, prospectuses[17], interim announcements, project reports, media reports, product reviews on business conditions, developments, R&D projects[1] , etc.
Analyzing predictive data based on expert knowledge, questionnaires[18]	project success rate, patent acquisition, product development cycle, development capital recovery cycle[15], technology demand, technology performance, technology improvement, technology migration, proportion of proprietary technology in product development, industry technology progress rate, etc.

After an exhaustive evaluation of primary data sources for pinpointing enterprise technology innovation risks, considering aspects such as the research area's features, data representativeness, accessibility, and the differentiation in the model's dimensions, the identification of technology risks is confined to external data. This encompasses information on the enterprise's operational status, R&D endeavors, and the procurement of intellectual property rights, which are more effective for forecasting potential technological risks.

For instance, financial health is intricately tied to technological investments. A company's financial stability often has a direct bearing on its technology investments. Sufficient financial resources allow for substantial R&D and innovation investments, thereby reducing technological risk. Conversely, financial difficulties can limit R&D expenditure, increasing technological risk. The significance of R&D activities is paramount. High levels of R&D activity are typically linked with elevated technological risk, as innovation efforts and the deployment of novel technologies carry the potential for failure. In competitive, technology-driven industries, companies often embrace the risks associated with technological uncertainty to gain a competitive edge. Moreover, initiatives like technology collaboration, patent licensing, and transfer are vital indicators of a company's technological innovation capacity. Possessing key patents denotes a company's prowess in technology and its ability to establish market barriers. The efficacy and popularity of patents can also affect a company's competitive position. On the flip side, a rival with a strong patent portfolio can limit a company's technological competitiveness, leading to challenges such as hefty royalty fees, market entry risks, and the possibility of technological obsolescence.

4 Empirical Study

4.1 Data collection and processing

The ICV sector, marking a new era in the automotive industry, features automobiles equipped with sophisticated sensors, controllers, actuators, and other devices. These vehicles, by integrating modern communication and network technologies, facilitate intelligent information exchange and sharing with various entities such as other vehicles, roads, people, and the cloud. The industry is characterized by advanced capabilities like complex environment perception, intelligent decision-making, and cooperative control, bridging key sectors like automobile manufacturing, telecommunications, and the internet, thus sparking a new wave of international scientific and technological competition. Analyzing the technical risks in this sector offers crucial insights for shaping China's automobile technology strategy, as well as the transformation of related industries and the value chain system.

In this research, we selected a sample of 4,238 enterprises (both listed and unlisted) from China's ICV industry, including 176 enterprises identified with technology risks. The data, spanning from December 31, 2017, to December 31, 2022, comprises corporate financial reports, R&D activities, patent information, and technology risk-related news reports. The first three data samples were sourced from the CSMAR China Economic and Financial Research Database and the Datago Finance and Economics Research Database. The fourth sample, pertaining to online big data collection, was uniformly integrated to create the enterprise risk dataset. Python 3.9.7 was utilized as the primary software tool for data processing and analysis.

Dataset description. The dataset compiled for this empirical study encompasses multi-source data, mainly including corporate finance, R&D, and patents. Each of these data elements offers specific insights that are crucial for analyzing the technology risks in the ICV industry. The details of these data elements, including their types, sources, and characteristics, are systematically presented in Table 2. This table provides a comprehensive overview of the dataset, facilitating an understanding of the range and depth of data used in the study.

Table 2 dataset elements

Variant	Data dimensions	Data sources	Data element field
Independent variable	Financial data	Table of major financial	total assets, total liabilities, intangible assets, net income attributable to

X_1 , ..., X_8		indicators of the company	shareholders of the company, net income, operating income, operating costs, operating profit
Independent variable X_9 , ..., X_{15}	R&D Data	Table of R&D investment	number of R&D personnel, number of R&D personnel as a percentage (%), amount of R&D investment, R&D investment as a percentage of operating revenue (%), amount of R&D investment (expenditure) expensed, amount of R&D investment (expenditure) capitalized, capitalized R&D investment (expenditure) as a percentage of R&D investment (%)
Independent variables X_{16} , X_{17}	Patent data	Table of Domestic and Foreign Patent Applications Acquired	number of patents filed, number of patents filed for inventions
Predictor variable Y	Technical risk data	Basic News Information Sheet	headline, full story, any mention of technology risks

Note: Data range is Dec. 31, 2017 to Dec. 31, 2022

In the samples for the empirical study, there are a total of 17 independent variable dimensions encompassing financial indicators, intellectual property rights, and R&D investment data. These dimensions are represented as X_i. The dependent variable, denoted as Y, is derived from mentions of technological risk events concerning the enterprises in news reports. These risk events include issues related to intellectual property rights, technological competition, R&D innovation, and technological substitution risk, among others. This forms the basis for constructing risk classification labels for the production enterprises. To reflect whether an enterprise has encountered risk events in the past, a one-hot variable is set up. This approach allows for a comparison between the model's predictions and the actual performance of the enterprise. Specifically, the setting works as follows: For the probability of risk occurrence calculated by the eight models, a positive sample (i.e., an event recognized or predicted by the model) corresponds to the enterprise experiencing a risk event. In this case, the default Y-value is set to 1. Conversely, if the enterprise does not experience a risk event, the Y-value is set to 0. This binary setup enables an effective way to analyze and predict the occurrence of technological risks in these enterprises based on the model's outputs.

Data pre-processing. To enhance the quality and usability of the data for the empirical study, various pre-processing methods were employed. These included data cleaning, feature screening, and data standardization and normalization. Among the 4,238 samples obtained from publicly available online information, 222 samples had missing values, constituting approximately 5.23% of the total dataset. Given the relatively small proportion of missing values, the decision was made to exclude these samples, resulting in 4,016 valid samples for analysis. Due to varying scales across different dimensional data, *Z-score* normalization was applied to the 17-dimensional X data to ensure consistent scaling across all dimensions. Within the pool of 4,016 valid samples, only 176 were positive samples, representing enterprises facing technological risks. This accounts for just 4.43% of the total, indicating a significant imbalance in the sample distribution. To mitigate potential biases in model training due to this imbalance, data equalization techniques were used. By employing random undersampling of the negative samples 10 times, 10 sample subsets were created. Each subset contained 384 negative samples. These subsets were then paired with the positive sample set, which included 176 records. The combined datasets were then divided into training and test sets in a 6-4 ratio, ensuring a balanced representation of both positive and negative cases in the model training and evaluation process.

Characterization of datasets. To evaluate the interconnections between the assessment variables and establish their statistical significance, the pre-processed data underwent a detailed analysis focusing on correlation and significance. The results of this analysis are illustrated in Figure 2.

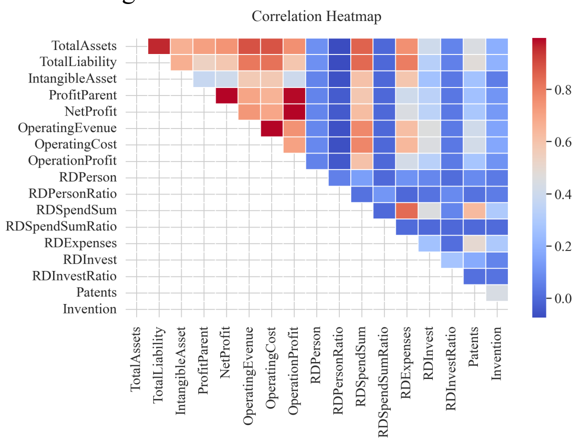

Fig. 2 Correlation and significance analysis (Note: *. Significantly correlated at the 0.05 level (two-sided))

The analysis of the 17 independent variables reveals nuanced relationships. Specifically, indicators like the number of R&D personnel and the capitalization of R&D investment (expenditure) have a correlation coefficient test probability p-value close to 0. This suggests that, at a 0.05 significance level, we should reject the null hypothesis of no correlation, indicating a general linear relationship between these factors. However, the Pearson correlation coefficient for these relationships is below 0.300, signifying that they are not strongly correlated. Conversely, for indicators with higher probability *p*-values, the null hypothesis is accepted at the 0.05 significance level, indicating no overall linear relationship between these variables.

4.2 Modeling

In the modeling phase of this research, several key factors were considered to ensure the effectiveness of the approach. These factors included the size of the dataset, data completeness, statistical properties, and the characteristics of the variables involved. A manual labeling method was employed for the risky training set to facilitate iterative training. Eight specific machine learning models were chosen for this purpose, each selected for their relevance and potential effectiveness in this context. These models are Logistic Regression, Support Vector Machine, K-Nearest Neighbors, Naive Bayes, Random Forest, Gradient Boosting Decision Tree, XGBoost, and LightGBM. To provide a clear understanding of the rationale behind the selection of these models, a comparative analysis of their respective advantages and disadvantages is presented in Table 3.

Table 3 Cross-sectional comparison of machine learning models

Model	Difficulty of model calculation	Strictness in problem type	Applicable data Dimensionality	Sample size limitations	Sensitivity to outliers	Tendency to Overfit
LR	low	linear	low	Small-scale	low	high
SVM	high	nonlinear	high	Large-scale	high	high
KNN	high	nonlinear	high	Large-scale	high	high
NB	low	linear	high	Small-scale	low	low
RF	high	nonlinear	high	Large-scale	low	high
GBDT	high	nonlinear	high	Large-scale	high	high
XGBoost	high	nonlinear	high	Large-scale	low	high
LightGBM	low	linear	high	Small-scale	high	low

All the parameters of each firm were individually used as feature inputs in the model. This was done to establish the functional relationship $Y = f(X_i)$, enabling the prediction of Y based on X_i values. The models were tasked with binary prediction for each set of input data. In total, 80 model training sessions (10 datasets × 8 models) were executed to thoroughly assess the performance of these models across various datasets and contexts. This comprehensive evaluation is crucial for understanding the effectiveness of each model in different scenarios and datasets.

4.3 Model evaluation

This research meticulously evaluates the applicability and effectiveness of the algorithmic model, primarily using metrics like accuracy, precision, recall, *AUC* (Area Under Curve) value in the *ROC* (Receiver Operating Characteristic) curve, and confusion matrix.

An accuracy heat map matrix, derived from the test results of enterprise data over the past decade, illustrates the high accuracy of different models, highlighting their strengths in identifying technology risks in this field. According to the prediction results displayed on the left side of Figure 3, models like Random Forest and three other gradient boosting models demonstrate an accuracy rate above 80% on the validation or test set, indicating robust prediction performance. Notably, a combined model utilizing

LightGBM, GBDT, and XGBoost achieves the highest average model accuracy of 82.59%.

The predictive ability of the eight models is further quantitatively assessed using *ROC* curves, a standard method for evaluating classification accuracy. These curves are constructed by sorting data according to the probability of the model predicting a positive category, and then calculating the True Positive Rate (*TPR*) and False Positive Rate (*FPR*) for each potential threshold. The *ROC* curve is plotted with *FPR* on the horizontal axis and *TPR* on the vertical axis. The *AUC* value, represented by the area enclosed between the *ROC* curve and the 45-degree line, provides insight into the overall classification performance of the model. Observing the *ROC* curve visualizations on the right side of Figure 3 reveals some variation in the models' classification performance. While the *AUC* values of single models differ, six out of the eight algorithms exhibit *AUC* values over 0.7, signifying high accuracy, with the exception of MLP and KNN. The *ROC* curves of the composite models, as depicted in the figure, demonstrate exceptional performance.

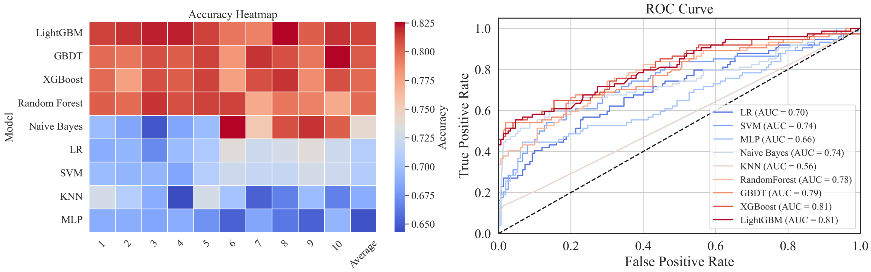

Fig. 3 Model accuracy and ROC curve

Based on the actual prediction results reflected in the confusion matrix presented in Table 4, the precision rate calculation reveals that the model successfully identifies 88.89% of actual genuine cases from all the samples predicted as genuine. However, the model's recall rate is lower, only able to correctly identify 54.05% of genuine cases. This indicates that although the model is quite precise in its predictions, it still misses a significant portion of actual genuine cases. Therefore, there is a notable opportunity for enhancing the model's ability to identify genuine examples, which is essential for reducing the risk underreporting rate. This improvement could lead to a more balanced and effective model, ensuring both high precision and recall.

Table 4 0_9 samples Confusion matrix

sample volume	Actual Values (0)	Actual Values (1)
Predicted Values (0)	145	34
Predicted Values (1)	5	40

In the risk prediction scenario of this research, the threshold activation model operates through classification equations based on a composite model. Each risk indicator's predicted probability is associated with a threshold, and the threshold activation function generates an output when this predicted probability surpasses the threshold. The activation threshold is typically set based on the model's calculation of the predicted probability of a future risk occurrence, denoted as p. Commonly, the baseline level for risk threshold activation, especially in relation to risk volatility, is set at 0.5.In the case of a binary threshold activator, the function outputs a value of 1 when the risk metric's value is greater than the threshold t (p > t), and 0 otherwise. Figure 4 compares the model accuracies at different thresholds horizontally and finds that the model accuracies reach local maximum when the thresholds are set to 0.46 and 0.52. For enterprises that are actually at risk, this research aims to increase the proportion of correct predictions, effectively enhancing the recall rate. Therefore, setting the threshold at 0.46 makes the model more likely to predict a sample as 1 (indicating risk). This setting represents a more stringent approach, where the model leans towards identifying potential risks more readily, thus potentially increasing its sensitivity to actual risk instances.

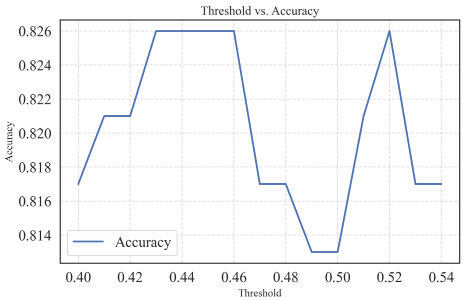

Fig. 4 Activation thresholds of the model at different accuracy rates

4.4 Analysis of model output results

As the automotive industry rapidly evolves towards electrification, connectivity, and intelligence, the ICV industry is at a pivotal stage of technological innovation, industrial transformation, and product upgrading. For some enterprises identified as at-risk and successfully predicted by the model, a technology risk profile encompassing five key dimensions is formed. These dimensions reflect various types of technological challenges faced by the enterprises.

Uncertainty in the Application of Emerging Technologies. ICVs have a wide array of applications in sectors such as governance, public welfare, finance, and intelligent manufacturing. If market demand falls short of expectations, these enterprises might face risks of overcapacity and market miscalculation.

Supply Risks of Core Components and Parts. With the large data processing requirements of ICVs and high demand for sophisticated automotive chips, the market for in-vehicle AI chips is expected to grow continuously. However, the widespread issue of chip shortages globally could lead to reduced production or even shutdowns by manufacturers, directly threatening the production capacity and profit margins of enterprises. In the long run, core technologies across the industry chain, like in-vehicle chips and operating systems, may encounter critical 'bottleneck' technological risks.

Risks of Delayed Technological Updates and Iteration in Response to New Technologies. ICVs integrate various advanced technologies, including sensing, decision-making, communication, and execution. If enterprises fail to timely adjust their product structures or effectively strategize their technology layouts to adapt to or lead technological advancements, they risk falling behind the industry's pace of change, placing them at a competitive disadvantage in technological evolution.

Risks of Technological Innovation and Competition. As downstream industries undergo transformation, many traditional interaction, control, and display technologies have become less applicable. To enhance the technological content of their products, enterprises need to actively strengthen R&D and strategically layout patents.

5 Conclusions

This research developed a semi-automatic, machine-assisted risk identification model that employs automated methods for quantitatively analyzing the regularity of industry risk activation. It identifies enterprises with potential technological risks by analyzing variables from financial, R&D, patent, and news-related data. In the constructed classification model, a combined approach using LightGBM, GBDT, and XGBoost achieves a prediction accuracy of 82.59%. This indicates the model's proficiency in identifying technology risk enterprises, though decision-makers should balance the trade-off between missed reporting and hit rates in practical scenarios. The training variables of the model highlight technical risk factors that could positively or negatively impact enterprise development, showing significant predictive effectiveness. The model supports the identification of risk variables highly correlated with enterprise risks, aiding decision-makers in understanding shifts in business risks within the external environment. A threshold parameter of 0.46 provides a reference for setting reasonable risk thresholds, effectively avoiding issues of under-alerting or false alarms, and is operationally meaningful for efficiently and accurately recognizing weak signals of risk.

Data selection is a pivotal element in the applicability of this research's model. Throughout the data collection phase, a broad spectrum of firms within a specific industry were taken into consideration, encompassing startups, medium-sized, and large enterprises. Consequently, the model's utility is not confined to the ICV industry, the primary focus of this research, but also extends to other strategic, emerging, and technology-intensive sectors such as biotechnology and renewable energy. Given that the ICV market is still developing, the model's applicability might be limited by variations in data distribution due to market dynamics and by disparities in data accessibility and quality in other sectors, potentially stemming from opacity or insufficient market regulation. The model designed in this research is highly versatile and can be tailored to meet the unique requirements of different industries. It allows for customization of risk assessment parameters based on specific risk elements pertinent to each industry, such as incorporating clinical trials in the healthcare sector. Furthermore, the model can be integrated with other analytical techniques like reinforcement learning and intelligent agent simulation, enhancing the precision and efficiency of its predictions. This adaptability renders the ETRTA model not only suitable for current industry research but also capable of evolving to meet the demands of future technology trend analysis and risk management.

Nonetheless, the model's accuracy is subject to certain limitations. It depends on publicly accessible data, which might impact prediction accuracy, particularly in specialized sectors where data availability is limited. The model's current data dimensions may not fully encompass all risk factors, suggesting a need for future research to broaden data sources and indicators. This would enable a more holistic risk assessment, improving the model's fit with real-world influencing factors. Future plans include gathering more varied data samples for in-depth training, which could boost the model's predictive accuracy and consistency. Exploring domain transfer, like adapting training benchmarks for similar fields in enterprise technology risk assessment, such as the fuel cell and hydrogen energy vehicle sectors, offers potential for further enhancing the model's practical utility.

References

1. Song, S., Zhong, Y., Liu, J., et al.: Progress and hot topics of research on corporate technology innovation risk. Technology and Innovation Management 44(5), 525–540 (2023).
2. Hao, L.: Philosophical reflections on the risk of technological innovation–a study of the beginning and end of iridium innovation. Comparative Research on Cultural Innovation 33(3), 31–32 (2019).
3. Zhu, A., Gu, J., Wei, M., et al.: Research on risk response strategy of Xinjiang cotton industry under the perspective of crisis management. China Cotton Processing 234(4), 4–9 (2022).
4. Lei, C.: Research on complex network model of industrial economic risk diffusion based on knowledge meta. Dalian University of Technology, Dalian (2018).
5. Yang, J.: Research on risk propagation mechanisms and mitigation strategies in industrial symbiotic networks. Nanjing University of Information Engineering, Nanjing (2021).
6. Lin, J.: Research on risk identification and evaluation of new energy vehicle industry in China. Harbin Engineering University, Harbin (2015).
7. Wang, S., Xiao, B., Zhou, X.: Measuring risk contagion in China's banking and real estate industries from the perspective of external shocks. Journal of Management 9(7), 968–974 (2012).
8. Xia, Z.: On the transmission mechanism of corporate financial risk. Finance and Economics Monthly 29(7), 19–22 (2009).
9. Yang, Y., Sun, Y., Zhou, L.: FSP:A risk-based security policy. Computer Engineering and Applications 46(13), 82–86 (2010).
10. Yang, X., Xu, C.: Optimal thresholds and forecasting performance analysis of financial risk early warning indicators in China. Journal of Guangdong Institute of Finance 27(2), 121–128 (2012).
11. Hu, L.: Methodology for determining indicator thresholds in risk early warning. Financial Electronic 192(9), 43–45 (2011).
12. Mu, J., Peng, G., Maclachlan, D.: Effect of risk management strategy on NPD performance. Technovation 29(3), 170–180 (2009).
13. Ge, X.: Key element identification in cooperative technological innovation risk on social network analysis. In: Proceedings of the 2014 Seventh International Joint Conference on Computational Sciences and Optimization, pp. 316–319. IEEE Computer Society, Washington (2014).
14. Li, Y., Zhou, S.: Early warning research on enterprise patent risk based on big data perspective. Journal of Intelligence 40(10), 21–26 (2021).
15. Zong, M.: Enterprise technology risk and its analysis. Journal of Kunming University of Science and Technology: Science and Engineering Edition 26(1), 118–122 (2001).
16. Liang, N., Yao, C., Wang, Z., et al.: Research on enterprise risk information extraction method based on triple dimensions. Journal of Intelligence 38(12), 1241–1249 (2019).

17. Rao, X.: Patent risk identification and countermeasures of enterprises to be listed on the science and technology innovation board - an empirical analysis based on 205 prospectuses of enterprises to be listed on the science and technology innovation board. Technology and Law 148(6), 31–40 (2020).
18. Masár, M., Hudáková, M.: Project risk identification of a new product in enterprise in Slovakia. In: CBU International Conference Proceedings 2018, pp. 343–349. CBU Research Institute, Prague (2018).