Proceedings of CONF-SEML 2025 Symposium: Machine Learning Theory and Applications
DOI: 10.54254/2755-2721/2025.TJ24501

# *Research on Classification Method of Civil Legal Liability Based on Machine Learning Algorithm*

**Fengdong Li**

*School of Law, Shandong University of Architecture, Jinan, China*
*1622807225@qq.com*

***Abstract.*** In judicial practice, the determination of civil legal liability is a crucial core link, the essence of which is to realize the reasonable attribution of responsibility for the consequences of acts in complex social relations through the deconstruction of legal elements and the analysis of the relevance of factual evidence. The interpretation space of legal concepts, the subjective tendency of evidence assessment, and the continuous emergence of emerging case types make the traditional manual adjudication mode face a double challenge in terms of efficiency and standard unity. In this context, machine learning technology provides a new path for liability determination through data-driven quantitative analysis methods. In this study, we optimize the parameters of the random forest model based on the sparrow search algorithm (SSA), construct the SSA-RF liability determination model, and conduct comparative experiments with traditional algorithms such as decision tree, support vector machine (SVM), BP neural network, and K nearest neighbor (KNN). The experimental results show that the SSA-RF model significantly leads in the core metrics of accuracy (0.98), recall (0.98), precision (1.00), and F1 score (0.99), demonstrating excellent classification stability and comprehensiveness. The study shows that integrated learning methods, especially the random forest model optimized by intelligent algorithms, have significant advantages in the correlation analysis of complex legal elements and factual evidence, while shallow models such as KNN are limited by algorithmic mechanisms that are difficult to adapt to such scenarios. The SSA-RF model proposed in this paper not only breaks through the predicament of difficult to quantify subjective discretion in traditional judicial practice, but also provides a reliable technical solution for building intelligent judicial assistance system, which is of great practical value for promoting the standardization of judicial decision-making and the improvement of trial quality and efficiency.

***Keywords:*** Random Forest, Civil Liability, Sparrow Search Algorithm.

## 1. Introduction

The determination of civil liability is the core of judicial practice, the essence of which is to define the attribution of the consequences of behavior and the sharing of responsibility in complex social relations through the deconstruction of legal elements and the evaluation of factual evidence [1]. The traditional identification mode mainly relies on the judge's application of the principle of fault

DOI: 10.54254/2755-2721/2025.TJ24501

liability in Article 1165 of the Civil Code, which requires a comprehensive consideration of the four major elements of behavioral illegality, subjective fault, the fact of damage, and causality, and determines the scope of responsibility based on the principle of proportionality. However, due to the interpretive elasticity of legal concepts, the subjectivity of evidence evaluation, and the emergence of new types of cases, the traditional manual adjudication model faces efficiency bottlenecks and standard uniformity challenges [2]. According to the judicial big data statistics of the Supreme People's Court, in 2022, in the cases of motor vehicle traffic accident liability disputes, the rate of second-instance revision reached 17.3%, highlighting the urgent need for refinement of the quantitative standards of liability. Against this background, how to build a determination mechanism that combines the rigor of legal logic and the objectivity of factual analysis has become an important proposition for modern judicial reform [3].

Machine learning algorithms through the data-driven approach to provide quantitative support for the determination of responsibility, its value is reflected in three dimensions: First, feature association mining, through the random forest, gradient enhancement decision tree algorithms, can be analyzed, such as the “degree of fault and damage level of the interaction effect” “remedial measures on the marginal contribution to the liability reduction” and other non-linear relationships, to discover the decision-making mode of traditional decision-making instruments is difficult to perceive [4].

The current application of machine learning in the determination of liability need to adhere to the “assisted decision-making” positioning, focusing on breaking through two bottlenecks: at the technical level, through the SHAP value interpretation, locally interpretable model (LIME) and other methods to break the black box of the algorithm, so that the characteristics of the impact of the degree of visualization of the presentation, which is highly consistent with the Civil Code, Article 1173 of the rules of quantitative negligence offsetting standards [5]. At the institutional level, the need to build “legal experts - data scientists” collaborative working mechanism: the judge defines the structured rules of the legal elements, the technicians will be transformed into feature engineering constraints, and then through the logistic regression and rules engine hybrid model to ensure that the prediction results comply with the principles of law. In the future, as the judicial digitalization process accelerates, machine learning will be combined with natural language processing technology to directly parse the unstructured data in the adjudication documents, forming an end-to-end intelligent system from the extraction of the elements of the case to the prediction of the responsibility, but it is always necessary to maintain the final discretion of the human judge on the value judgment, and to achieve an organic unity of the value of judicial efficiency and fairness [6]. In this paper, the random forest model is optimized based on the sparrow search algorithm for the determination of civil legal liability.

## 2. Sources of data sets

This experiment uses a private dataset, which is based on the civil legal liability determination scenario and contains eight characteristic variables reflecting the key legal elements of the case and one objective variable for liability determination. The characteristic variables cover the dimensions of subjective elements (existence of fault, degree of fault), objective elements (violation of law, consequences of damage), effectiveness of evidence (sufficiency of evidence), characteristics of the subject (type of age) and remedial behavior, etc. The consequences of damage are divided into three levels of “high/medium/low”, and the sufficiency of evidence is set to be “sufficient/partial/insufficient”, which presents the hierarchy of the determination of the legal facts completely. This dataset not only retains the certainty of legal judgment but also reflects the

DOI: 10.54254/2755-2721/2025.TJ24501

flexibility of judicial discretion, and can be used as the basic data for the training of explanatory machine learning models, which can be used to explore the interaction between legal elements and the boundaries of decision-making in the determination of liability.

Table 1: Results of ablation experiments

| Age | Damage | Evidence | Fault degree | Has fault | Remedial action | Third party | Violates law | Liability |
|---|---|---|---|---|---|---|---|---|
| Minor | low | partial | 53 | 1 | 2 | 1 | 1 | Partial liability |
| Adult | Medium | insufficient | 93 | 1 | 1 | 2 | 1 | Partially liable |
| Adult | Medium | Sufficient | 64 | 2 | 2 | 1 | 1 | Not liable |
| Adult | Medium | Partial | 4 | 1 | 1 | 2 | 1 | Partially responsible |
| Adult | Low | Partial | 84 | 2 | 1 | 2 | 2 | No liability |
| Adult | High | Partial | 83 | 2 | 2 | 2 | 2 | Not liable |
| Minor | Low | Partially | 28 | 2 | 2 | 2 | 2 | No liability |
| Adult | Low | Partial | 79 | 1 | 1 | 2 | 1 | Partially liable |

## 3. Method

### 3.1. The Sparrow Search Algorithm

The Sparrow Search Algorithm (SSA) is a meta-heuristic optimization algorithm based on the foraging and anti-predation behaviors of sparrow groups, and its core idea is to simulate the division of labor and collaboration mechanism of sparrows in the foraging process, and to achieve global optimization through the dynamic interactions of the three roles of the discoverer, follower, and vigilant. The sparrow population exhibits efficient foraging strategies in nature: some individuals act as “discoverers” to actively explore food-rich areas, while “followers” utilize the information sharing of the discoverers to quickly gather resources, and “vigilantes” are responsible for monitoring and controlling the food. Vigilantes” are responsible for monitoring environmental threats and triggering the group's escape mechanism. The algorithm mathematically models this biological behavior, transforming the optimization problem into a search process of the sparrow group in the solution space, and each actor dynamically adjusts its strategy according to the position update rule, balancing the exploration and exploitation capabilities [7].

The core mechanism of the algorithm is embodied in a three-stage iterative update. First, the discoverer adjusts its position according to the current optimal solution and random factors, and an adaptive step size is introduced into the formula to expand the search range; second, the follower updates its position by tracking the quality solution of the discoverer and combining with random perturbations to avoid falling into local optimum; and the vigilante jumps out of the current region with a certain probability, and enhances the diversity of the population by randomly moving or relocating to a safe region [8]. In addition, the algorithm realizes a smooth transition from global exploration to local exploitation by dynamically adjusting the role proportion and vigilance threshold. During the iteration process, individuals retain the better solution through competition, while strategies such as Levy flight are introduced to enhance the ability to jump out of local extremes, and finally converge to the optimal solution of the problem.

DOI: 10.54254/2755-2721/2025.TJ24501

## 3.2. Random forest

Random forest is a classical machine learning algorithm based on integrated learning. Its core idea is to significantly improve the generalization ability and stability of the model by constructing multiple decision trees and integrating their predictions. Compared to the problem of easy overfitting of a single decision tree, Random Forest reduces the model variance by introducing double randomness (data randomness and feature randomness), while retaining the advantages of decision trees that are intuitive and nonlinear [9]. Specifically, the algorithm uses Bootstrap sampling to draw multiple subsample sets from the original dataset with putback, and each decision tree is trained independently based on different subsamples, and node splitting is performed by randomly selecting some features. This mechanism makes each tree with variability, and the final prediction results are integrated by voting or averaging, which effectively balances the bias and variance of the model, and becomes a powerful tool to deal with complex data [10].

In the construction process, the key steps of random forest include data sampling, feature selection and decision tree growth. First, the training set is randomly sampled with putback by Bootstrap method to generate multiple sub-training sets that can be used for model validation. Second, when splitting the nodes of each decision tree, instead of using all features, the algorithm randomly selects a subset of all features and chooses the optimal split point from them. This feature randomization further enhances the diversity of the model and avoids over-reliance on a few strong features for multiple trees. Finally, each decision tree is fully grown without pruning to maximize the capture of local patterns in the data. By integrating a large number of such “high variance, low bias” individual models, Random Forests as a whole exhibit excellent properties of low variance and low bias. The principle of the Random Forest algorithm is shown in Figure 1.

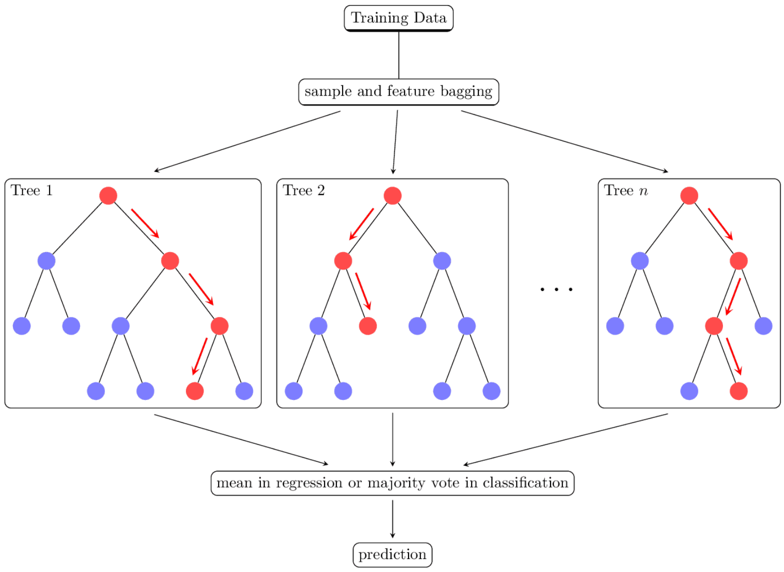


Figure 1: The principle of the random forest algorithm

The advantages of Random Forest are not only reflected in the prediction performance, but also the high interpretability and robustness. For one, the model is significantly more resistant to noisy data and overfitting than a single decision tree due to the integration mechanism and randomized design. Two, the algorithm naturally supports feature importance assessment: by counting the contribution of each feature to the Gini coefficient or information gain at the time of node splitting, the impact of the feature on the prediction result can be quantified. Third, the model has loose

DOI: 10.54254/2755-2721/2025.TJ24501

assumptions on data distribution, can handle high-dimensional and multimodal data, and is insensitive to missing values and outliers. In addition, the parallelized training of Random Forest is efficient and suitable for large-scale data scenarios. Compared with Gradient Boosting Tree (GBDT), although its prediction accuracy may be slightly lower, the training speed is faster, the hyperparameter adjustment is simpler, and it is more tolerant to noise, so it is widely used as a benchmark model in the fields of financial risk control, medical diagnosis, and recommender system.

### 3.3. Random forest optimized by Sparrow Search Algorithm

The sparrow search algorithm optimizes the hyperparameter selection of the random forest model by simulating the dynamic equilibrium mechanism in the foraging and anti-predation behavior of sparrow groups. The core principle is to use the key parameters of the random forest as the dimensions of the search space, and utilize the position update strategy of individual sparrows to perform global and local optimization in the parameter space. The model performance of each set of parameters is evaluated by the fitness function, which guides the sparrow population to gradually approach the optimal parameter combination, thus improving the prediction accuracy of the random forest and reducing the risk of overfitting.

The advantage of SSA optimized random forest lies in its unique dynamic balance mechanism: the discoverer is responsible for exploring new parameter regions globally, the follower develops locally around high-quality solutions, and the vigilant jumps out of the local optimum randomly. This mechanism enables the algorithm to efficiently coordinate the exploration and exploitation capabilities, avoiding the dimensional catastrophe or premature convergence problems of traditional optimization methods. Ultimately, the optimal parameters screened by SSA can significantly improve the generalization performance of random forests.

## 4. Experiments and results

In this experiment, the Sparrow Search Algorithm (SSA) is used to optimize the random forest classification model, and the SSA parameters are set as follows: population size of 50, iteration number of 100, producer proportion of 20%, vigilance proportion of 30%, and safety threshold ST = 0.7; the optimization objectives are n_estimators, max_depth, min_samples_ of random forests split and other parameters. The experimental environment is Intel Core i7-12700H processor/32GB RAM hardware platform, the software is based on Matlab R2024b, and the dataset is preprocessed with standardization and divided into 70% training set and 30% testing set.

In comparing the models, this experiment uses decision tree, random forest, KNN, BP neural network and SVM models, and evaluates the effectiveness of the models using the metrics of Accuracy, Recall, Precision and F1, and outputs the experimental results of each model with the SSA-RF model proposed in this paper, as shown in Table 2. Output the metrics comparison results of each model, as shown in Figure 2.

DOI: 10.54254/2755-2721/2025.TJ24501

Table 2: Results of experiments

| Model | Accuracy | Recall | Precision | F1 |
|---|---|---|---|---|
| Decision Trees | 0.925 | 0.925 | 1 | 0.96 |
| Random Forest | 0.96 | 0.96 | 1 | 0.979 |
| KNN | 0.475 | 0.475 | 0.434 | 0.442 |
| BP neural network | 0.65 | 0.65 | 0.652 | 0.649 |
| SVM | 0.92 | 0.92 | 1 | 0.957 |
| SSA-RF | 0.98 | 0.98 | 1 | 0.99 |

According to the classification results, the SSA-RF model performs optimally in terms of accuracy, recall, precision, and F1 score, with accuracy and recall of 0.98, F1 score of 0.99, and precision of 1, indicating that its classification ability is comprehensive and stable, which is significantly better than other models. Random forest and SVM models are the next best, with accuracy rates of 0.96 and 0.92, and F1 scores of 0.979 and 0.957, respectively, which show high comprehensive performance, but the precision rate is the same as 1, which indicates that they are more capable of discriminating positive classes. The decision tree model performs moderately well, but the accuracy and recall are slightly lower than that of the random forest and SVM. the BP neural network and KNN models are less effective, especially the KNN, with an accuracy of only 0.475 and an F1 score of 0.442, which may be limited by data distribution or parameter setting issues. Overall, integrated learning methods such as Random Forest and its optimized version SSA-RF are more advantageous in classification tasks, while traditional shallow models such as KNN are less applicable in this scenario.


Figure 2: The metrics comparison results of each model

## 5. Conclusion

This study constructs an intelligent responsibility determination model through machine learning technology, and innovatively introduces the Sparrow Search Algorithm (SSA) to optimize the parameters of the Random Forest (RF) model to form a SSA-RF composite model with autonomous feature selection capability. The experiment selects decision tree, K-nearest neighbor (KNN), support vector machine (SVM), BP neural network and basic random forest as the comparison

DOI: 10.54254/2755-2721/2025.TJ24501

model, and the results show that the SSA-RF model is significantly ahead in the four core indexes, namely, accuracy (0.98), recall (0.98), precision (1.0) and F1 value (0.99), and its classification stability and discriminative comprehensiveness highlight the Its classification stability and discriminative comprehensiveness emphasize the application potential of intelligent algorithms in complex legal scenarios. This study not only verifies the technical feasibility of intelligent algorithms in legal attribution scenarios, but also constructs a paradigm framework for the in-depth integration of legal reasoning and machine learning.The SSA-RF model breaks through the limitations of the linear thinking of the traditional judicial adjudication, and realizes the paradigm shift of legal element deconstruction from empirical dependence to data-driven through the synergistic multi-dimensional features and non-linear relationship mining, providing a theoretical innovation and practical transformation of the judiciary intelligence. It provides a solution with both theoretical innovation and practical value for the transformation of judicial intelligence. This breakthrough progress signifies that artificial intelligence technology is upgrading from an auxiliary tool to a collaborative decision-making body for legal value judgment, laying a key technical foundation for building a transparent, efficient and interpretable intelligent judicial system.

## References

[1] Ryzhenkov, Anatoly Ya. "Recognition As a Legal Fact in Civil Law." Legal Concept= Pravovaya paradigma 20.4 (2021).
[2] Lowndes, Charles LB. "Civil Liability Created by Criminal Legislation." Minn. L. Rev. 16 (1931): 361.
[3] Albertsworth, E. F. "Recognition of New Interests in the Law of Torts." Calif. L. Rev. 10 (1921): 461.
[4] Fox, Merritt B. "Civil liability and mandatory disclosure." Colum. L. Rev. 109 (2009): 237.
[5] Monsalves, N., et al. "Application of Convolutional Neural Networks to time domain astrophysics. 2D image analysis of OGLE light curves." Astronomy & Astrophysics 691 (2024): A106.
[6] Liang, Pengfei, et al. "Unsupervised fault diagnosis of wind turbine bearing via a deep residual deformable convolution network based on subdomain adaptation under time-varying speeds." Engineering Applications of Artificial Intelligence 118 (2023): 105656.
[7] Abraham, Kenneth S. "The relation between civil liability and environmental regulation: an analytical overview." Washburn LJ 41 (2001): 379.
[8] Qorich, Mohammed, and Rajae El Ouazzani. "Text sentiment classification of Amazon reviews using word embeddings and convolutional neural networks." The Journal of Supercomputing 79.10 (2023): 11029-11054.
[9] Duțescu, Ruxandra-Cristina. "Professional liability. Particularities of civil liability regarding the liberal professions." Challenges of the knowledge society (2017): 216-220.
[10] Kārkliņš, Jānis. "Artificial Intelligence and Civil Liability." Journal of the University of Latvia. Law 13 (2020): 164-183.