https://doi.org/10.31449/inf.v49i34.9232

Hybrid Churn Prediction Model Using SMOTE-SVM Resampling and Genetic Algorithm Optimized ELM

Xuan Chen
Xinyang University, Xinyang 464000, China
Email: apple82cx@163.com

Keywords: SVM, SMOTE, consumer, churn prediction, GA, ELM

Received: May 14, 2025

Consumer churn prediction and early warning have emerged as a major area of study in enterprise customer relationship management due to the advancement of big data technologies and intelligent decision-making systems. To improve the accuracy and stability of the churn prediction model, a novel resampling algorithm based on support vector machines and the synthetic minority class oversampling technique was designed. This technique alleviates the category imbalance problem by extracting support vector boundary samples and using the synthetic minority class oversampling technique to interpolate new samples in their neighborhoods. Next, a genetic algorithm is introduced to optimize the input weights and hidden layer bias of the extreme learning machine globally. Then, a classifier is constructed to improve the model's generalization performance and convergence stability. Finally, the resampling algorithm is integrated with the classifier to construct a complete consumer churn prediction model. The suggested churn prediction model had the highest mean average precision and F1 value of 98.5% and 0.98 on the training set, according to the performance test results. The lowest mean square error was 0.025 and 0.028 for the training set and the test set, respectively. The results of practical application tests indicated that the prediction accuracy of the proposed prediction model in five typical datasets was up to 93.17%, and the shortest average prediction time was only 0.83 seconds. In summary, the suggested methodology can successfully raise enterprise churn customer identification's accuracy and real-time performance. It provides powerful support for intelligent customer management and precise marketing decision-making in practical application scenarios.

Povzetek: Hibridni model SSGE-CCPM združuje SMOTE-SVM vzorčenje in GA-optimiziran ELM in omogoča bolj kvalitetno napoved odtujitve strank, saj presega modele ELM, HGWO-ELM in DE-RS-ELM po točnosti, stabilnosti in hitrosti.

1 Introduction

Businesses that compete in the market frequently deal with consumer churn (CC), and acquiring new consumers is far more expensive than keeping hold of current ones. Therefore, enhancing business operational efficiency and refining customer management strategies greatly depend on precise forecasting and prompt warning of probable churn customers. [1]. Currently, traditional churn prediction methods rely primarily on rule-based statistical analysis and traditional machine learning models, such as logistic regression, decision trees, and support vector machines (SVMs). Although these methods have the advantages of simple design and fast computation, they still face two key challenges in practical applications: First, the dataset has a serious category imbalance problem. This leads to the model's inability to accurately identify a small number of churned customers. Second, the classification performance and convergence stability of traditional models are limited in the face of high-dimensional heterogeneous data [2]. Data imbalance processing and classifier performance optimization are currently the focus of extensive research by numerous academics both domestically and internationally. Gurung N et al. conducted an in-depth study on the CCP problem. This study applied the random forest and decision tree algorithms to construct a consumer churn prediction (CCP) model, and systematically explored the key factors affecting the probability of churn. The findings illustrated that the classification accuracy of the model reached up to 96.25%, which provided important data support for the subsequent development of customer retention strategies in government enterprises [3]. Wu K H et al. conducted a study on CC in telecommunication industry by extracting topic-sentiment information using document-based heuristic phrase rule algorithm approach. Moreover, the sentiment score was calculated to reveal the characteristics and portraits of potential churn customers. Finally, the mined churn tendency features were compared and verified with the actual churn customer data of the enterprise [4]. Toor A A et al. proposed an optimized bilateral cumulative detection method for the conceptual drift and category imbalance problems prevalent in telecom CCP. The method introduced a dynamic adjustment mechanism for the sliding window error rate, aiming to improve the robustness of churn prediction in unbalanced data environments. The findings indicated that the approach performed well in terms of assessment efficiency and prediction accuracy (PA), offering new methodological support for the astute detection and prompt action of churn clients [5].

Meanwhile, in recent years, extreme learning machine (ELM) has received more and more attention from scholars by virtue of its advantages such as fast training speed and simple structure. Moreover, it is applied in the fields of CC, financial risk assessment, and so on. Lalwani P et al. proposed a prediction method consisting of six stages for the problem of CCP in telecommunication industry. To increase the model accuracy, the study first preprocessed and analyzed the data, then used the gravitational search algorithm to choose features, and lastly added boosting and integrated learning approaches [6]. Matuszelański K et al. suggested a method based on the fusion of data from multiple sources for the problem of CCP in the field of e-commerce. The study combined multi-class statistical data with feature extraction and spatial clustering analysis using latent Dirichlet allocation and density-based spatial clustering algorithms. According to experimental findings, the approach produced favorable outcomes for both churn feature resolution and model performance [7]. Singh P P et al. addressed the problem of CC in banks by modeling and analyzing customer data using various ML algorithms to predict potential churn users and comparing the model performance with different evaluation metrics. Moreover, the study developed an RShiny-based data visualization application, which was used to assist bank management in CC trend monitoring and decision support. The outcomes revealed that the method helped to identify high-risk customers in a timely manner and improve customer retention [8]. A summary comparison of the methods is shown in Table 1.

Table1: Summary comparison of various methods

Authors	Core method	Dataset used	Accuracy (%)	F1 value	Runtime
Gurung N et al. [3]	Random forest+decision tree	UCI ML customer data	96.25	Not reported	Not reported
Wu K H et al. [4]	Heuristic phrase rules+sentiment scoring	Telecom social media data	Not specified	Not reported	Not reported
Toor A A et al. [5]	Sliding window dynamic error detection	Telecom imbalance data stream	Good (not quantified)	Not reported	High (sliding window)
Lalwani P et al. [6]	Elm+gravitational search+ensemble learning	Telecom customer data	81.71	Not reported	Not reported
Matuszelański K et al. [7]	Multi-source fusion+gbdt+logistic regression	E-commerce user data	Good (not quantified)	Not reported	Not reported
Singh P P et al. [8]	Multi-model learning+rshiny visualization	Banking customer data	Good (multi-model comparison)	Not reported	Not reported

In summary, existing CCP methods have made some progress in both sampling strategy optimization and classifier performance enhancement, and the following gaps still exist in current research. On the one hand, most methods may introduce noisy samples when dealing with extreme category imbalances. This can affect model training. On the other hand, although some optimization strategies (e.g., differential evolution) improve classification, they have slow convergence and long prediction time. This limits the real-time application of the model in real business scenarios. Therefore, the study aims to address the following key issues. First, it examines whether the model can effectively increase the number of samples from the minority class without introducing too much noise. This improves the model's ability to recognize churned customers. Second, the study will examine whether the genetic algorithm (GA) can effectively optimize the parameters of the ELM to improve its classification accuracy and training stability. Third, the study will evaluate whether the proposed fusion model can outperform existing advanced models in terms of evaluation metrics such as accuracy, F1 value, mean squared error (MSE), and prediction time on multiple typical industry customer churn datasets. To address the above problems, the study first proposes a synthetic minority over-sampling technique-SVM (SMOTE-SVM), which aims to improve the model's ability to detect minority categories. On this basis, GA is utilized to optimize ELM parameters, and GA-optimized ELM (GA-ELM) is proposed. Finally, the CCP model is built by combining SMOTE-SVM and GA-ELM. The innovation of the study is to improve the data quality by introducing the synthetic minority over-sampling technique (SMOTE) to generate minority samples in the SVM neighborhood. Meanwhile, combining GA adaptive search to optimize ELM weights and biases improves the model convergence speed and PA. The goal of this study is to give businesses a reliable, accurate, and efficient early warning system for customer attrition.

2. Methods and materials

2.1 Design of consumer data sample resampling algorithm based on SMOTE-SVM

In the enterprise product CCP problem, there is an extreme imbalance in the distribution of sample categories. In general, there are a lot fewer churn customers than non-churn clients. As a result, the training set (TrS) has a very low percentage of churn customer samples and a preponderance of non-churn customer samples [9-10]. In this imbalanced environment, traditional classifiers often tend to optimize the overall classification accuracy during the training process and neglect the identification of a few classes of samples. In real-world applications, this results in a notable decline in the model's capacity to forecast important churn clients. This not only reduces the practical value of the churn early warning system, but also directly affects the effectiveness of the enterprise's subsequent precision marketing and customer relationship management. For the sample category imbalance problem, SVM, as a classical classification model, can realize the effective segmentation of samples in the feature space by constructing an optimal hyperplane [11-13]. Based on this, the study first combines SVM and SMOTE to design the sample resampling algorithm. The optimal hyperplane classification process of SVM is shown in Figure 1.

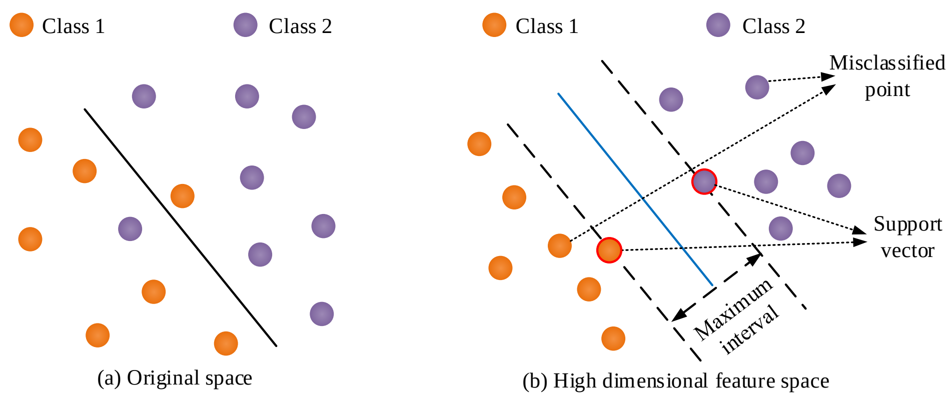


Figure 1: SVM optimal hyperplane classification diagram

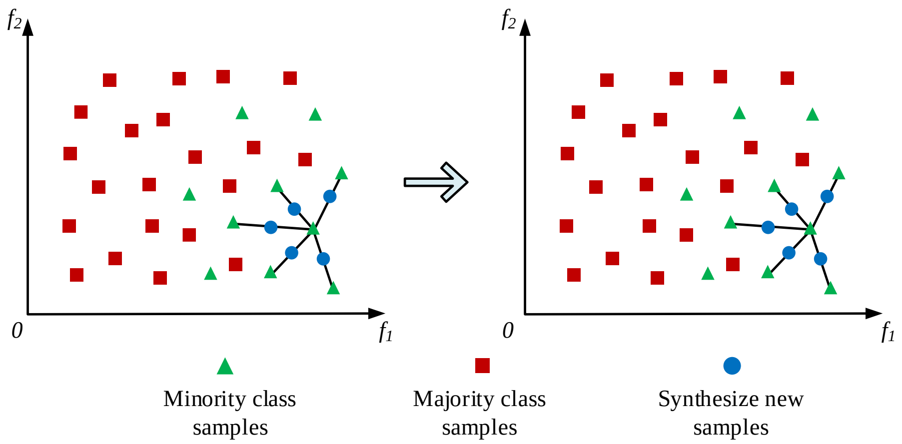


Figure 2: SVM optimal hyperplane classification diagram

Figure 1(a) shows the original feature space classification of SVM, and Figure 1(b) shows the high-dimensional feature space classification of SVM. SVM may efficiently discriminate between several sample categories by mapping to a high-dimensional space and creating an ideal hyperplane that maximizes the category spacing. However, in the CCP task, the SVM tends to optimize the overall classification accuracy during the training process, which leads to a decision boundary biased towards non-churn customers. This makes the prediction model's ability to recognize churn customers decrease significantly [14-16]. Based on this status quo, the study introduces SMOTE to improve the sample category distribution, which in turn improves the model's ability to recognize a small number of categories. The SMOTE algorithm can linearly interpolate between minority category samples to synthesize new ones, thereby increasing the number of minority category samples and alleviating the category imbalance problem. Its sample generation mechanism is shown in Figure 2.

In Figure 2, the freshly created synthetic samples are shown by the red squares on the right side of the graph, while the original data distribution is shown on the left. In the sample generation process, the blue dots are made to indicate the majority class samples and the black pentagrams indicate the minority class samples. The study uses an SVM model to train the boundary samples, construct SMOTE neighborhoods, and generate new samples through linear interpolation within the k-nearest neighbors of the minority class support vectors. To prevent noise points from appearing near the decision boundary, a minimum distance threshold and center-of-mass offset constraints are set. These constraints ensure that the interpolated samples are located in a reasonable region and improve the discriminative validity of the synthesized samples. In the resampling process, the study first trains an SVM model and extracts all the minority support vectors as representative boundary samples. Then, it only uses these support vectors as interpolation centers and selects other minority samples within their k-nearest neighbors for linear interpolation to generate new samples. This avoids the redundancy problem that may be introduced by standard SMOTE, which interpolates all the minority samples. Thus, it improves discriminative and boundary stability. The expression of the decision function is represented in Equation (1) [17-19].

$$f(x) = sign\left(\sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + b\right) \tag{1}$$

In Equation (1), α_i denotes the Lagrange multiplier. b denotes the bias term. x denotes the eigenvector of the input sample to be classified. x_i denotes the eigenvector of the i th support vector. y_i denotes the category label corresponding to this support vector. N denotes the number of support vectors. $K(x, x_i)$ denotes the kernel function between the input samples and the support vectors. Equation (2) illustrates the usage of the radial basis function as the KF.

$$K(x_i, x_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right) \tag{2}$$

In Equation (2), γ denotes the kernel width parameter, which is used to control the scale of the feature space mapping. Moreover, its numerical magnitude directly affects the similarity decay speed and the sensitivity of the classifier. In SVM classifier, the optimal decision boundary is mainly obtained by minimizing the weighted sum of the norm of hyperplane and classification error. Equation (3) displays the expression of its objective function.

$$\min_{\omega,b,\xi} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N}\xi_i \tag{3}$$

In Equation (3), ω denotes the normal vector of the hyperplane. ξ is the slack variable. ξ_i is the slack variable of the i th sample. C is the penalty factor. The constraints of the objective function in Equation (3) are shown in Equation (4).

$$subject\ \ to\ \ y_i\left(\omega^T \phi(x_i) + b\right) \geq 1 - \xi_i \tag{4}$$

In Equation (4), $\phi(x_i)$ denotes the feature mapping function. ω^T denotes the transpose of normal vector. In SVM classifier, the expression of the original TrS is made to be shown in Equation (5).

$$D = \left\{(x_i, y_i) \middle| x_i \in R^n, y_i \in \{+1, -1\}\right\} \tag{5}$$

In Equation (5), D represents the original TrS. +1 represents churned customers and -1 represents non-churned customers. n denotes the feature dimension. To improve the category distribution, the SMOTE algorithm will interpolate within the minority class sample neighborhood to generate new samples. The interpolation formula is shown in Equation (6).

$$x_{new} = x_i + \lambda(x_k - x_i) \tag{6}$$

In Equation (6), x_{new} denotes the new sample obtained by interpolation and x_k denotes one of the k nearest neighbor minority class samples. λ denotes a random number uniformly distributed in the range (0,1). To avoid generating anomalous samples in the over-sparse region, the generation of new samples also needs to satisfy the distance constraint to the center of mass of the original set of samples, as shown in Equation (7).

$$\left\| x_{new} - x_{center} \right\|_2 \leq d_{\max} \tag{7}$$

In Equation (7), x_{center} denotes the center of mass in the original sample set. $d_{\max}$ is the preset maximum distance threshold. $\|\cdot\|_2$ denotes the Euclidean paradigm. The formula for x_{center} is shown in Equation (8).

$$x_{center} = \frac{1}{N_p}\sum_{i=1}^{N_p} x_i' \tag{8}$$

In Equation (8), N_p denotes number of minority samples, i.e., the quantity of churned customers. x_i' denotes the feature vector of each churned customer sample. Finally, the updated dataset after sampling by SMOTE algorithm is shown in Equation (9).

$$D' = D \cup \{(x_{new}, +1)\} \tag{9}$$

In Equation (9), D' denotes the updated dataset. Combining SVM and SMOTE, the operation flow of

SMOTE-SVM is obtained as shown in Figure 3.

In Figure 3, firstly, the SMOTE-SVM algorithm will input the original training dataset and extract a few classes of support vectors based on the SVM model training. Secondly, in the k-nearest neighbor set of the support vectors, the SMOTE algorithm is utilized to complete the interpolation and generate new samples. Next, Euclidean distance constraints are set to avoid generating anomalous samples. Finally, the synthesized new samples are combined with the original dataset to form a balanced dataset and complete the improved resampling process.

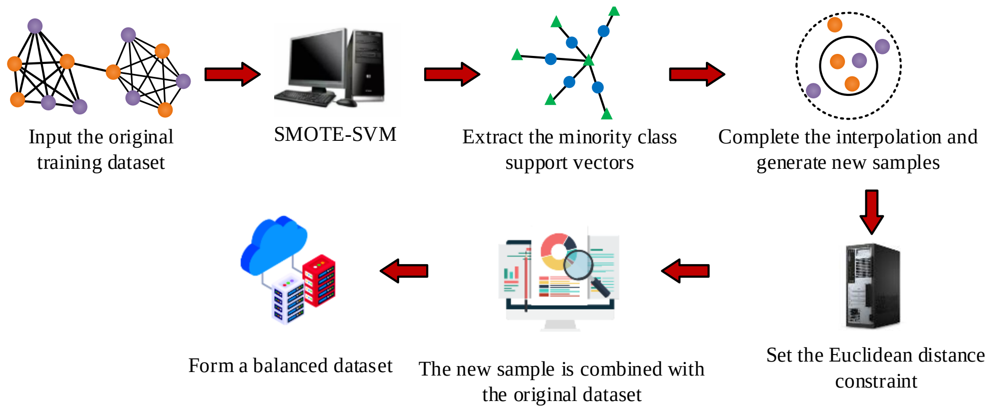


Figure 3: SMOTE-SVM flow chart

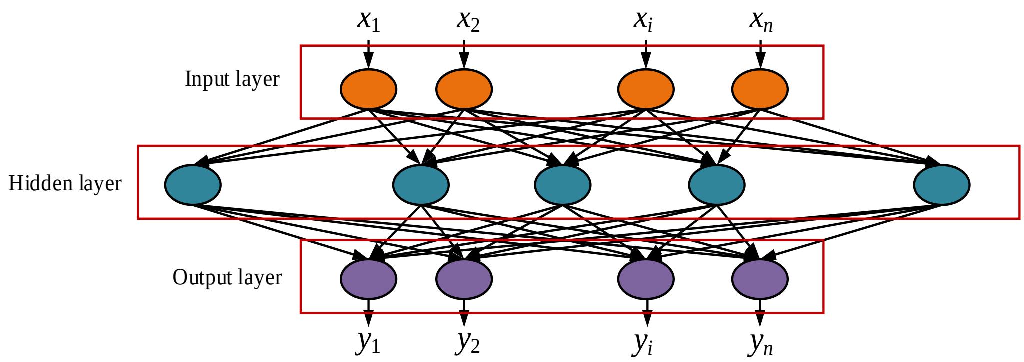


Figure 4: ELM assumption diagram

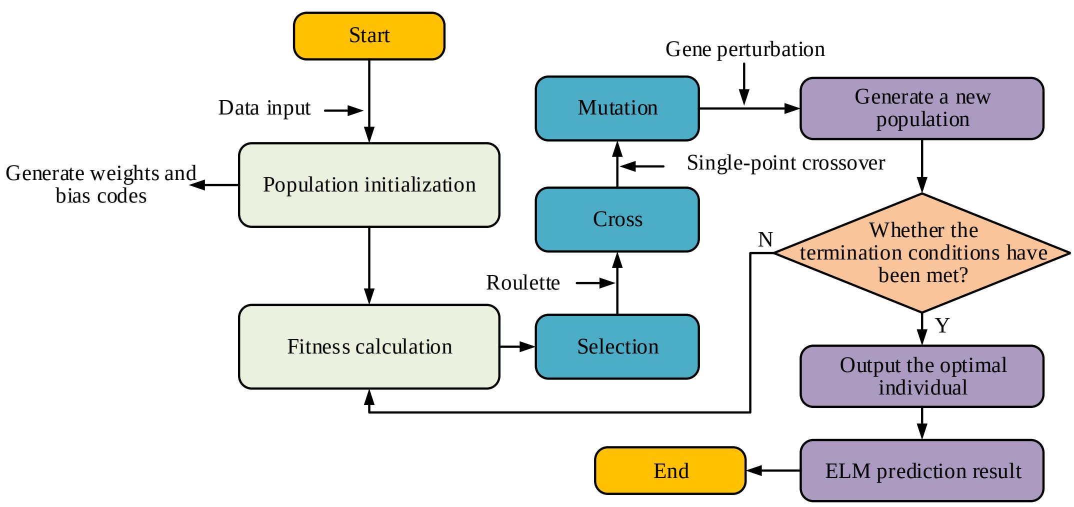


Figure 5: GA-ELM flow chart

2.2 Combining SMOTE-SVM and GA-ELM CCP model construction

To further improve the classification performance and stability of the CCP model, the study presents the GA-ELM optimized by GA to build the final churn prediction model based on the SMOTE-SVM sampling strategy after the category balancing process of sample data resampling is finished. ELM is a common single hidden layer (HL) feedforward neural network which is often used for classification and regression problems. The method has incredibly fast training speed, powerful generalization capabilities, and straightforward parameter selection. However, the initial parameters of ELM models are randomly generated, which easily leads to unstable model performance. Especially when facing the complex churn prediction task, the unreasonable initialization of parameters can affect the classification accuracy and model convergence. Therefore, to further enhance the model's PP and stability, the study presents GA, which globally optimizes the input weights and implicit layer bias of ELM. Figure 4 displays the ELM's structural makeup.

Figure 4 illustrates the basic structure of ELM. In ELM, the input features are mapped to the HL by a randomly initialized weight matrix. After nonlinear activation, then the output weights are determined by the least squares method (LSM) to finally realize the classification prediction. ELM lacks focused optimization because the input layer (IL) weights and HL bias settings are determined entirely at random, despite its benefits of quick training and easy implementation. This leads to a certain volatility in its classification performance and stability in the face of complex feature distributions [20-22]. Therefore, the study introduces GA to optimize the parameters of ELM model. Through the adaptive evolutionary search mechanism, the optimal initial weights and bias combinations that can improve the model performance are searched. The operation flow of GA-ELM is shown in Figure 5.

In Figure 5, firstly, the GA-ELM algorithm initializes the population individuals based on the original data, so that each individual code corresponds to a set of ELM IL weights and HL bias. Second, the fitness function determines each individual's prediction error, and the population evolves based on selection, crossover, and mutation processes. Next, the population is optimized iteratively and terminated after reaching the maximum number of iterations. Finally, the individual with the best fitness is selected as the initial parameter of the ELM model to complete the training and prediction process. In the initialization stage of the population, the input weight matrix is made to be W_k, and the HL bias is b_k. The expression of individual coding is obtained as shown in Equation (10) [23-25].

$$X_k = \{W_k, b_k\} \quad (k = 1, 2, \cdots, M) \tag{10}$$

In Equation (10), M denotes the quantity of individuals generated. X_k denotes the individual. For each X_k, the output matrix (OM) of its HL is shown in Equation (11).

$$H^{(k)} = g(ZW_k + B_k) \tag{11}$$

In Equation (11), Z denotes the matrix consisting of all input samples. $H^{(k)}$ denotes the OM of the HL. B_k denotes the bias matrix. $g(\cdot)$ denotes the activation function. The LSM is used to solve the output layer weights based on the HL output as shown in Equation (12).

$$\beta_k = \left(H^{(k)}\right)T' \tag{12}$$

In Equation (12), β_k denotes the output layer weights. T' denotes the target output vector. The fitness function is set for each individual as shown in Equation (13).

$$F(X_k) = \frac{1}{N}\sum_{i=1}^{N}\left(t_i - \left(h_i^{(k)}\beta_k\right)\right)^2 \tag{13}$$

In Equation (13), $F(X_k)$ denotes the fitness function. $h_i^{(k)}$ is the output of the i th sample in the HL. t_i denotes the i th target output vector. In the population evolution stage, single point crossover with Gaussian variation is used to generate new individuals. The expression of single point crossover operation is shown in Equation (14).

$$X_{child} = \{X_{parent1}(1:c), X_{parent2}(c+1:end)\} \tag{14}$$

In Equation (14), X_{child} denotes the newly generated zygotic individual. $X_{parent1}$ and $X_{parent2}$ denote parent individual 1 and parent individual 2, respectively. c denotes the position of single point crossover. $(1:c)$ denotes the gene segment from position 1 to position c in the coding of parent individual 1. $(c+1:end)$ denotes the gene segment from the $c+1$ position to the last position in the coding of parent individual 2. Next, a normal perturbation is added to achieve Gaussian variation as shown in Equation (15).

$$X_{mutated} = X_{child} + \sigma \times \delta(0,1) \tag{15}$$

In Equation (15), $X_{mutated}$ denotes the mutated offspring individuals. σ denotes the magnitude of variation. $\delta(0,1)$ denotes standard normal distribution. After completing ELM training using GA, the study combines SMOTE-SVM sample resampling with GA-ELM parameter optimization algorithm to build the final CCP model. The model is denoted as CCP model based on SMOTE-SVM sampling and GA-ELM optimization (SSGE-CCPM). Its structure is shown in Figure 6.

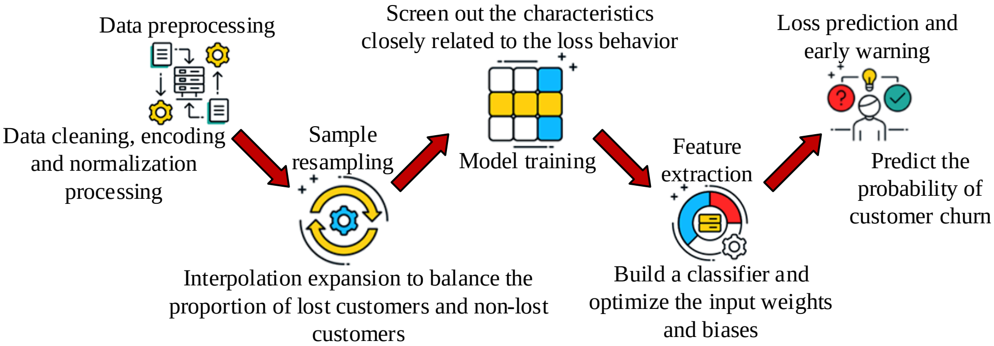


Figure 6: SSGE-CCPM prediction model structure diagram

The SSGE-CCPM model in Figure 6 can be divided into five main modules as a whole, which are data preprocessing module, sample resampling module, feature extraction module, model training module, and churn prediction and warning module. First, the data preprocessing module is responsible for cleaning, coding, and normalizing the original customer data to ensure the consistency and quality of the subsequent modeling data. Second, the sample resampling module adopts the SMOTE-SVM algorithm to interpolate and expand the samples for a few classes to balance the ratio of the number of churned and non-churned customers. Next, the feature extraction module filters features closely related to churning behavior from the preprocessed data. It uses Pearson correlation analysis and principal component analysis to reduce the number of high-dimensional, redundant features and ultimately retains 13 main features for model training. This improves training efficiency and generalization ability. The model training module then builds the classifier using the GA-ELM algorithm and utilizes GA to optimize the input weights and biases in an attempt to increase the ELM model's PP and stability. Finally, the churn prediction and early warning module uses the optimized GA-ELM model to predict the probability of CC. It also generates an early warning based on the prediction results to assist enterprises in formulating intervention strategies.

3 Results

3.1 Algorithm performance testing

The study builds a unified experimental platform for the SSGE-CCPM model to ensure the fairness and reliability of the results. The experimental operating system is Windows 10, the GPU used is NVIDIA GeForce RTX 3080, and the CPU is Intel Core i9-11900K with 64GB DDR4 memory. All experiments are run in Python 3.9 environment in conjunction with PyTorch 1.10 deep learning framework (DLF). The study uses the publicly available CCP dataset Telco CC for model validation. The experiment randomly splits the dataset into TrS and test set (TeS) in an 8:2 ratio to guarantee data division's unpredictability and consistency. Table 2 displays the particular experimental apparatus, dataset details, and model parameters.

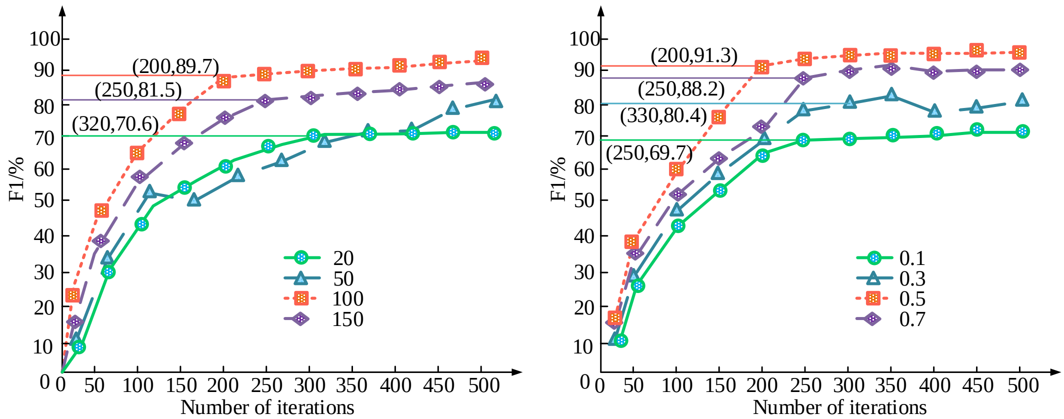
Table 2 lists the model parameter settings, hardware configuration, and basic dataset information for the investigation. The study first choose two types of hyperparameters that affect the model the most for the selected value test. The test results are shown in Figure 7.

Table 2: Experimental configuration information table

Category	Parameter/equipment	Description
Experimental setup	Operating system	Windows 10
	GPU	NVIDIA GeForce RTX 3080
	CPU	Intel Core i9-11900K
	Memory	64GB DDR4
	DLF	PyTorch 1.10
	Programming language	Python 3.9
Model parameters	GA population size	50
	Maximum iterations	100
	Crossover rate	0.8
	Mutation rate	0.05
	Hidden layer neurons	500
	SVM kernel type	Radial basis function
Dataset information	Dataset name	Telco customer churn
	Total samples	7032

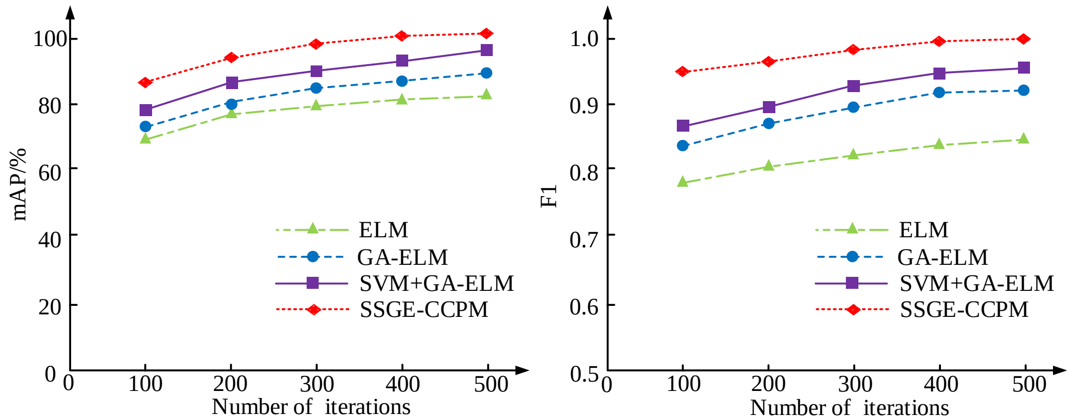
	Feature dimensions	21
	Positive samples	1869
	Negative samples	5163

Figure 7(a) shows the results of the value selection test for population size. Figure 7(b) shows the results of the value selection test for crossover rate. As illustrated in Figure 7(a), with an increase in population size, the F1 value of the model at each iteration stage exhibits an overall upward trend, particularly at the inception of the iteration. A notable enhancement in convergence speed and final performance of the model is observed when the population size is increased (e.g., 100 and 150). When the population size is 100, the F1 value has stabilized after the 250th iteration and finally reaches 91.5%, while the 150 group has weakened relative marginal gains in the enhancement effect. In contrast, the parameter combinations of sizes 20 and 50 converges more slowly and ends up with significantly lower F1 values than the larger sizes. Figure 7(b) shows the effect of different crossover rates on model performance. When the crossover rate is 0.5, the model performs optimally in most stages with a final F1 value of 91.3% and has largely converged after the 200th iteration. The model with a crossover rate of 0.7 is slightly faster in the initial stage, but slower in the final stage, with an F1 value of 88.2%. In contrast, models with crossover rates below 0.5 (e.g., 0.1 or 0.3) are slow to train and exhibited low convergence values. This indicates that an excessively low crossover rate hinders the ability to perform effective gene recombination. In summary, the final selects GA parameters are population size 100 and crossover rate 0.5, which take into account the PA and convergence efficiency of the model. To verify the performance of each module of the SSGE-CCPM model, the study uses mean average precision (mAP) and F1 value as indicators. The model is first tested for ablation as shown in Figure 8.



(a) Population size (b) Crossover rate

Figure 7: Hyperparameter selection test results



(a) The mAP of different models in the training set (b) The F1 of different models in the training set

Figure 8: The mAP and F1 values of different models

The performance differences between the original ELM, GA-ELM, SVM+GA-ELM, and the full SSGE-CCMP models are compared in the ablation experiments. Among them, SVM+GA-ELM refers to the direct training of the GA-ELM model using the resampled data generated by the SMOTE-SVM module as input, while SSGE-CCMP further integrates the feature extraction and center-of-mass constraint mechanisms based on the module. It optimizes the distribution of interpolated samples and the feature quality of the training data, thus achieving higher model stability and generalization ability. Figure 8 demonstrates the trend of mAP value and F1 value for different models on the TrS as the iteration increases. In Figure 8(a), the mAP values of each model show an increasing trend as the training iteration increases, but there is a significant difference in the magnitude of the enhancement. The original ELM model has the lowest overall mAP value and grows more slowly, eventually reaching only 82.1%. Following the implementation of GA optimization, the GA-ELM model performs better than the original ELM at every stage, demonstrating that the optimization approach can successfully raise the model's overall recognition accuracy. The SVM+GA-ELM model, which further combines SMOTE-SVM sampling, shows more obvious advantages in the middle and high iteration stages. In contrast, the SSGE-CCPM model proposed in this study consistently maintains the highest mAP value throughout the training process. Moreover, the improvement is more obvious, and finally the mAP reaches 98.5% at 500 iterations, which is far more than the other compared models. In Figure 8(b), SSGE-CCPM leads the F1 value at all stages, and finally reaches 0.98, which is significantly better than other combinations. The results of the ablation test demonstrate that the SSGE-CCPM model's accuracy and stability are significantly increased by the SMOTE-SVM sampling and GA-ELM optimization techniques. Additionally, there is a synergistic optimization impact when the two are combined.

ELM, extreme learning machine based on hybrid grey wolf optimizer (HGWO-ELM), extreme learning machine based on differential evolution and resampling strategy (DE-RS-ELM) are further selected as comparison algorithms. Figure 9 displays the four methods' loss curves on the TrE and TeS. DE-RS-ELM employs a random under-sampling (RUS) strategy to achieve class balance by randomly deleting samples in most classes. HGWO-ELM does not explicitly resample, but it mitigates the impact of imbalance indirectly by searching the feature space optimally with gray wolves. The original ELM model does not use any resampling techniques and is only trained directly on the original TrS.

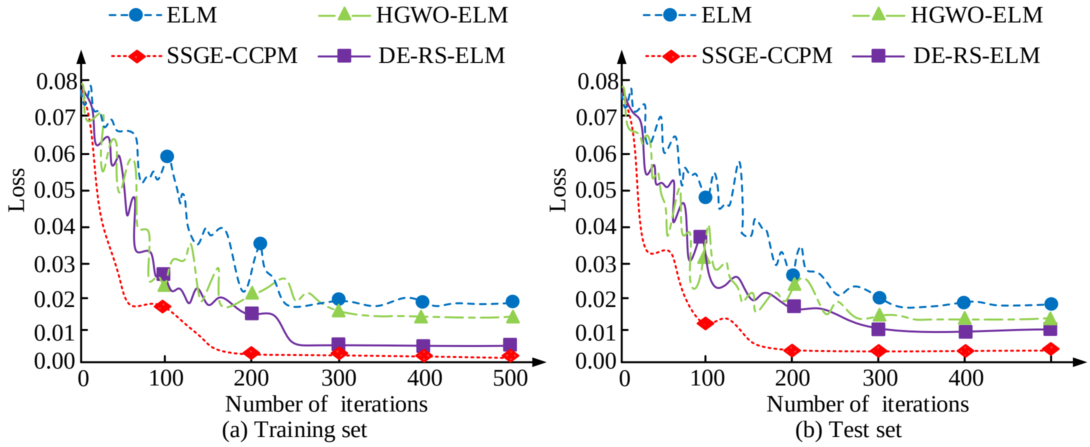


Figure 9: Loss curves for different algorithms

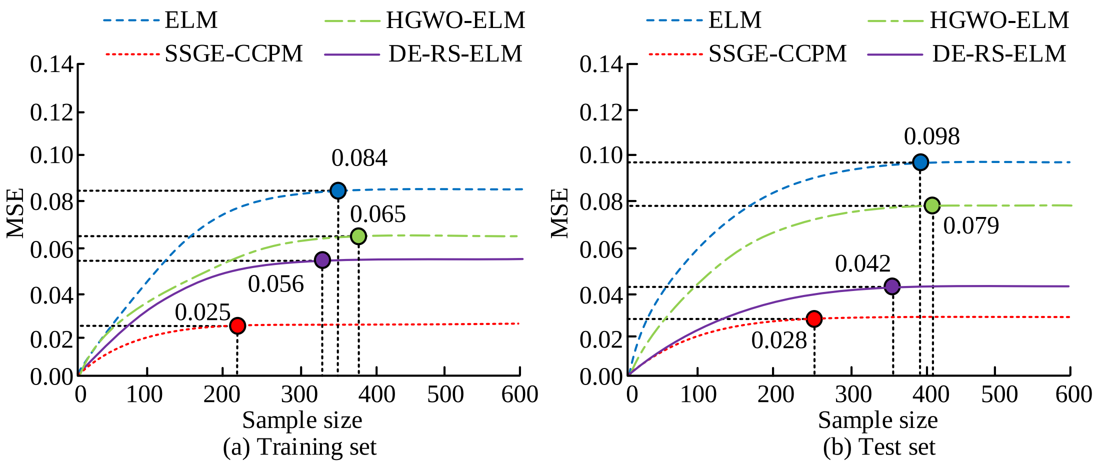


Figure 10: MSE values of different algorithms

Figure 9 shows the Loss variation curves of different models on the TrE and TeS. In Figure 9(a), the original ELM model has the slowest convergence and the largest fluctuation on the TrS. In contrast, the SSGE-CCPM model shows a faster descent rate at the beginning of the iteration and converges optimally with the Loss converging to 0 after 185 iterations. In Figure 9(b), SSGE-CCPM maintains the lowest Loss on the TeS at all stages, and finally almost converges to 0. The outcomes display that SSGE-CCPM not only converges faster during the training process, but also exhibits better generalization ability and stability on the TeS. Comparison of the MSE of the four algorithms on the two datasets is shown in Figure 10.

Figures 10(a) and 10(b) display the MSEs of different models on the TrE and TeS, respectively. In Figure 10(a), the original ELM model has the highest MSE of 0.084 on the TrS. The MSEs of HGWO-ELM and DE-RS-ELM decrease to 0.065 and 0.056, respectively. In contrast, SSGE-CCPM performs best on the TrS with an MSE of only 0.025, which is significantly better than the other compared methods. In Figure 10(b), on the TeS, the MSEs of ELM, HGWO-ELM, and DE-RS-ELM are 0.098, 0.079, and 0.042, respectively, while SSGE-CCPM has the lowest MSE of 0.028. The results show that SSGE-CCPM has lower prediction errors on both the TrE and TeS, verifying its good fitting ability and strong generalization performance. In order to troubleshoot the risk of overfitting and to validate statistical significance, the study conducts a five-fold cross-validation, recording the TrS and TeS performance for each fold separately. The results are shown in Table 3.

Table 3: 5-fold cross-validation results

Fold	Dataset	ELM	GA-ELM	SVM+GA-ELM	SSGE-CCPM	*p*-value and SSGE-CCPM (ELM/GA/SVM)
1	Train	0.842	0.8894	0.9226	0.9561	
1	Test	0.8191	0.8691	0.9063	0.9381	0.0000 / 0.0000 / 0.0002
2	Train	0.8381	0.8922	0.9181	0.9481	
2	Test	0.821	0.8623	0.8931	0.9328	0.0000 / 0.0000 / 0.0002
3	Train	0.8359	0.8913	0.9164	0.9444	
3	Test	0.8259	0.8691	0.9003	0.9293	0.0000 / 0.0000 / 0.0002
4	Train	0.8378	0.8904	0.9154	0.9515	
4	Test	0.8176	0.8688	0.8976	0.9424	0.0000 / 0.0000 / 0.0002
5	Train	0.8399	0.8858	0.9233	0.9451	
5	Test	0.8208	0.8622	0.8947	0.9358	0.0000 / 0.0000 / 0.0002

Table 4: Predictive accuracy of different models

Dataset	Method	Accuracy (%)
Telco Customer Churn	ELM	88.12
	HGWO-ELM	89.35
	DE-RS-ELM	89.72
	SSGE-CCPM	**92.68**
Bank Churn	ELM	86.57
	HGWO-ELM	87.92
	DE-RS-ELM	88.15
	SSGE-CCPM	**91.30**
Retail Subscription Churn	ELM	85.41
	HGWO-ELM	86.68
	DE-RS-ELM	87.09

	SSGE-CCPM	**90.22**
Insurance Policy Cancellation	ELM	87.83
	HGWO-ELM	88.76
	DE-RS-ELM	89.25
	SSGE-CCPM	**92.01**
Online Service Churn	ELM	89.05
	HGWO-ELM	90.02
	DE-RS-ELM	90.31
	SSGE-CCPM	**93.17**

In Table 3, the SSGE-CCPM model significantly outperforms the other three comparison models (ELM, GA-ELM, and SVM+GA-ELM) in terms of F1 values on all folds of the TeS under five-fold cross-validation. Specifically, the SSGE-CCPM model achieved the highest F1 value in all folds, and its paired t-test p-values are all less than 0.001. This indicates that the performance improvement obtained is highly statistically significant. Additionally, the differences between the F1 values of the training and TeSs are minimal, and the fluctuations in standard deviation are well-controlled. This further demonstrates that the proposed model has strong generalization capabilities and a low risk of overfitting.

3.2 Analysis of the effect of model application

To further validate the effectiveness of the SSGE-CCPM model in real CC early warning tasks, the study selected five public customer datasets from various industries as test objects. The study uniformly adopted the hyper-parameter configurations tuned on the Telco dataset, rather than making separate adjustments for each dataset, to test the model's ability to generalize and its robustness under cross-dataset conditions. The selected datasets cover a wide range of domains such as telecom, finance, retail, insurance, and subscription services. The churn distribution ranges from 15% to 40%, and the sample size ranges from thousands to tens of thousands, which is widely representative and challenging. The PA of the four models, ELM, HGWO-ELM, DE-RS-ELM, and SSGE-CCPM, is compared on the five types of datasets, as shown in Table 4.

Table 4 lists the classification accuracy comparison results of the four models, ELM, HGWO-ELM, DE-RS-ELM and SSGE-CCPM, on five different domain CCP datasets. The overall performance shows that SSGE-CCPM achieves the highest classification accuracy on all datasets, demonstrating a clear performance advantage. On the telecom churn dataset, the SSGE-CCPM model achieves an accuracy of 92.68%, which is a 2.96% improvement over the 89.72% of DE-RS-ELM. In the financial and retail datasets, SSGE-CCPM also achieves an accuracy rate of 91.30% and 90.22%, respectively. In the subscription service dataset, SSGE-CCPM has the highest PA of 93.17%, which is a 2.86% improvement over the 90.31% of DE-RS-ELM. Comparatively, DE-RS-ELM outperforms ELM and HGWO-ELM, but does not exceed SSGE-CCPM in all datasets. It shows that there are still limitations in relying solely on the differential evolution with resampling strategy for enhancement. Whereas the performance of ELM and HGWO-ELM is closer, but the overall accuracy is significantly lower than that of DE-RS-ELM and SSGE-CCPM. Overall, SSGE-CCPM maintains a stable performance advantage in different application scenarios. The test continues with the average prediction time as an indicator, and the results are shown in Figure 11.

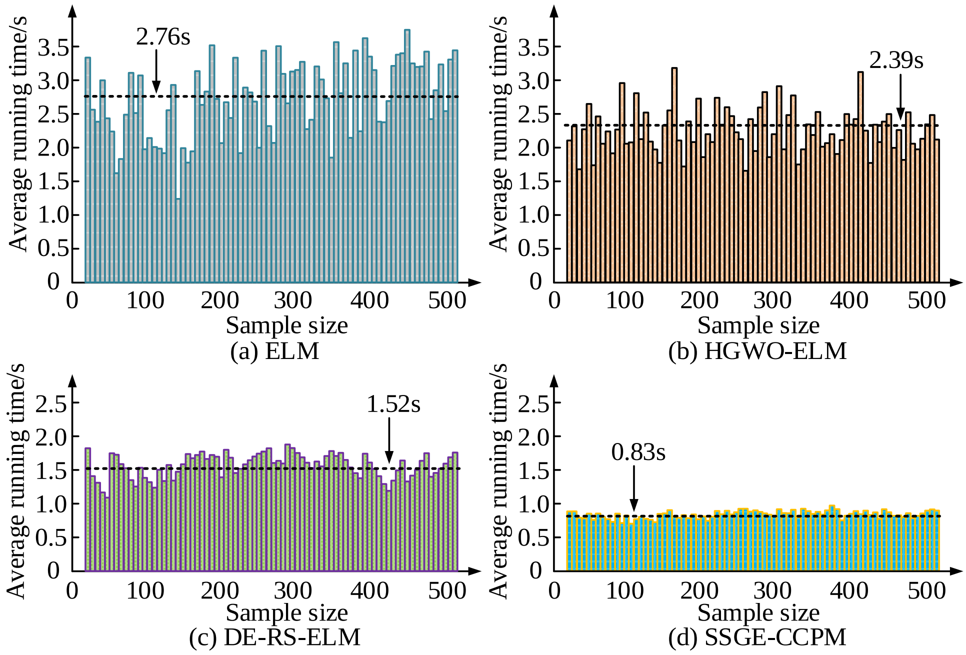


Figure 11: Average prediction time for different models

Figure 11 demonstrates the variation of the average prediction time of the four models with different sample sizes. In Figs. 11(a) and 11(b), the original ELM and HGWO-ELM models have higher average prediction times, which are stable at 2.76s and 2.39s, respectively. The prediction time of the DE-RS-ELM model in Figure 11(c) decreases, which is about 1.52s. In contrast, the SSGE-CCPM model in Figure 11(d) under each sample size has the has the lowest prediction time with an average of only 0.83s and the smallest fluctuation. It verifies its advantages of simple structure and high parallelism in the inference stage.

4 Discussion

The experimental results showed that the SSGE-CCPM model outperformed existing methods on all five public datasets in terms of F1 mean and prediction efficiency, achieving up to 98.5%. The performance of the TeS was highly consistent with that of the TrS, with small standard deviation fluctuations, indicating the model's good generalization ability. This performance improvement was mainly due to two key strategies: On the one hand, this study improved the distribution density of minority class boundary samples by bootstrap oversampling the minority class support vector neighborhood through the SMOTE-SVM mechanism. Its effectively alleviated the class imbalance problem and improves the recognition sensitivity of the model. On the other hand, this study optimized the global search of ELM network parameters by GA, which improved the stability and convergence efficiency of the model and significantly reduced the prediction time (only 0.83 s on average). However, the method still faced challenges in dealing with sample distribution noise. In particular, the synthetic samples generated by SMOTE in the neighboring regions of the support vectors could introduce uncontrollable boundary noise, thus affecting the robustness of the model. In the future, a combination of denoising strategies or boundary confidence mechanisms was needed to further suppress pseudo-sample interference. Moreover, although the model was validated on several standard datasets, its adaptability to dynamic environments such as high-dimensional heterogeneous data and conceptual drift needed to be further evaluated in real business systems. Its scalability was still limited by its dependence on specific data structures and a priori parameters.

On the other hand, in applying SSGE-CCPM to the actual system of customer churn prediction, the potential ethical and social risks needed to be fully considered. Because the model used historical behavioral data and service feedback to select features, it might unintentionally amplify systematic biases based on certain group characteristics (e.g., geography, gender, or consumption frequency). This could result in "algorithmic discrimination" or "unfair rejection" when judging churn among groups. This can lead to problems such as "algorithmic discrimination" or "unfair rejection" of churn judgment results among different groups. In addition, high-precision predictive models might be used for high-intensity marketing, differentiated pricing, or customer screening, all of which might violate users' rights to information and fair trade. Therefore, when deploying such models, enterprises and developers should introduce an interpretability mechanism that could trace back and transparently disclose the logic behind churn determination. This could help avoid potential harm to users' interests caused by algorithmic black boxes. At the same time, it was recommended that compliance strategies be combined with data ethics guidelines and that a dual safeguard mechanism be established for "minimizing deviation and controlling intervention" in training data collection, user profile modeling, and result usage scenarios. This could help achieve a balance between technological innovation and users' rights and interests.

5 Conclusion

Aiming at the problem of sample category imbalance and insufficient classification performance in CCP, the study proposed a novel SSGE-CCPM fusing data mining and intelligent neural network based on SMOTE-SVM sample resampling and GA-ELM.

The performance test results indicated that the complete SSGE-CCPM model outperformed other ablation combinations in terms of both mAP and F1 value on the TrS. When the number of iterations reached 500, the SSGE-CCPM model had the highest mAP and F1 value, which were 98.5% and 0.98, respectively. Compared to the three comparative models, ELM, HGWO-ELM, and DE-RS-ELM, the SSGE-CCPM was more stable on the dataset and performed better in terms of error. In the TrS, the SSGE-CCPM model had less fluctuation during the iteration process and needed only 185 iterations to satisfy the Loss value converging to 0, indicating its optimal convergence. In the TeS, the MSE values of the four models ELM, HGWO-ELM, DE-RS-ELM, and SSGE-CCPM were 0.098, 0.079, 0.042, and 0.028, respectively, after they reached the steady state. In the practical application tests of the models, SSGE-CCPM achieved the highest classification accuracy on all datasets. Among them, on the online service subscription churn dataset, SSGE-CCPM achieved the highest accuracy of 93.17%. In addition, in terms of the average prediction time, the running time of SSGE-CCPM is as low as 0.83 seconds, which is significantly better than that of the comparison model, reflecting the high prediction efficiency. The thorough findings demonstrate that the suggested model not only successfully raises the churn customer identification accuracy but also considers the twin needs of stability and real-time in real-world applications. Although the study achieves good results in improving churn PA and efficiency, there are still some limitations in handling extreme imbalance data, modeling complex behavioral characteristics and generalizing applications across industries. Future research will further optimize the robustness and adaptive capability of the model and explore other multimodal feature fusion and semi-supervised learning methods to enhance the application breadth and PA in more practical scenarios.

Funding

This study is supported by Soft Science Research Plan of Henan Province in 2025: "Study on Spatial Dislocation and Optimization Path of Tourism Competitiveness and Network Attention in Henan Province" (No. 252400410180).

References

[1] Yang W. Research on early warning of customer churn based on mutual information and integrated learning-taking ctrip as an example. Academic Journal of Computing & Information Science, 2022, 5(3): 785-800. https://doi.org/10.25236/AJCIS.2022.050303

[2] Adekunle B I, Chukwuma-Eke E C, Balogun E D, Ogunsola K O. Improving customer retention through machine learning: A predictive approach to churn prevention and engagement strategies. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2023, 9(4): 507-523. https://doi.org/10.32628/IJSRCS

[3] Gurung N, Hasan M R, Gazi M S, Chowdhury F R. AI-based customer churn prediction model for business markets in the USA: exploring the use of ai and machine learning technologies in preventing customer churn. Journal of Computer Science and Technology Studies, 2024, 6(2): 19-29. https://doi.org/10.32996/jcsts.2024.6.2.3x

[4] Wu K H, Chen P Y, Chiu C. A social media-based profiling approach for potential churning customers: An example for telecom industry. Journal of Internet Technology, 2022, 23(7): 1565-1571. https://doi.org/10.53106/160792642022122307011

[5] Toor A A, Usman M. Adaptive telecom churn prediction for concept-sensitive imbalance data streams. The Journal of Supercomputing, 2022, 78(3): 3746-3774. http://dx.doi.org/10.1007/s11227-021-04021-x

[6] Lalwani P, Mishra M K, Chadha J S, Sethi P. Customer churn prediction system: a machine learning approach. Computing, 2022, 104(2): 271-294. http://dx.doi.org/10.1007/s00607-021-00908-y

[7] Matuszelański K, Kopczewska K. Customer churn in retail e-commerce business: Spatial and machine learning approach. Journal of Theoretical and Applied Electronic Commerce Research, 2022, 17(1): 165-198. http://dx.doi.org/10.3390/jtaer17010009

[8] Singh P P, Anik F I, Senapati R, Sinha A, Sakib N, Hossain E. Investigating customer churn in banking: A machine learning approach and visualization app for data science and management. Data Science and Management, 2024, 7(1): 7-16. http://dx.doi.org/10.1016/j.dsm.2023.09.002

[9] Liu Y, Fan J, Zhang J, Yin X, Song Z. Research on telecom customer churn prediction based on ensemble learning. Journal of Intelligent Information Systems, 2023, 60(3): 759-775. http://dx.doi.org/10.1007/s10844-022-00739-z

[10] Wang G Y. Churn prediction for high-value players in freemium mobile games: using random

under-sampling. Statistika: Statistics and Economy Journal, 2022, 102(4): 443-453. http://dx.doi.org/10.54694/stat.2022.18

[11] Ali I M S, Hariprasad D. Hyper-heuristic salp swarm optimization of multi-kernel support vector machines for big data classification. International Journal of Information Technology, 2023, 15(2): 651-663. https://doi.org/10.1007/s41870-022-01141-2

[12] G Mehdi, H Hooman, Y Liu, S Peyman and R. Arif. Data mining techniques for Web mining: A survey. Artificial Intelligence and Applications, 2022, 1(1):3-10. https://doi.org/10.47852/bonviewAIA2202290

[13] Pichaimani T, Ratnala A K, Parida P R. Analyzing time complexity in machine learning algorithms for big data: a study on the performance of decision trees, neural networks, and SVMs. Journal of Science & Technology, 2024, 5(1): 164-205. https://thesciencebrigade.com/jst/article/view/454.

[14] Junaid M, Ali S, Siddiqui I F, Nam C, Qureshi N M F, Kim J, Shin D R. Performance evaluation of data-driven intelligent algorithms for big data ecosystem. Wireless Personal Communications, 2022, 126(3): 2403-2423. http://dx.doi.org/10.1007/s11277-021-09362-7

[15] Kurani A, Doshi P, Vakharia A, Shah M. A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. Annals of Data Science, 2023, 10(1): 183-208. http://dx.doi.org/10.1007/s40745-021-00344-x

[16] Neu D A, Lahann J, Fettke P. A systematic literature review on state-of-the-art deep learning methods for process prediction. Artificial Intelligence Review, 2022, 55(2): 801-827. https://doi.org/10.1007/s10462-021-09960-8

[17] Wu F, Lyu F, Ren J, Yang P, Qian K, Gao S, Zhang Y. Characterizing internet card user portraits for efficient churn prediction model design. IEEE Transactions on Mobile Computing, 2023, 23(2): 1735-1752. http://dx.doi.org/10.1109/TMC.2023.3241206

[18] Miao M, Miao T, Long Z. User churn prediction hierarchical model based on graph attention convolutional neural networks. China Communications, 2024, 21(7): 169-185. http://dx.doi.org/10.23919/JCC.fa.2024-0104.202407

[19] Nurcahyawati V, Mustaffa Z. Improving sentiment reviews classification performance using support vector machine-fuzzy matching algorithm. Bulletin of Electrical Engineering and Informatics, 2023, 12(3): 1817-1824. http://dx.doi.org/10.11591/eei.v12i3.4830

[20] Oktavia D, Ramadahan Y R, Minarto M. Analisis sentimen terhadap penerapan sistem e-tilang pada media sosial twitter menggunakan algoritma support vector machine (SVM). KLIK: Kajian Ilmiah Informatika dan Komputer, 2023, 4(1): 407-417. https://doi.org/10.30865/klik.v4i1.1040

[21] Pamungkas C A, Azril M F. Optimizing support vector machine for imbalanced datasets by combining posterior probability and correlation methods. International Journal of Software Engineering and Computer Systems, 2025, 11(1): 16-31. http://dx.doi.org/10.15282/ijsecs.11.1.2025.2.0134

[22] Boonmunewai J, Chantaraksa T, Rodjanadid B. Evaluating machine learning methods for solving class imbalance in banking customer data: a comparative study. KKU Science Journal, 2024, 52(3): 349-362. https://doi.org/10.14456/kkuscij.2024.27

[23] Kumar V, Kumar R K, Singh S K. Evaluation and enhancement of standard classifier performance by resolving class imbalance issue using smote-variants over multiple medical datasets. SN Computer Science, 2025, 6(3): 1-30. http://dx.doi.org/10.1007/s42979-025-03775-y

[24] Jia Y, Wang Y, Yang Y, Huang j, Xiao J. Complex-valued GMDH-based data characteristic-driven adaptive decision support system for customer classification. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2023, 54(1): 403-415. http://dx.doi.org/10.1109/TSMC.2023.3309709

[25] Koc K, Budayan C, Ekmekcioğlu Ö. Predicting cost impacts of nonconformances in construction projects using interpretable machine learning. Journal of Construction Engineering and Management, 2024, 150(1): 4023-4027. http://dx.doi.org/10.1061/JCEMD4.COENG-13857