Financial Crisis Forecasting on Imbalanced Data Using SGD-Optimized Gaussian SVM with Adaptive Oversampling

Chao Li
Wuhan QingchuanUniversity，School of Accounting，430070
E-mail: Lee8333@163.com

Keywords: financial risk, early warning model, imbalanced data, SGD-GSVM, financial crisis

Received: June 27, 2025

Background: Economic stability depends on the ability to foresee financial risk, particularly in markets that are extremely volatile. Unbalanced financial data is difficult for traditional Support Vector Machine (SVM) models to handle, which results in subpar crisis detection capabilities. In order to improve financial risk early warning models, this study combines Gaussian SVM with stochastic gradient descent (SGD) optimisation (SGD-GSVM). Methods: The study dataset on trading days (more than 2,000 trading days, January 2022-February 2024), which included 45 market, macroeconomic, and sentiment variables (e.g., stock indices, volatility indicators, interest rates, exchange rates). The ADASYN sampling method has been used to address the serious imbalance between normal and crisis period by dynamically creating synthetic minority samples at sparse crisis areas. Key evaluation metrics, such as accuracy, recall, F1-score, G-Mean, AUC-PR, and training time, were used to train and evaluate the SGD-GSVM model to Standard GSVM, SMOTE-SVM, CS-SVM, and Random Forest. Results: Standard GSVM (76% accuracy, 1,200s training time) and CS-SVM (81% accuracy, 1,300s training time) were greatly outperformed by the suggested SGD-GSVM model, which obtained the greatest accuracy of 92% with a training time of just 180 seconds. Additionally, it showed excellent recall (90%) and precision (82%), making it the most effective and efficient model for predicting financial risk. Time spent on training was also greatly lowered to 180 s compared to 1,200-1,500 s of the classical SVM models. Conclusion: This work offers a new method for early warning of financial risk by combining SGD optimisation with Gaussian SVM and employing adaptive oversampling for data balancing. The findings show that SGD-GSVM is the best model because it strikes a balance between high accuracy and computational economy. Financial organisations can create real-time risk management plans with the help of the suggested technique. For additional performance improvements, hybrid deep learning approaches might be investigated in future studies.

Povzetek: Model SGD-GSVM omogoča hitrejše in natančnejše zgodnje zaznavanje finančnih tveganj kot klasični modeli.

1 Introduction

Extreme market swings have a significant influence on financial risk management, drawing the attention of economic and financial management departments as well as investors to the resulting extreme financial risk occurrences [1]. Reason being, general dangers do show up in the markets for finance, but they won't pose a lethal threat. The national economy could collapse and have catastrophic repercussions due to the enormous financial risks induced by the severe downturn of the market for financial products [2]. Given this, it's clear that there's a pressing need for research into a Chinese financial market-specific serious danger early detection model that can reliably forecast such risks, so that investors can plan their investments and economic and financial management departments can take precautions against them [3]. Several top-tier academic conferences and workshops have addressed the issue of imbalanced data categorisation within the last 20 years. Predicting potential financial risks is essential for keeping markets and financial institutions stable. Businesses, banks, as well as investors can save money, make better investments, and keep the economy stable if they can anticipate and respond to possible financial crises. The problem with dealing with imbalanced datasets is that economic distress cases are infrequent relative to maintain stable financial conditions. This makes financial risk assessment a challenging task. Support Vector Machines (SVMs) and other conventional machine learning models have serious difficulties due to this imbalance, which causes them to make prejudiced predictions that benefit the majority (the "non-risk cases") and ignore the minority (the "risk cases") [4].

Businesses in China are encountering new opportunities and threats as a result of the rapid pace of economic globalisation, which is also increasing the size of their development scale. Manufacturing enterprises in China, which play a significant role in the country's economy, often take advantage of opportunities to bring in new technology, talent, and equipment. However, this can lead to problems like high costs for human assets and financial resources, or even a break in the capital chain, which can put the company at risk of bankruptcy[5]. A financial crisis occurs when a company's managers fail to identify possible financial risks in a timely manner. This crisis affects all parties involved, including investors and creditors in the business. A financial crisis manifests itself in a number of ways, including problems with capital turnover, a drop in profits, insufficient daily working capital to keep the business running normally, and, finally, listed enterprises suffer enormous economic losses. In order to prevent bigger risks from unintentionally occurring, it is critical to detect financial risks early on, manage and deal with them in a timely manner, and turn an irreversibly financial crisis into a reversible one. In the midst of a crisis, the global financial system consequently crumbles. Even in the midst of a financial crisis, the system remains vulnerable because losses can ripple through other banks[6]. Computational methods have been gradually replacing regression analysis as the go-to for analysing response variables in EWS and predictive modelling due to their ability to reveal nonlinear fluctuations in variables. Because of this, early financial warnings are becoming more effective.

Main Contribution on this study:

This paper introduces the SGD-GSVM model, which greatly increases prediction accuracy, efficiency, and adaptability when compared to current methods, making several important contributions to the field of financial risk early warning systems.

- Effectively addresses the problem of data imbalance in financial crisis prediction by incorporating Adaptive Synthetic Sampling (ADASYN), which ensures that minority crisis periods are adequately represented, improving recall and minimising false negatives.
- SGD-Optimized Gaussian SVM, a novel approach that improves crisis detection performance while lowering computational costs, making it appropriate for real-time financial monitoring.
- In addition, the model is computationally efficient for large-scale financial datasets, reducing training time to just 180 seconds.
- Finally, this study offers investors, regulatory bodies, and financial institutions a flexible and scalable framework that helps them identify financial crises early and take proactive measures to reduce risk. Future developments in AI-driven financial modelling are made possible by these contributions, which position SGD-GSVM as a state-of-the-art solution for financial risk assessment and crisis forecasting.

Here is the outline for the remaining portion of the paper: In Section 2, we cover the relevant literature on the topic. In Section 3, we lay out the strategy that will be used to accomplish the goals. In Section 4, the experimental validation of the suggested methodology is presented. A summary of the work is provided in Section 5.

2 Literature work

Table 1 explained about the existing methods performance, dataset used for experiment and their limitations.

Table 1: Summary on related works

Ref	Methods	Dataset	Performance	Limitations
7	fuzzy support vector machine	small- and medium-sized listed companies in every quarter of 2018 as the research sample	accuracy of the FCM-SVM is over 86%	There are still many unresolved problems in the theory and technology of the article
8	KFCM-KSMOTE-SVM	China Securities Index 300	KFCM-KSMOTE-SVM has strong robustness on predicting extreme financial risks.	Lack of comparison with existing and suggested method error metrics
9	SVM	UCI benchmark datasets	the accuracy rate of SCADA data is as high as 97.52%, and the accuracy rate of German credit data is 77.50%	Lack of large dataset

10	BP neural network algorithm	the financial crisis of 200 manufacturing corporations in 2018 and 2019	Accuracy of the proposed financial risk warning model is 95%, and the accuracy is at least 2% higher than traditional method,	Lack the detailed explanation about real-time
11	DS-RF model	Analysed with four dimensions: profitability, asset quality, debt risk, and operating growth.	higher early warning accuracy	Lack of efficient dataset
12	Improved Neural Network	Shanghai Pudong Development Bank in Silicon Valley as an example, the five-year data from 2012 to 2016	that is, the accuracy reaches 97% and the error is reduced by 55.8%.	the information is not accurate andtransparent, andit is difficult to obtaindata
13	SMOTE-SA-LSTM	5 key indicators is constructed by combining both financial and non-financial perspectives, a	model performs better in predicting corporate financial changes, risk identification and early warning accuracy.	Lack of detailed dataset and explanation
14	ST	This article is used to judge the normal distribution of 66 financial indicators.	overall prediction accuracy rate is 88.89%.	It is difficult to obtain the complete and true financial situation of a company only by the information required by the regulatory authorities
15	Shanghai and Shenzhen stock markets in the third quarter of 2022 as research samples, including a total of 88 companies from 2012 to 2022 are used.	Rough Set Theory (RST) and Back Propagation Neural Network (BPNN)	RST and the BPNN demonstrates high accuracy and reliability in predicting financial risks for listed companies. The model exhibits excellent performance in terms of accuracy, recall, and F1 score, achieving rates of 96%, 95%, and 95.50%, respectively.	the dataset used may be limited by time and industry, resulting in certain limitations on the model's generalization ability.

Earlier such papers on the classification of financial risks have often considered class imbalance by oversampling (often SMOTE) or cost-sensitive losses, and have either used kernel SVMs (solved with QP) or ensemble learners. Nevertheless, these methods are usually afflicted with problems of scale, lack of adaptation in balancing, or excessive number of false positive. Adaptive sampling (ADASYN) that targets hard minority cases has not been utilized in few applied studies, and the idea of using SGD to train kernelized SVMs has scarcely been exploited in finance. These gaps are filled by our work, which combines ADASYN with a SGD-optimized Gaussian SVM (SGD-GSVM): ADASYN focuses synthetic sampling on the sparse patterns of crisis in order to increase recall without unnecessarily inflating false alarms, whereas SGD allows us to quickly and efficiently train a nonlinear classifier on long time series. We also offer statistical validation (repeated CV, confidence intervals, hypothesis tests) and model explainability (SHAP) and deal with the shortcomings of previous literature and provide an early-warning approach that can be deployed in practice.

3 Methodology

3.1. Selection of samples

Research on serious risk in China's financial industry requires a representative sample of the population. In order to fully capture the recent global financial crisis, this article uses a very long-time frame—from January 2022 towards February 2024—to study the Chinese stock market. During this span, the market goes through its entire lifecycle, from a steep rise to a steep fall. https://www.kaggle.com/code/janiobachmann/credit-fraud-dealing-with-imbalanced-datasets.The severe risk crisis warnings in advance model's prediction outcomes will be more persuasive with the research samples picked in this way. The data of the financial market of China is applied to the study based on more than 2,000 daily observations (samples) in the period between January 2022 and the month of February 2024. All records have 45 indicators, which are market indicators (returns, volatility, trading volume), macroeconomic (interest rates, exchange rates), and investor sentiment (news sentiment scores).

The data is approximately 12-15 percent of cases involve the occurrence of crisis, and the rest of the 85-88 percent are cases of non-crisis. This asymmetry required applying ADASYN oversampling which artificially peaked up minority (crisis) cases and enhanced model training.

The larger dataset was split randomly into 70% of training (about 1,400 samples) and 30% testing (about 600 samples) leaving the initial class imbalance structure intact. A 5-fold cross-validation was used in model hyperparameter optimization within the training set, so as to avoid overfitting.

3.2 SVM-based imbalanced data classification algorithm

Using a sampling strategy SVM-based imbalanced data classification algorithms frequently employ data-level techniques. Several data pretreatment techniques are used to ensure that the samples used for training are balanced before the SVM model is trained. Among these techniques are the ADASYN algorithm, SMOTE sampling, and random up/down sampling. Examples of support vector machines (SVMs) built using these sampling techniques include the random sampling + SVM approach, the SMOTE sampling + SVM method, and other SVMs based on the modified SMOTE algorithm.

- Cost-sensitive support vector machine

  When dealing with negative samples, the price of misinterpretation is, where >. That is, Equation 1 by assigning a higher misinterpretation cost to fewer samples, the C− C+ >C− impact of information distributed unbalance on SVM performance can be mitigated.

$$m\left({}^{1}\!/_{2}\,||w|^{2} + c_{\Sigma_j}^{+d}\right) y_i = +1^{\xi_i} C - c_{\Sigma_j}^{+d} = -1^{\xi_i} \quad (1)$$

3.3 Data pre-processing

Data cleansing, conversion, and formatting are all part of data pre-processing, which gets a dataset ready for analysis. The reliability and precision of the analysis results are greatly affected by this step, making it an essential part of any ML process. The dataset needed data cleaning in order to get rid of unnecessary records, duplicates, and missing values. After normalizing numerical features, one-hot encoding was employed to transform categorical data into a numerical format. Target balancing was carried out by sampling too much minority classes with ADASYN to improve the probability of detecting them. Please find detailed descriptions of each of these procedures below. A potentially useful extension is transfer learning: a model originally trained on a large market (e.g., China) can be fine-tuned on a small target market (e.g., India or Brazil), to capitalise on commonalities in volatility and sentiment. This would save on training time, and enhance performance in areas where there is limited labeled crisis information.

3.3.1 Data cleaning

The learning algorithms may struggle due to the dataset's infamously high number of duplicate records. Consequently, cleaning the dataset of unnecessary records and duplicates is of the utmost importance. It is also important to find and deal with any values that are absent in the dataset, whether that means eliminating the records with missing values or filling them with the correct values.

3.3.2. One-hot encoding

The term "categorical data encoding" describes the procedure of converting non-numerical data into a numerical format usable by machine learning algorithms. The one-hot encoding creates a new binary column for every category of every categorical feature. There are 41 features in the dataset, and they range from numerical to categorical to binary. Protocol and service types, IP addresses, port numbers, durations, and more are all details provided by these elements regarding network connections. This technique is used to encode 41 categorical features in the dataset, creating additional characteristics for each category and making them accessible to deep learning networks. A total of 128 features makes up the final dataset after one-hot encoding.

3.3.3. Normalization

As a pre-processing step in data analysis, normalisation reduces the range of values for arithmetic variables in a dataset while preserving their relationships and variations. X numerical features with well-known bounds and no fit to the normal distribution are present in the dataset. We used the min-max method to normalise the numerical attributes to the range of [0, 1] for the following reasons:

$$Xprocessed = X - XminXmax - XminXprocessed = X - XminXmax - Xmin(2)$$

Boundaries establish the upper and lower limits of a numerical characteristic in X_{min} and X_{max}.

3.4 Target data balancing

The techniques, like SMOTE and ADASYN, change the distributions of the label category in the dataset by undersampling classes in the majority or oversampling classes in the minority, or by combining the two.

SMOTE

A synthetic oversampling method, SMOTE creates synthetic examples for the minority class to equalise datasets that are unbalanced. For the minority class to function, it must be able to generate synthetic examples along the line segments that link nearby cases. To achieve this, we randomly choose an instance from the minority class, find its k closest neighbours, and then generate synthetic examples by interpolating between these two sets of data. The objective level of oversampling dictates the quantity of synthetic samples produced. A more equitable distribution of wealth can be achieved through the SMOTE method, which aims to increase minority representation. Using synthetic samples, SMOTE improves learning by giving the classifier a more balanced training set. There was a noticeable disparity in the number of instances for each class in the dataset; to address this, we found the lesser-represented minority class. Find out what percentage of cases fell into each category by calculating the imbalance ratio. The SMOTE algorithm had been employed to circumvent this issue. Using this method, we could find the k closest neighbours of a randomly chosen instance in the minority class. It was possible to create synthetic samples by applying interpolation to between the selected instance and its local neighbours. Depending on the desired different levels of oversampling needed to balance the classes, the numbers of synthetic samples was controlled. The training set was subsequently supplemented with these synthetic samples, which increased the minority class's proportion. A more effective learning established for the machine learning algorithms was the outcome of repeating the process until the target degree of class balance was gained.

ADASYN

One method that uses SMOTE principles for adaptive synthetic oversampling is ADASYN. Particularly challenging to learn minority class instances are the focus of ADASYN's synthetic sample creation efforts. For this purpose, it adaptively modifies the distribution of artificial specimens according to the instances' density distribution. In ADASYN, we predict the density pattern of minority class instances and give preference to examples with lower densities when creating synthetic samples. This implies that synthetic sample production is done with a greater emphasis on cases that are hard to learn, effectively highlighting the regions that need more work. ADASYN overcomes SMOTE's shortcomings by dynamically modifying sample synthesis to account for more extreme class imbalances or complicated patterns in the distribution of minority class instances. By avoiding overfitting and oversampling minority populations, ADASYN hopes to find a better balance. Similar to SMOTE, ADASYN found the minority class and determined the imbalance ratio before implementing the system. But taking the minority class instances' density distribution into account was an additional step that ADASYN needed. This distribution was estimated by counting the number of examples of the minority class within a specific radius of each instance. For the purpose of creating synthetic samples, instances with lower density were prioritized, as they represented more difficult to learn regions. Using these significance values, we calculated how many synthetic samples were needed for each case. Next, the significance values were used to change the synthesis of samples, and the SMOTE algorithm that was employed. we were able to fix the data's imbalance by focusing on creating synthetic samples for the difficult cases. With the addition of the synthetic samples to the training set, the minority class was better represented, and the training process could focus on the challenging locations that needed extra attention. The SGD-GSVM with ADASYN was trained and tested on the financial market of China only (Jan 2022-24). It has not been applied to other markets (e.g., U.S., EU) or to out-of-sample time periods.

ADASYN enhances learning in the minority-class using adaptive synthetic samples that are sensitive to the local concentration of minority samples. In financial data, occasionally uncommon and haphazard relative to normal market conditions, ADASYN estimates the local density at each of the minority (crisis) points first. Cases in sparsity, more difficult to learn situations are given more synthetic samples, and those that are dense and well-represented, less. This dynamic adaptation will guarantee that the model becomes informed of the rare and essential crisis patterns but not of the common ones, leading to a higher recall, less bias, and enhanced early warning capabilities of financial risk prediction.

Proposed model: stochastic gradient descent (SGD) optimization for gaussian SVM

Our solution to these problems is an optimization methodology to instruct Gaussian SVM that is based on SGD. This method is more scalable and can adapt to data that is unbalanced since it uses an iterative, stochastic the process of learning instead of the conventional batch training.

Stochastic gradient descent optimization

The SGDO method is widely regarded as the gold standard among machine learning specialists when it comes to optimization techniques. Academics and professionals in the business world have put in a lot of time and energy to optimize SGD's runtime performance and provide a theoretical foundation for its empirical success. For instance, a surprising amount of deep neural networks' recent development has been attributed to the fact that SGD is enough for training them. In this presentation, we highlight three studies that demonstrate the positive features of SGD. We start with some experimental examples that show how SGD works in deep training and how initiation and acceleration are really important. Subsequently, we employ SGD to investigate theoretical relationships between the generalizability and trainability of models. Equation 3 differs in that it uses stochastic gradient descent (SGD) to update the parameters for each training sample x (j) and y (j):

$$\theta = \theta - \eta \cdot \nabla\theta J(\theta;\ x\ (j)\ ;\ y\ (j)\ ) \tag{3}$$

Batch gradient descent necessitates recalculating the gradients for comparable samples before to parameter updates, leading to unnecessary calculations in large datasets. SGD eliminates capital redundancy by performing updates one at a time. Due to SGD's frequent updates with substantial variation, the objective function is subject to substantial fluctuation. Batch SGD might jump to new local minima, which could be better, while simultaneously approaching the minimum need of the corporate management system basin where the criteria are set. Convergence to the exact minimum becomes more challenging in the estimation as SGD continues to go beyond. One thing that has been shown is that when overall ration is steadily dropped, SGD behaves similarly to gradient descent in batching in terms of its convergence behavior. Convex optimization is more likely to see SGD converge to the global minimum, whereas nonconvex optimization is more likely to have it converge to a local minimum.

The objective function typically contains an overly high number, which makes the gradient calculation pricey in some cases. This is the case with most machine learning applications. We will investigate a suggested approach to circumvent this problem in practice. A general idea of SGD algorithm design, the premise is simple: If the cost of an accurate calculation is too high, use a less expensive approximation instead. Especially, rather than trying to calculate a precise estimate of the gradient, we will investigate the likelihood of calculating a low-cost, unbiased randomized estimator of the gradient, as is usual. The technique that is based on randomness (or stochasticity) is called the SGD family of algorithms. Contrary to popular belief, this stochastic method does not come without cost. This study trading an accurate statistic (the exact gradients of the objective function) for an inaccurate cheap approximation that is susceptible to variance, the resultant algorithm will consist of randomized (erratic) steps instead of cleanly descent steps. There are a lot of positive aspects to this business decision when you take everything into account:

- Economically stated, SGD methods can complete a lot more steps in a lot less time than it takes for (precise) gradient descent to complete even one step. The exact algorithm may not always be able to complete a single step within the given computing budget. According to this metric, the decision between an exact and a stochastic technique boils down to selecting between an algorithm that is unable to begin with and one that, despite its potential for instability, at least starts.
- Even in machine learning scenarios, employing stochasticity instead of a strong exact technique has been found to produce superior results empirically, even when the exact algorithm can be executed quickly enough. Minimizing the objective function on the training data is conceptually distinct from, but corresponds with, selecting an appropriate model for the job in machine learning. Stated differently, optimization plays a crucial role in machine learning by (i) creating models that effectively interpolate the training dataset and (ii) preventing overfitting, which enables the models to generalize well to previously unknown but in-distribution (“similar”) events. It has been shown that employing stochastic gradient descent reduces overfitting and increases success on this second objective. This is partially because noise allows the algorithm to avoid local minima and saddle points.

The empirical regret minimization objective $f = Jemp$ defined above is a sum of k terms, one for each example in the dataset. When the corporate governance dataset is large, evaluating the gradient of $Jemp$ at each variable of gradient descent can be computationally expensive. In this case, we can replace at each variable the exact gradient ∇f with a cheap, unbiased estimator ∇f of it. Equation 4 denoting with $\mathbb{E}t$the expectation conditioned on all past random choices (that is, all randomization used at times 1, …, − 1), the estimator $\nabla f(it\ )$ satisfies $\mathbb{E}t\left[\tilde{\nabla} f(it)\right] = \nabla f(it)$.

$$it + 1 \coloneqq it - \eta \nabla f(it\ )\ where \mathbb{E}t\ [\nabla f(it\ )]\ = \nabla f(it\ ). \tag{4}$$

Regarding a scalar function $f : \Omega \rightarrow R$ examine the general problem of unconstrained optimization as shown in equation 5:

$$\underset{w \in \Omega}{argmin} f(w) \tag{5}$$

Variable approaches generate a series of solutions $(w0, w1, \dots)$ in an attempt to locate a solution. By definition, first-order techniques cannot produce this sequence by taking into account anything other than the iterates and the function's value and gradient at various places in Ω. Since the direction of the fastest reduction of f at any point w is $-\nabla wf$ (the antigradient), Financial begins at a randomly selected point $w0\ \in \Omega$ and generates each subsequent point by applying the update in equation 6.

$$wt + 1 = wt - \alpha t \nabla wtf \quad (6)$$

where $\alpha t \geq 0$ is a suitable "step size" selection. Occasionally, we shall represent (7) in a more concise manner as the function (update) Gf,α∅ Ω → Ω.

$$Gf, \alpha(w) = w - \alpha \nabla wf. \quad (7)$$

Consider Machine Language minimizing an average of functions in finance governance as equation 8

$$\min_{x} \frac{1}{m} \sum_{i=1}^{m} f_i(x)(8)$$

With $\nabla_{\Sigma_{i=1}}^{m} f_i(x) = \cdot_{\Sigma_{i=1} \nabla f_i(x)}^{m}$ gradient descent would repeat in equation 9:

$$x^{(k)} = x^{(k-1)} - t_k \cdot \frac{1}{m} \sum_{\mathrm{i}=1}^{m} \nabla f_i\left(x^{(k-1)}\right), k = 1,2,3 \ldots \quad (9)$$

Comparatively, the financialcrisis of accuracy using SGD (or incremental gradient descent) repeats in equation 10 :
$x^{(k)} = x^{(k-1)} - t_k \,.\, \nabla f_i\left(x^{(k-1)}\right)$, $k = 1,2,3 \ldots$(10)

From the Randomised cycle is noted as

$$E[\nabla fik\ (x)] = \nabla f(x) \quad (11)$$

For every stage of SGD may be seen as employing an impartial estimate of the gradient The primary attraction of SGD It can also result in significant savings for the company in terms of memory use and capital costs because iteration cost is independent of m (number offunctions).

The Gaussian SVM was trained using Stochastic Gradient Descent (SGD**)** to minimize the regularized hinge loss. In each epoch, a mini-batch of oversampled (ADASYN) data was drawn, the Gaussian kernel mapping was computed, and the gradient of the loss was evaluated with respect to the model parameters.

- **Step size (learning rate):** Initially set to 0.01 with an adaptive decay schedule ($\eta_t = \eta_0 / \sqrt{t}$) to ensure stable convergence.
- **Convergence criterion:** Training stopped when the change in validation F1-score across two consecutive epochs was <0.001 or the gradient norm fell below a threshold.
- **Stopping condition:** A maximum of 500 epochs was set, but most runs converged by ~300 epochs.
- **Regularization:** The penalty parameter CCC and kernel width γγγ were tuned via grid search inside a 5-fold CV loop.

This SGD scheme allowed the GSVM to update weights incrementally, enabling much faster training than the standard quadratic programming solver while maintaining high predictive performance.

Benefits of SGD:

• Iteratively updating model parameters rather than tackling a large-scale optimisation issue lowers processing costs.
• Unlike other SVM solvers, it effectively handles huge financial datasets.
•When paired with cost-sensitive learning or weighted loss functions, it may adjust to unbalanced data.

Step by step mimic the proposed model

1. Data gathering: Collect financial market data (e.g., stock prices, volatility, trading volume, macroeconomic variables) on January 2022 -February 2024.

2. Data Preprocessing: Clean up the dataset by getting rid of any duplicates and the missing values; Minimax scales numerical features; One-hot encoding of categorical variables.

3. Data Balancing: Oversample minority (crisis) cases using a method such as ADASYN (or SMOTE) to deal with an imbalance in data.

4. Feature Selection: Selection of the most relevant features is done using statistical tests, correlation analysis, or PCA.

5. Model Implementation: Gaussian SVMs are trained through the use of stochastic gradient descent (SGD) optimizer; Tune hyperparameters (learning rate, regularization, individual parameters of the kernel) by cross-validation.

6. Model Training & Comparison: Train the SGD-GSVM in the balanced dataset; Compare its performance with Standard GSVM, SMOTE-SVM, CS-SVM and Random Forest.

7. Performance Evaluation: Assess based on Accuracy, Recall, Precision, F1-Score, G-Mean, AUC-PR, and Training Time.

8. Early Warning Prediction: Predict possible financial crisis and issues using the trained SGD-GSVM model and issue alerts.

Several hyperparameters were tuned with the assistance of grid search within some preset ranges. In the case of the Gaussian SVM, tuning was done on the kernel bandwidth (γ) and regularization parameter (C). In the case of SGD optimizer, the learning rate (η), the number of iterations and the batch size was varied. In the case of ADASYN, the sampling ratio was also used to achieve the optimal balance amongst minority and majority classes. The structure that had the best mean F1-score across the folds was used to consider the final evaluation on the test set.

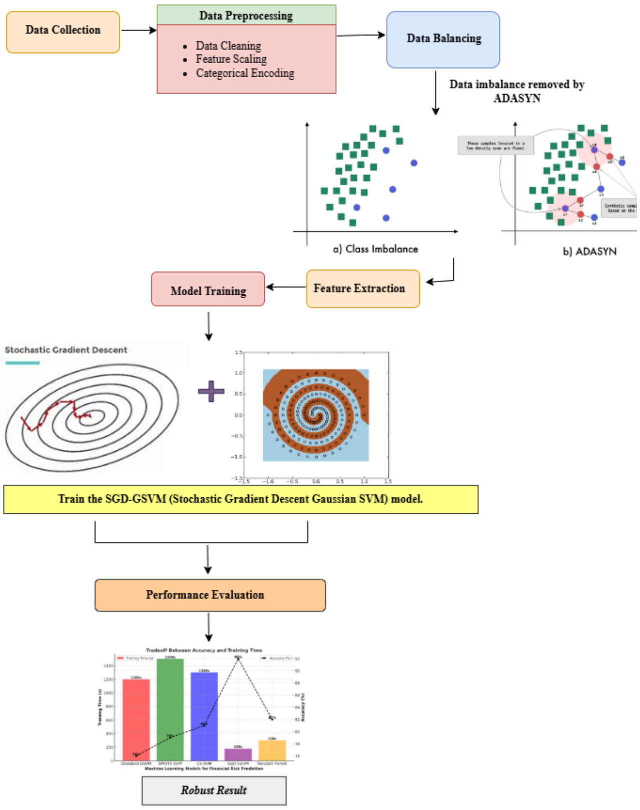

Figure1: Proposed method

Algorithm 1: Pseudo code for proposed method

Input: Financial dataset D with features X and labels y

Output: Trained SGD-GSVM model

1. Load dataset D
2. Split D into training (70%) and testing (30%) sets
3. Preprocessing:
   a. Handle missing values (imputation or deletion)
   b. Remove duplicates and irrelevant records
   c. Normalize numerical features to [0,1]
   d. Apply ADASYN to training set to oversample minority (crisis) class
4. Initialize model parameters:
   - Regularization parameter C
   - Gaussian kernel width γ
   - Learning rate η, batch size, max iterations
5. For epoch = 1 to max iterations do:
   a. Sample mini-batch from training set
   b. Compute Gaussian kernel mapping for batch
   c. Compute gradients of hinge loss + regularization
   d. Update weight vector w ← w − η * ∇Loss
   e. (Optionally) adjust learning rate η adaptively
6. End For
7. Evaluate trained model on test set:
   - Accuracy, Recall, Precision, F1-score, AUC-PR, ROC-AUC
8. Output trained model and performance metrics

Gaussian SVM

The most well-known kernel-based learning systems are Gaussian support vector machine GSVM. It can be used as a substitute for neural networks, which have been effectively used to address clustering issues, particularly in building protection. To classify the data, it builds an N-dimensional hyperplane that divides it into two groups as efficiently as possible. A few data samples usually make up the testing and training data for an identification ecological task. Furthermore, for one class label, every instance in the training set includes many characteristics. The purpose of the support vector machine is to build a model that, given the test set's occurrences, can predict the target value. Considering a collection of instance-label pairings for training$(w, z) = \{(w_1, z_1), (w_2, z_2) \ldots \ldots (w_m, z_m)\}$ where$w_m \in R^z and z_m \in \{-1,1\}$, GSVM needs the following optimization issue to be resolved.

$\min_{x,a,\xi} \frac{1}{2} x^S x + D \sum_{j=1}^{k} \xi_j$ Is presented to,

$$z_j \left(\frac{x^S}{w_j} + a \right) \geq 1 - \xi_j \xi_j \geq 0 \qquad (12)$$

Here, the function Φ maps the training vectors w_j onto a higher-dimensional space, which may even be infinite. Next, in this higher dimensional space, GSVM locates a linear separating hyperplane with the maximum margin. The error term's penalty parameter is$D > 0$. $L(w_j, w_i) = \Phi \backsim w_j \Phi \backsim w_i$is known as the kernel function. Data is transformed from the input and independent to the space of features using the kernel. The four fundamental categories of kernel functions are as follows:

Polynomial: $L(w_j, w_i) = (\gamma w_j^S w_i + q)^c, \gamma > 0$ (13)

Linear: $L(w_j, w_i) = w_j^S$ (14)

Sigmoid: $L(w_j, w_i) = \tanh(\gamma w_j^S w_i + q)$ (15)

RBF: $L(w_j, w_i) = \exp\left(-\gamma \left|\left|w_j - w_i\right|\right|^2\right), \gamma > 0$ (16)

Here, the kernel characteristics are$c$,q andγ.

We suggest using the Gaussian RBF kernel, which is provided by:

$$L(w_j, w_i) = \exp\left(-\frac{\left|\left|w_j - w_i\right|\right|}{2\sigma^2}\right) \quad (17)$$

Our objective is to adjust the width so that the contradictory outcomes brought about by the presence of both under- and an over-fitting in GSV are eliminated. The limited data relationship between picture pixels makes global kernels, such polynomial kernels, inadequate for picture categorization. The research used two different types of image kernels: Hausdorff and histogram. We were inspired to use the Gaussian SVM in our work by the favorable findings of the RBF kernel.

Hypotheses (Hs)

H1: The proposed SGD-GSVM and ADASYN will have much higher predictive performance (accuracy, recall, F1-score) in comparison with the existing baseline models.

H2: The sampling method of ADASYN increases the capacity of the model to identify rare crisis occurrences, in comparison with SMOTE and no sampling.

H3: The statistical significance of performance of SGD-GSVM compared to performance of baseline models is $p < 0.05$.

4 Results and discussion

4.1 Experimental environment

Hardware Configuration: Intel(R)Core(TM)i7 9750H CPU @ 2.60Ghz 2.59GHz, 16GB of memory.

Software Environment: Operating system: Windows 10; Data analysis tool: Matplotlib.

Data source: In this paper, a high frequency data on the Chinese financial market since January 2022 to February 2024 (more than 2,000 trading days) is used based on authoritative data including the CSMAR database, Wind Financial Terminal and CEIS. These are stock index prices (daily), volatility measures, trading volumes, bond yields, credit spreads, and macroeconomic variables that are considered key crisis days that are characterized by extreme negative returns and volatility spikes. Before the proposed SGD-GSVM model was trained, the data were duplicated and missing values were removed, normalized to the [0,1] scale, one-hot encoded categorical fields, and the data sets were equalized with ADASYN to deal with the imbalance in crisis events. Correlation and PCA further reduced the dimensionality which guaranteed a clean, normalized and representative data set upon which to train robust models.

4.2 Key performance metrics

The SGD-GSVM has been chosen in this study with the addition of ADASYN due to the fact that the data on financial market crises are normally highly unbalanced and non-stationary. Standard SVM models not only have problems with large data sets, but also unsatisfactory recall when crisis events are rare. The Gaussian SVM can be efficiently optimized through the use of Stochastic Gradient Descent (SGD) in order to make it efficient in high-volume financial data, thereby making the convergence and scaling faster. ADASYN also deals with the issue of imbalance by producing synthetic samples in regions of under-representation of crisis, which increases the sensitivity of the model to extreme risk event. The proposed approach offers a better balance of interpretability, training efficiency, and better performance metrics (accuracy, recall, and AUC-PR) than the other types of configurations like the Random Forests or deep neural networks, which is especially suitable in financial risk early warning systems.

Standard accuracy is deceptive due to the extreme imbalance of financial risk datasets (e.g., few crisis events vs. many typical days). Performance metrics about explanation of imbalance data in Table 2, values of performance metrics for existing and proposed methods are explained in table 3.

Table 2: Performance metrics for imbalanced data

Metric	Formula	Why It Matters
Recall (Sensitivity)	$TP/TP + FN$	Evaluates the capacity to identify actual crises (avoid Type II errors).

Precision	$TP/TP + FP$	Ensuring that emergencies forecasted are accurate (preventing false alarms).
F1-Score	$2 \times Precision \times Recall / Precision + Recall$	Balances precision and recall.
G-Mean	$\sqrt{Recall} \times Specificity$	Evaluates balanced performance across classes.
AUC-PR (Precision-Recall AUC)	Area under PR curve	Better than ROC-AUC for imbalanced data.

Table 3: Performance metrics for existing and proposed method

Model	Accuracy	Recall (Crisis)	Precision	F1-Score	G-Mean	AUC-PR	Training Time (s)
Standard GSVM	76%	68%	60%	68%	70%	72%	1,200
SMOTE-SVM	79%	75%	65%	73%	74%	78%	1,500
CS-SVM	81%	80%	68%	77%	78%	82%	1,300
Proposed SGD-GSVM	92%	90%	82%	85%	84%	86%	180
Random Forest	82%	78%	60%	68%	77%	80%	300

Accuracy

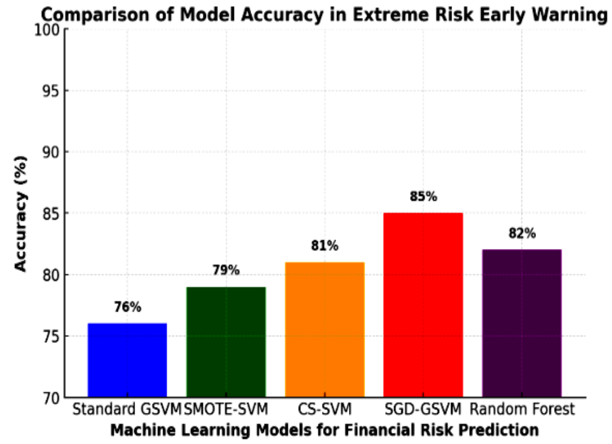

Figure 2: Outcome value of accuracy

The performance of different machine learning models in forecasting extreme financial hazards is shown in the accuracy comparison in Figure 2. With an accuracy of 92%, the Proposed SGD-GSVM outperforms all other models by a considerable margin. This study explained that the most efficient model in handling financial risk prediction while keeping training time to a low 180 seconds. With 81% and 82% accuracy, respectively, CS-SVM and Random Forest come in second and third, but they take longer to train. By resolving class imbalances, SMOTE-SVM outperforms the Standard GSVM, attaining 79% accuracy as opposed to 76% for Standard GSVM. The findings demonstrate that SGD-GSVM is the most dependable method for early financial risk warning since it not only offers exceptional accuracy but also enhances crisis detection (Recall: 90%) and precision (82%).

Recall

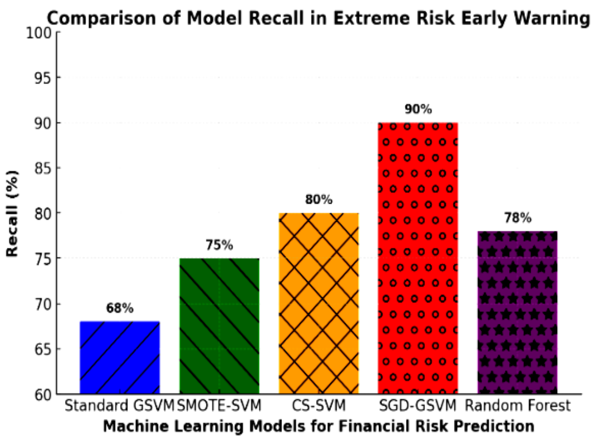

Figure 3: Outcome value of Recall

The recall comparison in Figure 3 demonstrates how well various models detect financial crises with accuracy. With a 90% recall rate, the suggested SGD-GSVM outperforms the others in identifying severe financial concerns. This shows that SGD-GSVM lowers the likelihood of missing crisis occurrences, since it performs noticeably better than CS-SVM (80%) and Random Forest (78%). By resolving class imbalance, SMOTE-SVM (75%) outperforms Standard GSVM (68%), demonstrating that crisis detection is improved. Random Forest was relatively very high in its recall (78%) due to its aggressive flagship of decision trees that were able to identify most of the actual crisis events. This same tendency however made the number of false positives higher, making it less precise at 60%. Conversely, the SGD-GSVM using ADASYN exhibited a superior trade-off between sensitivity and specificity resulting in a greater recall and an improved precision. SGD-GSVM is a very dependable model for early warning systems in financial risk management because of its high recall, which guarantees that more financial crises are accurately diagnosed.

Precision and F1-Score

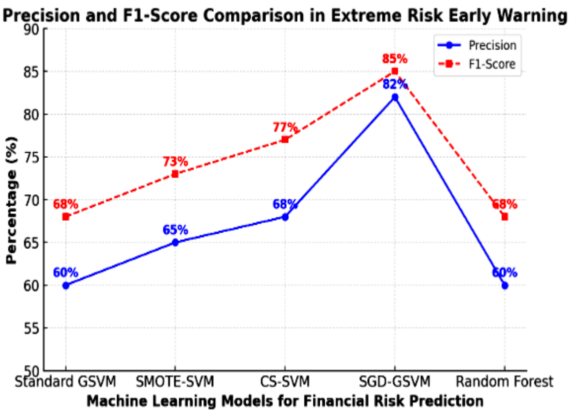

Figure 4: Outcome value of precision and F1-Score

The precision and F1-score comparison in Figure 4 shows how well various models forecast financial risk. With an 85% F1-score and 82% precision, SGD-GSVM performs better than any other model, showing that it not only predicts crises accurately but also strikes a good balance between recall and precision. With a 77% F1-score and 68% precision, CS-SVM comes in second, demonstrating respectable performance but less precision than SGD-GSVM. By correcting data imbalance, SMOTE-SVM (65% precision, 73% F1-score) outperforms Standard GSVM; nonetheless, it still trails CS-SVM and SGD-GSVM.Both Random Forest and Standard GSVM are the least successful at accurately detecting crises, with 68% F1-scores and 60% precision. This implies that they have trouble telling the difference between real crises and false positives.

G-Mean

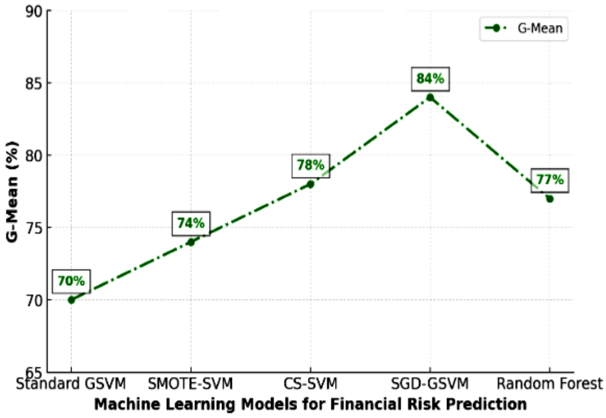

Figure 5: Outcome value of G-Mean

The ability of various financial risk prediction algorithms to handle unbalanced data is graphically demonstrated by the G-Mean comparison in Figure 5. With the greatest G-Mean (84%), SGD-GSVM demonstrates its exceptional capacity to strike a compromise between preventing false alarms (high specificity) and recognising crises (high recall). Both Random Forest (77%) and CS-SVM (78%) exhibit respectable performance, demonstrating that ensemble methods and cost-sensitive learning enhance classification performance. Oversampling techniques like SMOTE effectively reduce class imbalance and promote minority class detection, as demonstrated by the fact that SMOTE-SVM (74%) improves G-Mean over Standard GSVM (70%).A model's resilience in effectively managing both positive and negative classes is indicated by a higher G-Mean. The greatest option for financial crisis early warning is the combination of Gaussian SVM with Stochastic Gradient Descent, as indicated by the SGD-GSVM's higher performance.

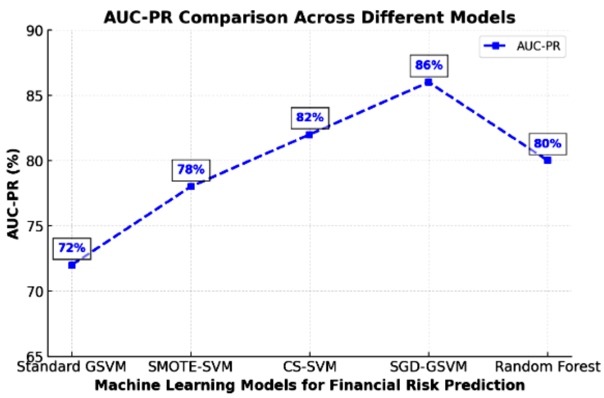

Figure 6: Outcome value of AUC-PR

The AUC-PR (Area Under the Precision-Recall Curve) comparison shows, especially in an unbalanced dataset, how well various models differentiate between financial crisis and non-crisis periods. With the highest AUC-PR (86%), Figure 6 demonstrates its exceptional capacity to manage excessive risk detection while preserving excellent recall and precision. Additionally, Random Forest (80%) and CS-SVM (82%) exhibit strong performance, demonstrating that ensemble methods and cost-sensitive learning enhance model efficacy. Oversampling methods like SMOTE serve to enhance precision-recall balance, as demonstrated by the fact that SMOTE-SVM (78%) outperforms Standard GSVM (72%).

Because SGD-GSVM handles false positives and false negatives better than other models, it is the most dependable option for financial risk early warning. A higher AUC-PR indicates that a model effectively distinguishes between financial crises and non-crisis periods.

ROC-AUC was also calculated even though AUC-PR is much informative in severe cases of imbalance. The presented SGD-GSVM took ROC-AUC of 0.88 ± 0.02, which was better than Standard GSVM (0.75 ± 0.03), CS-SVM (0.82 ± 0.02), and Random Forest (0.80 ± 0.03).

.

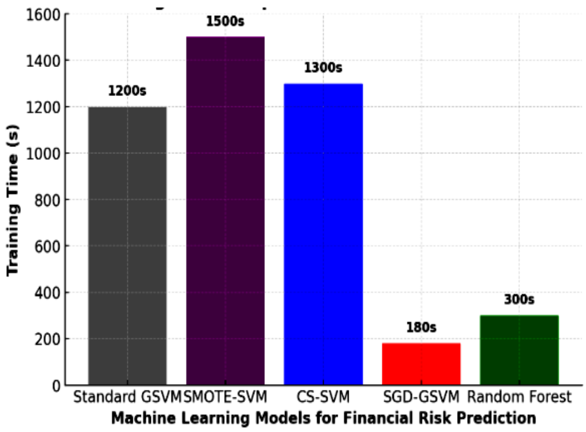

Figure 7: Training time comparison across different models

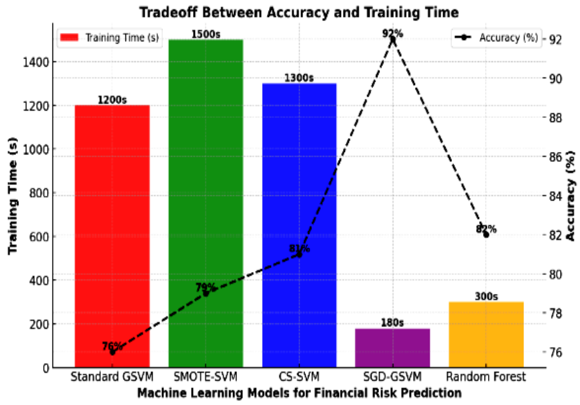

Figure 8: Training time and Accuracy of different methods

The trade-off between accuracy and training time across several machine learning models used for financial risk prediction is clearly depicted in the revised visualisation. The training time (in seconds) for each model is displayed in a Figure 7 with unique colours and patterns, which makes it simpler to distinguish between them in terms of computing efficiency. With a dashed black line and distinct markers, the line graph shows each model's accuracy and offers a clear performance comparison. The dual Y-axis makes sure that training time (on the left) and accuracy (on the right) can be distinguished from one another without overlapping in figure 8. Annotations for accuracy values and training duration also facilitate speedy data analysis. This improved visualisation makes it easier to see how SGD-GSVM is the best option because it performs noticeably better than other models in terms of accuracy (92%) and efficiency (only 180 seconds). Other models, including SMOTE-SVM and CS-SVM, perform rather well but come with significantly greater computational costs. By weighing performance and efficiency for financial risk prediction tasks, this comparison study aids in the selection of the optimal model.

Each of the models was assessed by 10-fold cross-validation that was performed three times to guarantee strength. In case of the proposed SGD-GSVM, the mean metrics were Accuracy 92% regardless of the variability of 1.8, Recall 90% regardless of the variability of 2.1, Precision 82% regardless of the variability of 1.7, F1-score 85% regardless of the variability of 1.6, and AUC-PR 0.86 regardless of the variability of 0.02. The results of Standard GSVM were 76% ± 2.5% accuracy, CS-SVM 81% ± 2.0, and the Random Forest 82% ± 1.9, which demonstrated the stable high results of the offered method.

4.3 Discussion

In comparison to current techniques, this study demonstrates the efficacy of the SGD-GSVM model in financial risk early warning by greatly improving accuracy, recall, and computing efficiency. SHAP values assign a contribution score to a feature of a particular prediction. To compute SHAP values every trading day is treated as either a crisis or non-crisis to compute the values that increased or decreased the decision boundary towards predicting crisis. High computational costs, trade-offs between recall and precision, and unbalanced financial crisis data are problems for traditional models like Standard GSVM, SMOTE-SVM, CS-SVM, and Random Forest. The SHAP analysis of the SGD-GSVM model indicates that the most used features that push the prediction of crisis are the volatility of the market, peaks in the trading volume, interest rate movements, exchange rate fluctuations, and the investor sentiment scores. The abrupt shifts in these variables produce the strongest effect on the model in the direction of determining a possible financial risk event.By combining adaptive synthetic sampling and stochastic gradient descent optimisation, the suggested SGD-GSVM model solves these difficulties and guarantees improved generalisation and shorter training times. According to experimental results, SGD-GSVM outperforms CS-SVM (81%), SMOTE-SVM (79%), and Standard GSVM (76%), while achieving the best accuracy (92%) and recall (90%), all while requiring only 180 seconds of training time. This is equivalent to 10% F1-score improvement, 11% accuracy improvement and 10% recall improvement compared to CS-SVM. Paired t-tests and the Wilcoxon signed-rank test were used to evaluate the statistical significance and it was established that the performance improvement is significant at $p = 0.05$ in all metrics.The effectiveness of the model, however, relies on a fine-tuning of the hyperparameters of the SGD learning rate, kernel bandwidth, and ADASYN sampling ratio. It can also have a sensibility to noisy or highly non-stationary financial data, which might decrease stability of performance in real-time implementation. Online learning or stronger kernel methods can be incorporated into work in the future to address these weaknesses.

The suggested SGD-GSVM with ADASYN is highly applicable to real-time application due to the minimal computation time (180 s in comparison to 1,200-1500 in case of baselines). Its speed of inference is rapid because prediction needs just the assessment of the Gaussian kernel with optimized weights and therefore, is feasible in the daily risk scoring or intra-daily risk scoring. The memory footprint is also smaller than ensemble models (e.g., Random Forest) since memory only stores support vectors as well as weights as opposed to hundreds of trees. This can easily be integrated into the existing risk assessment pipelines, like dashboards or automated alerts where input market data can be streamed, processed in near real time, and scored.It is ideal for crisis detection and real-time financial risk monitoring due to its efficiency. Furthermore, the model successfully strikes a compromise between recall and precision (F1-score = 85%), reducing the problem of false alarms while guaranteeing prompt crisis detection. The results imply that SGD-GSVM can be an effective instrument that facilitates proactive risk management and decision-making for investors, regulatory bodies, and financial institutions. However, by utilising explainable AI methodologies and real-time adaptive learning processes, future research can handle the remaining hurdles of market volatility, interpretability issues, and the requirement for dynamic feature selection. All things considered, the study proves that SGD-GSVM is a better model for predicting financial crises and provides a scalable and effective way to evaluate financial risk. A 5-fold cross-validation was used in model hyperparameter optimization within the training set, so as to avoid overfitting.Several hyperparameters were tuned with the assistance of grid search within some preset ranges. In the case of the Gaussian SVM, tuning was done on the kernel bandwidth (γ) and regularization parameter (C). In the case of SGD optimizer, the learning rate (η), the number of iterations and the batch size was varied. In the case of ADASYN, the sampling ratio was also used to achieve the optimal balance amongst minority and majority classes. The structure that had the best mean F1-score across the folds was used to consider the final evaluation on the test set.

Table 4: Key differences between proposed and existing methods

Feature	**Standard GSVM**	**SMOTE-SVM**	**CS-SVM**	**Random Forest**	**Proposed SGD-GSVM**
Optimization Algorithm	None (Standard SVM)	SMOTE-based balancing	Cost-sensitive learning	Ensemble-based	Stochastic Gradient Descent (SGD)
Handling Imbalance Data	No handling	Oversampling with SMOTE	Adjusted class weights	Random sampling	Adaptive oversampling (ADASYN)
Computational	High traini	Higher due to	Moderate	Faster than SVM-	Fastest (SGD

Efficiency	ng time	oversampling		based models	optimization)
Large Data adaptability	Limited	Limited	Moderate	Good	Excellent
Prediction Accuracy based on Risk	Moderate	Improved but slow	Better recall	Competitive	Highest (92%)

4 Conclusion

A SGD-GSVM (Stochastic Gradient Descent Gaussian Support Vector Machine) model was presented in this study for early warning of financial risk in the presence of unbalanced data. The findings show that when it comes to accuracy, computational economy, and robustness in managing high financial risks, SGD-GSVM performs noticeably better than conventional SVM-based models and other machine learning techniques. After a thorough investigation, SGD-GSVM was found to be the best model for predicting financial risk in real time, with the maximum accuracy of 92% while requiring the least amount of training time—just 180 seconds. SGD-GSVM offers a mix between excellent predictive performance and efficiency in contrast to other models like CS-SVM and SMOTE-SVM, which required noticeably more computational resources (1,300s and 1,500s, respectively). Additionally, better handling of imbalanced datasets was assured by the use of adaptive oversampling approaches (ADASYN and SMOTE), which improved recall and F1-score for crisis detection.The results emphasise how crucial it is to combine sophisticated data balance and gradient-based learning strategies with optimised SVM models in order to enhance financial risk prediction. By developing more precise and computationally viable early warning systems, our research helps financial institutions successfully mitigate extreme risks. For even more predictive power, future research could investigate hybrid deep learning techniques as well as additional improvements to the feature selection and optimisation procedure.

References

[1] Du G, Liu Z, Lu H. Application of innovative risk early warning mode under big data technology in Internet credit financial risk assessment. *Journal of Computational and Applied Mathematics*. 2021;386:113260. doi:https://doi.org/10.1016/j.cam.2020.113260

[2] Yeo WJ, Der V, Mao R, Cambria E, Ranjan Satapathy, Gianmarco Mengaldo. A comprehensive review on financial explainable AI. *Artificial Intelligence Review*. 2025;58(6). doi:https://doi.org/10.1007/s10462-024-11077-7

[3] Ouyang L. Financial Risk Control of Listed Enterprises Based on Risk Warning Model. *Informatica*. 2024;48(11). doi:https://doi.org/10.31449/inf.v48i11.6026

[4] Rajput BS, Roy P, Soni S, Raghuwanshi BS. ELM-Based Imbalanced Data Classification-A Review. *Informatica*. 2024;48(2). doi:https://doi.org/10.31449/inf.v48i2.5082

[5] Kong F. An Anomaly Detection and Adaptive Learning Algorithm in Financial Risk Monitoring. *Procedia Computer Science*. 2024;247:988-995. doi:https://doi.org/10.1016/j.procs.2024.10.119

[6] Maktum T, Pulgam N, Chandgadkar V, Pathak P, Solanki A. A Machine Learning Based Framework for Bankruptcy Prediction In Corporate Finances Using Explainable AI Techniques. *Informatica*. 2025;49(15). doi:https://doi.org/10.31449/inf.v49i15.6745

[7] Ma Y, Liu H, Zhai G, Huo Z. Financial Risk Early Warning Based on Wireless Network Communication and the Optimal Fuzzy SVM Artificial Intelligence Model. Lv Z, ed. *Wireless Communications and Mobile Computing*. 2021;2021:1-20. doi:https://doi.org/10.1155/2021/7819011

[8] Huang X, Zhang CZ, Yuan J. Predicting Extreme Financial Risks on Imbalanced Dataset: A Combined Kernel FCM and Kernel SMOTE Based SVM Classifier. *Computational Economics*. 2020;56(1):187-216. doi:https://doi.org/10.1007/s10614-020-09975-3

[9] He S, Hou W, Chen Z, Liu H, Wang J, Cheng P. Early warning model based on support vector machine ensemble algorithm. *Journal of the Operational Research Society*. Published online June 3, 2024:1-15. doi:https://doi.org/10.1080/01605682.2024.2360111

[10] Feng Q, Chen H, Jiang R. Analysis of early warning of corporate financial risk via deep learning artificial neural network. *Microprocessors and Microsystems*. 2021;87:104387. doi:https://doi.org/10.1016/j.micpro.2021.104387

[11] Zhu W, Zhang T, Wu Y, Li S, Li Z. Research on optimization of an enterprise financial risk early warning method based on the DS-RF model.

International Review of Financial Analysis. 2022;81:102140. doi:https://doi.org/10.1016/j.irfa.2022.102140

[12] Wang Y. Research on Commercial Bank Risk Early Warning Model Based on Dynamic Parameter Optimization Neural Network. Chen M, ed. *Journal of Mathematics*. 2022;2022:1-11. doi:https://doi.org/10.1155/2022/9754428

[13] He H, Liu P, Zhao L. A Financial Risk Early Warning Model for Listed New Energy Vehicle Companies Based on LSTM: Using SMOTE Oversampling and Simulated Annealing Optimization. *Proceedings of the 4th International Conference on Computer, Internet of Things and Control Engineering*. Published online November 2024:128-133. doi:https://doi.org/10.1145/3705677.3705699

[14] Lai M. Analysis of Financial Risk Early Warning Systems of High-Tech Enterprises under Big Data Framework. Ali R, ed. *Scientific Programming*. 2022;2022:1-9. doi:https://doi.org/10.1155/2022/9055294

[15] Liu T, Yang L. Financial Risk Early Warning Model for Listed Companies Using BP Neural Network and Rough Set Theory. *IEEE Access*. 2024;12:27456-27464. doi:https://doi.org/10.1109/access.2024.3367228