Comparative Analysis of Support Vector Machine Performance Across Geophysics, Finance, and Medical Diagnostics

Ciobanu Sergiu-Tudor

Abstract

This report presents a comparative performance evaluation of the Support Vector Machine (SVM) across three distinct domains: medical diagnostics, financial risk assessment, and geophysical warning. The methodology involved rigorous data preprocessing, outlier mitigation, and Stratified K-Fold Cross-Validation to ensure reliability. The SVC was optimized by tuning the regularization parameter to establish complex decision boundaries across the feature sets. Results demonstrated high effectiveness in the medical diagnosis and geophysical domains, achieving final Accuracies of 0.9918 and 0.8846, respectively. Nevertheless, the model showed significant limitations in the financial sector, achieving a low Recall (0.66). This weakness shows that the SVC struggles with imbalanced data and situations where the target classes are difficult to separate.

1. Introduction

Extracting information from raw datasets has become a new standard in recent years, providing a significant source of knowledge by enabling the understanding of various domains and their complexities. Bringing out their valuable insights from the data is the key to uncovering potential and hidden patterns that are not so perceptible to humans. This process comprises several steps and methodologies, such as data collection, data storage, data cleaning, processing, and analysis, but most importantly by choosing a machine learning model that can interpret all values and calculate their best outcome. Machine learning (ML), which successfully enables systems to make predictions and informed decisions without being explicitly programmed, is the crucial final stage.

To address this problem, the scope of this study is to employ a supervised ML technique, trained on labeled data where the input features correspond to known output labels. Therefore, we will delve into the analysis of performance and application of the Support Vector Machine (SVMs) for three different real-world classification tasks. Prominent for their strong classification proficiency, SVMs exhibit a remarkable ability to find optimal decision boundaries, effectively separating data points even in complex, high-dimensional, or non-linearly separable scenarios. This unique characteristic, alongside their strong generalization capability from training to unseen data, makes them adaptable and robust tools in machine learning.

The main purpose of this report is to conduct a comparative performance evaluation of the SVM model across three classification challenges originating from distinct, unrelated sectors of application, for instance, Geophysical Disaster Warning, Financial Risk Assessment, and Medical Diagnostics. This diverse experimental design fulfills the requirement for testing SVM's generalizability against heterogeneous data structures and domain-specific analytical challenges. The three real-world problems selected for this study are: checking patient biomarkers to verify the presence of the Hepatitis C Virus (HCV), predicting credit card defaulting within the banking sector, and developing an accurate system for earthquake alert prediction, necessary for early warning and disaster planning.

Following this introductory section, Section 2, Literature Review and Theoretical Framework, will detail the mathematical foundations of the Support Vector Machine, including the concepts of the optimal hyperplane, margin maximization, and kernel. Section 3, Data and Experimental Methodology, will provide an in-depth overview of the preprocessing stage, applied individually to the HCV, Credit Card Default, and Earthquake datasets. Section 4, Experimental Results, will present the performance metrics for the SVM on each of the three datasets. Section 5, Discussions of Results and Performance Analysis, will interpret the findings, compare the SVM's relative success across the three domains, and discuss the implications of the prediction errors. Finally, Section 6 will present the Concluding Remarks.

2. Literature Review and Theoretical Framework

2.1 General Machine Learning Context

Classification is a foundational, extremely common task in supervised machine learning. Especially, the aim is to construct a predictive model that can correctly assign an unknown data point to its corresponding category, or class label, based on the observed features. By learning the intricate connections between these features and the labels, the model gains the necessary competence to successfully generalize its knowledge when facing new, real-world scenarios. This capability to generalize and make informed decisions is what makes classification absolutely vital across fields like medical diagnostics, financial risk assessment, and global disaster warning systems.

2.2 Support Vector Machine Architecture

Introduced by Vladimir Vapnik in 1995 [1], SVMs stand out as a set of related eager inductive supervised learning methods used for both classification and regression. The algorithm aims to find a hyperplane that best separates data points into two different classes. The challenge is that there are infinitely many possible hyperplanes that satisfy this condition. Therefore, the goal of the SVM is to identify the hyperplane that achieves the maximum margin. This margin is defined as the largest distance separating observations belonging to one class from those of the other class.

The data points that fall exactly on this margin are termed support vectors, as these specific cases alone define the unique boundary solution. The separating boundaries geometry depends on the number of input features: with two predictor variables, the boundary is a line; with three, it is a plane; and with more than three predictors, it is universally referred to as a separating hyperplane. The SVM adheres to the Structural Risk Minimization (SRM) principle, which equips the model with a superior ability to generalize its findings effectively to unseen data.

A principal advantage of the SVM lies in its ability to handle data that is not linearly separable. This is achieved through the use of a Kernel Function, a computational method that takes input data and transforms it into the required form for processing. By mapping the data into a higher-dimensional feature space, the kernel effectively makes the previously non-linear data separable by a hyperplane [3]. There are multiple types of kernels; however, for the purpose of this research, the Radial Basis Function (RBF), often called the Gaussian Kernel, will be briefly reviewed. The RBF is particularly effective because it implicitly maps the data into an infinite number of dimensions, making it well-suited for situations where class boundaries are highly non-linear and irregular. It is also important to note that while powerful, the SVM algorithm can be computationally expensive with larger imbalanced datasets.

2.3 Evaluation and Validation

2.3.1 Cross-Validation Techniques

The dataset must be carefully partitioned into training and testing sets. This splitting process can generally be executed using two methods: subject-wise division and record-wise division. Subject-wise division ensures that all data points belonging to any single individual or subject are allocated entirely to either the training set or the testing set, thus ensuring the independence of subjects between the two sets. Contrary, record-wise division involves splitting the dataset randomly without considering whether records from the same subject might be shared between the training set and the testing set. When dealing with repeated measurements from the same entities, subject-wise division is critical because employing record-wise division allows the model to see a subject's data in both training and testing sets, resulting in data leakage and misleadingly high, biased performance estimates.

A common method for robust validation is K-fold cross-validation (CV). This technique divides the entire dataset into k equal-sized subsets, or folds. The model is then trained iteratively k times. In each iteration, $k-1$ folds are used for training, and the single remaining fold is reserved for validation. This process ensures that every data point serves as part of the validation set exactly once, providing a more comprehensive measure of the model's transferability across the data. A more advanced and often necessary variant of this is Stratified K-fold cross-validation [2]. This technique is particularly useful for datasets exhibiting class imbalance. Stratified CV ensures that each of the k folds maintains the same class distribution as the original dataset. Within K-fold cross-validation, a class could have an unequal distribution, with certain folds having more instances of that class than others.

2.3.2 Evaluation Metrics

To assess the performance and effectiveness of the ML model, a set evaluation metrics is necessary. The most reliable method for visualizing the relationship between the model's predictions and the actual class labels is by constructing a Confusion Matrix. This table summarizes the classification results, typically with predicted labels placed on one axis and actual labels on the other. It provides a detailed analysis of performance by quantifying four outcomes:

- True Positives (TP): The number of positive instances correctly classified as positive.
- True Negatives (TN): The number of negative instances correctly classified as negative.
- False Positives (FP): The number of negative instances incorrectly classified as positive.
- False Negatives (FN): The number of positive instances incorrectly classified as negative.

From this matrix, several key metrics are derived. Accuracy measures the proportion of all correct predictions made by the model out of the total number of predictions. Precision quantifies the proportion of instances predicted as positive that were actually positive. Recall measures the proportions of all actual positive instances that were correctly identified by the model. F1-Score represents the harmonic mean of Precision and Recall, giving equal importance to both metrics. A high F1-Score indicates that the model is successfully balancing both precision and recall, while a low score suggests poor performance in one or both areas.

2.4 Related Work

The selection of the Support Vector Machine (SVM) for this comparative study is supported by its demonstrated effectiveness in similar critical real-world classification tasks. In medical diagnostics, Safdari et al. [4] applied various machine learning models to classify suspected Hepatitis C Virus (HCV) infection, where the SVM achieved a high Accuracy of 94.59% alongside a balanced Recall of 87.27% and a Precision of 88.64%. Furthermore, in financial risk assessment, Bhandary et al. [5] confirmed the SVM's utility in Credit Card Default Prediction, noting an Accuracy of 81% and a Precision of 70% among the six models tested. Although the SVM's Recall was lower at 23% in the financial context, the model's overall performance in these distinct areas, diagnostics and finance, justifies its selection as a robust model capable of handling the data and varying complexities inherent in this research.

3. Data and Experimental Methodology

This section outlines the practical execution of the research, detailing the specific procedures used to prepare the raw data and the exact configuration of the experimental environment.

3.1. Data Description and Preprocessing

Before the raw data can be introduced to the machine learning algorithm, it must be converted into the required format. Preprocessing is a critical step that ensures the data used for modeling is of high quality by systematically resolving common issues such as missing values, redundancy, and data noise.

3.1.1. Dataset A: [Earthquake Alert Prediction Dataset]

3.1.1.1 Dataset Description

Sourced from Kaggle, the Earthquake Alert Prediction Dataset comprises 1300 samples, each row representing a seismic event. To address the class imbalance problem, the authors of the dataset employed a method called SMOTE utilized for advanced oversampling techniques to enhance the dataset records, in order to achieve an optimal class balance. The dataset contains 6 feature columns: 5 numerical input measurements or characteristics that describe an earthquake, such as magnitude, depth, community reports for the intensity (cdi), ground shaking intensity (mmi), or significance (sig). The numbers scientists record about how big, how deep, and how intense the earthquake was. The target categorical variable named “alert” represents the severity of the earthquake’s alert level (green, yellow, orange, red)

3.1.1.2 Data Analysis and Preprocessing

Initial inspection confirmed that all feature columns are numeric and have no missing values, so there is no need to do any imputations. Another observation is that the target variable has exactly 325 samples for each category, confirming that the dataset is perfectly balanced. From the correlation matrix, it can be interpreted that depth is weakly negatively correlated with cdi (-0.37) and moderately negatively correlated with mmi (-0.57), indicating that deeper earthquakes tend to have lower reported and instrumental intensities. Cdi and mmi have a strong positive correlation (0.68), which makes sense because both measure earthquake intensity from community reports and instruments.

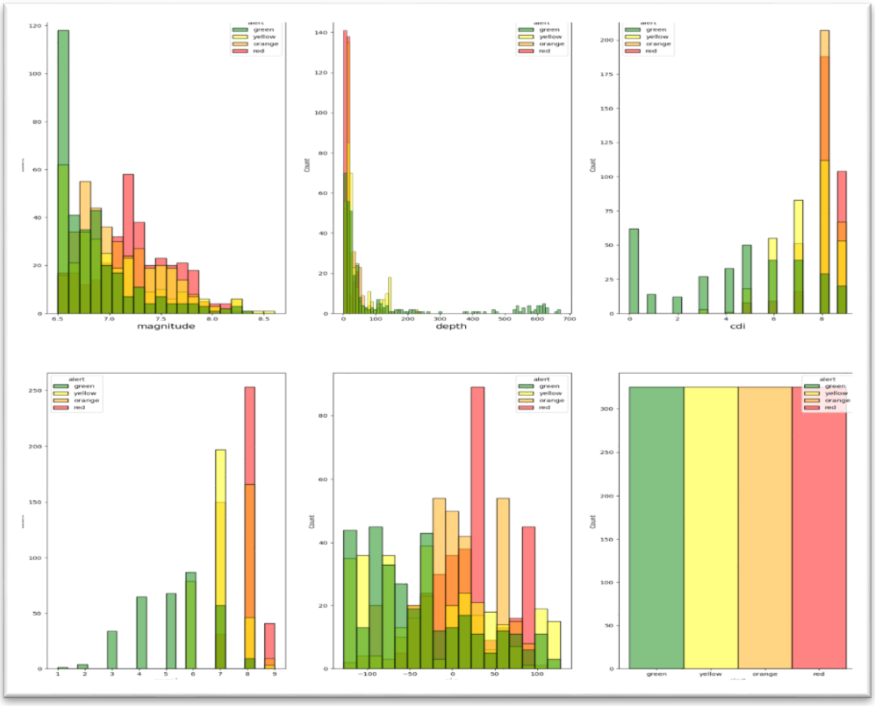


Figure 1. Distribution of Key Seismic Features by Alert Category

3.1.2. Dataset B: [Default of Credit Card Clients Dataset]

3.1.2.1 Dataset Description

The Default of Credit Card Clients Dataset was sourced from the UCI Machine Learning Repository [6]. This collection comprises records for 30,000 credit card customers in Taiwan and covers a six-month window, specifically from April 2005 to September 2005. The information details a variety of factors, including demographic profiles, credit limitations, payment histories, and monthly bill statements. The central focus of the data is on individuals who were either significantly overdue on their payments or failed to settle their accounts entirely.

The dataset includes 25 unique features, such as client IDs, credit limits, sex, level of education, marriage status, ages, billing statement balances, and payment amounts. The target variable indicates whether or not the customer is anticipated to miss their next payment. Regarding the feature LIMIT_BAL (Amount of given credit in NT dollars), the data exhibits significant variance: while the average value is 167,484 NTD, the maximum value is 1,000,000 NTD, resulting in an unusually large standard deviation. These extreme values were intentionally retained as they are not anomalies but rather represent vital information, necessary for the model to accurately assess the full range of financial risk.

3.1.2.2 Data Preprocessing

An initial discovery was that no feature contained missing data. However, the dataset presented a few structural inconsistencies. Initially, 68 records were excluded as they were identified as duplicates. Specifically, some categorical features contained undocumented category numbers that had to be addressed. Records associated with these uninformative categories were excluded from the dataset, resulting in the final dataset containing 29,566 records. Following this cleaning step, the relevant categorical features were transformed using One-Hot Encoding to convert them into a numerical format suitable for the Support Vector Machine (SVM) algorithm.

3.1.3. Dataset C: [HCV Dataset]

3.1.3.1 Dataset description

This data utilizes The UCI HCV dataset [7]. For this experiment, the research focuses on decision-making based on patient health status. The dataset comprises demographic and biochemical laboratory test data for 615 patients at different stages of HCV infection. It includes thirteen features, consisting of twelve input features such as Age, Sex, various proteins (e.g., Albumin), compounds (e.g., Bilirubin), and enzymes (e.g., Aspartate Aminotransferase, Alkaline Phosphatase), alongside other blood biometrics. The Category feature functions as the target class and encompases five diagnostic labels: 0 (Blood Donor), 1 (Suspect Blood Donor), 2 (Hepatitis), 3 (Fibrosis), and 4 (Cirrhosis). To convert this into a binary classification problem for identifying patients as healthy or unhealthy, the first and second categories were assigned as healthy, and the remaining three categories were assigned as diagnosed.

3.1.3.2 Data Preprocessing

Several biochemical markers, including ALB, ALP, ALT, CHOL, and PROT, exhibited notable levels of missing data, a factor critical to robust data preparation and ensuring model resilience. To address these absent values—specifically the recorded missingness in ALB (0.16%), ALT (0.16%), PROT (0.16%), CHOL (1.63%) and ALP (2.93%) a mode imputation methodology was applied. For every feature with missing data, the most frequent value (the mode) calculated from the available instances was used for replacement.

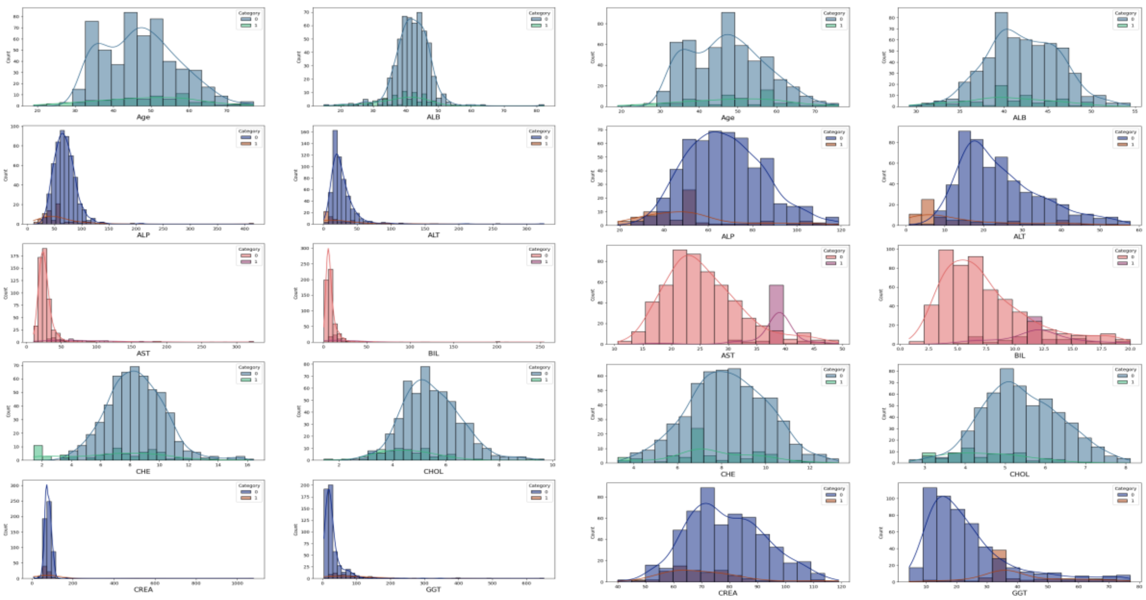


Figure 2: Comparison of Feature Distributions Before and After Outlier Mitigation

Outliers were identified using the Interquartile Range (IQR) method, employing a standard threshold multiplier of 1.5 to establish the bounds. The number of detected outliers varied across the features: Age (1 outlier), ALP (10 outliers), CHOL (12 outliers), CREA (12 outliers), PROT (20 outliers), CHE (24 outliers), ALB (27 outliers), ALT (36 outliers), BIL (47 outliers), AST (64 outliers), GGT (65 outliers). To mitigate the disproportionate impact of extreme values while mentaining the central tendency of the data, outliers were managed by applying class-wise median imputation, computed from non-outlier instances belonging to the respective diagnostic category. This involved replacing the detected outliers with the median value calculated exclusively from the non-outlier instances belonging to that specific diagnostic category.

3.2 Experimental Setup and SVM Implemenetation

The model implementation procedure was standardized across all three datasets to ensure a fair comparative analysis.

First, Standardization was applied to all numerical features within every dataset. This crucial step centered the feature values around zero with a variance in the same order of magnitude, which is mandatory for the Support Vector Machine (SVM) model, as its performance is highly dependent on distance calculations. Before being fed into the model, the data was partitioned into training and testing sets using an 80%–20% ratio.

The classification task was handled by the Support Vector Classifier (SVC), utilizing the implementation from the scikit-learn library in Python. The SVC was configured with the Radial Basis Function (RBF) kernel and the default kernel coefficient, gamma('scale'). The initial value for the regularization parameter C was set to 1.

For the HCV Data, Stratified K-fold Cross-Validation was employed during the training phase. This approach was essential to ensure that each fold maintained a proportional representation of the class outcomes, thereby mitigating the risk associated with the dataset's inherent class imbalance. During hyperparameter tuning, it was determined that the optimal value for the regularization parameter C varied for the financial and geophysical domains: the best performance for the Credit Card Default Dataset was achieved with C=10, while the Earthquake Alert Dataset yielded the best values at C=25.

4. Experimental Results

This section presents the factual outcomes of applying the optimized Support Vector Classifier (SVC) to the three distinct datasets. The results are presented individually for each domain to facilitate a clear comparative assessment.

For each experiment, the model's performance on the dedicated test set is rigorously evaluated. The primary focus of the analysis will be a detailed presentation of the Confusion Matrix for each dataset, which visually summarizes the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) rates. Additionally, to provide insight into the model's decision-making process, the Feature Importances for each model will be presented and interpreted. This analysis helps to explain which specific biomedical, financial, or seismic features were most critical to the SVM's predictive output in each unique domain.

This table (Table 1) summarizes the performance of the optimized Support Vector Classifier (SVC) on the test sets for the HCV, Credit Card Default, and Earthquake Alert datasets. The results immediately indicate exceptional performance on the HCV Data, while the Credit Card Default task presented the most significant challenge in achieving balanced metrics.

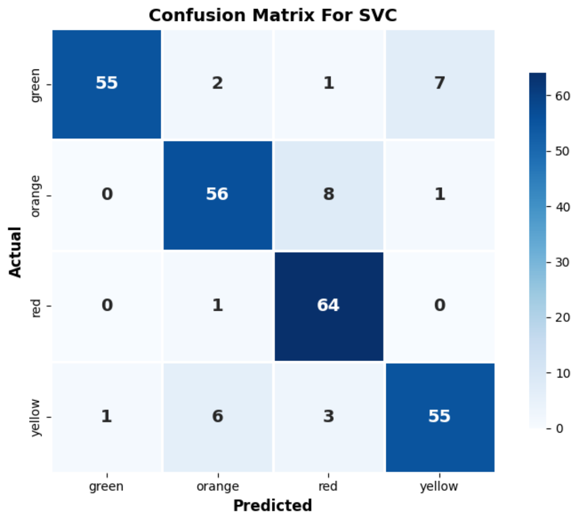
Dataset	Accuracy	Precision	Recall	F1-Score
HCV Data	0.9918	0.9919	0.9918	0.9917
Credit Card Default Data	0.8192	0.3515	0.6666	0.4603
Earthquake Alert Data	0.8846	0.8904	0.8846	0.8847

Table 1: Comparative Performance Metrics of the Support Vector Classifier (SVC) Across Three Datasets

4.1 Performance on the Earthquake Alert Data

The Confusion Matrix (Figure 3, Left) indicates that the SVM model achieved highly effective multi-class classification, successfully distinguishing between the four alert levels (green, yellow, orange, red). The model demonstrated a strong capability to correctly predict the most severe alert level, with 64 instances of the 'red' alert correctly classified (True Positives). The primary classification errors are minor: for instance, 7 'green' alerts were mistakenly classified as 'yellow', and 6 'yellow' alerts were misclassified as 'orange'. The high concentration of correct predictions along the main diagonal validates the previously reported high Accuracy for this domain.

The Feature Importance bar plot (Figure 3, Right) reveals the relative influence of the seismic features on the model's predictions. The analysis indicates that mmi (0.34) and sig (0.25) are the most critical features, meaning the model relies heavily on these two factors to determine the alert levels. Cdi (0.23) and magnitude (0.21) are less influential but still contribute significantly to the predictions. Depth (0.13) plays the least significant role in the classification. The model prioritizes combined significance and instrumental intensity over magnitude or community-reported intensity. This aligns with the idea that alert levels depend on multiple interacting seismic factors, not just one measurement.



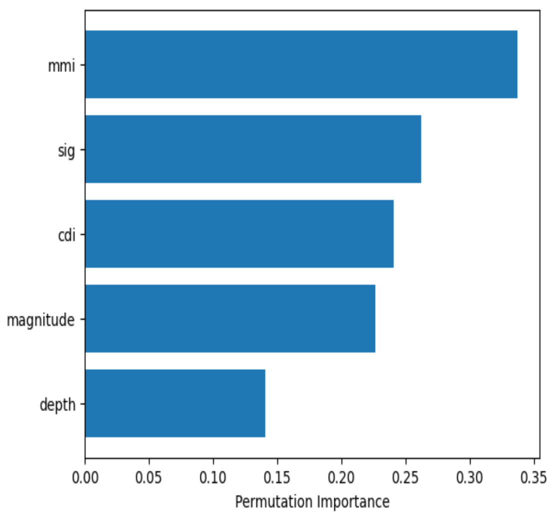


Figure 3: SVM Confusion Matrix (Left) and Permutation Feature Importance (Right) for Earthquake Alert Data

4.2 Performance on the Credit Card Default Data

The Confusion Matrix (Figure 4, Left) highlights the significant challenge presented by the imbalanced nature of the credit default task. The model correctly identified 4,389 clients who did not default (True Negatives), a high number which contributes substantially to the overall high Accuracy of 0.8192 (as seen in Table 1). However, the model struggled significantly with the minority class: only 456 actual defaulters were correctly identified (True Positives), while a high number of 841 actual defaulters were incorrectly predicted as non-defaulters (False Negatives). This substantial rate of False Negatives explains the corresponding low Recall (0.6666) and Precision (0.3515).

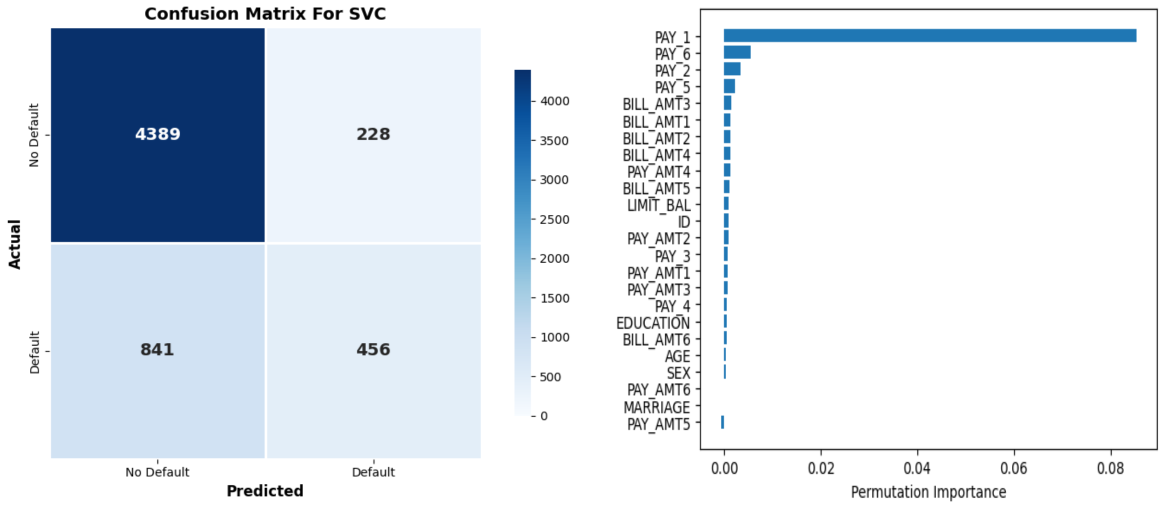


Figure 4: SVM Confusion Matrix (Left) and Permutation Feature Importance (Right) for Credit Card Default Data

The Permutation Feature Importance (Figure 4, Right) clearly indicates that the model relies almost entirely on the Repayment Status variables, confirming that recent financial behavior is the most critical factor in predicting future default. Specifically, the features tracking repayment status from September down to May (PAY_1, PAY_6, PAY_2, PAY_5) are the top predictors. This demonstrates that the history of past payment delays is vastly more influential than any other available data. Conversely, demographic features such as AGE, SEX, EDUCATION, and MARRIAGE, along with the client's total LIMIT_BAL (credit limit), were assigned very low importance by the model. In summary, the SVM operates primarily as a sophisticated classifier based on recent payment behavior, achieving high accuracy by correctly classifying non-defaulters, but struggling to achieve a balanced performance across all metrics due to the difficulty inherent in distinguishing between complex defaulting clients and the dominant non-defaulting clients.

4.3 Performance on the HCV Data

The Confusion Matrix (Figure 5, Left) confirms the exceptional performance of the model in this domain, which resulted in the highest overall accuracy seen in the study. The model achieved near-perfect classification on the test set, correctly identifying 108 healthy patients with zero misclassifications (False Positives). For the minority class, the model correctly identified 14 diagnosed patients (True Positives). Crucially for a medical diagnostic task, the SVM produced only 1 False Negative, meaning only a single diagnosed patient was incorrectly labeled as healthy.

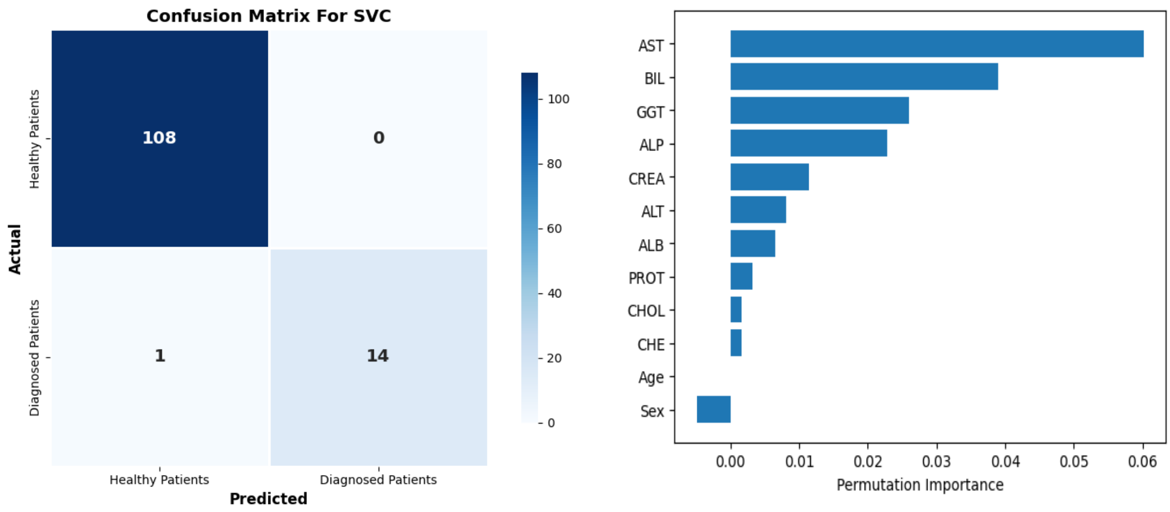


Figure 5: SVM Confusion Matrix (Left) and Permutation Feature Importance (Right) for HCV Data

The Permutation Feature Importance (Figure 5, Right) reveals that the model's high accuracy is heavily dependent on specific biochemical markers, which aligns closely with clinical expectations. The amount of Aspartate Aminotransferase (AST) in a patient's blood contributes significantly more to the model's decision-making than any other feature. Other highly influential markers include BIL (Bilirubin), GGT (Gamma-Glutamyl Transferase), ALP, and CREA. The amount of certain proteins and amino acids in the patient's blood is a strong indicator of the patient's risk of liver disease, and the model learned to prioritize these specific combinations. Conversely, demographic factors like Age and Sex were found to be the least influential features in the diagnostic classification task.

5. Discussion of Results and Performance Analysis

5.1 High Efficacy and Adaptability

The SVM model demonstrated remarkable success in both the HCV Data and the Earthquake Alert Data domains. This success was achieved by strategically adjusting the regularization parameter C, which instructs the optimization on the penalty level for misclassifying training examples: a large C value, such as the C=25 used for the Earthquake Data, forces the model to choose a smaller margin, more complex hyperplane to reduce errors, thus significantly raising accuracy from 0.77 to 0.8846. Furthermore, the critical role of the Stratified K-Fold Cross-Validation approach was also validated. This technique was employed specifically for the HCV Data to overcome the lower accuracy 0.92 seen in standard data splits, ensuring reliable results for the minority classes.

5.2 Limitations

The primary limitations of the SVM model were clearly demonstrated when applied to the Credit Card Default Data. Even with a moderate overall Accuracy of 0.8192, the model showed poor performance in predicting actual defaults, achieving a low Recall 0.66. This discrepancy confirms that using Accuracy is misleading for this imbalanced problem, as the score is artificially high due to the correct classification of the majority (non-defaulter) class.

This failure arises from the model's results difficulty in handling noisy data where target classes exhibit overlapping properties in the feature space. Due to the nature of the algorithm, the search for a perfect decision boundary can result in the model finding several local optima instead of the global best solution, especially with high-dimensional data. The SVM is highly sensitive to the size and balance of the data, a core weakness of the soft margin optimization. Even when the Credit Card dataset was downsampled, the model produced almost identical results. This suggests that the model's inability to achieve better metrics is due to the low separability of the underlying financial features, rather than a failure to scale efficiently with large dataset size.

6. Conclusion

This experimental study successfully performed a comparative evaluation of the Support Vector Machine classifier across three highly distinct real-world classification problems: medical diagnostics (HCV Data), financial risk assessment (Credit Card Default Data), and geophysical disaster warning (Earthquake Alert Data). The initial phase of data description and processing was critical, involving rigorous steps like subject-wise division, careful outlier mitigation via class-wise median imputation, and feature standardization. This meticulous preparation, which revealed valuable insights into feature structures before modeling, ensured the reliability of the final experimental setup. The initial phase of data description and processing was critical, involving steps like subject-wise division, outlier mitigation via class-wise median imputation, and feature standardization.

The results confirmed the SVM's strong resilience and high effectiveness in domains characterized by clear feature separability. The model achieved near-perfect performance on the HCV Data and showed significant improvement on the Earthquake Alert Data, a success due to the effective tuning of the regularization parameter C and the use of Stratified K-Fold Cross-Validation. However, the study also exposed the SVM's major limitations when dealing with high-noise, imbalanced datasets, specifically in the Credit Card Default task. Here, the model's low Recall demonstrated that high Accuracy can be a misleading metric, confirming the weakness of the soft margin optimization problem in non-separable financial feature spaces.

In summary, the Support Vector Classifier turned out to be a really effective model, especially when the data was clean enough to let it find a clear boundary. This made it perfect for the high-precision tasks like the HCV diagnostics and the Earthquake prediction. However, the study also showed its main weakness: the SVC is very sensitive to imbalanced data and low separability. That's why it struggled with the complex Credit Card Default problem compared to other models that might handle noisy financial data better.

Bibliography/References

[1] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

[2] Bergstra, J. and Bengio, Y. (2012) Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13, 281-305.

[3] Evgeniou, T. & Pontil, M. In *Support Vector Machines: Theory and Applications* (eds Paliouras, G. et al.) 249–257 (Springer, Berlin, 2001).

[4] Safdari, R., Deghatipour, A., Gholamzadeh, M., & Maghooli, K. (2022). Applying data mining techniques to classify patients with suspected hepatitis C virus infection. Intelligent Medicine, 2(4), 193–198.

[5] Bhandary, R., & Ghosh, B. K. (2025). Credit Card Default Prediction: An Empirical Analysis on Predictive Performance Using Statistical and Machine Learning Methods. Journal of Risk and Financial Management, 18(1), 23.

[6] I-Cheng Yeh. (2009). Default of Credit Card Clients [Dataset]. UCI Machine Learning Repository

[7] Unknown. (2020). HCV data [Dataset]. UCI Machine Learning Repository.

Acknowledgement:
This work is the result of my own activity, and I confirm I have neither given, nor received unauthorized assistance for this work.
I declare that I used generative AI or automated tools in the creation of content or drafting of this document.

I utilized a generative AI tool exclusively for correcting grammatical errors, improving vocabulary, and rewriting content to establish a more formal and academic tone.

All the text impacted by generative AI has been explicitely marked as such in this document.