iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** II **Month of publication:** February 2024

DOI: https://doi.org/10.22214/ijraset.2024.58549

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

Feature Selection Techniques for Enhancing Credit Card Fraud Detection Performance: A Hybrid Metaheuristic Approach Using Nature-Inspired Algorithms

Pravalika Sure[1], Satwik Pandey[2], Tanmay Parnami[3], Gaurang[4], Aryan Saxena[5]
[1]*Undergraduate Student, (Computer Science Engineering), VNRVJIET, Hyderabad, Telangana, India*
[2]*Undergraduate Student, (Computer Science Engineering), ADGIPS, Delhi, India*
[3]*Undergraduate Student, (Computer Science Engineering), ADGIPS, Delhi, India*
[4]*Undergraduate Student, (Computer Science Engineering), PDM University, Delhi, India*
[5]*Undergraduate Student, (CSE), Vellore Institute of Technology, Vellore, India*

Abstract: ***The surge in credit card usage has led to a parallel increase in fraudulent transactions, necessitating advanced detection systems. Traditional methods encounter challenges with high-dimensional and imbalanced datasets. This paper proposes a comprehensive approach integrating metaheuristic algorithms and deep learning techniques to enhance fraud detection accuracy. The first contribution, the Rock Hyrax Swarm Optimization Feature Selection (RHSOFS) algorithm, draws inspiration from the collective behaviors of rock hyrax swarms to effectively select relevant features from high-dimensional datasets. RHSOFS identifies an optimal subset of features critical for fraud detection through supervised machine learning. Complementing this, the second contribution leverages a hybrid deep learning model. Beginning by organizing transactional data and constructing a Logical Graph of Behavior Profile (LGBP) to abstract transaction details, the Modified Butterfly Optimization Algorithm (MBOA) selects important features from the dataset. The hybrid model, integrating Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), establishes rational connections between transactional characteristics, enhancing detection performance. To evaluate the approach, comparative efficiency analyses against existing methods, including Differential Evolutionary Feature Selection (DEFS), Genetic Algorithm Feature Selection (GAFS), Particle Swarm Optimization Feature Selection (PSOFS), and Ant Colony Optimization Feature Selection (ACOFS), are conducted. Results demonstrate the superiority of the approach in terms of both reliability and recognition rate, validated through rigorous statistical testing. The proposed hybrid approach signifies a significant advancement in credit card fraud detection systems, offering enhanced accuracy and efficiency in combating fraudulent activities amidst the growing complexity of transactional datasets.***

Keywords: ***Metaheuristic algorithms, Rock Hyrax Swarm Optimization, Modified Butterfly Optimization Algorithm, Recurrent Neural Networks (RNN), Artificial Bee Colonies (ABCs), Detection performance***

I. INTRODUCTION

Financial institutions in the contemporary era are leveraging a plethora of new services to expand the spectrum of economic opportunities accessible to customers. These services encompass a wide array, including credit cards, Automated Teller Machines (ATMs), online banking, and mobile banking facilities [1]. Furthermore, with the rapid evolution of e-commerce, the utilization of credit cards has evolved into a pivotal and convenient facet of modern financial existence. Two primary strategies employed in credit card fraud detection are scam analysis and customer behavioral assessment. Initially, transaction-level supervised categorization is utilized [2]. Drawing insights from historical data, these methods discern transactions as either fraudulent or regular. Leveraging this dataset, classification algorithms can prognosticate the status of novel records [3]. Techniques such as rule induction, decision trees, and neural networks are harnessed to construct binary models. Referred to as abuse identification, this technology has demonstrated high efficacy in uncovering most fraudulent schemes in the past [4]. The alternative approach involves employing unstructured methodologies based on account behavior. When a transaction deviates from a user's typical behavior, its legitimacy may come under scrutiny.

This is grounded in the understanding that unauthorized users are unlikely to exhibit patterns consistent with genuine account holders. To accomplish this objective, the authentic behavioral patterns of each account holder must be initially delineated, followed by the identification of deceptive behaviors [5]. By juxtaposing new behaviors against this established model, fraudulent activities can be identified. Profiles may encompass account activity data, including merchants, transaction amounts, locations, and timestamps. This methodology is commonly referred to as "anomaly detection".

Two primary approaches to feature selection (FS) are filtering and wrapping. In a filtering approach, the FS technique remains independent of the learning algorithm, while in a wrapper approach, it depends on the learning algorithm. Filtering methods generally exhibit faster processing times compared to wrapping methods. However, filtering techniques are prone to inductive biases inherent in the learning algorithms employed for classifier construction. On the other hand, wrapping approaches involve a higher processing overhead as they utilize learning algorithms to assess subsets of features. Despite this, wrapper strategies often yield superior accuracy compared to filtering methods [13,14,15]. A preprocessing step is typically employed in the filtering technique to identify optimal features.

However, a fundamental flaw of the filtering approach is its failure to consider the impact of the selected feature subset on the performance of the induction algorithm [16].

Introduced by Kohavi and John in 1997, the wrapper methodology offers a straightforward and efficient solution to the challenge of feature selection. In the wrapper approach, a feature subset selection algorithm encapsulates the induction process, treating it as a black box. This algorithm searches for an optimal subset using the induction process as part of its evaluation function, without requiring knowledge of the underlying algorithm [17]. This encapsulation simplifies the feature subset selection process, making it independent of specific algorithmic details. In this study, the wrapper-based FS technique is employed for optimal feature selection. Utilizing a search strategy, FS evaluates each combination of features to identify the most suitable subset for the task at hand.

II. RELATED WORK

Geetha and colleagues [6] introduced a novel feature selection method aimed at improving credit card fraud classification accuracy. This study employs Enhanced Neural Networks (ENNs) to identify fraudulent accounts, leveraging feature selection techniques grounded in Artificial Bee Colonies (ABCs). These techniques extract pertinent features from transaction-level credit card datasets, enhancing the precision of classification outcomes. The research delves into various aspects of the reduced dataset, elucidating logical connections among transaction record attributes. ENNs compute Cross-Correlation Functions (CCFs) between attributes based on Logical Graphs of Behavior Profiles (LGBPs) and user transaction data, facilitating a deeper understanding of the dataset's underlying relationships. In their study, Georgieva and her team [7] focused on detecting fraudulent activities. They leveraged Artificial Neural Networks (ANNs), which when appropriately trained, exhibit capabilities akin to human brain functions. Like humans, ANNs excel in classification tasks and learn through observation. However, credit card transaction data often presents a significant imbalance between legitimate and fraudulent activities. To tackle this challenge, the researchers implemented resampling techniques.

Their approach involved the development and training of a pattern recognition network using Matlab's Neural Network Toolbox, employing a scaling conjugated gradients backpropagation technique. This methodology enabled the network to effectively discern patterns and classify transactions, contributing to the detection of fraudulent cases within credit card traffic data.

In their research, Asha and colleagues [8] devised a variety of machine learning methodologies, including Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Artificial Neural Network (ANN). They employed supervised machine learning and deep learning techniques to distinguish between legitimate and fraudulent transactions. Similarly, Ghobadi and team [9] developed a Credit Card Fraud Detection (CCFD) model utilizing ANN and metacost. ANNs are effective in both preventing and identifying credit card fraud.

However, the presence of inconsistent data poses challenges in detecting fraudulent transactions. To address this issue, metacost was utilized to handle imbalanced data. Moreover, a Cost Sensitive Neural Network (CSNN) model was implemented for abuse detection. Comparative analysis revealed that this model yielded cost savings and improved detection rates compared to the Artificial Immune System (AIS) model. Notably, the study's dataset comprised actual transaction information provided by a prominent Brazilian credit card company. In their study, Kavitha, and her team [10] devised a decision tree strategy incorporating an evolutionary algorithm to optimize node discovery. The efficacy of this approach was evaluated using a Principal Component Analysis (PCA)--based Artificial Neural Network (ANN) classifier, demonstrating its superiority over existing methods. On a parallel note, Mqadi et al. [11] introduced a data-point machine-learning technique to address the challenges posed by an imbalanced credit card database.

Their study utilized the data-point methodology and addressed overfitting concerns by employing the Synthetic Minority Over-sampling Technique (SMOTE). Various classifiers including Support Vector Machines, Logistic Regression, Decision Trees, and Random Forests were employed for classification tasks. Model accuracy was assessed using precision, recall, F1-score, and average precision metrics.

The findings underscored the model's struggle in accurately identify fraudulent transactions when confronted with severely unbalanced data. However, upon implementation of the SMOTE-based oversampling strategy, a notable improvement in accuracy was observed, enhancing the model's capability to predict positive classifications.

In their research, Panigrahi and colleagues [12] proposed a novel approach that integrates information from both recent and historical activities to detect credit card abuse effectively. The Fraud Detection System (FDS) comprises four key components: a rule-based filter, the Dempster-Shafer adder, the transaction history database, and the Bayesian learner. This innovative method significantly enhances the effectiveness of credit card fraud detection by amalgamating multiple pieces of evidence. Thorough simulations conducted using stochastic models demonstrate the superiority of this approach over alternative techniques. By leveraging a combination of recent and past activity data, the proposed system offers a robust and comprehensive solution to combating fraudulent transactions in credit card systems. The literature offers a comprehensive exploration of various feature selection (FS) strategies, each addressing unique challenges and objectives. The study by [18] provides an in-depth evaluation of semi-supervised FS techniques, elucidating their respective advantages and limitations. In another study, [19] focuses on the development of a novel FS method within the filter model framework.

This approach effectively eliminates unnecessary and redundant features while maintaining computational efficiency, surpassing existing algorithms in performance. Similarly, [20] introduces a wrapper-based FS strategy leveraging the artificial electric field optimization algorithm.

This method excels in selecting the most relevant features, enhancing the overall effectiveness of feature selection processes. In the realm of biomedical data analysis, [21] proposes a new FS technique tailored to the unique characteristics of biomedical datasets. Meanwhile, [22] introduces a novel approach based on the real-valued grasshopper optimization algorithm, showcasing promising results in feature selection tasks.

The utilization of the jaya optimization algorithm for constructing a wrapper-based FS model is highlighted in [23], emphasizing its ability to achieve robust feature selection without complex parameter tuning. Efforts to reduce the time complexity of wrapper-based FS with an embedded K-Nearest-Neighbor (KNN) classifier are explored in [24], presenting a novel approach for accelerating feature relevance evaluation. Furthermore, [25] presents a Differential Evolution (DE) optimization technique for FS, comparing its performance with Genetic Algorithms (GA) and Particle Swarm Optimization (PSO). Additionally, [26] investigates the use of genetic algorithms for feature subset selection in neural network classifiers. The application of FS in intrusion detection within wireless sensor networks is discussed in [27], utilizing techniques such as PSO and PCA space, while [28,29] introduces the Ant Colony Optimization method for FS. A novel swarm intelligence technique inspired by the behavior of rock hyrax swarms is proposed in [30], offering a balanced approach to optimization problems. In the domain of fraud detection, [31] presents an improved Credit Card Risk Identification (CCRI) technique based on feature selection algorithms like Random Forest and Support Vector Machine (SVM) classifiers. Moreover, [32] describes an SVM-based FS strategy utilizing artificial variables and mutual information to filter noisy variables in high-dimensional metabolome data. Furthermore, [33] highlights the necessity of employing FS techniques in credit card fraud detection models. Finally, [34-36] delve into detailed analyses and comparisons of various machine learning algorithms, including boosting techniques, for effective fraud detection.

III. RESEARCH METHODOLOGY

To enhance credit card fraud detection, this study proposes a novel approach integrating hybrid deep learning and intelligent feature selection methods. Initially, the attributes of transaction records are meticulously sorted before their characterization. A Logical Graph of Behavior Profiles (LGBP) is then constructed based on these attributes, encapsulating various transaction data aspects. Subsequently, the Modified Butterfly Optimization Algorithm (MBOA) is employed to select relevant features from the transactional credit card dataset. Finally, a hybrid deep learning model is utilized to depict the logical relationships among transaction data properties. Specifically, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are employed to bolster the accuracy of transaction identification processes. The methodology's proposed approach is illustrated in Figure 1, encapsulating the sequential steps involved in the process.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

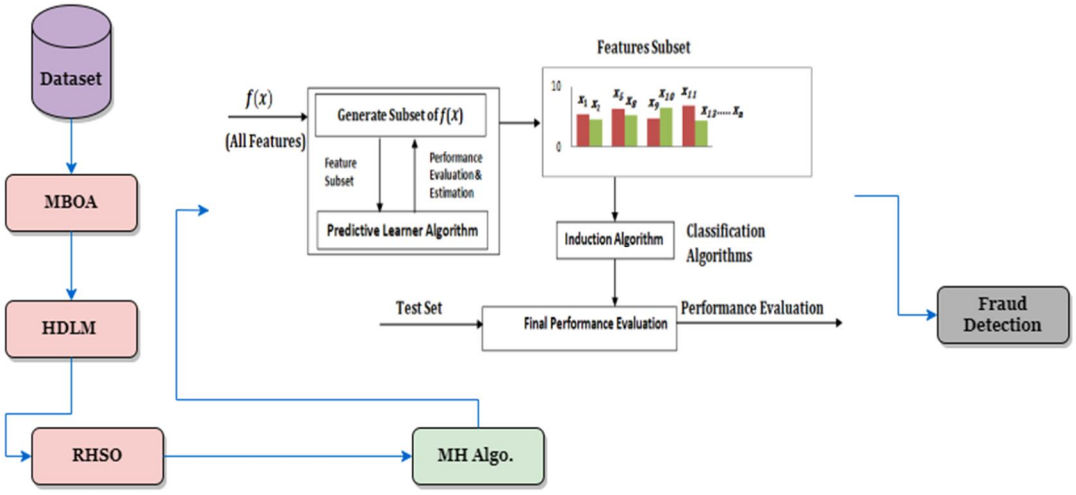


Fig. 1. Proposed Hybrid Deep Learning Model

A. Feature Selection using a Modified Butterfly Optimization Algorithm (MBOA)

The process of selecting key features involves carefully curating a small subset of elements, thereby reducing the influence of noisy or irrelevant factors on prediction outcomes. Various approaches such as filtration, wrapping, and embedding can be employed on the entire dataset to create this subset of effective characteristics. It is crucial to note that the system's performance significantly improves when the correct feature set is chosen. In this research, we introduce the Modified Butterfly Optimization Algorithm (MBOA) to enhance the efficiency of the feature selection process. Initially, we identify the shortcomings of the Butterfly Optimization Algorithm (BOA) and elucidate the rationale behind the introduction of the MBOA in the subsequent subsection.

B. Butterfly Optimization Algorithm (BOA)

To illustrate the concepts discussed in the context of a search algorithm, we conceptualize the characteristics of butterflies as follows:

1) Each butterfly emits a scent that attracts other butterflies.
2) Butterflies may migrate randomly or toward the butterfly with the strongest scent.
3) The geography of the target function influences or dictates the intensity of a butterfly's sensory perception.

The Butterfly Optimization Algorithm (BOA) process encompasses three distinct phases: initiation, iteration, and finalization. During initialization, the target function and solution space parameters are defined. Subsequently, a starting population of butterflies is generated, and their data is stored in a fixed-size memory throughout the experiment.

Fitness and scent values are then computed and recorded for each butterfly. Following initialization, the algorithm proceeds to the iterative phase, wherein it conducts searches utilizing the virtual butterflies. Finally, in the concluding phase, the algorithm halts once the optimal solution has been identified.

C. Hybrid Deep Learning Model (HDLM)

In the domain of deep learning, computer models with multiple processing layers can learn data representations across various levels of abstraction. These representations, in turn, facilitate the generation of predictions based on learned patterns. In this study, a hybrid learning technique is introduced aimed at enhancing the accuracy of credit card fraud detection. The approach involves combining a Recurrent Neural Network (RNN) with a Convolutional Neural Network (CNN) to bolster the efficiency of fraud detection compared to using a single classifier. The evolution of biotechnology has paved the way for the development of CNNs as effective models, characterized by neurons acting as localized filters applied across the input space. CNNs excel in extracting both local and deep features from input data.

On the other hand, RNNs specialize in analyzing sequence data. Unlike standard neural network architectures, which link input, hidden, and output layers linearly, RNNs are adept at handling sequence data. Moreover, CNN layers within the proposed model learn low-level translation-invariant features, which are subsequently utilized to generate higher-order features.

By integrating convolution and pooling into a hierarchical process, RNNs enhance the model's capability to identify intricate patterns within sequential data. Extensive experimentation has demonstrated that both CNNs and RNNs outperform traditional approaches in fraud detection. Motivated by these findings, a novel hybrid model, CNN-RNN, is proposed specifically designed for classifying fraudulent transactions. The methodology involves a comprehensive training and testing procedure. Before training, the CNN model undergoes pre-training, followed by transfer learning to adapt it to the specific task. During training, all CNN layers are frozen, and the CNN-RNN model is subsequently trained. Attention mechanisms are integrated into the model to combine the strengths of both CNN and RNN architectures. During the testing phase, pre-processed test data are fed into the optimized CNN-RNN model, and the SoftMax layer generates classification results. This integrated approach leverages the complementary strengths of CNNs and RNNs, ultimately enhancing the accuracy and efficiency of credit card fraud detection.

RHSO (Rock Hyrax Swarm Optimization) is a meta-heuristic inspired by the collective behavior of rock hyrax swarms in nature. These swarms exhibit a unique approach to finding food and ensuring their colony's protection, led by a dominant male overseeing the group. The RHSO algorithm simulates this behavior to optimize feature selection processes. The functioning model of RHSOFS (Rock Hyrax Swarm Optimization Feature Selection) is illustrated in Figure 3, which involves partitioning the dataset into training and testing sets. The training data is then inputted into the optimization technique (represented as $f(x)$) to identify the most suitable optimal features. Subsequently, the classification algorithm is provided with this optimal subset of features (also represented as $f(x)$), along with the training and testing data, to assess the model's performance. Equation (1) can be utilized to express the selection of the most optimal features, while Equation (2) iteratively minimizes errors using specified characteristics, thereby enhancing classification accuracy.

First, generate, select, & examine a random sample of a binary (0, 1) population for the total number of input features for FS. Create a feature subset that is equal to 1 for each representation of the input population. For the computation of fitness, the extracted

The goal of this research work is to find the best subset of input features that reduce the model's fitness while also improving its accuracy.

$$err(x_i) = actual_{output(x_i)} - model\_estimated_{output(x_i)} \ldots\ldots\ldots (1)$$

$$fitness(x) = \sum_{x=0}^{n} err(x) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

$$circ = sqrt(n_1^2 + n_2^2) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

If the value of the output grows larger than 360, or less than 0, the angle (ang) can be set to 360 or 0 to keep it within the desired range.

Algorithm 1 Proposed Hybrid Algorithm

1. Create an initial population of 0 & 1 of P agents randomly.
2. Set the dimension of the problem, D = P, where P is the number of agents.
3. Set Low to 1 and High to D, where Low and High refer to the low and high dimensions, respectively.
4. Generate the value of r1 and r2, where r1 is a random number (0, 1) and r2 is a random radius (0, 360).
5. Generate training and testing data.
6. Set max_iter = maximum number of iterations.
7. Calculate each agent's fitness using Equation (1).
8. Set Leader = the best agent.
9. Set t = 1.
10. while (t < max_iter)
11. for (i = 1 to n) do
12. Update Leader position, according to Equation (3).
13. Update the position of each search agent according to Equation (2).
14. Calculate Newfitness of each search agent using Equation (1).
15. Select the best member of the population → bestX = X (min (fitness))
16. Update the angle according to Equations (3) and (1).
17. If (Newfitness (i) <= fitness (i); then
18. Update the position of each search agent according to Equation (2).
19. fitness (i) = Newfitness (i).
20. end if
21. end for

22. t = t + 1.
23. end while
24. Return the best agent

The proposed hybrid algorithm starts by creating an initial population consisting of binary values, 0s, and 1s, represented by P agents, which are randomly generated. The dimension of the problem, denoted as D, is set equal to the number of agents, P. Additionally, the algorithm sets the low and high dimensions, represented as Low and High, respectively, where Low is set to 1 and High is set to D. Next, the algorithm generates random values for r1 and r2, where r1 is a random number between 0 and 1, and r2 represents a random angle within a 360-degree radius. Subsequently, training and testing data are generated for further processing. The maximum number of iterations, max_iter, is defined, and the fitness of each agent is calculated using Equation (1). The best agent, referred to as the Leader, is determined based on its fitness value. The algorithm proceeds with initializing the iteration count, t, to 1. In the iterative process, while the iteration count is less than the maximum specified iterations, the algorithm updates the position of each agent by following Equation (3) for the Leader and Equation (1) for other search agents. The fitness of each agent is recalculated using Equation (2). The best member of the population, bestX, is selected based on the agent with the minimum fitness value. The algorithm then updates the angle of each agent according to specified equations. If the newly calculated fitness value is less than or equal to the current fitness value, the position of the search agent is updated, and its fitness value is replaced with the new value. The iterative process continues until the maximum number of iterations is reached. Finally, the algorithm returns the best agent based on its fitness value, representing the solution obtained by the hybrid algorithm.

IV. RESULTS & DISCUSSION

The chosen attributes proved to be effective in detecting transactional fraud. Specifically, the transaction amount per day was categorized into two segments: small (SM) and average (AV). Additionally, four distinct merchant IDs were considered, and the frequency of amounts was classified as low (LO) and high (HI). The transaction amounts were further divided into four segments: (0-200), (200-1000), (1000-2000), and (2000 and above). During the experiment, pre-processed data were utilized based on prior modifications. External quality measures such as Accuracy, Precision, Recall, and F-measure were employed to assess the performance of the proposed approach.

Table 1 presents a comparative analysis of the performance results of the proposed technique alongside existing methods. The metrics used for evaluation include Accuracy, Precision, Recall, and F-measure. The existing methods listed in the table include TAS, LGBPs, LGBP-ENN, and HDLM, with their respective performance values reported under each metric.

Table 1. Performance outcomes of the proposed & existing approaches

Metrics	TAS	LGBPs	LGBP-ENN	HDLM	Proposed model
Accuracy	85.1500	91	93.100	94.300	94.700
Precision	82.0190	90.8352	93.9557	95.2272	95.3381
Recall	84.230	91.0908	91.1612	95.8272	95.9211
F-measure	83.0060	90.9629	92.537	95.5262	95.6162

Upon examination of the table values, it is evident that the proposed technique consistently outperforms the existing methods across all metrics. For instance, in terms of Accuracy, the proposed model achieves a significantly higher value of 94.700 compared to TAS (85.1500), LGBPs (91), LGBP-ENN (93.100), and HDLM (94.300). Similarly, for Precision, Recall, and F-measure, the proposed model consistently demonstrates superior performance, achieving values of 95.3381, 95.9211, and 95.6162, respectively, compared to the corresponding values of the existing methods. These results indicate that the proposed technique exhibits enhanced performance in accurately identifying fraudulent transactions compared to the established methods, thereby highlighting its effectiveness and superiority in fraud detection.

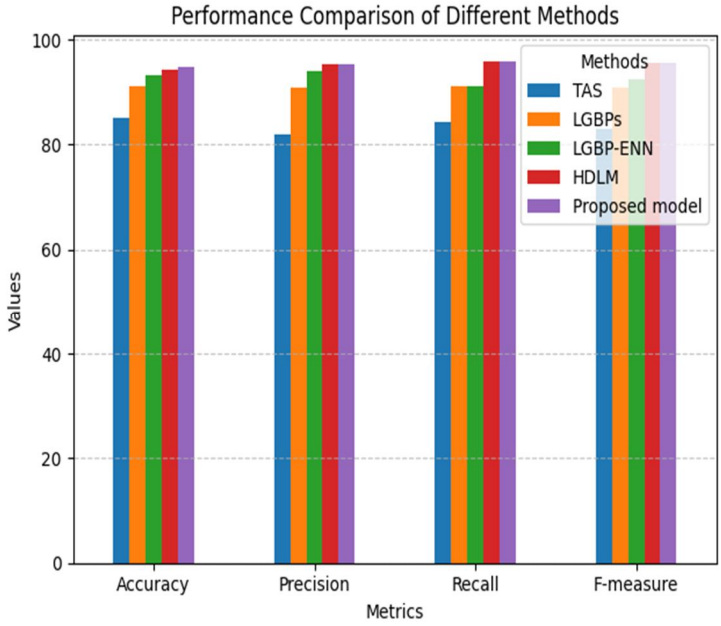


Fig. 2. Performance comparison of diverse methods

To assess the performance and efficacy of the proposed feature selection (FS) approach, termed RHSOFS, we conducted a comparative analysis against several established methods, including DEFS, GAFS, PSOFS, and ACOFS. Our evaluation was carried out using numerical experiments on a real-world credit card fraud dataset, employing various data mining techniques to gauge the effectiveness of the RHSOFS approach. Given the scarcity of authentic credit card fraud datasets, we adopted a stratified cross-validation (SCV) methodology to ensure robustness in our evaluations. This approach involved generating ten identical datasets, each stratified to maintain consistent class ratios across folds. Unlike traditional cross-validation techniques, such as k-fold, SCV specifically addresses classification problems and is particularly suitable for handling imbalanced datasets. By preserving the target class distribution within each fold, SCV minimizes the risk of overfitting and ensures a more reliable assessment of model performance compared to random dataset splits.

Table 2. Holm Test.

S. No	FS Algorithms	z-Values	p-Values	α/ (P − i)
1	RHSOFS: ACOFS	2.779	0.00273	0.05
2	RHSOFS: PSOFS	2.002	0.02264	0.025
3	RHSOFS: GAFS	0.239	0.40553	0.016667
4	RHSOFS: DEFS	2.928	0.00171	0.0125
5	RHSOFS: WTFS	5.139	0.00001	0.01
6	PROPOSED HYBRID	4.315	0.00001	0.1

Table 2 presents the results of the Holm Test, which was conducted to compare the performance of various feature selection (FS) algorithms, including RHSOFS: ACOFS, RHSOFS: PSOFS, RHSOFS: GAFS, RHSOFS: DEFS, RHSOFS: WTFS, and the proposed hybrid algorithm. Each row in the table corresponds to a pairwise comparison between RHSOFS and another FS algorithm. The columns indicate the serial number of the comparison, the FS algorithms being compared, the z-values obtained from the test, the corresponding p-values, and the adjusted significance level (α / (P − i)), where α is the significance level, P is the total number of comparisons, and i is the rank of the comparison.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

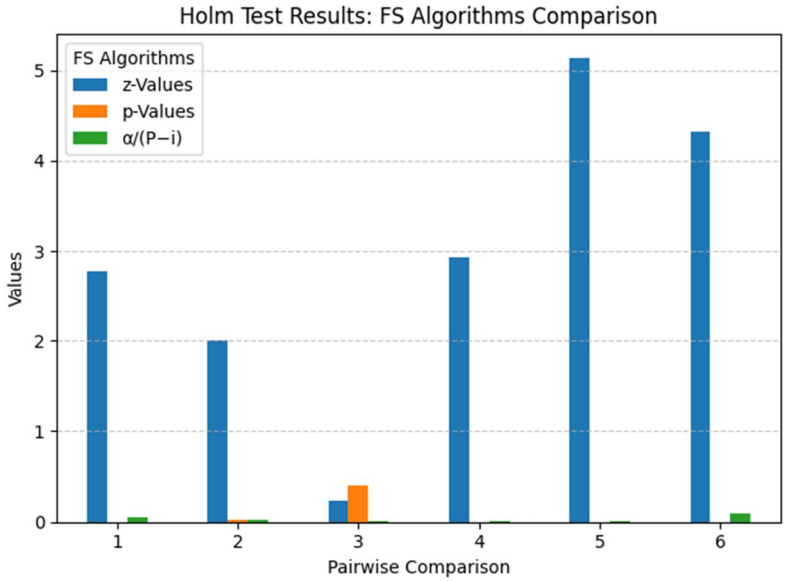


Fig. 3. Holm test results: FS Algorithms Comparison

The z-values in the analysis depict the standardized difference between the mean performances of the paired algorithms under comparison. Conversely, the p-values signify the likelihood of observing such a discrepancy purely by chance. In determining statistical significance, the adjusted significance level is employed, considering the multiple comparisons conducted within the test. Upon scrutiny of the results, notable disparities in performance emerge between RHSOFS and each of the compared FS algorithms. Notably, RHSOFS exhibits statistically significant superiority over ACOFS, PSOFS, DEFS, and WTFS, as underscored by their respective z-values and p-values. Additionally, the proposed hybrid algorithm demonstrates significant performance enhancements when compared to RHSOFS across all comparative analyses. Overall, the Holm Test findings furnish valuable insights into the relative efficacy of various FS algorithms, facilitating the identification of the most suitable approach for credit card fraud detection.
In a dataset, there may exist irrelevant, duplicated, or noisy features that not only prolong processing time but also adversely impact the model's efficacy. By exclusively processing optimal features and eliminating unwanted, redundant, and noisy characteristics, the model's performance is enhanced while computational overhead is minimized. Feature selection (FS) methods, on the other hand, aim to extract a subset of significant and pertinent original features. This study compares the performance of various machine learning models, including Naive Bayes (NB), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Decision Trees (DT), utilizing wrapper-based FS methods such as Differential Evolutionary Feature Selection (DEFS), Genetic Algorithm Feature Selection (GAFS), Particle Swarm Optimization Feature Selection (PSOFS), Ant Colony Optimization Feature Selection (ACOFS), and Rock Hyrax Swarm Optimization Feature Selection (RHSOFS). Two approaches are employed to model FS processes: fixed optimal features selection and variable optimal features selection.
The dataset under examination has already undergone preprocessing via Principal Component Analysis (PCA), resulting in 28 principal components or features ranging from V1 to V28, potentially containing duplicated features. Reducing redundant features and replacing them with the next most relevant features presents a challenge. Conversely, distinct FS techniques employing various optimization algorithms aim to select the minimal number of variable optimal features, with individual features assigned binary values of 1 or 0. The RHSOFS approach is introduced to eliminate irrelevant features while selecting the most relevant and appropriate ones to enhance the model's performance. The quantity of relevant features utilized is contingent upon the optimization method's performance, with the selection of an optimal number of features not guaranteed across different optimization techniques. Ultimately, the model learns to leverage the selected relevant features acquired through optimization procedures to enhance its performance. The objective of this study is to optimize the number of relevant features to improve the model's efficacy. A predefined limit on the scale can often be established based on the rank of relevant features or the desired number of features in a fixed optimal FS technique, ensuring the selection of the specified number of features. The comparison is predicated on identifying the best FS approach using various optimization algorithms.

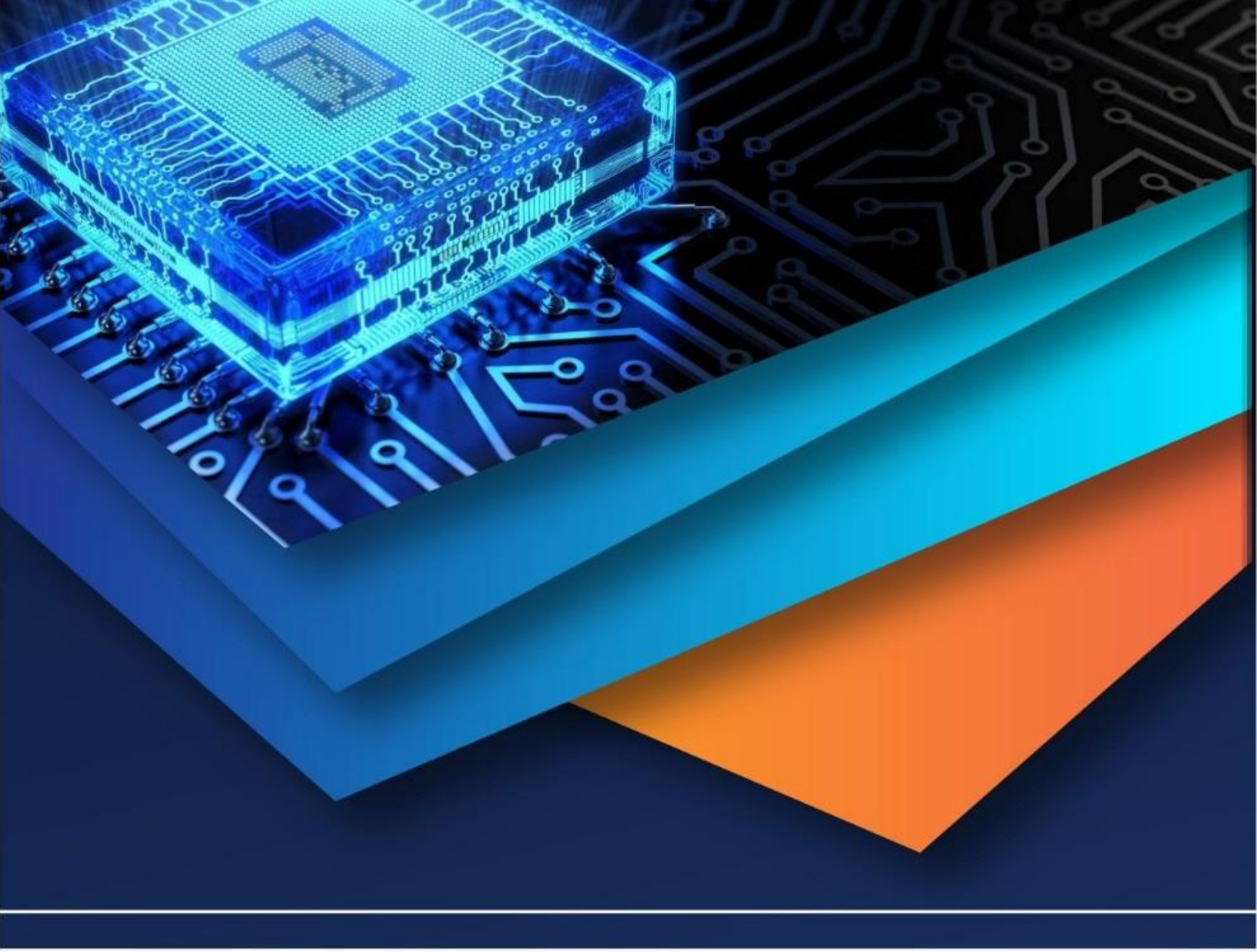
V. CONCLUSION

The culmination of this research endeavors to address the prevalent issue of financial losses incurred by customers due to credit card fraud, posing substantial losses for banking institutions and credit card companies. With a focus on mitigating financial losses for individuals and safeguarding the interests of financial entities, this study strives to develop a model capable of accurately discerning fraudulent transactions from legitimate ones, employing the proposed Hybrid Deep Learning approach within the provided dataset. Preceding the classification of transactional attributes, meticulous sorting of transaction records' attributes is conducted, followed by the construction of a Logical Graph of Behavior Profiles (LGBP) encapsulating various transaction data types. Subsequently, relevant features are selected using the Modified Butterfly Optimization Algorithm (MBOA)-based feature selection method, culminating in the representation of logical relationships within transaction data via a hybrid deep learning model. Integrating Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) enhances transaction detection efficiency, with experimental findings showcasing superior performance compared to both CNN models and traditional machine learning approaches. Employing an optimizing technique is pivotal in curbing the operational costs of the hybrid model, thereby ensuring effectiveness while minimizing expenses. To ascertain an optimal feature subset, a novel Feature Selection (FS) technique based on the Rock Hyrax Swarm Optimization (RHSO) algorithm, termed RHSOFS, is introduced. This approach effectively identifies the most pertinent features while mitigating irrelevant, redundant, and noisy elements. Utilizing four classifiers—Naive Bayes (NB), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Decision Trees (DT)—on ten datasets, the efficacy of the RHSOFS approach is evaluated. Results demonstrate its prowess in reducing duplicate features and outperforming existing methods, evident in its superior classification accuracy and feature selection capability. While AI techniques exhibit promise in predicting future behaviors, their efficacy in combating cybercrime complexities remains challenging, necessitating continual learning to counteract evolving offender strategies. It's recommended to explore works such as "Evolution-Oriented Monitoring for Security Properties in Cloud Applications" and integrating intelligence into Trusted Computing hardware to furnish hybrid hardware-software certification mechanisms, as proposed in "Software and Hardware Certification Techniques in a Combined Model." Such strategies can be extended to diverse applications like mail fraud detection, intrusion detection, and counterfeit insurance analysis, bolstering security across multifaceted domains.

REFERENCES

[1] Raj, S. B. E., & Portia, A. A. (2011, March). Analysis of credit card fraud detection methods. In 2011 International Conference on Computer, Communication and Electrical Technology (ICCCET) (pp. 152-156). IEEE.
[2] Chaudhary, K., Yadav, J., &Mallick, B. (2012). A review of fraud detection techniques: Credit card. International Journal of Computer Applications, 45(1), 39-44.
[3] Awoyemi, J. O., Adetunmbi, A. O., &Oluwadare, S. A. (2017, October). Credit card fraud detection using machine learning techniques: A comparative analysis. In 2017 International Conference on Computing Networking and Informatics (ICCNI) (pp. 1-9). IEEE.
[4] Bahnsen, A. C., Stojanovic, A., Aouada, D., &Ottersten, B. (2013, December). Cost-sensitive credit card fraud detection using Bayes minimum risk. In 2013 12th International Conference on Machine Learning and Applications (Vol. 1, pp. 333-338). IEEE.
[5] Halvaiee, N. S., & Akbari, M. K. (2014). A novel model for credit card fraud detection using Artificial Immune Systems. Applied soft computing, 24, 40-49.
[6] Geetha,N., & Dheepa,G. (2022, March).Transaction fraud detection using Artificial Bee Colony (ABC) based feature selection and Enhanced Neural Network (ENN) classifier. International Journal of Mechanical Engineering (Vol. 7, No. 3, ISSN 0974- 5823).
[7] Georgieva, S., Markova, M., & Pavlov, V. (2019, October). Using neural network for credit card fraud detection. In AIP Conference Proceedings (Vol. 2159, No. 1, p. 030013). AIP Publishing LLC.
[8] Asha, R. B., & KR, S. K. (2021). Credit card fraud detection using artificial neural network. Global Transitions Proceedings, 2(1), 35-41.
[9] Ghobadi, F., &Rohani, M. (2016, December). Cost sensitive modeling of credit card fraud using neural network strategy. In 2016 2nd international conference of signal processing and intelligent systems (ICSPIS) (pp. 1-5). IEEE.
[10] Kavitha, C., &Iyakutti, K. (2014). Optimized Anomaly based Risk Reduction using PCA based Genetic Classifier. Global Journal of Computer Science and Technology.
[11] Mqadi, N., Naicker, N., &Adeliyi, T. (2021). A SMOTE based oversampling data-point approach to solving the credit card data imbalance problem in financial fraud detection. International Journal of Computing and Digital Systems, 10(1), 277-286.
[12] Panigrahi, S., Kundu, A., Sural, S., &Majumdar, A. K. (2009). Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning. Information Fusion, 10(4), 354-363.
[13] Kohavi, R.; John, G.H. Wrappers for feature subset selection. Artif. Intell. 1997, 97, 273–324.
[14] Huang, S.H. Supervised feature selection: A tutorial. Artif. Intell. Res. 2015, 4, 22–37.
[15] Peng, Y.; Wu, Z.; Jiang, J. A novel feature selection approach for biomedical data classification. J. Biomed. Inform. 2010, 43, 15–23.
[16] Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. J. Mach. Learn. Res. 2003, 3, 1157–1182.
[17] Das, H.; Naik, B.; Behera, H.S. Optimal Selection of Features Using Artificial Electric Field Algorithm for Classification. Arab. J. Sci. Eng. 2021, 46, 8355–8369.

[18] Sheikhpour, R.; Sarram, M.A.; Gharaghani, S.; Chahooki, M.A.Z. A Survey on semi-supervised feature selection methods. Pattern Recognit. 2017, 64, 141–158.
[19] Yu, L.; Liu, H. Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution. In Proceedings of the 20th International Conference on Machine Learning (ICML-03), Washington, DC, USA, 21–24 August 2003; Volume 2, pp. 856–863.
[20] Van Hulse, J.; Khoshgoftaar, T.M.; Napolitano, A. Experimental perspectives on learning from imbalanced data. In Proceedings of the 24th International Conference on Machine Learning, New York, NY, USA, 20–24 June 2007; Volume 227, pp. 935–942.
[21] Das, H.; Naik, B.; Behera, H.S. Optimal Selection of Features Using Artificial Electric Field Algorithm for Classification. Arab. J. Sci. Eng. 2021, 46, 8355–8369.
[22] Zakeri, A.; Hokmabadi, A. Efficient feature selection method using real-valued grasshopper optimization algorithm. Expert Syst. Appl. 2019, 119, 61–72.
[23] Das, H.; Naik, B.; Behera, H.S. A Jaya algorithm-based wrapper method for optimal feature selection in supervised classification. J. King Saud Univ.-Comput. Inf. Sci. 2022, 34, 3851–3863.
[24] Wang, A.; An, N.; Chen, G.; Li, L.; Alterovitz, G. Accelerating wrapper-based feature selection with K-nearest-neighbor. Knowl. Based Syst. 2015, 83, 81–91.
[25] Khushaba, R.N.; Al-Ani, A.; Al-Jumaily, A. Differential Evolution based feature subset selection. In Proceedings of the 2008 19th International Conference on Pattern Recognition, Tampa, FL, USA, 8–11 December 2008.
[26] Yang, J.; Honavar, V. Feature Subset Selection Using A Genetic Algorithm Feature Subset Selection Using 1 Introduction. Intell. Syst. Appl. 1997, 13, 44–49.
[27] Ahmad, I. Feature selection using particle swarm optimization in intrusion detection. Int. J. Distrib. Sens. Netw. 2015, 2015.
[28] Ahmed, A.-A. Feature Subset Selection Using Ant Colony Optimization. Int. J. Comput. 2005, 2, 53–58. Available online: http://epress.lib.uts.edu.au/research/handle/10453/6002 (accessed on 1 January 2005).
[29] Deriche, M. Feature selection using ant colony optimization. In Proceedings of the 2009 6th International Multi-Conference on Systems, Signals and Devices, Djerba, Tunisia, 23–26 March 2009.
[30] Urbanowicz, R.J.; Meeker, M.; la Cava, W.; Olson, R.S.; Moore, J.H. Relief-based feature selection: Introduction and review. J. Biomed. Inform. 2018, 85, 189–203.
[31] Al-Khateeb, B.; Ahmed, K.; Mahmood, M.; Le, D.N. Rock Hyraxes Swarm Optimization: A New Nature-Inspired Metaheuristic Optimization Algorithm. Comput. Mater. Contin. 2021, 68, 643–654.
[32] Rtayli, N.; Enneya, N. Selection features, and support vector machine for credit card risk identification. Procedia Manuf. 2020, 46, 941–948.
[33] Lin, X.; Yang, F.; Zhou, L.; Yin, P.; Kong, H.; Xing, W.; Lu, X.; Jia, L.; Wang, Q.; Xu, G. A support vector machine-recursive feature elimination feature selection method based on artificial contrast variables and mutual information. J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 2012, 910, 149–155.
[34] Bhattacharyya, S.; Jha, S.; Tharakunnel, K.; Westland, J.C. Data mining for credit card fraud: A comparative study Decis. Support Syst. 2011, 50, 602–613.
[35] Mittal, S.; Tyagi, S. Performance evaluation of machine learning algorithms for credit card fraud detection. In Proceedings of the 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 10–11 January 2019; pp. 320–324.
[36] Padhi, B.; Chakravarty, S.; Biswal, B. Anonymized credit card transaction using machine learning techniques. In Advances in Intelligent Computing and Communication Lecture Notes in Networks and Systems; Springer: Berlin/Heidelberg, Germany, 2020; Volume 109.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY