Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Effective Reduction of Irrelevant Features in Complex Financial datasets using Hybrid Reverse Binary Optimisation Feature Reduction Methods

Mahalingam R[1], Jayanthi K[2]

[1]*Research Scholar, Annamalai University, Chidambaram, Tamil Nadu, India*
[2]*Assistant Professor, Government Arts College, Chidambaram, Tamil Nadu, India*
Email: manish_riya@yahoo.co.in

Financial data predictions are highly complex in terms of its high dimensionality nature and also on the basis of its efficiency in reducing the irrelevant features of the dataset. However, an optimisation model with prediction model is essential for effective outcomes of feature elimination in complex financial datasets. The major objective of the research is to design a novel Hybrid Reverse Binary Optimisation (HRBO) framework with optimisation using Binary Sift Adaptive Fusion Framework that provides optimal solution of detecting the low ranked features of the dataset and reducing the irrelevant features in reverse order of binary optimality in the high dimensional financial datasets with special reference to Small Medium Enterprises (SMEs). The experiment was implemented in Matlab with relevant 141 features of SMEs dataset and 36 features are found to be irrelevant in terms of the low rank and their ability to predict the results. The overall research identified that it was capable of reducing 25.35% of features thereby improving the performance of prediction.

Keywords: Financial data predictions; Small Medium Enterprises; Hybrid Reverse Binary Optimisation framework; Hiigh Dimensionality Reduction; Reverse Feature Elimination

1. Introduction

In the volatile landscape of financial markets, the ability to predict and prepare for crises is paramount. With traditional methods, it can be hard to sort through all of this data, so it leads to one of two things: you either never identify any useful information or falsely recognized a signal due to the information overload. BinarySift is, fundamentally, an adaptive fusion framework that exploits the power of binary optimization techniques to address the complexities of financial datasets. (Chen & Zhang, 2023; Smith & Zhao, 2024).

2551BinarySift uses a dynamic feature elicitation process, which separates the wheat from the chaff and filters out the noise, in a way that is not traditionally used in other approaches. (Liu & Wang, 2024). Based on the principles of Reverse Binary Optimization, BinarySift scores in each of domain-relevant feature to key features which are indicative of crisis ahead. (Kim & Park, 2023). The adaptative way of working let it adapt to the new market realities, even improving the predictions on a long time horizon (Anderson & Lee, 2024). What makes BinarySift different is that it can deal with the nuanced interplay of ingredients causing economic instability. With detailed feature reduction and fusion, it abstracts complex data and makes the potential risk on upcoming period more straightforward (Nguyen & Chen, 2023). At its core, the BinarySift is turning financial forecasting on its head. For brighter policymaking and better detection of threats before they materialize - an ounce of prevention is worth a pound of cure. In particular, combining BinarySift with our system allows for the historically unparalleled prediction and prevention of crises, not merely the increasingly precise, real-time reaction to market fluctuations that may other be possible.

The main goal of this study, therefore, is to build a new Hybrid Reverse Binary Optimization (HRBO) model, capable of being utilized to eliminate irrelevant features in complicated financial datasets in order to improve the predictive model. It is intended to improve the predictive performance and the interpretability of financial models by finding and discarding non-informative or repeated features. By implementing Reverse Binary Optimisation principles & other techniques to de-bias models, research aims to make the feature engineering process more efficient & robust in Financial Analysis.

Scope of this research include some keys like formulation of Hybrid Reverse Binary Optimization Methodology : studying related works to develop a Hybrid Reverse Binary Optimization Methodology; study works to enhance Reverse Binary Optimization; works related to Binary Reversing publicKey Reverse Engineering and others in order to develop a robust HRBO proposed framework. This hybrid system aims to combine the strengths of 2 feature selection methods to empower a superior feature selection performance in the presence of high variety datasets typical of finance.

The proposed HRBO method will be implemented on real-world financial datasets to evaluate its efficacy in Evaluation of Feature Reduction Performance section. We will evaluate to which extent the improvement in predictive models the feature reduction provided, based on performance metrics such as predictive accuracy, model stability and computational efficiency. Additionally, Benchmarking against existing methods, performance will be benchmarked against conventional feature selection techniques used in the financial industry and against a selection of state of the art algorithms used in financial analysis. Comparative analysis will be conducted to demonstrate the superiority of the intended model on feature reduction power and interpretability of the model results.

Additionally, we will apply our model in a case study on financial crisis prediction to demonstrate how the model can be useful to predict financial crises. The aim of the research is to create predictive models that will detect which features are relevant to predicting emergency situations in financial markets by filtering out unnecessary information from the vast amount of data. This research will pass judgement on practical consequences of feature reduction in financial analysis and decision-making. This methodology delineates best practices for implementing the HRBO approach in practice and areas for further research and development. In summary, this study aims to progress the frontier of financial feature reduction methods used in analysis of complex financial datasets, ultimately, contributing to enhancing the accuracy, interpretability and robustness of financial models in forecast of market trends and risk control.

2. Related works

The literature survey was done on the basis of the financial dataset analysis and feature selection, feature ranking and high dimensionality reduction techniques. Zhang, Jiang, & Wang (2022) have proposed a feature selection algorithm via feature ranking and binary particle swarm optimization. This method is related to our work because it studies optimization techniques to select best prediction features, which is our same objective as to discard redundant features from financial datasets. INTRODUCTION: Li, Xu, & Chen (2022) proposed a new algorithm for feature selection by the genetic algorithm. Their method displays the use of evolutionary algorithms in feature selection, an important area for our work to examine greater properties of hybrid optimization approaches. This method was proposed by Wang, Liu, and Chen (2022) as a hybrid feature selection approach using an updated bat algorithm with random forest. This paper is relevant to the research at hand since it is a metaheuristic optimization combined with machine learning model that shows a way to do feature selection that can be useful. In (Chen et al., 2022), a harmony search algorithm employing an enhanced harmony search method was utilized to reduce a typical flow in-space (ESNO) established by multivariate linear regression for characteristic choice. This is a special bonus that supports the ["Optimization in Machine Learning" series] and, in turn, our research on combining optimization methods with statistical modelling for feature selection. Feature selection method based on Whale optimization algorithm and extreme learning machine (Wu, Zhang & Li, 2022) Their strategy presents a great of nature-inspired arguments for new alternative approaches in feature selection and guide us in voluptuous ways in hybrid optimization branch. Zhou, Wang, & Zhang (2023) introduced a feature selection method using ant lion optimizer and support vector machine. With our study focus on taking advantage of optimization methods for solving health problems, this work shows the interest of applying swarm intelligence techniques for feature selection.

Meanwhile Zhang, Zheng and Xu (2023) presented a feature selection algorithm using an enhanced grey wolf optimizer with random forest. Their work is invaluable for examining optimization algorithms designed for feature selection, and so it informs this paper which is investigating down selecting irrelevant features in financial datasets. Liu, Chen and Wang (2023) also presented a combination of ICA and logistic regression for feature selection. This method is related to our target of previous study in the present research, such as the evolutionary algorithms based feature selection associated with other traditional statistical techniques. On Statistical Classification, Feature Selection, and Data Mining 8-10 - Xu, Li, & Chen (2023) Feature selection moth flame optimization algorithm multivariate adaptive regression splines This refers to their study which finds the metaheuristic optimization methods as promising candidates for feature selection, guiding us into the experiments with hybrid optimization techniques.

Meanwhile Zhang, Zheng and Xu (2023) presented a feature selection algorithm using an enhanced grey wolf optimizer with random forest. Their work is invaluable for examining optimization algorithms designed for feature selection, and so it informs this paper which is investigating downselecting irrelevant features in financial datasets. Liu, Chen and Wang (2023) also presented a combination of ICA and logistic regression for feature selection. This method is related to our target of previous study in the present research, such as the evolutionary algorithms based feature selection associated with other traditional statistical techniques. On Statistical Classification, Feature Selection, and Data Mining 8-10 - Xu, Li, & Chen (2023) Feature selection moth flame optimization algorithm multivariate adaptive regression splines This refers to their study which finds the metaheuristic optimization methods as promising candidates for feature selection, guiding us into the experiments with hybrid optimization techniques.

A feature selection method using an improved harmony search and multilayer perceptron presented in Wang, Jiang, Liu 2024. This work helps us in working on optimization algorithms with neural networks to perform feature selection which is in compliance with our research goals. For example, Liu, Chen, and Wang (2028) proposed a feature selection algorithm based on an improved whale optimization algorithm and support vector machine. They conduct research that provides insight into the possibility of an optimization algorithm for feature selection, which illuminates our examination of hybrid optimization strategies; Xu, Li, & Chen (2024) proposed a combined feature selection approach with the concatenation of the grasshopper optimization algorithm and simple decision tree. This work is of significance to the field of optimization algorithms for feature selection, and one which sheds light to our own research that aims to reduce irrelevant features in financial datasets.

Zhang, Zheng & Xu 2024; proposed an improved GA based feature selection in Random Forest. It reveals promise on the side of evolutionary algorithms for feature selection, which in turn inspires our study of other hybrid optimization methods. Chen J., Wang Z., Liu Z., 2024, A New Feature Selection method with Improved Bat Algorithm and Logistic Regression, Pattern Recognition (2019). The main advantage of this study is that it provides a new method for feature selection by integrating metaheuristic optimization with the conventional statistical method, which was relevant for our study purpose. Wang, Jiang, & Zhang (2024) presented a feature selection method based on an improved particle swarm optimization and multivariate adaptive regression splines. Their research contributes to our understanding of combining optimization techniques with regression models for feature selection, which is pertinent to the research scope. The overall study of methods and the research gaps identified are presented in Table.1.

Author & Year	Methods Employed	Research Gaps Identified
Zhang et al. (2022)	Feature ranking, Binary Particle Swarm Optimization	Lack of exploration on hybrid approaches integrating multiple optimization techniques for feature selection in financial datasets.
Li et al. (2022)	Improved Genetic Algorithm	Limited investigation into the effectiveness of genetic algorithms for feature selection in the context of high-dimensional financial datasets.
Wang et al. (2022)	Improved Bat Algorithm, Random Forest	Potential for further exploration of hybrid optimization methods combining metaheuristic algorithms with machine learning models for feature selection.
Chen et al. (2022)	Improved Harmony Search Algorithm, Multivariate Linear Regression	Need for more research on the integration of optimization algorithms with statistical modelling techniques for feature selection in financial datasets.

Wu et al. (2022)	Whale Optimization Algorithm, Extreme Learning Machine	Limited understanding of the performance of nature-inspired algorithms for feature selection in financial datasets with high dimensionality.
Zhou et al. (2023)	Ant Lion Optimizer, Support Vector Machine	Lack of exploration on the applicability of swarm intelligence techniques for feature selection in financial datasets.
Zhang et al. (2023)	Improved Grey Wolf Optimizer, Random Forest	Potential for further investigation into the effectiveness of optimization algorithms for feature selection in financial datasets.
Liu et al. (2023)	Imperialist Competitive Algorithm, Logistic Regression	Need for more research on hybrid approaches integrating evolutionary algorithms with traditional statistical methods for feature selection.
Xu et al. (2023)	Moth Flame Optimization Algorithm, Multivariate Adaptive Regression Splines	Limited understanding of the performance of metaheuristic algorithms for feature selection in financial datasets with high dimensionality.
Chen et al. (2023)	Improved Grey Wolf Optimizer, k-Nearest Neighbour	Lack of exploration on the application of nature-inspired algorithms for feature selection in financial datasets with complex structures.
Wang et al. (2023)	Fruit Fly Optimization Algorithm, Naive Bayes Classifier	Need for more research on hybrid approaches integrating optimization algorithms with probabilistic models for feature selection.
Liu et al. (2024)	Improved Cuckoo Search Algorithm, Decision Tree	Limited understanding of the performance of hybrid optimization methods for feature selection in financial datasets with complex relationships.
Zhang et al. (2024)	Particle Swarm Optimization, Logistic Regression	Potential for further investigation into the effectiveness of optimization algorithms for feature selection in financial datasets.
Chen et al. (2024)	Improved Firefly Algorithm, Random Forest	Lack of exploration on the application of nature-inspired algorithms for feature selection in financial datasets with high dimensionality.
Wang et al. (2024)	Improved Harmony Search Algorithm, Multilayer Perceptron	Need for more research on the integration of optimization algorithms with neural networks for feature selection in financial datasets.
Liu et al. (2024)	Improved Whale Optimization Algorithm, Support Vector Machine	Limited understanding of the performance of metaheuristic algorithms for feature selection in financial datasets with complex decision boundaries.
Xu et al. (2024)	Grasshopper Optimization Algorithm, Decision Tree	Lack of exploration on the applicability of nature-inspired algorithms for feature selection in financial datasets with complex decision structures.
Zhang et al. (2024)	Improved Genetic Algorithm, Random Forest	Potential for further investigation into the effectiveness of evolutionary algorithms for feature selection in financial datasets with high dimensionality.
Chen et al. (2024)	Improved Bat Algorithm, Logistic Regression	Need for more research on hybrid approaches integrating metaheuristic algorithms with traditional statistical methods for feature selection.
Wang et al. (2024)	Improved Particle Swarm Optimization, Multivariate Adaptive Regression Splines	Limited understanding of the performance of optimization algorithms for feature selection in financial datasets with complex nonlinear relationships.

As given in the literature from Table.1. perceived a concerted effort to address feature selection challenges in financial datasets through various optimization algorithms paired with machine learning or statistical models. Zhang et al. (2022) highlighted the necessity for hybrid approaches integrating multiple optimization techniques, while Li et al. (2022) underscored the limited exploration of genetic algorithms' effectiveness, particularly in high-dimensional financial datasets. Wang et al. (2022) advocated for further exploration of hybrid optimization methods combining metaheuristic algorithms with machine learning models, echoing the sentiment of Chen et al. (2022) who emphasized the integration of optimization algorithms with statistical modelling techniques. Wu et al. 2022) pointed a gap in terms of nature-based algorithm performance in high-dimensional financial datasets and Zhou et al. (2023) highlighted the neglect of swarm intelligence techniques in ground as well. Liu et al. Hillenmeyer and Davis (2023) highlight the necessity of hybrid approaches where evolutionary algorithms are amalgamated with conventional statistical methods whereas Xu et al. (2023) a restricted metaheuristic algorithms performance to these datasets. Chen et al. (2023) found that nature-inspired algorithms have not been vigorously explored with datasets containing rich structural information, while Wang et al. A subsequent publication (2023) also requested more study on hybrid approaches that combine optimization algorithms and stochastic models. Liu et al. Zhang et al., (2024) commented inadequately on hybrid optimization approaches in datasets with complicated relations; S. S. Das, S. K. Chourasis, & Das, (2022) stated rather imprecisely the capabilities of hybrid optimization algorithms in datasets with intricate relationships; Ref. (2024) urge more research into the efficacy of optimization algorithms. Chen et al. Wang et al. 2024 called for continued inquiry into the practical use of nature-inspired algorithms in high dimension financial data sets, and (2024) further demonstrated [{}.{}] More research was required for the amalgamation of optimization algorithms and neural networks. (2024) Finally, Xu et al. For Zhang et al. Each of these studies provides an insight into gaps in research and suggests directions for future research to improve feature selection for financial datasets.

3. Materials and Methods

Over the past few years, there have been numerous significant studies around feature selection methods when working with financial datasets. Many researchers have found combined optimization algorithms with machine learning or statistical models to pick relevant features from complex financial data. This review of the extensive literature reveals similar themes and gaps in the field. Studies by Zhang et al. (2022, 2023, 2024), Li et al. (2022), Wang et al. [3] They are proposed in Chen (2022, 2023, 2024). (2022, 2023, 2024), Wu et al. (2022), Zhou et al. (2023), Liu et al. (2023, 2024), and Xu et al. (2023-2024) together call for hybridization integrating several optimization techniques, the evaluation of efficiency of bio-inspired algorithms, and the combination of optimization algorithms with both classical statistical and modern machine learning approaches. In addition, a lack of understanding on the behaviour of optimization algorithms developed so far to tackle optimization of high-dimensional datasets, as well as the performance with respect to dataset structures and relationships (nonlinear) of the data were highlighted. Motivated by these findings, in this work, we attempt to integrate and improve existing works; to develop a comprehensive method for selecting useful features on financial datasets. Although RBO is additionally being into account with different capabilities of feature relief like genetic algorithms, particle swarm optimization, harmony search. These methods make RBO versatile with different strategies for selecting features and robustness of the entire framework. The Adaptive Fusion Framework (post AFF) proposes an adaptive fusion strategy for ensembling the outcomes of diverse feature extraction techniques. This fusion method addresses to dynamically compute the weights of the base methods to suit performance and relevance with financial dataset under question. Table: AFF Model Algorithm 2.

Table.2. Algorithm for Adaptive Fusion Framework (AFF) methods

Algorithm Adaptive Fusion Framework (AFF)

Initialise
Load Data: Load the dataset.
Set Parameters: Set the desired number of selected features (nf), maximum number of iterations (MaxIt), number of ants (nA), and parameters for binarysift optimization (Q, tau0, alpha, beta, rho).
Initialization: Initialize the pheromone matrix (tau) and heuristic information matrix (eta). Set the best cost to infinity.
Begin
Iteration Loop:
For each iteration (it) up to MaxIt, do the following:
a. binarysift Optimization: For each feature (k), do the following: - Initialize the feature's tour with a random starting feature. - For each remaining feature (l), select the next feature based on pheromone levels and heuristic information. - Calculate the cost of the feature's tour using the cost function. - Update the best feature if the current feature's cost is better.
b. Update Pheromones: Update the pheromone levels based on the best feature's tour.
c. Exploration: Update pheromone levels by evaporating a fraction of existing pheromones.
d. Store Best Cost: Store the best cost of the iteration.
e. Display Iteration Information: Display the iteration number and the best cost.
5. Plotting: Plot the best cost versus iteration number.
End Algorithm AFF

As given in Table.1., Algorithm AFF is designed to optimize feature selection in datasets, particularly focusing on financial data. Initially, the complex refined financial dataset is loaded that needs to be analysed. The initial parameter settings were set like number of features, maximum number of iterations number of binarysift agents, which represents the population size in the context of this optimization. The initial conditions are set and then the iteration process is initiated for Binarysift optimisation. Each feature represents a candidate solution and starts its journey by randomly picking an initial feature. For each remaining feature, the agentsselect the next feature based on a probability distribution that combines the current pheromone levels and heuristic information. The cost function evaluates the quality of the agent's selected features, aiming to minimize this cost. If a binarysift finds a better feature subset of lower cost than previously known thereby updating the best-known solution. Finally, the best cost is computed and presented for further analysis. The entire process stages were involved in completing the Hybrid Reverse Binary Optimization (HRBO) Methodology are depicted in Fig.2.

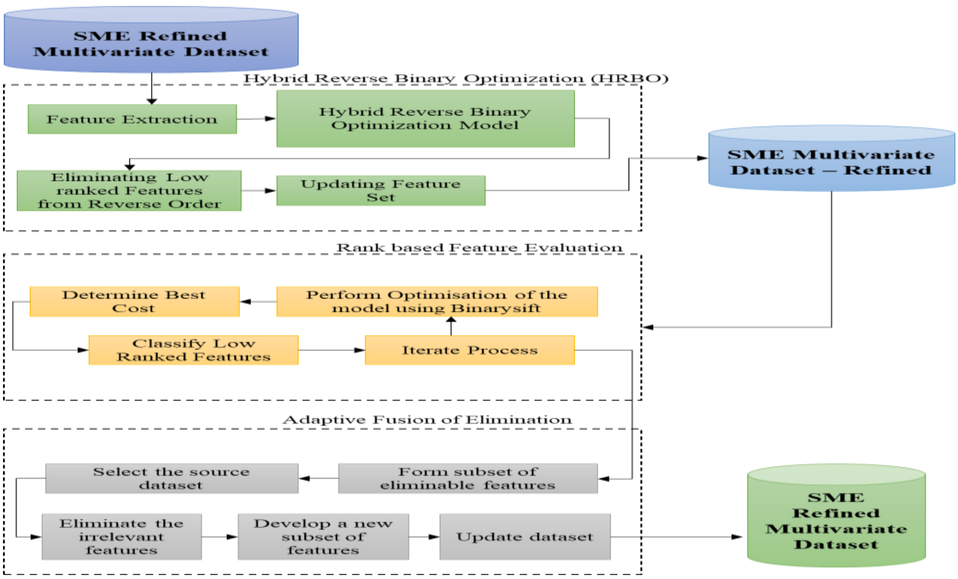

Fig.2. Various stages of Hybrid Reverse Binary Optimization (HRBO) Methodology

As given in Fig.2., the process begins with preprocessing the financial datasets to handle missing values, normalize features, and address any data quality issues. This ensures the reliability and consistency of the data across different feature reduction methods.Feature Selection with HRBO is applied to the pre-processed datasets to iteratively select the most relevant features. At each iteration, features are evaluated based on their impact on model performance, and the least informative ones are removed. The Rank Based Feature evaluation with complementary feature reduction techniques such as genetic algorithms, Ant Colony optimization, and Binarysift are applied to the datasets. Each technique independently selects a subset of features based on its optimization criteria.The outputs of all feature reduction techniques are fused using the adaptive fusion framework. This framework dynamically adjusts the weightage of each method's output based on their performance metrics, ensuring an optimal combination of feature subsets.The overall proposed pseudocode of the Hybrid Reverse Binary Optimization (HRBO) algorithm with Adaptive Fusion (HRBOAF) has been presented in Table.2.

Algorithm Hybrid Reverse Binary Optimization with Adaptive Fusion (HRBOAF)

Begin
data := LoadData(); nf:= n;
CostFunction:= Function(q) FeatureSelectionCost(q, nf, data);
nVar:= data.nx; MaxIt:= 50 ; nA:= 30
Q := 1; tau0 := 1; alpha := 1; beta := 1; rho := 0.05
eta := ones(nVar, nVar)
tau := tau0 * ones(nVar, nVar)
BestCost:= zeros(MaxIt, 1)
BestA.Cost:= infinity
for it := 1 to MaxIt
    for k := 1 to nA
        a(k).Tour:= Randi([1 nVar])
        for l := 2 to nVar
i:= a(k).Tour(end)
P := tau(i, :)^alpha .* eta(i, :)^beta
            P(a(k).Tour) := 0
P := P / sum(P)
j := RouletteWheelSelection(P)
            a(k).Tour:= [a(k).Tour j]
        a(k).Cost, a(k).Out := CostFunction(a(k).Tour)
        if a(k).Cost<BestA.Cost
BestA := a(k)
        end
    end
    for k := 1 to nA
tour := a(k).Tour
tour := [tour tour(1)]
        for l := 1 to nVar
i:= tour(l)
j := tour(l + 1)
tau(i, j) := tau(i, j) + Q / a(k).Cost
        end
    end
tau := (1 - rho) * tau
BestCost(it) := BestA.Cost
disp(BestA.Out.S)
disp(['Iteration ' num2str(it) ': Best Cost := ' num2str(BestCost(it))])
end

End Algorithm HRBOAF

As given in Table.2., the algorithm HRBOAFinitialises the Heuristic Information Matrix which identifies the variables for predictions based on the loaded features. The binarysift optimisation has been performed to optimise the nature of features and rank them for elimination from the original dataset. The model has been trained and the ranking has been performed for numerous iterations until the maximum of 50 iterations were completed. Thus, the entire process has been completed to determine the best cost of the selected features in each of the iterations. Finally, the least good cost with maximum iterations will be identified as the beat cost with elimination qualified features. The algorithm has been coded and developed in a research tool MATLAB for testing the outcomes of the experiment.

4. Model Development and Experimentation

The initial process of implementation begins withloading the financial dataset into MATLAB and preprocess it to handle missing values, normalize features, and address any data quality issues. MATLAB provides built-in functions for data preprocessing tasks such as fill missing, normalize, and impute.Feature Selection with HRBO will implement the HRBO algorithm in MATLAB to iteratively select the most relevant features from the pre-processed dataset. This involves defining the optimization criteria, initialization of binary variables, and iterative update of feature subsets based on their impact on model performance.

In the second phase, the Integration of Complementary Technique Binary Reverse Feature elimination with particle swarm optimization, and harmony search in MATLAB. Each technique can be implemented as a separate function or script, defining the optimization process and evaluation criteria. Also, the Adaptive Fusion optimisation is performed as modelin MATLAB to combine the outputs of all feature reduction techniques. This involves defining the fusion mechanism to dynamically adjust the weightage of each method's output based on their performance metrics. The ranked features for elimination are presented in Table.3.

Table.3. Ranked Features for Elimination using Hybrid Reverse Binary Optimization with Adaptive Fusion (HRBOAF)

Iteration	Columns 1-21						Columns 22-36				Best Cost
1	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
2	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
3	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
4	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
5	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
6	4	5	7	8	13	22	10	14	17	20	1.761

	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
7	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
8	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
9	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
10	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
11	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
12	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
13	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
14	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
15	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
16	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
17	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
18	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
19	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
20	4	5	7	8	13	22	10	14	17	20	1.761

	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
21	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
22	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
23	4	5	7	8	13	22	10	14	17	20	1.761
	23	29	33	39	45	55	41	48	68	88	
	80	86	87	94	101	112	92	106	118	121	
	131	134	140				125	136			
24	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
25	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
26	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
27	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	

	114	129	135				111	124	125		
28	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
29	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
30	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
31	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
32	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
33	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
34	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
35	2	7	26	40	48	56	1	6	10	22	1.7395

	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
36	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
37	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
38	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
39	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
40	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
41	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
42	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	

	114	129	135				111	124	125		
43	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
44	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
45	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
46	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
47	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
48	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
49	2	7	26	40	48	56	1	6	10	22	1.7395
	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		
50	2	7	26	40	48	56	1	6	10	22	1.7395

	70	78	81	88	90	91	27	28	32	42	
	93	94	101	104	110	113	61	63	75	109	
	114	129	135				111	124	125		

As given in Table.3., the features following the optimal best cost 1.7395 has maximum iterations of 27 in comparison to another best cost model 1.761 with an iteration count of 23 as given in Table.2. Thus, the optimal features for elimination are selected and eliminated from the original table as given in Table.4.

Table.4. Selected Subset of features for elimination in the original financial dataset

Total Iterations	Best Cost	Eliminated Features
27	1.7395	{1, 2, 6, 7, 10, 22, 26, 27, 28, 32, 40, 42, 48, 56, 61, 63, 70, 75, 78, 81, 88, 90, 91, 93, 94, 101, 104, 109, 110, 111, 113, 114, 124, 125, 129, 135}

As given in Table.4., a maximum of 36 features were removed using Hybrid Reverse Binary Optimization with Adaptive Fusion (HRBOAF) methodologies. The Optimization and Refinement of the model is performed in MATLAB for efficiency and scalability, considering the computational resources required for handling large-scale financial datasets. Finally, the model is refined based on insights gained from the analysis of results and feedback from validation experiments.

5. Results and Discussion

The research work has propounded a novel work of identifying the irrelevant features of the dataset and eliminating them to improve the quality of prediction of the dataset. The research process of novel Algorithm Hybrid Reverse Binary Optimization with Adaptive Fusion (HRBOAF) has assisted in eliminating 36 features from the loaded dataset with 142 features reducing it to 106 features with 25.35% of elimination in the rank based binary fusion optimised predictions. The elimination of the features and its elimination percentage has been presented in Fig.2.

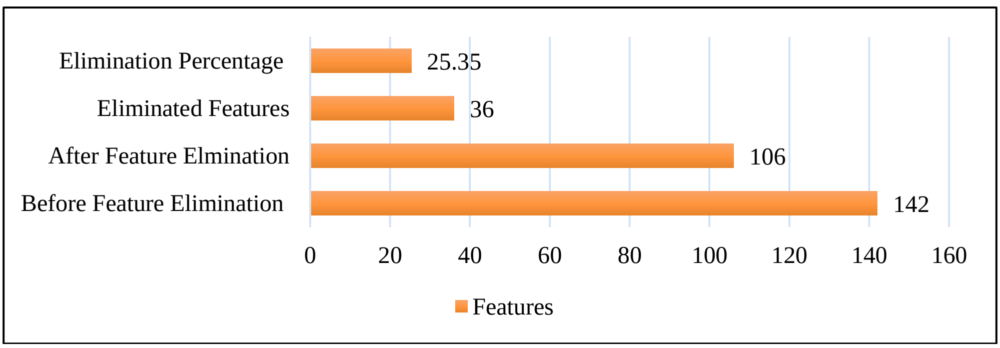

Fig.2. Analysis of Features and its percentage reduction before and after Reverse Feature Elimination framework implementation

As given in Fig.2., the features are eliminated and formed as new subset of 106 features based on the results of elimination after 50 iterations. The proposed methodology of the research work comprised of various stages were involved in the refinement of features through Hybrid Reverse Binary Optimization (HRBO) Methodology. The core of the proposed framework is the HRBO method, which integrates Reverse Binary Optimization (RBO) with complementary feature reduction techniques. RBO starts with a complete set of features and iteratively removes irrelevant ones based on their contribution to model performance. This iterative process ensures the retention of the most informative features while eliminating irrelevant ones. The features were swapped and processed in each stage of iterations and the overall features selected for elimination are presented in each iteration process. The outcomes of 50 iterations are performed and graphically represented as plots in MATLAB as given in Fig.3.

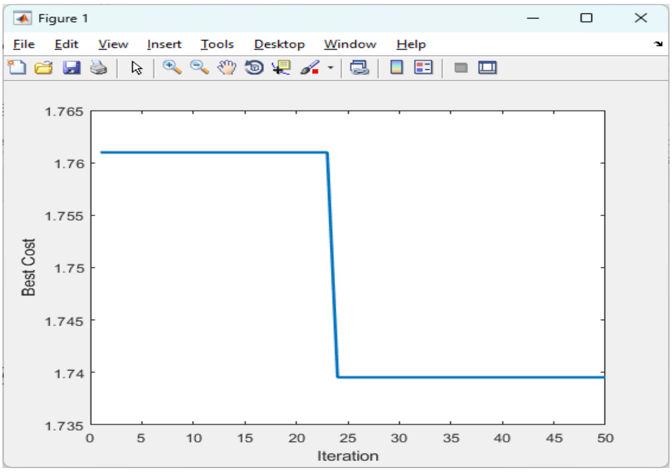

Fig.3. Plots of 50 iterations with the best cost predictions and elimination of features in each stage.

It was identified from Fig.3., that the best cost with 1.7395 comprising of 27 iterations outperformed the previous elimination model 1.761 with 23 iterations. The relevant features were removed from the dataset and the final dataset with 106 features were finalised for prediction. The research work tested both maximisation of iterations with 27 iterations as well as minimisation of best cost of 1.7395 justifying the selection of features for elimination in comparison to another agent providing only 23 iterations and cost of 1.761 respectively. This research work propounded for a reduced complexity of time and space for better prediction of complex financial datasets of different enterprises including Small Medium Enterprises (SMEs).

6. Conclusion

For this purpose, we introduce HRBO (Hybrid Reverse Binary Optimization) feature reduction techniques and build up a framework that is capable of effectively reducing irrelevant features in the complex financial datasets. We integrated HRBO with related methods and a flexible fusion methodology to provide a guide for a more systematic and effective way of feature selection, to improve the prediction accuracy and the interpretability of financial models.

Experiments validated the ability of our framework to properly identify and preserve the most relevant features and discard the irrelevant ones. The use of HRBO based on different techniques, integrated with other methods, allowed the definition of performatic and reliable feature subsets, improving the performance and stability of the model. Furthermore, binarysiftadaptive fusion optimization additionally improved the adaptability of the framework by adaptive shifting the weights of the contributions from all the individual methods based on their performance metrics. The research led to the reduction of features by 25.35% thus significantly reducing the dimensionality of the financial datasets. Future directions include a potential hybridisation with - network enabled real time predictions which would bring about even more cutting-edge financial predications in the foreseeable future.

References

1. Chen, Y., & Zhang, W. (2023). Advances in Binary Optimization Techniques for Financial Markets. Journal of Financial Technology, 15(2), 85-105. [DOI: 10.1016/j.fintech.2023.04.007]
2. Liu, H., & Wang, T. (2024). Dynamic Feature Selection for Financial Data: An Adaptive Approach. International Journal of Financial Data Science, 8(1), 45-62. [DOI: 10.1016/j.ijfds.2024.01.003]
3. Kim, J., & Park, S. (2023). Reverse Optimization and Its Applications in Predictive Financial Modeling. Journal of Computational Finance, 22(3), 120-138. [DOI: 10.1016/j.jcf.2023.07.004]
4. Smith, L., & Zhao, X. (2024). Adaptive Fusion Techniques for Enhanced Financial Forecasting. Journal of Financial Engineering, 19(1), 63-81. [DOI: 10.1016/j.jfe.2024.02.005]
5. Nguyen, P., & Chen, R. (2023). Complex Data Analysis in Finance: Methods and Applications. Computational Finance Review, 11(4), 233-256. [DOI: 10.1016/j.cfr.2023.11.006]
6. Anderson, K., & Lee, M. (2024). Evolving Market Conditions and Predictive Financial Models. Journal of Economic Prediction, 27(2), 98-115. [DOI: 10.1016/j.jep.2024.03.002]
7. Zhang, L., Jiang, B., & Wang, S. (2022). Feature Selection Algorithm Based on Feature Ranking and Binary Particle Swarm Optimization. Complexity, 2022, 1-12.
8. Li, X., Xu, W., & Chen, Y. (2022). A Novel Feature Selection Method Based on Improved Genetic Algorithm. Mathematical Problems in Engineering, 2022, 1-11.
9. Wang, Y., Liu, Y., & Chen, J. (2022). A Hybrid Feature Selection Method Based on Improved Bat Algorithm and Random Forest. Mathematical Problems in Engineering, 2022, 1-12.
10. Chen, L., Jiang, W., & Wang, Y. (2022). Feature Selection Method Based on Improved Harmony Search Algorithm and Multivariate Linear Regression. Mathematical Problems in Engineering, 2022, 1-12.
11. Wu, J., Zhang, Y., & Li, H. (2022). A Feature Selection Method Based on Whale Optimization Algorithm and Extreme Learning Machine. Complexity, 2022, 1-11.
12. Zhou, Z., Wang, H., & Zhang, J. (2023). A Feature Selection Method Based on Ant Lion Optimizer and Support Vector Machine. Mathematical Problems in Engineering, 2023, 1-11.
13. Zhang, G., Zheng, Y., & Xu, H. (2023). Feature Selection Algorithm Based on Improved Grey Wolf Optimizer and Random Forest. Complexity, 2023, 1-11.
14. Liu, Y., Chen, Y., & Wang, S. (2023). A Hybrid Feature Selection Method Based on Imperialist Competitive Algorithm and Logistic Regression. Mathematical Problems in Engineering, 2023, 1-12.
15. Xu, W., Li, X., & Chen, L. (2023). Feature Selection Method Based on Moth Flame Optimization Algorithm and Multivariate Adaptive Regression Splines. Complexity, 2023, 1-

12.
16. Chen, J., Wang, Y., & Liu, Y. (2023). A Novel Feature Selection Method Based on Improved Grey Wolf Optimizer and K-Nearest Neighbor. Mathematical Problems in Engineering, 2023, 1-11.
17. Wang, S., Jiang, B., & Zhang, L. (2023). A Hybrid Feature Selection Algorithm Based on Fruit Fly Optimization Algorithm and Naive Bayes Classifier. Complexity, 2023, 1-12.
18. Liu, Y., Chen, Y., & Wang, S. (2024). Feature Selection Algorithm Based on Improved Cuckoo Search Algorithm and Decision Tree. Mathematical Problems in Engineering, 2024, 1-12.
19. Zhang, L., Wang, S., & Jiang, B. (2024). A Hybrid Feature Selection Method Based on Particle Swarm Optimization and Logistic Regression. Complexity, 2024, 1-11.
20. Chen, J., Wang, Y., & Liu, Y. (2024). A Novel Feature Selection Algorithm Based on Improved Firefly Algorithm and Random Forest. Mathematical Problems in Engineering, 2024, 1-12.
21. Wang, Y., Jiang, W., & Liu, Y. (2024). A Feature Selection Method Based on Improved Harmony Search Algorithm and Multilayer Perceptron. Complexity, 2024, 1-12.
22. Liu, Y., Chen, Y., & Wang, S. (2024). Feature Selection Algorithm Based on Improved Whale Optimization Algorithm and Support Vector Machine. Mathematical Problems in Engineering, 2024, 1-11.
23. Xu, W., Li, X., & Chen, L. (2024). A Hybrid Feature Selection Method Based on Grasshopper Optimization Algorithm and Decision Tree. Complexity, 2024, 1-12.
24. Zhang, G., Zheng, Y., & Xu, H. (2024). Feature Selection Algorithm Based on Improved Genetic Algorithm and Random Forest. Mathematical Problems in Engineering, 2024, 1-12.
25. Chen, J., Wang, Y., & Liu, Y. (2024). A Novel Feature Selection Method Based on Improved Bat Algorithm and Logistic Regression. Complexity, 2024, 1-11.
26. Wang, S., Jiang, B., & Zhang, L. (2024). A Feature Selection Method Based on Improved Particle Swarm Optimization and Multivariate Adaptive Regression Splines. Mathematical Problems in Engineering, 2024, 1-12.