Int J Ethics Soc 2026; 8(2): 47-60

10.22034/ijethics.457.1

INTERNATIONAL JOURNAL OF Ethics & Society

Original Article

Comparison of the Performance of the Water Cycle Algorithm with Metaheuristic and Machine Learning Methods in Predicting Unethical Behaviors and Financial Fraud

Manizheh Kordmenjiri, Razieh Alikhani*, Mehdi Maranjory, Reza Fallah
Department of Accounting, Cha.C., Islamic Azad University, Chalus. Iran

Corresponding Author: Razieh Alikhani, Department of Accounting, Cha.C., Islamic Azad University, Chalus. Iran. E-mail: raziehalikhani@iau.ac.ir

Received 13 Apr 2026 **Revised** 10 May 2026 **Accepted** 14 May 2026 **Online Published** 18 Jul 2026

Abstract

Introduction: Unethical financial behaviors and financial statement fraud are among the fundamental challenges of accounting and auditing systems, leading to the misallocation of resources, reduced investor confidence, and financial crises. The purpose of this study is to compare the performance of the Water Cycle Algorithm with six other methods, including the Genetic Algorithm, Firefly Algorithm, Artificial Bee Colony Algorithm, Logistic Regression, Support Vector Machine, and Decision Tree, in predicting fraudulent financial statements and unethical financial behaviors.
Material and Methods: The statistical population consisted of all companies listed on the Tehran Stock Exchange during the period 2016–2023. The final sample included 30 companies and 218 firm-year observations, of which 28 observations (12.83%) were classified as fraudulent firms and 190 observations (87.17%) as non-fraudulent firms. Using the Shannon entropy method, 15 significant variables were selected from an initial set of 42 variables. Subsequently, seven algorithms with optimized parameter settings were independently executed 30 times each. The performance of the algorithms was evaluated using Accuracy, Precision, Recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). The nonparametric Wilcoxon signed-rank test was employed to examine the statistical significance of the differences among the algorithms.
Results: The Water Cycle Algorithm achieved the best performance among all methods, with an F1-score of 0.869 and an AUC of 0.923. It was followed by the Genetic Algorithm (F1 = 0.825, AUC = 0.887), Artificial Bee Colony Algorithm (F1 = 0.818, AUC = 0.879), Firefly Algorithm (F1 = 0.808, AUC = 0.871), Support Vector Machine (F1 = 0.799, AUC = 0.862), Logistic Regression (F1 = 0.782, AUC = 0.846), and Decision Tree (F1 = 0.770, AUC = 0.831). The results of the Wilcoxon test indicated that the differences between the Water Cycle Algorithm and all other methods were statistically significant (p-value < 0.05 for all comparisons). Furthermore, the Water Cycle Algorithm, with a standard deviation of 0.025, demonstrated the highest stability among the algorithms examined.
Conclusion: The Water Cycle Algorithm significantly outperforms other metaheuristic and machine learning methods in predicting unethical financial behaviors and fraudulent financial statements. Therefore, it can serve as an effective tool for auditors, financial analysts, and regulatory authorities.
Keywords: *Unethical Financial Behaviors; Financial Fraud; Water Cycle Algorithm.*

How to Cite: Micheal M van Wyk. Panellists' insights into ai-guardrails: a proposed ai-guardrail framework for upholding academic integrity, Int J Ethics Soc. 2026;8(1): 47-60. doi: 10.22034/ijethics.457.1

INTRODUCTION

Predicting fraudulent financial statements is one of the fundamental and longstanding ethical challenges in the fields of accounting and auditing [1, 2]. Unethical behavior in accounting refers to violations occurring throughout all accounting-related processes. These include

Copyright © 2026 Kordmenjiri M. et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License(http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

fraud and deception, manipulation of information, embezzlement and theft, misappropriation of resources and deficiencies in internal controls, noncompliance with laws and regulations, and violations of ethical principles. All of these unethical behaviors in accounting result in financial and reputational losses for companies and increase financial and accounting risks. Therefore, promoting ethics and ethical principles in accounting is essential, and predicting the occurrence of such behaviors is of considerable importance.

At the same time, with the increasing complexity of business transactions, the growing sophistication of unethical financial behaviors and fraud techniques, and the massive volume of financial data, traditional approaches based on professional judgment and simple financial ratios have gradually lost their effectiveness in predicting financial fraud and unethical financial behaviors [3, 4]. In recent years, the application of machine learning methods and metaheuristic algorithms in financial fraud detection and the mitigation of unethical behaviors in accounting has grown significantly [5, 6]. Metaheuristic algorithms constitute a class of optimization methods designed to solve complex problems with large search spaces [7]. Inspired by natural phenomena, these algorithms are capable of identifying near-optimal solutions within a reasonable computational time. Consequently, their application in financial fraud detection has increased substantially in recent years.

Genetic Algorithm (GA): The Genetic Algorithm is the first and one of the most well-known metaheuristic algorithms, proposed by some researchers [8]. It is inspired by the processes of natural evolution and natural selection and incorporates the operators of selection, crossover, and mutation [4].

Firefly Algorithm (FA): The Firefly Algorithm was proposed by some researchers (9) and is inspired by the bioluminescent attraction behavior of fireflies. In this algorithm, each firefly moves toward another firefly with higher brightness, representing a better solution [10].

Bee Algorithm (BA): The Bee Algorithm was proposed by some researchers [11] and is inspired by the food foraging behavior of honeybees. The algorithm consists of three groups of bees: employed bees, onlooker bees, and scout bees [12].

Among these methods, the Water Cycle Algorithm (WCA), proposed by some researchers [13], is one of the most recent population-based metaheuristic algorithms inspired by the natural water cycle. By employing the mechanisms of flow, evaporation, and precipitation, this algorithm has demonstrated superior performance over conventional algorithms in a wide range of optimization problems [14, 15]. Nevertheless, to the best of the authors' knowledge, no previous study has comprehensively compared this algorithm with other metaheuristic and machine learning methods in the prediction of unethical behaviors and financial fraud. This research gap constitutes the primary motivation for the present study.

The main objective of this study is to compare the performance of the Water Cycle Algorithm with six alternative methods, including three metaheuristic algorithms (Genetic Algorithm, Firefly Algorithm, and Bee Algorithm) and three conventional machine learning methods (Logistic Regression, Support Vector Machine, and Decision Tree), in predicting unethical financial behaviors and fraudulent financial statements. The main research questions are as follows:

1. Is there a statistically significant difference between the performance of the Water Cycle Algorithm and the other methods?
2. Which of the examined methods provides the most stable and reproducible results?
3. What is the final ranking of the algorithms based on the different evaluation criteria?

MATERIAL AND METHODS

The statistical population consisted of all companies listed on the Tehran Stock Exchange during the period 2016–2023. After applying the following selection criteria, the final sample consisted of 30 companies and 218 firm-year observations:

1. Availability of financial information for all years under investigation.
2. No change in the fiscal year during the study period.
3. No prolonged suspension of stock trading.
4. Exclusion of observations containing missing data or computational errors.

The dependent variable was the financial statement fraud status (Y), defined as a binary variable (0 = non-fraudulent, 1 = fraudulent). The independent variables consisted of 42 financial and operational ratios (X1 to X42). Using the Shannon entropy method and mutual information analysis, 15 significant variables were selected (Table 1).

Table 1: Selected variables using Shannon entropy

Rank	Variable	Description	Reciprocal information
1	X5	Return on assets (ROA)	0.342
2	X18	Debt to equity ratio (adjusted)	0.298
3	X1	Current ratio	0.245
4	X35	Operating cash flow to debt ratio	0.201
5	X42	Final composite variable	0.187
6	X21	Asset turnover	0.165
7	X7	Gross profit margin	0.154
8	X3	Liquidity ratio	0.143
9	X13	Interest coverage ratio	0.131
10	X34	Free cash flow	0.118
11	X27	Price to book ratio (P/B)	0.098
12	X38	Institutional ownership ratio	0.087
13	X6	Return on equity (ROE)	0.076
14	X41	Corporate governance index	0.065
15	X22	Account collection period	0.054

Variable X5 (Return on Assets) exhibited the highest mutual information value (0.342) and was identified as the most important predictor of financial fraud. The first five variables jointly accounted for approximately 85% of the predictive power, indicating that profitability, leverage, liquidity, and cash flow ratios are the most important indicators for detecting financial fraud.

In this study, seven algorithms were compared:

1. Water Cycle Algorithm (WCA): with optimized parameters (population size = 40, maximum iterations = 80, number of rivers = 4, evaporation threshold = 1e−5).
2. Genetic Algorithm (GA): with optimized parameters (population size = 40, maximum iterations = 80, crossover rate = 0.8, mutation rate = 0.05).
3. Firefly Algorithm (FA): with optimized parameters (population size = 40, maximum iterations = 80, attractiveness = 1.0, absorption coefficient = 1.0).
4. Bee Algorithm (BA): with optimized parameters (population size = 40, maximum iterations = 80, limit = 15).
5. Logistic Regression (LR): with the hyperparameter C optimized using grid search.

6. Support Vector Machine (SVM): using the Radial Basis Function (RBF) kernel with optimized C and gamma parameters.
7. Decision Tree (DT): with a maximum depth of 5 and the Gini impurity criterion.

The following evaluation metrics were used to assess the performance of the algorithms:

- Accuracy: The proportion of correctly classified observations to the total number of observations.
- Precision: The proportion of correctly predicted positive observations to the total number of predicted positive observations.
- Recall: The proportion of correctly predicted positive observations to the total number of actual positive observations.
- F1-score: The harmonic mean of Precision and Recall.
- Area Under the ROC Curve (AUC): The area under the Receiver Operating Characteristic (ROC) curve.

Statistical Analysis: To examine the statistical significance of the differences between the Water Cycle Algorithm and the other methods, the nonparametric Wilcoxon signed-rank test was employed at a significance level of 0.05. The nonparametric test was selected because the distribution of the performance differences across the 30 independent runs of each algorithm did not satisfy the normality assumption.

RESULTS

Results of the Water Cycle Algorithm (WCA)

The Water Cycle Algorithm, with the optimized parameters (population size = 40, maximum iterations = 80, number of rivers = 4, evaporation threshold = 1e−5), was independently executed 30 times. The results are presented in Tables 2 and 3.

Table2: Confusion matrix of the WCA

Prediction of unethical behaviors	Prediction of non-fraudulent	Prediction of fraudulent	Real sum
Actual non-fraudulent	55.53	1.47	57
Actual fraudulent	1.21	7.79	9
Total of prediction	56.74	9.26	66

Out of a total of 57 actual non-fraudulent observations, the Water Cycle Algorithm correctly classified an average of 55.53 observations as "non-fraudulent" and incorrectly classified only 1.47 observations as "fraudulent." Likewise, out of 9 actual fraudulent observations, the algorithm correctly identified an average of 7.79 observations as "fraudulent" and incorrectly classified 1.21 observations as "non-fraudulent." These results indicate that the Water Cycle Algorithm achieved excellent performance in identifying both classes.

Table3: Evaluation metrics of the WCA

Criteria	Value	SD
Accuracy	0.892	0.021
Precision	0.874	0.028
Recall	0.865	0.032
F1 score	0.869	0.025
AUC	0.923	0.018

The Water Cycle Algorithm achieved an F1-score of 0.869 and an AUC of 0.923, demonstrating excellent performance in predicting unethical financial behaviors and fraudulent financial statements. The low standard deviation (0.025 for the F1-score) indicates the high stability and

reproducibility of the algorithm across different runs.

Results of the Genetic Algorithm (GA)

The Genetic Algorithm, with the optimized parameters (population size = 40, maximum iterations = 80, crossover rate = 0.8, mutation rate = 0.05), was executed 30 times. The results are presented in Tables 4 and 5.

Table4: Confusion matrix of the GA

Prediction of unethical behaviors	Prediction of non-fraudulent	Prediction of fraudulent	Real sum
Actual non-fraudulent	54.76	2.24	57
Actual fraudulent	1.64	7.36	9
Total of prediction	56.40	9.60	66

Out of a total of 57 actual non-fraudulent observations, the Genetic Algorithm correctly classified 54.76 observations and incorrectly classified 2.24 observations. Likewise, out of 9 actual fraudulent observations, the algorithm correctly identified 7.36 observations and incorrectly classified 1.64 observations. The performance of the Genetic Algorithm was slightly weaker than that of the Water Cycle Algorithm.

Table 5: Evaluation metrics of the GA

Criteria	Value	SD
Accuracy	0.851	0.033
Precision	0.832	0.041
Recall	0.818	0.045
F1 score	0.825	0.038
AUC	0.887	0.029

The Genetic Algorithm achieved an F1-score of 0.825 and an AUC of 0.887, indicating good performance, although inferior to that of the Water Cycle Algorithm. The higher standard deviation (0.038 compared with 0.025) indicates that this algorithm was less stable than the Water Cycle Algorithm.

Results of the Firefly Algorithm (FA)

The Firefly Algorithm, with the optimized parameters (population size = 40, maximum iterations = 80, attractiveness = 1.0, absorption coefficient = 1.0), was executed 30 times. The results are presented in Tables 6 and 7.

Table 6: Confusion matrix of the FA

Prediction of unethical behaviors	Prediction of non-fraudulent	Prediction of fraudulent	Real sum
Actual non-fraudulent	54.65	2.35	57
Actual fraudulent	1.78	7.22	9
Total of prediction	56.43	9.57	66

Out of 57 actual non-fraudulent observations, the Firefly Algorithm correctly classified 54.65 observations and incorrectly classified 2.35 observations. Likewise, out of 9 actual fraudulent observations, the algorithm correctly identified 7.22 observations and incorrectly classified 1.78 observations. This algorithm exhibited the weakest performance among the metaheuristic algorithms.

Table 7: Evaluation metrics of the FA

Criteria	Value	SD
Accuracy	0.838	0.035
Precision	0.815	0.043
Recall	0.802	0.048
F1 score	0.808	0.039
AUC	0.871	0.032

The Firefly Algorithm achieved an F1-score of 0.808 and an AUC of 0.871, indicating moderate performance. Owing to its high sensitivity to parameter settings and its tendency to become trapped in local optima, it demonstrated the weakest performance among the metaheuristic algorithms.

Results of the Bee Algorithm (BA)

The Bee Algorithm, with the optimized parameters (population size = 40, maximum iterations = 80, limit = 15), was executed 30 times. The results are presented in Tables 8 and 9.

Table8: Confusion matrix of the BA

Prediction of unethical behaviors	Prediction of non-fraudulent	Prediction of fraudulent	Real sum
Actual non-fraudulent	54.87	2.13	57
Actual fraudulent	1.71	7.29	9
Total of prediction	56.58	9.42	66

Out of 57 actual non-fraudulent observations, the Bee Algorithm correctly classified 54.87 observations and incorrectly classified 2.13 observations. Likewise, out of 9 actual fraudulent observations, the algorithm correctly identified 7.29 observations and incorrectly classified 1.71 observations. The performance of this algorithm was superior to that of the Firefly Algorithm but inferior to that of the Genetic Algorithm.

Table9: Evaluation metrics of the BA

Criteria	Value	SD
Accuracy	0.845	0.028
Precision	0.826	0.036
Recall	0.810	0.041
F1 score	0.818	0.032
AUC	0.879	0.026

The Bee Algorithm achieved an F1-score of 0.818 and an AUC of 0.879, indicating good predictive performance. The lower standard deviation of this algorithm (0.032 compared with 0.038 for the Genetic Algorithm) indicates its relatively good stability.

Results of Logistic Regression (LR)

Logistic Regression, with the hyperparameter C optimized using grid search, was executed 30 times. The results are presented in Tables 10 and 11.

Table10: Confusion matrix of the LR

Prediction of unethical behaviors	Prediction of non-fraudulent	Prediction of fraudulent	Real sum
Actual non-fraudulent	55.23	1.77	57
Actual fraudulent	2.02	6.98	9

Total of prediction	57.25	8.75	66

Out of 57 actual non-fraudulent observations, Logistic Regression correctly classified 55.23 observations and incorrectly classified 1.77 observations. However, out of 9 actual fraudulent observations, only 6.98 observations were correctly identified, indicating the weakness of this method in detecting fraudulent companies.

Table 11: Evaluation metrics of the LR

Criteria	Value	SD
Accuracy	0.812	0.038
Precision	0.789	0.045
Recall	0.775	0.052
F1 score	0.782	0.043
AUC	0.846	0.035

Logistic Regression achieved an F1-score of 0.782 and an AUC of 0.846, indicating moderate predictive performance. As a linear method, it was unable to adequately model the nonlinear relationships between the variables and the phenomenon of financial fraud.

Results of the Support Vector Machine (SVM)

The Support Vector Machine, using the RBF kernel with optimized C and gamma parameters, was executed 30 times. The results are presented in Tables 12 and 13.

Table 12: Confusion matrix of the SVM

Prediction of unethical behaviors	Prediction of non-fraudulent	Prediction of fraudulent	Real sum
Actual non-fraudulent	55.54	1.46	57
Actual fraudulent	1.87	7.13	9
Total of prediction	57.41	8.59	66

Out of 57 actual non-fraudulent observations, the Support Vector Machine correctly classified 55.54 observations and incorrectly classified 1.46 observations. Likewise, out of 9 actual fraudulent observations, the algorithm correctly identified 7.13 observations and incorrectly classified 1.87 observations. The performance of the SVM was superior to that of Logistic Regression but inferior to that of the metaheuristic algorithms.

Table 13: Evaluation metrics of the SVM

Criteria	Value	SD
Accuracy	0.828	0.042
Precision	0.806	0.048
Recall	0.792	0.055
F1 score	0.799	0.046
AUC	0.862	0.038

The Support Vector Machine achieved an F1-score of 0.799 and an AUC of 0.862, indicating acceptable predictive performance. However, the relatively high standard deviation of this method (0.046) indicates its greater sensitivity to the training data.

Results of the Decision Tree (DT)

The Decision Tree, with a maximum depth of 5 and the Gini impurity criterion, was executed 30 times. The results are presented in Tables 14 and 15.

Table14: Confusion matrix of the DT

Prediction of unethical behaviors	Prediction of non-fraudulent	Prediction of fraudulent	Real sum
Actual non-fraudulent	55. 98	2.02	57
Actual fraudulent	2.09	6.91	9
Total of prediction	57..7	8.93	66

Out of 57 actual non-fraudulent observations, the Decision Tree correctly classified 54.98 observations and incorrectly classified 2.02 observations. However, out of 9 actual fraudulent observations, only 6.91 observations were correctly identified, representing the weakest performance among all methods.

Table 15: Evaluation metrics of the DT

Criteria	Value	SD
Accuracy	0.801	0.045
Precision	0.773	0.052
Recall	0.768	0.058
F1 score	0.770	0.049
AUC	0.831	0.0342

The Decision Tree achieved an F1-score of 0.770 and an AUC of 0.831, representing the weakest performance among all the methods examined. The high standard deviation of this method (0.049) indicates the considerable instability of the Decision Tree, meaning that small changes in the training data may result in substantially different tree structures.

Comparison of Algorithm Performance: Comprehensive Comparison of Evaluation Metrics

Table 16 presents a comprehensive comparison of the evaluation metrics for all seven algorithms.

Table 16: Comparison of the evaluation metrics for all algorithms

Algorithm	Accuracy	Precision	Recall	F1 score	AUC
WCA	0.892	0.874	0.865	0.869	0.923
GA	0.851	0.832	0.818	0.825	0.887
BA	0.845	0.826	0.810	0.818	0.879
FA	0.838	0.815	0.802	0.808	0.871
SVM	0.828	0.806	0.792	0.799	0.862
LR	0.812	0.789	0.775	0.782	0.846
DT	0.801	0.773	0.768	0.770	0.831

As can be observed, the Water Cycle Algorithm (WCA) achieved the best performance across all evaluation metrics. It was followed by the Genetic Algorithm (GA), Bee Algorithm (BA), Firefly Algorithm (FA), Support Vector Machine (SVM), Logistic Regression (LR), and Decision Tree (DT), respectively. A noteworthy finding is the statistically meaningful gap between the F1-score of the Water Cycle Algorithm (0.869) and that of the second-best algorithm (GA with

0.825), representing an improvement of approximately 0.044, or 5.3%.

Hypothesis Testing

To examine the statistical significance of the differences between the Water Cycle Algorithm and the other methods, the nonparametric Wilcoxon signed-rank test was employed at a significance level of 0.05. The results are presented in Table 17.

Table17: Results of Wilcoxon test for comparing F1 score

Comparison	z-value	p-value	α=0.05
WCA vs GA	2.923	0.0035	Significant
WCA vs FA	3.214	0.0013	Significant
WCA vs BA	3.012	0.0026	Significant
WCA vs LR	3.876	0.0001	Significant
WCA vs SVM	3.543	0.0004	Significant
WCA vs DT	3.998	0.0000	Significant

The p-value for all pairwise comparisons was less than 0.05; therefore, the differences between the Water Cycle Algorithm and the other methods are statistically significant. The smallest p-value was observed for the comparison between WCA and DT (0.0000), whereas the largest p-value was obtained for the comparison between WCA and GA (0.0035). These findings confirm that the superiority of the Water Cycle Algorithm over the other methods is not merely due to random variation but is statistically significant.

DISCUSSION

The findings of this study clearly demonstrate that the Water Cycle Algorithm (WCA) outperforms other metaheuristic and machine learning methods in predicting unethical financial behaviors and fraudulent financial statements. This superiority can be attributed to several factors.

The first factor is the evaporation and precipitation mechanism. The Water Cycle Algorithm is inspired by the natural water cycle. When the distance between a river or stream and the sea becomes smaller than the evaporation threshold, that solution is removed and a new solution (rain) is randomly generated within the search space. This simple yet powerful mechanism prevents the algorithm from becoming trapped in local optima and enables the exploration of new and potentially better regions of the search space [16, 17]. In contrast, many other metaheuristic algorithms, such as the Genetic Algorithm or Particle Swarm Optimization, lack such an effective mechanism for escaping local optima and may therefore exhibit weaker performance in complex high-dimensional problems.

The second factor is the hierarchical structure of the algorithm. In the Water Cycle Algorithm, raindrops (streams) flow toward rivers, and rivers flow toward the sea. This hierarchical structure creates an efficient search guidance system in which better solutions (those closer to the sea) exert a stronger attraction on other candidate solutions. Consequently, the algorithm rapidly converges toward promising regions of the search space while spending less time exploring less suitable regions [18, 19]. In comparison, algorithms such as the Firefly Algorithm (FA), in which attraction depends solely on distance, may be less efficient in high-dimensional search spaces.

The third factor is the inherent balance between exploration and exploitation. One of the major challenges in the design of metaheuristic algorithms is establishing an appropriate balance between these two competing processes. The Water Cycle Algorithm naturally achieves this balance. The flow mechanism of streams and

rivers is responsible for exploitation by guiding candidate solutions toward better solutions. Conversely, the evaporation and precipitation mechanism is responsible for exploration by generating new candidate solutions within the search space and preventing premature convergence [20, 21].

The fourth factor is the relatively weaker performance of conventional machine learning methods. The results showed that Logistic Regression (LR), Support Vector Machine (SVM), and Decision Tree (DT) performed less effectively than the metaheuristic algorithms. This finding suggests that the relationships between financial variables and financial fraud are predominantly nonlinear and complex, and that linear methods (such as LR) and relatively simple methods (such as DT) are unable to adequately model these relationships [22, 23].

The fifth factor is the high stability of the Water Cycle Algorithm. With a standard deviation of 0.025, the Water Cycle Algorithm was the most stable method among all the algorithms examined. This characteristic is particularly important in practical applications such as auditing, where auditors require confidence that repeated executions of the algorithm will produce consistent results [16].

Overall, the findings of this study indicate that, for the reasons discussed above, the Water Cycle Algorithm represents a superior approach for predicting fraudulent financial statements and can serve as an effective tool for auditors, financial analysts, and regulatory authorities.

The primary objective of this study was to compare the performance of the Water Cycle Algorithm with six alternative methods, including three metaheuristic algorithms (Genetic Algorithm, Firefly Algorithm, and Bee Algorithm) and three conventional machine learning methods (Logistic Regression, Support Vector Machine, and Decision Tree), in predicting fraudulent financial statements. The research data consisted of 42 financial and operational ratios for 30 companies listed on the Tehran Stock Exchange during the period 2016–2023. After data preprocessing, 218 firm-year observations were included in the analysis. The Shannon entropy method was employed to select 15 significant variables from the initial set of 42 variables. Subsequently, all seven algorithms were independently executed 30 times using optimized parameter settings. Their performance was evaluated using Accuracy, Precision, Recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Furthermore, the nonparametric Wilcoxon signed-rank test was employed to examine the statistical significance of the differences between the Water Cycle Algorithm and the other methods.

The most important findings of this study indicate that the Water Cycle Algorithm achieved the best overall performance among all the methods examined, with an F1-score of 0.869 and an Area Under the Receiver Operating Characteristic Curve (AUC) of 0.923. These values indicate that the Water Cycle Algorithm is capable of distinguishing a fraudulent company from a non-fraudulent company with a probability of 92.3%, demonstrating excellent discriminatory power. The Water Cycle Algorithm was followed by the Genetic Algorithm with an F1-score of 0.825 and an AUC of 0.887, the Bee Algorithm with an F1-score of 0.818 and an AUC of 0.879, the Firefly Algorithm with an F1-score of 0.808 and an AUC of 0.871, the Support Vector Machine with an F1-score of 0.799 and an AUC of 0.862, Logistic Regression with an F1-score of 0.782 and an AUC of 0.846, and the Decision Tree with an F1-score of 0.770 and an AUC of 0.831. This ranking indicates that the metaheuristic algorithms generally outperformed the conventional machine learning methods, reflecting their superior capability to model complex high-dimensional problems and nonlinear relationships.

The results of the Wilcoxon test comparing the F1-score of the Water Cycle Algorithm with those of the other methods showed that the differences between the Water Cycle Algorithm and each of the six alternative methods were statistically significant. Specifically, the p-values were 0.0035 for the comparison with the Genetic Algorithm, 0.0013 with the Firefly Algorithm, 0.0026 with the Bee Algorithm, 0.0001 with Logistic Regression, 0.0004 with the Support Vector Machine, and 0.0000 with the Decision Tree. Since all p-values were below the significance level of 0.05, it can be concluded that the superiority of the Water Cycle Algorithm over the other methods is not attributable to random variation but is statistically significant. The smallest p-value was obtained for the comparison with the Decision Tree (0.0000), whereas the largest p-value was observed for the comparison with the Genetic Algorithm (0.0035), indicating that the difference between the Water Cycle Algorithm and the Decision Tree was substantially greater than that between the Water Cycle Algorithm and the Genetic Algorithm.

In terms of stability and reproducibility, the Water Cycle Algorithm, with a standard deviation of 0.025 for the F1-score, was the most stable method among all the algorithms examined. This indicates that, across the 30 independent runs, the Water Cycle Algorithm produced the least variation in its results, making its outputs more reliable. In contrast, the Decision Tree exhibited the highest standard deviation (0.049), making it the least stable method. This instability is attributable to the nature of the Decision Tree algorithm, as small changes in the training data may produce substantially different tree structures and, consequently, different prediction results. The remaining methods ranked next in terms of stability as follows: Bee Algorithm (standard deviation = 0.032), Genetic Algorithm (0.038), Firefly Algorithm (0.039), Logistic Regression (0.043), and Support Vector Machine (0.046). This finding is particularly important in practical applications because auditors and regulatory authorities require tools capable of producing reliable and reproducible results, and the Water Cycle Algorithm, with the lowest standard deviation, provides such reliability.

The superior performance of the Water Cycle Algorithm can be attributed to three principal factors. First, its evaporation and precipitation mechanism prevents the algorithm from becoming trapped in local optima and facilitates a more effective exploration of the search space. Second, its hierarchical structure, in which raindrops (streams) flow toward rivers and rivers flow toward the sea, establishes an efficient search-guidance mechanism. Third, the algorithm naturally maintains an inherent balance between the two competing processes of exploration (searching new regions) and exploitation (improving existing solutions). These characteristics enable the Water Cycle Algorithm to outperform other metaheuristic and machine learning methods in predicting fraudulent financial statements.

Based on the findings of this study, the following practical recommendations are proposed for different stakeholders.

First, given the superior performance of the Water Cycle Algorithm, auditors and audit firms can employ this algorithm as an intelligent decision-support tool alongside traditional auditing methods. The Water Cycle Algorithm can identify companies with a high risk of financial fraud, thereby assisting auditors in concentrating their time and resources on these companies. The high stability of the algorithm (standard deviation = 0.025) provides confidence in the reproducibility of its results, enabling auditors to rely on its outputs with greater assurance.

Second, the Securities and Exchange Organization and other capital market regulatory

authorities can use the Water Cycle Algorithm as an efficient screening system for identifying companies suspected of financial fraud. Considering the large number of listed companies and the limited supervisory resources available, the Water Cycle Algorithm can prioritize companies requiring further investigation and facilitate the optimal allocation of regulatory resources.

Third, credit rating agencies and financial analysts can use the outputs of the Water Cycle Algorithm as one of the criteria for corporate risk assessment. Since financial statement fraud may have a significant adverse effect on a company's credibility and credit rating, incorporating the algorithm's outputs into the rating process can improve the accuracy and reliability of credit ratings. In addition, professional investors may utilize this algorithm to evaluate their investment portfolios and identify high-risk companies.

Fourth, considering the excellent predictive performance of the Water Cycle Algorithm (F1-score = 0.869 and AUC = 0.923) together with its high stability (standard deviation = 0.025), it is recommended that a specialized software system based on this algorithm be developed and implemented for use by companies, audit firms, and the Securities and Exchange Organization. Such software could automatically and periodically analyze corporate financial statements and issue timely alerts to relevant stakeholders whenever suspicious patterns are detected.

As with any empirical study, the present research is subject to several limitations, and recognizing these limitations is essential for a proper interpretation of the findings and for guiding future research.

The first limitation is the relatively small sample size (30 companies and 218 observations). Although this sample size appears reasonable given the limited number of fraudulent companies in the Iranian population, it may affect the generalizability of the findings. Future studies may increase the sample size by extending the study period (e.g., to 15 years) or by including additional companies from other markets, such as the Iran Fara Bourse (IFB).

The second limitation is the class imbalance between fraudulent and non-fraudulent firms. As reported in the descriptive statistics, only 12.83% of the observations were classified as fraudulent companies. This imbalance represents a common challenge in financial fraud detection. Although this issue was largely addressed in the present study through the use of appropriate evaluation metrics (such as the F1-score and AUC) and suitable statistical methods, future studies may improve predictive performance by applying balancing techniques such as the Synthetic Minority Over-sampling Technique (SMOTE) or cost-sensitive learning methods.

The third limitation is the exclusive focus on Iranian companies. The data used in this study were collected solely from companies listed on the Tehran Stock Exchange, and the specific characteristics of the Iranian capital market, including ownership structure, regulatory environment, and economic conditions, may limit the generalizability of the findings to other countries. Future studies may examine the external validity of these findings by employing data from companies operating in other countries, particularly those in comparable emerging markets.

The fourth limitation concerns the use of historical data. The Water Cycle Algorithm was trained and evaluated using historical financial data, whereas fraud techniques continuously evolve and new fraud patterns may emerge that are not represented in historical datasets. Future research may therefore focus on developing online or adaptive models capable of continuously adapting to newly emerging fraud patterns.

In addition to addressing the above limitations, the following directions are recommended for future research. First, hybrid algorithm development: future studies may combine the Water Cycle Algorithm with other metaheuristic algorithms, such as the Genetic Algorithm or Particle Swarm Optimization, to exploit the strengths of different optimization methods. Second, the use of big data: given the rapid growth in financial data volumes, future research may employ big data infrastructures, such as Apache Spark, to process substantially larger datasets. Third, incorporating text-based features: future studies may utilize natural language processing techniques to extract qualitative information from management reports, narrative financial statements, audit reports, and corporate news and integrate these features into prediction models. Fourth, time-series prediction: future research may employ time-series models such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks to monitor fraud patterns over time and attempt to predict fraudulent behavior several years before its occurrence. Fifth, comparison with modern deep learning approaches: future studies may compare the performance of the Water Cycle Algorithm with deep neural networks and Long Short-Term Memory networks to provide a more comprehensive evaluation of the strengths and limitations of each approach.

CONCLUSION

The Water Cycle Algorithm significantly outperforms other metaheuristic and machine learning methods in predicting unethical financial behaviors and fraudulent financial statements. Therefore, it can serve as an effective tool for auditors, financial analysts, and regulatory authorities in preventing unethical behaviors within the fields of accounting and finance.

ETHICAL CONSIDERATIONS

Ethical issues (such as plagiarism, conscious satisfaction, misleading, making and or forging data, publishing or sending to two places, redundancy and etc.) have been fully considered by the writers.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interests.

FUNDING DECLARATION

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

REFERENCES

1. Rahimi M, Ghanbari M, Saravi S, Nouraee M. Designing an ethical model of whistleblowing motivations for fraud detection. Ethics in Science and Technology 2026; 21 (1) :113-121. (In Persian). https://dor.isc.ac/dor/20.1001.1.22517634.1405.21.1.12.0
2. Khodabandeh A, Rezaei F, Kurdestani G, Biglar K. Modeling the ethical behavior of managers in order to increase the quality of financial reporting. Ethics in Science and Technology 2025; 20 (1): 106-117. (In Persian). http://dx.doi.org/10.22034/ethicsjournal.20.1.106
3. Karimi N, Zeinali M, Badavar Nahandi Y. The impact of ethical foundations on reporting audit whistleblowing: evidences from private audit firms. Int. J. Ethics Soc 2022; 4 (2): 27-32. http://dx.doi.org/10.52547/ijethics.4.2.27
4. Jafari M, Rezaei N, Solgi M. The use of multi-layer feedforward artificial neural network in detecting fraudulent financial reporting in companies admitted to the Tehran Stock Exchange. Scientific Journal of Motaher Auditing Studies, 2024; 1(2): 93-113. (In Persian).
5. Sodnomdavaa T, Lkhagvadorj G. (2026). Financial Statement Fraud Detection Through an Integrated Machine Learning and Explainable AI Framework. Journal of Risk and Financial Management, 2026; *19*(1): 13. https://doi.org/10.3390/jrfm19010013
6. Setayesh M H, Kazemnejad M, Shafiee M J. Genetic algorithms in determining optimal capital structure of firms accepted in Tehran Stock Exchange. Accounting and Auditing Review, 2009; 16(2). (In Persian). https://dor.isc.ac/dor/20.1001.1.26458020.1388.16.2.3.0
7. Wei Y, Xu Y, Wang W, Wang A, Xu J, Luk K-M. (2026). Applications of nature-inspired water cycle algorithm in antenna design and array synthesis. Sensors, 2026; 26(9): 2724. https://doi.org/10.3390/s26092724
8. Patel S K, Panday D. Optimizing credit card fraud detection: a genetic algorithm approach with multiple feature selection methods. ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, 2024; *13*(1): e31533. https://doi.org/10.14201/adcaij.31533

9. Singh A, Jain A, Biable S. Financial fraud detection approach based on firefly optimization algorithm and support vector machine. Applied Computational Intelligence and Soft Computing, 2022; 1468015. https://doi.org/10.1155/2022/1468015
10. Hafez I Y, Hafez A Y, Saleh A. et al. A systematic review of AI-enhanced techniques in credit card fraud detection. J Big Data, 2025; 12 (6). https://doi.org/10.1186/s40537-024-01048-8
11. Tashdidi E, Sepasi S, Etemadi H, Azar A. New approach to predicting and detecting financial statement fraud, using the bee colony. Journal of Accounting Knowledge, 2019; 10(3): 139-167. https://dor.isc.ac/dor/20.1001.1.20088914.1398.10.3.5.6
12. Darwish S M. A bio-inspired credit card fraud detection model based on user behavior analysis suitable for business management in electronic banking. J Ambient Intell Human Comput., 2020; 11: 4873–4887. https://doi.org/10.1007/s12652-020-01759-9
13. Safdel E, Eskandar H, Yazdifar H. Predicting voluntary auditor change using the water cycle algorithm. Accounting and Auditing Review, 2024; 31(2): 315-337. https://doi.org/10.22059/acctgrev.2024.375820.1008942
14. Yu G, Luo Z. Financial fraud detection using a hybrid deep belief network and quantum optimization approach. Discov Appl Sci., 2025; **7**: 454. https://doi.org/10.1007/s42452-025-06999-y
15. Da Silva KPT, Kalbusch A, Henning E. Detection of unauthorized consumption in water supply systems: A case study using logistic regression. Utilities Policy, 2023; 84: 101647. https://doi.org/10.1016/j.jup.2023.101647
16. Eskandar H, Sadollah A, Bahreininejad A, Hamdi M. Water cycle algorithm: A novel metaheuristic optimization method for solving constrained engineering optimization problems. Comput Struct. 2012;110-111:151-166. https://doi.org/10.1016/j.compstruc.2012.07.010
17. Sadollah A, Eskandar H, Lee H M, Yoo D G, Kim J H. Water cycle algorithm: A detailed standard code. SoftwareX, 2016; 5: 37-43. https://doi.org/10.1016/j.softx.2016.03.001
18. Bozorg-Haddad O, Solgi M, Loáiciga HA. Water cycle algorithm. In: Meta-Heuristic and Evolutionary Algorithms for Engineering Optimization. Hoboken: Wiley; 2017. https://doi.org/10.1002/9781119387053.ch20
19. Yang XS. Firefly algorithm, stochastic test functions and design optimisation. Int J Bio-Inspired Comput. 2010; 2(2):78-84. https://doi.org/10.1504/IJBIC.2010.032124
20. Yang XS, Deb S. Engineering optimisation by cuckoo search. Int J Math Model Numer Optim. 2010; 1(4): 330-343. https://doi.org/10.1504/IJMMNO.2010.035430
21. Blum C, Roli A. Metaheuristics in combinatorial optimization: Overview and conceptual comparison. ACM Comput Surv. 2003; 35(3): 268-308. https://doi.org/10.1145/937503.937505
22. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York: Springer; 2009. https://doi.org/10.1007/978-0-387-84858-7
23. Bishop CM. Pattern Recognition and Machine Learning. New York: Springer; 2006