Ingénierie des Systèmes d'Information
Vol. 30, No. 4, April, 2025, pp. 855-864

Journal homepage: http://iieta.org/journals/isi

A Bio-Inspired Grey Wolf Approach to Enhancing Fake Profile Detection in Online Social Media

Nawel Sekkal[1*], Nadir Mahammed[2], Zouaoui Guellil[3]

[1] LRIT Lab., Faculty of Sciences, University Abou Bekr Belkaid, Tlemcen 13000, Algeria
[2] LabRI-SBA Laboratory, Ecole Supérieure en Informatique Sidi Bel Abbès, Sidi Bel Abbès 22000, Algeria
[3] LabRI-SBA Laboratory, University Hassiba Benbouali of Chlef, Chlef 2000, Algeria

Corresponding Author Email: n.sekkal@esi-sba.dz

https://doi.org/10.18280/isi.300403

Received: 26 September 2024
Revised: 8 April 2025
Accepted: 22 April 2025
Available online: 30 April 2025

Keywords:
computational intelligence, data science, online social network, fake profiles detection, metaheuristic, bio-inspired algorithm, machine learning, dataset

ABSTRACT

The rapid increase in user activity on online social networks (OSNs) has gathered widespread attention. Despite this development, it faces significant challenges due to the rise of fake accounts, which misrepresent real users and invade on privacy regulations within these digital communities. As a result, it is essential to detect and eliminate such profiles to improve the security of social media users. In response, recent research has increasingly focused on machine learning techniques to address this issue. Numerous studies have explored and compared various machine learning-based approaches, yet there remains a gap in the literature, particularly in terms of a comprehensive analysis across different social media platforms. Furthermore, bio-inspired algorithms have received minimal attention in this context. Our study introduces a novel perspective by performing an extensive comparative analysis of different fake profiles detection methods on social media. Our findings demonstrate that both supervised and unsupervised machine learning models are effective in identifying fraudulent accounts on social media platforms. However, the application of suitable bio-inspired metaheuristics has the potential to surpass the results of existing approaches.

1. INTRODUCTION

Online social networks (OSNs) have become one of the most widely used applications, serving as a vital tool for connecting people globally and facilitating the sharing of diverse content such as videos, photos, and messages. The expansion of OSNs has accelerated significantly in recent decades, driven by technological advancements. As of today, approximately 5.35 billion people worldwide have access to the Internet, with 5.04 billion actively engaged on social media platforms (https://datareportal.com/reports/digital-2024-global-overview-report). However, this rapid growth has also brought a substantial rise in fake accounts, which pose serious challenges and encourage various forms of harmful behavior, including political manipulation, the spread of misinformation, misleading advertisements, terrorist propaganda, and hate speech. These fraudulent profiles can be categorized into multiple types, such as compromised profiles, cloned or duplicated profiles, and bots [1]. Commonly referred to as "fake profiles," these profiles represent a significant threat to the security and privacy of OSNs. In addition, platforms like Facebook, Twitter, Instagram, and others have seen an alarming increase in the number of fake users, many of whom were created with malicious intent [2].

The credibility and reputation of OSNs have been significantly impacted due to various security challenges, particularly concerning the protection of users' privacy from fake profiles. In response, researchers have recently turned to machine learning (ML) algorithms to automate and improve the detection of fraudulent profiles. Algorithms like K-Nearest Neighbor (K-NN), Support Vector Machine (SVM), Decision Trees (DT), and Random Forest (RF) have been employed for this purpose. Given that most OSNs make their user data publicly accessible by default [2], they have become the primary focus in recent studies, surveys, and reviews aimed at examining and comparing different ML-based approaches [3]. Despite numerous studies on machine learning techniques for fake profile detection, a significant gap exists regarding comprehensive analyses across various social media platforms, particularly involving bio-inspired algorithms. Additionally, while some research has explored the use of metaheuristics algorithms, few studies have integrated bio-inspired algorithms to enhance the existing detection methodologies and approaches.

This study aims to explore and evaluate the application of bio-inspired algorithm and ML algorithm for detecting fake profiles in OSNs. Specifically, we propose a comprehensive model that combines the optimization capabilities of a bio-inspired algorithm with the classification ability of ML algorithms. The proposed model is evaluated using a well-known bio-inspired algorithm, the Gray Wolf Optimizer (GWO) [4] and a selection of ML algorithms.

Advancing existing research, this work presents a comparative analysis of various algorithms used for detecting fake profiles in OSNs. Our study encompasses multiple platforms, including Facebook, Twitter, and Instagram, while also examining the potential of bio-inspired algorithms.

This research aims to achieve several key objectives:

•Establish a model that integrates bio-inspired optimization with ML classification algorithms.

•Analyze the performance of GWO in optimizing both feature selection and model parameters.

•Conduct a comparative analysis of different ML algorithms for classifying profiles.

•Test the proposed model using real-world datasets from platforms such as Facebook, Twitter, and Instagram.

The remainder of this paper is organized as follows: Section 2 provides a review of related work on fake profiles detection. Section 3 introduces the materials and methodologies used in this study, including datasets, preprocessing steps, and implementation details. Section 4 presents the results and discusses the performance of the proposed model. Finally, Section 5 concludes the paper and outlines potential future research directions.

2. RELATED WORK

This section reviews various approaches for identifying fake profiles on OSNs, highlighting the use of numerous ML algorithms.

The framework for detecting fake profiles on OSNs [5] employs open-source big data tools and Long Short-Term Memory (LSTM) networks for analysis. It integrates the Dispersive Flies Optimization (DFO) metaheuristic to enhance feature selection from the dataset. Additionally, the approach underscores the importance of ethical data collection, taking into account both public and private user attributes.

The issue of detecting fake profiles on Facebook was addressed through a hybrid methodology [6]. This approach involves a two-phase process: first, the Satin Bowerbird Optimization Algorithm (SBO) is used to establish initial clusters and determine optimal centroids for profile classification. Then, the K-means clustering algorithm is applied to categorize each profile as either real or fake.

A bio-inspired algorithm called the Fire Hawk Optimizer (FHO) was introduced to address the issue of fake profile detection [7]. Various feature groups from a Twitter dataset were evaluated to assess their effectiveness in identifying fake profiles. Based on their findings, they recommended the use of the Gradient Boosting Classifier (GBC) as the Fitness function.

The research [8] dealt with the problem of fake profiles on social media by introducing a novel method that integrates various ML algorithms to evaluate user behavior and profile information. This approach, termed "ensemble," employs a Majority Voting Technique (MVT) to classify profiles as either fake or real. The findings indicate that this method holds significant potential for improving the security of social media platforms.

Machine learning algorithms were applied to detect fake profiles on OSNs, with multiple models evaluated for effectiveness in identifying fraudulent accounts [9]. The study emphasized the importance of diverse evaluation approaches, such as confusion matrices and error rate analyses, to determine the optimal model.

In the context of existing research, Mahammed et al. [10] assessed how effectively the Fire Hawk Optimizer (FHO) detects fake social media profiles. The authors tested FHO's performance. Their goal was to identify the most effective feature subsets from a Facebook dataset to differentiate between real and fake profiles.

A novel method was introduced utilizing an extensive feature set to examine profile information, network connections, and user behavior for fake profile detection. An adjustable Bagged Tree Algorithm (BTA) was also proposed, enhancing decision tree models by eliminating irrelevant branches to improve accuracy and efficiency [11].

An ML model combining Logistic Regression (LR) with Gradient Descent Optimization algorithm (GBO) was introduced to identify fake profiles on OSNs [12]. When tested on an Instagram dataset, the model demonstrated strong performance.

Raghavendra et al. [13] presented detailed solution framework for detecting and eliminating fake profiles on OSNs. This solution utilizes ML algorithms to examine different aspects of user behavior and identify related anomalies. It highlights the importance of user education, ongoing updates, and privacy-respecting practices. The approach incorporates a variety of algorithms, including ANN, RF, Extreme Gradient Boost (XGBoost), LSTM, and a Voting Classifier (VC).

Table 1 offers a detailed synopsis of the latest advances in detecting fake profiles across different social media using ML algorithms.

Table 1. Related work synopsis

Ref.	OSN	ML	Meta	Spec	Dataset Size	Acc
[5]	Facebook	LSTM	DFO	-	-	0.979
[6]	Facebook	k-means	SBO	Hybrid	1244	0.989
[7]	Twitter	GBC	FHO	Hybrid	17350	0.996
[8]	Twitter	MVT	-	Combo	6825	0.991
[9]	Instagram	RF	-	Compo	6868	0.997
[10]	Facebook	GBC	FHO	Hybrid	1244	0.998
[11]	Facebook	BTA	-	-	-	0.999
[12]	Instagram	LR, GBO	-	Combo	7500	0.927
[13]	Facebook	LSTM	DFO	-	-	0.979

Table 1 provides a comprehensive overview of the information. ML: Machine Learning, Meta: Metaheuristic, Spec: Specification, Combo: Combination, Compo: Composition, Acc: Accuracy. It indicates that various OSNs are employed in the detection of fake profiles, with Facebook being the most frequently examined platform, followed by Twitter and Instagram. GBC and LSTM are the most commonly used ML algorithms, each has demonstrated strong performance in detecting fake profiles. Other algorithms employed less frequently but still contribute valuable insights.

Metaheuristic techniques, particularly bio-inspired algorithms, are less common compared to traditional ML algorithms. FHO, SBO and BTA are featured in numerous studies. Although less prevalent, these metaheuristics offer alternative approaches to improving detection accuracy.

Hybrid and combination techniques are prominent, with three studies using hybrid methods that integrate various ML algorithms and bio-inspired algorithms. These approaches often lead to high accuracy rates, reflecting their effectiveness. Additionally, two studies employ combination techniques to enhance detection accuracy.

Dataset sizes in these studies vary, from relatively small datasets of around a thousand to 7,500 accounts to much larger datasets, such as 17,350 accounts [7]. Despite this variability, the results are notable for their high accuracy.

Table 1 clearly demonstrates that both ML and bio-inspired algorithms play decisive roles in the detection of fake profiles on OSNs. ML algorithms such as GBC is highly effective and widely used, achieving high accuracy rates in distinguishing between fake and real profiles. Bio-inspired algorithms, though less common, offer valuable optimization capabilities and can significantly enhance detection systems when combined with traditional ML algorithms.

The present work is positioned at the intersection of ML and bio-inspired algorithm combinations for fake profile detection, as deduced from Table 1. The materials and methods section introduces the proposed model based on the well-known bio-inspired GWO, utilizing an ML algorithm as the Fitness function for feature selection. The model is then applied to different datasets, using various ML algorithms to assess its performance.

3. MATERIAL AND METHOD

This section begins with a detailed examination of the datasets used, followed by an overview of the preprocessing techniques applied to prepare the data. Subsequently, a concise review of the machine learning algorithms considered for this task is presented. The section concludes by introducing the core focus of this work: bio-inspired algorithm. This includes a discussion of their design principles, operational mechanisms, and the Fitness functions employed in the context of detecting fake profiles.

3.1 Datasets

The chosen datasets from Facebook, Twitter, and Instagram represent diverse user behaviors and profile characteristics commonly targeted by fake profiles, thus providing a comprehensive evaluation across different social media environments.

3.1.1 Facebook dataset

The dataset utilized in this study was sourced from Facebook [14]. It consists of 1244 instances, characterized by 14 attributes (Table 2). The dataset is categorized as follows:

Real Accounts: Comprising 1043 accounts.

Fake Accounts: Including 201 fake profiles.

Table 2. Facebook dataset features

Feature	Description
Name-Id	Unique identifier assigned to each profile.
Profile Picture	Image associated with the user's profile.
Likes	Indicates user appreciation for specific content.
Number of Likes	Specifies whether the number of likes is mentioned.
Number of groups joined	Total groups the user has joined.
Number of friends	Count of mutual friends the user has.
Education status	Indicates if the user's education is mentioned.
Work	Specifies whether work information is provided.
Living place	Indicates if the user's location is mentioned.
Relationship	Indicates if the user's relationship status is mentioned.
Checkin	Feature allowing users to share their location.
Number of posts	Total content shared by the user.
Number of tags	Feature enabling users to tag other users.
Profile intro	Introductory information used to identify the account.

3.1.2 Twitter dataset

This dataset was sourced from the social media platform Twitter [15]. It comprises 1000 instances, each described by 16 attributes (Table 3). The dataset is divided into the following categories:

Real Accounts: Consisting of 499 real profiles.

Fake Accounts: Including 501 fake profiles.

Table 3. Twitter dataset features

Feature	Description
Description	Length of the user-generated string that describes the account.
Protected	Indicates whether the user has chosen to protect their Tweets (true/false).
Followers count	The total number of followers the account currently has.
Friends count	The number of users this account is following.
Statuses count	Total number of Tweets (including retweets) posted by the user.
Favorites count	The total number of Tweets the user has liked throughout the account's history.
Listed count	The number of public lists in which this user is included.
Verified	Indicates if the user's account is verified (true/false).
Background image	Indicates whether the user has opted to use their uploaded background image (true/false).
Contributors enabled	Specifies if the user has enabled "contributor mode" for their account.
Default profile	Indicates whether the user has kept the default profile theme or background (true/false).
Default profile image	Indicates if the user has not uploaded a custom profile picture, using a default image instead (true/false).
Is translator	Specifies if the user is a member of Twitter's translator community (true/false).
Hashtags average	The average number of hashtags used by the user in their last 20 tweets.
Mentions average	The average number of mentions included in the user's last 20 tweets.
URLs average	The average number of URL links included in the user's last 20 tweets.

3.1.3 Instagram dataset

This dataset, sourced from Kaggle (https://www.kaggle.com/free4ever1/instagram-fake-spammer-genuineaccounts), originates from the OSN Instagram. It comprises 696 instances, characterized by 15 attributes (Table 4). The dataset is categorized into the following groups:

Real Accounts: Comprising 348 real profiles.

Fake Accounts: Including 348 false profiles.

Table 4. Instagram dataset features

Feature	Description
Profile pic	Indicates whether the user has a profile picture.
Nums/length username	Ratio of numerical characters to the total length of the username.
Full name words	Number of word tokens in the user's full name.
Nums/length full name	Ratio of numerical characters to the total length of the full name.
Username	Indicates whether the username and full name are identical.
Description length	Length of the biography in characters.
External URL	Indicates whether the user’s profile includes an external URL.
Private	Specifies whether the user’s profile is set to private.
Posts	Total number of posts made by the user.
Followers	Total number of followers the user has.
Follows	Total number of accounts the user follows.
Username	Ratio of numerical characters to the total length of the username.
Full name tokens	number of word tokens present in the full name.
Full name num ratio	Ratio of numerical characters to the total length of the full name.
Full name match	Indicates whether the username and full name are exactly the same.

3.2 Dataset preprocessing

Data preprocessing plays an essential role in data analysis. It involves converting raw data, which is frequently noisy and inconsistent, into a format that is ready for analysis [16]. This process is particularly crucial for machine learning, as algorithms rely on clean and well-structured data.

Table 5. Dataset preprocessing steps

Data Preparation	Step	Description
Textual Cleanup	Eliminate unnecessary elements.	Remove URLs, user mentions, hashtags, and special symbols.
	Text replacement.	Substitute emoticons and emojis with their textual descriptions.
	Language standardization.	Convert abbreviations and slang into their complete forms.
	Error correction.	Detect and fix spelling errors.
	Expand contractions.	Change contractions.
	Character normalization.	Adjust elongated characters.
Text Normalization	Punctuation removal.	Delete all punctuation marks.
	Lowercasing.	Convert the entire text to lowercase.
	Word tokenization.	Split text into individual words.
	Numeric data removal.	Exclude numbers from the text.
	Stop word removal.	Eliminate common words that lack significant meaning.
	Lemmatization.	Reduce words to their root forms.

By tackling issues related to data quality, preprocessing ensures that statistical modeling and algorithm implementation can proceed effectively [17]. the preprocessing steps applied to each dataset, as detailed in Table 5.

Table 5 outlines the data preparation process, divided into two primary categories: Textual Cleanup and Text Normalization.

•Textual Cleanup focuses on refining the text by removing unnecessary elements such as URLs, user mentions, hashtags, and special symbols. It involves replacing emoticons and emojis with their textual counterparts, standardizing language by converting abbreviations and slang, correcting spelling errors, expanding contractions, and normalizing elongated characters.

•Text Normalization involves standardizing the text further by removing punctuation, converting text to lowercase, tokenizing words into individual components, excluding numeric data, removing stop words that do not add significant meaning, and lemmatizing words to their root forms. These steps ensure that the data is clean, consistent, and ready for analysis.

3.3 Selected bio-inspired algorithm

In this study, the Grey Wolf Optimization Algorithm (GWO) has been selected for its nature-inspired approach. Mimicking the hunting strategies of grey wolves, GWO is designed to solve complex problems by simulating a hierarchical system. GWO was selected due to its hierarchical structure, which outperforms Particle Swarm Optimization (PSO) and Fire Hawk Optimizer (FHO) algorithms in terms of convergence speed, particularly when dealing with high-dimensional feature spaces typical of social media data. This structure fosters effective cooperation and maintains a balance between exploration and exploitation, enhancing its capability to find optimal solutions.

3.3.1 Origin and inspiration

GWO finds its roots in the social behavior and hunting tactics of grey wolves. The hierarchical structure and cooperative behaviors of wolf packs were computationally modeled through a bio-inspired method [4]. The algorithm is modeled after the natural dynamics of wolves working collectively to hunt prey, leveraging the principles of leadership and teamwork inherent in their social structure to address complex optimization challenges.

3.3.2 Operating mechanism

GWO mimics the hierarchical organization of a wolf pack. It categorizes wolves into three levels: alpha, beta, and omega. Alpha wolves are the leaders, beta wolves act as subordinates, and omega wolves hold the lowest rank.

Algorithm 1. GWO pseudo-code

```
Initialize population
Evaluate fitness of each wolf
Identify alpha, beta, and delta wolves
While not termination condition:
    Update coefficients A and C
    Update wolf positions
    Evaluate fitness of updated population
    Identify new alpha, beta, and delta wolves
Return best solution (alpha)
```

As shown in Algorithm 1, search agents, representing these wolves, start with random initial positions. Through iterative updates, the positions of these agents are adjusted based on the positions of the alpha, beta, and omega wolves [5]. The classification of wolves is based on their fitness values: the wolf with the highest fitness is designated as the alpha, followed by the beta and omega wolves. The search agents move according to the guidance of these wolves, progressively converging towards the optimal solution.

3.3.3 Transition from nature to artificial

The adaptation of GWO from its natural origins to artificial applications is outlined in Table 6. It provides a summary of how GWO has been adapted to address the issue of fake profiles detection.

Table 6. Natural vs. artificial aspects of GWO

Feature	Natural	Artificial
Hunting	Grey wolves pursuing prey in their natural habitat.	Users are categorized as "Real" or "Fake".
Behavior	Wolves collaborate in hunting to find optimal solutions.	Binary classification: distinguishing "Real" from "Fake" profiles.
Environment	Wilderness.	Online social networks.
Individual	A single grey wolf.	An individual user of a social network.
Population	A pack of grey wolves.	A group of social network users.
Best solution	The wolf in the group that successfully captures the prey.	The "Alpha" user represents the best solution found.
Distance	Space between the wolf and its prey.	Calculated as $D = \|C * pos - wolves[i]\|$ determining proximity to the solution.

Table 6 highlights the parallels between GWO for fake profiles detection and the natural behavior of grey wolves. In this context, users are categorized as either "Real" or "Fake," much like grey wolves' pursuit of prey. This classification mirrors how wolves cooperate in hunting to reach the best outcomes. In the artificial environment, it is online OSNs, while in nature, wolves hunt in the wilderness. Individual social network users are analogous to single wolves, and groups of users resemble wolf packs. The "alpha" wolf, which represents the best solution in the optimization process, parallels the leading wolf that successfully captures prey in the wild.

The distance metric in GWO measures the proximity to the optimal solution, similar to how wolves gauge their distance from prey while hunting.

3.4 Chosen fitness function

To address the challenge of fake profiles detection, GWO implements a Fitness function for feature selection using Logistic Regression (LR) as shown in Algorithm 2. LR was chosen as the fitness function due to its robustness, interpretability, computational efficiency, and strong suitability for binary classification tasks, particularly in distinguishing between real and fake profiles.

Algorithm 2 outlines the Fitness function utilized by GWO. It is designed to assess the effectiveness of a solution within the bio-inspired algorithm, specifically for feature selection and classification tasks. This evaluation process helps determine the quality of the selected features and their impact on classification accuracy:

Algorithm 2. Fitness function pseudo-code

```
Input: X: Feature set, y: Labels, Feature_Mask: Binary mask for selecting features
Select the features from X using Feature_Mask
Initialize a logistic regression model
Train the model using Selected_Features and y
Predict the labels using the trained model
Calculate the accuracy score: Fitness_Value = Accuracy_Score(y, Predicted_Labels)
Return Accuracy score as Fitness_Value.
```

Classifier initialization: The algorithm initializes a LR model. The latter is a widely used classification algorithm that models the probability of a binary outcome (e.g., fake or real profile). LR is the ideal choice for this task because it can handle binary classification problems and provides probabilistic outputs [18].

Feature selection: It is a crucial step in ML, especially when dealing with high-dimensional data [19]. By selecting the most relevant features, the algorithm can improve model performance, reduce computational complexity. Algorithm 2 shows that LR employs a feature selection mechanism using a binary mask. This mask determines which features from the original feature set (X) will be included in the model training. The selected features are then used to train the LR model.

Fitness evaluation: The Accuracy score between the predicted labels and the actual labels is calculated. This Accuracy score serves as the fitness value for the algorithm.

LR and bio-inspired algorithms are increasingly recognized as robust options for detecting fake profiles on OSNs due to their complementary strengths:

•Enhanced performance: Combining LR with bio-inspired algorithms leverages the strengths of both approaches. LR offers binary classification, interpretability, and efficiency, while bio-inspired algorithms provide optimization, and adaptability. This synergistic combination leads to superior performance in detecting fake profiles.

•Optimized feature selection: Bio-inspired algorithms effectively identify the most relevant features for classification, increasing both the accuracy and efficiency of the overall model.

•Parameter optimization: These algorithms also facilitate tuning of LR parameters, further boosting the model's overall performance.

•Addressing complex relationships: Bio-inspired algorithms can capture intricate relationships between features that are challenging to model with traditional techniques.

•In the context of fake profiles detection, GWO is instrumental in selecting the most informative features that effectively distinguish between real and fake profiles. Once these optimal features are identified, LR serves as a powerful classifier, leveraging the refined feature set to accurately and efficiently distinguish between the two categories.

3.5 Chosen machine learning algorithms

In fake profile detection, various ML algorithms are selected for their ability to complement each other by identifying patterns and anomalies effectively.

Decision Tree (ID3): This algorithm constructs interpretable models by establishing distinct decision boundaries, making it valuable for analyzing how specific profile features (such as activity levels or group memberships) correlate with a profile being classified as real or fake [20].

Support Vector Machine (SVM): Known for its ability to identify nuanced patterns in high-dimensional social datasets, such as behavioral metrics, SVM is highly effective for binary classification tasks, making it particularly suited for distinguishing between "real" and "fake" profiles [21].

Naive Bayes (NB): Its simple and straightforward approach and computational efficiency make it well-suited for handling large datasets, particularly in instances where specific features, such as profile descriptions or posting behaviors, strongly indicate fake accounts [22].

Random Forest (RF): Its ensemble structure enables it to model complex interactions between various features, making it highly effective in detecting a wide range of fake profiles by analyzing diverse behavioral patterns [23].

K-Nearest Neighbor (K-NN): This algorithm classifies profiles by assessing their similarity to existing real or fake profiles, making it efficient at identifying outliers and unusual accounts by measuring their proximity to clusters of typical user behavior [24].

K-Means: Although it is mainly a clustering technique, it helps group users based on behavioral traits, facilitating the unsupervised detection of fake profiles that display patterns distinctly different from those of real users [25].

These algorithms offer a diverse set of approaches to deal with the fake profiles detection problem, each leveraging different aspects of user data and behavior patterns to improve overall model's detection.

3.6 Implementation parameters

This section presents the different implementation parameters defined for GWO (Table 7) and the chosen ML algorithms (Table 8).

Table 7. GWO implementation parameters

Parameter	Description	Value
Population size	The total number of wolves (solutions) in the population.	50
Maximum iterations	The maximum number of iterations the algorithm will run.	50
Alpha (α)	Controls the influence of the alpha wolf in guiding the search.	0.5
a	Determines how many top wolves (alpha, beta, delta) are updated per iteration.	2
Coefficient C	Coefficient vector used to calculate the distance between wolves and prey.	Random [0, 2]
Coefficient A	Controls the exploration and exploitation balance by adjusting how wolves position themselves.	Random [-a, a]
Convergence criteria	The condition for terminating the optimization process (e.g., iteration count).	50
Prey position (Best solution)	Represents the optimal solution found so far, corresponding to the prey's position.	Dynamic
Exploration rate	The rate at which wolves explore new areas of the search space.	0.6
Exploitation rate	The rate at which wolves refine their search around the best solution.	0.7

Table 8. ML implementation parameters

Algorithm	Parameter	Default Value
ID3	Maximum depth	None (Unlimited)
	Minimum samples split	2
	Criterion	"entropy"
SVM	Kernel	"rbf"
	Regularization (C)	1.0
	Gamma	"scale"
NB	Prior probabilities	None
	Var smoothing	1,00E-09
RF	Number of trees	100
	Maximum depth	Unlimited
	Criterion	"gini"
K-NN	Number of neighbors	Different values
	Weights	"uniform"
	Algorithm	"auto"
K-Means	Number of clusters	2
	Initialization	random
	Max iterations	300

4. RESULTS AND DISCUSSION

This section presents the experimental results and offers a detailed discussion on the performance of the proposed bio-inspired algorithm in detection fake profiles. The results are evaluated based on metrics such as Accuracy, Precision and convergence, and efficiency also. Additionally, comparison and analysis of key parameters affecting the exploration/exploitation balance of GWO is presented.

Table 9 presents the confusion matrix for the different datasets used in the experiment. It provides valuable insight into the effectiveness of the proposed model which coupled with GWO and its LR-based Fitness function.

•The confusion matrix presents a good classification performance of both real and fake profiles:

•1025 true positives and 200 true negatives with Facebook dataset.

•464 true positives and 498 true negatives with Twitter dataset.

•326 true positives and 328 true negatives with Instagram dataset.

Table 9. Confusion matrix with GWO

OSN	Actual Label	Predicted Real	Predicted Fake
Facebook	Actual Real	1025	1
	Actual Fake	18	200
Twitter	Actual Real	464	3
	Actual Fake	35	498
Instagram	Actual Real	326	22
	Actual Fake	20	328

This implies that only 18, 35, and 20 profiles are detected as false negatives for Facebook, Twitter, and Instagram, respectively. It demonstrates balanced fake profiles detection. These results highlight the model's Precision at 0.97, 0.93, and 0.94 for Facebook, Twitter, and Instagram, respectively (Table 10), meaning it is highly effective at minimizing false positives, with the best results achieved using the Facebook dataset. With Recall values of 0.96, 0.93, and 0.94 (for Facebook, Twitter, and Instagram, respectively), the model also performs well in detecting a significant proportion of fake profiles.

As shown in Table 9, GWO played a key role in enhancing the model's performance. By mimicking the hierarchical hunting strategies of grey wolves, which is essential in optimization tasks like feature selection for fake profiles detection. Thereafter, through the iterative process of position updates guided by alpha, beta, and omega wolves, GWO effectively identifies the most relevant features. This ensures that the model focuses on the most distinguishing characteristics of real and fake profiles, contributing to its high classification Accuracy of 0.97, 0.93, and 0.94 for Facebook, Twitter, and Instagram, respectively (Table 10). Logistic Regression (LR) serves as the fitness function within GWO algorithm to assess how well a set of selected features can classify profiles correctly. This is done by evaluating the quality of feature subsets during the optimization process. LR is particularly well-suited for binary classification tasks, such as distinguishing between real and fake profiles. It is also important to point out that LR's natural ability to handle regularization helps control overfitting, as demonstrated across various datasets with different features and sizes.

Table 10. Results overview

Dataset	Ref.	Metric	NB	DT	SVM	RF	K-NN	K-Means	GWO
Facebook	[14]	Accuracy	-	0.97	0.95	-	0.91	-	-
		Precision	-	0.98	0.97	-	0.95	-	-
		Recall	-	0.98	0.96	-	0.93	-	-
		F1-score	-	0.93	0.91	-	0.81	-	-
	Our model	Accuracy	0.96	0.96	0.95	0.95	0.96	0.82	**0.97**
		Precision	0.91	0.92	0.91	0.92	0.93	0.42	**0.97**
		Recall	0.95	0.95	0.94	0.93	0.93	0.48	**0.96**
		F1-score	0.93	0.93	0.963	0.94	0.94	0.45	**0.95**
Twitter	[15]	Accuracy	0.91	-	-	-	-	-	-
		Precision	0.88	-	-	-	-	-	-
		Recall	0.93	-	-	-	-	-	-
		F1-score	0.91	-	-	-	-	-	-
	Our model	Accuracy	0.84	0.88	0.84	0.9	0.87	0.75	**0.93**
		Precision	0.75	0.89	0.85	0.89	0.88	0.5	**0.93**
		Recall	0.76	0.88	0.84	0.9	0.87	0.55	**0.93**
		F1-score	0.76	0.87	0.84	0.9	0.87	0.52	**0.93**
Instagram	Dataset	Accuracy	0.91	-	0.92	-	-	-	-
		Precision	0.87	-	0.93	-	-	-	-
		Recall	0.95	-	0.91	-	-	-	-
		F1-score	0.91	-	0.92	-	-	-	-
	Our model	Accuracy	0.89	0.88	0.89	0.92	0.89	0.73	**0.94**
		Precision	0.6	0.88	0.9	0.92	0.89	0.48	**0.94**
		Recall	0.6	0.87	0.89	0.93	0.89	0.53	**0.94**
		F1-score	0.6	0.87	0.89	0.92	0.89	0.5	**0.93**

Note: Dataset https://www.kaggle.com/free4ever1/instagram-fake-spammer-genuineaccounts

Table 10 compares the results obtained by the proposed model with those from studies providing datasets from OSNs such as Facebook, Twitter, and Instagram. The performance metrics evaluated include Accuracy, Precision, Recall, and F1-score. The results can be interpreted in terms of GWO's performance, including its use of LR as a fitness function, its exploration and exploitation capabilities, and its convergence behavior across the diverse solution spaces presented by the datasets.

For the Facebook dataset, the combination of GWO and LR emerges as a highly effective solution for fake profile detection. GWO's capabilities, coupled with its rapid convergence, enable it to select the most relevant features for LR, resulting in superior performance across all metrics. In spite of that, the previously reported model [14] achieved a slightly higher scores with DT with Accuracy, Precision and Recall reaching 0.97, 0.98 and 0.98 respectively compared to 0.97, 0.96, 0.96 obtained by the proposed model. This suggests that while DT might capture more true positives, it also introduces more false positives, which GWO mitigates. However, the proposed model surpasses the previous model's performance in terms of F1-score, achieving 0.95 compared to the earlier result of 0.93. It is worth to notice that the previous reported model [14], which employed SVM and K-NN showed lower Accuracy, Precision, Recall, and F1-score compared to GWO, indicating that they are less effective for this dataset.

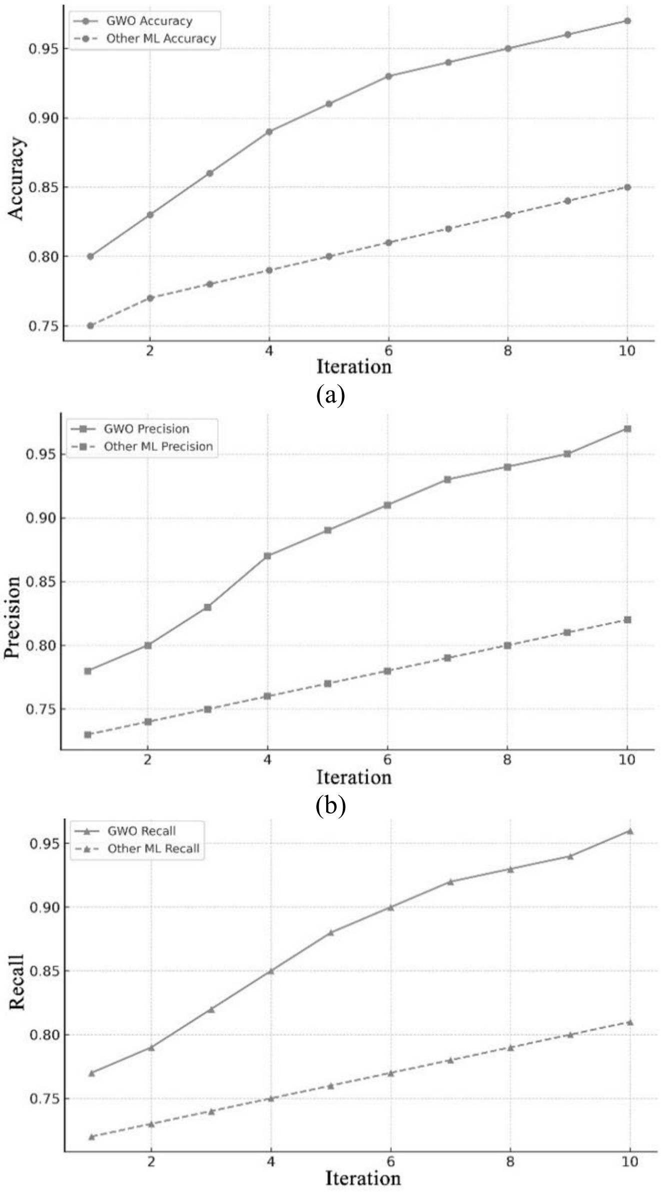
On Facebook dataset, the high Precision (0.97) and Recall (0.96) demonstrate that GWO effectively balances the search for new solutions with refining existing ones. Its exploration ability allows it to search through a large solution space, identifying relevant features that may be scattered across various behavioral patterns, interaction data, and content attributes. Once promising feature subsets are found, exploitation ensures that the algorithm reduces its scope of action on refining the most optimal feature set.

On Facebook dataset, GWO exhibits rapid convergence, as showed by its Accuracy score of 0.97. This rapid convergence enables the proposed model to efficiently identify the most relevant feature subsets, leading to superior Precision and Recall, which is valuable when working with a dataset like Facebook's.

For the Twitter dataset, GWO/LR combination proves to be a very efficient as solution for the detection of fake profiles. GWO's capabilities (exploration, exploitation and rapid convergence), facilitate its identification of the most relevant features for LR, consistently outperforming in all aspects of evaluation. While the previously reported model [15] achieved an equal Recall score of 0.93 obtained by NB, the proposed model consistently outperforms it in terms of Accuracy (0.93 vs. 0.91) and Precision (0.93 vs. 0.88).

On Twitter dataset, in order to avoid overfitting, GWO\LR's capabilities ensure that selected features generalize well. GWO’s exploration is particularly important, given the rapid pace and diversity of user activity. The model excels at analyzing dynamic data (interactions, mentions and retweets) to find important features. Then, GWO's exploitation is used to refine the feature set, as shown by the Precision and Recall scores of 0.93.

On Twitter dataset, GWO's convergence is also evident in its Accuracy score of 0.93, which surpasses other algorithms. The proposed model converges toward a solution that balances false positives and false negatives effectively, achieving the highest F1-score of 0.93. It is clear that the proposed model has the ability to handle dynamic and varied data as Twitter dataset.



(a)

(b)

(c)

Figure 1. GWO's convergence performance, (a) accuracy, (b) precision, (c) recall

Figure 1 visually illustrates GWO's convergence behavior over iterations in terms of accuracy, precision, and recall, clearly outperforming other ML algorithms. The convergence results, illustrated in Figure 1, highlight the effectiveness of the GWO-based approach compared to traditional machine learning (ML) algorithms for fake profile detection on social media. In Figure 1(a), the accuracy of GWO rises steadily from 80% to 97%, clearly outperforming baseline ML models that plateau at 85%. This improvement reflects GWO's capability to guide effective feature selection, resulting in more accurate classification. Similarly, Figure 1(b) shows that GWO achieves precision values exceeding 96%, compared to ~82% for standard ML. This indicates a marked reduction in false positives, enhancing the model's reliability in distinguishing fake profiles from real ones. Figure 1(c) shows recall improving from 77% to over 95% with GWO, highlighting its strength in reducing false negatives. This is crucial in security contexts where missing fake profiles can lead to harmful consequences.

A detailed error analysis across the three datasets—Facebook, Twitter, and Instagram—revealed distinct patterns in misclassification. The majority of false negatives occurred within the Instagram dataset, largely due to sparse and inconsistent profile attributes, which limit the model's ability to extract distinguishing features. Unlike Facebook and Twitter, where users tend to share more structured and informative data (e.g., friend counts, bio text, post activity), Instagram profiles often lack depth, making fake profiles harder to differentiate from real ones. In the Twitter dataset, most false positives were associated with verified or bot-like real accounts that exhibited abnormal behavior, such as excessive posting or retweeting, mimicking fake profiles. For Facebook, errors were more balanced between false positives and false negatives, often due to duplicated or minimalistic user information, which blurred the decision boundary. These findings suggest the need for platform-specific feature engineering and the possible integration of behavioral or temporal features to further improve model accuracy, especially in data-sparse environments like Instagram.

Also, Table 10's results can be interpreted in terms of datasets specifications. Each dataset provides distinct sets of features that highlight the behavioral patterns and profile characteristics of users. These features, along with the distribution of real and fake profiles, directly influence the model's ability to detect fake profiles.

Size and distribution: Facebook dataset has a much larger proportion of real profiles compared to fake ones (16%), which could make detecting fakes more challenging. When, Twitter and Instagram datasets are perfectly balanced, which allows for better performance evaluation in fake profile detection as the algorithm has equal exposure to real and fake profiles.

Features set: Facebook dataset focuses on profile information and social activity, which gives a lot of data related to social behaviors but lacks some of the detailed engagement metrics (hashtags and mentions). Whereas, Twitter highlights account behavior, including follower-following ratios and tweet activity, which provides rich behavioral data that could help for detecting fake profiles. While, Instagram focuses more on profile creation characteristics, such as username, bio length, and followers-to-follows ratio, as well as private settings and presence of an external URL. This could make it particularly suitable for detecting fake profiles, such as those created automatically.

Data imbalance: Facebook dataset presents an imbalanced dataset where the real profiles significantly outnumber the fake ones (5:1 ratio), which can introduce biases in training models. Twitter and Instagram datasets, being balanced datasets, do not suffer from this issue, allowing algorithms to train on an equal representation of real and fake profiles.

Complexity features: Facebook's feature set is broader in terms of social behavior (groups, friends, education, check-ins), making it a more complex dataset for models that need to understand diverse aspects of user interaction and profile details. In turn, Twitter dataset adds complexity with its focus on engagement metrics, such as hashtags, mentions, and tweet activity (behavior-based features), which are essential to understanding how users engage with the platform. Yet, Instagram dataset is a mix of both profile structure and activity. Its emphasis on profile features (e.g., numerical characters in usernames) could make it easier to detect fake accounts, but also lacks some of Twitter's engagement data.

Despite the promising results achieved by the proposed model, certain shortcomings persist. Therefore, it is imperative to propose solutions to address these issues.

Additional evaluation metrics: While Table 10 focuses on common metrics like Accuracy, Precision, Recall, and F1-score, other evaluation metrics [26], such as Area Under the ROC Curve (AUC), Matthews Correlation Coefficient (MCC), or Specificity, could provide deeper insight into model performance, especially in imbalanced datasets.

Incorporating more advanced feature engineering: The datasets (especially Facebook) include various profile-related features, but additional behavioral features and text analysis (e.g., sentiment analysis of posts or descriptions [27]) could help improve the algorithms' ability to differentiate between real and fake profiles.

Balancing false positives and negatives: While some algorithms achieve high Recall, they suffer from false positives (low Precision). Cost-sensitive learning techniques [28] could help minimize false positives by adjusting the classification threshold or weighting false positives more heavily in the loss function.

Addressing imbalance between precision and recall: Several models, especially NB on Instagram dataset, show high Recall but suffer from lower Precision. This indicates that while true positives are identified well, false positives remain problematic. A solution to overcome this situation is to go for hybrid models.

Algorithm-specific tuning: Certain algorithms, like K-means, consistently underperform and may require specialized tuning or adjustments for binary classification tasks. Applying feature selection techniques or exploring enhanced clustering methods (e.g., SK-NSGAII [29]) could help improve K-means' performance. The exploration of additional metaheuristics, not only bio-inspired optimization algorithms [30], but also those inspired by different approaches [31], aims to enhance performance.

It is important to acknowledge that the iterative feature selection process in GWO increases computational runtime compared to traditional ML methods. However, this additional computational overhead is justified by the substantial improvements in classification accuracy.

5. CONCLUSION

This study presents a comprehensive analysis of fake profile detection on social media platforms using bio-inspired algorithms, specifically the Grey Wolf Optimization algorithm, in conjunction with ML algorithms. The proposed model, which integrates GWO with Logistic Regression as a Fitness function for feature selection matter, demonstrates higher performance across datasets from Facebook, Twitter, and Instagram.

The key findings of this research are as numerous. First, the GWO-based model consistently outperforms traditional ML algorithms in terms of Accuracy, Precision, Recall, and F1-score. This highlights the potential of bio-inspired optimization techniques in enhancing feature selection and classification accuracy for fake profile detection. Second, the proposed model shows robust performance across varied datasets, effectively handling both imbalanced (Facebook) and balanced (Twitter, Instagram) datasets. Next, the study highlights the critical role of relevant feature selection in improving model performance to enhance differentiation between real and fake profiles.

While the proposed model delivers promising results, there is scope for future research directions. The exploration of additional metaheuristics, not only bio-inspired optimization algorithms, but also those inspired by different approaches, aims to enhance performance. Also, the development of real-time detection models capable of efficiently identifying fake profiles without losing accuracy or precision. Additionally, it would be interesting to conduct cross-platform analyses to develop more generalized detection models applicable across multiple social networks. Additionally, since datasets differ in characteristics (e.g., Facebook emphasizes social behavior while Instagram highlights profile details), using personalized solutions for each dataset may be more effective. Adapting algorithms or their parameters to the specific features of each platform, like interaction patterns for Twitter or multimedia content for Instagram, could enhance fake profiles detection accuracy. Lastly, it is important to address ethical considerations, including privacy concerns and the potential impact of false positives on legitimate users, especially given the current global context.

REFERENCES

[1] Montag, C., Demetrovics, Z., Elhai, J.D., Grant, D., et al. (2024). Problematic social media use in childhood and adolescence. Addictive behaviors, 153: 107980. https://doi.org/10.1016/j.addbeh.2024.107980

[2] Bokolo, B.G., Liu, Q. (2024). Artificial intelligence in social media forensics: A comprehensive survey and analysis. Electronics, 13(9): 1671. https://doi.org/10.3390/electronics13091671

[3] Brown, A., Gupta, M., Abdelsalam, M. (2024). Automated machine learning for deep learning based malware detection. Computers & Security, 137: 103582. https://doi.org/10.1016/j.cose.2023.103582

[4] Mirjalili, S., Mirjalili, S.M., Lewis, A. (2014). Grey wolf optimizer. Advances in Engineering Software, 69: 46-61. https://doi.org/10.1016/j.advengsoft.2013.12.007

[5] Varuna, W.R., Shalini, K., Roy, M.E.A. (2022). An efficient framework for fake profile identification using metaheuristic and deep learning techniques. Journal of Positive School Psychology, 6(4): 3741-3750.

[6] Mahammed, N., Bennabi, S., Fahsi, M., Klouche, B., Elouali, N., Bouhadra, C. (2022). Fake profiles identification on social networks with bio inspired algorithm. In 2022 First International Conference on Big Data, IoT, Web Intelligence and Applications (BIWA), Sidi Bel Abbes, Algeria, pp. 48-52. https://doi.org/10.1109/BIWA57631.2022.10037927

[7] Mahammed, N., Klouche, B., Saidi, I., Khaldi, M., Fahsi, M. (2023). Bio-inspired algorithms for effective social media profile authenticity verification. In 6th International Hybrid Conference on Informatics and

Applied Mathematics, Guelma, Algeria, pp. 109-119.

[8] Patil, D.R., Pattewar, T.M., Punjabi, V.D., Pardeshi, S.M. (2024). Detecting fake social media profiles using the majority voting approach. EAI Endorsed Transactions on Scalable Information Systems, 11(3): 1-18. https://doi.org/10.4108/eetsis.4264

[9] Soorya Ramdas, A.N.N. (2024). Leveraging machine learning for fraudulent social media profile detection. Cybernetics and Information Technologies, 24(1): 118-136. https://doi.org/10.2478/cait-2024-0007

[10] Mahammed, N., Klouche, B., Saidi, I., Khaldi, M., Fahsi, M. (2023). Enhancing social media profile authenticity detection: A bio-inspired algorithm approach. In International Conference on Machine Learning for Networking, Paris, France, pp. 32-49. https://doi.org/10.1007/978-3-031-59933-0_3

[11] Kumar, C., Bharati, T.S., Prakash, S. (2023). Online social networks: An efficient framework for fake profiles detection using optimizable bagged tree. In International Conference on Data & Information Sciences, Agra, India, pp. 255-264. https://doi.org/10.1007/978-981-99-6906-7_22

[12] Raja, E.V.S., Aditya, B.L., Mohanty, S.N. (2024). Fake profile detection using logistic regression and gradient descent algorithm on online social networks. EAI Endorsed Transactions on Scalable Information Systems, 11(1): 8. https://doi.org/10.4108/eetsis.4342

[13] Raghavendra, M.S., Prasad, P.P., Neha, E.S., Meghana, K., Buela, C.H. (2024). Utilizing machine learning techniques to eradicate fake profiles. Bulletin for Technology and History Journal, 24(4): 324-332. https://doi.org/10.37326/bthnlv22.8/1618

[14] Albayati, M., Altamimi, A. (2019). MDFP: A machine learning model for detecting fake Facebook profiles using supervised and unsupervised mining techniques. International Journal of Simulation: Systems, Science & Technology, 20(1): 1-10. https://doi.org/10.5013/IJSSST.a.20.01.11

[15] Erşahin, B., Aktaş, Ö., Kılınç, D., Akyol, C. (2017). Twitter fake account detection. In 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, Turkey, pp. 388-392. https://doi.org/10.1109/UBMK.2017.8093420

[16] Palakurti, N.R., Kanchepu, N. (2024). Machine learning mastery: Practical insights for data processing. In Practical Applications of Data Processing, Algorithms, and Modeling, pp. 16-29. https://doi.org/10.4018/979-8-3693-2909-2.ch002

[17] Naseem, U., Razzak, I., Eklund, P.W. (2021). A survey of pre-processing techniques to improve short-text quality: A case study on hate speech detection on twitter. Multimedia Tools and Applications, 80: 35239-35266. https://doi.org/10.1007/s11042-020-10082-6

[18] Das, A. (2024). Logistic regression. In Encyclopedia of Quality of Life and Well-Being Research, pp. 3985-3986. https://doi.org/10.1007/978-3-031-17299-1_1689

[19] Theng, D., Bhoyar, K.K. (2024). Feature selection techniques for machine learning: A survey of more than two decades of research. Knowledge and Information Systems, 66(3): 1575-1637. https://doi.org/10.1007/s10115-023-02010-5

[20] Marudi, M., Ben-Gal, I., Singer, G. (2024). A decision tree-based method for ordinal classification problems. IISE Transactions, 56(9): 960-974. https://doi.org/10.1080/24725854.2022.2081745

[21] Kavitha, S.S., Kaulgud, N. (2024). Quantum machine learning for support vector machine classification. Evolutionary Intelligence, 17(2): 819-828. https://doi.org/10.1007/s12065-022-00756-5

[22] Veziroğlu, M., Veziroğlu, E., Bucak, İ.Ö. (2024). Performance comparison between Naive Bayes and machine learning algorithms for news classification. In Bayesian Inference-Recent Trends. https://doi.org/10.5772/intechopen.1002778

[23] Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M., Liang, X. (2024). An improved random forest based on the classification accuracy and correlation measurement of decision trees. Expert Systems with Applications, 237: 121549. https://doi.org/10.1016/j.eswa.2023.121549

[24] Wang, N., Zhao, E. (2024). A new method for feature selection based on weighted k-nearest neighborhood rough set. Expert Systems with Applications, 238: 122324. https://doi.org/10.1016/j.eswa.2023.122324

[25] Huang, Z., Zheng, H., Li, C., Che, C. (2024). Application of machine learning-based k-means clustering for financial fraud detection. Academic Journal of Science and Technology, 10(1): 33-39. https://doi.org/10.54097/74414c90

[26] Rainio, O., Teuho, J., Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. Scientific Reports, 14(1): 6086. https://doi.org/10.1038/s41598-024-56706-x

[27] Tan, K.L., Lee, C.P., Lim, K.M. (2023). A survey of sentiment analysis: Approaches, datasets, and future research. Applied Sciences, 13(7): 4550. https://doi.org/10.3390/app13074550

[28] Araf, I., Idri, A., Chairi, I. (2024). Cost-sensitive learning for imbalanced medical data: A review. Artificial Intelligence Review, 57(4): 80. https://doi.org/10.1007/s10462-023-10652-8

[29] Mahammed, N., Bennabi, S., Bekka, A., Klouche, B., Fahsi, M., Guellil, Z. (2021). An attempt to enhance NSGA-II with a clustering approach. In 2021 International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain, pp. 1143-1149. https://doi.org/10.1109/DASA53625.2021.9682408

[30] Maarouk, C., Haouassi, H., Malik, M.M. (2024). Discrete black widow optimization algorithm for multi-objective IoT application placement in fog computing environments. Revue d'Intelligence Artificielle, 38(4): 1077-1088. https://doi.org/10.18280/ria.380403

[31] Kusuma, P.D., Novianty, A. (2024). A new metaheuristic algorithm called treble opposite algorithm and its application to solve portfolio selection. Mathematical Modelling of Engineering Problems, 11(3): 807-816. https://doi.org/10.18280/mmep.110326