A Modified Binary Arithmetic Optimization Algorithm for Feature Selection

RAJESH RANJAN, JITENDER KUMAR CHHABRA
Computer Engineering Department,
National Institute of Technology,
Kurukshetra, Haryana, 136119
INDIA

Abstract: - Feature selection chooses the optimal subset from the feature set without scarifying the information carried by the dataset. It is considered a complex combinatorial problem, so classical optimization techniques fail to solve it when the feature set becomes larger. Meta-heuristic approaches are well known to solve complex optimization problems; hence these algorithms have been successfully applied to extract optimal feature subsets. The arithmetic Optimization Algorithm is a newly proposed mathematics-based meta-heuristic search algorithm successfully applied to solve optimization problems. However, it has been observed that AOA experiences a poor exploration phase. Hence in the present work, a Modified Binary Arithmetic Optimization Algorithm (MB-AOA) is proposed, which solves the poor exploration problem of standard AOA. In the MB-AOA, instead of utilizing a single best solution, an optimal solution set that gradually shrinks after each successive iteration is applied for better exploration during initial iterations. Also, instead of a fixed search parameter (μ), the MB-AOA utilizes a variable parameter suitable for binary optimization problems. The proposed method is evaluated over seven real-life datasets from the UCI repository as a feature selection wrapper method and compared with standard AOA over two performance metrics, Average Accuracy, F-score, and the generated feature subset size. MB-AOA has performed better in six datasets regarding F-score and average accuracy. The obtained results from the simulation process demonstrate that the MB-AOA can select the relevant features, thus improving the classification task's overall accuracy levels.

Key-Words: - Feature Selection, Meta-heuristic Algorithm, Machine Learning, Wrapper

Received: July 19, 2022. Revised: June 2, 2023. Accepted: June 28, 2023. Published: July 24, 2023.

1 Introduction

Technological advancement, mainly in the digital domain, has led to an enormous volume of raw data. These raw datasets must be pre-processed to extract valuable information from them. A dataset may contain several features that may not be significant for all the tasks; some may be redundant and correlated, so a subset of features must be selected for a particular task, [1]. Feature selection chooses a feature subset from the original feature to improve the desired accuracy and authenticity of the information carried out by the dataset. It is the most significant pre-processing step in supervised and unsupervised machine learning, [2]. The researcher has proposed several methods to extract the relevant subset of features, which broadly falls into three categories, [3].

1. Filter Approach: Filtering techniques choose features using the data's inherent characteristics. According to various statistical criteria, the filter typically calculates each feature's scores before selecting the features with the highest scores.

2. Wrapper Approach: The wrapper strategy employs a learning technique to determine the importance of a specific collection of features/attributes. The wrapper strategy often yields superior results than the filter approach, although it is computationally more costly.

3. Embedded Approach: In this method, the choice of which features to use is built into the learning algorithm. The feature selection and learning algorithms are made simultaneously by the embedded process. It keeps the model from being too well-fitted but takes longer than the wrapper approach.

Several search mechanisms, such as exhaustive, random, and greedy approaches, have been proposed, but as the feature size increases, the feature selection task gradually becomes a computationally expensive, time-consuming, complex optimization task, [4]. Recently, several nature-inspired algorithms have been successfully applied to solve complex non-linear optimization tasks. Through their intrinsic property of exploration and exploitation mechanism, these meta-heuristic approaches avoid optimal local solutions and hence do not suffer from premature convergence. So, considering the complexity of the feature selection task, meta-heuristic methods are well suited to solve it while maintaining the accuracy level of the model.

Recently, several nature-inspired algorithms have been employed to solve the feature selection task either through the wrapper approach or in the hybrid form, along with filter techniques in the machine learning domain. Researchers have designed and are still working to find several new meta-heuristic methods to solve various optimization techniques, including the feature selection problem. Genetic Algorithm (GA), [5], Particle Swarm Optimization (PSO), [6], Ant Colony Optimization (ACO), [7], Crow Search Algorithm (CSA), [8], and Differential Evolution (DE), [9], are some of the approaches which have been successfully applied to feature selection tasks in various problems in their original as well as hybrid form.

The arithmetic Optimization Algorithm is a recently proposed meta-heuristic search algorithm that works on the principles of basic mathematical functions Addition, Subtraction, Multiplication, and Division, [10]. The AOA solves several real-life optimization problems from various domains, [11].

Since feature selection is considered an optimization problem so, in the present work, the AOA is modified to solve the binary feature selection problem. The explore and exploit the whole solution space, AOA only utilizes the best solution obtained; hence in specific scenarios, it fails to explore the entire search space and is thus stuck to the optimal local solution. The present works propose a Modified Binary Arithmetic Optimization Algorithm (MB-AOA) by introducing a variable search operator and a set of optimal solutions to delve into the search space. The performance of MB-AOA is demonstrated through three evaluation criteria, average accuracy, F-score, and feature subset size over seven real-life datasets, and is compared to standard AOA.

The rest of the paper is structured as follows section-2 represents a brief literature review; section-3 describes the overall methodology of standard AOA, its drawbacks, and MB-AOA and application of MB-AOA as a wrapper method for feature selection task. Section-4 discusses the experimental parameters, datasets, and the obtained results. Finally, section-5 concludes the whole work and the present work's prospect.

2 Literature Review

Feature selection has become the most prominent step in domains like bioinformatics, pattern recognition, machine learning, and various disciplines with large feature sets. Accordingly, researchers have done multiple studies in the past and still proposing different new approaches due to the emergence of the huge volume of data. In the past, several meta-heuristic techniques have been applied as a wrapper method for feature selection problems. In this section, we have studied some modified implementations of AOA approaches successfully applied to feature selection problems.

In [12], the authors, have proposed two binary variants of AOA, BAOA-V, and BAOA-S, for feature selection for high-resolution image data for tumor detection. The BAOA-V hyperbolic tangent and the BAOA-S sigmoid functions transform standard AOA into binary form for the feature selection problem. Even within BAOA-V and BAOA-S, BAOA-S performs better by selecting small and more relevant feature subsets than BAOA-V.

In another recent work, [13], hybridized AOA with Simulated Annealing (SA) and combined the hybrid approach with a filter method for feature selection in a high-dimensional cancer gene-expression dataset. The crossover concept is further applied to enhance the exploratory capability of the hybrid approach. The proposed approach is used over ten gene-expression datasets to evaluate the performance of the hybrid method.

In, [14], the authors have applied the AOA used to optimize SVM to detect and categorize the defects over the chip surfaces. Here AOA is used to determine the optimal kernel function for the SVM, which is further applied for categorizing and detecting defects over the chips.

In, [15], the authors have proposed k-NN-AOA for detecting fake news spread during the covid-19 pandemic by improving the k-NN classifier accuracy level by selecting relevant feature subsets. The proposed approach is applied to the real-life Koirala dataset. The proposed work is further compared with other similar techniques for feature selection using the k-NN classifier, and the obtained result shows that the proposed technique outperforms different approaches used for comparison.

Recently, more stress has been given to the hybrid approach for numerous optimization problems, including feature selection problems in classification and clustering. In, [16], the authors have modified Coronavirus Herd Immunity Optimizer with a greedy crossover approach and applied the algorithm as a wrapper for feature selection over 23 medical datasets using a k-NN classifier. The proposed method is compared with several filters and recently proposed wrapper approaches for the feature selection problem. In another work, in, [17], the authors enhanced the Moth Flame Optimization (MFO) algorithm in two ways. The initial step involves the generation of eight binary variants by applying eight transition functions. The LBMFO-V3 is a modified version of the MFO algorithm that includes a Lévy flight operator in conjunction with the transition functions. The study demonstrated that the LBMFO V3 technique, as proposed, exhibits superior performance compared to multiple established wrapper methods in 83% of the datasets.

Alweshah utilized a hybrid approach by combining AOA with Great Deluge Algorithm (GDA) and AOA-GD to select pertinent features in actual medical datasets. The performance of AOA has been improved by AOA-GD, resulting in significantly better performance compared to Binary Moth Flame Optimizer (MFO) and Coronavirus Herd Immunity Optimizer, [18].

The previous studies show that the AOA is a recently proposed meta-heuristic approach, so only a few works have been reported on the feature selection problem. A vast scope is available for modifying and hybridizing the standard AOA with other methods for various optimization techniques, including the feature selection task.

3 Methodology

The various steps involved in the present work are discussed in this section. Standard Arithmetic Optimization Algorithm and its drawbacks are discussed after that Modified Binary Arithmetic Optimization Algorithm is discussed, and finally, the wrapper-based feature selection using MB-AOA is discussed.

3.1 Arithmetic Optimization Algorithm

AOA is a population-based meta-heuristic approach proposed by, [10]. Arithmetic is a subfield of mathematics that deals with adding, subtracting, multiplying, and dividing numbers and their related operations. The AOA search technique consists of two stages exploration and exploitation common to other metaheuristic algorithms. Multiplication and division are utilized to update the search agents' locations during the exploration stage, whereas addition and subtraction are employed during the exploitation stage. Depending on the formulation, AOA may tackle small or big optimization problems due to its population-based, gradient-free nature. The Hierarchy of Arithmetic Operators is presented in Figure 1.

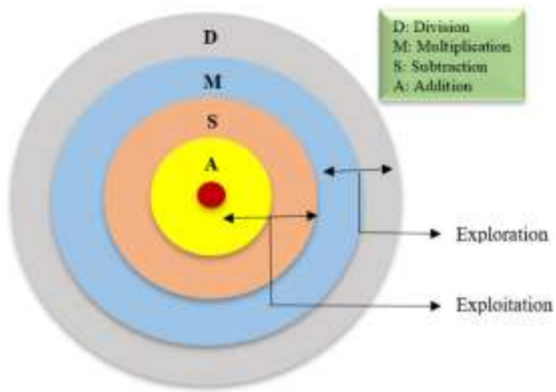

Fig. 1: Hierarchy of Arithmetic Operators

3.1.1 Working

AOA applies basic arithmetic operations to solve the optimization task. Initially, the number of candidate solutions is generated randomly. After that, with the help of Math Optimizer Accelerator (MOA) functions, the AOA decides to search the solution space for global exploration or local exploitation. MOA is mathematically defined as given in Eq. (1). Depending upon the value obtained through Eq. (1) and a random number (r1) the AOA switches between the exploration and exploitation phase.

$$MOA = MOA\_Min + CUR\_ITER \times \left(\frac{\text{MOA\_Max}-\text{MOA\_Min}}{\text{MAX\_ITER}}\right) \quad (1)$$

$$Search = \begin{cases} Exploration, & if\ r1 > MOA, \\ Exploitation, & Otherwise, \end{cases} \quad (2)$$

Mathematical computations, through the division and multiplication operator, produce highly dispersed values committed to the exploratory search process, as stated by the Arithmetic operators. Hence D and M operators are used in the exploration stage of the AOA.

$$x_{i,j}(CUR\_ITER + 1) = \begin{cases} best(x_j) \div (MOP + \varepsilon) \times \left((Ub_j - Lb_j) \times \mu + Lb_j \right) \\ best(x_j) \times (MOP) \quad \times \left((Ub_j - Lb_j) \times \mu + Lb_j \right) \end{cases} \quad (3)$$

$$MOP(CUR\_ITER) = 1 - \frac{CUR\_ITER^{\frac{1}{\alpha}}}{MAX\_ITER^{\frac{1}{\alpha}}} \quad (4)$$

Math Optimizer Probability (MOP) is a function defined mathematically as given in Eq. (4); here, α represents the exploration strategy and is taken as 5. $\mathrm{Ub_j}$ and $\mathrm{Lb_j}$ represent the upper and lower limit values of the j^{th} feature and μ is a search parameter whose value is 0.5 in the standard AOA. The $\mathrm{best}(\mathrm{x_j})$ represents the j^{th} feature value of the best particle obtained.

Within the exploitation phase, the i^{th} particle updates its position through a Subtraction (S) or Addition (A) operation, decided randomly through a random number. In contrast to other operators, however, S and A have such little dispersion that they may come quite close to the target. So, the exploitation search identifies the nearly optimum solution, which may be derived after several different attempts (iterations). Eq. (5) represents the exploitation phase of AOA. The Flowchart of AOA is presented in Figure 2.

$$x_{i,j}(CUR\_ITER + 1) = \begin{cases} best(x_j) - (MOP) \times \left((Ub_j - Lb_j) \times \mu + Lb_j \right) \\ best(x_j) + (MOP) \times \left((Ub_j - Lb_j) \times \mu + Lb_j \right) \end{cases} \quad (5)$$

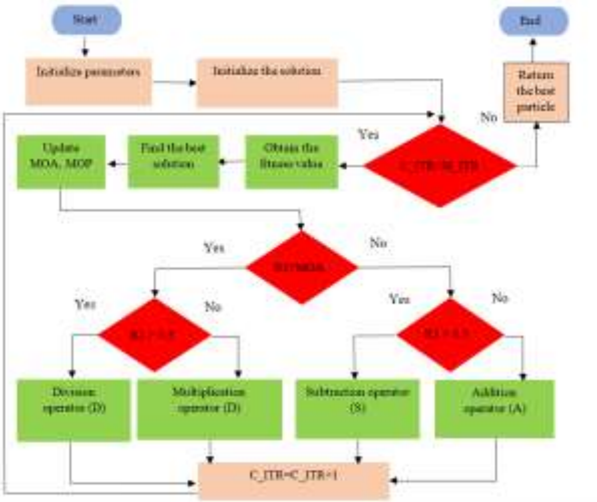

Fig .2: Flowchart of AOA

3.2 Modified Binary Arithmetic Optimization Algorithm

The exploitation and exploration stages in the AOA target only the best particle obtained. As a result, the AOA fails to fully explore the whole search space. Besides this, in the binary form of the AOA, while searching for optimal feature subsets, the upper and lower bound are 1 and 0 for all the features in the dataset. So, to overcome these shortcomings, three modifications have been applied.

3.2.1 Optimal Solution Set

In the initial phase best 15% of the total population size in terms of the fitness function is taken as the Initial Solution Size (ISS). During each successive iteration, the size of the optimal solution set is gradually decreased by applying Eq. (6), and after that, a random particle is selected from the solution set for further exploration and exploitation phase. In this way, the MB-AOA has different options to explore during initial iterations, which decrease after each successive iteration. As per the working of various similar metaheuristic approaches in the initial phase of the search, more preference is given to the exploration phase; subsequently, the search shifts from exploration to exploitation, and stress over local search is shown in the later stage. Based on the above principle, instead of following a single best solution, a set of solutions is given preference, and that set gradually shrinks in size after successive iterations.

$$Solution\_Set = ISS - CURR\_ITER \times \left(\frac{ISS}{MAX\_ITER} \right) \quad (6)$$

$$Best\_P = \text{Randomly selected value from Solution Set} \quad (7)$$

3.2.2 Variable Search Parameter (μ)

In the standard AOA, the search parameter (μ) is taken as a constant variable whose value is taken as 0.5. In the binary form of AOA, the upper and lower bound is fixed to 1 and 0, so a constant search parameter only partially explores and exploits the solution space. Two separate search parameters for the exploration and exploitation phases are defined and given in Eq. (8), which helps overcome the shortcoming discussed to a certain extent. It can be seen from the new value assigned to the search parameter (μ) that more randomness is preferred for the exploration phase and the exploitation stage, and a more structured value is preferred, which gradually increases.

$$\mu = \begin{cases} Random\ Number, & Exploration, \\ 0.5 \times exp\left(0.25 \times \frac{CUR\_ITER}{MAX\_ITER}\right), & Exploitation, \end{cases} \quad (8)$$

3.3 MB-AOA as a Wrapper for Feature Selection

After updating the AOA with the above changes, the MB-AOA is applied as a wrapper method for the feature selection problem. All the particles are randomly initialized between 0 and 1. It has been proved that for feature selection problem KNN classifier works better than other classification problems, [19], hence in the present work, the KNN algorithm is used as a classifier. The number of nearest neighbours used in this work is kept to 5. Datasets are divided into 70 % for the training phase and 30% for the testing phase. Within the training dataset, 5-fold internal cross-validation is performed.

4 Experimentation and Result

This section discusses details about the datasets, the parameters of the algorithms, and the results obtained after the simulation. The MB-AOA has been compared to standard AOA over seven real-life datasets with varying classes, instances, and features for feature selection problems using the k-NN classifier. The details of the dataset are given in Table-1. The datasets are taken from the UCI repository, [20]. In the simulation number of particles is kept to 30, and the total number of iterations is set to 50; for both AOA and MB-AOA, the MOA_Max and MOA_Min are taken as 0.9 and 0.1, respectively. The exploration strategy (α) is taken as 5 in both AOA and MB-AOA, whereas the variable search parameter (μ) is taken as 0.5 for AOA and MB-AOA; it is given in Eq. (7).

Table 1. Dataset Details

Dataset	Features	Instances	Classes
Cleveland (D1)	13	297	5
Dermatology (D2)	34	366	6
ParkinsonC (D3)	753	755	2
Sonar (D4)	60	208	2
SpectefHeart (D5)	43	266	2
Vehicle (D6)	18	846	4
WDBC (D7)	30	569	2

In the present work, two performance metrics, Accuracy, and F-score, are used to evaluate the performance of both binarized forms of standard AOA and MB-AOA. Accuracy is the ratio of correctly identified data instances to their respective class label to the total number of data instances used in testing the classifier. Mathematically, it is given in Eq. (9)

$$Accuracy = \frac{TP + TN}{TP + Tn + FP + FN} \quad (9)$$

F-score is the harmonic mean of the Precision and Recall measure. Here Precision is defined as the ratio of actual relevant (True Positive) data instances to all the data instances identified as positive (True Positive+ False Positive) by the classifier.

The recall is defined as the ratio of actual relevant data instances (True Positive) out of total relevant data instances (True Positive+ False Negative) identified by the classifier. Thus F-score balances the Precision and recall performance metrics. Mathematically, it is given in Eq. (10)

$$F - score = \frac{TP}{TP + \frac{(FP+FN)}{2}} \quad (10)$$

Table 2 represents the Average accuracy, feature size, and F-score of the AOA and MB-AOA over seven datasets over 20 independent runs. From the result, it can be seen that out of seven datasets, MB-AOA has obtained better results in 6 datasets. Due to an almost similar approach Feature subset obtained has a similar size for both MB-AOA and AOA. F-score balances both precision and recall performance metrics defined above, especially in the case of multi-class classification hence representing a better way to express the obtained results. Thus, a higher value of the F-score represents better results in terms of classification. It can be seen from the obtained results that out of seven datasets, MB-AOA has performed better in 6 datasets. In the case of the vehicle dataset, the AOA has performed slightly better than MB-AOA.

Table 2. Accuracy, Feature Size, and F-score

	Accuracy		Feature Size		F-score	
Dataset	**MB-AOA**	**AOA**	**MB-AOA**	**AOA**	**MB-AOA**	**AOA**
D1	**57.88**	56.13	3.66	**3.42**	**0.52**	0.50
D2	**97.31**	96.71	22.95	**22.14**	**0.97**	0.96
D3	**85.84**	85.19	**163.19**	163.81	**0.85**	0.84
D4	**81.40**	79.13	**22.76**	23.52	**0.81**	0.79
D5	**76.66**	75.13	**17.57**	19.47	**0.75**	0.74
D6	72.42	**72.55**	11.00	**8.52**	0.71	**0.71**
D7	**97.04**	96.71	17.95	**13.90**	**0.97**	0.96

MB-AOA (A1) is further compared with two similar metaheuristic approaches, CHIO-GC (A2), [16] and LBMFO-V3 (A3), [17] applied as a wrapper for the feature selection problem.

Table 3. Accuracy and Feature Size

Accuracy	Feature Size

Dataset	A1	A2	A3	A1	A2	A3
D1	57.88	**59.66**	53.33	**3.66**	6.68	6.80
D2	**97.31**	80.06	84.42	22.95	18.49	18.35
D3	**85.84**	84.00	81.90	163.19	365.83	369.10
D4	81.40	*N.A.*	*N.A.*	22.76	*N.A.*	*N.A.*
D5	**76.66**	73.03	70.13	17.57	21.00	20.45
D6	72.42	*N.A.*	*N.A.*	11.00	*N.A.*	*N.A.*
D7	**97.04**	90.33	91.00	17.95	13.37	13.99

The comparison is made over two performance metrics, average accuracy, and the obtained feature subset size. The details of obtained results are given in Table 3.

5 Conclusion

The newly introduced arithmetic optimization approach has been refined to feature selection problems in the supervised machine learning approach. The AOA is a recently proposed algorithm with several scopes for further improvement according to the problem to be solved. The present work has introduced two significant changes to the original AOA: a better exploration opportunity and a variable search parameter to solve the feature selection task. MB-AOA is tested over seven significant real-life datasets, and the result obtained is compared with standard AOA over three performance metrics: average accuracy level, F-score, and accepted feature subset size. The MB-AOA has produced better results when compared with the standard AOA in terms of F-score and mean accuracy level. MB-AOA can be combined with similar algorithms to create a hybrid approach that can produce more robust and sustainable results. Further, introducing specific changes can apply MB-AOA to more complex continuous optimization problems. Besides this, MB-AOA can be extended to a multi-objective method for optimizing more than one problem.

References:

[1] Tang, J., Alelyani, S., & Liu, H. (2014). Feature selection for classification: A review. Data classification: Algorithms and applications, 37

[2] Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. Artificial intelligence, 97(1-2), 273-324.

[3] Dash, M., & Liu, H. (1997). Feature selection for classification. Intelligent data analysis, 1(1-4), 131-156.

[4] Liu, H., & Motoda, H. (2012). Feature selection for knowledge discovery and data mining (Vol. 454). Springer Science & Business Media.

[5] Leardi, R., Boggia, R., & Terrile, M. (1992). Genetic algorithms as a strategy for feature selection. Journal of chemometrics, 6(5), 267-281.

[6] Chuang, L. Y., Chang, H. W., Tu, C. J., & Yang, C. H. (2008). Improved binary PSO for feature selection using gene expression data. Computational Biology and Chemistry, 32(1), 29-38.

[7] Kashef, S., & Nezamabadi-pour, H. (2015). An advanced ACO algorithm for feature subset selection. Neurocomputing, 147, 271-279.

[8] Ouadfel, S., & Abd Elaziz, M. (2020). Enhanced crow search algorithm for feature selection. Expert Systems with Applications, 159, 113572.

[9] Hancer, E. (2019). Differential evolution for feature selection: a fuzzy wrapper–filter approach. Soft Computing, 23, 5233-5248.

[10] Abualigah, L., Diabat, A., Mirjalili, S., Abd Elaziz, M., & Gandomi, A. H. (2021). The arithmetic optimization algorithm. Computer methods in applied mechanics and engineering, 376, 113609.

[11] Kaveh, A., & Hamedani, K. B. (2022, January). Improved arithmetic optimization algorithm and its application to discrete structural optimization. In Structures (Vol. 35, pp. 748-764). Elsevier.

[12] Bansal, P., Gehlot, K., Singhal, A., & Gupta, A. (2022). Automatic detection of osteosarcoma based on integrated features and feature selection using binary arithmetic optimization algorithm. Multimedia Tools and Applications, 81(6), 8807-8834.

[13] Pashaei, E., & Pashaei, E. (2022). Hybrid binary arithmetic optimization algorithm with simulated annealing for feature selection in high-dimensional biomedical data. The Journal of Supercomputing, 78(13), 15598-15637.

[14] Chen, K., Yao, H., & Han, Z. (2022, November). Arithmetic optimization algorithm to optimize support vector machine for chip defect Identification. In 2022 28th International Conference on Mechatronics and Machine Vision in Practice (M2VIP) (pp. 1-5). IEEE.

[15] Zivkovic, M., Stoean, C., Petrovic, A., Bacanin, N., Strumberger, I., & Zivkovic, T. (2021, December). A novel method for covid-19 pandemic information fake news detection

based on the arithmetic optimization algorithm. In 2021 23rd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC) (pp. 259-266). IEEE.

[16] Alweshah, M., Alkhalaileh, S., Al-Betar, M. A., & Bakar, A. A. (2022). Coronavirus herd immunity optimizer with greedy crossover for feature selection in medical diagnosis. Knowledge-Based Systems, 235, 107629.

[17] Abu Khurmaa, R., Aljarah, I., & Sharieh, A. (2021). An intelligent feature selection approach based on moth flame optimization for medical diagnosis. Neural Computing and Applications, 33, 7165-7204.

[18] Alweshah, M. (2022). Hybridization of arithmetic optimization with great deluge algorithms for feature selection problems in medical diagnosis. Jordanian Journal of Computers and Information Technology, 8(2).

[19] Maleki, N., Zeinali, Y., & Niaki, S. T. A. (2021). A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection. Expert Systems with Applications, 164, 113981.

[20] Asuncion, A., & Newman, D. (2007). UCI machine learning repository.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

Rajesh Ranjan: Conceptualization, Methodology, Investigation, Software, Writing - original draft.

Dr. Jitender Kumar Chhabra: Visualization, Formal analysis, Validation, Supervision, Writing - review & editing.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

Self-Funding

Conflict of Interest

The authors have no conflict of interest to declare.