TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 20, No. 5, October 2022, pp. 1046~1054
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v20i5.23515

Hybrid features selection method using random forest and meerkat clan algorithm

Hayder Adnan Saleh[1], Rana Amer Sattar[1], Enas Mohammed Hussein Saeed[1], Dalael Saad Abdul-Zahra[2]
[1]Department of Computer Science, College of Education, Al-Mustansiriya University, Baghdad, Iraq
[2]Department of Medical Physics, Hilla University College, Babylon, Iraq

Article Info

Article history:

Received Mar 05, 2022
Revised Jul 21, 2022
Accepted Jul 29, 2022

Keywords:

Feature selection
Machine learning
Meerkat clan
Random forest

ABSTRACT

In the majority of gene expression investigations, selecting relevant genes for sample classification is considered a frequent challenge, with researchers attempting to discover the minimum feasible number of genes while yet achieving excellent predictive performance. Various gene selection methods employ univariate (gene-by-gene) gene relevance rankings as well as arbitrary thresholds for selecting the number of genes, are only applicable to 2-class problems and use gene selection ranking criteria unrelated to the algorithm of classification. A modified random forest (MRF) algorithm depending on the meerkat clan algorithm (MCA) is provided in this work. It is one of the swarm intelligence algorithms and one of the most significant machine learning approaches in the decision tree. MCA is used to choose characteristics for the RF algorithm. In information systems, databases, and other applications, feature selection imputation is critical. The proposed algorithm was applied to three different databases, where the experimental results for accuracy and time proved the superiority of the proposed algorithm over the original algorithm.

Corresponding Author:

Hayder Adnan Saleh
Department of Computer Science, College of Education, Al-Mustansiriya University
Baghdad Directorate of Education, Rusafa III, Ministry of Education, Baghdad, Iraq
Email: haider8883@uomustansiriyah.edu.iq

1. INTRODUCTION

Machine learning is an important method of artificial intelligence. Machine adaptation seeks to indicate the data structure as well as the suitability of that information in a way that people can use and understand [1]. Machine learning takes into account computer calculations, preparation of information, and the use of statistical analysis, taking into account the ultimate goal of a particular field. The main aim of machine learning (ML) is learning from data related to a particular function in maximizing performance [2], [3]. With large amounts of data available, there are many real reasons why smart data analysis is more prevalent in technological progress. It allows calculating and understanding data-driven decisions, rather than programming them to perform a task. And adjust the actions taken accordingly. After adequate training, the system may be able to provide targets for any new inputs. Machine learning contains many techniques that are used for data analysis such as decision trees and random forest (RF) [4], [5]. Creating a predictor with several versions to create a predictor group. When the classifier or predictor gives equal voting when the class predicts, the majority votes represent the predicted class [6].

Trees might be built using a variety of partitioning techniques that rely more on data rather than building trees. As the number of trees grows, so does the precision of all partitioning mechanisms [7], don't most attributes can be used to construct trees, decide what attribute is most relevant and apply to the class or

provide a great deal of knowledge. Inductive learning is utilized in tree construction synchronization. Trees influence the process of the selection for building trees [8].

Classification algorithms RF strategies depend on the compound of many predicator trees. The function of such ensemble of classifier (EoC) is that its tree is dependent upon the random growing. RF is a general norm of the randomized aggregates as this idea has been accepted [9], [10]. RF has made many adjustments in recent years, in this work development of random forest using meerkat clan algorithm (MCA) which is one of the swarm intelligence algorithms is presented [11]. Meerkat has three types of behavior sentry, foraging, and baby-sitter. Through them, the algorithm is built by dividing the solution groups into 2 groups, and all operations have been carried out on the foraging group. Sentry is the optimal solution [12], [13]. There are many early studies that were carried out with the aim of identifying relevant samples and excluding weak, repetitive, and noiseless samples because of their significant impact on classification processes and the accuracy of cadastral joins ,they are listed in order from oldest to newest.

In 2018 Aonpong *et al.* [14], in this study, we offer a new random forest technique that is tailored to the problem of land cover mapping. Pixel-based, neighbor-looking, and a mix of both techniques are studied. When using the pixel-based technique, we take advantage of the fact that all decision trees are unique, whereas when using the neighbor-looing strategy, we use the judgments from surrounding pixels when the RF decisions are unclear. Our results indicated that our new RF techniques beat the traditional one on both simulated and real-world data sets.

In 2019 Tyralis *et al.* [15], this research paper aims to make RF and its variants more accessible to practical water scientists, as well as to examine associated methods and concepts that have gotten less attention from water science and the hydrologic community. This work analyzed RF applications in the resources of water, indicate the potential of the original method and its derivatives, and evaluate RF exploitation degree in a variety of applications as a result of this work. RF implementations in the RF programming language, along with the associated techniques and concepts, are also explored.

In 2019 Georganos *et al.* [16], the geographical random forest (GRF) is a new geographical version of RF that may be used as an exploratory and a predictive tool for the estimation of the population as an RF covariates' function. GRF can be defined as a geographical disaggregation of RF in a form of local sub-models. This work indicates that the GRF could be better predictive in the case where a suitable spatial scale has been used to represent data, with lower residual auto-correlation values, based on the first empirical results. Lastly, and perhaps most importantly, GRF might be utilized as an exploratory tool to show the relationship between independent and dependent variables, indicating significant local variations and allowing a better understanding of mechanisms that might be creating spatial heterogeneity.

In 2020 Kang *et al.* [17], using an RF technique, this research paper provides statistical machines for predicting excitation energies and corresponding oscillator strengths of a particular molecule. The emission spectrum and quantum yield of the fluorophores are closely associated with excitation energy and oscillator strengths, respectively. From the feature importance analysis regarding this RF approach, this work uncovered certain molecular fragments and substructures which govern the oscillator strengths of molecules. This finding is intended to serve as a new design principle for new fluorophores.

In 2021 Brophy *et al.* [18], this research paper provides data removal-enabled (DaRE) forests, a type of RF which allows training data to be removed with little retraining. Model updates are precise for each DaRE tree in the forest, which means that deleting instances from a DaRE model produces the same model as retraining from scratch on updated data. This study discovered that DaRE forests remove data orders of magnitude faster compared to retraining from scratch, sacrificing no or little predictive power in trials on one synthetic dataset and 13 real-world datasets.

In 2021 Antoniadis *et al.* [19], using RF as a non-parametric method for generating meta-models which allows for rapid sensitivity analysis. Aside from its ease of application to regression problems, RF has several strong benefits, including the capability for implicitly dealing with correlations and high dimensional data, dealing with variable interactions, and recognizing informative inputs with the use of a permutation-based RF variable importance index that is simple and quick to compute. Also, this work discussed a suitable set of tools for measuring variable relevance, which is after that used for decreasing the model's dimension, allowing previously impossible sensibility analysis investigations to be conducted. To demonstrate the efficiency of such an approach, numerical results from many simulations and data exploration on a real dataset are shown.

In this paper, it is proposed to modify the random forest algorithm for selecting features using the MCA to improve and increase the performance of the RF algorithm depending on choosing a group of features that give the highest percentage of accuracy and by turning the solution group into two groups of foraging and caring. Most operations are on the forage set and replace the worst solutions with the best solution. The worst solution is dropped into the care setting and a randomly generated solution is added. These results show the amazing performance of the algorithm in achieving optimal or near-perfect solutions

at a very fast rate. The rest of this paper contains related work in section 2, RF algorithm in section 3, MRF in section 4, experimental results in section 5, and conclusions in section 6.

- Random forest algorithm

Random forest's versatile and easy machine-learning technology provides great results even without defining the super parameters and is the most used because of its simple advantages. It is simple and can be used for classification and regression functions [15]. Random forest, first suggested by Tin Kam Ho of Bell Labs in 1995, is a classification and regression learning technique.

The idea behind (RF) is defined as a general principle for random groups of decision trees. Leo Breiman and Adele Cutler created a random forest induction algorithm (2001) key concept is to create a greater number of decision trees. In this way, the error association between classifiers is that. By using a random set of features at each node to be separated. Advantages RF can be easily constructed and predicted faster, immune to overwork and overfit results, ability to handle data without preprocessing or recalibration, resistant to outliers, and can handle null values [20], [21].

To determine the best feature, RF employs the gain index or information gain, which assesses how well a certain feature classifies or separates target classes through the calculation of entropy reduction. The optimal feature has been selected as the feature that has the highest information gain. Entropy, in simple terms, represents a measure of disorder, and the entropy of a dataset is the measure of disorder in the dataset's target feature. Entropy is 0 if all values in the target column are homogeneous (i.e. similar) and 1 in the case where the target column has an equal number of values for both classes in binary classification (when the target column has just two types of classes). In this work, multiple trees are constructed without the use of information to identify new solutions and trees (random forest). The RF algorithm's primary steps are listed:

a) Take a sample of the original data for the bootstrap
b) Grow a tree using data from step 1. When growing the tree, at every node in the tree, determine the optimal split for the node using $m < p$ randomly selected variables. Grow a tree so each terminal node does not contain less than n & 1 cases.
c) Repeat steps 1–2, $B > 1$ time independently
d) Combine the B trees to form the ensemble predictor. Data points to the category that wins the majority votes [22], [23].

- Meerkat clan algorithm

It is the swarm Intelligence algorithms resulting from meerkat behavior in the Kalahari Desert in South Africa. Meerkat lives in gatherings where each gathering consists of 20 to 50 male and female partners. Meerkat has three types of behaviors.

Foraging and sentry and baby sitter. By splitting the solution groups into two groups, the algorithm is generated and all operations are performed in the foraging group. The best solution is sentry behaviour, meerkat establishes good behavior within its colonies at least one sentry (lookout) will be there, while the others will search or play. The goal of the sentry is to inform them of the dangers and threats. Foraging behavior is a natural social mongoose behavior. When the animals spread for search, and maintain visual and sound contact separately [24], [7], [25].

It carefully fills the feed within the home, taking a different path every day. Babysitter behavior, meerkat is concerned with the main auxiliary measures which are monitoring children, and the assistants remain with the puppies 25 inside the package, while the rest of the group searches for food, where the assistants give a measure of food to the puppies while searching for food [13]. The algorithm 2 shows the basic phases of the meerkat clan algorithm as shown:

a) Read parameters N = 20–50 clan, FS size foraging where $FS < n$, CS size care $N - FS - 1$.
b) locate worst rate (WR) of foraging, lowest rate (LR) of care, L solution for a neighbour.
c) Generate random solution clan (N).
d) Calculate clan solution fitness by sentry = best clan solution.
   - Split the two clan (foraging and care) groups.
   - Call neighbor-generate (L, sentry, foraging i, best-one).
   - Foraging i = best neighbor of L.
   - Swap the solution worst for WR in the best group foraging.
e) Choose the best one to call it bestforg.
f) End.

2. METHOD

This research focuses on creating an algorithm to solve the problem of the large volume of data entering the processing process, which is sometimes weak and prone to repetition. Figure 1 illustrates the

proposed algorithm stages for the features selection technique. Where a proposed technique required several steps to solve the problem of large size of samples, the exclusion of weak, repetitive, and noise, and the extraction of strong relevant features that achieve better results in classification processes. By the random forest algorithm as:

a) Generation random sampling of features.
b) Split the dataset into several blocks.
c) Build random forest (trees) for each block.
d) Refinement the accuracy for each block via the proximity matrix.
e) Application of the MCA algorithm to the RF to improve the selection process of the RF by selecting the best blocks that achieve the optimized solutions.

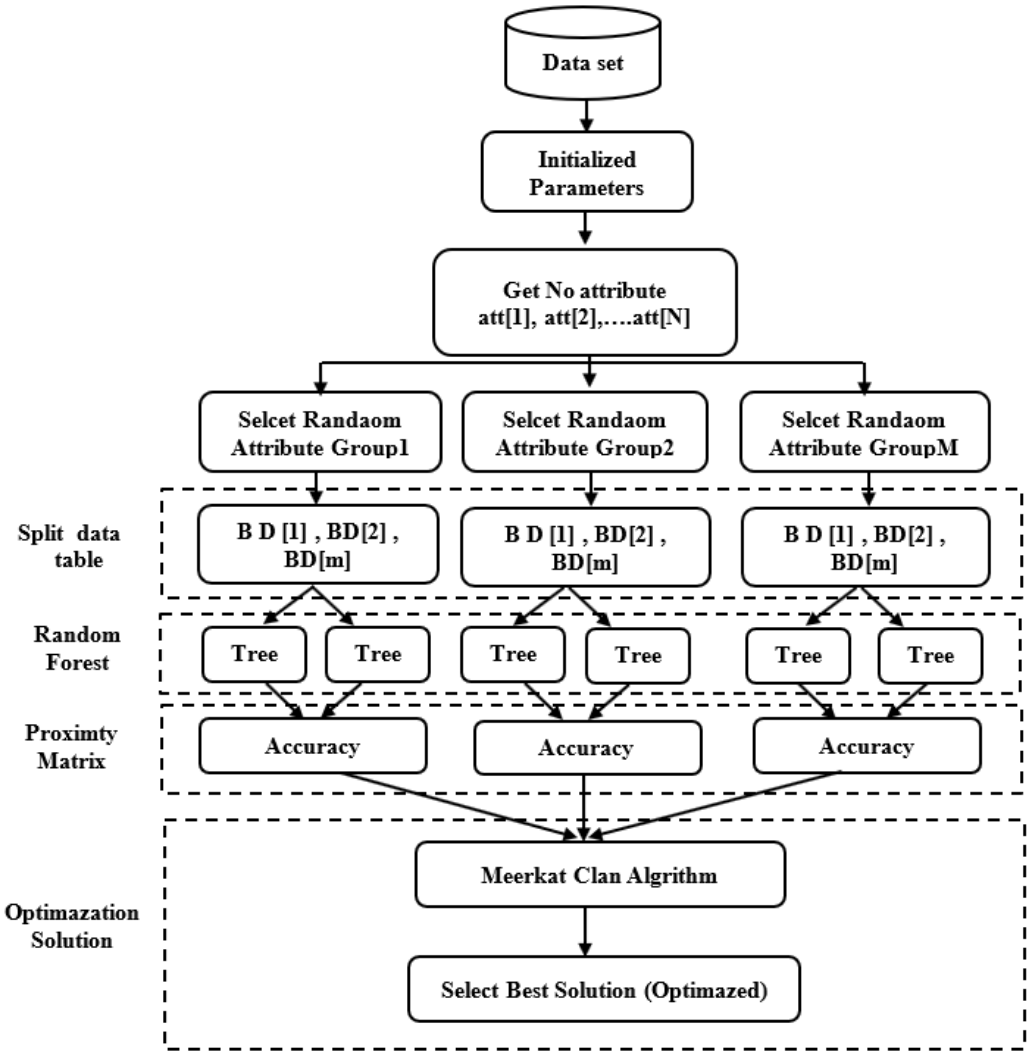

Figure 1. Block diagram of the proposed algorithm

2.1. Initialize parameters

A proposed technique requires initialized parameters and split dataset to number of blocks in the first stage. where initialized parameters include an input size block, number of trees, number of iterations, and number of neighbors. Bagging, or bootstrap aggregating, is the process of training each learner on multiple bootstrapped subsets of the data and after that averaging the predictions. The RF other major notion is that just a subset of all attributes is taken into account when dividing each node in each decision tree. Algorithm 1 explains the total features required to create various tees for classification (excluding those with null values).

Algorithm 1. Generating random samples

Input: Dataset
Output: Sample of the dataset
Begin
- Step 1: // Dataset analysis
Select attribute of miss value // attribute name include miss value
Set no of attributes // number attribute of table
Set all attributes // set attribute without missing value
Set no of tree // number of tables
Input block size
- Step 2:
No of blocks size = (table name) / block size
Splitting data set into blocks
- Step 3: For each block
Building no of trees
For each tree
Generate a set of random attributes to which the attribute of the
the missing value is combined
Fill data from table name
End for
End for
- Step 4: Return sample of the dataset
- Step 5: End

2.2. Build a random forest

The process of building RF depends mainly on the creation of multiple groups of decision trees that work separately to decide the existing data. The differences in trees in terms of the roots of them such each decision tree starts with tree different and thus there are structures for different trees and different results. The final decision is depending on voting in that are adopted as a final value. Random-forest as explained in algorithm 2.

Algorithm 2. Building of random forest

Input: Set of blocks
Output: Random forest // build tree for each block
Begin
- Step 1:
Get no attributes from block // number attribute of table
Get name of attributes in block // attributes name
- Step 2:
For each tree
A select random subset of attributes
For each attribute
Build tree
End for
End for
End

2.3. Apply meerkat clan algorithm on random forest

MCA splits the set solution into two sets foraging and care. Most of the operations on the foraging set and the worst solutions replacement by the best in the care solution. The worst solution in the care setting is dropped and a randomly generated solution is added. These results show the algorithm's incredible performance in achieving optimal or near-optimal solutions at an extremely fast rate. The MCA process is used to improve the performance of the algorithm random forest via the feature selection process by selecting better block data that represent good tree solutions and relying on them instead of relying on all trees and characteristics. Through its reliance on finding an initial solution that represents the best solution, and then doing the process of repetition in the search neighborhood and comparing it with the old solution and choosing the best one as explained in algorithm 3.

Algorithm 3. Apply MCA on random forest

Input: Instances of accuracy rate // for all block
Blocks data // for each accuracy rate
No of neighbors
Max no of iterations

Output: Optimum solution
Begin
- Step 1: // Initialized MCA solution
Initialized set of random forests from random attribute with its targets
Evaluate each random forest // accuracy
- Step 2: // Split solution
Select the best solution
The select random solution from the reminder
Find the neighborhoods of solution
Swapping the solution tree with block neighbor solution trees
Evaluating results
- Step 3: // Select local best
If the result is better than the old solution
Swap block random forest trees
Else
Don't change
End if
- Step 4: // Do specific iteration
For each iteration applying steps 2 and step3
Update the total results depend on evaluation (accuracy)
End for
- Step 5: // Find best
Return best solutions
End algorithm

3. RESULTS AND DISCUSSION

Three different datasets were utilized for implementing and testing the suggested approach on a real dataset from the UC Irvine ML repository [19]. "Adult income" or simply "adult" is a standard imbalanced ML dataset see Table 1, contraceptive method choice (CMC) is a sub-set of the 1987 national Indonesia contraceptive prevalence survey see Table 2, and credit approval is concerned with credit card applications see Table 3. The values and names of all attributes were replaced with meaningless symbols. According to the results in Figure 2, the suggested equation is superior to the measurements identified in the standard of RF classics.

Table 1. Adult data set description

Data set characteristics	Multivariate	Number of instances	48842	Area	Social
Attribute characteristics:	Categorical, integer	Number of attributes:	15	Date donated	1996-05-01
Associated tasks:	Classification	Missing values?	Yes	A number of web hits:	1889759

Table 2. CMC data set description

Data set characteristics	Multivariate	Number of instances	1473	Area	Life
Attribute characteristics:	Categorical, integer	Number of attributes:	9	Date donated	1997-07-07
Associated tasks:	Classification	Missing values?	No	A number of web hits:	194684

Table 3. Credit approval data set description

Data set characteristics	Multivariate	Number of instances	690	Area	Financial
Attribute characteristics:	Categorical, integer, Real	Number of Attributes:	15	Date donated	N/A
Associated tasks:	Classification	Missing values?	Yes	A number of web hits:	425129

Table 4. Accuracy comparison result in three datasets

Dataset used	Random forest			Random forest via MCA		
	SB = 100	SB = 200	SB = 300	SB = 100	SB = 200	SB = 300
Adult	79.29%	79.79%	80.11%	97.26%	98.93%	99.12%
CMC	77.80%	79.34%	80.18%	97.12%	96.49%	97.50%
Credit approval	78.17%	79.31%	80.89%	99.06%	99.08%	99.12%

Table 5. Time estimation comparison result in three datasets

Dataset used	Random forest			Random forest via MCA		
	SB = 100	SB = 200	SB = 300	SB = 100	SB = 200	SB = 300
Adult	06.266	20.054	36.627	04.359	13.457	30.706
CMC	02.094	03.250	09.879	01.155	02.875	11.985
Credit approval	03.547	12.873	29.018	02.616	11.452	24.813

The result of the comparison between standard random forest algorithm and modified random forest via MCA is explained in Table 4 and Table 5. Where the number of iterations is 100, the number of trees is 3, the number of blocks is 8, the number of neighbors is 2, and the size of a block is 100, 200, and 300. The results showed that the proposed algorithm has a more accurate result than the random forest in all selected block sizes as shown in Figure 2.

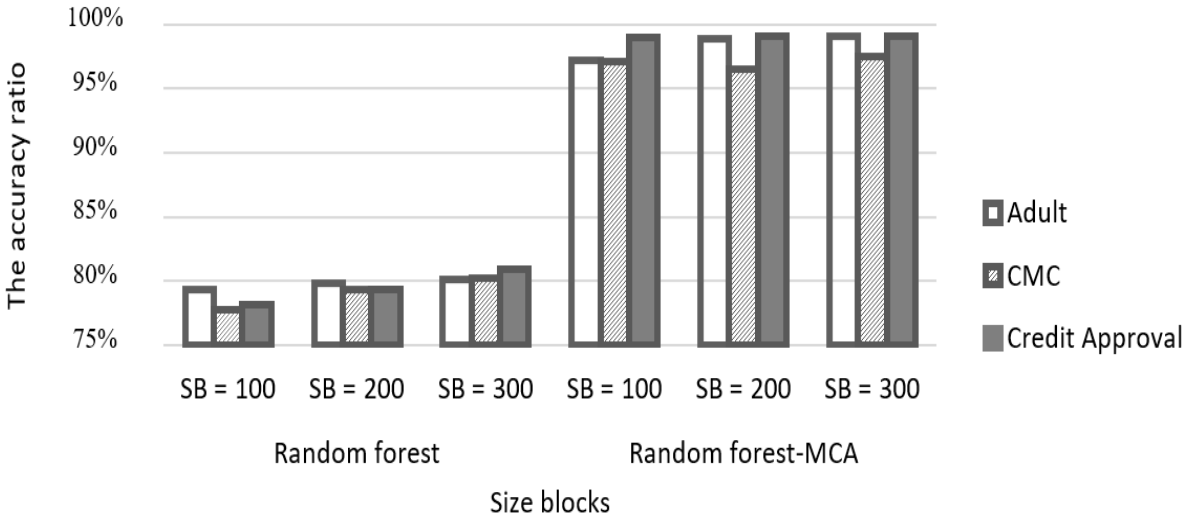

Figure 2. Displays accuracy for both RF and RF-MCA algorithms

4. CONCLUSION

Selecting relevant genes for sample classification has become a common task in most gene expression studies, the proposed algorithm has been tested and proven very successful to find the smallest possible set of genes that can still achieve good predictive performance. The accuracy is improved by the RF-meerkat because it trained within features more than the original RF. By using 100 iterations in RF-meerkat the accuracy is good also the time is less than the original RF. But in 200 and 300 iterations the time complexity increases with some accuracy. The increase in the size of the blocks in RF-meerkat is leading to an increase in the accuracy of the null value imputation. The increase in the number of trees in each dataset will not increase the accuracy rate estimation, it depends on the kind of dataset. Using several types of null values (categorical and numerical) makes the proposed method more adaptable and could be tested on other kinds of datasets.

REFERENCES

[1] I. H. Sarker, "Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions," *SN Computer Science*, vol. 2, no. 420, 2021, doi: 10.1007/s42979-021-00815-1.
[2] C. Robert, "Machine Learning, a Probabilistic Perspective," *Chance*, vol. 27, no. 2, pp. 62–63, 2014, doi: 10.1080/09332480.2014.914768.

[3] J. Schmidt, M. R. G. Marques, S. Botti, and M. A. L. Marques, "Recent advances and applications of machine learning in solid-state materials science," *npj computational materials*, vol. 5, no. 83, 2019, doi: 10.1038/s41524-019-0221-0.
[4] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Computer Science*, vol. 2, no. 160, 2021, doi: 10.1007/s42979-021-00592-x.
[5] S. Lemm, B. Blankertz, T. Dickhaus, and K. R. Müller, "Introduction to machine learning for brain imaging," *Neuroimage*, vol. 56, no. 2, pp. 387–399, 2011, doi: 10.1016/j.neuroimage.2010.11.004.
[6] E. M. H. Saeed, H. A. Saleh, and E. A. Khalel, "Classification of mammograms based on features extraction techniques using support vector machine," *Computer Science and Information Technologies*, vol. 2, no. 3, pp. 121–131, 2021, doi: 10.11591/csit.v2i3.p121-131.
[7] A. Baylar, D. Hanbay, and E. Ozpolat, "An expert system for predicting aeration performance of weirs by using ANFIS," *Expert Systems with Applications*, vol. 35, no. 3, pp. 1214–1222, 2008, doi: 10.1016/j.eswa.2007.08.019.
[8] Tin Kam Ho, "The random subspace method for constructing decision forests," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832-844, Aug. 1998, doi: 10.1109/34.709601.
[9] Y. Amit and D. Geman, "Shape quantization and recognition with randomized trees," *Neural Computation*, vol. 9, no. 7, pp. 1545–1588, 1997. [Online]. Available: https://galton.uchicago.edu/~amit/Papers/shape_rec.pdf
[10] A. Choi, A. Shih, A. Goyanka, and A. Darwiche, "On Symbolically Encoding the Behavior of Random Forests," *arXiv*, pp. 7–9, 2020, doi: 10.48550/arXiv.2007.01493.
[11] F. Provost and T. Fawcett, "Data science and its relationship to big data and data-driven decision making," Big data, vol. 1, no. 1, pp. 51–59, 2013, doi: 10.1089/big.2013.1508.
[12] H. A. AlSudani, E. M. Hussain, and E. A. Khalil, "Classification the Mammograms Based on Hybrid Features Extraction Techniques Using Multilayer Perceptron Classifier," *Al-Mustansiriyah Journal of Science*, vol. 31, no. 4, pp. 72–79, 2020, doi: 10.23851/mjs.v31i4.902.
[13] S. Zhou and L. Mentch, "Trees, Forests, Chickens, and Eggs: When and Why to Prune Trees in a Random Forest," 2021, [Online]. Available: http://arxiv.org/abs/2103.16700.
[14] P. Aonpong, T. Kasetkasem, P. Rakwatin, I. Kumazawa, and T. Chanwimaluang, "The Random forest approach for land cover mapping," *2016 7th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES)*, 2016, pp. 1-6, doi: 10.1109/ICTEmSys.2016.7467112.
[15] H. Tyralis, G. Papacharalampous, and A. Langousis, "A brief review of random forests for water scientists and practitioners and their recent history in water resources, " Water, vol. 11, no. 5, 2019, doi: 10.3390/w11050910.
[16] S. Georganos *et al*., "Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling," *Geocarto International*, vol. 36, no. 2, pp. 121–136, 2021, doi: 10.1080/10106049.2019.1595177.
[17] B. Kang, C. Seok, and J. Lee, "Prediction of Molecular Electronic Transitions Using Random Forests," *Journal of Chemical Information and Modeling*, vol. 60, no. 12, pp. 5984–5994, 2020, doi: 10.1021/acs.jcim.0c00698.
[18] J. Brophy and D. Lowd, "Machine Unlearning for Random Forests," *arXiv*, 2020, [Online]. Available: http://arxiv.org/abs/2009.05567
[19] A. Antoniadis, S. L. -Lacroix, and J. -M. Poggi, "Random forests for global sensitivity analysis: A selective review," *Reliability Engineering & System Safety*, vol. 206, p. 107312, 2021, doi: 10.1016/j.ress.2020.107312
[20] Y. Jin, H. Wang, and C. Sun, *Data-Driven Evolutionary Optimization, Integrating Evolutionary Computation, Machine Learning and Data Science*. Switzerland AG: Springer Nature, 2021, doi: 10.1007/978-3-030-74640-7
[21] R. Houborg and M. F. McCabe, "A hybrid training approach for leaf area index estimation via Cubist and random forests machine-learning," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 135, pp. 173–188, 2018, doi: 10.1016/j.isprsjprs.2017.10.004.
[22] G. Manimannan, R. L. Priya, K. J. Reena, and S. K. Priya, "Climate Changes of Tamilnadu Based on Rainfall Data Using Data Mining Model Evaluation and Cross Validation," IOSR Journal of Computer Engineering (IOSR-JCE), vol. 20, no. 3, pp. 32-38, 2018, [Online]. Available: https://www.iosrjournals.org/iosr-jce/papers/Vol20-issue3/Version-2/E2003023238.pdf"
[23] O. Villacampa, "Feature Selection and Classification Methods for Decision Making: A Comparative Analysis," ProQuest Diss. Theses, College of Engineering and Computing, Nova Southeastern University, no. 63, p. 188, 2015, [Online]. Available: https://nsuworks.nova.edu/cgi/viewcontent.cgi?article=1064&context=gscis_etd
[24] J. Tooby and L. Cosmides, "Conceptual Foundations of Evolutionary Psychology," *The Handbook of Evolutionary Psychology*, pp. 5–67, 2015, doi: 10.1002/9780470939376.ch1.
[25] S. Fischer, M. B. -Nettelbeck, A. Kotrschal, and B. Taborsky, "Rearing-group size determines social competence and brain structure in a cooperatively breeding cichlid," *The American Naturalist*, vol. 186, no. 1, pp. 123–140, 2015, doi: 10.1086/681636.

BIOGRAPHY OF AUTHORS

Hayder Adnan Saleh received a Bachelor's degree in Computer Science from Ibn Al-Haytham College of the Pure Sciences / University of Baghdad in the year 2007, holds a master's degree from the Department of Computer Science at the College of Education, Al-Mustansiriya University in 2021. I am currently working as an assistant teacher in the Third Rusafa Directorate of Education, the Iraqi Ministry of Education. He can be contacted at email: haider8883@uomustansiriyah.edu.iq.

Rana Amer Sattar Birth: 1982, Baghdad, Certificate: Bachelor of computer sciences- Al-Mustansiriya University, Baghdad.Iraq, 2005, and a master's degree in computer (artificial intelligence) -Al-Mustansiriya University, Baghdad, Iraq. 2021. Occupation: Working in the Ministry of Education / Directorate of Education Rusafa First as an assistant teacher at Al-Rafidain Secondary School for Girls. She can be contacted at email: salihirana@gmail.com.

Enas Mohammed Hussein Saeed received the B.Sc. in computer science from Mustansiriyah University, Baghdad, Iraq in 1998, M.Sc. in computer science from Mustansiriyah University, Baghdad, Iraq in 2004, and Ph.D. in computer science from the University of Technology, Baghdad, Iraq in 2014. Current position & Functions: computer science from Mustansiriyah University. She can be contacted at email: drenasmohammed@uomustansiriyah.edu.iq and enas.saeed38@gmail.com.

Dalael Saad Abdul-Zahra received the B. Sc Mathematical from Babylon University, Babylon, Iraq, in 2015 and M.SC. Mathematical (Factional Analysis) from Kufa University, Najaf, Iraq, in 2020, I am currently working at Hilla University College as an assistant teacher in the Department of Medical Physics Interested in studying topics related to applied and pure mathematics, rubber engineering, and some special interests in computers and basic programming languages. She can be contacted at email: dalael20175@gmail.com.