https://doi.org/10.56452/0038

Step: 4 Output the global best firefly

Firefly Algorithm with Inversion uses the concept of reversing the firefly, and choosing the best between the original and reverse based on the fitness value and number of selected features. Firefly Algorithm without Inversion does not construct the inverse.

Termination Condition : The proposed algorithm terminates after the specified number of iterations is completed.

Objective function: f(x)

Step1: Scanning the firefly to select those attributes which are set by the corresponding position in the firefly

Step2: Based on selected attributes transform the dataset resulting in reduced dimension

Step3: Generating a Classifier model using reduced dataset

Step4: Calculating the accuracy score of the classifier

Fig.3. Objective Function outline

The sorting of fireflies is done based on the objective function output and count of selected features. During every iteration locally best firefly is identified and checked with the global best firefly for similarity. If the locally best is found to improved solution than global then global best firefly is updated to local.

4 EXPERIMENTAL RESULTS AND DISCUSSION

The proposed methodology is implemented using python language in Anaconda3 Jupyter Notebook. Using four dataset as described in Table I , the performance of the algorithm is evaluated. All the datasets used were assumed to be noise free and with no missing values. In the experiments conducted the execution time was not found to be too varying because of the size of the dataset was not too big but showed better results in terms of accuracy.

The Decision tree classifier model is used in proposed algorithm with the train-test split ratio 70:30 and random State = 10.

TABLE I. DATASET DESCRIPTION

S.No	Dataset Information		
	Name	***Instances***	***Attributes***
1	MobilePrice.csv ; Source : https://www.kaggle.com	1400	21+1
2	Winequality-white.csv ; Source : https://www.kaggle.com	4898	11+1
3	WDBC.csv; Source : https://www.kaggle.com/uciml/breast-cancer-wisconsin-data	569	32
4	spambase.csv ; Source : https://archive.ics.uci.edu/ml/datasets/spambase	4601	58

In this article SelectKBest available in scikit-learn library in Python is used which scores the features against the target variable using Chi-squared statistics. Principal Component Analysis (PCA) is a another dimensionality-reduction method is used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information found in the large set where data is projected into a subspace (principal components). The fraction of variance is way to measure performance of PCA. The count of selected features from proposed algorithm will be taken as input to PCA which indicates the number of principal components to be generated from original feature space. Then based on the principal components the dataset would be transformed and fit into a Decision Tree Classifier model to check its performance in terms of accuracy score and explained variation.Thus the accuracy score of SelectKBest, Principal Component Analysis and the proposed algorithm is compared and modified firefly is found to be superior.The proposed algorithm has some parameters to be set which controls the movement of fireflies in exploring and exploiting the search space as specified in Table II.

TABLE II. PARAMETERS USED IN MODIFIED FIREFLY ALGORITHM

S.No	Parameter	INITIALI VALUE
1	Number of Fireflies	10
2	MAX_ITERATION	20
3	Beta0	0.2
4	Gamma	1
5	Alpha	0.5

https://doi.org/10.56452/0038

The following tables and graphs depicts the evaluation of the proposed methodology with other methods with respect to each dataset depicting the progress of firefly algorithm along the iterations in terms of accuracy score , accuracy obtained using SelectKBest and the selected attributes, accuracy obtained using Principal Component Analysis – (PCA) method and the explained variance and Accuracy of Decision Tree classifier considering all the features (i.e., without feature selection)

TABLE III ACCURACY COMPARISON - : WINEQUALITY-WHITE.CSV

Methods	Dataset : winequality-white.csv		
	No. of Features	Accuracy %	Inference
Decision Tree Classifier	11	58.02	All the features were considered
SelectKBest	6 / 11	59.86	**Selected Attributes Indices** [1, 2, 4, 5, 6, 10]
PCA	6 PC	58.70	**Variance of PCs** [9.11312392e-01 7.73858874e-02 1.04363597e-02 5.13705969e-04 3.26074596e-04 8.75115973e-06]
Modified Firefly	6 / 11	61.29	**Selected Attributes Indices** [0 1 3 5 6 10]

Table III depicts the performance of the suggested method experimented on winequality-white dataset and compared with other feature selection methods. It is inferred that performance is near to the method when all features of dataset are considered.

TABLE IV ACCURACY COMPARISON - MOBILEPRICE DATASET

Methods	Dataset : MobilePrice.csv		
	No. of Features	*Accuracy %*	Inference
Decision Tree Classifier	21	80.23	All the features were considered
SelectKBest	9 / 21	80.95	**Selected Attributes Indices** [0 1 7 9 12 13 14 16 17]
Principal Component Analysis	9 PC	84.52	**Variance of PCs** [6.15630930e-01 1.50304460e-01 1.02199334e-01 8.21454322e-02 4.88303712e-02 6.52398618e-04 1.70246904e-04 2.39451178e-05 1.59354488e-05]
Modified Firefly	9 / 21	84.047	**Selected Attributes Indices** [1, 4, 5, 7, 11, 12, 13, 14, 15]

Table IV depicts the performance of the suggested method experimented on MobilePrice dataset and compared with other feature selection methods. It is inferred that performance is near to the method when all features of dataset are considered.

https://doi.org/10.56452/0038

TABLE V ACCURACY COMPARISON - SPAMBASE DATASET

Methods	**Dataset :** spambase.csv		
	No. of Features	***Accuracy %***	Inference
Decision Tree Classifier	57	90.51	All the features were considered
SelectKBest	34 / 57	90.15	**Selected Attributes Indices** [3 4 5 6 7 8 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 32 34 36 41 43 44 45 51 52 54 55 56]
PCA	34 PC	86.38	**Variance of Principal Components** [9.68278064e-01 3.06877420e-02 9.50108548e-04 2.70412368e-05 1.02334109e-05 6.56532010e-06 3.97609637e-06 3.41298123e-06 2.95145271e-06 2.81139523e-06 2.58740802e-06 2.19156268e-06 2.02497787e-06 1.94133447e-06 1.74100535e-06 1.41179790e-06 1.30353275e-06 1.05294082e-06 9.99237080e-07 9.44644356e-07 8.90522489e-07 7.68671938e-07 7.13870623e-07 6.34628759e-07 6.13453225e-07 5.92131080e-07 5.68607015e-07 4.87057264e-07 4.66427083e-07 4.26988579e-07 3.85059062e-07 3.58275120e-07 3.46110219e-07 3.18809109e-07]
Modified Firefly	34/57	93.26	**Selected Attributes Indices** [4, 5, 6, 8, 9, 10, 12, 13, 15, 17, 19, 20, 22, 24, 25, 26, 27, 30, 33, 34, 35, 36, 37, 39, 41, 44, 45, 46, 47, 49, 52, 53, 55, 56]

TABLE VI ACCURACY COMPARISON - WDBC DATASET

Methods	**Dataset :** WDBC.csv		
	No. of Features	Accuracy %	Inference
Decision Tree Classifier	31	92.39	All the features were considered
SelectKBest	10 / 31	91.22	Selected Attributes Indices [0 1 3 4 13 14 21 22 23 24]
PCA	10 PCs	94.15	**Variance of PCs** [1.00000000e+00 2.47308925e-11 4.42650282e-13 4.73483824e-14 2.93772913e-15 2.23365907e-15 1.86190927e-16 1.02637986e-16 2.26307710e-17 9.56742320e-18]
Modified Firefly	10 / 31	97.6	Selected Attributes Indices [7, 10, 11, 15, 18, 21, 24, 25, 26, 28]

Table V, V1 depicts the performance of the suggested method experimented on spambase and WDBC datasets and compared with other feature selection methods. It is inferred that performance is near to the method when all features of dataset are considered

https://doi.org/10.56452/0038

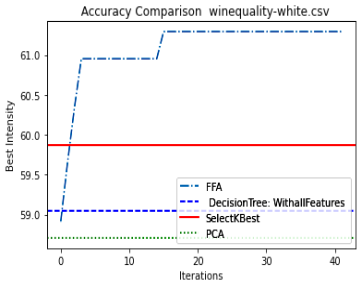

FIG.4 ACCURACY COMPARISON CHART- WINEQUALITY-WHITE.CSV

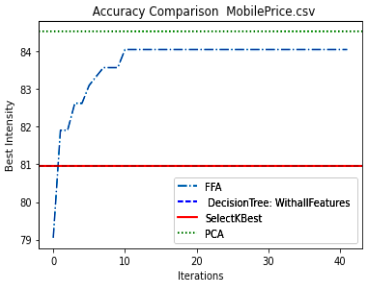

FIG 5. ACCURACY COMPARISON CHART- MOBILEPRICE DATASET

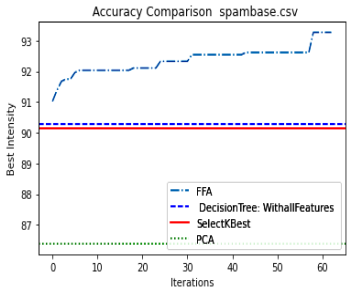

FIG 6. ACCURACY COMPARISON CHART - SPAMBASE DATASET

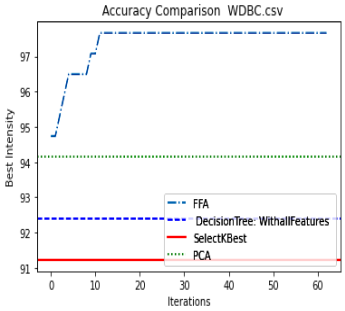

FIG 7. ACCURACY COMPARISON CHART - WDBC DATASET

In Fig 4,5,6,7 the graph represents the progress of proposed modified firefly algorithm with respect to the iterations representing the direction of searching for optimal solution and compared with the accuracy obtained It is evident that the proposed method is superior when compared to using SelectKBest and Principal Component Analysis(PCA).

Conclusion

In this research work, a Metaheuristics algorithm – evolutionary firefly algorithm is applied to determine a set of persistent features for the given dataset which exploits the benefits of swarm intelligence along with the evolutionary algorithmic mutation and selection techniques. To demonstrate the success of the algorithm's performance, four different datasets were used for testing and found to outperform in selecting features leading to the same/better accuracy when considered with all the features. Therefore, it can be considered as a better tool for high-dimensional data reduction problems. It is observed during experimentation that results fluctuate given the stochastic nature of the algorithm or evaluation procedure, or differences in numerical precision. Therefore, to get a consistent output , to discover optimal solution and to exploit the solution space in a better way the algorithm has to incorporate larger number of iteration and fireflies overcoming the time complexity issue, providing a scope for further research.

References

[1] J. H. Holland, "Genetic algorithms and the optimal allocation of trials ", SIAM Journal on Computing, vol. 2, no. 2, pp. 88–105, 1973.

[2] Sotirios K. Goudos, Christos Kalialakis, Raj Mittra, "Evolutionary Algorithms Applied to Antennas and Propagation: A Review of State of the Art", International Journal of Antennas and Propagation, vol. 2016, Article ID 1010459, 12 pages, 2016, https://doi.org/10.1155/2016/1010459.

[3] Manik Sharma, Prableen Kaur, A Comprehensive Analysis of Nature-Inspired Meta-Heuristic Techniques for Feature Selection Problem , Archives of Computational Methods in Engineering , 2019, https://doi.org/10.1007/s11831-020-09412-6.

[4] T. Ajay Adithyan, V. Sharma, B. Gururaj and C. Thirumalai, "**Nature inspired algorithm**," 2017 International Conference on Trends in Electronics and Informatics (ICEI), Tirunelveli, India, 2017, pp. 1131-1134 doi: 10.1109/ICOEI.2017.8300889.

[5] Slowik and H. Kwasnicka, "Nature Inspired Methods and Their Industry Applications—Swarm Intelligence Algorithms," in IEEE Transactions on Industrial Informatics, vol. 14, no. 3, pp. 1004-1015, March 2018, doi:10.1109/TII.2017.2786782.

[6] S. Agarwal, A. P. Singh and N. Anand, "Evaluation Performance Study Of Firefly Algorithm, Particle Swarm Optimization And Artificial Bee Colony Algorithm For Non-Linear Mathematical Optimization Functions," 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Tiruchengode, India, 2013, pp. 1-8, doi: 10.1109/ICCCNT.2013.6726474.

[7] Mohammed Abdulrazaq Kahya, Suhaib Abduljabbar Altamir & Zakariya Yahya Algamal (2019) Improving firefly algorithm-based logistic regression for feature selection, Journal of Interdisciplinary Mathematics, 22:8, 1577-1581, DOI: 10.1080/09720502.2019.1706861.

[8] S. MAZA and D. ZOUACHE, "Binary Firefly Algorithm for Feature Selection in Classification," 2019 International Conference on Theoretical and Applicative Aspects of Computer Science (ICTAACS), Skikda, Algeria, 2019, pp. 1-6, doi: 10.1109/ICTAACS48474.2019.8988137.

[9] Pan, X., Xue, L. & Li, R. A New And Efficient Firefly Algorithm For Numerical Optimization Problems. Neural Comput & Applic 31, 1445–1453 (2019). https://doi.org/10.1007/s00521-018-3449-6.

[10] Abdullah A, Deris S, Anwar S, Arjunan SNV (2013) "An Evolutionary Firefly Algorithm for the Estimation of Nonlinear Biological Model Parameters", PLoS ONE 8(3): e56310 , https://doi.org/10.1371/journal.pone.0056310.

[11] X. Zhang, Z. Shi, X. Liu and X. Li, "A Hybrid Feature Selection Algorithm For Classification Unbalanced Data Processsing," 2018 IEEE International Conference on Smart Internet of Things (SmartIoT), Xi'an, China, 2018, pp. 269-275, doi: 10.1109/SmartIoT.2018.00055.

[12] Fan, X., Sayers, W., Zhang, S. et al. Review and Classification of Bio-inspired Algorithms and Their Applications. J Bionic Eng 17, 611–631 (2020). www.springer.com/journal https://doi.org/10.1007/s42235-020-0049-9.

[13] Lindfield, G., & Penny, J. (2017). The Firefly Algorithm. Introduction to Nature- Inspired Optimization,Chapter-5,85–100. doi:10.1016/b978-0-12-803636-5.00005-0.

[14] K. K. Ghosh, P. K. Singh, J. Hong, Z. W. Geem and R. Sarkar, "Binary Social Mimic Optimization Algorithm With X-Shaped Transfer Function for Feature Selection," in IEEE Access, vol. 8, pp. 97890-97906, 2020, doi: 10.1109/ACCESS.2020.2996611.

[15] Timea Bezdan, Dusan Cvetnic, Luka Gajic, Miodrag Zivkovic, Ivana Strumberger, and Nebojsa Bacanin. 2021. Feature Selection by Firefly Algorithm with Improved Initialization Strategy. In 7th Conference on the Engineering of Computer Based Systems (ECBS 2021). Association for Computing Machinery, New York, NY, USA, Article 8, 1–8. DOI:https://doi.org/10.1145/3459960.3459974

[16] Wang, L.; Gao, Y.; Gao, S.; Yong, X. A New Feature Selection Method Based on a Self-Variant Genetic Algorithm Applied to Android Malware Detection. *Symmetry* 2021, *13*, 1290. https://doi.org/10.3390/sym13071290

[17] Seyedali Mirjalili, Andrew Lewis,S-shaped versus V-shaped transfer functions for binary Particle Swarm Optimization,Swarm and Evolutionary Computation,Volume 9,2013,Pages 1-14,ISSN 2210-6502,https://doi.org/10.1016/j.swevo.2012.09.002.

(https://www.sciencedirect.com/science/article/pii/S2210650212000648)