

Hybrid ANOVA and DOA-GA Feature Selection for EEG-Based Emotion Recognition

Ghassan G. Kanaan*, Mu'taz Al-Zghoul, Tarek Kanaan

Department of Artificial Intelligence-Faculty of Science and Information Technology,
Al-Zaytoonah University of Jordan, Amman, Jordan

Abstract—Using a Dhole Optimization Algorithm-Genetic Algorithm (DOA-GA) wrapper, this study reduces feature numbers from an ANOVA filter retaining 500 features. DOA and GA utilize half of the iteration budget, and the fitness function balances subset size and error using K-nearest neighbors. EEG Brainwave and HBN datasets were 4-channel and 128-channel, respectively. The final ANOVA filter-reduced feature sets were 232 and 241 features, yielding 90.9% and 95.9% reductions, respectively. In the Results section, XGBoost models had accuracies of 95.53% and 97.82% on the EEG Brainwave data and HBN data, respectively. Some overlap of participants and sessions may have biased the results. The validity of HBN labels and matched-size comparisons are still undetermined; the results of this study do not prove subject-independent emotion recognition or isolate the wrapper effect. The evaluation also compares optimization time, retained feature counts, and downstream classifier performance across six learning algorithms, providing a descriptive assessment of the trade-offs between dimensionality reduction and predictive accuracy.

Keywords—*Electroencephalography; emotion recognition; feature selection; metaheuristic optimization; dhole optimization algorithm; genetic algorithm*

I. INTRODUCTION

To create an emotion-detecting technology, we must identify aspects of human experience that can still be measured if a person tries to control their emotions or hide their feelings. One such technology is electroencephalography (EEG). EEG can measure activity in the brain and interpret nuances in processing and its relation to mood and emotion. The device has a relatively quick sampling speed and millisecond precision [1]–[3]. There are, however, several artifacts and variable distractors that can get in the way of the data collection and can make EEG difficult to interpret [4], [5].

The biggest concern in this regard is when equipment used to collect data has a different density of electrodes. It has been shown that a wearable EEG device can create thousands of signals from only a handful of electrodes while a clinical EEG device can create the same signals from 64 or 128 electrodes. Training machines to decode the signals from EEG can help serve the goal of emotion detection, but not at the rate that is required in today's computing environment [6], [7]. Decreasing the number of signals created from the EEG can help serve the machine learning goal of emotion detection, and it can aid in creating an interactive, portable, and adaptable system.

Filter methods are fast, but don't capture interactions with the final classifier. Wrapper methods, on the other hand, are too

computationally expensive. However, recent works combining Wrapper and Metaheuristic methods have started to address limitations of combinatorial search, particularly partial convergence and a lack of stability. Most works based on the combination of Wrapper and Metaheuristic methods use only one of the search techniques and evaluate only one acquisition setting, which leaves a considerable amount of uncertainty regarding the effectiveness of a feature subset under different levels of channel density and substantive variation of the original feature set.

This paper investigates feature construction, ANOVA filtering, and a sequential DOA-GA wrapper. Its contribution is the evaluation of this combination across two EEG acquisition settings using repeated optimizer runs and six classifiers. Filter-wrapper and exploration-exploitation approaches already exist [9], [10], [17]; no first-of-its-kind claim is made.

II. RELATED WORK

A. EEG Emotion Recognition

EEG emotion-recognition pipelines commonly combine artifact removal, segmentation, feature extraction, feature reduction, and classification. Statistics like time-domain, spectral band power, and entropy are related to emotional valence and arousal; hence, their use in EEG analysis [2, 5, 13]. It has been shown that deep learning models can classify EEG data with high accuracy [6, 14]. However, when comparing the classification performance of conventional classifiers with deep learning models, the choice of features could play a greater role than selection effects [24].

B. Feature Selection and Metaheuristic Search

Filters such as ANOVA and mutual information select variables before classification; principal component analysis instead transforms the representation. ANOVA is attractive because it ranks each variable using between-class and within-class variation with minimal overhead [15]. Wrappers evaluate subsets using a learner and can capture predictive interactions, but their repeated evaluations are expensive. Evolutionary and swarm optimizers, including GA, particle swarm optimization (PSO), enhanced Coati optimization, and channel-wise approaches, have therefore been used to approximate optimal subsets [8]–[12], [16]. Prior work examines feature reduction [21] and hybrid GWO-XGBoost selection [22]. A limitation is the exploration-exploitation trade-off. Delaying convergence with exploration takes longer to find solutions. Premature exploitation can get stuck around local solutions. Combining complementary search behaviors is the core idea of most hybrid

*Corresponding author.

approaches [10], [17]. The current implementation uses a fixed mid-search handoff. The search first takes place through leader-guided stochastic movement to map candidate spaces. A genetic algorithm is then used to perform crossover, mutation and elitism to fine-tune solutions in the mapped region.

Sharma et al. studied the use of evolutionary algorithms for mental stress detection using EEG [25]. Awadallah et al. presented an improved binary Rat Swarm Optimizer for feature selection [26]. These works inspired the use of hybrid search but are not directly comparable under the present conditions.

III. MATERIALS AND METHODS

A. Datasets and Feature Construction

Two public datasets were selected to test acquisition heterogeneity. Dataset 1 is the EEG Brainwave dataset introduced by Bird et al. [18]. It contains 2,133 four-channel EEG samples associated with positive, negative, and neutral emotional states. After data-quality checks, 2,548 engineered variables were available. Dataset 2 is drawn from the Healthy Brain Network EEG resource [19]. Recordings from ten participants and 128 channels produced 3,951 samples for the experimental protocol. Signals were filtered from 0.5 to 40 Hz, segmented using event annotations, standardized, and represented by six time-domain and 40 frequency-domain features per channel, yielding 5,888 variables per sample. Table I summarizes the sample count, channel density, class count, and initial feature dimensionality of both datasets. For Dataset 2, the mapping from HBN event annotations to emotion labels and the per-class counts have not been documented. Its three-class results are therefore provisional and cannot yet be interpreted as validated emotion recognition.

TABLE I. DATASET CHARACTERISTICS

Dataset	Samples	Channels	Classes	Initial features
EEG Brainwave	2,133	4	3	2,548
Healthy Brain Network	3,951	128	3 (labels unverified)	5,888

Both datasets adopted a 75/25 random stratified split of windowed samples, and an internal 80/20 training split for wrapper evaluation. For both, normalization and ANOVA were fit to the outer training data. Note that windows belonging to the same participant or session may occur in both outer training data partitions. These are sample-level estimates and should not be considered results that are independent of subjects. Results from grouped evaluation depend on the learned preprocessing and feature selection being fitted to the corresponding inner training folds [23].

B. ANOVA Filtering

For every feature, a one-way ANOVA score compares variation among the emotion classes with variation within the classes. Features with larger F statistics carry stronger class-discriminative evidence. The 500 highest-ranked training variables were retained in both datasets. This fixed intermediate dimension controls wrapper cost while allowing the same optimization budget to be applied to substantially different original spaces.

C. DOA-GA Wrapper

A candidate solution is a binary vector in which one denotes a selected variable. K-nearest neighbors with $k = 5$ evaluate each subset inside the wrapper. The objective combines validation error and subset size: $\text{Fitness} = 0.9 \times \text{Error} + 0.1 \times (\text{Selected features} / 500)$. The higher weight on error prevents excessive sparsity from dominating predictive quality. The 0.9/0.1 weights are fixed design choices; their influence on the accuracy–sparsity trade-off has not been tested.

DOA initializes a population of binary candidates and updates them relative to the current leader while injecting Gaussian perturbations. Its convergence factor decreases linearly across the allocated iterations, enabling large exploratory moves early and finer leader-centered moves later [20]. At the midpoint, the best DOA population is transferred to GA. Single-point crossover recombines candidates, mutation is applied with probability 0.10, and elitism retains the globally best solution. The fixed handoff is a design choice whose benefit has not been isolated. The complete data-processing and feature-selection workflow is illustrated in Fig. 1. Reproduction requires the implemented binary transfer rule and update equations, boundary handling, GA parent selection, crossover rate, mutation semantics, and empty-subset handling; these details must be verified from the code.

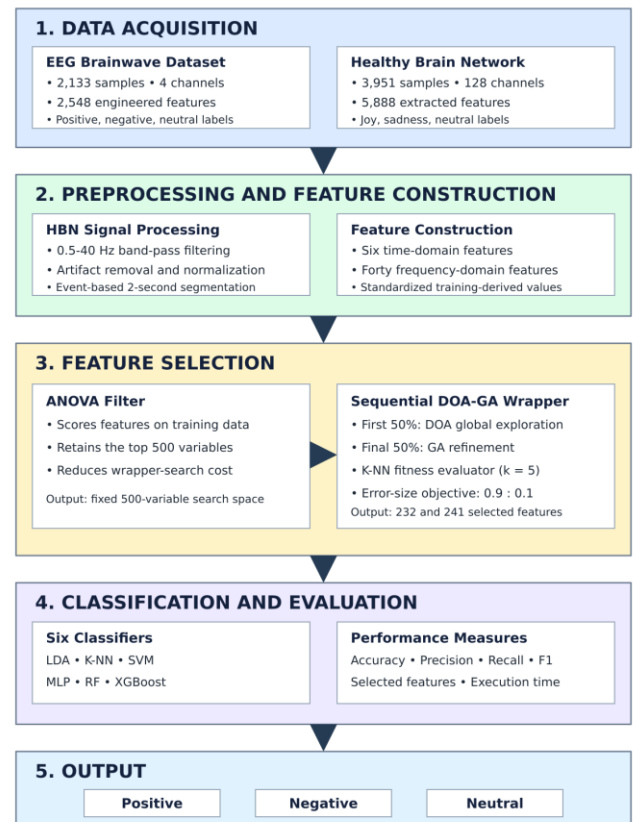


Fig. 1. Five-stage processing and feature-selection workflow.

As illustrated in Fig. 1, the proposed framework consists of five main stages. First, EEG data are acquired from the EEG Brainwave and HBN datasets. Second, the signals undergo dataset-specific preprocessing, segmentation, normalization,

and time- and frequency-domain feature extraction. Third, feature selection is performed through ANOVA filtering followed by the proposed DOA-GA wrapper. Fourth, the selected subsets are evaluated using six machine-learning classifiers and multiple performance measures. Finally, samples are classified into three target classes. For HBN, their interpretation as positive, negative, and neutral emotions remain subject to label validation.

D. Experimental Protocol

Population sizes of 20, 40, 60, 80, and 100 and iteration budgets of 10, 30, 70, and 100 were examined. Four standalone optimizers (DIO, PSO, GA, and DOA) and five hybrids (DIO-DOA, DIO-GA, DIO-PSO, DOA-PSO, and DOA-GA) were compared. Each configuration was run 30 times on one fixed outer split to assess optimizer variability. Final subsets were evaluated using linear discriminant analysis (LDA), K-nearest neighbors (K-NN), support vector machines (SVM), multilayer perceptron (MLP), random forest (RF), and XGBoost (XGB). Accuracy, weighted precision, weighted recall, weighted F1, selected-feature count, and execution time were recorded. Table II lists the principal data-splitting, filtering, fitness, and optimization settings used in all experiments. The full name and source of DIO must be confirmed from the implementation. Optimizer and classifier settings must be selected without using outer test outcomes.

TABLE II. CORE OPTIMIZATION SETTINGS

Component	Setting	Value
Split	Outer train/test	75/25
ANOVA	Retained variables	500
Wrapper	Evaluator	K-NN, k=5

Fitness	Error : size	0.9 : 0.1
Search	Population	20-100
Search	Iterations	10-100
Hybrid	DOA : GA	50% : 50%
GA	Mutation	0.10

IV. RESULTS

A. Dimensionality Reduction

ANOVA reduced 2,548 and 5,888 features to 500, removing 80.4% and 91.5%, respectively. DOA-GA then retained reported mean counts of 232 and 241, removing 53.6% and 51.8% of the 500-feature spaces. Total reductions were 90.9% and 95.9%. Thus, much of the compression came from ANOVA rather than the wrapper alone (Table III).

TABLE III. DIMENSIONALITY REDUCTION

Dataset	Original	After ANOVA	After DOA-GA (total reduction)
EEG Brainwave	2,548	500	232 (90.9%)
HBN	5,888	500	241 (95.9%)

B. Comparison of Search Algorithms

The hybrids listed in Table IV contain DOA-GA as their optimizer. The mean accuracy across samples for DOA-GA in Dataset 1 is 0.940, and 0.875 for Dataset 2. The corresponding F1 scores are 0.941 and 0.872. However, DOA-GA retains more features than the other alternatives. This, therefore, raises a confounding variable with respect to the choice of optimizer and subset size and does not isolate wrapper's superiority. The time vs. accuracy reports of the optimizers are presented in Fig. 2.

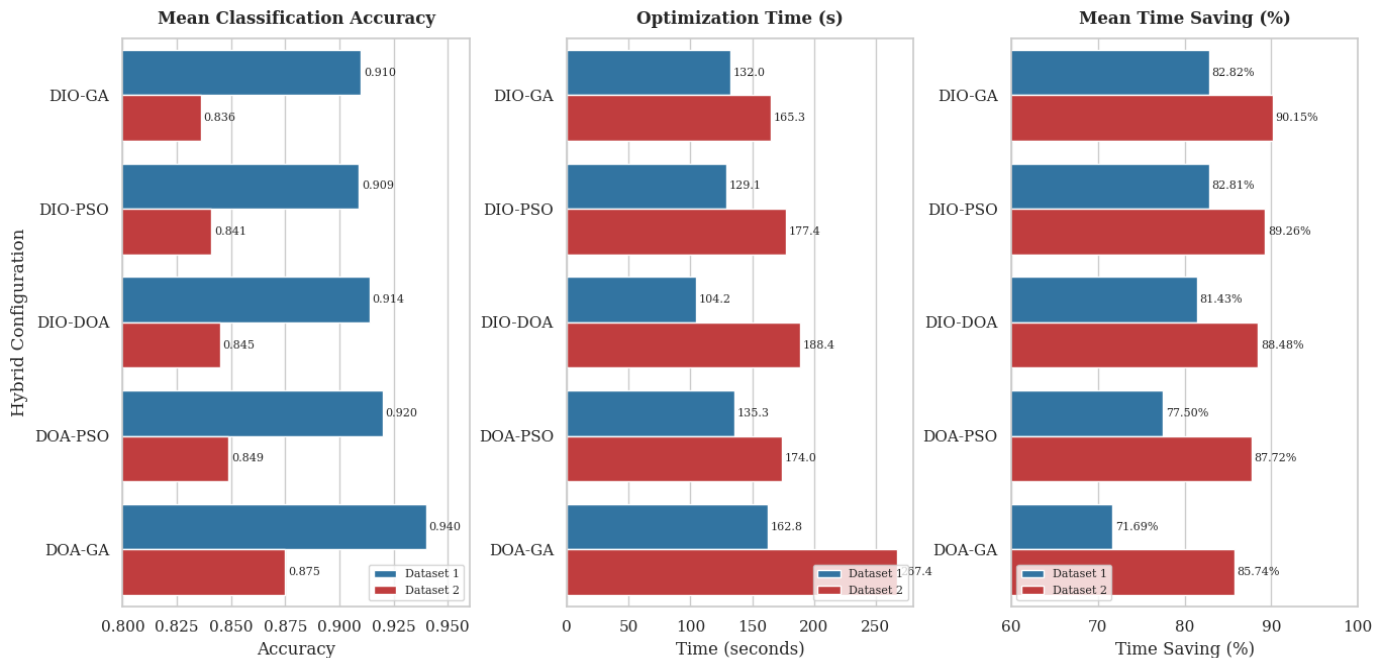


Fig. 2. Accuracy, optimization time, and time-saving trade-offs for hybrid optimizers.

TABLE IV. HYBRID PERFORMANCE AVERAGED ACROSS CLASSIFIERS

Method	Accuracy D1	Accuracy D2	Features D1/D2
DIO-DOA	0.914	0.845	43 / 144
DIO-GA	0.910	0.836	14 / 90
DIO-PSO	0.909	0.841	22 / 124
DOA-PSO	0.920	0.849	125 / 174
DOA-GA	0.940	0.875	232 / 241

C. Repeated-Run Descriptive Results

The 30-run accuracy summaries reported were 0.9387 ± 0.0038 for Dataset 1 and 0.8713 ± 0.0036 for Dataset 2. The source of these (i.e., whether they come from inner K-NN validation or outer tests) as well as the meaning of the dispersion needs to be explained by the logs. As provided in Table V, the gains and effect sizes at the run level will remain as provisional descriptive results. The previous conclusions at $p < 0.001$ and the effect at the population level are no longer valid. Optimizer seeds on one split are not independent evidence of generalization. Similar accuracy across runs does not establish stable feature selection; selection frequencies and subset overlap were not measured.

D. Downstream Classifier Performance

The Results section reports XGBoost accuracy and weighed F1 of 95.53% and 95.39% for EEG Brainwave (Dataset 1), and 97.82% and 97.40% for the HBN-derived task (Dataset 2). This assignment corrects the submitted abstract but must be reconciled with the original predictions and Fig. 3–5. These sample-level results do not demonstrate generalization to unseen participants; HBN emotion-label validity remains unresolved. These downstream comparisons do not establish classifier-independent optimality, since the subsets were selected using K-NN fitness.

TABLE V. PROVISIONAL RUN-LEVEL COMPARISON

Dataset	Baseline	Reported run gain	Run effect (provisional)
D1	DOA	0.0107	2.93
D1	PSO	0.0132	3.38
D1	DIO	0.0292	6.83
D2	DOA	0.0212	5.73
D2	PSO	0.0213	5.75
D2	DIO	0.0316	7.97



Fig. 3. Classifier performance and improvement after DOA-GA feature selection.

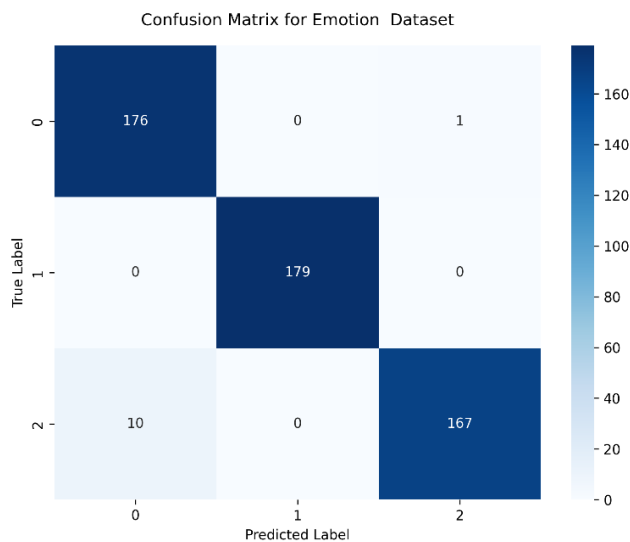


Fig. 4. Confusion matrix for XGBoost using the selected EEG Brainwave features.

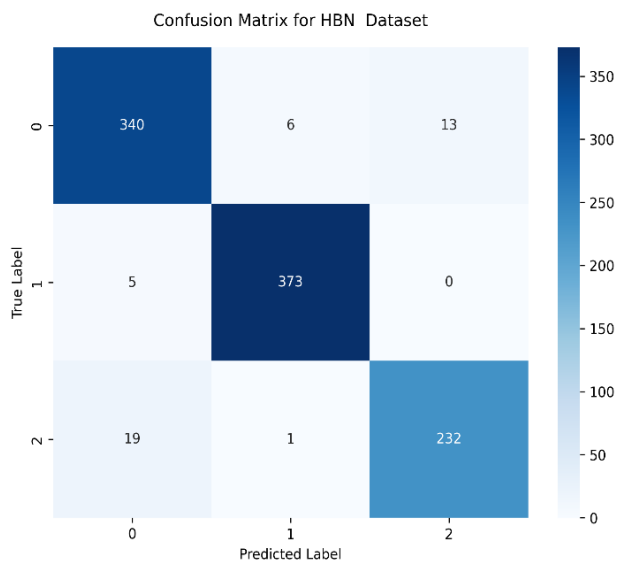


Fig. 5. Preliminary confusion matrix for the HBN-derived task; emotion-label validity remains unverified.

V. DISCUSSION

ANOVA ranks individual features, whereas the wrapper repeatedly evaluates candidate subsets using K-NN. DOA and GA provide sequential search operations; DOA-GA should not be attributed to the behavior of the DIO comparator. Differences between the datasets include participants, tasks, and feature construction, so these experiments do not isolate channel-density effects or demonstrate transfer learning. Wrapper superiority remains unconfirmed. Selected features have not been summarized by channel, frequency band, or feature family; their physiological relevance therefore remains unestablished.

Reduced classifier inputs may lower downstream cost, but this does not establish faster feature extraction or real-time deployment. End-to-end timing must account for preprocessing, feature construction, search, and inference. Matched-size

controls are needed to distinguish the wrapper contribution from ANOVA filtering. Feature reduction also does not demonstrate fewer physical electrodes or simpler acquisition hardware.

VI. LIMITATIONS

Randomly split EEG windows may share participant or session information and artificially inflate accuracy. For the main analysis, subject-independent or session-wise validation should be done. Sample-level analysis may be provided in addition. The HBN emotion classifier requires a documented event to label mapping and class counts. In the absence of such a mapping, the HBN emotion classifier should be either removed or re-framed. The requirements include matched-size analysis of variance (ANOVA), mutual information, recursive feature elimination (RFE), XGBoost ranking, DOA/GA, and handoff ablations. Also required are complete implementation details and comparable literature and deep-learning baselines. These experiments have not been performed in this revision.

VII. CONCLUSION

This study analyzed ANOVA followed by DOA-GA and noted final feature counts of 232 and 241 out of 2,548 and 5,888 variables. Total reduction includes ANOVA. Reported sample-level accuracies for XGBoost for EEG Brainwave and the HBN-derived task were 95.53% and 97.82%, respectively. These require verification against the original outputs. The results are preliminary. Generalizable improvements from DOA-GA depend on a combination of factors. The first is a grouped validation where, in addition to independently verifying HBN labels, DOA-GA and other processes are applied to datasets of equal size.

FUNDING

This research received no external funding.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY

The datasets analyzed in this study are publicly available from the sources cited in [18] and [19].

ETHICS STATEMENT

This study used publicly available, anonymized secondary datasets and involved no new recruitment or data collection from human participants.

DECLARATION ON GENERATIVE AI

Generative AI tools were used to assist language refinement and manuscript organization. The authors reviewed and edited the manuscript and remain fully responsible for the study design, data, analyses, interpretations, citations, originality, and integrity of the work.

REFERENCES

- [1] S. Afzal, H. Khan, M. Piran, and J. Lee, "A comprehensive survey on affective computing: Challenges, trends, applications, and future directions," *IEEE Access*, vol. 12, pp. 96150-96168, 2024, doi: 10.1109/ACCESS.2024.3422480.

- [2] E. Houssein, A. Hammad, and A. Ali, "Human emotion recognition from EEG-based brain-computer interface using machine learning: A comprehensive review," *Neural Computing and Applications*, vol. 34, pp. 12527-12557, 2022, doi: 10.1007/s00521-022-07292-4.
- [3] K. Varbu, M. Naveed, and Y. Muhammad, "Past, present, and future of EEG-based BCI applications," *Sensors*, vol. 22, no. 9, p. 3331, 2022, doi: 10.3390/s22093331.
- [4] Z. Xue et al., "Instrumentation, measurement, and signal processing in EEG-based brain-computer interfaces: Situations and prospects," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1-28, 2024, doi: 10.1109/TIM.2024.3417598.
- [5] R. Yuvaraj, P. Thagavel, J. Thomas, J. Fogarty, and F. Ali, "Comprehensive analysis of feature extraction methods for emotion recognition from multichannel EEG recordings," *Sensors*, vol. 23, no. 2, p. 915, 2023, doi: 10.3390/s23020915.
- [6] F. Lotte et al., "EEG classification algorithms: A review," *Journal of Neural Engineering*, vol. 15, no. 3, p. 031005, 2018.
- [7] K. Hossain, M. Islam, S. Hossain, A. Nijholt, and M. Ahad, "Status of deep learning for EEG-based brain-computer interface applications," *Frontiers in Computational Neuroscience*, vol. 16, 2023, doi: 10.3389/fncom.2022.1006763.
- [8] W. Chen, Y. Cai, A. Li, Y. Su, and K. Jiang, "EEG feature selection using maximum information coefficient," *Scientific Reports*, vol. 13, p. 14515, 2023, doi: 10.1038/s41598-023-41682-5.
- [9] B. Nakisa, M. Rastgoo, D. Tjondronegoro, and V. Chandran, "Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors," *Expert Systems with Applications*, vol. 93, pp. 143-155, 2018, doi: 10.1016/j.eswa.2017.09.062.
- [10] E. Houssein, A. Hammad, M. Emam, and A. Ali, "An enhanced Coati Optimization Algorithm for global optimization and feature selection in EEG emotion recognition," *Computers in Biology and Medicine*, vol. 173, p. 108329, 2024, doi: 10.1016/j.combiomed.2024.108329.
- [11] I. Wijayanto, S. Hadiyoso, A. Safitri, and T. Rahmiani, "Hybrid metaheuristic algorithms for EEG feature selection," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 2, pp. 366-379, 2025.
- [12] J. Zhang, S. Liu, H. Wu, Z. Zhang, and J. Long, "EEG feature selection in emotion recognition using a fuzzy information-theoretic based optimization approach," *IEEE Transactions on Fuzzy Systems*, vol. 33, no. 3, pp. 2675-2688, 2025, doi: 10.1109/TFUZZ.2025.3574786.
- [13] X. Wu, W. Zheng, Z. Li, and B. Lu, "Investigating EEG-based functional connectivity patterns for multimodal emotion recognition," *Journal of Neural Engineering*, vol. 19, no. 1, 2020, doi: 10.1088/1741-2552/ac49a7.
- [14] A. Topic and M. Russo, "Emotion recognition based on EEG feature maps through deep learning network," *Engineering Science and Technology, an International Journal*, vol. 24, no. 6, pp. 1442-1454, 2021, doi: 10.1016/j.jestech.2021.03.012.
- [15] B. Raufi and L. Longo, "Comparing ANOVA and PowerShap feature selection methods via Shapley additive explanations of models of mental workload built with theta and alpha EEG band ratios," *BioMedInformatics*, vol. 4, no. 1, pp. 853-876, 2024, doi: 10.3390/biomedinformatics4010048.
- [16] X. Xu et al., "Embedded EEG feature selection for multi-dimension emotion recognition via local and global label relevance," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, pp. 514-526, 2024, doi: 10.1109/TNSRE.2024.3355488.
- [17] S. Marjit, P. Das, U. Talukdar, and S. Hazarika, "Hybrid feature selection for EEG emotion recognition," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 37, pp. 871-895, 2024.
- [18] J. Bird, L. Manso, E. Ribeiro, A. Ekart, and D. Faria, "A study on mental state classification using EEG-based brain-machine interface," in *Proc. 9th International Conference on Intelligent Systems*, 2018, pp. 795-800, doi: 10.1109/IS.2018.8710576.
- [19] S. Shirazi et al., "HBN-EEG: The FAIR implementation of the Healthy Brain Network electroencephalography dataset," *bioRxiv*, 2024, doi: 10.1101/2024.10.03.615261.
- [20] B. Mohammed, H. Aghdasi, and P. Salehpour, "Dhole optimization algorithm: A new metaheuristic algorithm for solving optimization problems," *Cluster Computing*, vol. 28, no. 7, 2025, doi: 10.1007/s10586-024-05005-1.
- [21] L. Trujillo, D. Hernandez, A. Rodriguez, O. Monroy, and O. Villanueva, "Effects of feature reduction on emotion recognition using EEG signals and machine learning," *Expert Systems*, vol. 41, no. 8, 2024, doi: 10.1111/exsy.13577.
- [22] H. Asemi and N. Farajzadeh, "Improving EEG signal-based emotion recognition using a hybrid GWO-XGB feature selection method," *Biomedical Signal Processing and Control*, vol. 99, p. 106795, 2025, doi: 10.1016/j.bspc.2024.106795.
- [23] Q. She, X. Shi, F. Fang, Y. Ma, and Y. Zhang, "Cross-subject EEG emotion recognition using multi-source domain manifold feature selection," *Computers in Biology and Medicine*, vol. 159, p. 106860, 2023.
- [24] A. Nawaz, K. Cheah, H. Nisar, and V. Yap, "Comparison of different feature extraction methods for EEG-based emotion recognition," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 910-926, 2020, doi: 10.1016/j.bbe.2020.04.005.
- [25] L. D. Sharma, V. K. Bohat, M. Habib, A. M. Al-Zoubi, H. Faris, and I. Aljarah, "Evolutionary inspired approach for mental stress detection using EEG signal," *Expert Systems with Applications*, vol. 197, p. 116634, 2022, doi: 10.1016/j.eswa.2022.116634.
- [26] M. A. Awadallah, M. A. Al-Betar, M. S. Braik, A. I. Hammouri, I. A. Doush, and R. A. Zitar, "An enhanced binary Rat Swarm Optimizer based on local-best concepts of PSO and collaborative crossover operators for feature selection," *Computers in Biology and Medicine*, vol. 147, p. 105675, 2022, doi: 10.1016/j.combiomed.2022.105675.