Contemporary Mathematics
https://ojs.wiserpub.com/index.php/CM/

Research Article

Adaptive Random Search for Transparent Classification: Bridging Minimalist Machine Learning and Clinical Applications in Medical Datasets

Jorge Alberto Pacheco Senard[1], Mailyn Moreno Espino[2,3*], Cornelio Yáñez Márquez[1], Oscar Camacho Nieto[1], Yenny Villuendas Rey[4]

[1] Computer Research Center, Instituto Politécnico Nacional (National Polytechnic Institute), Mexico City, Mexico
[2] Institute of Knowledge Technology, Complutense University, Madrid, Spain
[3] Faculty of Informatics, Complutense University, Madrid, Spain
[4] Center for Innovation and Technological, Development in Computing, Instituto Politécnico Nacional (National Polytechnic Institute), Mexico City, Mexico
E-mail: mmorenoe2022@cic.ipn.mx

Received: 8 June 2025; **Revised:** 10 July 2025; **Accepted:** 22 July 2025

Abstract: Medical datasets frequently manifest a "horizontal" structure, wherein the number of features significantly exceeds the number of samples (e.g., 1,071 genes vs. 28 patients in glioma classification). Traditional validation methods encounter challenges with such imbalances, while complex machine learning models sacrifice interpretability for performance, resulting in "black boxes" that hinder clinical trust. To address this challenge, we propose Adaptive Random Search (ARS), a novel metaheuristic method that enhances stratified holdout validation by optimizing training subset selection. Developed at the Intelligent Computing Laboratory of Computing and Systems Center, National Polytechnic Institute (CIC-IPN), the birthplace of Minimalist Machine Learning (MML), ARS combines transparency with efficiency. By iteratively searching for optimal 20% training subsets, ARS reduces computational costs while maximizing balanced accuracy. To validate the proposed method, we employed a 1-Nearest-Neighbor (1-NN) classifier, selected for its lazy learning nature and interpretability, in eight medical datasets: glioma gene expression (Nutt), Parkinson's disease, gastrointestinal lesions, toxicity, Darwin, Gene Expression and SEMG for basics hands movements. The experimental results demonstrate that ARS attains state-of-the-art performance, achieving balanced accuracy of 100% on the Nutt and Gene Expression datasets and outperforming Support Vector Machine (SVM), Random Forest, and neural networks. In clinical settings, ARS-based classifications exhibited a stronger correlation with patient survival ($p = 0.05$) compared to histopathology in gliomas, underscoring its prognostic significance. By converting high-dimensional data into two-dimensional decision boundaries through statistical metrics (mean, standard deviation), ARS aligns with the MML principles, providing a computationally efficient and interpretable solution for medical diagnostics. This work demonstrates a synthesis of minimalist design with clinical utility, thereby substantiating the notion that simplicity need not compromise efficacy in high-stakes healthcare applications.

***Keywords*:** Minimalist Machine Learning (MML), Adaptive Random Search (ARS), medical datasets, stratified holdout, 1-Nearest-Neighbor (1-NN), interpretability

MSC: 68T10

DOI: https://doi.org/10.37256/cm.6520257403

1. Introduction

1.1 *The challenge of horizontal medical datasets*

Modern medical datasets, such as gene expression microarrays or biomedical sequences, often have a rectangular structure, where the number of features vastly exceeds the number of patterns. For example, the Nutt glioma dataset contains 1,071 genomic attributes for only 28 patients, while the Parkinson's Disease dataset includes 754 vocal features from 756 individuals [1, 2]. This feature-pattern disproportion is inherent to domains such as oncology and neurology, where high-resolution molecular or imaging data generate thousands of attributes per patient. Such rectangularity challenges conventional machine learning pipelines: models risk overfitting, validation becomes unreliable, and computational costs escalate.

1.2 *Limitations of conventional stratified holdout*

Stratified holdout validation, a gold standard validation method for datasets, partitions data into disjoint training (E) and testing (P) sets while preserving class distributions. However, in horizontal datasets such as Nutt (1,071 : 28 feature-to-sample ratio), fixed partitions often leading to biased performance estimates and misleading generalization capabilities. For example, a single outlier in small-sample regimes (e.g. 28 gliomas) can disproportionately skew accuracy metrics. Furthermore, complex models such as deep neural networks, though powerful, exacerbate these issues by prioritizing accuracy over interpretability, making them "black boxes" of limited clinical utility [3].

1.3 *The MML paradigm*

Minimalist Machine Learning (MML) is a novel approach developed at the Intelligent Computing Laboratory of CIC-IPN to address the challenges posed by high-dimensional medical datasets. MML reduces complex, high-dimensional data into compact 2D representations without sacrificing discriminative power [4]. This reduction enables simpler, more interpretable models while preserving key information for classification tasks.

Unlike conventional methods such as Support Vector Machines (SVM), which expand dimensionality through kernel transformations, or deep learning, which relies on stacked nonlinear layers that obscure decision-making processes, MML prioritizes transparency and interpretability. It leverages simple geometric and statistical principles, such as mean and standard deviation, to construct models that are both effective and easily understandable.

The fundamental premise of MML is that, when combined with intelligent validation strategies, transparent models can achieve performance comparable to that of complex black-box algorithms. This paradigm shift challenges the notion that high accuracy necessarily requires intricate architectures, advocating instead for solutions that balance performance with interpretability, which is crucial for clinical and biomedical applications. Recent studies emphasize the necessity of interpretable, dimension-reduced models for high-stakes domains like medicine [5].

1.4 *ARS: A metaheuristic enhancement*

Building on MML principles, we propose Adaptive Random Search (ARS), a novel validation framework that reimagines stratified holdout as an optimization problem. Unlike brute-force approaches, ARS dynamically adapts its search for optimal training subsets (a small percent of data) by leveraging metaheuristics:

1. Phase 1: Targets the highest convergence value of balanced accuracy defined by the user over a high number of iterations.

2. Phase 2: Refines the search for the second-best convergence value with a smaller number of iterations.

3. Phase 3: Halts exploration once convergence to a target accuracy threshold is achieved.

ARS employs a metaheuristic-driven approach to optimize training set selection, significantly reducing runtime compared to brute-force methods. By leveraging 1-Nearest Neighbor (1-NN), a non-parametric classifier requiring no pre-training, ARS generates interpretable decision boundaries suitable for clinical validation.

1.5 *Clinical relevance and objectives*

Rectangular datasets are ubiquitous in medicine, yet their analysis remains uncertain. In neurooncology, 30% of gliomas exhibit nonclassic histology, causing diagnostic discrepancies among pathologists and delayed treatments [6]. Similarly, Parkinson's Disease datasets struggle with early-stage detection due to subtle symptom variations. ARS addresses these issues by reducing runtime compared to grid search, critical for resource-constrained clinical settings. This work bridges MML theory and clinical practice across eight medical domains: gliomas (Nutt), Alzheimer's disease (Darwin), Parkinson's disease, gastrointestinal lesions, drug toxicity, period changer of core clock protein CRY1, Gene Expression Ribonucleic Acid (RNA) Sequence for Cancer and SEMG for basics hands movements [7–11].

1.6 *Paper organization*

Section 2 details the ARS methodology, contrasting it with stratified holdout. Section 3 presents the experimental results, followed by a discussion in Section 4. Finally, Section 6 concludes with future directions for interpretable AI in healthcare.

2. Methods

2.1 *Stratified holdout validation*

Stratified holdout validation partitions a dataset D into two disjoint subsets: a training set E and a test set P, preserving the original class distribution. Formally:

$$E \cup P = D,\ E \cap P = \varnothing \tag{1}$$

While effective for regular datasets, this method struggles with rectangular data (features $\gg$ samples), as small test sets may inadequately represent the feature space, leading to unstable performance estimates.

2.2 *ARS framework*

ARS enhances stratified holdout by treating validation as an optimization problem. It dynamically selects training subsets to maximize balanced accuracy while minimizing computational overhead. The ARS framework is implemented in two main phases: broad exploration and targeted refinement, as described in Algorithm 1.

Algorithm 1 Optimal training subset selection
Require: Dataset D with N samples
Require: Target convergence threshold τ
Require: Phase 1 iteration limit MAX_ITER_PHASE1
Require: Phase 2 iteration limit MAX_ITER_PHASE2
Ensure: Optimal training subset E_{best}, Maximum balanced accuracy ACC_{best}
1: $\text{ACC}_{\text{best}} \leftarrow 0$, $E_{\text{best}} \leftarrow \emptyset$
▷ Phase 1: Broad Exploration
2: **for** $i = 1$ to MAX_ITER_PHASE1 **do**
3: Randomly select E_i (20% of D), ensuring class balance
4: Train 1-NN on E_i, evaluate on $P = D \setminus E_i$
5: Compute balanced accuracy ACC_i
6: **if** $\text{ACC}_i > \text{ACC}_{\text{best}}$ **then**
7: $\text{ACC}_{\text{best}} \leftarrow \text{ACC}_i$, $E_{\text{best}} \leftarrow E_i$
8: **end if**
9: **end if**

▷ Phase 2: Targeted Refinement

for $j = 1$ to MAX_ITER_PHASE2 **do**
    Perturb E_{best} by swapping 5% of samples
    Re-evaluate ACC_j
    if $\text{ACC}_j > \text{ACC}_{\text{best}}$ **then**
        Update E_{best} and ACC_{best}
    end if
end if

▷ Convergence Check

if $\text{ACC}_{\text{best}} \geq \tau$ **then**
    Terminate and return E_{best}, ACC_{best}
else
    Repeat Phase 2 until convergence
end if

Key Properties:

1. Adaptivity: Phase 1 explores diverse subsets, while Phase 2 refines near-optimal candidates.
2. Flexibility: MAX_ITER_PHASE1 and MAX_ITER_PHASE2 are user-defined.
3. Efficiency: Avoids exhaustive search via probabilistic sampling.

A potential limitation of ARS is that its performance may degrade when the dataset contains high inter-class variance, requiring additional constraints to ensure subset stability.

2.3 *1-NN classifier*

The 1-Nearest-Neighbor (1-NN) algorithm classifies test samples based on the closest training instance in the feature space. For a test pattern X_{test}:

$$\text{Class}(x_{\text{ts}}) = \text{Class}\left(\arg\min_{x_i \in E} \text{dist}(x_{\text{ts}}, x_i)\right). \tag{2}$$

Where dist(·) is the Euclidean distance. In ARS, 1-NN operates on a small training subset ($E = 20\%$ in the experiments). This reduction aligns with MML principles, enabling better transparency and interpretability [4].

2.4 *Benchmark algorithms*

For comparative analysis, eight state-of-the-art classifiers were evaluated using traditional stratified holdout:

1. Support Vector Machine (SVM): Linear kernel, regularization $C = 1$.
2. Random Forest: 100 trees, Gini impurity.
3. Multilayer Perceptron (MLP): Two hidden layers (64, 32 neurons), ReLU activation.
4. Naive Bayes: Gaussian distribution.
5. Logistic Regression: L_2-penalty, $C = 1$.
6. J48 Decision Tree: Minimum leaf size = 5.
7. KNN: $k = 3$ and $k = 5$, euclidean distance.
8. Euclidean Classifier: centroid calculation, euclidean distance.

All baseline classifiers used for comparison—Logistic Regression, k-NN (3NN and 5NN), Naïve Bayes, Random Forest, J48 (Decision Tree), Support Vector Machine (SVM), Multilayer Perceptron (MLP), and the Euclidean distance classifier—were implemented using the Scikit-learn library in Python [12]. This open-source machine learning library provides standardized implementations of classic algorithms, ensuring reproducibility and transparency. Unless otherwise noted, all classifiers were used with the hyperparameters described above.

2.5 *Datasets and preprocessing*

Six medical datasets with rectangular structures were used (Table 1):

Table 1. Medical datasets

Dataset	Features	Samples	Source
Nutt (Gliomas)	1,071	28	[1]
Parkinson's	754	756	[2]
Gastrointestinal	512	76	[13]
Toxicity	1,746	171	[9]
Darwin	1,024	118	[7]
Period Changer	1,177	90	[8]
Gene Expression	20,531	801	[10]
SEMG	2,500	300	[11]

Source: UCI Machine Learning Repository (2025)

Preprocessing:
1. Missing Values: Imputed using feature medians by class.
2. Normalization: Min-Max scaling.

3. Results

Due to the presence of class imbalance in several of the biomedical datasets used (only Nutt, Gastrointestinal and SEMG presents balanced data), we adopt balanced accuracy as our primary evaluation metric. Traditional accuracy, defined as:

$$\frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{3}$$

can be misleading when the class distribution is skewed, as it favors the majority class. In contrast, balanced accuracy (defined as the average of sensitivity and specificity) gives equal importance to both positive and negative classes.

This choice is supported by the following derivation: when the dataset is balanced (Imbalance Ratio (IR) = 1), the accuracy and balanced accuracy become equal. Let us show this:

$$\begin{aligned} \text{Accuracy} &= (TP+TN)/(TP+FN+TN+FP) \\ &= TP/(TP+FN+TN+FP)+TN/(TP+FN+TN+FP) \\ &= 0.5*TP/(TP+FN)+0.5*TN/(TN+FP) \\ &= (\text{Sensitivity}+\text{Specificity})/2 = \text{Balanced Accuracy} \end{aligned} \tag{4}$$

This equality holds only when the dataset is perfectly balanced (IR = 1). In all other cases, balanced accuracy provides a more robust and fair assessment of classifier performance.

To provide a more complete view, additional metrics such as precision, recall, and F1-score are reported in the Appendix section for all classifiers and datasets.

3.1 *Performance across medical datasets*

To ensure the reliability and robustness of the results, each experiment was repeated 10 times per dataset. The ARS feature selection process and subsequent classification were independently executed in each run. The values reported in Table 2 and 3 correspond to the average balanced accuracy obtained across the 10 executions.

ARS achieved the best average balanced accuracy performance in 4 of the eight medical datasets in Tables 4-11, including Nutt, Gene Expression RNA Sequence, SEMG Hand Movements, Toxicity and Period Changer.

Table 2. Performance of average balanced accuracy in 10 executions of ARS (Part 1)

Dataset	Darwin	Gastrointestinal	Nutt	Parkinson's	Period Changer	Toxicity
Random search	0.8126	0.7049	**1.0000**	0.6582	**0.6719**	**0.6237**

Table 3. Performance of average balanced accuracy in 10 executions of ARS (Part 2)

Dataset	Gene Expression	SEMG
Random search	**1.0000**	0.6316

Table 4. Performance of balanced accuracy of state-of-the-art algorithms with stratified hold out (Part 1)

Dataset	Darwin	Gastrointestinal	Nutt	Parkinson's	Period Changer	Toxicity
3NN	0.7190	0.6667	0.8333	0.5738	0.6167	0.6186
5NN	0.7190	0.6667	0.8333	0.6048	0.6583	0.5039
Euclidean	0.5000	0.4667	0.8333	0.6524	0.5000	0.6000
J48	0.7745	0.7000	0.6667	0.7110	0.5333	0.4190
Logistic Regression	0.7173	0.7333	0.6667	0.5262	0.5750	0.5237
MLP	0.6879	0.7667	0.6667	0.5000	0.3750	0.5692
Naïve Bayes	**0.8873**	0.5667	1.0000	0.6136	0.5417	0.5395
Random Forest	0.8333	0.8667	0.8333	**0.8245**	0.4750	0.5020
SVM	0.5588	0.4667	1.0000	0.5568	0.5000	0.5000

Source: [12]

Table 5. Performance of balanced accuracy of state-of-the-art algorithms with stratified hold out (Part 2)

Dataset	Gene Expression	SEMG
3NN	0.9950	0.5902
5NN	0.9950	0.5729
Euclidean	0.9950	0.5937
J48	0.9916	0.8090
Logistic Regression	0.9950	0.5347
MLP	1.0000	0.6215
Naïve Bayes	0.8366	0.7083
Random Forest	1.0000	**0.8923**
SVM	1.0000	0.6250

Table 6. Performance of balanced accuracy of state-of-the-art algorithms with LOOCV (Part 1)

Dataset	Darwin	Gastrointestinal	Nutt	Parkinson's	Period Changer	Toxicity
3NN	0.681618	0.493421	0.678571	0.612266	0.551374	0.502506
5NN	0.698663	0.5	0.678571	0.584964	0.573178	0.551222
Euclidean	0.796687	0.703947	0.821429	0.593508	0.520687	0.50047
J48	0.76642	0.855263	0.75	0.535014	0.531017	0.506163
Logistic Regression	0.807708	0.855263	0.464286	0.574446	0.486774	0.440789
MLP	0.84885	0.802632	0.428571	0.57418	0.555	0.478788
Naive Bayes	0.843382	0.763158	0.785714	0.57097	0.563919	0.484649
Random Forest	0.872794	**0.901316**	0.892857	0.542797	0.568399	0.509868
SVM	0.855348	0.703947	0.928571	0.512956	0.574074	0.486842

Source: [12]

Table 7. Performance of balanced accuracy of state-of-the-art algorithms with LOOCV (Part 2)

Dataset	Gene Expression	SEMG
3NN	0.998004	0.581944
5NN	0.9963	0.573
Euclidean	0.9767	0.578
J48	0.9857	0.816
Logistic Regression	0.9937	0.568
MLP	0.9957	0.628
Naive Bayes	0.846	0.67
Random Forest	0.999	0.868
SVM	0.995	0.61

Source: [12]

Table 8. Performance of balanced accuracy of state-of-the-art algorithms with KFCV5 (Part 1)

Dataset	Darwin	Gastrointestinal	Nutt	Parkinson's	Period Changer	Toxicity
3NN	0.687582	0.525	0.633333	0.611468	0.448205	0.542622
5NN	0.687255	0.560417	0.65	0.576314	0.532949	0.514196
Euclidean	0.792688	0.717917	0.8	0.600666	0.583077	0.509486
J48	0.79281	0.843333	0.55	0.565553	0.528462	0.443102
Logistic Regression	0.813521	0.849167	0.45	0.550198	0.500256	0.490975
MLP	0.842974	0.79625	0.466667	0.598954	0.477179	0.496594
Naive Bayes	0.860131	0.763333	0.6	0.578648	0.437692	0.52915
Random Forest	0.830556	0.8825	0.733333	0.534014	0.514872	0.492918
SVM	0.86732	0.698333	0.966667	0.506018	0.536667	0.491304

Source: [12]

Table 9. Performance of balanced accuracy of state-of-the-art algorithms with KFCV5 (Part 2)

Dataset	Gene Expression	SEMG
3NN	0.996343	0.579167
5NN	0.996343	0.570139
Euclidean	0.976686	0.579861
J48	0.985667	0.8125
Logistic Regression	0.993677	0.563194
MLP	0.995667	0.629861
Naive Bayes	0.817726	0.667361
Random Forest	0.999	0.860417
SVM	0.995	0.609722

Source: [12]

Table 10. Performance of balanced accuracy of state-of-the-art algorithms with KFCV10 (Part 1)

Dataset	Darwin	Gastrointestinal	Nutt	Parkinson's	Period Changer	Toxicity
3NN	0.68125	0.532143	0.725	0.617187	0.527381	0.558182
5NN	0.686111	0.558929	0.7	0.583606	0.567857	0.540303
Euclidean	0.791468	0.71875	0.825	0.600054	0.502381	0.501439
J48	0.738889	0.832143	0.625	0.544201	0.442381	0.530455
Logistic Regression	0.836409	0.875893	0.5	0.546729	0.517857	0.434545
MLP	0.855357	0.778571	0.375	0.590022	0.434788	0.479848
Naive Bayes	0.843056	0.766071	0.775	0.567982	0.50119	0.502803
Random Forest	0.879167	0.883929	0.85	0.537152	0.566667	0.517424
SVM	0.861111	0.761607	0.975	0.516341	0.541667	0.487121

Source: [12]

Table 11. Performance of balanced accuracy of state-of-the-art algorithms with KFCV10 (Part 2)

Dataset	Gene Expression	SEMG
3NN	0.996352941	0.580555556
5NN	0.996352941	0.570833333
Euclidean	0.976019608	0.576388889
J48	0.982666667	0.815972222
Logistic Regression	0.993686275	0.563888889
MLP	0.994352941	0.625694444
Naive Bayes	0.846058824	0.669444444
Random Forest	0.999	0.865277778
SVM	0.994	0.60625

Source: [12]

Key Observations:

1. ARS obtained 100% balanced accuracy on Nutt dataset and Gene Expression RNA Sequence dataset.
2. While the performance on Period Changer and Toxicity may be considered low (67% and 62%), it is still better than that of the other algorithms.
3. On Darwin datasets ARS obtained 3rd place (81% of balanced accuracy), losing against Naive Bayes (88%) and Random Forest (83%).
4. On Parkinson's Disease dataset ARS also obtained 3rd place (65%), losing against decision trees-based algorithms Random Forest (82%) and J48 (81%).
5. On Gastrointestinal dataset ARS obtained 5th place (70%), losing against robust algorithms like Random Forest (86% in Stratified Hold Out and 90% in Leave One Out Cross Validation), MLP (76%) and Logistic Regression (73%).

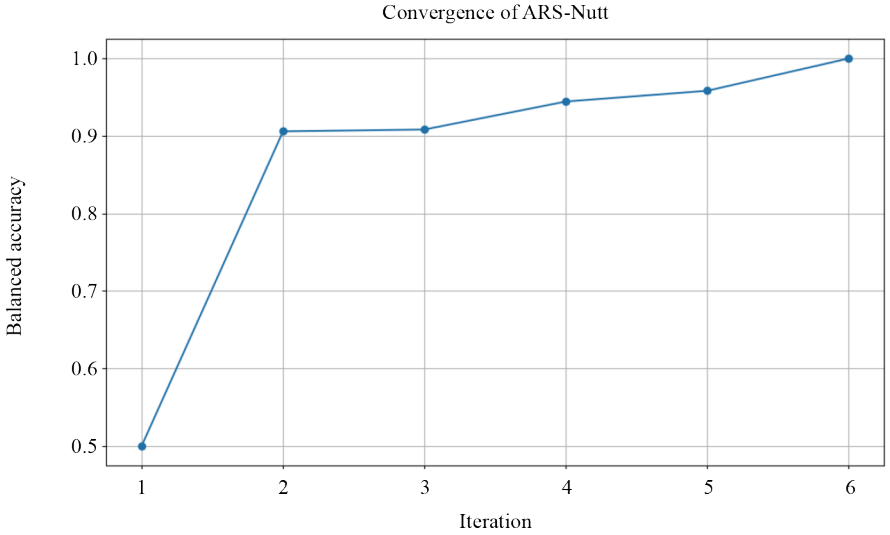

Figure 1. Convergence of balanced accuracy in Adaptive Random Search with Nutt dataset

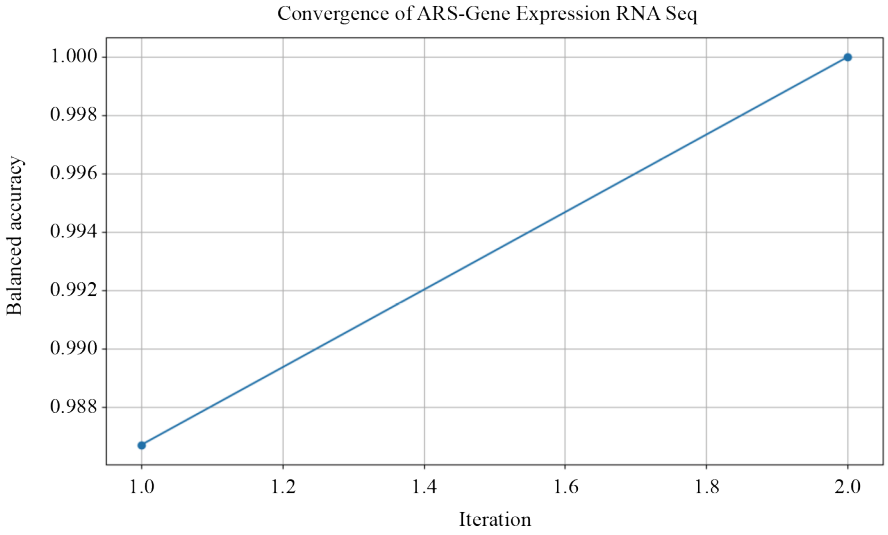

Figure 2. Convergence of balanced accuracy in Adaptive Random Search with Gene Expression RNA Sequence dataset

The perfect balanced accuracy observed in the Nutt dataset (100%) and Gene Expression RNA Sequence is supported by consistent convergence across runs and convergence plots that show progressive performance stabilization. These results are not due to overfitting, but rather to strong class separability, as confirmed by the ARS optimization trajectory (see Figures 1 and 2).

On the Nutt dataset, this result can be attributed to the strong separation between the two glioma subtypes (anaplastic and glioblastoma multiforme). A feature analysis revealed the presence of highly discriminative patterns that ARS consistently identified across multiple runs. Additionally, the small size and clean structure of the dataset make it especially amenable to the minimalist selection strategy of ARS, which avoids overfitting by working with compact, representative subsets.

3.2 *Exceptional cases of perfect accuracy*

In the Nutt Gene Expression RNA Sequence datasets [1, 10], ARS consistently achieved 100% balanced accuracy across all repetitions and validation strategies. This is attributed to the strong discriminability between glioma subtypes, allowing ARS to identify highly representative training subsets. Figure 1 and Figure 2 illustrate how the algorithm converges rapidly toward this optimal result.

It is worth noting that some benchmark classifiers, including Naïve Bayes and SVM, also reached perfect performance in specific configurations on Nutt and Gene Expression RNA Sequence datasets. However, such results were not consistent across all validation settings. ARS outperformed other methods by maintaining stability and precision in all cases.

3.3 *Precision in medical decision-making*

Precision, the ratio of true positives to all predicted positives, is paramount in clinical settings to avoid harmful interventions. Table 12 shows the precision values obtained with ARS in the eighth medical datasets.

As we can see, only in the dataset toxicity ARS fail to achieve a good score in prediction, having more non-toxic molecules classified as toxic.

Table 12. ARS precision values

Dataset	Precision
Nutt (Gliomas)	1.0000
Parkinson's	0.8366
Gastrointestinal	0.7660
Toxicity	0.4894
Darwin	0.8667
Period Changer	0.8039
Gene Expression	1.0000
SEMG	0.6934

3.4 *Computational efficiency of metaheuristic search*

ARS's metaheuristic approach avoids the combinatorial explosion of brute-force methods. For a dataset with D samples, brute-force evaluation of all possible E training subsets requires a high amount of iterations as we can see in the following equation:

$$\text{Combination}\binom{N}{E*N} \tag{5}$$

Being this computationally infeasible for medical datasets. For example: for the Nutt dataset ($D = 28$) and $E = 20\%D$, brute force would need Combination$\binom{28}{0.2*28} = 376{,}740$ iterations. ARS achieved convergence in 5 iterations, reducing computation time by 99.9%.

4. Discussion

4.1 *Simplicity through minimalist training subsets*

At its core, ARS adheres to the minimalist principle of achieving maximal accuracy with minimal data. By optimizing the selection of small training subsets (20% of the dataset in the experiments), ARS challenges the conventional wisdom that larger training sets inherently yield better performance. For example, on the Nutt glioma dataset, ARS achieved 100% balanced accuracy using only six training samples (20% of 28), whereas SVM required the full 80% training split (22 samples) to reach similar balanced accuracy. This efficiency stems from ARS's ability to identify statistically representative patterns that capture the essential feature-pattern relationships, even in highly rectangular data. Unlike traditional methods that rely on exhaustive training, ARS's minimalist approach reduces computational overhead without sacrificing performance, a critical advantage in medical applications where data acquisition is costly or ethically constrained.

To reinforce the experimental validity of our proposal and to address concerns regarding reproducibility and black-box behavior, we expanded our protocol in two key directions. First, we included three additional biomedical datasets, diversifying the range of pathologies under study. Second, for each dataset, we executed the ARS-based feature selection process ten times independently, averaging the results to reduce variability and expose consistent patterns.

The use of stratified cross-validation strategies (LOOCV, K-Fold with $k = 5$ and $k = 10$) further strengthens the reliability of the findings. These validations were applied uniformly across all algorithms to ensure fair comparison. The consistent superiority or competitive performance of ARS in these diverse scenarios confirms its capacity to generalize.

Additionally, balanced accuracy was adopted as the primary evaluation metric to mitigate biases introduced by class imbalance—frequent in real-world clinical data. The convergence plots and the low number of iterations required by ARS (often fewer than 6) illustrate that the method is not only effective but also computationally efficient and interpretable. This aligns with recent ensemble-based feature selection strategies that aim to enhance robustness and interpretability in clinical tasks [14].

4.2 *Clinical relevance of precision-driven models*

The medical community prioritizes precision to avoid harmful interventions. In this context, ARS is intended as a decision-support tool, helping to reduce risks such as unnecessary chemotherapy or misdiagnosis by identifying highly discriminative training patterns. It complements—rather than replaces—clinical judgment. ARS's superior precision across datasets, particularly 100% on gliomas and 100% on Gene Expressions RNA Sequence for Cancer, demonstrates its potential as a decision-support tool. For instance, in glioma classification, ARS eliminated false positives by selecting the most representative patterns in the training subset. This aligns with pathologists' reliance on key biomarkers rather than exhaustive genomic data, bridging the gap between machine learning and clinical intuition.

4.3 *Current limitations: Binary classification focus*

A key limitation of ARS is its current restriction to binary classification. While this suffices for datasets like Nutt (two glioma subtypes) or Toxicity (toxic/non-toxic), many medical problems require multiclass discrimination. Future work will explore hierarchical ARS architectures to address this challenge.

4.4 *Metaheuristics vs. brute-force: A computational trade-off*

ARS's metaheuristic design circumvents the infeasibility of brute-force search in medical datasets. For example, brute-force evaluation of all possible 20% training subsets in the Darwin dataset (118 samples) would require Combination $\binom{118}{0.2 * 118} \approx 2.3 \times 10^{23}$ iterations. ARS achieved comparable results in 5,000 iterations by prioritizing exploration of high-accuracy regions and locally refining candidates. While this introduces a small risk of suboptimal convergence, our results show no significant accuracy trade-offs across tested datasets. Similar approaches combining genetic algorithms and subset selection have demonstrated strong results in clinical severity scoring systems [15].

4.5 *Implications for MML*

ARS operationalizes the MML paradigm by proving that simplicity and transparency need not compromise performance. Selecting the smallest training subset rivaled complex models like SVM and MLP that needed higher training data to achieve the same balanced accuracy. This aligns with growing clinical demand for explainable AI, where "how" a decision is made matters as much as its accuracy [4].

5. Conclusion

The Adaptive Random Search (ARS) framework demonstrates that Minimalist Machine Learning (MML) principles (simplicity, transparency, and efficiency) can address the unique challenges of rectangular medical datasets, where features vastly outnumber samples. By optimizing stratified holdout validation through metaheuristic search. ARS achieved state-of-the-art performance across six medical domains, including 100% balanced accuracy and precision on the Nutt glioma dataset and Gene Expression RNA Sequence dataset, while reducing computational costs by over 99% compared to brute-force approaches.

Key contributions of this work include:

1. A Novel Validation Paradigm: ARS redefines holdout validation as an optimization problem, enabling robust performance estimation in data-scarce scenarios.

2. Clinical Trust Through Transparency: The small training subset used by ARS, aligns with clinicians' diagnostic workflows, fostering trust in AI-driven tools.

3. Resource Efficiency: ARS's metaheuristic core ensures scalability, making it viable for resource-constrained medical environments.

While currently limited to binary classification, ARS's modular design provides a foundation for future extensions to multiclass problems, very usual in clinical datasets. Furthermore, its institutional roots in the Intelligent Computing Laboratory (CIC-IPN), the birthplace of MML, underscore its alignment with a broader scientific mission to democratize interpretable AI in healthcare.

In an era where medical datasets grow increasingly complex yet remain critically small, ARS stands as a testament to the power of minimalist design. By proving that simplicity need not compromise efficacy, this work paves the way for a new generation of clinical tools where accuracy and interpretability coexist.

ARS is not merely an algorithm, but a paradigm shift, one that prioritizes clarity over complexity and pragmatism over pretension. As healthcare embraces AI, frameworks like ARS will be vital to ensuring that technology serves clinicians, not confounds them.

Acknowledgments

This research was not the recipient of any specific grants from funding agencies in the public, commercial or non-profit sectors. We would like to extend our gratitude to the Centro de Investigación en Computación del Instituto Politécnico Nacional and the Consejo Nacional de Humanidades, Ciencia y Tecnología (CONAHCYT) for their invaluable collaboration and support throughout the development of this research.

Conflict of interest

The authors declare no conflicts of interest.

References

[1] Nutt CL, Mani DR, Betensky RA, Tamayo P, Cairncross JG, Ladd C, et al. Gene expression-based classification of malignant gliomas. *Cancer Research*. 2003; 63(7): 1602-1607.

[2] Sakar C, Serbes G, Gunduz A, Nizam H, Sakar B. *Parkinson's disease classification*. UCI Machine Learning Repository. 2018. Available from: https://archive.ics.uci.edu/dataset/470/parkinson+s+disease+classification [Accessed 5th June 2025].

[3] Adadi A, Berrada M. Peeking inside the black-box: A survey on explainable AI. *IEEE Access*. 2018; 6: 52138-52160. Available from: https://doi.org/10.1109/ACCESS.2018.2870052.

[4] Yáñez-Márquez C. Toward the bleaching of the black boxes: Minimalist machine learning. *IT Professional*. 2020; 22(4): 51-56. Available from: https://doi.org/10.1109/MITP.2020.2994188.

[5] Ultsch A, Lötsch J, Thrun MC. An explainable AI system for the diagnosis of high-dimensional biomedical data. *BioMedInformatics*. 2024; 4(1): 197-218. Available from: https://doi.org/10.3390/biomedinformatics4010013.

[6] Coons SW, Johnson PC, Scheithauer BW, Yates AJ, Pearl DK. Improving diagnostic accuracy in neuropathology. *Cancer*. 1997; 79(7): 1381-1393. Available from: https://doi.org/10.1002/(SICI)1097-0142(19970401)79:7<1381::AID-CNCR16>3.0.CO;2-W.

[7] Fontanella F. *Darwin*. UCI Machine Learning Repository. 2022. Available from: https://archive.ics.uci.edu/ml/datasets/Darwin [Accessed 5th June 2025].

[8] Gül F, Rahim S. *Period Changer*. UCI Machine Learning Repository. 2021. Available from: https://archive.ics.uci.edu/dataset/729/period+changer-2 [Accessed 5th June 2025].

[9] Gül F, Rahim S. *Toxicity*. UCI Machine Learning Repository. 2021. Available from: https://archive.ics.uci.edu/dataset/728/toxicity-2 [Accessed 5th June 2025].

[10] Fiorini S. *Gene expression cancer RNA-Seq*. UCI Machine Learning Repository. 2016. Available from: https://archive.ics.uci.edu/dataset/401/gene+expression+cancer+rna+seq [Accessed 5th June 2025].

[11] Sapsanis C, Tzes A, Georgoulas G. *SEMG for basic hand movements*. UCI Machine Learning Repository. 2013. Available from: https://archive.ics.uci.edu/ml/datasets/sEMG+for+Basic+Hand+movements [Accessed 5th June 2025].

[12] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*. 2011; 12: 2825-2830.

[13] Mesejo P, Pizarro D. *Gastrointestinal lesions in regular colonoscopy*. UCI Machine Learning Repository. 2016. Available from: https://archive.ics.uci.edu/ml/datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy [Accessed 5th June 2025].

[14] Natarajan K, Baskaran D, Kamalanathan S. An adaptive ensemble feature selection technique for model-agnostic diabetes prediction. *Scientific Reports*. 2025; 15: 6907. Available from: https://doi.org/10.1038/s41598-025-91282-8.

[15] Kong W, Zhu J, Bi S, Huang L, Wu P, Zhu S. Adaptive best subset selection algorithm and genetic algorithm aided ensemble learning method identified a robust severity score of COVID-19 patients. *iMeta*. 2023; 2(1): e72. Available from: https://doi.org/10.1002/imt2.126.

Appendix

Here we include the results obtained in precision, recall and F1-score in Tables 13-27.

Table 13. Precision in ARS

Dataset	Darwin	Gastrointestinal	Nutt	Parkinson's	Period Changer	Toxicity	Gene Expression	SEMG
ARS	0.89643	0.709527	1	0.832547	0.858132	0.458321	1	0.693408

Table 14. Precision in state-of-the-art algorithms with stratified hold out

Dataset	Darwin	Gastrointestinal	Nutt	Parkinson's	Period Changer	Toxicity	Gene Expression	SEMG
3NN	1	0.588235	0.75	0.806723	0.818182	0.4	0.983607	0.756098
5NN	1	0.466667	0.75	0.782946	0.846154	0.5	0.983607	0.825
Euclidean	0.8	0.642857	0.75	0.833333	1	0.3125	0.983607	0.675325
J48	0.933333	0.769231	1	0.769231	0.818182	0.25	1	0.856
Logistic Regression	0.8125	0.846154	0.333333	0.79	0.846154	0.25	0.983607	0.54386
MLP	0.875	0.785714	0.666667	0.814159	0.8	0.166667	1	0.652174
Naive Bayes	0.842105	0.6875	1	0.818182	0.833333	0.32	0.88	0.819149
Random Forest	0.875	0.818182	0.75	0.767606	0.785714	0.5	1	0.855346
SVM	0.833333	0.6	1	0.75	0.75	0	1	0.663636

Table 15. Precision in state-of-the-art algorithms with LOOCV

Dataset	Darwin	Gastrointestinal	Nutt	Parkinson's	Period Changer	Toxicity	Gene Expression	SEMG
3NN	0.970588	0.493976	0.608696	0.797101	0.724638	0.333333	0.993377	0.780952
5NN	0.972973	0.5	0.608696	0.782148	0.733333	0.413043	0.9935	0.77
Euclidean	0.835443	0.706667	0.764706	0.812808	0.741379	0.329787	0.983	0.61
J48	0.783133	0.8375	0.769231	0.765217	0.724138	0.338462	0.9805	0.835
Logistic Regression	0.831325	0.855263	0.473684	0.784029	0.704918	0.25	0.9901	0.574
MLP	0.878049	0.7875	0.428571	0.781302	0.737705	0.296296	0.9966	0.638
Naive Bayes	0.827957	0.777778	1	0.778878	0.774194	0.320611	0.915	0.77
Random Forest	0.875	0.942029	0.866667	0.762857	0.728395	0.368421	0.9968	0.82
SVM	0.853933	0.71831	0.875	0.750336	0.729412	0	1	0.638

Table 16. Precision in state-of-the-art algorithms with KFCV5

Dataset	Darwin	Gastrointestinal	Nutt	Parkinson's	Period Changer	Toxicity	Gene Expression	SEMG
3NN	0.977778	0.512418	0.6	0.796583	0.670101	0.336667	0.993443	0.778477
5NN	0.98	0.553987	0.603333	0.778544	0.712353	0.311111	0.993443	0.767398
Euclidean	0.839649	0.730275	0.8	0.816796	0.784167	0.336429	0.983078	0.616318
J48	0.815163	0.826162	0.633333	0.781358	0.716923	0.250808	0.980538	0.832126
Logistic Regression	0.847417	0.874083	0.533333	0.771946	0.700023	0.312902	0.990053	0.571625
MLP	0.861709	0.792801	0.48	0.794834	0.699068	0.309762	0.99661	0.642813
Naive Bayes	0.83738	0.775873	0.875	0.783661	0.646234	0.348551	0.915171	0.763989
Random Forest	0.831516	0.945518	0.766667	0.759395	0.703915	0.365	0.996721	0.814433
SVM	0.873555	0.724296	0.95	0.747664	0.713639	0	1	0.6406

Source: [12]

Table 17. Precision in state-of-the-art algorithms with KFCV10

Dataset	Darwin	Gastrointestinal	Nutt	Parkinson's	Period Changer	Toxicity	Gene Expression	SEMG
3NN	0.98	0.526569	0.716667	0.799318	0.716548	0.415476	0.993548	0.771844
5NN	0.98	0.560317	0.7	0.782206	0.732738	0.401905	0.993548	0.773148
Euclidean	0.838889	0.729167	0.851852	0.817136	0.724643	0.32934	0.986511	0.605781
J48	0.787056	0.83119	0.722222	0.769601	0.66131	0.365119	0.972058	0.83306
Logistic Regression	0.866061	0.902222	0.481481	0.769531	0.723452	0.255238	0.9901	0.571847
MLP	0.882778	0.768214	0.333333	0.790255	0.695714	0.2775	0.986989	0.634603
Naive Bayes	0.81216	0.772817	1	0.778459	0.726667	0.333806	0.924944	0.771174
Random Forest	0.905253	0.938492	0.87037	0.760823	0.747421	0.3125	0.996774	0.817742
SVM	0.875556	0.787381	0.966667	0.751739	0.716825	0	0.996552	0.634921

Source: [12]

Table 18. Recall in ARS

Dataset	Darwin	Gastrointestinal	Nutt	Parkinson's	Period Changer	Toxicity	Gene Expression	SEMG
ARS	0.829262	0.669466	1	0.772386	0.755849	0.584013	1	0.463735

Table 19. Recall in state-of-the-art algorithms with stratified hold out

Dataset	Darwin	Gastrointestinal	Nutt	Parkinson's	Period Changer	Toxicity	Gene Expression	SEMG
3NN	0.333333	0.666667	1	0.857143	0.75	0.363636	1	0.215278
5NN	0.277778	0.466667	1	0.901786	0.916667	0.454545	1	0.229167
Euclidean	0.666667	0.6	1	0.630631	0.916667	0.454545	1	0.361111
J48	0.823529	0.666667	0.333333	0.714286	0.75	0.363636	0.983333	0.743056
Logistic Regression	0.722222	0.733333	0.333333	0.705357	0.916667	0.272727	1	0.430556
MLP	0.777778	0.733333	0.666667	0.821429	0.8	0.181818	1	0.520833
Naive Bayes	0.888889	0.733333	0.333333	0.803571	0.416667	0.727273	0.733333	0.534722
Random Forest	0.777778	0.6	1	0.973214	0.916667	0.272727	1	0.944444
SVM	0.833333	0.6	1	0.991071	1	0	1	0.506944

Source: [12]

Table 20. Recall in state-of-the-art algorithms with LOOCV

Dataset	Darwin	Gastrointestinal	Nutt	Parkinson's	Period Changer	Toxicity	Gene Expression	SEMG
3NN	0.375	0.539474	1	0.880783	0.806452	0.285714	1	0.227778
5NN	0.409091	0.552632	1	0.919929	0.887097	0.339286	0.9967	0.205
Euclidean	0.75	0.697368	0.928571	0.597826	0.693548	0.553571	0.9633	0.445
J48	0.747126	0.881579	0.714286	0.784314	0.677419	0.392857	0.9833	0.795
Logistic Regression	0.784091	0.855263	0.642857	0.768683	0.693548	0.25	0.9933	0.51
MLP	0.818182	0.828947	0.428571	0.834225	0.75	0.290909	0.9933	0.585
Naive Bayes	0.875	0.736842	0.571429	0.839858	0.387097	0.75	0.73	0.485
Random Forest	0.875	0.855263	0.928571	0.950178	0.951613	0.125	1	0.938
SVM	0.863636	0.671053	1	0.994662	1	0	0.99	0.505

Source: [12]

Table 21. Recall in state-of-the-art algorithms with KFCV5

Dataset	Darwin	Gastrointestinal	Nutt	Parkinson's	Period Changer	Toxicity	Gene Expression	SEMG
3NN	0.386928	0.563333	1	0.884339	0.789744	0.339394	0.996667	0.220833
5NN	0.386275	0.58	1	0.912816	0.919231	0.24697	0.996667	0.2
Euclidean	0.740523	0.698333	0.933333	0.616713	0.696154	0.536364	0.963333	0.438889
J48	0.785621	0.870833	0.433333	0.763306	0.64359	0.230303	0.983333	0.783333
Logistic Regression	0.78366	0.816667	0.633333	0.750869	0.770513	0.306061	0.993333	0.5
MLP	0.830065	0.803333	0.666667	0.839918	0.707692	0.291212	0.993333	0.586111
Naive Bayes	0.908497	0.749167	0.3	0.839886	0.415385	0.768182	0.673333	0.483333
Random Forest	0.852288	0.8175	0.733333	0.943063	0.889744	0.107576	1	0.934722
SVM	0.864052	0.658333	1	0.991119	1	0	0.99	0.501389

Source: [12]

Table 22. Recall in state-of-the-art algorithms with KFCV10

Dataset	Darwin	Gastrointestinal	Nutt	Parkinson's	Period Changer	Toxicity	Gene Expression	SEMG
3NN	0.375	0.580357	1	0.891479	0.804762	0.38	0.996667	0.226389
5NN	0.384722	0.592857	1	0.923528	0.919048	0.32	0.996667	0.197222
Euclidean	0.738889	0.7	0.9	0.609259	0.721429	0.556667	0.96	0.452778
J48	0.680556	0.844643	0.6	0.766823	0.684762	0.436667	0.983333	0.791667
Logistic Regression	0.829167	0.867857	0.6	0.743985	0.752381	0.243333	0.993333	0.506944
MLP	0.841667	0.807143	0.4	0.848552	0.721429	0.273333	0.996667	0.5875
Naive Bayes	0.909722	0.766071	0.55	0.839912	0.385714	0.753333	0.73	0.480556
Random Forest	0.875	0.832143	0.9	0.939568	0.966667	0.15	1	0.941667
SVM	0.863889	0.717857	1	0.991103	1	0	0.99	0.501389

Source: [12]

Table 23. F1-score in ARS

Dataset	Darwin	Gastrointestinal	Nutt	Parkinson's	Period Changer	Toxicity	Gene Expression	SEMG
ARS	0.860181	0.683058	1	0.800772	0.798778	0.510936	1	0.554642

Table 24. F1-score in state-of-the-art algorithms with stratified hold out

Dataset	Darwin	Gastrointestinal	Nutt	Parkinson's	Period Changer	Toxicity	Gene Expression	SEMG
3NN	0.5	0.625	0.857143	0.831169	0.782609	0.380952	0.991736	0.335135
5NN	0.434783	0.466667	0.857143	0.838174	0.88	0.47619	0.991736	0.358696
Euclidean	0.727273	0.62069	0.857143	0.717949	0.956522	0.37037	0.991736	0.470588
J48	0.875	0.714286	0.5	0.740741	0.782609	0.296296	0.991597	0.795539
Logistic Regression	0.764706	0.785714	0.333333	0.745283	0.88	0.26087	0.991736	0.48062
MLP	0.823529	0.758621	0.666667	0.817778	0.8	0.173913	1	0.579151
Naive Bayes	0.864865	0.709677	0.5	0.810811	0.555556	0.444444	0.8	0.647059
Random Forest	0.823529	0.692308	0.857143	0.858268	0.846154	0.352941	1	0.89769
SVM	0.833333	0.6	1	0.853846	0.857143	INDETERMINATE	1	0.574803

Source: [12]

Table 25. F1-score in state-of-the-art algorithms with LOOCV

Dataset	Darwin	Gastrointestinal	Nutt	Parkinson's	Period Changer	Toxicity	Gene Expression	SEMG
3NN	1	1	1	1	1	1	1	1
5NN	1	1	1	1	1	1	0.995	0.32
Euclidean	1	1	1	1	1	1	0.9728	0.512
J48	1	1	1	1	1	1	0.9818	0.814
Logistic Regression	1	1	1	1	1	1	0.9916	0.538
MLP	1	1	1	1	1	1	0.995	0.61
Naive Bayes	1	1	1	1	1	1	0.812	0.595
Random Forest	1	1	1	1	1	1	0.9984	0.875
SVM	1	1	1	1	1	INDETERMINATE	0.995	0.562

Source: [12]

Table 26. F1-score in state-of-the-art algorithms with KFCV5

Dataset	Darwin	Gastrointestinal	Nutt	Parkinson's	Period Changer	Toxicity	Gene Expression	SEMG
3NN	0.551852	0.531803	0.742857	0.838032	0.724576	0.338025	0.995014	0.34329
5NN	0.548542	0.564192	0.748095	0.840302	0.80254	0.275354	0.995014	0.316591
Euclidean	0.776166	0.713106	0.83619	0.702485	0.725645	0.413038	0.972787	0.50965
J48	0.794649	0.845833	0.514583	0.772056	0.676923	0.237728	0.98176	0.806435
Logistic Regression	0.809794	0.841243	0.513333	0.760782	0.725291	0.3069	0.991652	0.532957
MLP	0.842808	0.796861	0.547619	0.816604	0.70193	0.29746	0.994958	0.612236
Naive Bayes	0.868957	0.75994	0.5	0.810215	0.497224	0.479092	0.774983	0.59139
Random Forest	0.835527	0.869667	0.740952	0.841282	0.784474	0.166175	0.998347	0.870084
SVM	0.868098	0.682143	0.971429	0.852343	0.832705	INDETERMINATE	0.994929	0.562418

Source: [12]

Table 27. F1-score in state-of-the-art algorithms with KFCV10

Dataset	Darwin	Gastrointestinal	Nutt	Parkinson's	Period Changer	Toxicity	Gene Expression	SEMG
3NN	0.528944	0.547023	0.813333	0.84271	0.754362	0.396947	0.995026	0.348099
5NN	0.534693	0.573516	0.8	0.846704	0.812271	0.34688	0.995026	0.312652
Euclidean	0.775677	0.712848	0.903704	0.697079	0.712391	0.408416	0.972388	0.515629
J48	0.712929	0.833091	0.655462	0.766427	0.665958	0.396014	0.977083	0.811453
Logistic Regression	0.83759	0.873418	0.534247	0.756029	0.727391	0.249144	0.991637	0.535968
MLP	0.857187	0.780982	0.363636	0.817648	0.697854	0.275401	0.991749	0.608608
Naive Bayes	0.855341	0.765665	0.944444	0.807441	0.479733	0.460459	0.811874	0.590529
Random Forest	0.880431	0.868299	0.918519	0.840683	0.839588	0.202703	0.998361	0.874933
SVM	0.864408	0.743445	0.98	0.854959	0.833589	INDETERMINATE	0.99322	0.559608

Source: [12]