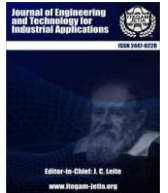




ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.12 n.59, p.148-155. May/June, 2026.

DOI: <https://doi.org/10.5935/jetia.v12i59.3348>

RESEARCH ARTICLE

OPEN ACCESS

INVESTIGATING BIO-INSPIRED FEATURE SELECTION FOR AN ENHANCED ENSEMBLE-BASED INTRUSION DETECTION SYSTEM

Souad Guessoum^{*1}, Oussama Kadri² and Nabih Azizi³

^{1,2,3}Department of Computer Science, Badji Mokhtar-Annaba University, P.O BOX 12 - 23000, Annaba, Algeria.

¹<http://orcid.org/0009-0006-7069-6682>, ²<http://orcid.org/0009-0009-6582-0863>, ³<http://orcid.org/0000-0002-2713-4792>

Email: ^{*}souad.guessoum@univ-annaba.dz, ²oussama.kadri07@gmail.com, ³nabiha.azizi@univ-annaba.dz

ARTICLE INFO

Article History

Recebido: 19 de janeiro de 2026

Revisado: 21 de Março de 2026

Aceito: 21 de abril de 2026

Publicado: 29 de maio de 2026

Keywords:

Intrusion Detection,

Feature Selection,

Bio-inspired Algorithm,

Decision Tree,

Logistic Regression.

ABSTRACT

An intrusion in a network traffic is a serious attack on the security of systems, which can sometimes collapse if their data is not protected enough or if the intrusion is not detected at the right time. The network traffic state is commonly represented in a detection intrusion system (IDS) by datasets with a wide range of parameters, which can considerably slow down the system's response time, and thus preventing detection at the right moment to deal with the attack. Hence the need to integrate a feature selection engineering module into any effective IDS. The current work revolves around the selection of the most relevant and representative characteristics of the network to be able to detect the intrusion with the best precision while minimizing the system computation time. An investigation of four bio-inspired models is made to extract the best subset according to each one, i.e. genetic algorithm, gray wolf algorithm, bat algorithm, and particle swarm optimizer. Results are gathered in a pool where two strategies are developed to decide on the final structure of the subset, which is validated by a classification process. Two strong classifiers namely logistic regression and decision tree, are used in an individual way, and then in fusion approach to benefit from the strengths of several machine learning models. The ensemble model showed that we can achieve highly accurate and reliable decisions (99,8%) while working only on reduced subsets.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Nowadays, humans are increasingly dependent on computer networks, whose ability to defend against intrusions must evolve almost continuously. Intrusion Detection Systems are more and more in demand, given the important protection they can offer to organizations. They aim to identify and report abnormal behavior that may indicate an attempt at unauthorized access, or the exploitation of system vulnerabilities for a given purpose. Their evolution depends on the techniques used for detection, and that's why we are increasingly finding Data Analysis or Artificial Intelligence (AI) applications dedicated to intrusion detection. Numerous researchers in the literature are focused on the investigation of accurate machine learning models and most informative features to generate powerful IDS. In the survey authored by [1], in-depth synthesis of the main studies on meta-heuristic approaches for building efficient IDS is described. It gives to readers an overview on advantages, challenges, and limitations of several works focused on detection intrusion field. In the context of traffic networks, most databases are generally high-dimensional, as the more data we have on the network state, the better the situation is represented. This dimensionality offers the advantage of reliable representation, but a lot of redundant and irrelevant features in the datasets poses a major drawback, mainly in terms of computation time and decision robustness. The quality of the IDS response is determined not only by accurate detection of the network status, but above all by the right timing, where the response time factor becomes very important. Indeed, an intrusion detected too late cannot be of significant use, and a Feature Selection (FS) module is becoming more than essential to integrate into any efficient IDS. This concept is a process used in machine learning models to filter the most relevant features to use in building more reliable models.

It is based on the idea of choosing an appropriate subset of features from a larger one, focusing on those that are most relevant and informative. This can be done to improve model performance, reduce computation time, or even to understand the correlation between the features and target variables which generally represent the diagnosis of the situation. The most commonly used methods for a relevant selection of features are those based on bio-inspired metaheuristics. We are particularly interested here in four families, namely Genetic Algorithms (GA), Gray Wolf Optimizer (GWO), Bat Algorithm (BA), and Particle Swarm Optimization Algorithm (PSO). These methods are based on the natural behavior of living beings, on principles of selection and reproduction. They are powerful approaches to solving optimization problems, such as FS by helping to find the best subset of primitives that maximize the model performance. An overview is provided in the following section, based on a selection of recent and interesting articles published in renowned journals and focused on the exploitation of these bio-inspired algorithms in FS field, (see section 2 for more details). Based on different domains and different studies, each metaheuristic mentioned above and used for FS has at least one field application where it provides the better feature subset than others, while there is no guarantee it still giving the best one when we change the domain. This is in accordance with the idea of the No-Free-Lunch (NFL) theorem, no algorithm, no metaheuristic is good enough to solve all optimization problems, or even the same problem however the dataset is or however the domain is. The main purpose of this work is to try answering to this question: how to build a high-performance model in the generalization task by taking advantage of the benefits of strengths and complementarities of several models.

This study serves three objectives. The first one is to investigate and conduct a comparative analysis of the behavior of four bio-inspired algorithms (GA, GWO, BA and PSO) in feature selection applied to intrusion detection. Each of these models will generate a specified vector of features, which basically corresponds to a specified reduced dataset. The aim is to select the best diversified subsets that ensure the optimal performance and so, our second contribution which intends to improve the final decision of the IDS, is to propose a pool of strategies for producing new vectors based on those generated by the algorithms, aiming to benefit from the robustness and diversities of all methods rather than favoring one at the expense of another. To this end, there are traditional fusion methods, our idea is to propose a new way of fusing the outputs of the models while integrating the notion of the importance of selected features. Each strategy which will be described in detail later, generates a new subset according to its principle. These subsets will be used to train two well-known classifiers (Logistic Regression and Decision Tree) resulting in the creation of four different models. Our third contribution is to leverage the power of these four models by merging them into a single model in which the decision takes into account a very important parameter, namely the F1-score of each model. This paper is organized as follows: Section 2 provides a review of the literature on the feature selection problem and the use of bio-inspired algorithms to solve it in several fields of study but particularly in intrusion detection. Section 3 describes our methodology to generate a high-performance IDS based on FS module supplemented by a merger pool of decision and a fusion classification process. The section also details the conflict resolution strategy in this fusion module. In Section 4, experimentation results are given and accompanied by an established discussion and interpretation. The article ends with a conclusion and some perspectives in the last section.

II. RELATED WORKS

This section focuses on very recent researches that have demonstrated the importance of FS step in each machine learning model to enhance the generated decision. Since their introduction by Holland [2], genetic algorithms have continued to spark significant research interest to this day. Their concept has attracted the attention of researchers, who every year produce numerous studies employing this technique to solve optimization problems. The authors Bohrer and Dorn presented recently in [3] a hybrid feature selection approach using a novel multi-objective GA to reduce dimensionality and enhance classification performance on several datasets from different domains. They explore and combine results of classical feature selection methods which allows to generate optimized feature sets in a process that adapts to different data conditions. Experimentation done shows significantly a positive impact of the approach in high-dimensional use cases, offering improved classification. The domain of brain computer interface is known by datasets of high dimensionality. Here, FS becomes more than indispensable to optimize the classification performance and improve interpretability. In the paper authored by [4] this area is addressed by combining a modified genetic algorithm and support vector machine classifier and applied to hybrid EEG-EMG and EEG-fNIRS brain-computer interface systems. Results show a significant improvement at the average accuracy in 2 big datasets related to these 2 systems. Imani et al address in the article [5], the issue of high dimensionality in the multi-view datasets, since this type of data offer diverse forms of complementary data to enhance the model performance, while it increases considerably the dimensionality. The work proposed an approach called the multi-view multi-objective feature selection genetic algorithm (MMFS-GA) which performs simultaneous selection on the best subset within a view and between views in a unified framework.

The proposed model has been evaluated on 9 datasets between synthetic and real data. Published results show a significant improvement over the best baseline methods. Aksu and Aydin propose in their work presented in [6] a new intrusion detection framework based on feature selection and classification. The process begins with a modified genetic algorithm for selecting the best feature subset based on k-fold cross validation. After, 5 classifiers are used in the classification phase, namely: support vector, logistic regression, decision tree, k-nearest neighbors and linear discriminant analysis classifiers. Experimentation is done on 3 datasets, namely the HCRL-car hacking dataset, UNSW-NB15, and CIC-IDS2017 datasets. Results show that the proposed modified GA combined with decision tree classifier present a better performance in terms of several metrics. Fang et al enriches the literature with another recent work [7], where authors propose a GA based FS method for industrial control systems (ICS). Redundant features are eliminated by a feature ranking fusion mechanism used in GA. They used also the growing tree clustering idea to enhance the global merit-seeking speed, and design a new fitness function for ICS characteristics. The model effectiveness is demonstrated on a real ICS dataset. In this study, we are also interested in 3 other bio-inspired algorithms namely: Grey Wolf Optimizer algorithm which was proposed in 2014 by Mirajili in [8], Bat algorithm proposed in 2010 by Xin-She Yang in [9] and Particle Swarm Optimizer algorithm proposed by Kennedy and Eberhart in [10] in 1995. All these optimization algorithms are inspired by the collective behavior of animals (wolves/ bats/birds...), they have been exploited in different ways since their appearance to solve many optimization problems.

In order not to overwhelm the reader of this paper, we provide in Table 1 an overview of a selection of these works recently published dedicated to feature selection issue or applied in intrusion detection field.

Table 1: Bio-inspired algorithms for the FS/Intrusion Detection.

n°	year	Reference	Issue/field	Datasets	Bio-Algorithm	Combined with
1	2024	[11]	Intrusion detection	KDD99 and UNSW15	GWO	multi-layer perceptron
2	2025	[12]	FS/ Alzheimer's diagnosis	EEG	PSO /GWO	
3	2025	[13]	FS/ Soil Moisture	Agriculture dataset	PSO	KNN/ DT/RT
4	2024	[14]	FS/Detection intrusion	UNSW-NB15	PSO/GA	
5	2024	[15]	FS/Epileptic Seizure	EEG data	BA/ GA	NN
6	2023	[16]	FS/ diabetic retinopathy diagnosis	Fundus retinal eye/ messidor datasets	BA	deep forest technique

Source: Authors, (2026).

We can also cite some recent and interesting works that has targeted the FS issue using other techniques to solve it. In 2023, Hu et al published a work presented in [17] and focused on feature selection on the dataset with incomplete data. Authors precise that most feature selection techniques work well only with the complete data, but show poorly when it comes to missing data especially with high missing rate. They proposed a novel FS method namely non-negative latent factor incorporated duplicate maximal information coefficient (NLF-DMIC), fulfilled by 3 steps. Experiments done on an artificially synthetic dataset and 8 real datasets show that it outperforms some existing feature selection methods. Liu et al presented a work in [18] where they introduce SHapley Additive exPlanations (SHAP) value for FS of Parkinson's disease dataset. This field is known by its high feature dimensionality. in this work, SHAP value is combined with four classifiers, namely deep forest, extreme gradient boosting, light gradient boosting machine and random forest. The magnitude of contribution of SHAP value to the features is calculated, after the process selects the features with significant contribution in the classification task, and then the selected subset is used as input to classify the Parkinson's disease. Priyatno et al published recently a work [19] focused on the feature selection problem and dedicated to dropout prediction in higher education. They proposed a step-by-step framework to improve prediction accuracy trough out a process composed of three stages, namely Variance Threshold, Mutual Information, and Boruta, which allows them to achieve an F1-score of 87.63%. Feature selection was also the subject of a recent article by Nigam et al [20], applied to the prediction of heart attacks using several machine learning models such as logistic regression and decision trees, a choice that is common to our current work. Their results prove that decision trees ensure faster processing in classification, with very good accuracy for all the trained models thanks to the feature selection process.

III. METHODOLOGY

In order to provide bio-inspired models with a coherent, complete and balanced dataset, the functioning of our IDS is triggered by a pre-processing phase of the chosen original dataset. Given that the details of this phase are outside the scope of this paper, we summarize its steps as follow: A purification of redundant data is carried out, followed by an elimination of missing data, given that the present work does not manage a specific treatment for cases not completely filled in. This phase also performs a data balance analysis and remedies any imbalanced classes with a standard treatment. In this study, we are interested in whether or not there is an intrusion in the traffic, and therefore not in the nature of the attack. That said, the dataset has been encoded so that the label attribute is transformed into a binary decision: Attack/Normal state. Once the dataset has been cleaned and prepared, it is provided as an input to the bio-inspired models to activate the FS phase. To extract the optimal subset in terms of reduced dimensionality, while maintaining a reliable representation of the cases, we chose to analyze the behavior of the four bio-inspired models indicated above, in order to assess the relevance of each one. Our approach is based on the integration of a model wrapper responsible for filtering the fitness function of each one. This strategy aims to identify a subset of features by assessing the model's performance during fitness function estimation. By reviewing the main algorithms of the four methods, we can see that the fitness function plays an important role, and the choice of its criteria is crucial. Indeed, each time the philosophy of the fitness function is changed, the selected subset may change. The output of each algorithm depends closely of it and as such we have introduced two factors:

- The error of the model wrapper, i.e. the machine learning model used to evaluate the performance/importance of the selected features.
- The percentage of selected features relative to the total dimension of the dataset.

The four algorithms are trained on a large number of cases, and each generates a subset of features that locally provides the best value of the fitness function. The aim of this study is to take advantage of the diversity of the adopted models, and not to favor one model over the others. Indeed, in a context of generalization to new unknown cases, no model can guarantee permanent superiority. From this perspective, we have proposed a way of combining the different points of view of all the models. The next phase is to decide on the final structure of the best subset based on the outputs of all bio-inspired algorithms and so, the conceived model brings together the structures of the four generated subsets in a pool. Which one and through two strategies, takes charge of deciding on the final structure to be considered for the rest of the process. The first strategy is to consider all the characteristics that at least one algorithm has agreed on in a non-redundant way. The vision here is to consider the merging of all decisions without redundancy, in order to benefit from the full diversity of the strengths of the algorithms but adding a criterion that we deemed important: the weight of each feature, which is estimated according to the number of algorithms that have chosen it. The second strategy is chosen to be more restrictive, and consists of retaining only those features on which all the algorithms have agreed, and therefore, present in the intersection of the four generated vectors. The idea is to give presence to the attribute if all models have approved it, as this proves its weight in the decision.

In the following, the subsets outputs are named: DS_U (DataSet generated by the Union strategy) and DS_I (DataSet generated by the Intersection strategy). In DS_I , all features have the same weight = 1, as long as they were all chosen by all the algorithms. In DS_U , the features are weighted by a coefficient whose estimated values indicate the selection rate of each one by the algorithms. The values belong to the range [0.25, 1], where the extreme value is given when all the algorithms agree on the feature. This weighting will be used to manage confusion situations in the fusion model. Figure 1 shows the global process of the proposed model learning.

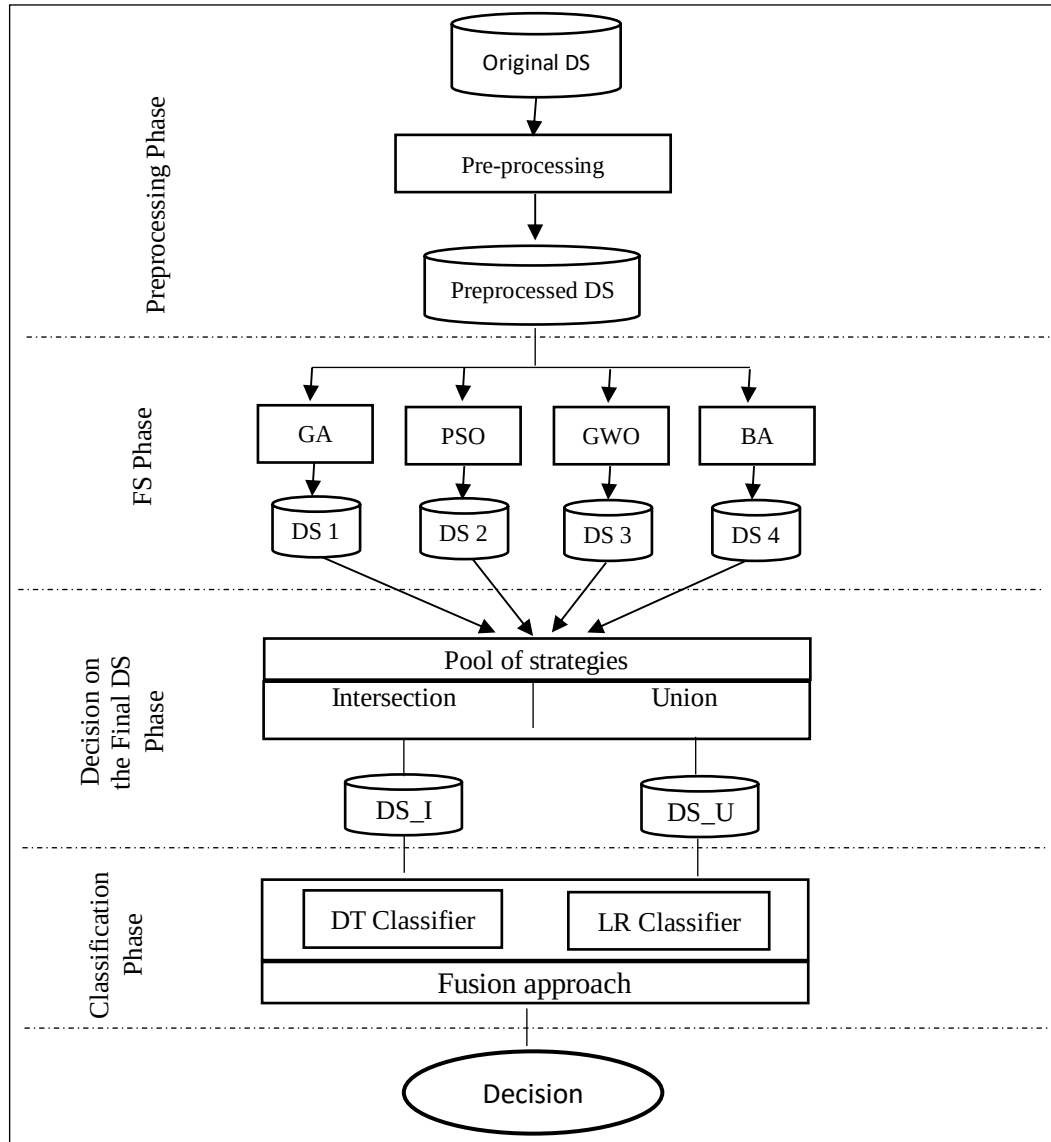


Figure 1: The proposed Learning Model.

Source: Authors, (2026).

To validate the new reduced subsets, a classification phase is performed. Among the best-known classification methods, we find Logistic Regression (LR) and Decision tree (DT). The first model is a powerful and simple statistical method for binary classification, based on the probability of belonging to a class. The interpretation of the probability value involves determining a threshold for belonging to the class. While the second model has the basic principle to use feature values to recursively divide the data set into smaller subsets. Its aim is to ensure that the examples in each subset belong to the same class or have similar output values. Regularization mechanisms, are integrated in order to reduce the risk of overfitting and improve the classifier's generalization ability on test subsets. So LR and DT are trained each one on two reduced datasets namely DS_I and DS_U , which means that four models are obtained in this classification phase. Each individual's performance metrics are retained, primarily the $F1_score$, which will be used as a model weighting factor in the ensemble approach, we have proposed. This approach aims to create a more accurate and robust classification model than each model individually by considering the synthesis of the weighted responses of the four models. The final model will choose the decision with the highest total weight $W_{decision}$ estimated by the equation (1) :

$$W_{decision} = \max \left(\sum_{j=1}^n F_j, \sum_{k=1}^m F_k \right) \quad (1)$$

Where:

$n + m$ = the number total of models

F_j : the $F1_score$ of a model that has decided on a “normal” state of the network,

F_k : the $F1_score$ of a model that has decided on a “attack” state of the network,

In the following, we denote:

- $M1$, the model result of the classifier LR trained on DS_I
- $M2$, the model result of the classifier LR trained on DS_U
- $M3$, the model result of the classifier DT trained on DS_I
- $M4$, the model result of the classifier DT trained on DS_U
- M , the model of the ensemble approach, based on $M1$, $M2$, $M3$ and $M4$ results

In the case where the two sums are equal, a confusing situation is generated and is handled by a specific management strategy to decide on the classification of the unknown case. The main idea in this strategy is to use the principle of similarity between cases to decide what response to take. The reduced (in features) datasets are filtered according to the first response, and a retrieval for the most similar case is triggered, then the process is repeated with filtering according to the second response, to finally decide in favor of the one giving the highest similarity. To ensure this retrieval process, similarity metrics must be defined. The equation (2) shows the estimation formula of the global similarity between 2 cases: *source* and *target*, while equations (3), and (4) give the estimation between two feature values.

$$Sim(source, target) = \frac{\sum_{a \in subset} w_a \times LocalSim_a(source.a, target.a)}{\sum_{a \in subset} w_a} \quad (2)$$

where:

- *source* represent a labeled case in the dataset
- *target*: the new case to be classified
- w_a : the feature weight estimated at the pool level.

$LocalSim_a$: is a function giving a local similarity between the same feature (a) in both source and target cases. When the type of a is Boolean or categorical, it is defined by the equation (3):

$$LocalSim_a(x, y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{else} \end{cases} \quad (3)$$

If the attribute a is of a numeric type, $LocalSim_a$ is defined by the equation (4):

$$LocalSim_a(x, y) = 1 - \frac{|y - x|}{B_a} \quad (4)$$

where B_a is the “breadth” of the range of a , i.e., it is the difference between the maximal and the minimal values of this range.

IV. RESULTS AND DISCUSSIONS

IV.1 THE USED DATASET

The NSL-KDD dataset was selected to evaluate the contribution of this work. It contains 25192 records of Internet traffic as seen by a simple intrusion detection network. This number is reasonable, making it affordable to run experiments on the complete set without the need to randomly select a small portion and so, the evaluation of results will be consistent and comparable. The dataset contains 43 features per record, 41 of the features refer to the traffic entry itself and the last two are “labels” and “score”. Labels inform on the state whether normal or attack type, i.e, gives the attack type: DoS (Denial of Service), Probe, U2R (User to Root) or R2L (Remote to Local), while the score gives the severity of the intrusion. The dataset attributes are of different type, we can find: 4 Categorical attributes, 6 Binary attributes, 23 Discrete attributes, and 10 Continuous attributes. In this work we are particularly interested in the Normal or Intrusion traffic situation. The Label attribute was adapted to our objective in a pre-processing phase for binary classification. Thus, we leave the consideration and treatment of attack types for a future work.

IV.2 THE FEATURE SELECTION PHASE

To generate the different individual models, each based on a specific FS method, a hyperparameter selection phase is necessary. After various empirical tests and analytical studies, we selected a set of hyperparameters for each one as mentioned in Table 2.

Table 2: Algorithm Hyperparameters.

Algorithm	Parameters
GA	$P_c = 0.2, m = 0.5$
GWO	$T = 10$
PSO	$w = 0.7, C1 = 0.6, C2 = 0.5, \text{iteration number} = 10$
BA	$\alpha = 0.8, A = 0.15, PRF \in [0, 2]$

Source: Authors, (2026).

The confusion matrix was used to validate the results performance. Based on four parameter rates, this matrix can show quickly whether the system manages to classify correctly. The classification quality is better when parameters *True_Positive (TP)* and *True_Negative (TN)* are maximal while *False_Negative (FN)* and *False_Positive (FP)* should be minimal. In addition, we used the performance metrics: *Accuracy*, *Precision*, *Recall*, and *F1-score* which are widely used to provide an overall assessment of a classification model. *Accuracy* measures the proportion of correctly classified instances among all predictions. *Precision* reflects the model's ability to minimize *FP* by quantifying the proportion of *TP* instances among all predicted positives. *The F1-score*, which ranges from 0 to 1, represents the harmonic mean of *precision* and *recall* and thus captures the trade-off between these two metrics. In an intrusion detection context, a high *F1-score* indicates that the model is effective not only in reducing false alarms (high *precision*) but also in correctly identifying the majority of actual intrusions (high *recall*). We give below some formulas for estimating model performance. The equations (5), (6), (7) and (8) show how the different rates of confusion matrix are estimated while the equations (9), (10), (11) and (12) give the estimation of the *F1-score*, Recall, Precision, and Accuracy respectively.

$$TruePositiveRate(TPR) = \frac{TP \times 100}{FN + TP} \quad (5)$$

$$TrueNegativeRate(TNR) = \frac{TN \times 100}{FP + TN} \quad (6)$$

$$FalseNegativeRate(FNR) = \frac{FN \times 100}{FN + TP} \quad (7)$$

$$FalsePositiveRate(FPR) = \frac{FP \times 100}{TN + FP} \quad (8)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (12)$$

At the end of the FS phase, we collected the results, which clearly show that the best selected subsets vary from one method to another. Table 3 illustrates the number of features selected by each algorithm, and by observing it, we note that PSO was able to select the smallest subset in dimensionality reducing the total number of features from 41 to 14 and therefore to less than 35%. GWO managed to reduce to 39%, while GA and BA achieved reductions of 43% and 56% respectively. The performance metrics obtained for each algorithm are reported in the same table and illustrated in a graph in Figure 2 to provide greater clarity when comparing the four algorithms.

Table 3: The dimensionalities returned from the FS phase.

Algorithm	Features number	Precision	F1-score	Accuracy	Recall
GA	18	0,85008635	0,76752744	0,80276248	0,69958496
GWO	14	0,46542827	0,63521126	0,46544242	0,99999999
PSO	23	0,97318419	0,86079401	0,88383784	0,77167548
BA	16	0,96878154	0,85425474	0,87867802	0,76394337

Source: Authors, (2026).

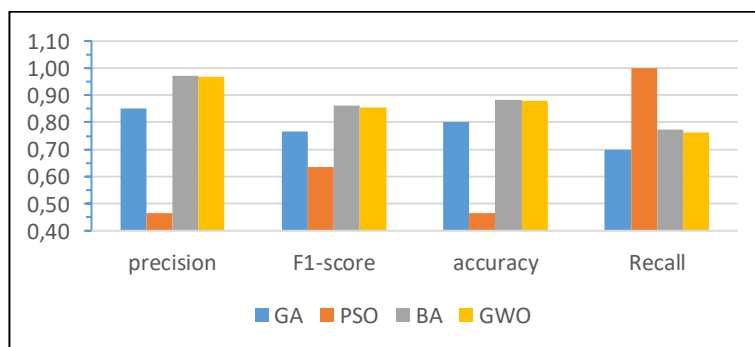


Figure 2: performance metrics obtained in the feature selection phase.

Source: Authors, (2026).

According to these illustrations by Table 3 and Figure 2, we note that scores vary widely between the results of the algorithms, ranging from 0.46542827 to 0.97318418 for *precision*. The subset given by PSO looks very weak here, while BA and GWO features prove their efficiency. The highest accuracy is 0.88383784, which is relatively good but not exceptional. We can easily notice a strict correlation between the number of features and the precision value. The largest subset given by the BA algorithm with 23 features brings the highest precision. This suggests that, at this level, an average reduction in the dimension provides a better representation of traffic and therefore allows for more accurate decision making. These new subsets are transmitted to the pool which creates two new subsets: *DS_I* and *DS_U* according to the principles detailed above. However, we cannot assert the effectiveness of subsets without validating them with robust classifiers such as LR or DT, which is the subject of the next phase. The classification process in this phase gives rise to four diverse and trained models: *M1*, *M2*, *M3* and *M4* as mentioned in the section (III). Table 4 highlights the F1-scores performance of these four models while Figure 3 presents a comparison of them with the F1-score obtained in the feature selection phase. We note a clear improvement in this parameter and therefore a clear improvement in the precision/recall balance when working on the subsets generated by our pool strategies. Generally, all the others metrics have improved significantly in this phase (a overview on the accuracy scores is given too in the Table 4).

Table 4: F1-score of the 4 trained models.

Model	F1-score	Accuracy
M1	0,99801068	0,99814775
M2	0,9978678	0,99801545
M3	0,99795326	0,99809483
M4	0,99786792	0,99801545

Source: Authors, (2026).

Our third contribution in this work is the ensemble approach which creates the model *M* based in the fusion whose details are presented in the section III. This model gave an F1-score of **0.99818099**, which improve the performance model more than the models *M_i*. ($i=1,4$) by offering the best *TP* and the best *FN* of all models discussed in this work. This shows the advantage of considering the results of the four models to benefit from their power and weighting them with their F1- score.

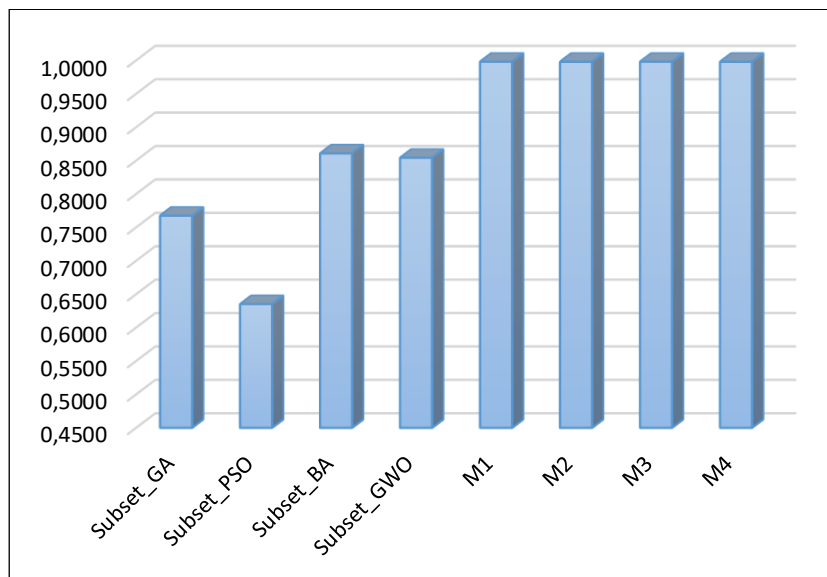


Figure 3: Comparison of the models F1-score.

Source: Authors, (2026).

V. CONCLUSIONS

This paper deals with the network detection intrusion problem. A field which is often represented by a wide dimensionality of characteristics. The trade-off for accurate and fast diagnosis in network traffic is always a significant one, because in fact, the better the situation is enriched by additional data, the better the accuracy is guaranteed, but this is at the detriment of response time, which is slowed down and exposes the network to risk. The work serves three objectives, for which, four bio-inspired algorithms are investigated to extract the best subset of features allowing the most accuracy classification of the network state. The use of *GA*, *PSO*, *BA* and *GWO*, has allowed a significative reduction in dimensionality by selecting the most representative features. In a second step, their subsets outputs have been introduced into a pool which includes two different proposed strategies to make the final decision on the subset structure that allows the weighting of each feature. Thus, two new subsets are created based on the outputs of feature selection phase. The effectiveness of these subsets has been validated using two strong classifiers namely *DT* and *LR* in an individual way; to produce four different machine learning models whose performance metrics showed a clear improvement over those of the four algorithms. After, and in a vision to take advantage of their benefits, they have been merged in a robust ensemble model augmented by a confusion management strategy. The performance metrics show that the compromise between the reduced subset and decision accuracy was successfully established by the final model achieving the rate of 99.8%. The proposed model is capable of dealing with datasets other than *NSL-KDD*. However, it may give different performances. A comparative study of results will be carried out in a next work by considering another intrusion detection dataset.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Souad Guessoum

Methodology: Souad Guessoum.

Investigation: Oussama Kadri.

Discussion of results: Souad Guessoum.

Writing – Original Draft: Souad Guessoum.

Writing – Review and Editing: Souad Guessoum and Nabih Azizi.

Resources: Souad Guessoum and Nabih Azizi.

Supervision: Souad Guessoum and Nabih Azizi

Approval of the final text: Souad Guessoum and Nabih Azizi

VII. REFERENCES

- [1] Y. P. khah, M. H. Shirvani, and J. Taheri, "A survey study on meta-heuristic-based feature selection approaches of intrusion detection systems in distributed networks", *Computer Standards Interfaces*, vol. 96, no. 104074, 2026. (<https://doi.org/10.1016/j.csi.2025.104074>)
- [2] J. H. Holland, "Genetic Algorithms and the Optimal Allocation of Trials", *SIAM Journal on Computing*, vol. 2, 1973.
- [3] J. da S. Bohrer and M. Dorn, "Enhancing classification with hybrid feature selection: A multi-objective genetic algorithm for high-dimensional data", *Expert Systems With Applications*, vol. 255, no. 124518, 2024. (<https://doi.org/10.1016/j.eswa.2024.124518>)
- [4] N. Mateen, M. Naeem, M. J. Khan, T. Yousaf, A. Ali, W. A. Altabay, M. Noori, and S. A. Kouritem, "Subject based feature selection for hybrid brain computer interface using genetic algorithm and support vector machine", *Results in Engineering*, vol. 27, no. 105649, 2025. (<https://doi.org/10.1016/j.rineng.2025.105649>)
- [5] V. Imani, C. Sevilla-Salcedo, E. Moradi, V. Fortino, and J. Tohka, "Multi-objective genetic algorithm for multi-view feature selection for the Alzheimer's Disease Neuroimaging Initiative", *Applied Soft Computing*, vol. 167, no. 112332, 2024. (<https://doi.org/10.1016/j.asoc.2024.112332>)
- [6] D. Aksu and M. A. Aydin, "MGA-IDS: Optimal feature subset selection for anomaly detection framework on in-vehicle networks-CAN bus based on genetic algorithm and intrusion detection approach", *Computers and Security*, vol. 118, no. 102717, 2022. (<https://doi.org/10.1016/j.cose.2022.102717>)
- [7] Y. Fang, Y. Yao, X. Lin, J. Wang, and H. Zhai, "A feature selection based on genetic algorithm for intrusion detection of industrial control systems", *Computers and Security*, vol. 139, no. 103675, 2024. (<https://doi.org/10.1016/j.cose.2023.103675>)
- [8] S. Mirjalili, S. M. Mirjalili, and A. Lewi, "Grey Wolf Optimizer", *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014. (<https://doi.org/10.1016/j.advengsoft.2013.12.007>)
- [9] Y. Xin-She, "A new metaheuristic bat-inspired algorithm", in *Nature Inspired Cooperative Strategies for Optimization (NICSO 2010)*, ser. Studies in Computational Intelligence. Berlin: Springer, 2010, vol. 284, pp. 65–74. (https://doi.org/10.1007/978-3-642-12538-6_6)
- [10] J. Kennedy and R. C. Eberhart, "Particle swarm optimization", in *Proceedings of the IEEE International Conference on Neural Networks (ICNN'95)*, vol. 4. Washington, DC, USA: IEEE Computer Society, 1995, pp. 1942–1948. (<https://doi.org/10.1109/ICNN.1995.488968>)
- [11] A. Ali, M. Assam, F. U. Khan, Y. Y. Ghadi, Z. Nurdaulet, A. Zhibek, S. Y. Shah, and T. J. Alahmadi, "An optimized multilayer perceptron-based network intrusion detection using Gray Wolf Optimization", *Computers and Electrical Engineering*, vol. 120, no. 109838, 2024. (<https://doi.org/10.1016/j.compeleceng.2024.109838>)
- [12] R. Wang, H. Xu, D. Yi, C. Song, and Y. Che, "Automatic detection of Alzheimer's disease from EEG signals using hybrid PSO-GWO algorithm", *Biomedical Signal Processing and Control*, vol. 107, no. 107798, 2025. (<https://doi.org/10.1016/j.bspc.2025.107798>)
- [13] P. Majumdar and S. Mitra, "Optimized Soil Moisture Prediction Using RF-PSO and Modified BACO Based Feature Selection", *SN Computer Science*, vol. 6, no. 662, 2025. (<https://doi.org/10.1007/s42979-025-04162-3>)
- [14] J. Samraj and K. Abirami, "Attack Detection with Optimal PSO Feature Selection with DTOHE Model", *SN Computer Science*, vol. 5, no. 562, 2024. (<https://doi.org/10.1007/s42979-024-02911-4>)
- [15] M. Pouryosef, R. Abedini-Nassab and S. M. R. Akrami, "A Novel Framework for Epileptic Seizure Detection Using Electroencephalogram Signals Based on the Bat Feature Selection Algorithm", *Neuroscience*, no. 5, pp. 35–49, 2024. (<https://doi.org/10.1016/j.neuroscience.2024.01.014>)
- [16] P. Modi and Y. Kumar, "Smart detection and diagnosis of diabetic retinopathy using bat-based feature selection algorithm and deep forest technique," *Computers Industrial Engineering*, vol. 182, no. 109364, 2023. (<https://doi.org/10.1016/j.cie.2023.109364>)
- [17] J. Hu, K. Pan, Y. Song, G. Wei, and C. Shen, "An improved feature selection method for classification on incomplete data: Non-negative latent factor-incorporated duplicate MIC," *Expert Systems With Applications*, vol. 212, no. 118654, 2023. (<https://doi.org/10.1016/j.eswa.2022.118654>)
- [18] Y. Liu, Z. Liu, X. Luo, and H. Zhao, "Diagnosis of Parkinson's disease based on SHAP value feature selection," *biocybernetics and biomedical engineering*, vol. 42, pp. 856-869, 2022. (<https://doi.org/10.1016/j.bbe.2022.06.007>)
- [19] A. M. Priyatno, Y. Ningsih, R. J. Farhas, F. I. Firmananda, R. K. Sari, and Aryadi, "Multi-stage feature selection optimizing student dropout prediction", *JETIA*, vol. 11 no.55, pp.86-95. September/october, 2025 (<https://doi.org/10.5935/jetia.v11i55.1910>)
- [20] R. Nigam, H. R. Jha and R. Ranjan, "Machine Learning Models Analysis using Multi-Factors for heart attack risk prediction", *JETIA*, vol. 11 no.56, pp.240-247. November/December, 2025, (<https://doi.org/10.5935/jetia.v11i56.2734>)