2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)
October 11-14, 2020. Toronto, Canada

Binary Hybrid Differential Evolution Algorithm for Multi-label Feature Selection

Nelson Vithayathil Varghese
Electrical, Computer, and Software Engineering
OntarioTech University
Oshawa, Canada
nelson.vithayathilvarghese@ontariotechu.net

Amritpal Singh
Electrical, Computer, and Software Engineering
OntarioTech University
Oshawa, Canada
amritpal.singh@ontariotechu.net

Ashwin Suresh
Electrical, Computer, and Software Engineering
OntarioTech University
Oshawa, Canada
ashwin.suresh@ ontariotechu.net

Shahryar Rahnamayan, SMIEEE
Nature Inspired Computational Intellegence(NICI) Lab,*Electrical, Computer, and Software Engineering*
OntarioTech University
Oshawa, Canada
shahryar.rahnamayan@uoit.ca

Abstract*— Driven by the recent technological advancements within the machine learning field, multi-label classification has been introduced as one of the challenging tasks to assign more than one label to each instance in a dataset. Feature selection is one of the predominant feature engineering methodologies which being extensively used as a vital step in predictive model construction to enhance the multi-label classification performance. Many metaheuristic algorithms have been tailored to choose the optimal subset of features in datasets but as a challenging problem, such algorithms suffer from a slow process during fine-tuning. Objective of this paper is to propose a hybrid mechanism by which an obtained feature subset from a Binary Differential Evolution (BDE) algorithm will be further enhanced to minimize the classification error using a local search methodology. Key motivation behind the proposed model is to address the weakness in exploitation of metaheuristic feature selection algorithms with the help of classical feature selection method such as Sequential Backward Selection (SBS) as a local search strategy. The classical feature selection method eliminates more redundant and irrelevant features of obtained subset using the BDE to decrease the classification error. The empirical results obtained on eight various multi-label datasets show that the proposed hybrid approach, which is a fusion of both evolutionary and classical feature selection methods, can minimize the classification error on the obtained feature subset using the BDE.

Keywords*— *Feature Selection, Binary Differential Evolution, Sequential Backward Selection, Multi-label classification, Hybrid Method, Memetic Differential Evolution.

I. INTRODUCTION

Traditionally single-label classification is associated with learning from a set of given examples that are related with a particular and only a single label (class) (Trohidis, Tsoumakas, Kalliris, & Vlahavas, 2008). When it comes to multi-label classification, the instances within the dataset are related to a set of labels $Y \subseteq L$, where L represents the set of entire disjoint labels. Nowadays multi-label classification has gained wide popularity in numerous real-world scenario-based applications including protein function classification, music categorization, and semantic scene classification. For instance, in the case of semantic scene classification, a photograph instance taken from the dataset could probably belong to more than one conceptual class, such as sunsets and beaches simultaneously (Boutell, Luo, Shen, & Brown, 2004). In a similar fashion, within the music categorization a song could probably belong to more than one genre class at the same time (Tsoumakas & Katakis, Multi-label classification: An overview, 2007).

Now a days data mining plays a critical role in knowledge discovery. The data collected will have often many features that are either redundant or irrelevant. In most cases, these extra features do not contribute to the knowledge discovery but rather they tend to raise the complexity of the data mining process (Mlambo, Cheruiyot, & Kimwele, 2016). Having said that, a preprocessing task such as feature selection is used in both data mining and machine learning for removing such redundant and irrelevant features to a great extent (Bidgoli, Ebrahimpour-Komleh, & Rahnamayan, 2019; Bidgoli A. A.-K., 2020). Feature selection not only leads to the improvement of the overall performance of classification but also, it decreases the complexity involved in constructing model.

The search for an optimal feature subset is often being considered as a NP-hard problem. As a result of this, an exhaustive search within the entire search space is required to acquire the optimal solution, which otherwise is not guaranteed to obtain. Application of metaheuristic based techniques facilitates to obtain fairly good solutions without having to explore the search space (Yusta, 2009). Even though metaheuristic techniques have the ability to find acceptable solutions in reasonable time with the help of either experience-based techniques or by guided search, they do not always guarantee that the optimal solution is found.

Metaheuristic-based techniques typically follow an iterative generation process that guides a subordinate heuristic for exploring the search space. Some of the popular metaheuristic-based techniques being used within the feature selection arena are predominantly based on methods such as genetic algorithm (Wang, et al., 2014), particle swarm optimization (Ramos, Chiachia, Papa, Souza, & Falcão, 2011), ant colony optimization (Boubezoul & Paris, 2012), and quantum-inspired evolutionary algorithm (Chen, Chen, & Chen, 2013). All these algorithms perform a sequence of iterative computations with an intention to evolve a population of individuals. Generally, techniques based on evolutionary algorithms are used to solve the feature selection (Leardi, Boggia, & Terrile, 1992). One of the areas where metaheuristic-based feature selection solutions are implemented is within the field of sentimental analysis which is based on natural language processing (Ahmad, Bakar, & Yaakub, 2015).

II. Background Review

A. Differential Evolution

Real-world problems are being solved with the help of an optimization algorithm widely known as differential evolution (DE). It can be viewed as a population oriented stochastic method with the aim of achieving global optimization (Chakraborty, 2008). DE algorithm brought forward by R. Storn and K. V. Price is primarily made up of three key evolutionary operators, namely mutation, crossover and greedy selection (Storn & Price, 1997). A population of search vectors that is variable is initialized, after which the mutation operator is used (Das & Suganthan, 2010). The operator for the mutation operation of the classical DE (i.e., DE/rand/bin/1) for an individual element j is denoted as:

$$V_{j,i} = X_{j,i1} + F.(X_{j,i2} - X_{j,i3}) \quad (1)$$

Mutation operator generates a new D-dimensional vector, V_i, by using three individuals, $X_{j,i1}$,$X_{j,i2}$ and $X_{j,i3}$ which are selected randomly from the current population (Bidgoli, Ebrahimpour-Komleh, & Rahnamayan, 2019). Mutation factor denoted by the parameter named F is used to scale the difference between two vectors, $X_{j,i2}$ and $X_{j,i3}$. Mutation operation is further followed by the crossover operation which induces the test vector by connecting the mutation vector with the population's selected parent vector. The selection operator as the name suggests is essentially used to find the better solution (Li, Ma, & Hu, 2017). The selection operator plays a vital role of selecting the individual members that will make the next generation from the current population as well as from the generated offsprings. In fact, a greedy selection determines which one of the parent or newly generated offspring should be forwarded to the next generation.

However, DE algorithm like any other algorithm, also has its drawbacks. Even though DE has good exploration ability, the local exploitation capability is rather weak. Along with this, DE mutation strategy and the algorithm's corresponding control parameters often play a critical role as performance of the algorithm is extremely sensitive to these control parameters (Feoktistov, 2006).

In order to tackle the binary feature selection, a binary version of differential evolution, BDE, was proposed which comprises of mutation, crossover, and selection operations (Bidgoli, Ebrahimpour-Komleh, & Rahnamayan, 2019). Feature selection is often treated as a binary optimization problem wherein number of features denotes the dimension of the problem. The differential evolution algorithm used in this research work uses a binary mutation operator for producing the candidate solutions (Bidgoli, Ebrahimpour-Komleh, & Rahnamayan, 2019). Criteria of selection for each individual feature for the newly generated candidate solution will be based off the status of that particular feature within those three candidate solutions. Depending on the agreement between two candidate solutions selected randomly, binary mutation operator alters the presence or absence of a specific feature within a subset of features. If a feature is selected only in one of the subsets, then the status quo for that particular feature will be maintained same among both the parent and the candidate solution. In a similar manner, if both selected solutions include the selection of a specific feature, so same would be made as part of in new candidate solution as well. This is due to the fact that there is an understanding on selection of that specific feature among the selected solutions. Finally, if any one of selected solutions doe not select a feature, then that particular feature would not be added as part of the candidate solution. This is based on the consensus that there is an understanding on not selecting that particular feature.

B. Objective Function of Multi-label Feature Selection

Primary objective of feature selection task in machine learning is to choose the best features which will increase the accuracy of classification with smaller number of meaningful features (Asilian Bidgoli, Ebrahimpour-Komleh, & Rahnamayan, 2020). It is quite necessary that algorithm requires enough number of important features, without which it will be underperforming. One major aspect within the multi-label feature selection process is the quality of the selected features for the intended classification task. For this purpose, hamming loss can be used as a reliable measure in the classification task to compute the quality of the features that are selected (Bidgoli, Ebrahimpour-Komleh, & Rahnamayan, 2019). It is given as:

$$hloss(h) = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{q} |h(xi) \Delta Yi| \quad (2)$$

where p denotes the number of samples and q represents the number of labels. The classification result of sample *i, (xi)* is represented by *h(xi)* and the labels of that sample are denoted by Yi. Hence the objective of the optimization algorithm in the selection is to minimize the hamming loss (Bidgoli, Ebrahimpour-Komleh, & Rahnamayan, 2019).

C. Heuristic Selection algorithms

Greedy Hill climbing algorithm (Caruana & Freitag, 1994), branch and bound method (Koontz, Narendra, & Fukunaga,, 1975), and beam search (Kumar, Vembu, Menon, & Elkan,

2013) are some of the popular heuristic methods of feature selection problem. To select the relevant features, Greedy hill climbing algorithm takes into consideration all the local changes [25]. Aforementioned local changes can be considered as either addition of a specific feature into the selected features or deletion of feature. SFS and SBS are two such kinds of those algorithms (Hall & Smith, 1997). SFS starts its operation by keeping selected features as an empty set and subsequently the most informative feature into the set. Whereas SBS functions in an opposite way by starting with the complete set of features and during each subsequent iteration, one of the most redundant or irrelevant features is dropped from the feature set. Proposed method relies on the SBS method wherein the solution generated by the binary differential evolution will be enhanced further to reduce the hamming loss as well as the size of the feature set.

III. Proposed Hybrid Method For the Enhacement of Single Objective Optimization Solution

Even though evolutionary algorithms are known for their great potential in exploration and reaching close to the optimal solutions in numerous real-world problems, it is often wise to make use of a specialized method to increase the chance of convergence to the global optimal solutions through a local search methodology (Zhu, Ong, & Dash, 2007). Since the classical methods such as SBS, SFS have good convergence capability through exploitation to local optimal solution, a hybrid method combining the benefits of these two approaches is a natural choice (Ahn, Kim, Kim, Lim, & An, 2010). In one approach, an evolutionary approach can be used to find good initial solutions for further enhancement by classical method, which then can make an attempt to find the solutions even closer to the global optimum. Because of this, the combined use of an evolutionary approach and an enhancement method forming a relay scheme may result in a saving of computational effort, if used properly. Going by this, hybrid methods of using evolutionary algorithms with classical method forming a relay scheme can be used in the context of single-objective optimization. Predominantly, there are at least three reasons why a hybrid method would be useful to solve a single-objective optimization problem. Firstly, it ensures better convergence to the global optimal solution. Secondly, often it demands smaller computational effort than each individual method applied alone to find the solution with the same quality of which hybrid finds. Finally, the hybrid methods can decrease the number of selected features with preserving or even decreasing the accuracy of classification (Ahn, Kim, Kim, Lim, & An, 2010).

Objective of forthcoming section is to provide details about the hybrid method applied towards the enhancement of single objective optimal feature selection. The hybrid method is applied on the best solution generated by a binary differential evolution algorithm.

A. *Hybrid method using SBS*

Proposed hybrid method can be viewed as a two-stage processing model, wherein during the first stage of feature selection process an evolutionary algorithm, such as BDE, will be employed to generate the best solution from the search space of a specific feature selection problem (Pampara, Engelbrecht, & Franken, 2006). In general, metaheuristic-oriented techniques are proven to be extremely good at exploration of feature set and efficient in obtaining acceptable solutions within a reasonable timeframe, though it does not guarantee that the optimum solution is found. In the second stage of feature selection process, which is more kind of an enhancement phase, the best solution generated from the first stage of feature selection will be used as the input. During this stage a classical approach like SBS will be adopted, where it sequentially starts removing irrelevant or redundant feature(s) to further enhance the solution in terms of the hamming loss measure. The hybrid model encompasses multi-label k-nearest neighbor (MLKNN) classification algorithm for the associated hamming loss calculation (Zhang & Zhou, 2038--2048). The MLKNN classifier plays a vital role in evaluating the relevance of features. The k-nearest neighbors (KNN) algorithm is a popular machine learning algorithm that is used to solve classification problems. The key assumption behind the KNN algorithm is based on the assumption that similar things exist in close proximity and is based on feature similarity algorithm functions by selecting K entries from the database which are closest to the new sample, and further finds the most common classification of these selected entries. Eventually, the new sample will be assigned with same class label of its neighbors. ML-KNN is a popular multi-label classifier model for the classification multi-label instances and based off extension of the single-label KNN classifier for multi-label classification support (Zhang & Zhou, 2038--2048).

SBS is a feature selection method based on wrapper method (Oh, Lee, & Moon, 2004). SBS initiates its operation with entire set of features, and subsequentially eliminates the feature x^-, the one that brings down the value of objective function $J(Y - x^-)$ by the least margin, where J denotes the objective function such as hamming loss. Elimination of a specific feature this method may actually increase objective function's value, $J(Y_k - x^-) > J(Y_k)$. SBS is found to be more efficient whenever the size of the optimal feature subset is relatively larger as it would give the local search strategy more room to visit the large subset search space. The major drawback of SBS is its inability to re-evaluate the effectiveness of a specific feature once it is eliminated from the feature set (Ahn, Kim, Kim, Lim, & An, 2010).

SBS algorithm will use hamming loss value as a baseline to decide which feature(s) needs to be dropped out. Same approach is also applied by adopting a hybrid method with proposed local search, where we randomly select and decide to drop a particular feature or set of features from the best solution generated by BDE algorithm if the elimination of that particular feature(s) can reduce the hamming loss further down.

In the single objective optimization solution enhancement scenario, the best solution from the BDE will be taken as X, then search for a particular feature x^- in such a way that removing that feature value would reduce the hamming loss than the value generated by BDE's best solution. This procedure can be re-iterated for a specified number of times, so that eventually SBS approach will help to achieve a set with lower hamming loss (cost) value and fewer number of features than the original input solution used from BDE's best output.

Algorithm 1 represents a high-level model for the proposed hybrid algorithm with key steps of operations involved during the functioning of hybrid model with the SBS.

Algorithm 1. High Level algorithm for the proposed hybrid model

Input: *Input Space with randomly initialized population*
Output: *Feature Subset having the lowest hamming loss*

1. *Repeat step 2 to 6 for maximum iteration value*
2. *Randomly select 3 candidate solutions from input space*
3. *Apply mutation operation with binary mutation operator*
4. *Apply crossover operation*
5. *Select the best candidate between parent and offspring solution*
6. *Store the candidate solution with best cost function value*

\\ Appling SBS algorithm
7. *Repeat steps 7 to 10 for pre-set iteration count*
8. *Remove one feature randomly from set resulted from BDE*
9. *Calculate the cost function value by using KNN classifier*
10. *Keep the new subset, if it is better.*
11. *Output the final solution's cost value (hamming loss)*
12. *end*

IV. Experiments and Result Analysis

A. Test Datasets and Settings

For the purpose of evaluating and measuring the proposed hybrid method's performance, eight various standard multi-label datasets are utilized. Table- I provides the details of all the datasets used and its related details such as the dataset domain, number of features, number of training and test sets used and the number of labels. All the datasets used within this experiment are benchmark datasets taken from multiple application domains such as image, biology, audio, and text. All of datasets used are divided into training and testing sets for model training and testing purposes. Origin of all the dataset are based on MULAN library from which they are taken from (Tsoumakas, Katakis, & Vlahavas, Mining multi-label data, 2009).

During the evaluation, optimization process is carried out on teach dataset's training set individually, and further on performance is measured on the test set. The hybrid method is compared with BDE algorithm. In order to maintain a fair comparison, both methods have utilized the same control parameter settings. For the single objective solution enhancement, the main focus was on the minimization of hamming loss (objective). During the model evaluation process, the same level of budget was allocated to both methods in terms of the number of fitness evaluation calls was evoked. For the evaluation, model parameters such as size of the population and number of evaluations on each dataset are fixed to 100 and 100,000, respectively. In order to evaluate all the possible combination of selected features, MLKNN is utilized. MLKNN classifier functions by predicting the label(s) of unseen data by looking into the similarity levels of its major K neighboring samples. K denotes the number of neighboring samples to look for similarity check when the classifier needs to identify the labels of a new data item. MLKNN classifier classifies a multi-label data sample to classes with which the distribution of sample neighbors' similarity checks is higher. K is set to 10 in our experiments. For the optimization efforts carried out for single objective enhancement experiments, the MATLAB R2019b version was used as the development platform.

TABLE I. STANDARD MULTI -LABEL DATASETS

Dataset	Domain	Features	Labels	Training set	Test set
CAL500	Music	68	174	352	150
Emotions	Music	72	6	391	202
Yeast	Biology	103	14	1500	917
Birds	Audio	260	19	322	323
Scene	Image	294	6	1211	1196
Enron	Text	1001	53	1123	579
Genbase	Biology	1186	27	463	199
Medical	Text	1449	45	645	333

B. Experimental Results

This section details the experimental results obtained by the BDE and hybrid method on eight various datasets. Key objective behind these experiment attempts was to optimize (minimize) the hamming loss further down with the proposed hybrid approach using SBS. Test results for the hybrid method was generated by applying the local search strategy on the best solution generated by the BDE. Based on the test results obtained with hybrid approach, it is quite clear that further minimization of hamming was achieved with SBS-based hybrid approach. The proposed approach could employ the local search strategy effectively to exploit the solution generated by BDE to further optimize the hamming loss. As it is evident from the Table II, empirical results indicate that hybrid approach applied with all the eight different datasets could minimize the hamming loss further down in comparison what was achieved with BDE algorithm.

Table II indicates the related statistics measures generated based on the test results obtained on eight datasets by applying BDE, and proposed SBS-based hybrid method over 10 independent runs. As it is evident from the statistics, it is conspicuous that the proposed hybrid methods able to outperform BDE approach in optimizing (minimizing) the hamming loss across all the eight datasets. By analyzing the test results achieved on datasets such as Enron, it is imperative that the proposed hybrid method has significant improvement when the dataset has relatively high number of features. Whereas on datasets that have relatively smaller number of features, such as CAL500 and emotions, the enhancement achieved by the proposed hybrid method is marginal as BDE is quite efficient in finding solutions which are quite better subset in terms of classification error and number of features.

TABLE II. THE COMPARISION OF PROPOSED HYBRID METHOD WITH BDE IN TERMS OF HAMMING LOSS AND FEATURE COUNT

Datasets		Hamming Loss		Number of Features	
		BDE	Hybrid	BDE	Hybrid
Birds	Mean	0.0402	**0.0373**	174	**169**
	Median	0.0395	0.0374		
	Std	0.0015	0.0007		
Enron	Mean	0.0421	**0.0400**	848	**823**
	Median	0.0417	0.0402		
	Std	0.0002	0.0008		
Emotions	Mean	0.1414	**0.1363**	53	**49**
	Median	0.1415	0.1366		
	Std	0.0011	0.0011		
Medical	Mean	0.0093	**0.0082**	759	**738**
	Median	0.0093	0.0083		
	Std	0.0002	0.0003		
CAL500	Mean	0.1357	**0.1256**	36	**24**
	Median	0.1368	0.1271		
	Std	0.0034	0.0036		
Scene	Mean	0.0609	**0.0527**	178	**153**
	Median	0.0612	0.0532		
	Std	0.0022	0.0016		
Yeast	Mean	0.1665	**0.1624**	76	**72**
	Median	0.1677	0.1620		
	Std	0.0027	0.0030		
Genbase	Mean	0.0032	**0.0025**	887	**859**
	Median	0.0028	0.0025		
	Std	0.0011	0.0001		
w/t/l		0/0/8	8/0/0	0/0/8	8/0/0

As a supplementary measure to access the impact of proposed hybrid method on the other related aspects of feature selection other than hamming loss minimization, we have also captured the number of features associated with each of optimized candidate solutions. The last column of Table-II indicates the number of features for resulted subset of BDE and hybrid method. As it is expected from the operation of SBS, while applying the local search to reduce the hamming loss, on each run, SBS attempts to reduce number of features if its elimination could reduce the hamming loss. Therefore, the number of features or size of optimal feature subset achieved with hybrid method is decreased on all datasets with even decreasing the hamming loss error. Reduction in the number of features implies reducing the computational complexity and also the problem dimension for the classification task. In light of the empirical results, it is reasonable to conclude that the proposed hybrid method is able to achieve a lower hamming loss and also low number of features, simultaneously.

V. CONCLUSION AND FUTURE WORKS

This paper proposed a hybrid approach for multi-label feature selection wherein obtained solutions from binary differential evolution algorithm are further modified using a classical method, SBS. Proposed hybrid approach creates a combination of BDE algorithm and the classic method, wherein best solution generated by the former will be enhanced further by the latter. In the current study, we focused on experiments towards single objective optimization associated with multi-label feature selection. To this end, the proposed approach uses the classical method as a local search method to increase the exploitation power of the initial solution generated by BDE. In this manner hybrid approach attempts to bring in the right balance of both exploration and exploitation into its enhanced solution. All the experimental results are captured to show the improvements for single objective optimization solution's enhancement. The resulted feature subsets have fewer number of features while they increase the accuracy of classification. The future work is planned to include the application hybrid method in the context of multi-objective feature selection. Furthermore, other classical feature selection methods can be also utilized as a local search strategy.

VI. REFERENCES

[1] K. Trohidis, G. Tsoumakas, G. Kalliris and I. P. Vlahavas, "Multi-label classification of music into emotions," in *ISMIR*, vol. 8, 2008, pp. 325--330.

[2] M. R. Boutell, J. Luo, X. Shen and C. M. Brown, "Learning multi-label scene classification," Pattern recognition, vol. 37, no. 9, pp. 1757--1771, 2004.

[3] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," International Journal of Data Warehousing and Mining (IJDWM), vol. 3, no. 3, pp. 1--13, 2007.

[4] N. Mlambo, W. K. Cheruiyot and M. W. Kimwele, "A survey and comparative study of filter and wrapper feature selection technique," International Journal of Engineering and Science (IJES), vol. 5, no. 8, pp. 57--67, 2016.

[5] A. A. Bidgoli, H. Ebrahimpour-Komleh and S. Rahnamayan, "A novel multi-objective binary differential evolution algorithm for multi-label feature selection," 2019 IEEE Congress on Evolutionary Computation (CEC), pp. 1588--1595, 2019.

[6] A. A. a. E.-K. H. a. R. S. Bidgoli, "An evolutionary decomposition-based multi-objective feature selection for multi-label classification," PeerJ Computer Science, vol. 6, no. 6, p. e261, 2020.

[7] S. C. Yusta, "Different metaheuristic strategies to solve the feature selection problem," Pattern Recognition Letters, vol. 30, no. 5, pp. 525--534, 2009.

[8] L. Wang, H. Ni, R. Yang, V. Pappu, M. B. Fenn and P. M. Pardalos, "Feature selection based on meta-heuristics for biomedicine," Optimization Methods and Software, vol. 29, no. 4, pp. 703--719, 2014.

[9] C. C. Ramos, G. Chiachia, J. P. Papa, A. N. Souza and A. X. Falcão, "A novel algorithm for feature selection using harmony search and its application for non-technical losses detection," Computers \& Electrical Engineering, vol. 37, no. 6, pp. 886--894, 2011.

[10] A. Boubezoul and S. Paris, "Application of global optimization methods to model and feature selection," Pattern Recognition, vol. 45, no. 10, pp. 3676--3686, 2012.

[11] B. Chen, L. Chen and Y. Chen, "Efficient ant colony optimization for image feature selection," Signal processing, vol. 93, no. 6, pp. 1566--1576, 2013.

[12] R. Leardi, R. Boggia and M. Terrile, "Genetic algorithms as a strategy for feature selection," Journal of chemometrics, vol. 6, no. 5, pp. 267--281, 1992.

[13] S. R. Ahmad, A. A. Bakar and M. R. Yaakub, "Metaheuristic algorithms for feature selection in sentiment analysis," in 2015 Science and Information Conference (SAI), 222--226, 2015, pp. 222--226.

[14] R. S. Wahono, N. Suryana and S. Ahmad, "Metaheuristic optimization based feature selection for software defect prediction," Journal of Software, vol. 9, no. 5, pp. 1324--1333, 2014.

[15] S. Fong, S. Deb, X.-S. Yang and J. Li, "Feature selection in life science classification: metaheuristic swarm search," IT Professional, vol. 16, no. 4, pp. 24--29, 2014.

[16] F. Glover, "Tabu search—part I," ORSA Journal on computing, vol. 1, no. 3, pp. 190--206, 1989.

[17] M. Rajkumar, P. Asokan, N. Anilkumar and T. Page, "A GRASP algorithm for flexible job-shop scheduling problem with limited resource constraints," International Journal of Production Research, vol. 49, no. 8, pp. 2409--2423, 2011.

[18] H. Salimi, "Stochastic fractal search: a powerful metaheuristic algorithm," Knowledge-Based Systems, vol. 75, pp. 1--18, 2015.

[19] A. A. Bidgoli, H. Ebrahimpour-Komleh and S. Rahnamayan, "A Many-objective Feature Selection Algorithm for Multi-label Classification Based on Computational Complexity of Features," in 14th International Conference on Computer Science & Education (ICCSE), Toronto, 2019.

[20] A. A. Bidgoli, H. Ebrahimpour-Komleh and S. Rahnamayan, "Opposition-based multi-objective binary differential evolution for multi-label feature selection," in International Conference on Evolutionary Multi-Criterion Optimization, Michigan, 2019.

[21] M. A. Hall, "Correlation-based feature selection for machine learning," University of Waikato Hamilton, 1999.

[22] B. Xue, M. Zhang, W. N. Browne and X. Yao, "A survey on evolutionary computation approaches to feature selection," IEEE Transactions on Evolutionary Computation, vol. 20, no. 4, pp. 606--626, 2015.

[23] I.-S. Oh, J.-S. Lee and B.-R. Moon, "Hybrid genetic algorithms for feature selection," IEEE Transactions on pattern analysis and machine intelligence, vol. 26, no. 11, pp. 1424--1437, 2004.

[24] A. Hussain and Y. S. Muhammad, "Trade-off between exploration and exploitation with genetic algorithm using a novel selection operator," Complex \& Intelligent Systems, pp. 1--14, Springer.

[25] U. K. Chakraborty, Advances in differential evolution, vol. 143, Springer, 2008.

[26] R. Storn and K. Price, "Differential evolution--a simple and efficient heuristic for global optimization over continuous spaces," Journal of global optimization, vol. 11, no. 4, pp. 341--359, 1997.

[27] S. Das and P. N. Suganthan, "Differential evolution: A survey of the state-of-the-art," IEEE transactions on evolutionary computation, vol. 15, no. 1, pp. 4--31, 2010.

[28] X. Li, S. Ma and J. Hu, "Multi-search differential evolution algorithm," Applied Intelligence, vol. 47, no. 1, pp. 231--256, 2017.

[29] V. Feoktistov, "Differential evolution," Springer, 2006.

[30] A. Asilian Bidgoli, H. Ebrahimpour-Komleh and S. Rahnamayan, "Reference-point-based multi-objective optimization algorithm with opposition-based voting scheme for multi-label feature selection," Information Sciences, vol. 547, pp. 1-17, 2020.

[31] R. Caruana and D. Freitag, "Greedy attribute selection," in Machine Learning Proceedings 1994, Elsevier, 1994, pp. 28--36.

[32] W. L. G. Koontz, P. M. Narendra and K. Fukunaga,, "A branch and bound clustering algorithm," IEEE Transactions on Computers, vol. 100, no. 9, pp. 908--915, 1975.

[33] A. Kumar, S. Vembu, A. K. Menon and C. Elkan, "Beam search algorithms for multilabel learning," Machine learning, vol. 92, no. 1, pp. 65--89, 2013.

[34] M. A. Hall and L. A. Smith, "Feature subset selection: a correlation based filter approach," Springer, 1997.

[35] Z. Zhu, Y.-S. Ong and M. Dash, "Wrapper--filter feature selection algorithm using a memetic framework," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 37, no. 1, pp. 70--76, 2007.

[36] C. W. Ahn, E. Kim, H.-T. Kim, D.-H. Lim and J. An, "A hybrid multiobjective evolutionary algorithm: Striking a balance with local search," Mathematical and Computer Modelling, vol. 52, no. 11-12, pp. 2048--2059, 2010.

[37] G. Pampara, A. P. Engelbrecht and N. Franken, "Binary differential evolution," in 2006 IEEE International Conference on Evolutionary Computation, IEEE, 2006, pp. 1873--1879.

[38] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," Pattern recognition, vol. 40, no. 7, pp. 2038--2048, 2038--2048.

[39] G. Tsoumakas, I. Katakis and I. Vlahavas, "Mining multi-label data," in Data mining and knowledge discovery handbook, Springer, 2009, pp. 667--685.

[40] Y. Zhang, D.-w. Gong,, X.-y. Sun and Y.-n. Guo, "A PSO-based multi-objective multi-label feature selection method in classificatio," Scientific reports, vol. 7, no. 1, p. 376, 2017.

[41] F. Tan, "Improving feature selection techniques for machine learning," tan2007improving, 2007.

[42] R. Kalaiselvi., P. Premadevi and M. Hamsathvani, "Survey on Semi-Supervised Feature Selection in Datamining," International Journal of Computer Science and Information Technologies, pp. 7114-7117, 2014.

[43] J. R. Koza, "Genetic programming," Citeseer, 1997.

[44] C. W. Ahn, E. Kim, H.-T. Kim, D.-H. Lim and J. An, "A hybrid multiobjective evolutionary algorithm: Striking a balance with local search," Mathematical and Computer Modelling, vol. 52, pp. 11-12, 2010.