

SUFFUSE: Simultaneous Fuzzy-Rough Feature-Sample Selection

Javad Rahimipour Anaraki

Department of Computer Science, Memorial University of Newfoundland, St. John's, NL, Canada

Email: jra066@mun.ca

Saeed Samet

Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada

Email: saeed.samet@med.mun.ca

Jeon-Hyun Lee and Chang-Wook Ahn

Department of Computer Engineering, Sungkyunkwan University, Suwon, South Korea

Email: {ljh08375, cwan}@skku.edu

Abstract—One of the most successful tools for modeling and dealing with uncertainty is Rough Set Theory. Based on this theory several Feature Selection methods have been proposed. As an extension, Fuzzy-Rough set has been introduced to deal with vagueness of both discrete and continuous data in Feature and Sample Selection methods. However, both Fuzzy-Rough Sample Selection and Simultaneous Fuzzy-Rough Feature-Sample Selection are investigated by few. This paper proposes a novel Simultaneous Fuzzy-Rough Feature-Sample Selection method based on Shuffled Frog Leaping Algorithm. The effectiveness of proposed method demonstrated and compared through its performance resulting from nine conventional as well as an improved mGP classifiers over fifteen UCI datasets. This work is also applied to a real world classification problem of noisy Functional Near-Infrared Spectroscopy neural signals. Experimental results show meaningful increase in classification accuracy, and decrease in dataset size according to non-parametric statistical analysis.

Index Terms—fuzzy-rough sets, simultaneous fuzzy-rough feature-sample selection, feature selection, sample selection

I. INTRODUCTION

The amount of raw data produced daily is much higher than the information extracted from them. Therefore, more cost and time are needed to process, save and maintain those data for later processing. Many problems in machine learning, data mining and pattern recognition involve big datasets. A high dimensional data in terms of number of features and samples needs huge effort to be processed. Therefore, Feature Selection (FS) methods can effectively reduce the size of datasets in one direction by selecting significant columns. These methods select most-informative features which are highly correlated to the outcome and loosely depended on other features in favor of minimizing further processing. Since the size of

datasets can also be decreased in terms of samples, Sample Selection (SS) methods have emerged to reduce size of datasets by removing irrelevant samples. Therefore, by employing FS and SS methods, datasets' dimensions can be lowered and further processing can be done more efficiently.

Raman and Ioerger [1], proposed a feature selection, and sample selection method. The former eliminates irrelevant features using a sequential search on feature space to maintain a balance between local hypotheses and prediction accuracy. The latter, uses Hamming distance to filter out samples, and naive bayes classifier to predict class labels based on the selected samples. Then each method has been applied on a same dataset to perform two dimensional selection. Rozsypal and Kubat [2] have introduced simultaneous feature-sample selection based on genetic algorithm with the aim of increasing classification accuracy and decreasing the number of selected features and samples. Chromosome designation has been established to accommodate two subsets of integers, each representing selected features and samples. The fitness function has been designed based on the number of retained features and samples, and also the number of misclassified examples.

Rough Set Theory (RST) [3] is one of the most successful mathematical tools in FS [4] which nowadays receives much of attention in SS. This theory has been applied to many real-world applications [5] since it allows minimal representation of the data while sustaining semantic of data with no human provided information. However, RST is only decent to deal with crisp and discrete data; therefore, a combination of RST and Fuzzy Set has been proposed in [6] to overcome this inadequacy. Stand on Fuzzy-Rough set (FR), some research has been conducted in FS [5], [7] and SS [8], and very few works have been done in simultaneous Fuzzy-Rough feature-sample selection [9].

Genetic Programming (GP) is capable of finding hidden relations in data and presenting them in terms of

mathematical functions [10]. This method has been widely used in tough classification problems and investigated by many researchers to develop classifiers for two- and multi-class problems. In [11], An et al. designed a new multi-tree GP (mGP) classifier by modifying crossover and mutation operators.

In this paper we have proposed a Simultaneous Fuzzy-Rough Feature-Sample Selection method (SUFFUSE) based on Shuffled Frog Leaping Algorithm (SFLA) [12], as well as an improved mGP. The rest of the paper is organized as follows: Section II describes preliminaries of FR, SFLA and mGP. Section III presents the proposed methods, SUFFUSE, and improved mGP. In Section IV, experimental results are shown. Application to noisy Functional Near-Infrared Spectroscopy (fNIRS) neural signals dataset and conclusion are placed in Sections V and VI, respectively.

II. PRELIMINARIES

Two fundamental components of feature, sample and feature-sample selections are Evaluation Metric and Search Method. In this work the former is based on Fuzzy-Rough Positive Region (FRPR), and the latter uses SFLA. Finally, an improved mGP classifier analyzes and builds data models to figure out capabilities of proposed methods. All basics are categorized as follows:

A. Evaluation Metric: Fuzzy-Rough Positive Region (FRPR)

In RST, data are organized in decision table. Let \mathbb{U} be the universe of discourse, R be the equivalence relation on \mathbb{U} , so approximation space is shown by (\mathbb{U}, R) . Let X be a subset of \mathbb{U} and P be a subset of A , which is a non-empty set of attributes. Approximating X using RST is done by means of lower and upper approximations. Objects in lower approximation ($\underline{P}X$) are the ones which are surely classified in X regarding the attributes in P . Upper approximation of X with regards to ($\overline{P}X$) contains objects which are possibly classified in X regarding the attributes in P . Based on these approximations, three different regions are defined as positive, negative and boundary that are shown by Equations 1, 2, and 3, respectively [13].

$$POS_p(Q) = \bigcup_{x \in \mathbb{U}/Q} \underline{P}X \quad (1)$$

$$NEG_p(Q) = \mathbb{U} - \bigcup_{x \in \mathbb{U}/Q} \overline{P}X \quad (2)$$

$$BND_p(Q) = \bigcup_{x \in \mathbb{U}/Q} \overline{P}X - \bigcup_{x \in \mathbb{U}/Q} \underline{P}X \quad (3)$$

B. Search Method: Shuffled Frog Leaping Algorithm (SFLA)

SFLA is a meta-heuristic search algorithm which is inspired by real frogs. The search starts by generating population over the search space. Then the population is divided into subpopulations called memplexes which are able to evolve separately. In each memplex, frogs participate in meme evolution due to infection by other

frogs. By meme evolution, each frog's performance is increased referring to the best frog in each memplex and poor ideas evolve toward new ideas. The frogs are infected both by best frogs in their memplex and the entire population. After specified number of evolutions, memplexes are mixed together and new memplexes are emerged by shuffling the population. This process migrates frogs to different regions of the swamp. Therefore they can share their experiences with other frogs. A modified binary form of SFLA has been applied to the problem of simultaneous selection.

C. Multi-tree Genetic Programming Classifier

In [11], individuals of a c -class problem are generated randomly with $c-1$ trees. Then all the individuals are evaluated using fitness function and top N individuals are selected based on τ -wise tournament selection. The classifier continues by applying crossover and mutation for generating new individuals. Then, the worst individuals are substituted with the newly generated best ones and the classifier continues until the stopping criterion is satisfied.

III. PROPOSED METHODS

A. Evaluation Metric: Fuzzy-Rough Positive Region (FRPR)

The FRSS [14] is based on FRPR as an evaluation measure, and SFLA as a search method. The length of each frog in population is equal to the number of samples in the dataset where their presence and absence are depicted by one and zero, respectively. As SFLA generates initial population, related dataset formations are constructed referring to each individual frog. Then, fitness of all frogs is calculated using FRPR as shown in Equation 1. Each frog's formation is shown in Fig. 1, where $s_j \in \{0,1\}$ and j is number of samples of dataset.

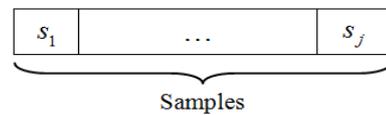


Figure 1. Each frog's formation in FRSS

Table I represents a dataset with two features and seven samples. Based on the table, a possible frog's formation and related dataset is presented in Fig. 2 and Table II, respectively.

TABLE I. A DECISION TABLE

Samples	Features		Class
	f_1	f_2	
s_1	0.65	0.59	Yes
s_2	0.93	0.88	No
s_3	0.48	0.73	No
s_4	0.70	0.43	Yes
s_5	0.49	0.76	No
s_6	0.05	0.23	Yes
s_7	0.54	0.60	No

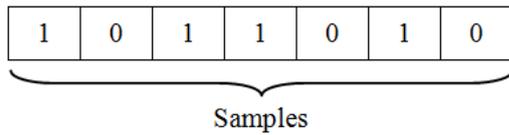


Figure 2. A possible frog's formation in FRSS

TABLE II. RESULTING DATASET REFERRING TO POSSIBLE FROG'S FORMATION

Samples	Features		Class
	f_1	f_2	
s_1	0.65	0.59	Yes
s_3	0.48	0.73	No
s_4	0.70	0.43	Yes
s_6	0.05	0.23	Yes

The SFLA continues until the stopping criterion, which is either maximum iteration or gaining the highest FRPR value, is satisfied. Feature and sample selections can be done either in order or simultaneously. Applying either feature or sample selection beforehand might have a huge effect on the final performance. Even if the first operation has a great efficiency, the outcome would be less desirable since each method acts independently. Thus, simultaneous selection would increase the quality of the outcome by considering ongoing two dimensional selections together.

At the starting point, a population consists of frogs with the length proportional to the number of features and samples is generated. Fig. 3 depicts each frog's formation. In this formation, each bit's value and position show the presence or absence of either a feature or sample that specifies the final structure of the extracted dataset from the original one, where $f_i, s_j \in \{0,1\}$, and i and j are the number of features and samples in each dataset, respectively.

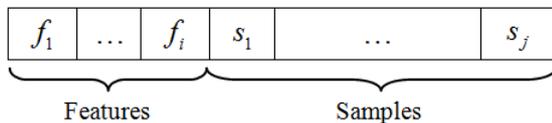


Figure 3. Each frog's formation with features and samples individuals

Fig. 4 shows the formation of each frog for aforementioned table. Since the first position is equal to one, therefore the proportional feature should participate in the new dataset. Similarly, those samples which corresponding bits are equal to one will form the output dataset.

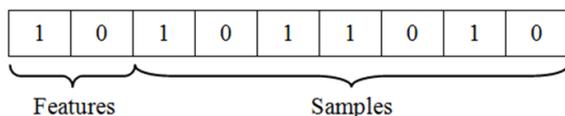


Figure 4. Possible frog's formation with features and samples individuals

Table III demonstrates the final dataset formation based on the original dataset in Table I and by referring to presence and absence of both features and samples in Fig. 4.

TABLE III. RESULTING DATASET OF POSSIBLE FROG'S FORMATION WITH FEATURE AND SAMPLES INDIVIDUALS

Samples	Features	Class
	f_1	
s_1	0.65	Yes
s_3	0.48	No
s_4	0.70	Yes
s_6	0.05	Yes

Since Rough Set could not deal with continuous values, the original Fuzzy-Rough set has been proposed by Dubois and Prade [6] to elude this lack. Later, a new definition was introduced by Radzikowska and Kerre [15] and then Shen and Jensen [5] modified the original definitions. In [7], final definitions of X -lower and X -upper approximations based on Fuzzy-Rough sets are presented as in Equations 4 and 5, where I is Łukasiewicz Fuzzy implicator, which is defined by $\min(1-x+y, 1)$ and T is Łukasiewicz Fuzzy t -norm, which is shown by $\max(x+y-1, 0)$.

$$\mu_{R_p X}(x) = \inf_{y \in U} I\{\eta_{R_p}(x, y), \mu_X(y)\} \quad (4)$$

$$\mu_{R_p X}^-(x) = \sup_{y \in U} T\{\eta_{R_p}(x, y), \mu_X(y)\} \quad (5)$$

$$\eta_{R_p}(x, y) = \bigcap_{a \in P} \{\eta_{R_a}(x, y)\} \quad (6)$$

In Equation 6, R_p is Fuzzy similarity relation and $\eta_{R_a}(x, y)$ is the degree of similarity between objects x and y , considering feature a [7]. A fuzzy similarity relation is shown in Equation 7, where σ_a the variance of feature a . Positive region in RST is defined as a union of lower approximations. Referring to extension principle [7], the membership of object x to a FRPR is defined in Equation 8.

$$\eta_{R_p}(x, y) = \max \left(\min \left(\frac{(a(y) - (a(x) - \sigma_a))}{\sigma_a}, \frac{((a(x) + \sigma_a) - a(y))}{\sigma_a} \right), 0 \right), \quad (7)$$

$$\mu_{POS_p(Q)}(x) = \sup_{x \in U/Q} \mu_{PX}(x). \quad (8)$$

If the equivalence class of which x belongs to, does not belong to the positive region, obviously x will not be a part of the positive region. Equation 8 is the fitness function of the search algorithm which measures the significance of the selected features-samples subset [8]. Finally, SFLA evaluates each final dataset corresponding to each frog by calculating FRPR. The best frog in each memplex infects other frogs, and as a result the whole

population moves toward the final goal, which is finding the lowest number of features and samples with the highest fitness value.

In the very first point, dataset is loaded and the number of its features and samples, specifies all parameters of SFLA. In SUFFUSE, SFLA and FRPR collaborate to find the best feature-sample subsets. Then the classification methods, which involve conventional classifiers as well as improved mGP, classify the datasets. The value of division of classification accuracies' mean by summation of the number of selected features and samples is calculated and compared with the results of the FRFS and FRSS. Fig. 5 shows the overall workflow of SUFFUSE.

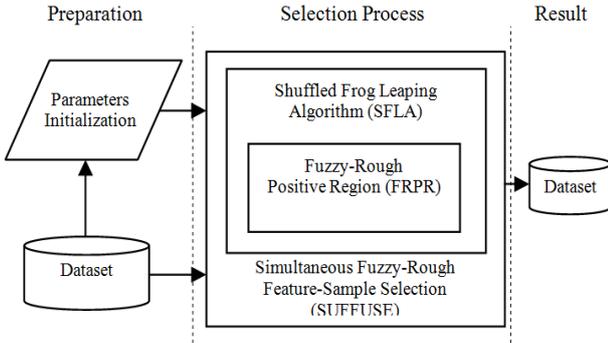


Figure 5. Simultaneous fuzzy-rough feature-sample selection workflow

B. Improved Multi-tree GP Classifier

This method is robust to noise since the voting system is inspired by honey bee migration that is less sensitive to noise. Fig. 6 describes the method. Fig. 7 shows the representation of each individual with its equation referring to the number of trees (m), which is specified by user and number of classes. For instance a three-class dataset would have two classifiers. In the proposed classifier four main parts have been modified as follows:

1) Fitness function

The new multi-modal fitness function is based on classification accuracy and variance. The goal is to maximize the classification margin, while decreasing intraclass similarities using Equation 9. Equation 10 calculates the centroid of each class to be used in Equation 9. Therefore fitness function is determined by the summation of Classification Accuracy (CA) and distance function as shown in Equation 11.

$$Distance = \sum_{i \in classA} \frac{|T_{m,x}^{(i)} - CentroidA|}{|Max(classA) - Min(classA)|} \quad (9)$$

$$CentroidA = \frac{\sum_{i \in classA} T_{m,x}^{(i)}}{\|classA\|} \quad (10)$$

$$Fitness = CA + Distance \quad (11)$$

2) Selection strategy

The selection process has three stages. At first top 3% of previous generation is selected to construct new generation, and if there were more than 3% individuals with highest ranking, top 10% will be selected. However, if two or more classifiers have the same fitness value, all of them will be used in the next generation. Then 65% of the new generation is selected based on pair-wise tournament selection. Finally the rest of the individuals will be randomly generated.

3) Mutation

The mutation process contains three policies for the internal mutation and one policy for the external one. In the internal mutation, a node can add, remove or exchange children. Thus the whole tree is reconstructed in the external mutation as Fig. 8 shows.

4) Crossover

The crossover is divided into the internal and external crossovers. In the former, trees are selected in each individual based on the internal crossover probability parameter. The latter is based on one-point crossover and it takes place among any trees by considering external crossover probability. Fig. 9 describes the crossover strategy.

IV. EXPERIMENTAL RESULTS

Fifteen UCI datasets [16] have been selected to measure the performance of the proposed methods. Parameter selection for SFLA has been formulated based on the number of features $|F|$, samples $|S|$ and feature-samples $|FS|$ using trial and error method. The results are mentioned in Table IV, in which m is the number of memplexes, n is the number of frogs in each memplex, N is the number of evolution processes, q is the number of frogs which are selected randomly from n frogs to form a memplex and S_{max} is the maximum step size allowed to be adopted after infection.

TABLE IV. SFLA PARAMETERS FOR FRFS, FRSS AND SUFFUSE

Method	m	n	N	q	S_{max}
FRFS	$ F \times 2.20$	$ F \times 0.70$	$ F \times 0.50$	$ F \times 0.45$	$ F \times 0.50$
FRSS	$ S \times 0.02$	$ S \times 0.01$	$ S \times 0.01$	$ S \times 0.50$	$ S \times 0.50$
SUFFUSE	$ FS \times 0.02$	$ FS \times 0.01$	$ FS \times 0.01$	$ FS \times 0.50$	$ FS \times 0.50$

Each algorithm runs ten times over the datasets and information-rich features, samples, and features-samples are selected by FRFS, FRSS and SUFFUSE, respectively. The best results over all iterations are chosen and presented in Table V in terms of the number of selected features and samples and overall model size. The number of samples are fix in the results of FRFS as it only selects features, whereas, the number of features are constant for FRSS since it just affects samples. The mean of ranking for each method is calculated and shown in Table VI, in which SUFFUSE performs 51% and 31% better than FRSS and FRFS, respectively.

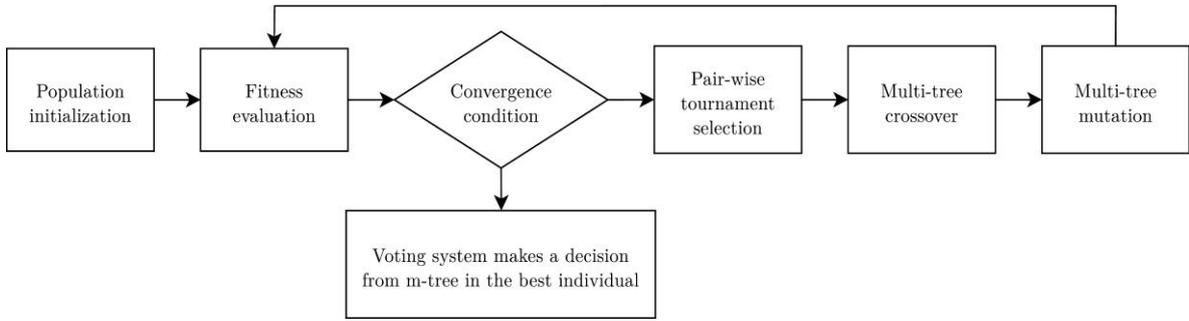


Figure 6. Improved Multi-Tree GP

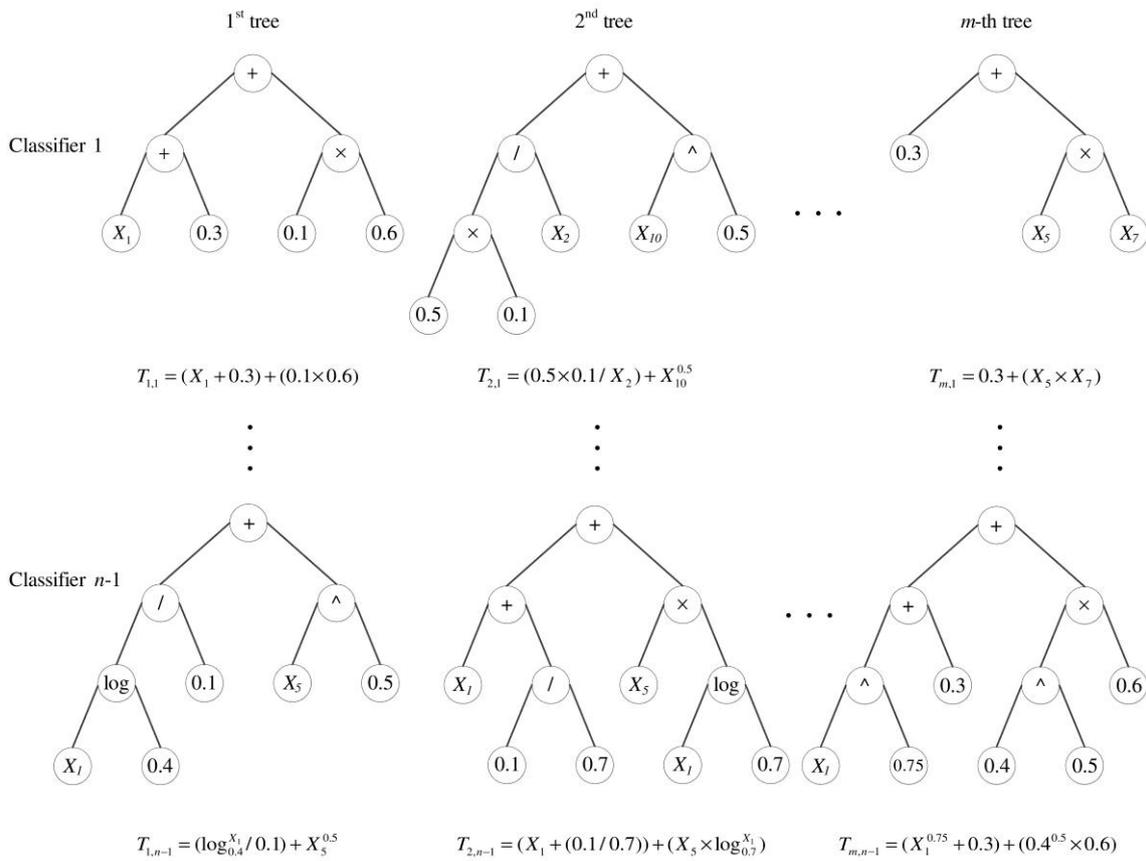


Figure 7. An individual with $m \times (n - 1)$ trees

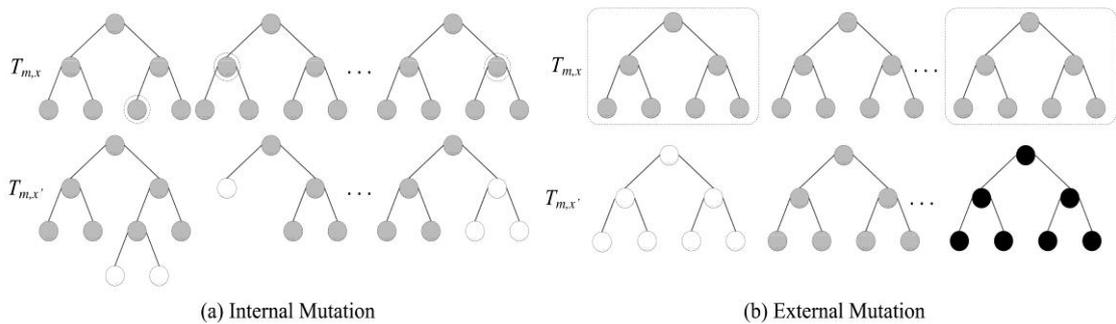


Figure 8. Proposed mutation operator

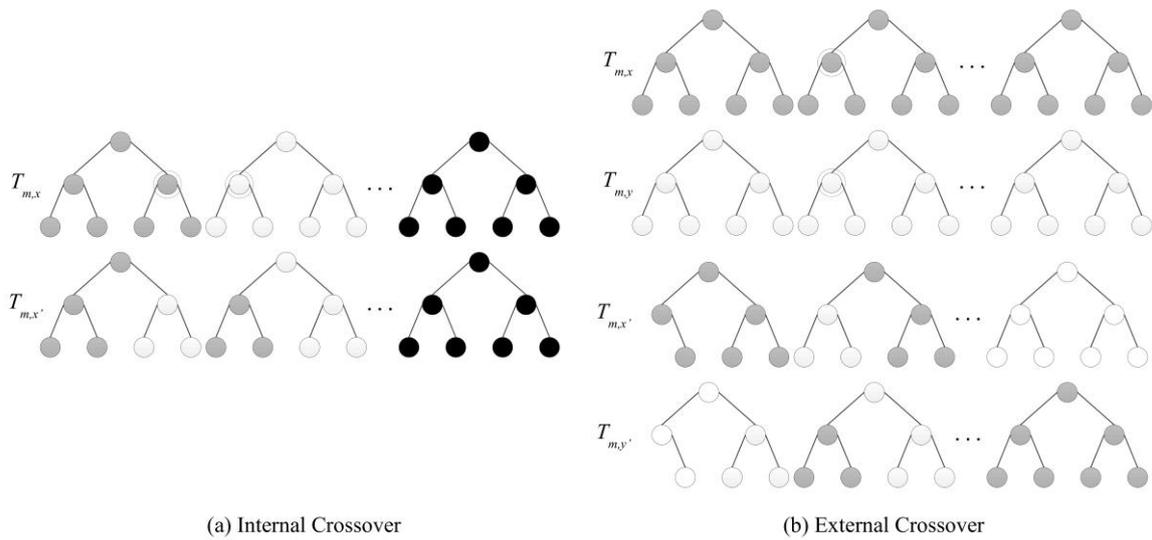


Figure 9. Proposed crossover operator

TABLE V. RESULTING REDUCTION AND MODEL SIZE BY FRFS, FRSS, & SUFFUSE

Dataset	Unreduced			FRFS			FRSS			SUFFUSE		
	S	F	Size	S	F	Size	S	F	Size	S	F	Size
Blood Trans.	748	4	2992	748	3	2244	264	4	1056	372	2	744
Breast Cancer	683	9	6147	683	7	4781	256	9	2304	357	6	2142
Breast Tissue	106	9	954	106	6	636	70	9	630	51	5	255
Cleveland	297	13	3861	297	7	2079	199	13	2587	108	2	216
Glass	214	9	1926	214	6	1284	144	9	1296	130	7	910
Heart	270	13	3510	270	7	1890	156	13	2028	166	9	1494
Ionosphere	351	33	11583	351	7	2457	115	33	3795	203	12	2436
Lung Cancer	27	56	1512	27	3	81	20	56	1120	10	25	250
Olitos	120	25	3000	120	5	600	81	25	2025	74	12	888
Parkinson	195	22	4290	195	6	1170	130	22	2860	111	10	1110
Pima Indian Dia.	768	8	6144	768	6	4608	256	8	2048	270	3	810
Sonar	208	60	12480	208	6	1248	140	60	8400	128	34	4352
Soybean	47	35	1645	47	2	94	31	35	1085	30	20	600
SPECTF Heart	80	44	3520	80	6	480	55	44	2420	38	29	1102
Wine	178	13	2314	178	5	890	115	13	1495	97	7	679

TABLE VI. RANKING OF FRFS, FRSS AND SUFFUSE BASED ON MODEL SIZE

Dataset	FRFS	FRSS	SUFFUSE
Blood Trans.	3	2	1
Breast Cancer	3	2	1
Breast Tissue	3	2	1
Cleveland	2	3	1
Glass	2	3	1
Heart	2	3	1
Ionosphere	2	3	1
Lung Cancer	1	3	2
Olitos	1	3	2
Parkinson	2	3	1
Pima Indian Dia.	3	2	1
Sonar	1	3	2
Soybean	1	3	2
SPECTF Heart	1	3	2
Wine	2	3	1
Mean	1.93	2.73	1.33

Table VII shows mean of the classification results for conventional classifiers (such as PART, JRip, Naive Bayes, Bayes Net, J48, BFTree, FT, NBTree and RBFNetwork, which are implemented in WEKA [17]) as well as improved mGP, and Fig. 10 presents the classification workflow process. The mean of accuracies

of conventional classifiers for our proposed method shows 3.55% increase comparing to unreduced datasets, as well as 2.55% and 1.58% improvement comparing with FRFS and FRSS, respectively. Whereas, the result of improved mGP for SUFFUSE shows 5.58%, 4.10% and 1.23% increase comparing to the results of improved GP for unreduced datasets, FRFS and FRSS. As the initial experiment results show, the fusion of SUFFUSE with improved mGP produces the simplest model which leads to the higher classification accuracies.

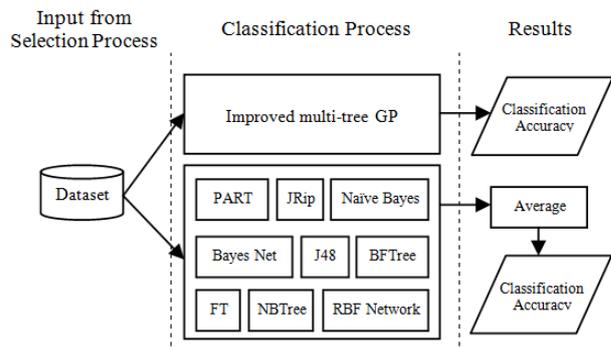


Figure 10. Classification workflow

TABLE VII. AVERAGE CLASSIFICATION ACCURACIES (%) OF CONVENTIONAL CLASSIFIERS (PART, JRIP, NAIVE BAYES, BAYES NET, J48, BFTREE, FT, NBTREE AND RBFNETWORK) AND IMPROVED MGP BASED ON FRFS, FRSS AND SUFFUSE RESULTS

Dataset	Unreduced		FRFS		FRSS		SUFFUSE	
	Conv.	mGP	Conv.	mGP	Conv.	mGP	Conv.	mGP
Blood Trans.	77.20	79.95	77.30	79.14	78.87	82.26	79.24	80.11
Breast Canc	96.18	96.93	96.40	97.95	95.14	98.05	96.70	98.04
Breast Tissu	66.46	69.81	68.66	73.58	65.56	77.14	69.93	82.35
Cleveland	50.13	41.28	50.88	52.53	52.26	57.79	55.86	57.41
Glass	61.89	53.74	66.87	70.09	66.82	71.53	65.47	71.54
Heart	79.55	82.96	73.61	81.85	80.56	84.62	84.54	86.14
Ionosphere	89.68	91.74	89.55	91.19	85.70	87.83	90.15	92.12
Lung Cancer	55.56	74.07	57.61	77.78	60.56	75.00	64.44	80.00
Olitos	69.81	75.00	69.07	72.73	77.23	76.54	73.87	79.73
Parkinson	82.34	88.72	85.64	87.69	84.10	90.77	84.99	92.79
Pima Indian Dia.	75.00	75.42	75.61	75.42	75.82	78.91	76.01	78.89
Sonar	67.47	78.85	74.73	79.81	74.13	86.43	73.61	83.59
Soybean	98.58	100.00	90.54	91.49	91.76	100.00	92.96	100.00
SPECTF Heart	73.06	78.75	73.47	78.75	73.74	80.36	78.65	78.65
Wine	85.52	93.82	93.51	93.26	95.70	99.16	95.30	97.94
Mean	75.23	78.74	76.23	80.22	77.20	83.09	78.78	84.32

A Non-parametric statistical analysis [18] is employed to compare the overall performance of each method based on the results of improved mGP in Table VII. The average ranks obtained by each method in the Friedman test are presented in Table VIII. As shown, SUFFUSE has gained the lowest ranking, which proves the effectiveness of the proposed method. Friedman statistic (distributed according to chi-square with 3 degrees of freedom) is 27.56, and p-value computed by Friedman test is $4e-6$. By referring to the post hoc comparison results in Table IX, the probability of FRFS and Unreduced to perform better than SUFFUSE is less $(5e-3)\%$ and $(8e-3)\%$, respectively. Also, the probability of FRSS to outrun SUFFUSE is less than 48%.

TABLE VIII. AVERAGE RANKINGS OF THE ALGORITHMS (FRIEDMAN)

Algorithm	Ranking
SUFFUSE	1.4667
FRSS	1.8000
Unreduced	3.3333
FRFS	3.4000

TABLE IX. POST HOC COMPARISON TABLE FOR $\alpha = 0:05$ (FRIEDMAN)

i	algorithm	$z = (r_0 - R_i) / SE$	p
3	FRFS	4.101219	0.000041
2	Unreduced	3.959798	0.000075
1	FRSS	0.707107	0.4795

V. APPLICATION TO FUNCTIONAL NEAR-INFRARED SPECTROSCOPY (FNIRS) NEURAL SIGNALS

To show the appropriateness of the proposed methods, a real world dataset called Neural Signal is used as a benchmark dataset. The neural signal acquisition has been done by a multi-channel optical brain imaging system (fNIR-300) and the levels of oxy-, deoxy- and total-haemoglobin have been specified using 16 signal channels at 2 Hz sampling rate.

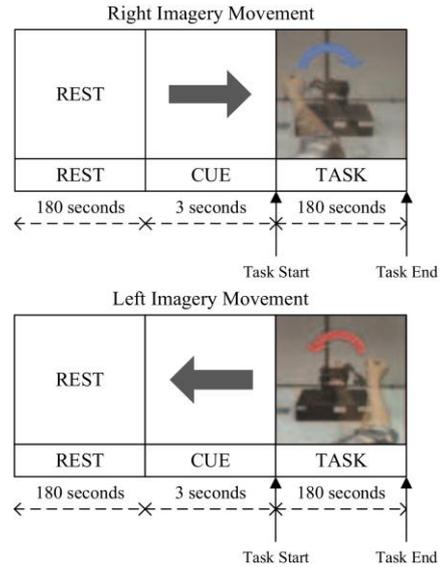


Figure 11. Experimental scenarios for acquiring fNIRS neural signals

The signals are collected through the optical fibers, which are attached to the pre-frontal cortex. As Fig. 11 shows, two cognitive activities of rest \rightarrow right imagery movement and rest \rightarrow left imagery movement have been sampled in a dataset with three classes, rest, right and left. The dataset has 280 samples and 45 features. Table X shows the average classification accuracies of applying FRFS, FRSS and SUFFUSE. It can be seen that SUFFUSE ends to higher classification accuracy comparing to unreduced, FRFS and FRSS, both by using conventional and improved mGP. The proposed classification system results 5.83% higher than the other classifiers.

TABLE X. AVERAGE CLASSIFICATION ACCURACIES (%) OF CONVENTIONAL CLASSIFIERS (PART, JRIP, NAIVE BAYES, BAYES NET, J48, BFTREE, FT, NBTREE AND RBFNETWORK) & IMPROVED MGP FOR UNREDUCED & REDUCED NEURAL SIGNAL DATASET USING FRFS, FRSS & SUFFUSE

Dataset	Unreduced		FRFS		FRSS		SUFFUSE	
	Conv.	mGP	Conv.	mGP	Conv.	mGP	Conv.	mGP
Neural Signal	75.40	82.86	78.02	81.79	75.04	81.34	83.68	89.51

VI. CONCLUSION

This paper proposes a novel Simultaneous Fuzzy-Rough Feature-Sample Selection (SUFFUSE), and an improved multi-tree GP (mGP). The SUFFUSE selects features and samples simultaneously by coding both in a single frog of SFLA, and use Fuzzy-Rough Positive Region (FRPR) as fitness function to evaluate selected subsets. An improved mGP classifier, classifies the results of proposed methods based on the new selection strategy, fitness function, mutation and crossover operators. Finally, the experimental results of SUFFUSE, Fuzzy-Rough Feature Selection (FRFS) and Fuzzy-Rough Feature Selection (FRSS) on fifteen UCI datasets show the effectiveness of the proposed methods, both in terms of classification accuracy and models size. As a real-world application, the proposed methods handle fNIRS neural signal dataset. It can be seen from the

results that SUFFUSE and mGP have a great impact on classification accuracy comparing to independent feature and sample selections. As a future work, we are so excited to apply improved version of SFLA, and perform broad comparisons among different evolutionary algorithms.

ACKNOWLEDGMENT

This work has been partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Research & Development Corporation of Newfoundland and Labrador (RDC).

REFERENCES

- [1] B. Raman and T. R. Iroger, "Enhancing learning using feature and example selection," Texas A&M University, College Station, TX, USA, 2003.
- [2] A. Rozsypal and M. Kubat, "Selecting representative examples and attributes by a genetic algorithm," *Intelligent Data Analysis*, vol. 7, no. 4, pp. 291–304, 2003.
- [3] Z. Pawlak, "Rough sets," *International Journal of Computer & Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.
- [4] J. R. Anaraki and M. Eftekhari, "Rough set based feature selection: A review," in *Proc. 2013 5th Conference on Information and Knowledge Technology*, May 2013, pp. 301–306.
- [5] Q. Shen and R. Jensen, "Selecting informative features with fuzzyrough sets and its application for complex systems monitoring," *Pattern Recognition*, vol. 37, no. 7, pp. 1351–1363, 2004.
- [6] D. Dubois and H. Prade, "Putting rough sets and fuzzy sets together," in *Intelligent Decision Support, ser. Theory and Decision Library*, R. Slowinski, Ed. Springer Netherlands, vol. 11, 1992, pp. 203–232.
- [7] R. Jensen and Q. Shen, "New approaches to fuzzy-rough feature selection," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 4, pp. 824–838, Aug 2009.
- [8] R. Jensen and C. Cornelis, "Fuzzy-rough instance selection," in *Proc. 2010 IEEE International Conference on Fuzzy Systems (FUZZ)*, July 2010, pp. 1–7.
- [9] N. Mac Parthalain and R. Jensen, "Simultaneous feature and instance selection using fuzzy-rough bireducts," in *Proc. 2013 IEEE International Conference on Fuzzy Systems (FUZZ)*, July 2013, pp. 1–8.
- [10] D. P. Muni, N. R. Pal, and J. Das, "A novel approach to design classifiers using genetic programming," *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 2, pp. 183–196, April 2004.
- [11] J. An, J. Lee, and C. Ahn, "An efficient gp approach to recognizing cognitive tasks from fmirs neural signals," *Science China Information Sciences*, vol. 56, no. 10, pp. 1–7, 2013.
- [12] M. Eusuff, K. Lansey, and F. Pasha, "Shuffled frog-leaping algorithm: a memetic meta-heuristic for discrete optimization," *Engineering Optimization*, vol. 38, no. 2, pp. 129–154, 2006.
- [13] J. Komorowski, Z. Pawlak, L. Polkowski, and A. Skowron, "Rough sets: A tutorial," in *Rough-Fuzzy Hybridization: A New Trend in Decision Making*, S. K. Pal and A. Skowron, Eds.

Secaucus, NJ, USA: Springer- Verlag New York, Inc., 1998, pp. 3–98.

- [14] J. R. Anaraki and C. Ahn, "Fuzzy-rough sample selection," in *Proc. 2014 4th International Workshop on Computer Science and Engineering (WCSE)*, August 2014, pp. 22–29.
- [15] A. M. Radzikowska and E. E. Kerre, "A comparative study of fuzzy rough sets," *Fuzzy Sets and Systems*, vol. 126, no. 2, pp. 137–155, 2002.
- [16] M. Lichman. (2013). UCI machine learning repository. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [17] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, Nov. 2009.
- [18] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, and S. García, "Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework," *Multiple-Valued Logic and Soft Computing*, vol. 17, no. 2-3, pp. 255–287, 2011.

Javad Rahimpour Anaraki is a member of e-Health Research Unit. He received the B.E. degree in Computer Software from the Shahid Ashrafi Esfahani University, Esfahan, Iran, and the M.E. degrees in Computer Engineering from the Shahid Bahonar University of Kerman, Kerman, Iran, in 2007 and 2012, respectively. In 2014, he joined the Department of Computer Science at Memorial University of Newfoundland as Ph.D. student. His current research interests include fuzzy-rough sets, feature and sample selection, and privacy-preserving methods.

Saeed Samet is a member of the Faculty of Medicine as an Assistant Professor at the e-Health Research Unit. Prior to that, he was a Post-Doctoral Fellow at the e-Health Information Laboratory at the Children Hospital of Eastern Ontario Research Institute (CHEO-RI) from 2010 to 2012. During his post-doc fellowship he designed and developed several secure protocols for various health applications. He received his Ph.D. in Computer Science from the University of Ottawa in 2010 and his thesis was "Privacy-Preserving Data Mining", in which he has proposed and designed different protocols on privacy-preserving methods for standard data mining and machine learning techniques.

Jeon-Hyun Lee is a member of Sungkyunkwan Evolutionary Algorithm Lab (SEAL) at Sungkyunkwan University (SKKU), Republic of Korea. In 2010, he joined the Department of Computer Science and Engineering at Sungkyunkwan University as Ph.D. student. His current research interests include genetic algorithm and swarm robotics.

Chang Wook Ahn received Ph.D. degree at the Department of Information and Communications, Gwangju Institute of Science and Technology (GIST), Republic of Korea, in 2005. He was a Visiting Scholar at the Illinois Genetic Algorithms Laboratory (IlligAL), University of Illinois at Urbana Champaign (UIUC), in 2003. From 2005 to 2007, he was a research staff member at the 4G Communication Research Group, Samsung Advanced Institute of Technology. From 2007 to 2008, he was a Research Professor in GIST. He is currently an Associate Professor in Sungkyunkwan University (SKKU), Republic of Korea. His research interests include genetic algorithms, multiobjective optimization, and the applications of evolutionary techniques to wireless networks.