Int. J. Advance Soft Compu. Appl, Vol. 10, No. 1, March 2018
ISSN 2074-8523

Predictive based Hybrid Ranker to Yield Significant Features in Writer Identification

Intan Ermahani A. Jalil[1], Siti Mariyam Shamsuddin[2], Azah Kamilah Muda[1]
Mohd Sanusi Azmi[1], and Ummi Raba'ah Hashim[1]

[1]Computational Intelligence and Technologies Lab
Faculty of Information and Communication Technology
Universiti Teknikal Malaysia Melaka (UTeM), Melaka Malaysia
e-mail: ermahani@utem.edu.my

[2]UTM Big Data Centre
Ibnu Sinar Institute for Scientific and Industrial Research
Universiti Teknologi Malaysia (UTM), Johor, Malaysia
e-mail: mariyam@utm.my

Abstract

The contribution of writer identification (WI) towards personal identification in biometrics traits is known because it is easily accessible, cheaper, more reliable and acceptable as compared to other methods such as personal identification based DNA, iris and fingerprint. However, the production of high dimensional datasets has resulted into too many irrelevant or redundant features. These unnecessary features increase the size of the search space and decrease the identification performance. The main problem is to identify the most significant features and select the best subset of features that can precisely predict the authors. Therefore, this study proposed the hybridization of GRA Features Ranking and Feature Subset Selection (GRAFeSS) to develop the best subsets of highest ranking features and developed discretization model with the hybrid method (Dis-GRAFeSS) to improve classification accuracy. Experimental results showed that the methods improved the performance accuracy in identifying the authorship of features based ranking invariant discretization by substantially reducing redundant features.

Keywords: *Features Ranking, Grey Relational Analysis, Predictive, Significant, Writer Identification*

1 Introduction

The research on the capability of any methods to predict the importance or relevancy of any features or attributes is currently an expanding challenge in the area of machine learning [5, 28]. Whereby most of the fields of study that relates to machine learning especially when handling with huge amount of data as such medical data [6, 11, 25,], stock exchange prediction [12], software fault or effort prediction [26, 31], traffic data [34] and writer identification [2, 23] are prone to find the most simplest and fastest way to retrieve significant information and eliminate unnecessary factor.

The famous method used to solve the problem is feature selection. Feature selection is capable of selecting features or attributes by determining their significance and effect towards classification performance. Feature selection is a process used to select the best subsets of features that can best representing the class model to maximally increase the performance [21]. It aims to merely select the subset of features without altering the original representation of the variables. Feature selection methods search through the subsets of features and try to find the best one among the competing features [15]. Large data scale can be reduced and provide better computational process if some of the features can be eliminated at the early stage by optimizing the feature selection algorithms. Feature selection techniques can be divided into three categories that are the filter methods, wrapper methods and the hybrid or embedded methods. The filter method relies on general characteristics of the data to evaluate and select feature subsets without involving any classification algorithm [5]. The wrapper method requires a pre-determined classification algorithm and uses its performance as the evaluation criterion [5]. It will search for features that are better suited with the classifier aiming to improve the performance. The hybrid method will exploit the evaluation criteria of the two models in different search stage that can benefit each other.

The features ranking method proposed by this study is under the filter method in the feature selection field of study. Filter techniques assess the relevance of features by looking at the intrinsic properties of the data. Feature relevance score is calculated and low scoring features are removed [21]. Some filter methods that can be considered are as such the distance measures, information measures, dependency measures and consistency measures. Features ranking method has the advantage of evaluating each data or features independently without having to concern of its classifier performance evaluation [28] as compared to the other feature selection method that is the wrapper methods. The most commonly used methods for features ranking in many fields include Chi-Squared, Gain Ratio, Information Gain, One R, Relief F and Symmetrical Uncertainty [29, 31]. Thus, this study proposed the Grey Relational Analysis (GRA) as the features ranking method for its predictive capability that able to determine the level of significant for each features without depending on any classifiers [26]. The scoring is made for each feature and the highest score produced by grey relational grade represent the most significant features.

2 Related Work

Features ranking is a procedure to predict and rank features or any attribute data to determine their significance level. The ranking is done by scoring the features in terms of their importance towards their class label. The method is aimed to select data that are being used as the input into classification model by using only the most significant features. The problem of data with high dimensionality has given too much disadvantages in terms of classification performance for several fields of study. Currently, features ranking procedures are adapted to solve the problem of too many features in medical data [9, 1], traffic congestion prediction [34], shellfish farms closure causes [20] and consumer product decision support [14] that are aimed to increase the classification performance by using only the most significant features by ranking.

Thus, one research has presented a new probability scoring method for traffic congestion prediction [34]. The task of prediction involves wide area correlation and high dimensionality of the data with large number of sensors. The relevancy of each sensor to the prediction task is 100 to 1. The performance is maintained although the data dimensionality is reduced in remarkably way. The method of ensemble feature ranking to determine the fault in shellfish farm closure has been proposed by Rahman [20]. This feature ranking algorithm is aimed to produce individual ranking for a number of subsets/bags by using the vector voting approach. They have determined that the factor of rain as the main cause of closure for most of the locations of the fish farms while the salinity factor has high probability for some locations.

Besides, the texture feature ranking method of Generalized Matrix Learning Vector Quantization (GMLVQ) has been proposed by Huber [9]. This method is aimed to solve the relevancy factor in texture features for lung disease pattern in HRCT images classification problem. There are 65 features that were used to determine their relevancy by ranking and selecting the features by implementing the GMLVQ. The best results were presented with the sets between 4 and 6 features for GMLVQ. The research involving High-Dimensional DNA microarray gene expression data by incorporating feature ranking and evolutionary method is done by Abedini [1]. They have proposed two methods based on the extension of the eXtended Classifier System (XCS) that include the feature selection for FS-XCS and GRD-XCS that incorporates probabilistic guided rule discovery mechanism for XCS. The research were given the result performance of GRD-XCS are better than FS-XCS in term of classification performance though both have performed much better than the original XCS. Thus, they suggest that by using informative features can improve the classification performance.

The research that proposed to ranking consumer's review on product features by using the method of linear regression with rule-based were proposed by Li [14]. This is aimed to present better suggestion to the future customer regarding the products. The features are extracted from the customers review on the product and services through various websites. A new approach to feature subset ranking were proposed by Xue [32] that involves two wrapper methods which are the single feature ranking that ranks the features according to their classification accuracy and the BPSO based feature subset ranking. The result obtained from their experiment have presented that with small number of top-ranked features have achieved better classification performance than using all features.

While the empirical study that comparing among 17 features ranking techniques is done by Wang [30]. This research proposed the ensemble techniques of features ranking for software measurement data reduction to predict software risk with high number of faults. These defect predictors are aimed to choose the most important features to improve their effectiveness. There are two, three and up to six combinations of rankers that have been manipulated to find their performances in this study. The researchers have come to conclusion that the combination of two rankers performed better than others.

Besides, the process of combining multiple features ranking into an ensemble features ranking framework was presented by Prati [19]. The research presented that by combining features ranking method has improved the method itself. The best aggregation method of all is SSD that is significantly better than any other features ranking individually or the aggregate rankings for the empirical evaluation using 39 UCI datasets, three different performance measures and three different learning algorithms. There are several features ranking that have been evaluated empirically [30, 31] that include Chi-Squared, Information Gain, Gain Ratio, ReliefF (RF and RFW) and Symmetrical Uncertainty. The Chi Squared – χ^2 (CS) is aimed to determine the distribution of the class to the target feature value [30]. This will evaluate the worth of each feature in regard towards their class. The feature is relevant to the class when the value of χ^2 statistics is larger that shows that the distribution values and classes are dependent.

3 Methodology

Feature Extraction procedure is one of the most important process in handwriting analysis and writer identification. This procedure is done to extract features and acquire information from handwriting image whether to determine the writer's characteristic or even the meaning of the words written. This study implemented the Higher-Order United Moment Invariant (HUMI) to construct the feature vectors for Global Features while the Local Features are extracted by the Edge based Directional (ED) method for the identification of author.

While the task of ranking features and select the most significant features involves two techniques that go through the process of hybridization to determine the best subsets of features. This task is aimed to select and reduce the number of features based on their level of significance in order to improve the performance accuracy with optimal amount of information to build the classifier model. The Grey Relational Analysis (GRA) as the features ranking technique is hybridized with the Feature Subset Selection (FSS). This process is aimed to produce the features based ranking and select the best subsets of significant features for this study through the hybridization of features ranking and feature subset selection (GRAFeSS).

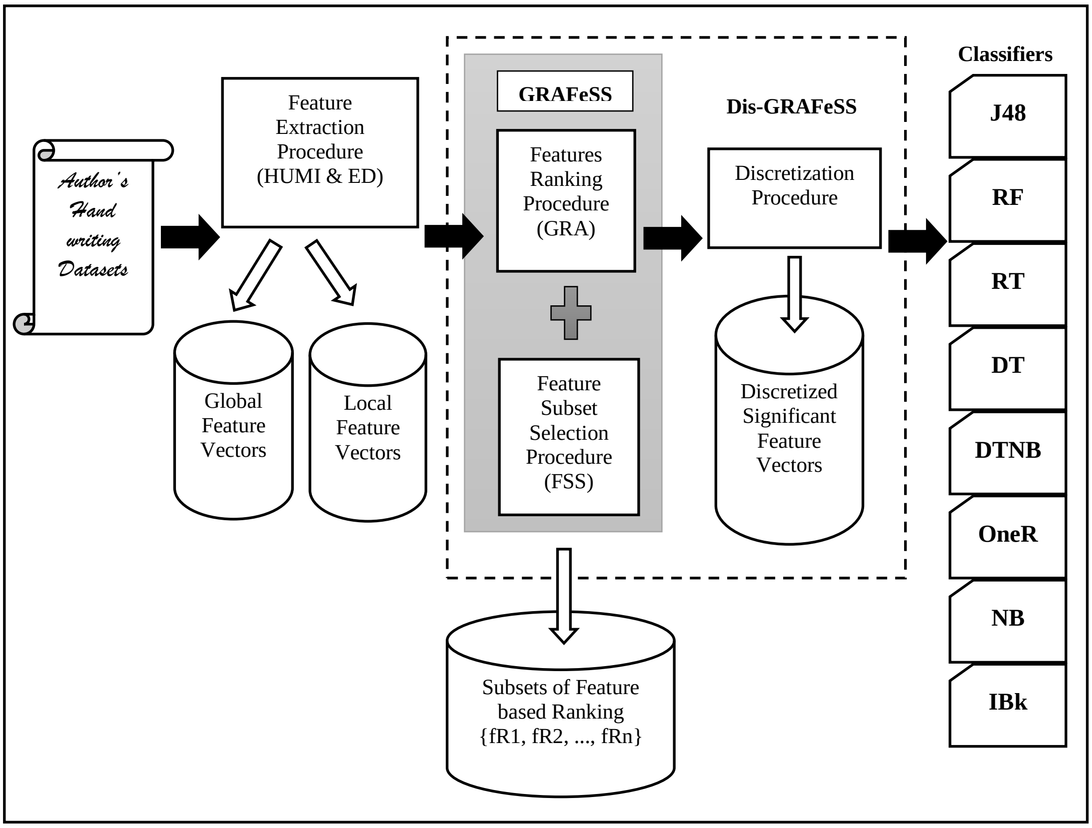


Fig. 1: The New Scheme of Discretized Features based Ranking for Writer Identification

This study also implemented the discretization procedure towards the proposed hybrid method of GRAFeSS. The task of discretization incorporates the process of transforming each features data into a general value that can represent certain feature through a certain common figure. The supervised discretization method of Equal Width Binning (EWB) [18] is deployed in this study. This method is implemented towards the features based ranking for both Global and Local

Features. This procedure is aimed to produce the discretized features based ranking as the invariant discretization for this study through the hybridization of features ranking and feature subset selection with discretization method that is named as Dis-GRAFeSS. Thus, the new scheme for writer identification is proposed in this study that is shown in Fig. 1 above to yield and select the most significant discretized features based ranking.

3.1 Grey Relational Analysis (GRA)

The most commonly used methods for features ranking in many fields include Chi-Squared, Gain Ratio, Information Gain, One R, Relief F and Symmetrical Uncertainty [29, 31]. Thus, the Grey Relational Analysis (GRA) are discussed here as the features ranking method for its predictive capability that able to determine the level of significant for each features without depending on any classifiers [26]. The scoring is made for each feature and the highest score produced by grey relational grade represent the most significant features.

The Grey Relational Analysis (GRA) that was first introduced by Julong [13] is used to measure the distance between two points as the degree of similarity or difference based on the grade of relation. The method contributions are expanded throughout different fields such as medical [10, 29], software prediction [3, 8, 27, 33] and system engineering [22]. The correlation degree of factors is measured by grey relational grade: higher similarities correspond to higher correlation of features. Measurements are obtained from the quantification of all the influences of various factors and the relationship among data series [26, 27]. The approach taken in this study is new in writer identification that it ranks the significance of features based on the grey possibility degree by using GRA. First, the *reference feature* and *comparative features* are determined. One feature is used as the reference feature, while the remaining is used as comparative features.

In the following, given features x_{ik}, $i = 0, 1, ..., n;$ $k = 1, \ldots, m;$ x_0, denotes the *reference feature* vector, and the reference features are x_{0k}, $k = 1, 2, \ldots, m;$ while the *comparative features* are denoted by x_{ik}, $i = 1, ..., n;$ $k = 1, \ldots, m;$ Let $D = \{x_1, x_2, ..., x_n\}$ be the handwriting data set, and $x_i = \{x_{i1}, x_{i2}, ..., x_{im}, x_0\},$ $i = 1, 2, ..., n;$ is a handwriting sample. x_{ik}, $k = 1, \ldots, m;$ are the features of handwriting sample of x_i. x_0, is the corresponding reference feature.

In matrix form, the data set D is as follows:

$$D = \begin{pmatrix} x_{10} & x_{11} & x_{12} & \ldots & x_{1m} \\ x_{20} & x_{21} & x_{22} & \ldots & x_{2m} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ x_{i0} & x_{i1} & x_{i2} & \ldots & x_{im} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ x_{n0} & x_{n1} & x_{n2} & \ldots & x_{nm} \end{pmatrix} \tag{1}$$

The steps to select the optimal feature subset using GRA are as follows:

Step 1 (Data series construction). Each column vector of the matrix D is viewed as a data series. There are a total of $m+1$ as follows:

$$\begin{aligned} x_0 &= \{x_{10}, x_{20}, \ldots, x_{n0}\}, \\ x_1 &= \{x_{11}, x_{21}, \ldots, x_{n1}\}, \\ x_2 &= \{x_{12}, x_{22}, \ldots, x_{n2}\}, \\ \ldots & \quad \ldots \quad \ldots \quad \ldots \quad \ldots \\ x_m &= \{x_{1m}, x_{2m}, \ldots, x_{nm}\}, \end{aligned} \tag{2}$$

Step 2 (Normalization). Data normalization is done in order to scale features into the same range to support their comparison. Here features are normalized by using equation (3).

$$x_{ik} \leftarrow \frac{x_{ik} - \min_i x_{ik}}{\max_i x_{ik} - \min_i x_{ik}}, \; i = 1, \ldots, n; \; k = 1, \ldots, m; \tag{3}$$

Step 3 (Find difference series). For each comparative feature, its difference series Δ_{ik}, is defined as the absolute difference between itself and the definite reference,

$$\Delta_{ik} = \left| x_{0k} - x_{ik} \right| \tag{4}$$

The following quantities are calculated next,

$$l_1(k) = \min_i \Delta_{ik}, \qquad L_1(k) = \max_i \Delta_{ik}$$
$$\text{and} \tag{5}$$
$$l = \min_k l_1(k), \qquad L = \max_k L_1(k)$$

Step 4 (Calculate relational coefficient). The relational coefficient, ξ_{ik}, for both reference and comparative feature is defined as follows:

$$\xi_{ik} = \frac{l + \rho L}{\Delta_{ik} + \rho L} \tag{6}$$

Where, the distinguishing coefficient $\rho \in [0,1]$ is usually set to 0.5 [13].

Step 5 (Calculate grey relational grade). The Grey Relational Grade (GRG), denoted by $\gamma_i \in [0,1]$, is the average of $\xi_{0k}, \ldots, \xi_{im}$.

$$\gamma_i = \frac{1}{m} \sum_{k=1}^{m} \xi_{ik} \tag{7}$$

Step 6 (Determine Grey Relational Rank (GRR)). Let $x_{(i)}$ and $\gamma_{(i)}$ denote the sequences x_i and γ_i respectively, considered in non-increasing order. That is, $x_{(i)}$ and $\gamma_{(i)}$ denote the ith largest values of x_i and γ_i respectively. Features are ranked by their grey relational grade. More precisely, the ith feature $x_{(i)}$ corresponds to the ith largest γ, i.e. $\gamma_{(i)}$. The optimal feature set consists of the highest ranked f features. Here, the cases $f = \{2,3\}$ are considered.

Combinations $(x_0, x_{1:f})$ feature subsets are then used with a selected classification algorithm and tested for their performance accuracy.

4 The Proposed Method of Hybrid Features Ranking

This study proposed to hybrid the two methods of Grey Relational Analysis (GRA) as the ranking procedure together with the Feature Subset Selection (FSS) method to select and combine the features based ranking. This proposed method is named as GRAFeSS. Fig. 2 shows the design flow of proposed hybrid method of GRAFeSS.

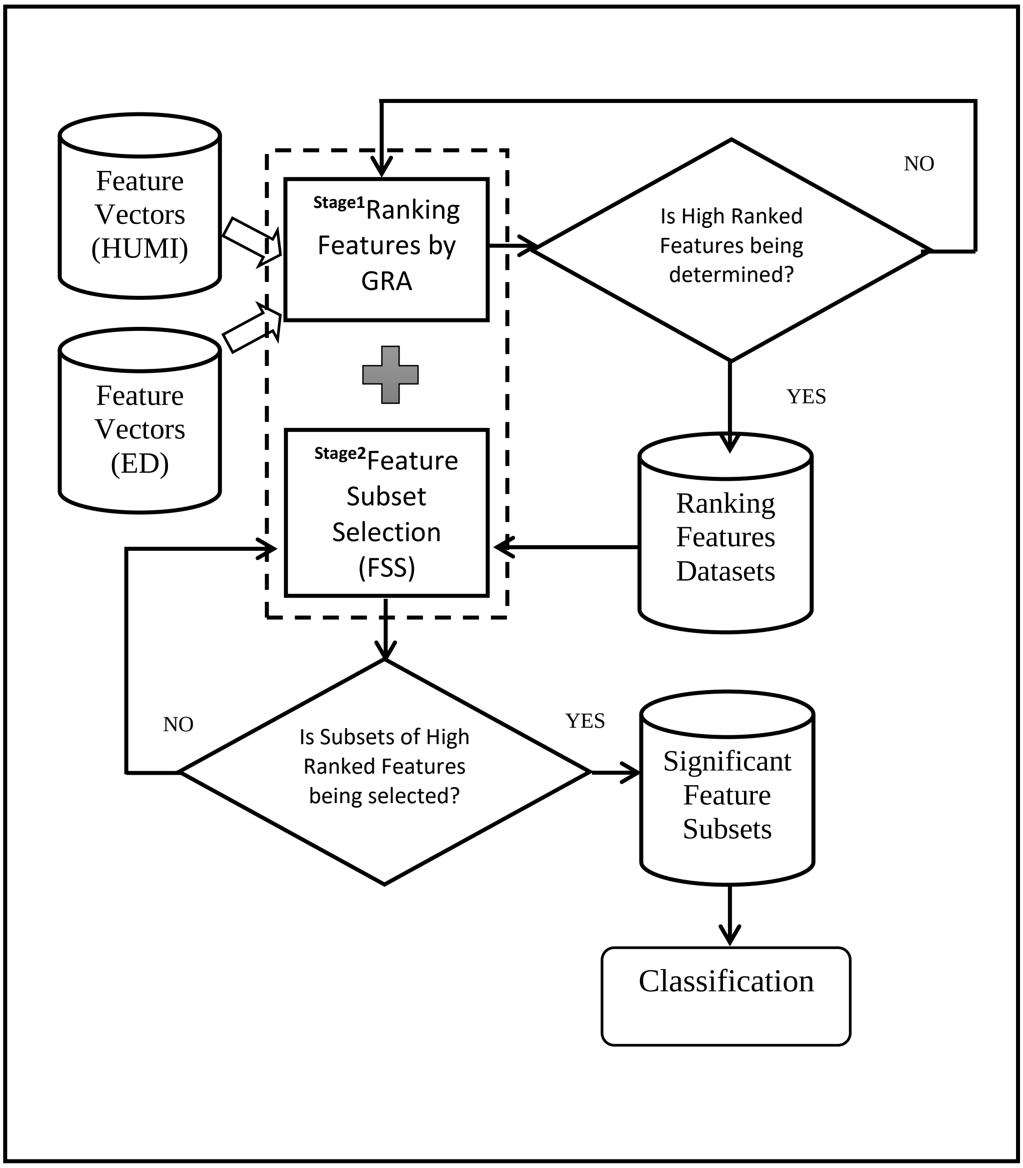


Fig. 2: Design Flow of Proposed Hybrid Method of GRAFeSS

The proposed hybrid method is implemented in two-stage manner. This two-stage of hybrid procedure is done by first implementing the features ranking procedure to determine the ranking score of each feature vectors. The features based ranking dataset is the output of the first stage and is used as the input to the next stage that is the selection and the combination of each high ranking feature. The feature subset selection procedure is done by selecting and combining the most significant features. This is aimed to produce the best subsets of the most significant features based ranking.

The proposed GRAFeSS algorithm for the hybridization of Grey Relational Analysis (GRA) and Feature Subset Selection (FSS) in this study is shown as Fig. 3.

Step 1

x_0, denotes the *reference feature* vector, and the reference features are x_{0k}, $k = 1, 2, \ldots, m$; while the *comparative features* are denoted by x_{ik}, $i = 1, \ldots, n$; $k = 1, \ldots, m$;

Step 2

The 6 steps to calculate the Grey Relational Analysis (GRA):-
Step 1 (Data series construction); Step 2 (Normalization);
Step 3 (Find difference series); Step 4 (Calculate relational coefficient); Step 5 (Calculate grey relational grade);
Step 6 (Determine Grey Relational Rank (GRR)).

Step 3

Construct features $F_i = \{f_1, f_2, \ldots, f_n\}$ with their ranks $R_i = \{fR_1, fR_2, \ldots, fR_n\}$ for Step A and Step B.

Step 3 A

Start with the subsets of the (n/2) highest rank $R_i = \{fR_1, fR_2, fR_3, fR_4\}$ then select and combine the features into the most significant feature subsets;

Step 3 B

For i=1 until (n/2) do:

i. For all f_i determine the feature rank then select the highest rank and add the feature into the most significant feature subsets; $S_i = \{fR_i, fR_{i+1}\}$

ii. Compare the combination of all the highest rank of features.

iii. For all $S_i \in R_i$ calculate the performance accuracy of each subset of the most significant features.

Step 4

Reiterate Steps 3 until complete all subsets.

Step 5

Finally select the best performance subsets of the most significant features.

Fig. 3: The Proposed Hybrid Method Algorithm of GRAFeSS

This process produces some combination of features to find whether the selected features will result in higher classification accuracy. The number of iteration to select and combine the features is based on the number of features that is divided by two. This will give the ability for the loop to be cut into half and expedite the process. It is also to avoid the problem of unknown number of iteration that can lead to exhaustive search process. The stopping criteria for the number of iteration must be set for feature subset selection procedure to stop the loop.

As a result, for example the two-highest-ranked combination of features based on their ranking is shown by Table 1. The subset of features will be tested to find their classification accuracy. This will produce a possible subset of features that could result in better classification accuracy or even higher than when using all features. Besides, Table 2 presents the example of high ranking feature subsets for HUMI and Edge based Directional. The features based ranking are constructed into ten (10) subsets of the most significant features that include the subsets of two (2), three (3) and four (4) features combination with features that ranking from first through fourth.

Table 1: Example of Subsets of Feature based Ranking for HUMI and Edge Features

Feature Subset	**Feature Combination**	**Feature Subset**	**Feature Combination**
S_1	$\{fR_1, fR_2\}$	S_4	$\{fR_2, fR_3\}$
S_2	$\{fR_1, fR_3\}$	S_5	$\{fR_2, fR_4\}$
S_3	$\{fR_1, fR_4\}$	S_6	$\{fR_3, fR_4\}$

Table 2: Example of High Ranking Feature Subsets for HUMI and Edge Features

Feature Subset $S_1, S_2, ..., S_k$	**Feature Combination** $\{fR_{n1}, fR_{n2}, .., fR_{nm}\}$	**Feature Subset** $S_1, S_2, ..., S_k$	**Feature Combination** $\{fR_{n1}, fR_{n2}, .., fR_{nm}\}$
S_1	$\{fR_1, fR_2\}$	S_6	$\{fR_3, fR_4\}$
S_2	$\{fR_1, fR_3\}$	S_7	$\{fR_1, fR_2, fR_3\}$
S_3	$\{fR_1, fR_4\}$	S_8	$\{fR_1, fR_2, fR_4\}$
S_4	$\{fR_2, fR_3\}$	S_9	$\{fR_2, fR_3, fR_4\}$
S_5	$\{fR_2, fR_4\}$	S_{10}	$\{fR_1, fR_2, fR_3, fR_4\}$

4.1 Significant Features by Ranking

The features significance is determined by deploying the features input data following all the six (6) steps of Grey Relational Analysis (GRA) as described before. The first step of GRA is to arrange the input data into data series constructions. The matrix D shows the matrix of data for HUMI global features and Edge local features in data series construction.

$$D = \begin{pmatrix} f_{10} & f_{11} & f_{12} & \cdots & f_{1m} \\ f_{20} & f_{21} & f_{22} & \cdots & f_{2m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ f_{i0} & f_{i1} & f_{i2} & \cdots & f_{im} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ f_{n0} & f_{n1} & f_{n2} & \cdots & f_{nm} \end{pmatrix} \quad (8)$$

Data series construction represents by each features are as shown below. HUMI has a total of eight (8) features while Edge has nine (9) features that are represented by the subscript number of each feature. The feature f_0 represents the reference features while other features $f_1, f_2, f_3 \cdots f_m$ represent the comparative features for either HUMI or Edge features. The value *m* represents the total number of comparative features that are seven (7) for global HUMI and eight (8) for local edge directional features while *n* represents the number of vectors for each feature.

$$\begin{aligned} f_0 &= \{f_{10}, f_{20}, \ldots, f_{n0}\}, \\ f_1 &= \{f_{11}, f_{21}, \ldots, f_{n1}\}, \\ f_2 &= \{f_{12}, f_{22}, \ldots, f_{n2}\}, \\ \cdots & \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\ f_m &= \{f_{1m}, f_{2m}, \ldots, f_{nm}\}, \end{aligned} \quad (9)$$

This technique first calculates the absolute difference $| X_0(k) - X_i(k) |$ between each feature vectors that is known as comparative with the reference feature that has been selected. The first feature $F1$ is selected as reference feature for HUMI while $F7$ is selected for Edge features. The first minimum, $\min_k | X_0(k) - X_i(k) |$ and maximum, $\max_k | X_0(k) - X_i(k) |$ absolute difference values are calculated for each comparative feature. Then, the second minimum,

$\min_k \max_k |X_0(k) - X_i(k)|$ and maximum, $\max_k \max_k |X_0(k) - X_i(k)|$ absolute difference values are calculated based on the first minimum and maximum values.

HUMI construct the values of 0 and 0.790141 for the second minimum and maximum respectively. While Edge gives the values of 6.30463e-006 for the second minimum and 0.998826 for the second maximum. All features use the same second minimum and maximum difference to calculate the relational coefficient values. Each feature is ranked based on their relational grade γ_i and give higher significance to features when the value is bigger.

As a result, the orders of each comparative feature for HUMI are:-

$$\gamma_1 > \gamma_3 > \gamma_2 > \gamma_7 > \gamma_5 > \gamma_4 > \gamma_6$$

The orders for each feature including the reference feature for HUMI are:-

$$f1 > f2 > f4 > f3 > f8 > f6 > f5 > f7$$

The results show $f1$ that has been chosen as the reference feature X_0 is in the highest rank as it is chosen for the reference of all other features. Among the comparative features, $f2$ that is represented by γ_1 is ranked first followed by $f4$ that is represented by γ_3 and $f3$ that is represented by γ_2. These three features together with the reference feature $f1$ are ranked in the four highest rank features. Besides, the other four lowest ranked features are the $f8$, $f6$, $f5$ and $f7$ in descending order.

While the result of ranking orders of each feature for Edge are:-

$$\gamma_8 > \gamma_3 > \gamma_7 > \gamma_2 > \gamma_1 > \gamma_4 > \gamma_6 > \gamma_5$$

The ranking orders for each feature including the reference feature for Edge are:-

$$f7 > f9 > f3 > f8 > f2 > f1 > f4 > f6 > f5$$

The results show $f7$ that has been chosen as the reference feature X_0 is in the highest rank as it is chosen for the reference of all other features. Among the comparative features, $f9$ that is represented by γ_8 is ranked first followed by $f3$ that is represented by γ_3 and $f8$ that is represented by γ_7. These three features together with the reference feature $f7$ are ranked in the four highest rank features.

5 Results, Analysis and Discussions

This section covers the comparison analysis between GRA and other feature ranking methods. The other six (6) feature ranking methods that have been considered are the Symmetrical Uncertainty, Chi Squared, Gain Ratio, Information Gain, ReliefF and OneR that was deployed by WEKA toolkit. GRA has determined that the four most significant features are F1, F2, F4 and F3 being F1 is the most significant followed by F2, F4 and F3 while the lowest four are F8, F6, F5 and F7 consecutively. Table 3 shows the comparison towards the other six (6) rankers that include the Symmetrical Uncertainty which has defined that the most significant features is F1 and the second highest rank is F2 given the same result as GRA but suggested differently with the third best feature that is F7 followed by F3 for the highest four subset while F5, F6, F8 and F4 are in the lowest four subset.

Besides, the other rankers of Chi Squared, Gain Ratio, Information Gain and ReliefF has suggested that the highest rank feature is F3 followed by F4, F1 and F2 as the highest four features while the lowest four are F7, F8, F5 and F6. This has determined that the subsets of four highest features between GRA and these four rankers are the same. Thus, the OneR ranker has given a slightest different result that proposed the best features of all 8 features is F2 and F1 falls to the second place followed by F4 and F7 while F3, F5, F8 and F6 fall to the lowest four rank.

Table 3: Ranking of Features by Other Rankers and GRA for HUMI

Rank	1st	2nd	3rd	4th	5th	6th	7th	8th
Rankers								
Grey Relational Analysis (GRA)	*F1*	*F2*	*F4*	*F3*	*F8*	*F6*	*F5*	*F7*
Chi Square	*F3*	*F4*	*F1*	*F2*	*F7*	*F8*	*F5*	*F6*
Gain Ratio	*F3*	*F4*	*F1*	*F2*	*F7*	*F8*	*F5*	*F6*
Information Gain	*F3*	*F4*	*F1*	*F2*	*F7*	*F8*	*F5*	*F6*
Symmetrical Uncertainty	*F3*	*F4*	*F1*	*F2*	*F7*	*F8*	*F5*	*F6*
OneR	*F2*	*F1*	*F4*	*F7*	*F3*	*F5*	*F8*	*F6*
Relief F	*F1*	*F2*	*F7*	*F3*	*F5*	*F6*	*F8*	*F4*

Table 4 shows that the most significant features for Local Features that has been determined by GRA are F7 followed by F9, F3 and F8 that comprised the set of four most significant features. The six (6) rankers that include the Symmetrical Uncertainty, Chi Squared, Gain Ratio, Information Gain, ReliefF and OneR have also been applied to determine the ranking for each Local Feature. Feature F3, F4, F1 and F2 are ranked as the four most significant features by Symmetrical Uncertainty, Chi Squared, Gain Ratio and Information Gain.

Table 4: Ranking of Features by Other Rankers and GRA for Edge

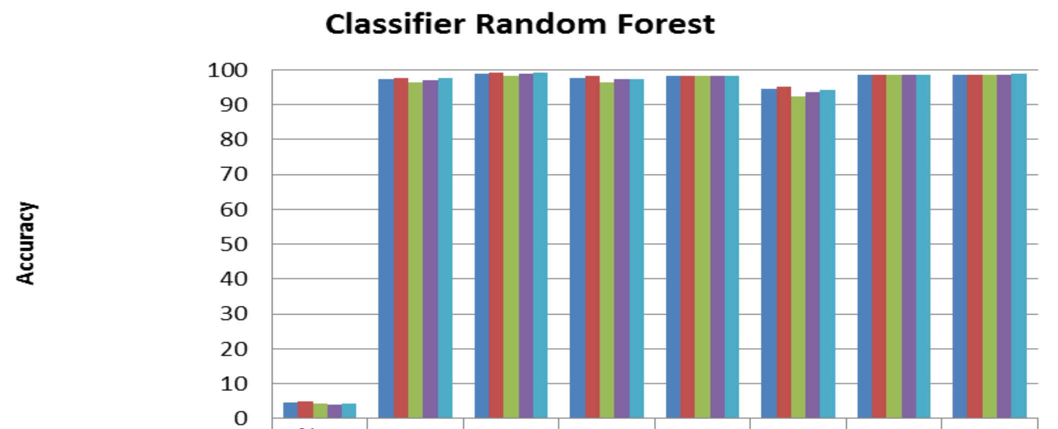
Rank	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}
Rankers									
Grey Relational Analysis (GRA)	*F7*	*F9*	*F3*	*F8*	*F2*	*F1*	*F4*	*F6*	*F5*
Chi Square	*F3*	*F4*	*F1*	*F2*	*F8*	*F9*	*F5*	*F7*	*F6*
Gain Ratio	*F3*	*F4*	*F1*	*F2*	*F8*	*F9*	*F5*	*F7*	*F6*
Information Gain	*F3*	*F4*	*F1*	*F2*	*F8*	*F9*	*F5*	*F7*	*F6*
Symmetrical Uncertainty	*F3*	*F4*	*F1*	*F2*	*F8*	*F9*	*F5*	*F7*	*F6*
OneR	*F4*	*F1*	*F5*	*F2*	*F3*	*F8*	*F7*	*F9*	*F6*
Relief F	*F5*	*F8*	*F9*	*F7*	*F1*	*F3*	*F2*	*F4*	*F6*

Besides, One R ranker has suggested that the feature F4 is the most significant followed by F1, F5 and F2 as four most significant features. This is a totally different result than the suggestion made by GRA. Thus, ReliefF has proposed almost the same result as GRA that given the four highest rank features includes F5 as the first in ranking followed by F8, F9 and F7. This has determined that at least three features in the four highest ranking are the same as GRA has proposed. Thus, the ranking of features for HUMI and Edge that have been proposed by GRA determined the feature subsets that are chosen to be implemented into the next procedure in the proposed schemes. This is aimed to improve the identification performance rate by using the smallest number of features determined by their significance level.

As a result, the proposed method of Dis-GRAFeSS is aimed to produce the most significant discretized feature based ranking that able to improve the performance accuracy by using the smallest number of features. Fig. 4 and 5 shows the comparison performance by using HUMI features for all discretize features with the two (2) combination features of the four (4) most significant features based ranking by GRA that included features of F1, F2, F3 and F4 by using classifiers Random Forest and Random Tree as the classifier scheme. The four most significant features have produced the subsets of two high ranking features that generated the following subsets;

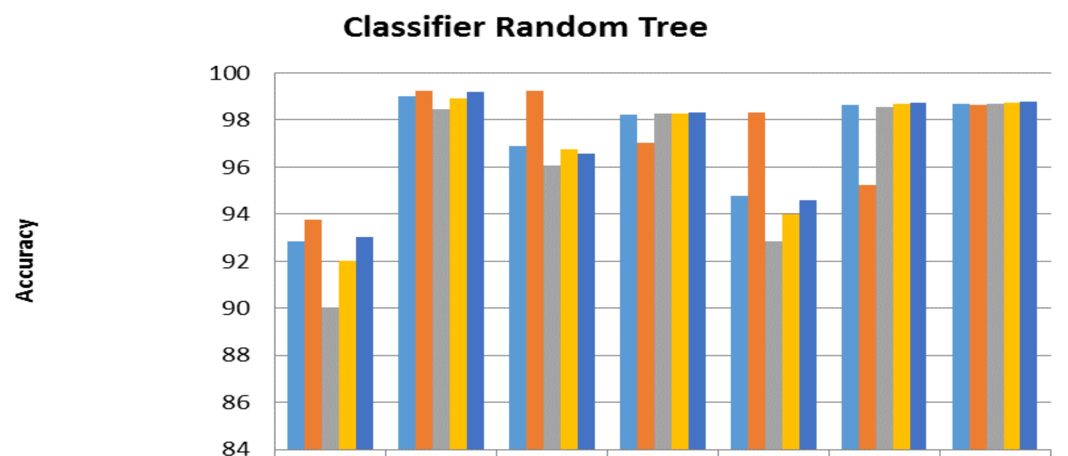
{f1, f2}, {f1, f3}, {f1, f4}, {f2, f4}, {f2, f3} and {f3, f4}.

The results for classifiers Random Forest and Random Tree are shown by Fig. 4 and Fig. 5 below.



	Non Dis	Dis	F1F2	F1F4	F1F3	F2F4	F2F3	F3F4
5 Folds	4.64	97.2833	99.02	97.7233	98.33	94.6333	98.6	98.6667
10 Folds	4.7933	97.7533	99.22	98.3333	98.3367	95.21	98.6233	98.6667
60% Train 40% Test	4.3417	96.575	98.3417	96.3	98.2583	92.4083	98.5833	98.675
70% Train 30% Test	3.9778	97.0333	98.8889	97.4222	98.3556	93.6889	98.6778	98.7222
80% Train 20% Test	4.4	97.6333	99.1	97.4833	98.3667	94.2333	98.75	98.7833

Fig. 4: Comparison Performance of the Four Highest Ranking of 2-Combination Global Features for Classifier of Random Forest



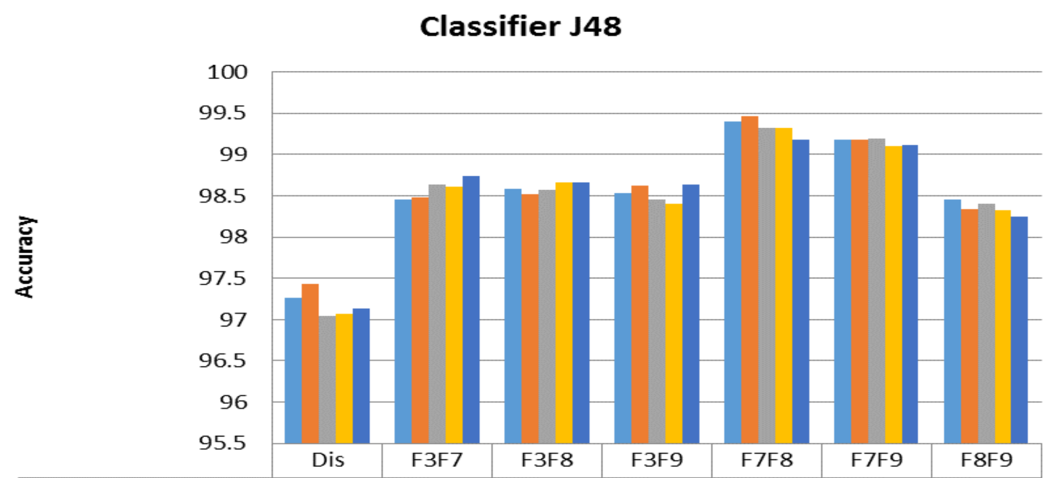
	Dis	F1F2	F1F4	F1F3	F2F4	F2F3	F3F4
5 Folds	92.8333	99.02	96.9167	98.2167	94.7933	98.62	98.6667
10 Folds	93.7633	99.22	99.2567	97.0333	98.3133	95.2233	98.6533
60% Train 40% Test	90.0417	98.475	96.05	98.2667	92.825	98.5583	98.675
70% Train 30% Test	92.0333	98.9222	96.7556	98.2556	93.9778	98.6778	98.7222
80% Train 20% Test	93.05	99.1833	96.55	98.3	94.6	98.75	98.7833

Fig. 5: Comparison Performance of the Four Highest Ranking of 2-Combination Global Features for Classifier of Random Tree

Both presented that the performance of the feature subsets of {f1, f2} that is produced by Dis-GRAFeSS has given better performance than by using all discretize features. The feature subset of {f1, f2} has given the performance accuracy of 99.22% for both classifiers with the ten (10) fold cross validation environment setup to be compared with 97.8% for Random Forest and 93.8% for Random Tree for all discretized features.

In another environment setup, Fig. 6 shows the comparison performance by using Edge features towards J48 as the classifier scheme for all discretize features with the two (2) combination features of the four (4) most significant features based ranking by GRA that included features of F7, F9, F8 and F3 that are ranking first through fourth. The four most significant features have produced the subsets of two high ranking features that generated the following subsets;

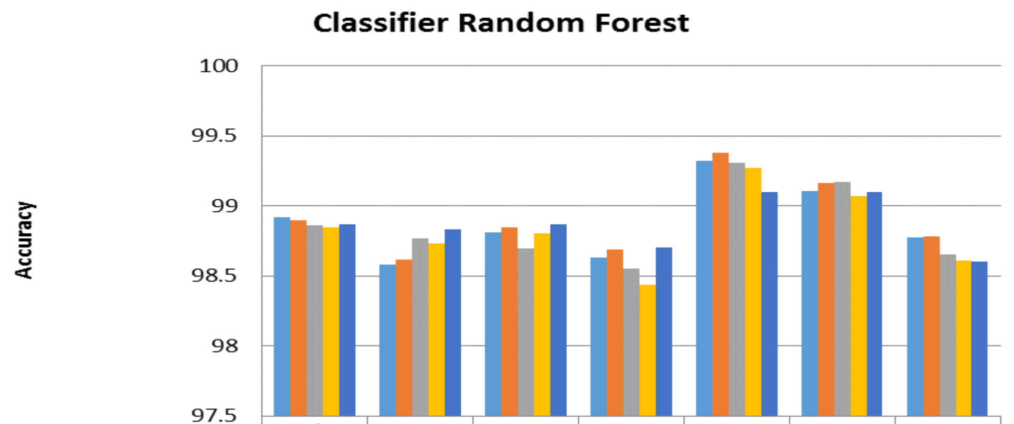
{f3, f7}, {f3, f8}, {f3, f9}, {f7, f8}, {f7, f9} and {f8, f9}.



	Dis	F3F7	F3F8	F3F9	F7F8	F7F9	F8F9
5 Folds	97.2567	98.4533	98.58	98.5333	99.3967	99.1733	98.45
10 Folds	97.43	98.4767	98.52	98.62	99.4567	99.1833	98.3333
60% Train 40% Test	97.0417	98.6333	98.5667	98.4583	99.325	99.1917	98.4
70% Train 30% Test	97.0667	98.6111	98.6667	98.4	99.3222	99.1	98.3222
80% Train 20% Test	97.1333	98.7333	98.6667	98.6333	99.1833	99.1167	98.25

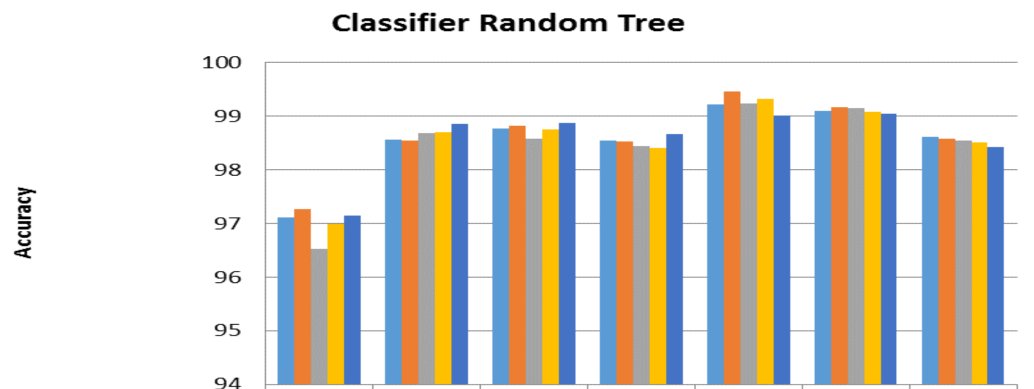
Fig. 6: Comparison Performance of the Four Highest Ranking of 2-Combination Local Features for Classifier of J48

The performance of the subsets of most significant discretized features based ranking has exceeded the performance of all discretized features. These performances include the classifiers of J48, Random Forest, Random Tree, Decision Tree and DTNB that are shown by Fig. 6 until Fig. 10. The feature subset of {f7, f9} has given the highest performance of 100% for all five (5) environment setup for classifier scheme of DTNB to be compared with the performance of all discretized features that reached 99.94% for the ten (10) fold cross validation environment setup for the same scheme.



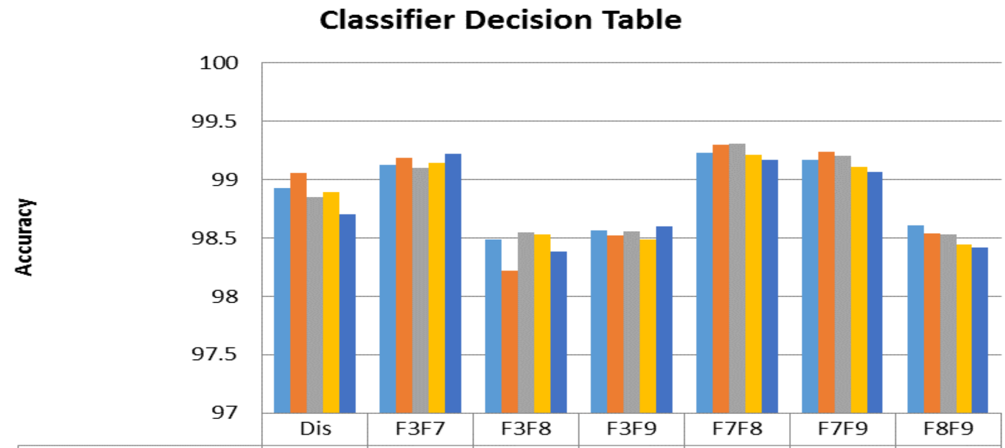
	Dis	F3F7	F3F8	F3F9	F7F8	F7F9	F8F9
5 Folds	98.9167	98.5833	98.8067	98.63	99.32	99.1033	98.7733
10 Folds	98.9	98.6167	98.8467	98.6867	99.3767	99.16	98.7833
60% Train 40% Test	98.8583	98.7667	98.6917	98.55	99.3083	99.1667	98.65
70% Train 30% Test	98.8444	98.7333	98.8	98.4333	99.2667	99.0667	98.6111
80% Train 20% Test	98.8667	98.8333	98.8667	98.7	99.1	99.1	98.6

Fig. 7: Comparison Performance of the Four Highest Ranking of 2-Combination Local Features for Classifier of Random Forest



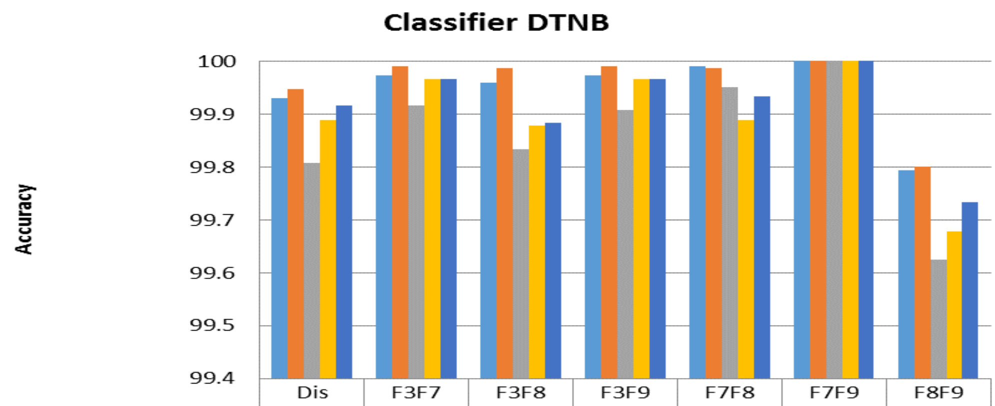
	Dis	F3F7	F3F8	F3F9	F7F8	F7F9	F8F9
5 Folds	97.1133	98.5633	98.7633	98.5333	99.2133	99.0967	98.6133
10 Folds	97.2667	98.5367	98.8133	98.5167	99.45	99.1567	98.5833
60% Train 40% Test	96.525	98.6833	98.575	98.4417	99.225	99.15	98.55
70% Train 30% Test	96.9889	98.6889	98.7556	98.4111	99.3111	99.0778	98.5111
80% Train 20% Test	97.15	98.85	98.8667	98.6667	99	99.05	98.4167

Fig. 8: Comparison Performance of the Four Highest Ranking of 2-Combination Local Features for Classifier of Random Tree



	Dis	F3F7	F3F8	F3F9	F7F8	F7F9	F8F9
5 Folds	98.9267	99.12	98.4833	98.56	99.23	99.1667	98.6067
10 Folds	99.0567	99.1833	98.22	98.5233	99.3	99.24	98.54
60% Train 40% Test	98.85	99.1	98.55	98.5583	99.3083	99.2	98.5333
70% Train 30% Test	98.8889	99.1444	98.5333	98.4889	99.2111	99.1111	98.4444
80% Train 20% Test	98.7	99.2167	98.3833	98.6	99.1667	99.0667	98.4167

Fig. 9: Comparison Performance of the Four Highest Ranking of 2-Combination Local Features for Classifier of Decision Table



	Dis	F3F7	F3F8	F3F9	F7F8	F7F9	F8F9
5 Folds	99.93	99.9733	99.96	99.9733	99.99	100	99.7933
10 Folds	99.9467	99.99	99.9867	99.99	99.9867	100	99.8
60% Train 40% Test	99.8083	99.9167	99.8333	99.9083	99.95	100	99.625
70% Train 30% Test	99.8889	99.9667	99.8778	99.9667	99.8889	100	99.6778
80% Train 20% Test	99.9167	99.9667	99.8833	99.9667	99.9333	100	99.7333

Fig. 10: Comparison Performance of the Four Highest Ranking of 2-Combination Local Features for Classifier of DTNB

Besides, the feature subset of {f7, f8} has also given higher performances than all discretized features with the performance accuracy of 99.99% for the same scheme of DTNB. This is also shown that both feature subsets of {f7, f9} and {f7, f8} have performed better than all discretized features in the classifiers scheme of J48, Random Forest, Random Tree, Decision Tree and DTNB that are also the highest performance that have been given by the feature subsets produced by Dis-GRAFeSS. This has proven that the proposed method of Dis-GRAFeSS has been able to rank and select the best feature subset based on their significant level to improve the classification accuracy.

6 Conclusion

The purpose of this study is to propose the hybrid method of GRA and Feature Subset Selection that is named as GRAFeSS and deployed the discretization model towards the hybrid method (Dis-GRAFeSS). This is aimed to construct the best subsets of the most significant features that contribute to improve the performance accuracy by using the smallest number of discretized feature subsets. The proposed method is implemented towards two types of features that are the Global Features extracted from Higher-Order United Moment Invariants (HUMI) and Local Features that are constructed by the Edge based Directional (ED). GRAFeSS has proposed that the four most significant features for global features are F1, F2, F4 and F3. Besides, the features F7, F9, F3 and F8 are determined with the four highest significance levels for local features. Thus, the subsets and combination of features are constructed based on their significance level resulting to the best subsets of discretized features. As a result, the proposed best subset of features is {f1, f2} that are defined for global features while local features are represented by {f7, f9}. Thus, this shows that the performance of the proposed method of Dis-GRAFeSS has succeeded to improve the accuracy rate to determine the writers with the feature based ranking invariant discretization by using only the most significant features with the smallest number of feature subsets. The best result obtained by Discretized Local Features based Ranking by using the feature subset of {f7, f9} for the classifier of DTNB with performance of 100% while Discretized Global Features based Ranking presenting the performance of 99.22% for the feature subset of {f1, f2} with classifier Random Forest.

ACKNOWLEDGEMENTS

This work is supported by UTM Big Data Centre (UTM BDC) and Computational Intelligence and Technologies Lab (CIT Lab). The researcher is sponsored by Ministry of Education, Malaysia, Universiti Teknikal Malaysia Melaka (UTeM) and Universiti Teknologi Malaysia (UTM).

References

[1] Abedini, M., & Kirley, M. (2013). An enhanced XCS rule discovery module using feature ranking. *International Journal of Machine Learning and Cybernetics*, *4*(3), 173–187.

[2] Amaral, A. M. M., de Almendra Freitas, C. O., & Bortolozzi, F. (2013, August). Feature selection for forensic handwriting identification. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on* (pp. 922-926). IEEE.

[3] Azzeh, M., Neagu, D., & Cowling, P. I. (2009). Fuzzy grey relational analysis for software effort estimation. *Empirical Software Engineering*, *15*(1), 60–90.

[4] Blumenstein, M., Verma, B., & Basli, H. (2003, August). A novel feature extraction technique for the recognition of segmented handwritten characters. In Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on (pp. 137-141). IEEE.

[5] Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1), 16-28.

[6] Dittman, D., Khoshgoftaar, T., Wald, R., & Napolitano, A. (2012, October). Similarity analysis of feature ranking techniques on imbalanced DNA microarray datasets. In *Bioinformatics and Biomedicine (BIBM), 2012 IEEE International Conference on* (pp. 1-5). IEEE.

[7] Hu, M. K. (1962). Visual pattern recognition by moment invariants. *IRE transactions on information theory*, 8(2), 179-187.

[8] Hsu, C. J., & Huang, C. Y. (2011). Comparison of weighted grey relational analysis for software effort estimation. Software Quality Journal, 19(1), 165-200.

[9] Huber, M. B., Bunte, K., Nagarajan, M. B., Biehl, M., Ray, L. A., & Wismüller, A. (2012). Texture feature ranking with relevance learning to classify interstitial lung disease patterns. *Artificial intelligence in medicine*, 56(2), 91-97.

[10] İçer, S., Coşkun, A., & İkizceli, T. (2012). Quantitative grading using grey relational analysis on ultrasonographic images of a fatty liver. *Journal of medical systems*, 36(4), 2521-2528.

[11] Inbarani, H. H., Azar, A. T., & Jothi, G. (2014). Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis. *Computer Methods and Programs in Biomedicine*, *113*(1), 175–185.

[12] Jiang, H., & He, W. (2012). Grey relational grade in local support vector regression for financial time series prediction. *Expert Systems With Applications*, *39*(3), 2256–2262.

[13] Julong, D. (1989). Introduction to Grey System Theory. *The Journal of Grey System*, 1 (1), 1–24.

[14] Li, S. K., Guan, Z., Tang, L. Y., & Chen, Z. (2012). Exploiting consumer reviews for product feature ranking. *Journal of Computer Science and Technology*, 27(3), 635-649.

[15] Liu, H., & Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on knowledge and data engineering*, 17(4), 491-502.

[16] Muda, A. K., Shamsuddin, S. M., & Abraham, A. (2009, June). Authorship invarianceness for writer identification. In *Biometrics and Kansei Engineering, 2009. ICBAKE 2009. International Conference on* (pp. 34-39). IEEE.

[17] Muda, A. K., Shamsuddin, S. M., & Darus, M. (2007, August). Embedded scale united moment invariant for identification of handwriting individuality. In *International Conference on Computational Science and Its Applications* (pp. 385-396). Springer, Berlin, Heidelberg.

[18] Muda, A. K., Shamsuddin, S. M., & Darus, M. (2008b). Invariants discretization for individuality representation in handwritten authorship. *Lecture Notes in Computer Science*, 5158, 218-228.

[19] Prati, R. C. (2012, June). Combining feature ranking algorithms through rank aggregation. In *Neural Networks (IJCNN), The 2012 International Joint Conference on* (pp. 1-8). IEEE.

[20] Rahman, A., D'Este, C., & McCulloch, J. (2013, December). Ensemble feature ranking for shellfish farm closure cause identification. In *Proceedings of Workshop on Machine Learning for Sensory Data Analysis* (p. 13). ACM.

[21] Saeys, Y., Inza, I., & Larrañaga, P. (2007a). A review of feature selection techniques in bioinformatics. *bioinformatics*, 23(19), 2507-2517.

[22] Samvedi, A., Jain, V., & Chan, F. T. (2012). An integrated approach for machine tool selection using fuzzy analytical hierarchy process and grey relational analysis. *International Journal of Production Research*, 50(12), 3211-3221.

[23] Sesa-Nogueras, E., & Faundez-Zanuy, M. (2013). Writer recognition enhancement by means of synthetically generated handwritten text. *Engineering Applications of Artificial Intelligence*, 26(1), 609-624.

[24] Shamsuddin, S. M., Sulaiman, M. N., & Darus, M. (2002). Invarianceness of higher order centralised scaled-invariants undergo basic transformations. *International Journal of Computer Mathematics*, 79(1), 39-48.

[25] Shilaskar, S., & Ghatol, A. (2013). Feature selection for medical diagnosis: Evaluation for cardiovascular diseases. *Expert Systems with Applications*,

40(10), 4146-4153.

[26] Song, Q., & Shepperd, M. (2011). Predicting software project effort: A grey relational analysis based method. *Expert Systems with Applications*, 38(6), 7302-7316.

[27] Song, Q., Shepperd, M., & Mair, C. (2005, September). Using grey relational analysis to predict software effort with small data sets. In *Software Metrics, 2005. 11th IEEE International Symposium* (pp. 10-pp). IEEE.

[28] Test, E., Kecman, V., Strack, R., Li, Q., & Salman, R. (2012, March). Feature ranking for pattern recognition: A comparison of filter methods. In *Southeastcon, 2012 Proceedings of IEEE* (pp. 1-5). IEEE.

[29] Van Hulse, J., Khoshgoftaar, T. M., & Napolitano, A. (2011, August). A comparative evaluation of feature ranking methods for high dimensional bioinformatics data. In *Information Reuse and Integration (IRI), 2011 IEEE International Conference on* (pp. 315-320). IEEE.

[30] Wang, H., Khoshgoftaar, T. M., & Napolitano, A. (2012). Software measurement data reduction using ensemble techniques. *Neurocomputing*, 92, 124-132.

[31] Wang, H., Khoshgoftaar, T. M., & Gao, K. (2010, August). A comparative study of filter-based feature ranking techniques. In *Information Reuse and Integration (IRI), 2010 IEEE International Conference on* (pp. 43-48). IEEE.

[32] Xue, B., Zhang, M., & Browne, W. N. (2012, January). Single feature ranking and binary particle swarm optimisation based feature subset ranking for feature selection. In *Proceedings of the Thirty-fifth Australasian Computer Science Conference*-Volume 122 (pp. 27-36). Australian Computer Society, Inc.

[33] Yang, C. C., & Chen, B. S. (2006). Supplier selection using combined analytical hierarchy process and grey relational analysis. J*ournal of Manufacturing Technology Management*, 17(7), 926-941.

[34] Yang, S. (2013). On feature selection for traffic congestion prediction. *Transportation Research Part C: Emerging Technologies*, 26, 160-169.

[35] Yinan, S., Weijun, L., & Yuechao, W. (2003, October). United moment invariants for shape discrimination. In *Robotics, Intelligent Systems and Signal Processing, 2003. Proceedings. 2003 IEEE International Conference on* (Vol. 1, pp. 88-93). IEEE.