International Journal of Engineering & Technology, 7 (4.31) (2018) 228-233

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Hybrid Feature Selection Technique for Classification of Group-based Holy Quran Verses

A. Adeleke[1]*, N. Samsudin[2]

Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia,
86400 Parit Raja, Batu Pahat, Johor, Malaysia.
**Corresponding author E-mail: hi170046@siswa.uthm.edu.my*

Abstract

Text classification problem is primarily applied in document labeling. However, the major setbacks with the existing feature selection techniques are high computational runtime associated with wrapper-based FS techniques and low classification accuracy performance associated with filter-based FS techniques. In this paper, a hybrid feature selection technique is proposed. The proposed hybrid technique is a combination of *filter-based information gain (IG) and wrapper-based CFS algorithms.* The specific purpose for this combination is to achieve both high classification *accuracy* performance (associated with wrapper) at *lower computational runtime* (associated with filter). The proposed *IG-CFS* technique is then applied to label Quranic verses of al-Baqara and al-Anaam from two major references, the English translation and commentary (tafsir). StringToWordVector with weighted TF-IDF method were used for preprocessing the textual data while four classifiers: naïve bayes, libSVM, *k*-NN, and decision trees (J48) were experimented. The overall highest classification accuracy of 94.5% was achieved at 3.89secs runtime with the proposed *IG-CFS* technique.

Keywords: *Feature Selection Techniques; Holy Quran; Text Classification Algorithms; AUC; ROC Curve*

1. Introduction

Automated text classification is the systematic way of organizing documents automatically into predefined categories [1]. The number of documents being processed over the years continually increased as a result of the massive technological growth and development in the important fields of artificial intelligence (AI) and machine learning (ML). With the increasingly demand for processing large documents within a short time, automating documents processing becomes a necessity.

In the field of machine learning, the specific task is developing models that give computing machines the capability of learning [2], [3]. Thereafter translates the acquired knowledge learnt into decision making automatically.

Important to text classification is feature selection: a dimensionality reduction method that helps to reduce the complexity of dimensionality usually associated with text [5]. To optimize the performance of the classification algorithms, techniques such as in feature selection are essentially needed in cleansing and selecting the most relevant feature subsets for the experimental work.

Quran is a unique text and a good source of information with 78,000 words systematically arranged into sections such as verses, chapters, quarters, parts etc., by religious scholars [1]. With such great importance and characteristics, features (or keywords) could be extracted from the divine text which helps to improve the literacy level of its readers and researchers.

In addition, there are multiple sources of the Quranic text available from the extensive academic works of the religious scholars. From such sources are the Quran translations done into almost all languages, Quran commentary etc. However, working on these sources independently has its setback. For example, the Holy Quran translation as a source may not be sufficient for the purpose of analyzing the Quranic verses for the labeling task. Thus, there is a need to combine the sources while extracting features for the classification task.

However, the existing FS techniques employed for this purpose have the limitations of high computational runtime as associated with the wrapper-based feature selection algorithms and lower classification accuracy performance associated with the filter-based FS algorithms. In view of this, the study proposes a hybrid feature selection technique.

The proposed technique is a two-step combination of filter-based *information gain (IG)* and wrapper-based *CFS* algorithms. The proposed *IG-CFS* technique will be used for the purpose of automating the classification of Quranic verses in chapters two and six of the Holy book. The input data (Quranic verses) are automatically labeled to one of the predefined labels: '*iman*, *ibadah*, *akhlak*'.

In organizing the paper, some related works were reviewed in section II while the detailed explanation of the methodology employed for the experimental work is presented in section III. Subsequently, section IV documented the results while section V summed up the research work with recommendations.

2. Related Work

The field of Holy Quran study have witnessed quite a number of research works among which are: text classification applications on the Holy Quran [1], [5], [6], [7], [8], [9]; ontology-based applications [10], [11], [12], [13]; digitized Holy Quran applications [14], [15], [16], [17].

From among the techniques that have been widely applied to text classification problems, including the Bayes probabilistic approach [4], decision trees [18], neural networks [19], support vector machines [20], and *k*-nearest neighbour [21].

The research work in [29] is based on classifying sonar targets using information gain FS algorithm for attribute evaluation. The experimental work focused on training networks for the purpose of discriminating between the sonar signals. The experimental results showed IG attribute evaluation significantly improved the classification task.

A Genetic algorithm wrapper-based feature selection method was proposed in [30] for classifying hyperspectral images using SVM classifier. The feature selection process involves three steps: creating the training and testing sets using ENVI software; setting up required parameters; running the model. The FS algorithm was used to optimize the kernel parameters and feature subsets.

[31] introduced in their work a feature selection method using support vector machine to find dependency between the attributes of high dimensional data extracted from UCI data repository and then decide the appropriate class attributes values. The results showed that the FS method had promising results in the classification tasks. In addition, from other research works in feature selection include *but not limited* to [32], [33].

In this study, a hybrid model is proposed to improve feature selection process for the Quranic text classification task. Hybrid approach to feature selection has been successfully experimented in classification problems [22] – [25]. The proposed FS technique combines information gain filter FS algorithm and correlation-based wrapper FS algorithm. Both algorithms are often experimented in text classification problems [26]. IG algorithm is simple, easy to use and has been effectively employed in classification problems [26], [24], [27]. Correlation-based (CFS) is a wrapper feature selection algorithm. CFS have also been experimented in classification problems [28], [5].

3. Method and Experiment

This study is about applying FS techniques in selecting most relevant features from the Quranic text needed for the classification problem. In this paper, three FS algorithms are experimented: *IG*, *CFS*, and the proposed hybrid technique. The experimental work employed four classifiers for classifying the input text to the desired labels.

The experimental design consists of four steps as shown in Fig. 1. The input data are Quranic verses collected from the combined sources of Holy Quran translation and tafsir. The resulting combined text data is otherwise termed 'grouped-data'.

3.1. Data Acquisition

The experimental datasets consist of 451 instances (or verses); 286 verses from chapter two (Surah al-Baqarah) and 165 verses from chapter six (Surah al-Anaam) of the Holy Quran. Table 1 shows the distribution of the percentage weight compositions of the three predefined classes (*Iman*, *Ibadah*, and *Akhlak*) with the experimental datasets.

Table 1: Percentage Composition of Class Labels

Datasets	No of Instances	Class Weight		
		Iman	Ibadah	Akhlak
Translation	451	343.0	44.0	64.0
Tafsir	451	345.0	42.0	64.0
Trans+Taf	451	345.0	42.0	64.0

3.2. Feature Generation

The experimental work is not possible to be performed except with features. These features needed are to be first extracted from the Quranic texts. To do this, the study employed standard StringToWordVector filter tool [5].

TF-IDF weighting method is thereafter applied to access and measure the degree of relevance of the extracted features.

Term frequency $Tf(t,d)$ is defined as the number of times a given term t (word/token) appears in a document d [1,5]. Mathematically, $(Tf(t,d))$ is defined as:

$$Tf(t,d) = 0.5 + \frac{0.5 \times f(t,d)}{Maximum\ Occurrences\ of\ words} \tag{1}$$

Where $Maximum\ Occurrences\ of\ words$ is denoted with: $Max\,\{ft^{|}, d\colon t^{|} \in d\}$.

Inverse-Document Frequency (IDF) is a measure of how much information a word provides. In other words, IDF is a method of evaluating if a term is common or rare across all documents in a collection [1,5]. Mathematically, IDF is given as:

$$idf(t,D) = log\frac{N}{|\{d \in D\colon t \in d\}|} \tag{2}$$

TF-IDF could then be given as:

$$tfidf(t,d,D) = tf(t,d) \cdot idf(t,D) \tag{3}$$

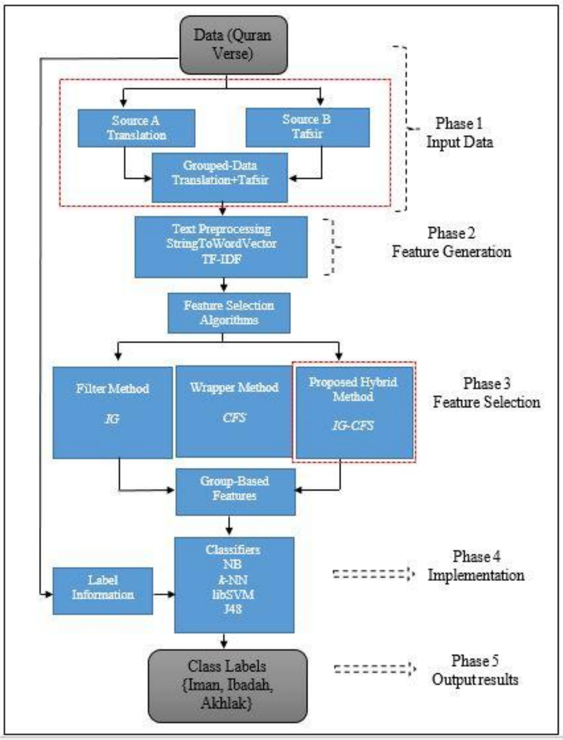
Data (Quran Verse)
Source A Translation
Source B Tafsir
Grouped-Data Translation+Tafsir
Phase 1 Input Data
Text Preprocessing StringToWordVector TF-IDF
Phase 2 Feature Generation
Feature Selection Algorithms
Filter Method IG
Wrapper Method CFS
Proposed Hybrid Method IG-CFS
Phase 3 Feature Selection
Group-Based Features
Label Information
Classifiers NB k-NN libSVM J48
Phase 4 Implementation
Class Labels {Iman, Ibadah, Akhlak}
Phase 5 Output results

Fig. 1: Proposed Framework

3.3. Feature Selection

To solve the complexity associated with text, feature selection technique is employed. The technique has significant influence on the classification process. Three feature selection algorithms are experimented namely: *IG*, *CFS*, and the proposed hybrid *IG-CFS*.

Furthermore, the filter method is less computationally expensive in comparison with the wrapper method. The filter method selects features independent of the classifiers. This makes the method simple, fast, and less computationally expensive. The wrapper method utilizes the performance of the classification algorithms to evaluate and select the feature subsets. Due to this dependency, the method has high accuracy performance but highly computationally expensive. Thus, to achieve both high classification accuracy at lower computational runtime, the study proposed a hybrid technique.

Information gain FS method measures the inter-dependency of the extracted features to the class labels. Given variables X and Y, the

FS method can be calculated as:

$$I(X{:}\,Y) = H(X) - H(X|Y) \qquad (4)$$

$H(X)$ represents entropy of discrete random variable X defined as:

$$H(X) = -\sum_{x_i \in X} P(x_i) \log\left(P(x_i)\right) \qquad (5)$$

x_i represents value of the variable X, $P(x_i)$ probability of x_i over all possible values of Xy.
$H(X|Y)$ could as well be defined as:

$$H(X|Y) = \sum_{y_j \in Y} P(y_j) \sum_{x_i \in X} P(x_i|y_j) \log\left(P(x_i|y_j)\right) \qquad (6)$$

$P(y_j)$ is the prior probability of y_j, $P(x_i|y_j)$ is the conditional probability of x_i given y_j.
IG is otherwise calculated as:

$$I(X{:}\,Y) = \sum_{x_i \in X} \sum_{y_j \in Y} P(x_i, y_j) log \frac{P(x_i, y_j)}{P(x_i)P(y_j)} \qquad (7)$$

CFS wrapper-based method finds the subsets of features that are individually highly correlated with the class but have low inter-correlation [5]. The method can be calculated using:

$$r_{zc} = \frac{k\overline{r_{z\imath}}}{\sqrt{k + k(k-1)\overline{r_{\imath\imath}}}} \qquad (8)$$

The experimental work of the proposed hybrid technique follows the following steps for the Holy Quran datasets.

Step 1: Input data (Holy Quran verse)
Step 2: Generate features from input data using StringToWordVector and TF-IDF
Step 3: Feature preprocessing using IG feature selection algorithm
Step 4: Implement features from (step 3) with the classifiers
Step 5: Select features from (step 2) using CFS algorithm
Step 6: Implement selected features from (step 5) with the classifiers
Step 7: Load the generated features from (step 2)
Step 8: Iterate (step 3)
Step 9: Apply CFS algorithm on selected features from (step 8)
Step 10: Implement resulting features from *IG-CFS* with the classifiers
Step 11: Evaluate results

3.4. Classifier

In this study, four classification algorithms: nearest neighbor (k-NN), support vector machines (libSVM), naïve bayes (NB), and decision trees (J48) classifiers were implemented using 10-fold cross validation method for the labeling task.
The k-NN classifier is an instance-based learning algorithm that has shown to be very simple but effective for text classification problem [1,5]. It is a non-parametric method used in classification and works by calculating the Euclidean distance between points.
In classifying a new document x, the algorithm ranks the document's neighbors in the training set, and then uses the class of k most similar neighbors to predict the class of a new document (also known as majority vote). The Euclidean distance is given as:

$$d(x, x_i) = \sqrt{\sum_{i=1}^{n} (x_j - x_{ij})^{\wedge}{}_2} \qquad (9)$$

Naïve bayes classifier greatly simplify learning by assuming that features are independent given class and has proven effective in many practical applications, including text classification [1,5]. The classifier is a simple probabilistic model based on the Bayes rule [5]. Given a class C, the probabilty of a particular document d to belong to C is given as:

$$P(C_i|\ d) = \frac{P(d|\ C_i) * P(C_i)}{P(d)} \qquad (10)$$

SVM is one of the most widely used and applied classification methods. It has been successfully applied to many application domains. SVMs are typically used for learning classification, regression, or ranking function. The algorithm works by searching a seperating hyperplane to seperate between samples with a maximal margin [1,5]. The equation for hyperplane is:

$$w^T x + b = 0 \qquad (11)$$

To classify an unseen document d, the sign of $w^T x + b$ must be known [1]. This is further shown as:

$$w^T x_i + b \geq 1 \ \text{ or } \ w^T x_i + b \leq 1 \qquad (12)$$

Decision tree is a way of representing a sequence of rules that leads to a class or value. It consists of three fundamentals: root node, internal node, and leaf node [5]. The algorithm is a tree like structure which classifies an input sample into one of its possible classes [1]. In decision tree classification algorithm, each node specifies a test to be performed on a single attribute [1]. The goal is to create a model that predicts the value of a target variable based on several input variables. The data generally takes the form:

$$(x, Y) = (x_1, x_2, x_3, \cdots, x_k, Y) \qquad (13)$$

3.5. Performance evaluation

Accuracy is one of the standard evaluation measures employed to validate the classifiers' results [5]. Given the confusion matrix, accuracy is obtained using:

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \qquad (14)$$

In addition, AUC values as a validation method also reflects the overall ranking performance of a classifier [5].

4. Results and Discussions

The FS algorithms experimented on the Quranic datasets produced varying results as shown in Tables 2 to 5. The datasets include: *QTrans, QTaf, and QTrans+Taf.*

Table 2: Classification Accuracy using NB classifier

FS Algorithm	QTrans		QTaf		QTrans+Taf	
	Time	ACC (%)	Time	ACC (%)	Time	ACC (%)
All Features	-	83.4	-	87.5	-	90.4
IG	1.30s	91	1.36s	88.3	1.37s	92.6
CFS	144.1s	91	112s	88.3	155.4s	92.5
IG-CFS	2.57s	91	3.53s	88.3	3.89s	92.5

Table 3: Classification Accuracy using libSVM classifier

FS Algorithm	QTrans		QTaf		QTrans+Taf	
	Time	ACC (%)	Time	ACC (%)	Time	ACC (%)
All Features	-	84	-	84.3	-	84.8
IG	1.30s	88.6	1.36s	86.6	1.37s	93.1
CFS	144.1s	90	112s	89.2	155.4s	94.5
IG-CFS	2.57s	90	3.53s	93.3	3.89s	94.5

Table 4: Classification Accuracy using Avg. *k*-NN classifier

FS Algorithm	QTrans		QTaf		QTrans+Taf	
	Time	ACC (%)	Time	ACC (%)	Time	ACC (%)
All Features	-	83.2	-	84.2	-	84.5
IG	1.30s	86.9	1.36s	87.5	1.37s	86
CFS	144.1s	87	112s	90.9	155.4s	89.4
IG-CFS	2.57s	86.9	3.53s	90.9	3.89s	89.4

Table 5: Classification Accuracy using J48 classifier

FS Algorithm	QTrans		QTaf		QTrans+Taf	
	Time	ACC (%)	Time	ACC (%)	Time	ACC (%)
All Features	-	82.3	-	85.7	-	84.1
IG	1.30s	85.7	1.36s	86.6	1.37s	87
CFS	144.1s	85.2	112s	86.7	155.4s	87.3
IG-CFS	2.57s	85.2	3.53s	86.7	3.89s	87.1

From the experimental results, the classifiers were implemented with the entire features generated from the Quranic datasets as well as with the selected features using IG, CFS, and the proposed IG-CFS algorithms. The significant influence of feature selection on classification algorithms could be seen.

In explaining how the complexity of text data could affect the classification process, the entire features generated were implemented on the classifiers. This had the least classification accuracy performance of 82.3% with decision trees *J48* algorithm using the *QTrans* dataset.

Consequently, applying feature selection technique on the feature set in reducing the dimensionality produced the overall highest classification accuracy of 94.5% with libSVM classifier using CFS and the proposed *IG-CFS* algorithms on *QTrans+Taf* dataset. Although, both feature selection algorithms jointly had the highest accuracy result; however, the proposed hybrid technique utilized lower computational runtime of 3.89secs in comparison with CFS with higher runtime of 155.4secs.

This explains why wrapper-based FS algorithms (such as *CFS*) are highly computationally expensive to run but however achieve better results. Furthermore, the results explained why filter-based FS algorithms (such as *IG*) perform less in terms of accuracy but are very simple and easy to run with lower computational runtime.

To sum it up, the study proposed a hybrid FS technique applicable in achieving high classification accuracy performance at less computational runtime. This specific goal as could be seen from the experimental results was achieved with the proposed *IG-CFS* algorithm. In addition, the study further tested the algorithms using AUC performance evaluation metric as visualized in Figures 2 to 5.

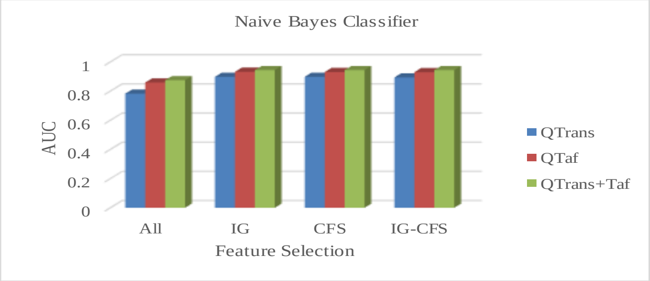

Fig. 2: AUC results with NB

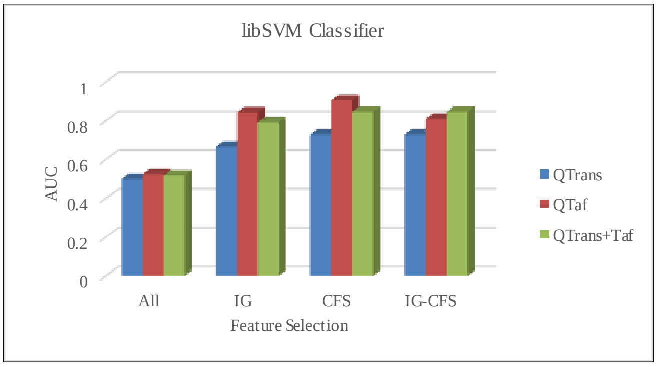

Fig. 3: AUC results with libSVM

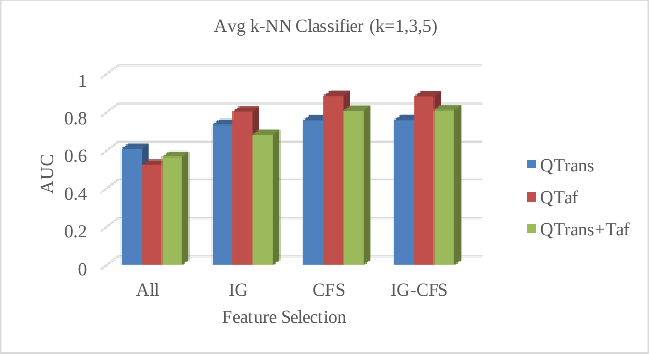

Fig. 4: AUC results with *k*-NN

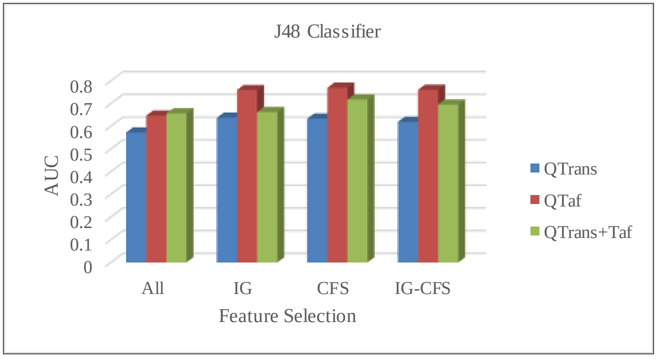

Fig. 5: AUC results with J48

Evaluating the performance of the algorithms with AUC evaluation metric generated promising results as could be seen in the above figures. Implementing with naïve bayes classifier produced the overall best AUC result of 0.944 with the *QTrans+Taf* dataset. The proposed *IG-CFS* algorithm consistently had satisfactory AUC values across the experimental Quranic datasets.

The results of the proposed *IG-CFS* algorithm on the Quranic datasets are further visualized in Figures 6 to 8. Employing ROC curve together with other standard performance metrics show better justifications of the experimental results.

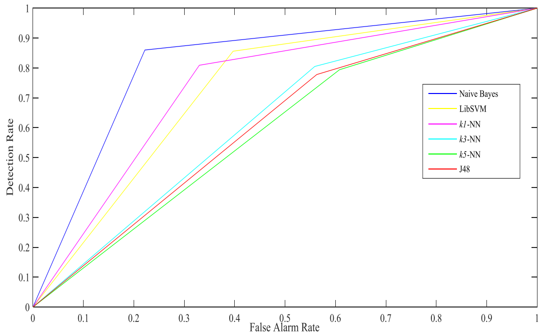

Fig. 6: *IG-CFS* algorithm on *QTrans* dataset

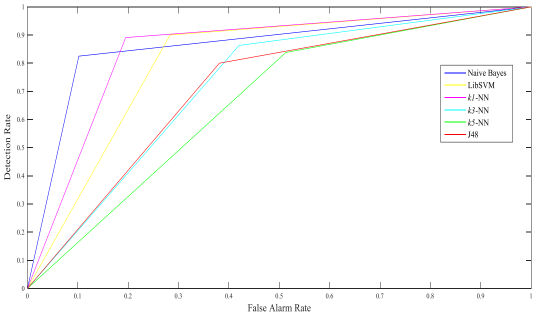

Fig. 7: *IG-CFS* algorithm on *QTaf* dataset

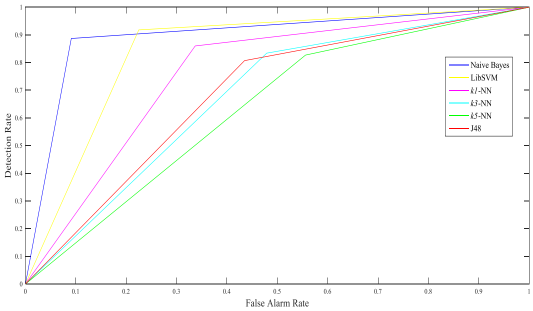

Fig. 8: *IG-CFS* algorithm on *QTrans+Taf* dataset

5. Conclusion

Selecting features is an integral phase of classification problems such as the Holy Quran verses labeling. A hybrid FS technique was proposed and further experimented with classification algorithms. The ultimate goal of the study is to automate the labeling of Quranic verses into the predefined classes: '*iman, ibadah, akhlak*'. Specifically, the study aimed at achieving high classification accuracy at less computational runtime with feature selection algorithms experimented.

The input data (Quranic verses) were normalized while information gain, CFS, and the proposed *IG-CFS* algorithms were adopted and experimented in the classification task.

Subsequently, the experimental results were validated using standard metrics and the proposed *IG-CFS* achieved the overall best performance of 94.5% accuracy result and AUC value of 0.944 at a lower computational runtime of 3.89secs. In addition, the experimental results of the proposed FS technique were further visualized using receiver operating characteristic curve. The experimental results have shown that the feature selection algorithms employed in the proposed approach had significant impacts on the classifiers implemented in the verse classification task.

Conclusively, the research study aims at extending the experimented hybrid FS technique to other classification problems.

Acknowledgement

This research study is supported by the Ministry of Higher Education Malaysia, Fundamental Research Grant Scheme Vot 1609.

References

[1] Adeleke AO, Samsudin NA, Mustapha A & Nawi NM (2017), Comparative Analysis of Text Classification Algorithms for Automated Labelling of Quranic Verses. *Int. J. on Advance Science, Engineering and Info. Tech*. 7, 1419-1427.

[2] Das S, Dey A, Pal A & Roy N (2015), Applications of Artificial Intelligence in Machine Learning: Review and Prospect. *J. of Computer Applications* 115, 31-41.

[3] Talwar A & Kumar Y (2013), Machine Learning: An Artificial Intelligence Methodology. *J. of Engineering and Computer Science* 2, 3400-3404.

[4] Tang J, Alelyani S & Lin H (2014), Feature Selection for Classification: A Review. *In Data Classification: Algorithms and Applications. CRC Press.*

[5] Adeleke AO, Samsudin NA, Mustapha A & Nawi NM (2018), A Group-Based Feature Selection Approach to Improve Classification of Holy Quran Verses. in *R. Ghazali et al. (eds.), Recent Advances on Soft Computing and Data Mining, Advances in Intelligent Systems and Computing 700*, 549, 282-297.

[6] Jamil NS, Ku-mahamud KR, Din AM, Ahmad F, Chepa N, Ishak WHW, Din R & Ahmad FK (2017), A subject identification method based on term frequency technique. *J. of Advanced Computer Research* 7, 103-110.

[7] Goudjil M, Bedda M, Koudil M, & Ghoggali N (2015), Using Active Learning in Text Classification of Quranic Sciences. *Int. Conf. on Advances in Information Technology for the Holy Quran and Its Sciences*, 209-213.

[8] Hassan GS, Mohammad SK & Alwan FM (2015), Categorization of Holy Quran Tafseer' using k-Nearest Neighbour Algorithm. *Int. J. of Computer Applications*, 129, 1-6.

[9] Ibrahim EAA, Ataelfadiel MAM & Atwel ES (2017), Provisions of Quran Tajweed Ontology (Articulations Points of Letters, UN Vowel Noon and Tanween). *Int. J. of Science and Research*, 6, 8, 756-761.

[10] Alqahtani M & Atwell E (2016), Arabic Quranic Search Tool Based on Ontology. *21st Int. Conf. on Applications of Natural Language to Information Systems*, 478-485.

[11] Hamed SK & Ab Aziz MJ (2016), A Question Answering System on Holy Quran Translation Based on Question Expansion Technique and Neural Network Classification. *J. of Computer Sciences*, 12, 3, 169-177.

[12] Abdelnasser H, Mohamed R, Ragab M, Mohamed A, Farouk B & El-Makky N (2014), Al-Bayan: An Arabic Question Answering System for the Holy. *Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing*, 57-64.

[13] Alrehaili SM & Atwell E (2014), Computational Ontologies for Semantic tagging of the Quran: A survey of past approaches. *Ninth Int. Conf. on Language Resources and Evaluation*.

[14] Abdelhamid Y, Mahmoud M & El-Sakka TM (2013), Using Ontology for Associating Web Multimedia Resources with the Holy Quran. *Taibah University Int. Conf. on Advances in Information Technology for the Holy Quran and its Sciences*, 266-271.

[15] Akkila AN & Abu Naser SS (2017), Teaching the right letter pronunciation in reciting the holy Quran using intelligent tutoring system. *Int. J. of Advanced Research and Development*, 2, 1, 64-68.

[16] Ahmed AH & Abdo SM (2017), Verification System of Quran Recitation Recordings. *Int. J. of Computer Applications*, 163, 4, 6-11.

[17] Aljaloud HO, Dahab M & Kamal M (2016), Stemmer Impact on Quranic Mobile Information Retrieval Performance. *Int. J. of Advanced Computer Science and Applications*, 7, 12, 135-139.

[18] Zharmagambetov AS & Pak AA (2015), Sentiment analysis of document using deep learning and decision trees. *Twelve IEEE Int. Conf. on Electronics Computer and Computation*, 1-4.

[19] Wang JH & Wang HY (2014), Incremental Neural Network Construction for Text Classification. *IEEE Int. Symposium on Computer Consumer and Control*, 970-973.

[20] Sabbah T & Selamat A (2014), Support Vector Machine based approach for Quranic words detection in online textual content. *8th IEEE Malaysian Software Engineering Conference*, Malaysia, 325-330.

[21] Townsend KR, Sun S, Johson T, Attia OG, Jones PH, and Zambreno J (2015), k-NN text classification using an FPGA-based sparse matrix vector multiplication accelerator. *IEEE Int. Conf. on Electro/Information Technology*, 257-263.

[22] Aladeemy M, Tutun S & Khasawneh MT (2017), A new hybrid approach for feature selection and support vector machine model selection based on self-adaptive cohort intelligence. *Expert Systems with Applications*, 88, 118-131.

[23] Wang H & Liu S (2016), An Effective Feature Selection Approach Using the Hybrid Filter Wrapper. *Int. J. of Hybrid Information Technology*, 9, 1, 119-128.

[24] Uysal AK (2016), An improved global feature selection scheme for text classification. *Expert Systems with Applications*, 43, 82-92.

[25] Ghareb AS, Abu Bakar A & Hamdan AR (2016), Hybrid feature selection based on enhanced genetic algorithm for text categorization," *Expert Systems with Applications*, 49, 31-47.

[26] Hancer E, Xue B & Zhang M (2017), Differential evolution for filter feature selection based on information theory and feature ranking," *Knowledge-Based Systems*, 000, 1-17.

[27] Feng PM, Ding H, Chen W & Lin H (2015), Naive Bayes Classifier with Feature Selection to Identify Phage Viron Proteins. *Computational and Mathematical Methods in Medicine.*

[28] Pashaei E & Aydin N (2017), Binary black hole algorithm for feature selection and classification on biological data. *Applied soft computing,* 56, 94-106.

[29] Novakovic J (2009), Using Information Gain Attribute Evaluation to classify Sonar Targets. *17th Telecommunications Forum,* 1351-1354.

[30] Zhuo L, Zheng J, Wang F, Li X, Ai B & Qian J (2008), A Genetic Algorithm based Wrapper Feature Selection method for Classification of Hyperspectral Images using Support Vector Machine, *The Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVII, 397-402.

[31] Veeraswamy A & Balamurugan SA (2013), An Effective Performance of Feature Selection with Classification of Data Mining Using SVM Algorithm, *Proceedings of the National Conf. on Recent Trends in Mathematical Computing*, 427-431.

[32] Mansoori TK, Suman A & Mishra SK (2014), Feature Selection by Genetic Algorithm and SVM Classification for Cancer Detection, *Int. J. of Advanced Research in Computer Science and Software Engineering,* 4, 357-365.

[33] Molano V, Cobos C, Mendoza M, Viedina EH & Manic M (2011), Feature Selection based on sampling and C4.5 Algorithm to improve the Quality of Text Classification using Naïve Bayes, *Springer*.