Improving handwritten digit recognition using hybrid feature selection algorithm

Fung Yuen Chin, Kong Hoong Lem and Khye Mun Wong
Department of Physical and Mathematical Science,
Universiti Tunku Abdul Rahman–Kampus Perak, Kampar, Malaysia

Abstract

Purpose – The amount of features in handwritten digit data is often very large due to the different aspects in personal handwriting, leading to high-dimensional data. Therefore, the employment of a feature selection algorithm becomes crucial for successful classification modeling, because the inclusion of irrelevant or redundant features can mislead the modeling algorithms, resulting in overfitting and decrease in efficiency.

Design/methodology/approach – The minimum redundancy and maximum relevance (mRMR) and the recursive feature elimination (RFE) are two frequently used feature selection algorithms. While mRMR is capable of identifying a subset of features that are highly relevant to the targeted classification variable, mRMR still carries the weakness of capturing redundant features along with the algorithm. On the other hand, RFE is flawed by the fact that those features selected by RFE are not ranked by importance, albeit RFE can effectively eliminate the less important features and exclude redundant features.

Findings – The hybrid method was exemplified in a binary classification between digits "4" and "9" and between digits "6" and "8" from a multiple features dataset. The result showed that the hybrid mRMR + support vector machine recursive feature elimination (SVMRFE) is better than both the sole support vector machine (SVM) and mRMR.

Originality/value – In view of the respective strength and deficiency mRMR and RFE, this study combined both these methods and used an SVM as the underlying classifier anticipating the mRMR to make an excellent complement to the SVMRFE.

Keywords Handwritten digit recognition, Feature selection, mRMR, SVMRFE

Paper type Research paper

1. Introduction

1.1 Handwritten digit recognition

High-dimensional data, no doubt, will cause a whole load of problems toward classification accuracy. A large number of features will only create unnecessary noise and affect the performance of predictive modeling [1]. Therefore, feature selection will be needed to select only features that are relevant, nonredundant and consistent. This will decrease the feature space and hence allow a more useful feature to build an effective model [2].

Feature selection plays an important role in the preliminary stage of classification. It is impractical to have a lot of irrelevant and redundant features present in the dataset because it reduces the efficiency of the model [3]. In actual practice, due to variations of handwriting style, strokes, resemblance in outline and other additional noise from individuals, a number of features for handwritten digits are often largely resulting in these data normally appearing to

be high dimensional too. Therefore, feature selection will come into play and reduce the number of handwritten digit features and improve the recognition speed. Existing feature selection such as support vector machine recursive feature elimination (SVMRFE) is able to build a predictive model which has high accuracy; however, this method is not able to rank the selected features according to their importance. Therefore, the first selected feature may not be the most important. Minimum redundancy and maximum relevance (mRMR) can select the most relevant features, and at the same time, this method might also select the redundant features [4]. When building the predictive model, the redundant features will increase the complexity of the model, and the model will tend to be overfitting as well.

The presence of distorted characters and high similarities between outlines of certain digits give rise to redundancy in classification. Therefore, in handwritten digit recognition, the implementation of one feature selection method alone might not be enough to yield an optimal classification accuracy. A hybrid feature selection method is proposed in this study to combine the advantages and overcome the shortage of the mRMR and the SVMRFE methods. The hybrid feature selection works better than a single feature selection algorithm in improving the performance of the predictive model using a small number of features.

1.2 Motivation and main contribution
SVMRFE algorithm generally repeatedly removes features having the lowest weighted values. However, the top-ranked feature (the lastly removed) is not necessarily the most relevant one [5]. This gives a drawback that unless many features are used, the algorithm might not perform well when only one or two features are used. On the other hand, mRMR is an effective method that uses mutual information to search for high-relevance and low-redundancy features. Nevertheless, there is a trade-off between relevance and redundancy. This has motivated us to combine the two methods, complementing their shortcomings mutually. In this study, we tried to embed the highly relevant features shortlisted by the mRMR in the SVMRFE hoping to alleviate the ranking issue of SVMRFE and the redundancy issue of mRMR. In addition, the goal is to create an approach, which can produce better classification by using only the first few most significant features in handwritten digit recognition.

The proposed hybrid idea was tested on the binary classification between digits "4" and "9" and between digits "6" and "8." The classification performance of the hybrid method outperformed the mRMR, the SVMRFE and the ReliefF methods in comparison.

The contribution of this article is as follows: (1) proposing a framework to combine a filter method with an embedded method in the area of feature selections that compensates for the weakness of each other; (2) creating a mRMR-SVMRFE hybrid algorithm in handwritten digit recognition, and it not only serves as a new alternative in handwritten digit recognition but may also be further applied to other classification problems besides handwriting and (3) analyzing the characteristic of the hybrid method shows that its strength lies in the ability to select and rank the most significant features, and it can give good classification performance only by using a few features. This is very valuable, for example, in the fields of feature selection in biomarker discovery, where more features will lead to more money and time.

The rest of this article is organized as follows: Section 2 gives a brief description of the related works, Section 3 introduces the proposed hybrid method, Section 4 presented the experimental results and Section 5 concludes the study and discusses potential extensions.

2. Literature review
2.1 Dimension reduction
The presence of high-dimensionality data has increased the cost and prolonged the time for classification and other data mining analysis [6]. The optimal solution is to use the dimension

reduction method as a data preprocessing step in reducing the complication and eliminating the redundant and irrelevant features in high-dimensional data. According to Pino and Morell [7], feature selection has been an ever-evolving problem due to the rise of big data in recent years. Feature selection aims to find the smaller number of essential features out of the high-dimensional data, containing the best subset features with the least number of dimensions to improve the classification accuracy [8]. The three main groups of feature selection consist of the filter method, wrapper method and embedded method. The filter method employs the statistical way of evaluating each subset without the dependence on the classifier [9]. The wrapper method, on the other hand, will be classifier dependent, and it utilizes a machine learning algorithm to find out the prediction power gained in the evaluated dataset. Therefore, it will cause computational complexity as the validation process takes place for every subset evaluated. The embedded method learns the best attributes for improving the accuracy of the predictive model when the model is set. The embedded method integrates the feature selection process with the model training process, and both processes are completed in an optimization process. The mRMR is a filter method, and the SVMRFE is an embedded method.

On the other hand, feature extraction is a process where it transforms the feature from a high-dimensional space into a lower-dimensional space by using the fusion of the first and original feature, thus keeping the most relevant information for further classification process [10]. Some examples of feature extraction methods include principal component analysis (PCA), latent semantic analysis (LSA), linear discriminant analysis (LDA), independent component analysis (ICA), partial least square (PLS), etc. Among the feature extraction methods, PCA, ICA and PLS stand out the most as they are the most effective methods in extracting important features [11].

2.2 Mutual information and mRMR
Mutual information (MI) measures the information shared between the two random discrete variables x and y. It can also be interpreted as how much does random variable talks about another. The complete formula for MI is defined as follows:

$$I(x,y) = \sum_{y} \sum_{x} p(x,y) \cdot \log\left(\frac{(p(x,y)}{p(x) \cdot p(y)}\right) \quad (1)$$

where $p(x,y)$ is the joint probability of x and y.

However, MI becomes less efficient whenever there is a large dimensional feature input vector, particularly when the number of samples and computational time is taken into consideration [12]. Battiti overcame the issue by adopting the MI feature selector (MIFS) method. MIFS is a greedy feature selection algorithm that considers the most relevant feature k out from the original set of features, n and also the mutual information to the output class. MIFS can solve the weakness in MI by optimizing the information about the class and subtracting the quantity proportional to MI with the previously selected feature.

Studies in Kwak and Choi [13] found out that there was still a limitation in the MIFS proposed by Battiti. [12]. They instead proposed a better solution method known as MIFS-U. MIFS-U is better in terms of obtaining a more precise estimation between input features and output class in MI than MIFS. Despite MIFS-U being a better feature selection algorithm than MIFS, there are still some limitations between these two methods [14].

The redundancy issue in MIFS-U was then minimized by using a method called mRMR proposed by Peng *et al* [4]. The maximal relevance of MI will enhance the minimum redundancy criterion to become more representative of the target features. However, it was also claimed that mRMR might select a high-relevant feature which also caused high

redundancy at the same time because the selection was based on the difference between relevancy and redundancy [15].

2.3 Feature ranking with recursive feature elimination (RFE)
In supervised learning, a predictive model often oversees the features inside a dataset, hence jeopardizing its ability to generalize well. When a predictive model includes the noise in a limited-size training dataset instead of focusing on learning the meaning behind the data features, its predictive power will decrease [16]. The recursive feature elimination (RFE) method which was first introduced by Guyon *et al.* [17] can effectively increase the accuracy by eliminating uncorrelated noise and irrelevant features. RFE is an embedded feature selection that recursively eliminates the features which are irrelevant and have small feature weight. In every iteration, RFE orderly discards the worst feature that affects the classification accuracy. RFE approach is frequently integrated with the support vector machine (SVM) classifier to form the SVMRFE [18].

2.4 Feature selection in handwritten digit recognition
Supplementary material at https://docs.google.com/document/d/1kNA-NVSVpNUc46p c1Zg_K1sXD8vMCrWE/edit?usp=sharing&ouid=106536917224200212284&rtpof=true& sd=true shows the previous studies in handwritten digits recognition [19–27]. The past research focused more on the use of machine learning algorithms such as artificial neural network (ANN), convolution neural network (CNN), *k*-nearest neighbor (KNN) and correlation features selection (CFS) in building the handwritten digits recognition predictive model. The ReliefF algorithm searches for a subset of features with a minimum error rate, while histogram of oriented gradients (HOG) is a preprocessing method to extract the image of handwritten digits before applying the filter method. The ReliefF algorithm uses the feature value to rank the features, where the feature value is the distance between the nearest neighbor pair of features.

The chemical reaction optimization (CRO) is a feature extraction method to select a subset of features with a minimum recognition rate and a minimum recognition cost. The Quantum *k*-neighbor algorithm transforms the classical information of handwritten into quantum information to speed up the computation time in building the handwritten digits recognition classification model. Memory-based histogram-oriented multiobjective genetic algorithm (M-HMOGA) uses a genetic algorithm, and it is an enhanced method that includes a memory to keep track of the best solutions in classification. Spiking neural network (SNN) is composed of three spiking neural layers and one output neuron.

Previous studies have used the machine learning algorithm to minimize the error rate or used filter methods to search the minimum number of feature subsets, but there are not many studies that combine machine learning algorithms with filter methods.

3. Material and methodology
3.1 The dataset
The dataset used in this paper was the multiple feature (MFEAT) dataset [28]. It was a dataset that consists of features of handwritten digits (0–9) extracted from a collection of Dutch utility maps. The rows represented the number of samples present in the dataset, and the columns represented the handwritten digits. This dataset contained a total of 649 features and 2,000 samples. The two datasets selected in this study were digits "4" and "9" and digits "6" and "8". Those sets of digits were selected due to the occurrence of the misleading contour of handwriting and the high resemblance between these two digits.

3.2 Minimum redundancy and maximum relevance (mRMR)

MI, $I(x, y)$ is used in mRMR to find the maximum dependency within a set of attributes and its given label class. There are two stages for mRMR in choosing the optimal subset of the feature. The first step is to apply the maximum relevance, which will be used to select a set of features (S) with features $\{x_i\}$ that contain the most relevant information to their class label, h [15]. The relevance formula is as follows:

$$\max\ V_i = \frac{1}{|S|} \sum_{i \in S} I(h, x_i) \tag{2}$$

where $|S|$ is the number of features in the set S.

The second step is to minimize the redundancy among the features because redundancy features provide no useful information for the classification model [4]. The minimum redundancy concept is to choose the features that have mutually dissimilar traits. The minimum redundancy condition is as follows:

$$\min\ W_i = \frac{1}{|S|^2} \sum_{i,j \in S} I(x_i, x_j) \tag{3}$$

A set of features of mRMR will be acquired based on the combination of equations (2) and (3) to form a single selection criterion in equation (4) known as the "minimum-redundancy-maximum relevance" criterion.

$$\text{mRMR} = V_i - W_i \tag{4}$$

3.3 Support vector machine recursive feature elimination (SVMRFE)

SVMRFE is a feature selection method that utilizes the criteria acquired from the SVM's coefficient to choose selected features and recursively remove features that contain fewer criteria or weight in a backward elimination manner. SVMRFE does not rely on cross-validation accuracy to determine the relevant features from the training data. The algorithm trains the model using every feature, meanwhile the contribution of each feature is evaluated. Less significant features are eliminated repeatedly until all features are traversed. Thus, it exhibits robustness to prevent overfitting even for data containing thousands of features [29].

Generally, the selection of relevant features for SVMRFE can be divided into three stages. First, the input data will be inserted into the classifier SVM for classification. The second stage will involve the process of calculation for all of the features in terms of ranking weights. The deletion of features that have a smaller ranking weight is performed at the last stage [30].

Under the SVM, let $X = [x_1, x_2, \ldots, x_k]^T$ be the input training data and $Y = [y_1, y_2, \ldots, y_k]^T$ be the class label of X, and the ranking score of the trained features will be computed according to the weight vector, w.

$$w = \sum_{k=1}^{n} a_k y_k x_k \tag{5}$$

Here, a_k is the Lagrange multiplier involved in maximizing the margin of separation of class labels and n is the number of features.

The ranking criterion C_k for the surviving feature will be computed by obtaining the square of the k-th feature of the weight vector, w.

$$C_k = w_k^2 \quad, \ k = 1, 2, 3, \ldots \tag{6}$$

The feature that has the smallest ranking criterion will be identified and eliminated. For each iteration of RFE, an SVM model is trained and the surviving features will be kept for the next iteration. The process keeps on repeating until all of the features are discarded, and then, they will be sorted according to the removal sequence. The later a feature being discarded, the more significant that feature is and will be given a higher rank. The process eventually produces an optimal feature subset [28].

3.4 Proposed hybrid method

The mRMR was applied to rank the features according to equation (4), and the shortlisted features contained the most relevant features. This process reduced the high-dimensional data to a smaller dataset. The weight, w of each feature from the shortlisted features was calculated. The weights of the features were then sorted in descending order, and the feature having a smaller weight value was eliminated from the list of surviving features. The process was repeated until all of the features with smaller ranking criteria were removed such that no more features were left for training. At the end of the iteration, the desired number of selected features will be obtained using RFE as a feature ranking mechanism. Figure 1 shows the flowchart of the proposed hybrid method.

In implementing the mRMR algorithm, the number of features to keep k has to be preset by the researcher. Here we arbitrarily took $k = 15$ throughout. The dataset was split into a training set and a test set according to the ratio of 7:3. After the splitting process, mRMR was applied to the training set to rank the features according to equation (4), and the most relevant features would be shortlisted. This process would reduce the high-dimensional data into lower-dimensional data which would decrease the computational time in SVMRFE. In SVMRFE, the weight of each shortlisted feature was calculated according to equations (5) and (6). The predicting model would then be built, and the test set would be used in this model to obtain the classification accuracy.

It has been proven that mRMR is good at selecting the most relevant features, but it also includes some redundant features in the process. On the other hand, SVMRFE as an embedded method will lead to high computational cost and time for high classification accuracy. Therefore, as a filter method that requires less computation time, mRMR can first screen the number of features to reduce the computation time of SVMRFE, and SVMRFE can solve the redundancy issue faced by mRMR. This motivates the intention to combine these two algorithms to obtain an optimal subset of features by complementing each other's constraints.

3.5 Performance metric and model comparison

To indicate the superiority of the proposed hybrid method, two extra predictive models, namely the mRMR and the SVMRFE, were built for comparison. The performance metrics
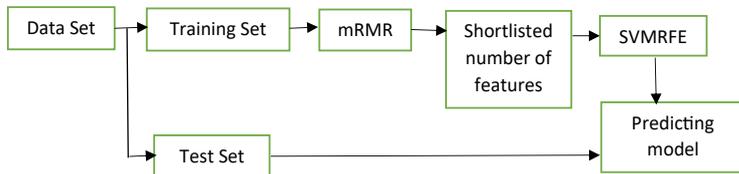

Figure 1.
Flowchart of the proposed hybrid method

used were (1) cross-validation accuracy, (2) test accuracy and (3) area under the curve (AUC). The accuracy is defined as follows:

$$accuracy = \frac{true\ positive + true\ negative}{false\ negative + false\ positive + true\ positive + true\ negative} \quad (7)$$

4. Experimental result and discussion

Four methods, namely the mRMR, SVMRFE, ReliefF and the hybrid mRMR + SVMRFE, were employed to perform the 4-9 classification and the 6-8 classification. The digits "4" and "9" and "6" and "8" were chosen because of the high similarity between these two numbers. The cross-validation accuracy and the test accuracy versus the number of features are presented in Figures 2 and 3, respectively. The accuracy curves for mRMR in Figures 2 and 3 showed up–down fluctuation when more features were included. This revealed the fact that while the mRMR method selects the most relevant features, it also includes some redundant features during the process.

The performance of the SVMRFE was good only when more features were included in the predictive model. It was obvious that SVMRFE gave the lowest accuracy compared to the other two methods if only the first feature was included. This showed that the first feature

Note(s): The comparison was for digits "4" and "9" (left) and for digits "6" and "8" (right)
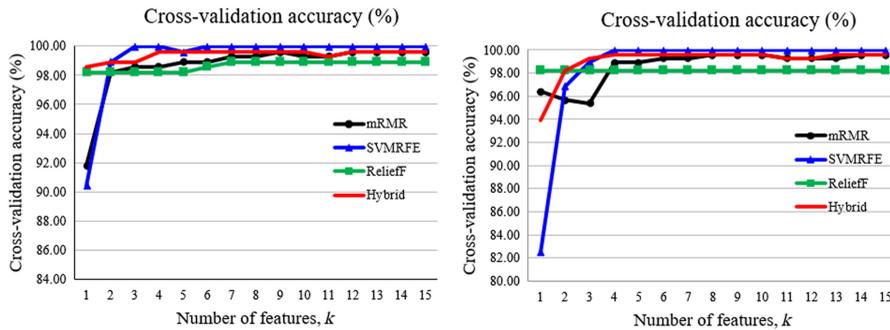
Figure 2.
Comparison of cross-validation accuracy among mRMR, SVMRFE, ReliefF and hybrid method (mRMR + SVMRFE)

Note(s): The comparison was for digits "4" and "9" (left) and for digits "6" and "8" (right)
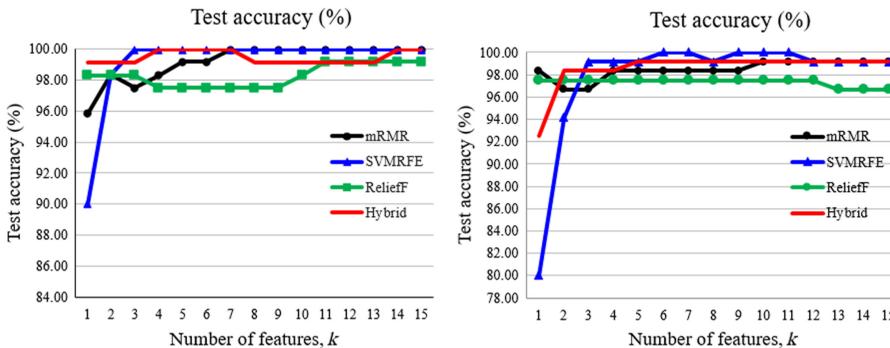
Figure 3.
Comparison of test accuracy among mRMR, SVMRFE, ReliefF and hybrid method (mRMR + SVMRFE)

from the SVMREF-selected feature subset was not necessarily the most significant one. The fact that the features selected by SVMREF are not ranked in the order of importance was disclosed here.

As an additional comparison, the ReliefF method showed slight up–down fluctuation in the accuracy curve of Figure 2 (left) and Figure 3 when additional features were included. This revealed the deficiency of the ReliefF method in removing irrelevant and redundant features. Therefore, when additional features were included in the model, compared with the SVMRFE method, it leads to the loss of accuracy consistency.

Among these methods, the proposed hybrid method using digits "4" and "9" yielded the highest accuracy when only one feature was selected as shown in Figure 2. In comparison, the proposed hybrid method using digits "6" and "8" could achieve the highest accuracy when two features were selected as observed in Figure 3. Unlike the mRMR and ReliefF methods, the hybrid method performed more stable when more features were added in. This hybrid method can effectively extract all high-relevance features using mRMR. When combined with SVMRFE, the predictive model can achieve high accuracy by using only one or two features. Results showed that the hybrid method managed to improve the performance of the classification by addressing the redundant features and the ranking issue in the SVMRFE.

The average AUC and the average accuracy of the test data for the four models using two sets of binary digits are summarized in Table 1. As can be seen from the table, the average classification accuracy of two sets of binary digits in the four models achieved more than 90%. The comparison showed that the hybrid model exhibited the highest classification accuracy among the four models for binary digits "4" and "9," with an accuracy of 99.45%, followed by SVMRFE (99.31%), then ReliefF (98.69%) and lastly mRMR (98.65%). Also, the hybrid model using binary digits "6" and "8" yielded the highest classification accuracy of 99.04%, followed by mRMR (98.65%), then SVMRFE (98.54%) and lastly ReliefF (98.25%). This was evidence that the feature selection combination of mRMR and SVMRFE outperformed the single feature selection.

Besides, the AUC for digits "4" and "9" had also been greatly optimized by the hybrid method to reach the value of 1. The average AUC for digits "6" and "8" achieved the highest average AUC value of 0.9993 as compared to mRMR, SVMRFE and ReliefF. As a whole, the implementation of the hybrid method had been proven to improve the handwritten digit feature classification accuracy compared to mRMR, SVMRFE and ReliefF.

5. Conclusion and future works

A hybrid method was proposed and tested on the 4-9 and 6-8 binary classification. It achieved relatively higher classification accuracy in terms of average AUC and average classification accuracy for the top 15 ranked features. It gave more stable results when more features were included. The hybrid approach can be a feasible option for better classification when using only a few most significant features.

Since datasets may not be linearly separable, SVM can be implemented on different kernels in which the performance of each kernel is compared to ensure classification accuracy

Table 1.
The average AUC and the average accuracy of test data among mRMR, SVMRFE, ReliefF and proposed hybrid method for the top 15 ranked features

MFEAT Dataset (digit)	Average AUC				Average accuracy (%)			
	mRMR	SVMRFE	ReliefF	Hybrid	mRMR	SVMRFE	ReliefF	Hybrid
"4" and "9"	0.9960	0.9973	0.9960	*1.0000*	98.65	99.31	98.69	*99.45*
"6" and "8"	0.9960	0.9933	0.9873	*0.9993*	98.65	98.54	98.25	*99.04*

and stability. In fact, apart from SVM, the hybrid strategy can be incorporated with other classifiers (such as KNN, decision tree, random forest, etc.) for further study.

The proposed method may not benefit low-dimensional data because the need does not arise for such data. For imbalanced data, it requires further studies to avoid classification bias and overfitting. Meanwhile, the implementation of microarray data analysis and biomarker discovery might be a potential future direction.

References

1. AlNuaimi N, Masud MM, Serhani MA, Zaki N. Streaming feature selection algorithms for big data: a survey. Appl Comput Inform. 2022; 18: 113-35. doi: 10.1016/j.aci.2019.01.001.

2. Serpush F, Rezaei M. Complex human action recognition using a hierarchical feature reduction and deep learning-based method. SN Comput Sci. 2021; 2: 94. doi: 10.1007/s42979-021-00484-0.

3. Zhou H, Wang X, Zhang Y. Feature selection based on weighted conditional mutual information. Appl Comput Inform. 2020. doi: 10.1016/j.aci.2019.12.003.

4. Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell. 2005; 27(8): 1226-38.

5. Jeon H, Oh S. Hybrid-recursive feature elimination for efficient feature selection. Appl Sci. 2020; 10(9): 3211.

6. Niedzielewski K, Marchwiany ME, Piliszek R, Michalewicz M, Rudnicki W. Multidimensional feature selection and high performance ParalleX. SN Comput Sci. 2020; 1: 40. doi: 10.1007/s42979-019-0037-5.

7. Pino A, Morell C. Analytical and experimental study of filter feature selection algorithms for high-dimensional datasets. In: Proceedings of the Fourth International Workshop on Knowledge Discovery, Knowledge Management and Decision Support; 2013. 339-49.

8. Kalina J, Schlenker A. A robust supervised variable selection for noisy high-dimensional data. Biomed Res Int. 2015; 2015: 1-10.

9. Sonowal G. Phishing email detection based on binary search feature selection. SN Comput Sci. 2020; 1: 191. doi: 10.1007/s42979-020-00194-z.

10. Aziz R, Verma C, Srivastava N. A novel approach for dimension reduction of microarray. Comput Biol Chem. 2017; 71: 161-9.

11. Velliangiri S, Alagumuthukrishnan S, Thankumar Joseph S. A review of dimensionality reduction techniques for efficient computation. Proced Comput Sci. 2019; 165: 104-11.

12. Battiti R. Using mutual information for selecting features in supervised neural net learning. IEEE Trans Neural Netw. 1994; 5(4): 537-50.

13. Kwak N, Choi C. Input feature selection for classification problems. IEEE Trans Neural Netw. 2002; 13(1): 143-59.

14. Estevez P, Tesmer M, Perez C, Zurada J. Normalized mutual information feature selection. IEEE Trans Neural Netw. 2009; 20(2): 189-201.

15. Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. J Bioinform Comput Biol. 2005; 3(2): 185-205.

16. Ying X. An overview of overfitting and its solutions. J Phys Conf Ser. 2019; 1168(022022): 1-6.

17. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Machine Learn. 2002; 46(1/3): 389-422.

18. Zhou Q, Hong W, Shao G, Cai W. A new SVM-RFE approach towards ranking problem. In: 2009 IEEE International Conference on Intelligent Computing and Intelligent Systems; 2009. 270-3. doi: 10.1109/ICICISYS.2009.5357684.

19. Durgabai R, Bhushan YR. Feature selection using ReliefF algorithm. Int J Adv Res Comput Commun Eng. 2014; 3: 8215-18.

20. Ghosh S, Bhowmik S, Ghosh KK, Sarkar R, Chakraborty S. A filter ensemble feature selection method for handwritten numeral recognition. Electron Med Rec. 2016; 7213: 007213.

21. Islam KT, Mujtaba G, Raj RG, Nweke HF. Handwritten digits recognition with artificial neural network. In: 2017 International Conference on Engineering Technology and Technopreneurship (ICE2T); 2017. 1-4.

22. Boni PK, Abir BS, Hasan HM, Islam MR. Handwritten Bangla digit recognition using chemical reaction optimization. In: 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT); 2018. 1-7.

23. Pratt S, Ochoa A, Yadav M, Sheta A, Eldefrawy M. Handwritten digits recognition using convolution neural networks. J Comput Sci Coll. 2019; 34(5): 40-46.

24. Abdulrazzaq MB, Saeed JN. A comparison of three classification algorithms for handwritten digit recognition. In: 2019 International Conference on Advanced Science and Engineering (ICOASE); 2019. 58-63.

25. Cilia ND, De Stefano C, Fontanella F, di Freca AS. A ranking-based feature selection approach for handwritten character recognition. Pattern Recognit Lett. 2019; 121: 77-86.

26. Guha R, Ghosh M, Singh PK, Sarkar R, Nasipuri M. M-HMOGA: a new multi-objective feature selection algorithm for handwritten numeral classification. J Intell Syst. 2020; 29(1): 1453-67.

27. Faghihi F, Alashwal H, Moustafa AA. A synaptic pruning-based spiking neural network for hand-written digits classification. Front Artif Intell. 2022; 5. doi: 10.3389/frai.2022.680165.

28. Blake C, Merz C. UCI repository of machine learning databases. Available from: http://archive.ics.uci.edu/ml/index.php.

29. Jeon H, Oh S. Hybrid-recursive feature elimination for efficient feature selection. Appl Sci. 2020; 10(9): 3211.

30. Yan K, Zhang D. Feature selection and analysis on correlated gas sensor data with recursive feature elimination. Sens Actuators B: Chem. 2015; 212: 353-63.

Corresponding author
Fung Yuen Chin can be contacted at: chinfy@utar.edu.my