2019 International Conference on Advanced Science and Engineering (ICOASE)
University of Zakho, Duhok Polytechnic University, Kurdistan Region, Iraq

A Comparison of Three Classification Algorithms for Handwritten Digit Recognition

Maiwan Bahjat Abdulrazzaq
Department of Computer Science
University of Zakho
Duhok, Kurdistan Region, Iraq
maiwan.abdulrazzaq@uoz.edu.krd

Jwan Najeeb Saeed
Department of Information Technology
Duhok Polytechnic University
Duhok, Kurdistan Region, Iraq
jwan.najeeb@dpu.edu.krd

***Abstract*—Handwritten digits recognition is considered as a core to a diversity of emerging application. It is used widely by computer vision and machine learning researchers for performing practical applications such as computerized bank check numbers reading. However, executing a computerized system to carry out certain types of duties is not easy and it is a challenging matter. Recognizing the numeral handwriting of a person from another is a hard task because each individual has a unique handwriting way. The selection of the classifiers and the number of features play a vast role in achieving best possible accuracy of classification. This paper presents a comparison of three classification algorithms namely Naive Bayes (NB), Multilayer Perceptron (MLP) and K_Star algorithm based on correlation features selection (CFS) using NIST handwritten dataset. The objective of this comparison is to find out the best classifier among the three ones that can give an acceptable accuracy rate using a minimum number of selected features. The accuracy measurement parameters are used to assess the performance of each classifier individually, which are precision, recall and F-measure. The results show that K_Star algorithm gives better recognition rate than NB and MLPas it reached the accuracy of 82.36%.**

***Keywords*—handwriting digit recognition, naïve Bayes, multilayer perceptron, k-star, and classification algorithms**

I. Introduction

Handwritten digit recognition (HDR) is a crucial issue in computer vision. Numerous studies have been done in this area to overcome some challenges of it because hand written digits are different in size, thickness and orientation [1, 2]. The fundamental object of such studies is to minimize classification error, and in some situations they attempt to reduce computational complexity[3]. The investigators get motivations on this research because HDR has a diversity of applications. For instance, postal system computerization, automated bank check processing, writer's verification, passports analysis, and plate number identification [4-6].

Kaushik et al. [7] presented a comparison for the performance evaluation of three different classifiers (NB, J48 , and SMO) they focused on the impact of feature selection techniques and engineering in the classification for handwritten text using 524features.In their comparison, MNIST dataset is used to train the classifiers. They conducted that both NB and SMO proved better enhancement in the term of accuracy than J48 relating to role of the way of selecting features as 71.19% and 89.95% respectively.

Pramanik et al. [8] proposed a way to get the number of components exists in a word image. Cumulative stretch and shadow based features are extracted from each of these components. Then they used a set of classifiers namely MLP, SVM, K-star, Naive Bayes, and Random Forest to find whether the component needs more segmentation or not. Their proposed method used 204 features. The Random Forest showed best result compared to others classifiers as it found 1.37% of misclassification rate.

Othman et al. [9] presented a comparative study among the NB statistical approach and MLP neural network approach. The key point of their study relied on the bias and variance of the error rate. Their work is implemented on simulations and handwritten digit images using MNIST dataset for only 3 digits from (0-3) each with 400 images utilizing 14 features. They suggest a novel stander for comparing neural and Bayesian classifiers. Its uniqueness lies on the stability comparison of each classifier. Their results show that the statistical approaches' have more stability rather than neural networks' one. Thus, in spite of the considerable advantages of the neural approach, the lack of control over its mathematical formulation shows the instability of its classification results.

In this paper a comparison method between three classification algorithms namely, Naive Bayes (NB), Multilayer Perceptron (MLP), and K_Star (K*) has been done on the NIST dataset based on correlation features selection. Since the number of selected features plays a vital role in the classification process, this comparison method considerably used a fewer features than what were used in [7] and [8].The number of digits and instances in this comparison are noticeably more than what had been used in [9].The organization of this paper is as follows: Section IIillustrates the methodology and depicts the dataset. The comparison method is presented in section III. Section IV shows the results and discussion. The last section includes the conclusion

II. Methodology

A. The NIST Dataset

The research uses NIST dataset which is a set of manually written digits data from National Institute of Standard and Technology as the original dataset. Gotten from digits written by high school students and employees of the United States Census Bureau[10, 11]. A random subset consists of (46080) images are selected to ensure that each digits from 0 – 9 has an

978-1-5386-9343-8/19/$31.00 ©2019 IEEE

equal numbers of instances. The size of image is 128 × 128 (16384 pixels). The images are converted to binary images then it resized to 16×16 pixels using. Samples of NIST dataset can be seen in Fig.1.

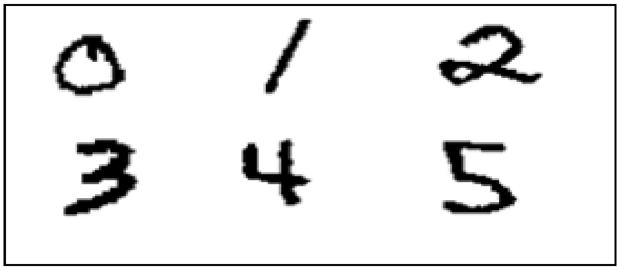

Fig. 1. Samples of the NIST dataset.

There are some confusing images in the NIST dataset that are hard to be recognized by the classifiers as it is human handwritten digits and each person has its own writing style as shown in Fig.2. This may effect on the classification process particularly when it is taken randomly.

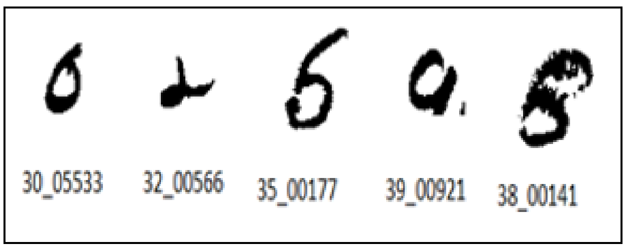

Fig. 2. Samples of some confusing images in NIST dataset.

B. Feature selection

Feature selection (FS) refers to the process of selecting the best featuresubset, which is crucial and popular method[12].Both rankers and subset evaluators are the two kinds of feature selection techniques. The first technique assesses each feature separately utilizing various statistical measures. While, subset-based feature selection techniques evaluate all subsets simultaneously, either using statistical measures (Filter-based subset selection) or using a classifier (wrapper-based feature selection)[13]. The majority of research work emphasizes on ranker-based feature selection techniques because they cost less in terms of computation than other techniques of feature selection[14].

Correlation-based feature selection (CFS) has been used in the comparison method. It ranks the features in sequential order. This means the best feature has the highest rank and so on. The first 37 features are selected to be used in the three classifiers. The selected 37 features can be considered as the minimum number of features that show the differences between the three classifiers with acceptable average accuracy. Table I. shows the correlation-based feature selection ranking of 37 attributes.

TABLE I. CORRELATION BASED FEATURES SELECTION RANKING

No.	Attribute	Rank
1	22	0.14122
2	23	0.15338
3	24	0.15712
4	25	0.15745
5	26	0.16062
6	27	0.14074
7	38	0.15677
8	39	0.1569
9	40	0.14765
10	41	0.17251
11	42	0.17481
12	53	0.14189
13	54	0.15188
14	57	0.1571
15	58	0.14228
16	73	0.14466
17	114	0.14692
18	130	0.15309
19	134	0.14358
20	146	0.14613
21	150	0.15637
22	165	0.15563
23	166	0.16085
24	181	0.17813
25	182	0.1428
26	196	0.1536
27	197	0.14857
28	211	0.15249
29	212	0.15479
30	216	0.14243
31	217	0.15292
32	225	0.14296
33	227	0.14288
34	228	0.15888
35	232	0.15357
36	233	0.14677
37	241	0.15323

C. Naive Bayes

Naive Bayes is an excessively used algorithm for classification with considerable impact[15]. Bayesian classifiers (BC) are regulated measurable classifiers. They are used to predict the class enrollment probabilities which are the likelihood that an instance has probability belonging to a specific class. Bayesian classifiers based on Bayes' hypothesis; Bayesian classifiers have high precision and speed. The Naïve Bayes relies upon the as assumption of independence of variables within the class. Given a dependent attribute vector(x_1, x_2, …x_n) and a class variable y, the theorem of Bayes states the following relationship[16]:

$$p\ (y|\ x_1,..,\ \ ...,x_n)= \frac{p(y)p(x_1,\ ...x_n|y)}{p(x_1,\ ...xn)} \tag{1}$$

Appling the assumption of the naive independence that

$$p\ (x_i|y,x_1,...x_n) = p\ (x_i|y) \tag{2}$$

By simplifying this relationship for all i, to:

$$p\ (y|\ x_1,..,\ \ ...,x_n)= \frac{p(y)\prod_{i=1}^{n} p(x_i|y)}{p(x_1,\ ...x_n)} \tag{3}$$

The rule of Bayes classification can be set as:

$$P\ (y|x_1,\dots\dots,\ xn)\ \propto p(y)\prod_{i=1}^{n} p(x_i|y) \tag{4}$$

If whole feature follows a normal distribution, the probability of attribute is given by:

$$p(x_i|y) = \frac{1}{\sqrt{2\pi\sigma y}} exp\ (-(x_i-\mu_y)^{2}/2\sigma_y^2) \tag{5}$$

The estimation of standard deviation σ_y and the mean μ_y are calculated from training data.

D. Multilayer Perceptron

Multilayer Perceptron (MLP) is a type of feed forward neural network (NN) which is proved its effectiveness in the field of classification. MLP consists of three layers. The input layer, the hidden layer, and the output layer [17-19]. It uses back propagation technique for training. A wide range of research applications such as speech recognition, character recognition and image recognition are used the techniques of MLP [20, 21]. The architecture of the MLP is shown in Fig.3. In this work the input layer consists of 37nodes. The second layer which is the hidden layer consist of20 fully connect nodes with the input layer. This means the total number of hidden layer's weights is 740. The third layer contains 10 fully connected output nodes (digits 0-9) which have 200 weights. The sigmoid activation function has been used.

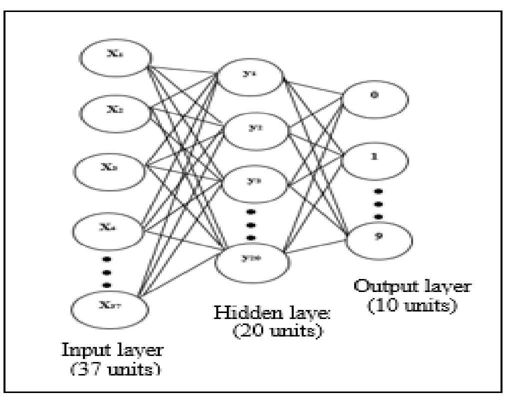

Fig. 3. Multilayer Perceptron Network.

E. K_Star

K_Star is instance-based classifiers. The classification is done after comparing the instance with an index of pre-classified samples. The entropy based distance measurement function is used in K-star algorithm to specify the class of a test instance by computing the distance between the test instance and with the similar training instances. Providing a reliable method for handling of symbolic features, actual valued features, and missing values are the key advantages of K-star. The probability of an instance x being in category Y by summing the likelihood from x to each instance that is a member of, as shown in equation (6) [8, 22]:

$$P^*(Y/x) = \sum_{b\in y} P*(x_1|x) \tag{6}$$

Where P* is the probability of all transformational paths from instance x to y.

F. Accuracy measurement

Three different classification algorithms (NB, MLP, K_Star) are used in this work. Confusion matrix is often used in data mining applications to evaluate the performance of the classifiers. The parameters and the formulas that are used to compare the performance of each classification algorithm are as follows[23, 24]:

- Accuracy: It is the number of correctly classified instances divided by the total number of instances
- Precision: It defines the proportion of those digits true to a specific class divided by overall digits classified as that class.
- Recall: It refers to the proportion of correctly classified digits divided by the total number of digits present in the class.
- F-measure: It is calculated by combining the measure of Precision and Recall.
- True Positive (TP) Rate: The rate of correctly classified digits.
- False Positive (FP) Rate: The rate of incorrectly classified digits.
- True Negative (TN) Rate: The rate of correctly rejected digits.
- False Negative (FN) Rate: The rate of incorrectly rejected digits

$$TP\ Rate\ =\ TP\ /\ (TP\ +\ FN) \tag{7}$$

$$FP\ Rate\ =\ FP\ /\ (FP\ +\ TN) \tag{8}$$

$$TN\ Rate\ =\ TN\ /\ (TN\ +\ FP) \tag{9}$$

$$FN\ Rate\ =\ FN\ /\ (FN\ +\ TP) \tag{10}$$

$$Precision\ =\ TP\ /\ (TP\ +\ FP) \tag{11}$$

$$Accuracy = (TP\ +\ TN)/\ (TP\ +\ TN\ +\ FP\ +\ FN) \tag{12}$$

$$Recall = TP\ /\ (TP+FN) \tag{13}$$

$$F\text{-}Measure=2*(Precision*Recall)/\ (Precision+Recall) \tag{14}$$

III. The Comparison Method

The block diagram of the comparison methods consisting of several steps that shown in Fig.4.

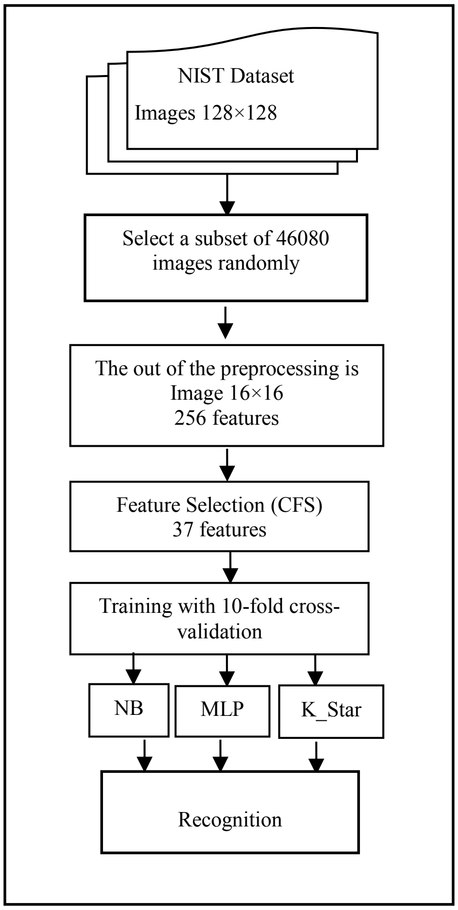

Fig. 4. The block diagram of the comparison method.

First of all a subset of the NIST dataset images are selected randomly. The number of each digit in the subset has been selected equally (4608) images for each one. So the total number of the chosen images are (46080) sized (128×128). In the second step, these images will transform into binary images (black and white).Then resizing them to (16×16) pixels. The choice of suitable classifier and features are the most important steps in handwritten digit recognition due to their impact on the results. Thus, the 256 features are ranked using correlation feature selection (CFS). In this paper, only the highest ranked 37 attributes are selected as a fewest number of attributes that can show the differences between the three classifiers to get an acceptable recognition rate. The 10-fold cross-validation method is used to train and test each classifier. The final step is the recognition task.

IV. Results and Discussion

This section shows the performance evaluation for the comparison method of the classification algorithms in handwritten digits recognition. This work is performed using MATLAB R2013a and WEKA on a computer with Intel® CoreTMi7-4700MQ processor @ 2.40 GHz speed with 64 bit. The 10-fold cross-validation technique is used for training and testing the three classifiers. The accuracy rate is used as assessment criteria for measuring the recognition performance of each classifier. The confusion matrixes of the three classifiers show that the digits (0, 6 and 5) are the most correctly classified instances respectively. Digit number (8) is the most incorrectly classified because of the similarity in digit shapes, specially, with the digits (3, 8), (5, 8). The rest of the digits (1, 2, 3, 4, 7, and 9) have different orders from the highest to the lowest rate of recognition in each classifier as follows:

The NB classifier shows that the orders of the correctly classified digits are (7, 9, 3, 4, 2, and 1).The total number of incorrectly classified digits are 15188 which considered as a highest one in all classifiers. Table II shows the confusion matrix of NB classifier.

The MLP classifier shows that the orders of the correctly classified digits are (4, 7, 3, 2, 1, and 9).The total number of incorrectly classified digits is 9976 which is less than NB. Table III, shows the confusion matrix of MLP classifier.

TheK_Star classifier shows that the orders of the correctly classified digits are (4, 1, 7, 3, 2, and 9).The total number of incorrectly classified digits is8129. This lowest number of incorrectly classified digits means that K_Star is better than the two other classifiers (NB and MLP). Table IV shows the confusion matrix of K-star classifier.

TABLE II. CONFUSION MATRIX OF NB CLASSIFIER WITH 10-FOLD CROSS-VALIDATION

0	1	2	3	4	5	6	7	8	9	class
3766	91	387	37	52	10	93	17	87	68	0
11	2626	79	231	62	264	224	503	501	107	1
228	129	2631	204	60	273	439	310	286	48	2
14	185	261	3044	32	115	48	584	134	191	3
169	309	81	30	2980	190	227	54	48	520	4
29	75	128	196	76	3731	180	62	74	57	5
34	227	248	24	121	175	3761	2	8	8	6
36	424	45	166	62	74	3	3342	66	390	7
258	679	350	135	83	326	132	346	1897	402	8
144	131	9	130	168	151	21	519	221	3114	9

TABLE III. CONFUSION MATRIX OF MLP CLASSIFIER WITH 10-FOLD CROSS-VALIDATION

0	1	2	3	4	5	6	7	8	9	class
3993	63	148	45	68	18	69	37	103	64	0
7	3477	104	175	125	123	147	109	266	75	1
120	61	3520	199	81	54	141	164	215	53	2
21	88	246	3605	49	127	22	184	133	133	3
79	83	145	32	3772	75	77	35	80	230	4
18	42	109	129	59	3948	96	26	111	70	5
39	166	118	38	91	84	3989	6	63	14	6
16	132	109	106	80	33	1	3728	156	247	7
118	297	291	161	129	210	88	270	2688	356	8
89	68	29	108	271	91	8	264	296	3384	9

TABLE IV. CONFUSION MATRIX OF K-STAR CLASSIFIER WITH 10-FOLD CROSS-VALIDATION

0	1	2	3	4	5	6	7	8	9	class
4301	36	60	5	28	5	74	22	20	57	0
12	3926	83	108	21	55	134	106	118	45	1
149	87	3574	136	80	36	167	220	128	31	2
11	118	153	3686	21	153	23	287	73	83	3
67	139	94	9	3946	42	61	41	38	171	4
11	80	64	155	31	4105	65	13	60	24	5
27	199	75	14	22	37	4216	1	12	5	6
10	171	87	48	64	18	1	3886	141	182	7
87	355	231	129	71	151	83	320	2842	339	8
94	83	35	79	206	80	20	318	224	3469	9

The average weights values of TP rate, FP rate, Precision, Recall and F-measure for all digits of the three algorithms are shown in table V.

TABLE V. CLASSIFIERS' ACCURACY AVERAGE WEIGHT

Classifier	TP-Rate	FP-Rate	Precision	Recall	F-Measure
NB	0.67	0.037	0.672	0.67	0.667
MLP	0.784	0.024	0.782	0.784	0.783
K-Star	0.824	0.02	0.824	0.824	0.822

The differences and similarities of correctly classified digits (0-9) in all classifiers are illustrated and summarized in Fig.5. It is obvious that the digit (0) has the highest number of correctly classified digits which reached to (4301) digits using K_Star classifier. While the digit (8) has the lowest number of correctly classified digits which is (1897) using NB classifier.

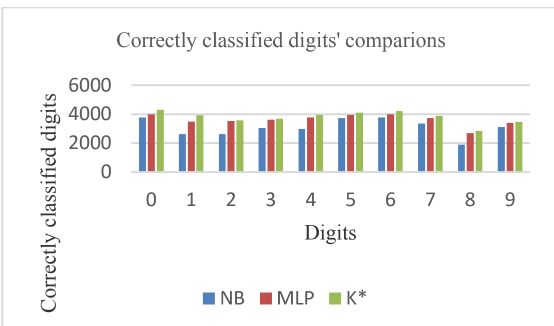

Fig. 5. A comparison of correctly classified digits for the three classifiers.

The comparison between the three classifiers for predicting the correctly classified and incorrectly classified digits is shown in Fig.6. The total number of correctly classified instances for all digits in K_star classifier is 37951 from the total subset (46080) digits. This means the accuracy rate is 82.36% which is the highest rate between the two others classifiers. The accuracy rate of MLP classifiers 78.35% and the total number of correctly classified instances is 36104 digits. The NB classifier gives the lowest accuracy rate as67.04% and 30892 correctly classified digits.

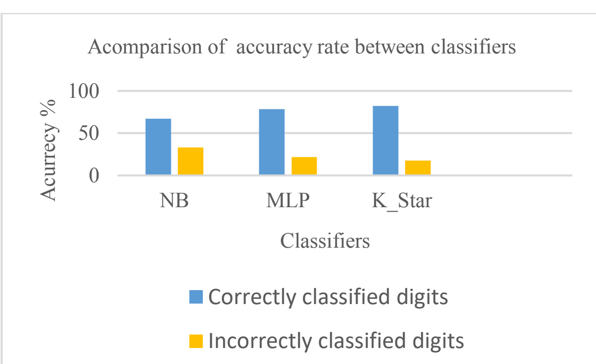

Fig. 6. A comparison between the classifiers.

The task of recognizing the handwriting of an individual from another is difficult as each person possess a unique handwriting style. This is one reason of why handwriting is considered as one of the main challenging studies. The NIST dataset images include some confusing images as they written manually. Since the selection of subset is done randomly this has a significant effect on the results due to increasing the incorrectly classified digits. Also the preprocessing (resizing) has an impact on the classification accuracy. These points can be considered as a challenge to find out the best classifier in these circumstances.

V. CONCLUSION

The comparison between three classification algorithms, Naive Bayes, MLP and K_Star has been done for the recognition of the NIST handwritten digits. Using (CFS), only 37 features are selected from 256 features to get an acceptable recognition rate. After performing an evaluation for each classifier on 46080 instances with 10-fold cross-validation for each one,K_Star presents the highest accuracy of 82.36% compared to NB 67.04%, and MLP 78.35%. In this comparison method K-Star algorithm works as the best classifier in handwritten digits recognition than the two others (NB and MLP) classifiers.

REFERENCES

[1] M. M. A. Ghosh and A. Y. Maghari, "A Comparative Study on Handwriting Digit Recognition Using Neural Networks," in *Promising Electronic Technologies (ICPET), 2017 International Conference on*, 2017, pp. 77-81.

[2] M. Ahmed, A. K. Paul, and M. Akhand, "Stacked auto encoder training incorporating printed text data for handwritten bangla numeral recognition," in *Computer and Information Technology (ICCIT), 2016 19th International Conference on*, 2016, pp. 437-442.

[3] M. H. Shammakhi, A. Mirzaei, P. Khavari, and V. Pourahmadi, "Combined mRMR-MLPSVM scheme for high accuracy and low cost handwritten digits recognition," in *Machine Vision and Image Processing (MVIP), 2015 9th Iranian Conference on*, 2015, pp. 155-158.

[4] A. Chatterjee, S. Malakar, R. Sarkar, and M. Nasipuri, "Handwritten Digit Recognition using DAISY Descriptor: A Study," in *2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT)*, 2018, pp. 1-4.

[5] K. T. Islam, G. Mujtaba, R. G. Raj, and H. F. Nweke, "Handwritten digits recognition with artificial neural network," in *Engineering Technology and Technopreneurship (ICE2T), 2017 International Conference on*, 2017, pp. 1-4.

[6] J. Wannenburg and R. Malekian, "Physical activity recognition from smartphone accelerometer data for user context awareness sensing," *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* vol. 47, pp. 3142-3149, 2017.

[7] A. Kaushik, H. Gupta, and D. S. Latwal, "Impact of feature selection and engineering in the classification of handwritten text," in *Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on*, 2016, pp. 2598-2601.

[8] R. Pramanik, V. Raj, and S. Bag, "Finding the optimum classifier: Classification of segmentable components in offline handwritten Devanagari words," in *2018 4th International Conference on Recent Advances in Information Technology (RAIT)*, 2018, pp. 1-5.

[9] I. B. Othman and F. Ghorbel, "A new criteria for comparing neural networks and Bayesian classifier," in *Computer Applications Technology (ICCAT), 2013 International Conference on*, 2013, pp. 1-5.

[10] Y. Hou and H. Zhao, "Handwritten digit recognition based on depth neural network," in *Intelligent Informatics and Biomedical Sciences (ICIIBMS), 2017 International Conference on*, 2017, pp. 35-38.

[11] X. Guo, A. A. Minai, and L. J. Lu, "Feature selection using multiple auto-encoders," in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017, pp. 4602-4609.

[12] Z. Li and J. Tang, "Unsupervised feature selection via nonnegative spectral analysis and redundancy control," *IEEE Transactions on Image Processing,* vol. 24, pp. 5343-5355, 2015.

[13] A. A. Shanab and T. Khoshgoftaar, "Filter-based subset selection for easy, moderate, and hard bioinformatics data," in *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, 2018, pp. 372-377.

[14] C. Yao, Y.-F. Liu, B. Jiang, J. Han, and J. Han, "LLE score: a new filter-based unsupervised feature selection method based on nonlinear manifold embedding and its application to image recognition," *IEEE Transactions on Image Processing,* vol. 26, pp. 5257-5269, 2017.

[15] Y. Wu, "A New Instance-weighting Naive Bayes Text Classifiers," in *2018 IEEE International Conference of Intelligent Robotic and Control Engineering (IRCE)*, 2018, pp. 198-202.

[16] S. Shakya and S. Sigdel, "An approach to develop a hybrid algorithm based on support vector machine and Naive Bayes for anomaly detection," in *Computing, Communication and Automation (ICCCA), 2017 International Conference on*, 2017, pp. 323-327.

[17] E. Genc, H. R. Shin, and J. S. Park, "Number Recognition of Parts Book Schematics using Convolutional Recurrent Neural Network," in *2018 18th International Conference on Control, Automation and Systems (ICCAS)*, 2018, pp. 1190-1192.

[18] M. Hagenbuchner, A. C. Tsoi, F. Scarselli, and S. J. Zhang, "A fully recursive perceptron network architecture," in *Computational Intelligence (SSCI), 2017 IEEE Symposium Series on*, 2017, pp. 1-8.

[19] J. A. Ahmed and A. M. A. Brifcani, "A new internal architecture based on feature selection for holonic manufacturing system," *International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering,* vol. 2, p. 1431, 2015.

[20] M. R. Mahmood, A. M. Abdulazeez, and Z. Orman, "Dynamic Hand Gesture Recognition System for Kurdish Sign Language Using Two Lines of Features," in *2018 International Conference on Advanced Science and Engineering (ICOASE)*, 2018, pp. 42-47.

[21] J. Si and S. L. Harris, "Handwritten digit recognition system on an FPGA," in *Computing and Communication Workshop and Conference (CCWC), 2018 IEEE 8th Annual*, 2018, pp. 402-407.

[22] R. Ade and P. Deshmukh, "Classification of students by using an incremental ensemble of classifiers," in *Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions), 2014 3rd International Conference on*, 2014, pp. 1-5.

[23] S. Sharma and S. K. Srivastava, "Feature Based Performance Evaluation of Support Vector Machine on Binary Classification," in *Computational Intelligence & Communication Technology (CICT), 2016 Second International Conference on*, 2016, pp. 170-173.

[24] D. Q. Zeebaree, H. Haron, and A. M. Abdulazeez, "Gene Selection and Classification of Microarray Data Using Convolutional Neural Network," in *2018 International Conference on Advanced Science and Engineering (ICOASE)*, 2018, pp. 145-150.