ILKOM Jurnal Ilmiah Vol. 15, No. 1, April 2023, pp. 186-192
Accredited 2nd by RISTEKBRIN No. 200/M/KPT/2020; E-ISSN 2548-7779 | P-ISSN 2087-1716

Research Article

Analysis of the Ensemble Method Classifier's Performance on Handwritten Arabic Characters Dataset

Abdul Rachman Manga’ [a,b,1]; Anik Nur Handayani [a,2]; Heru Wahyu Herwanto[a,3]; Rosa Andrie Asmara[c,4]; Yudha Islami Sulistya [b,5,*]; Kasmira [b,6]

[a] *Universitas Negeri Malang, Malang, Indonesia*
[b] *Universitas Muslim Indonesia, Makassar and 90231, Indonesia*
[c] *Information Technology Department, State Polytechnic of Malang, Malang, Indonesia*
[1] *abdulracman.manga@umi.ac.id;* [2] *aniknur.ft@um.ac.id;* [3] *heru_wh@um.ac.id;* [4] *rosa.andrie@polinema.ac.ad;* [5] *yudhaislamisulistya@umi.ac.id;*
[6] *kasmira.labfik@umi.ac.id;*
* *Corresponding author*

Article history: Received August 10, 2021; Revised October 11, 2021; Accepted December 14, 2022; Available online April 20, 2023

Abstract

Arabic character handwriting is one of the patterns and characteristics of each person's writing. This characteristic makes Arabic writing more challenging if the letter recognition process is based on a dataset of Arabic scripts. This Arabic script has been presented in a dataset totaling 16800, each representing a class of hijaiyah letters starting from alif to yes, consisting of 600 data for each class. The accuracy of the data used can be increased using the ensemble method. By using multiple algorithms at simultaneously, the ensemble technique can raise the level or result of a score in machine learning. This study's primary goal is to evaluate the ensemble method classifier's performance on datasets of handwritten Arabic characters. The classifier uses the ensemble method by applying the proposed soft voting to provide a multiclass classification of three machine learning algorithms, namely, SVM, Random Forest, and Decision Tree for classification. This research process produces an accuracy value for the voting classifier of 0.988 and several other SVM algorithms with an accuracy of 0.103, a random forest with an accuracy of 1.0, and a decision tree with an accuracy of 0.134. The test results used the confusion matrix evaluation model, including accuracy, precision, recall, and f1-score of 0.99.

Keywords: Ensemble Method; Voting Classifiers; Arabic Character Handwriting; Evaluation Model.

Introduction

Arabic consists of the Arabic/Hijaya composition found in the Al-Qur'an/Koran. The Koran is the holy book for Muslims and needs to be studied, and all Muslims should be able to read the Koran and recognize the letters of each Hijaiyah that make it up. In pattern recognition, numerical data and symbolic data (images, etc.) are grouped automatically by machines (computers) or better known as machine learning. Pattern recognition aims to recognize objects in digital images [1].

Machine learning (ML) is a field of science that can learn patterns based on data [2]. One part or method in machine learning that is widely used is ensemble. The ensemble method combines several tree classifiers to produce better predictive performance than a single tree classifier, thereby increasing the model's accuracy. One of them is that the random forest is a refinement of the C4.5 algorithm, so the selection of the method used proves the superiority of the random forest, which has higher accuracy than conventional classification methods of C4.5[3]. The ensemble method combines several classification algorithms such as SVM, Random Forest Classifier, and Decision Tree Classifier. This method is often known as the Voting Classifier [4].

Hence, based on the background above, the researchers propose an approach with the ensemble method in classifying Arabic character handwriting based on datasets collected independently. On the other hand, this research also aims to analyze performance using the confusion matrix model evaluation method. Moreover, this study also mainly focuses on performance, theory, and properties of learning models and algorithmic methods. A comprehensive approach has been proposed using a handwritten Arabic character classifier [5]. The Support Vector Machine, Random Forest Classifier, and Decision Tree Classifier algorithms are combined, and the performance of the synthetic method is better than the basic classifier. The algorithm's performance mentioned above is evaluated by taking accuracy, precision, recall, and f1 score as evaluation criteria.

http://dx.doi.org/10.33096/ilkom.v15i1.1357.186-192

Method

This study worked intensively to increase the accuracy and outcomes of Arabic character handwriting classification. In order to categorize handwritten Arabic characters that are divided into 28 classes or 28 hijaiyah letters starting from alif to yes and each data in a class of 600, which is frequently referred to as a balanced dataset, the author has proposed a set of machine learning algorithms using the Voting Classifier. Before supplying input to the model, data preparation and data augmentation are performed. **Figure 1** shows the flowchart for the ensemble technique approach suggested utilizing the Soft Voting Classifier.

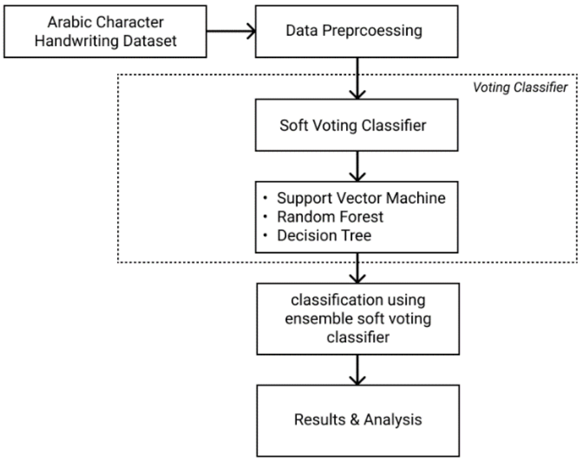
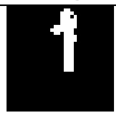
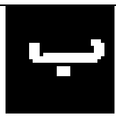
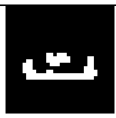
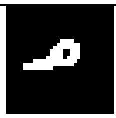
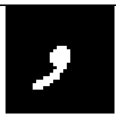
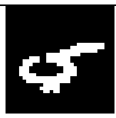


Figure 1. Flowchart of the proposed ensemble method using the Soft Voting Classifier

A. *Data Description*

The handwritten Arabic character dataset was carefully examined for testing in this work. The dataset has 7 features that have previously been processed by moment invariant and 1 label. There are 28 classes and 16800 data points in this dataset. Examples of the photos in the dataset are shown in **Table 1**.

Table 1. Some examples of images in the dataset (6 of 16800)

No	Image	Class
1		*Alif*
2		*Ba*
3		*Ta*
	...	
16798		*Hah*
16799		*Waw*
16800		*Ya*

B. Data pre-processing

This process is the most crucial part for converting data into a more manageable format so that it is very easy to process with machine learning. The first technique used for data pre-processing is data normalization. This technique is used to perform linear transformation of data. It is also called MinMaxScaler normalization[3]. Therefore, it is possible to create a balance of comparative values between data. Calculating MinMaxScaler can use formula (1) as follows [6]:

$$X_{sc} = \frac{X - Xmin}{Xmax - Xmin} \quad (1)$$

Information:

X_{sc} = Normalized value

X_i = Value value in the dataset

X_{min} = Minimum value in a column

X_{max} = Maximum value in a column

C. Model Architecture

In this proposed method, research employs a set of machine learning algorithms including Support Vector Machine, Random Forest, and Decision Tree. The algorithm mentioned above has been combined with a soft voting classifier to improve accuracy. This algorithm is briefly described below.

1. *Support Vector Machine*

The Support Vector Machine (SVM) method was proposed by Vapnik around 1992 to classify two classes [7]. In its development, this method can then be used to classify several classes. The main process carried out by SVM is to separate two groups of data using a vector space called a hyperplane. There are two types of SVM, namely linear and nonlinear, both of which are differentiated based on the type of multiplication function, including the radial basis function (RBF) and polynomial[8][9].

2. *Random Forest*

Random forest is one of the map classification methods and is widely used in various studies and case models. Moreover, it is a decision process based on tree branching or decision trees [3], [10]. Each branch contains questions to solve the ideal number of branches.

3. *Decision Tree*

Tree is a data structure consisting of nodes and edges. The nodes of a tree are divided into three, namely root nodes (root/node), branch/internal nodes (branch/internal nodes) and leaf nodes. A decision tree is a simplified representation of a classification technique for a finite number of classes, in which inner nodes and roots are labeled with attribute names, and slopes are labeled with values. Possible attributes and leaf nodes are marked with different classes [11].

4. *Ensemble Method* with *Voting Classifier*

This classifier serves as a super classifier for combining conceptually similar or different machine learning models for majority vote prediction. The voting rating class uses two voting techniques, hard and soft. In hard voting, the final prediction is made by majority vote, where the aggregator selects the class predictions that repeatedly appear among the underlying models. In the case of secret voting, the basic model must have a probe prediction method. Voting classifiers show better overall results than basic models because they combine predictions from different models. In the proposed model, SVM, Random Forest, and Decision Tree, a soft voting classifier is adopted, using the attribute predict probability column to provide the probability of each target variable [12]. It then randomizes the training data and data points fed into the SVM, Random Forest, and Decision Tree models. Each model calculates individual predictions using aggregator voting and classifier soft voting techniques, and the majority of votes are counted, leading to the final prediction.

Results and Discussion

The proposed methodology uses an ensemble of three machine learning models including Soft Voting Classifier, Random Forest, Decision Tree, and SVM. The experiment was carried out using a handwritten Arabic character dataset. The dataset comprises 7 feature columns and 16800 data points, with the median in place of the zeros in each column. The dataset is divided into 10-90% test and training datasets, respectively. Accuracy, precision, recall and f1-score are the most commonly used metrics to check robustness and efficiency.

A. Handwritten Arabic Character Dataset

In this study, Arabic characters written by hand were used, and each feature had a positive real number data type. The classes in this dataset consist of 28 representing hijaiyah letters in Arabic, each of which totals 600 data points for each class. In the **Figure 2** you can see Balanced character handwriting dataset. An illustration of a handwritten Arabic character dataset can be found in **Table 2**.

Table 2. Arabic handwriting dataset in moment invariant form (6 out of 16800)

No	h1	…	h7	Label
1	0.001804		5.710000e-06	1
2	0.001249		-3.770000e-06	1
3	0.001873		-3.130000e-07	1
		…		
4	0.001520		1.150000e-05	28
5	0.001074		1.690000e-07	28
6	0.001095		1.200000e-07	28

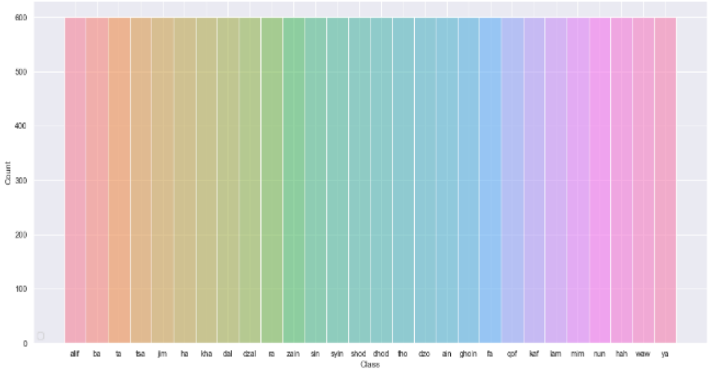


Figure 2. Balanced character handwriting dataset.

B. Data Pre-processing

At this stage it is used for simplification or normalization of features. The normalization in question is to provide a minimum value and a maximum value with the MinMaxScaler technique. The MinMaxScaler results can be seen in **Table 3**.

Table 3. Normalization with MinMaxScaler Arabic character handwriting dataset (6 of 16800 data)

No	h1	…	h7	Label
1	0.385872		0.640312	1
2	0.267154		0.640312	1
3	0.400489		0.640312	1
		…		
4	0.325103		0.640312	28
5	0.229613		0.640312	28
6	0.234186		0.640312	28

1. Testing with Voting Classifier

At this stage, the testing process is carried out using the ensemble method with the voting classifier algorithm. In the **Figure 3** you can see Results of several voting algorithms and classifiers. This algorithm is presented in stages in **Table 4**.

Table 4. Classifier voting algorithm with the ensemble method

Algortihm

```
model_dtc = DecisionTreeClassifier(max_depth=4)
model_rfc = RandomForestClassifier(random_state=1)
model_svc = SVC(kernel='rbf', probability=True)

final_model = VotingClassifier(estimators=[('svc',
model_svc), ('rf', model_rfc), ('dt', model_dtc)],
voting='soft')

model_svc.fit(X, y).score(X, y)
model_rfc.fit(X, y).score(X, y)
```

Algortihm
*model*_dtc.fit(X, y).score(X, y)
final_*model*.fit(X, y).score(X, y)

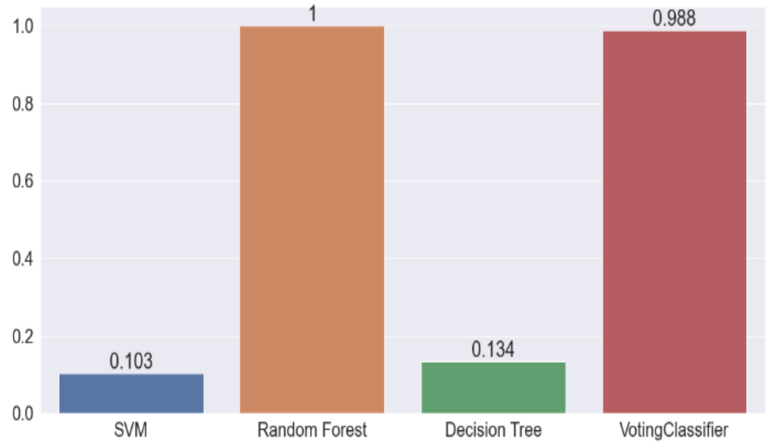


Figure 3. Results of several voting algorithms and classifiers

2. Testing with Confusion Matrix

In testing, this research uses the confusion matrix with the conditions True positive (tp) means that the value before the specified class is 1 and the actual class value is also 1. True negative (tn) means that the predicted class value is 0 and the actual class is also 0. Negative false (fn) and false positive (fps) occur when the predicted value is not consistent with the true value. Accuracy is an important safety measure, the ratio of the total number of correctly predicted observations to the total number of observations. Fit rate, accuracy, recall, and F1 score can be calculated using the following formulas[13]–[15]:

$$Accuracy = \frac{tp + tn}{all\ data} \quad (2)$$

$$Precision = \frac{tp}{tp + fp} \quad (3)$$

$$Recall = \frac{tp}{tp + fn} \quad (4)$$

$$F1Score = 2\ \frac{Precision * Recall}{Precision + Recall} \quad (5)$$

```
Prediction 25.0 : Actual 25.0
Prediction 8.0 : Actual 8.0
Prediction 19.0 : Actual 19.0
Prediction 28.0 : Actual 28.0
Prediction 4.0 : Actual 4.0
```

Figure 4. Example 5 of 1680 predictions and actuals

In **Figure 4**, you can see 5 of the 1680 predictions and actual values that machine learning predicts nun and is correct that it is nun, predicts dal and is correct that it is dal, predicts ghoin and is correct that it is ghoin, predicts yes and correctly that it is yes and the last prediction is tsa and it is true that it is tsa

The confusion matrix can be seen in **Figure 5**. The amount of testing data in this study is 1680 or 10 percent of the total data, where the more colorful the parts of the cell, the higher the prediction value. **Table 5** shows the classification matrix.

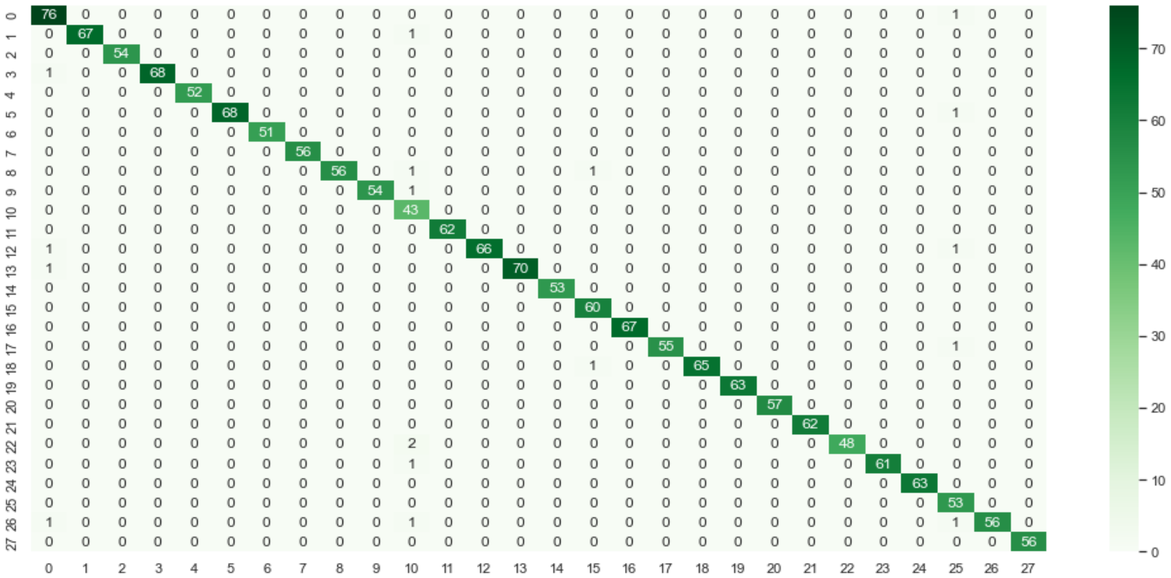


Figure 5. The results of testing the confusion matrix data

Table 5. Classification report

Class	*Precision*	*Recall*	*F1-score*	*Data point*
1	0.95	0.99	0.97	77
2	1.00	0.99	0.99	68
3	1.00	1.00	1.00	54
…	…	…	…	…
26	0.91	1.00	0.95	53
27	1.00	0.95	0.97	59
28	1.00	1.00	1.00	56
Accuracy			0.99	1680
macro avg	0.99	0.99	0.99	1680
weighted avg	0.99	0.99	0.99	1680

Conclusion

This study used a variety of machine learning techniques to examine the performance of the classifier ensemble method performed on datasets of handwritten Arabic characters. The dataset used is handwritten Arabic characters, which can be experimented with hijaiyah letter label categories from alif to yes. This research process produces an accuracy value for the voting classifier of 0.988 and several other SVM algorithms with an accuracy of 0.103, a random forest with an accuracy of 1.0, and a decision tree with an accuracy of 0.134. The test results used the confusion matrix evaluation model, including accuracy, precision, recall, and an f1-score of 0.99.

References

[1] N. Kasim and G. S. Nugraha, "Pengenalan pola tulisan tangan aksara Arab menggunakan metode Convolution Neural Network," *Jurnal Teknologi Informasi, Komputer, dan Aplikasinya (JTIKA)*, vol. 3, no. 1, pp. 85–95, 2021, doi: 10.29303/jtika.v3i1.136.

[2] P. Tesa Ananda, "Pengenalan Karakter Aksara Incung (Kerinci) ke Karakter Latin menggunakan metode Convolutional Neural Network," Universitas Jambi, Jambi, 2021. [Online]. Available: https://repository.unja.ac.id/28261/

[3] A. Eko Cahyo and A. Nilogiri, "Klasifikasi gangguan Autisme pada anak menggunakan Algoritma C4.5 Denganteknik Random Forest," Jember, 2021.

[4] S. Kumari, D. Kumar, and M. Mittal, "An Ensemble Approach for Classification and Prediction of Diabetes Mellitus using Soft Voting Classifier," *International Journal of Cognitive Computing in Engineering*, vol. 2, no. 2, pp. 40–46, 2021, doi: 10.1016/j.ijcce.2021.01.001.

[5] E. Nuranti Kusuma, "Identifikasi citra huruf Arab menggunakan metode Jaringan Syaraf Tiruan Kohonen," Universitas Islam Negeri Maulana Malik Ibrahim, Malang, 2015.

[6] D. A. Nasution, H. H. Khotimah, and N. Chamidah, "Perbandingan normalisasi data untuk klasifikasi wine menggunakan algoritma K-NN," *CESS (Journal of Computer Engineering System and Science)*, vol. 4, no. 1, p. 78, 2019, doi: 10.24114/cess.v4i1.11458.

[7] O. V. Putra, T. Harmini, and A. Saroji, "Outlier detection on graduation data of Darussalam Gontor University using One-Class Support Vector Machine," in *ENASAINS 2nd*, 2021, vol. 1, no. 2, pp. 1–4.

[8] C. Cortes and V. Vapnik, "Support-Vector Networks," *IEEE Expert-Intelligent Systems and their Applications*, vol. 7, no. 5, pp. 63–72, 1992, doi: 10.1109/64.163674.

[9] I. Wijayanto, A. Humairani, A. Rizal, and S. Hadiyoso, "Klasifikasi sinyal EKG menggunakan Ciri Statistik dan Parameter Hjorth dengan SVM dan k-NN," *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika* , vol. 10, no. 1, pp. 132–145, 2022.

[10] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, "How Many Trees in A Random Forest?," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7376 LNAI, pp. 154–168, 2012, doi: 10.1007/978-3-642-31537-4_13.

[11] A. H. Nasrullah, "Implementasi algoritma Decision Tree untuk klasifikasi produk laris," *Jurnal Ilmiah Ilmu Komputer*, vol. 7, no. 2, pp. 45–51, 2021, doi: 10.33480/pilar.v14i2.926.

[12] I. Indriati and A. Kusyanti, "Metode Ensemble Classifier untuk mendeteksi jenis Attention Deficit Hyperactivity Disorder (ADHD) Pada Anak Usia Dini," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 6, no. 3, pp. 301–308, 2019, doi: 10.25126/jtiik.201961313.

[13] M. Jumarlis, M. Mirfan, and A. R. Manga, "Classification of Coffee Bean Defects using Gray-Level Co-Occurrence Matrix and K-Nearest Neighbor," *ILKOM Jurnal Ilmiah*, vol. 14, no. 1, pp. 1–9, Apr. 2022, doi: 10.33096/ilkom.v14i1.910.1-9.

[14] S. Budiman, A. Sunyoto, and A. Nasiri, "Analisa Performa penggunaan feature selection untuk mendeteksi Intrusion Detection Systems dengan Algoritma Random Forest Classifier," *SISTEMASI:Jurnal Sistem Informasi*, vol. 10, no. 3, pp. 753–760, 2021, doi: 10.32520/stmsi.v10i3.1550.

[15] J. Eska, "Penerapan data mining untuk prediksi penjualan wallpaper menggunakan algoritma C4.5," *JURTEKSI(Jurnal Teknologi dan Sistem Informasi)*, vol. 2, no. 2, pp. 9–13, 2018, doi: 10.31227/osf.io/x6svc.