Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 7 No. 1 (2023) 94 - 104 ISSN Media Electronic: 2580-0760

Improved Classification of Handwritten Jawi Script Based on Main Part of Script Body

Safrizal Razali[1], Fitri Arnia[2], Rusdha Muharrar[3], Kahlil Muchtar[4], Akhyar Bintang[5]
[1]Department of Electrical and Computer Engineering, Faculty of Engineering, Syiah Kuala University
[1]safrizal.razali@unsyiah.ac.id

Abstract

Since the entry of Islam, many ancient relics in the archipelago were written using Jawi script. Due to human or natural factors, these ancient relics will be damaged or destroyed. To avoid the loss of this ancient heritage data, the data must be stored in digital documents. In order to convert digital documents into machine-readable text format, the use of Optical Character Recognition (OCR) technology is inevitable. In this research, OCR technology is implemented on isolated Jawi scripts. Freeman Chain Code (FCC) is used to extract the isolated Jawi script features. Subsequently, the FCC feature is fed into Support Vector Machine (SVM) in order to classify the character. The decision rule classification is applied to the class of SVM classification in the Jawi script form. The results of the SVM classification into 19 classes reached 81.58%, while the results for merging into 15 classes produced better results with the accuracy 84.21%. Feature extraction of dot location is divided into the top, middle, and bottom. Feature extraction of the number of dotss is done by counting the number of dots, while feature extraction of the presence of holes is carried out by detecting the presence of holes in the characters. These features are applied to the class of results from SVM classification with decision-making rules. The percentage of success in applying the decision rules to the results of the classification of incorporation into 15 classes by SVM reached 92.86%. Further research will be conducted to determine the effect of the feature of the location of the dot and the number of dots on the shape of the main part of the character.

Keywords: jawi script; FCC; SVM; the shape of the main part of the character; decision-making rules.

1. Introduction

Arabic script is a script used to write Arabic. There are several variants of the Arabic script used to write particular languages. For example, Urdu script is used to write Urdu and is the lingua franca of Pakistan. The Farsi script is used to write Persian, while the Jawi script is also a variant of the Arabic script and is used to write Malay [1]. The Malay language is used by people in the Nusantara. In Indonesia, especially the Aceh region, many ancient relics are written in Jawi script. Some of these ancient relics are in the form of tombstones, currency, books, and Islamic books.

The ancient relics in the Jawi script are even 600 years old [2]. These ancient relics need to be preserved. One way to preserve these ancient relics is to store them in the form of digital images. The Jawi script in the form of digital images can be converted into characters that can be read automatically by a computer using Optical Character Recognition (OCR) technology.

Research on OCR for the Jawi script is not much done, while research on other characters is quite a lot. Other Arabic variant scripts are generally still used as official state scripts, such as Urdu writing and Farsi writing, so they are more interesting for researchers. On the other hand, the Jawi script is no longer used in official activities of countries that use Malay as the lingua franca. Currently, the Jawi script is generally only used in Islamic teaching and learning activities in madrasas or Islamic schools.

Jawi script consists of 35 characters. These scripts consist of 29 pure Arabic scripts, and there are six additional characters to write Malay, which cannot be accommodated by Arabic script. Similar to Arabic script, Jawi script is also written in cursive from right to left. Several scripts have different shapes because the Jawi script can be connected at the beginning, middle, and end, as shown in Table 1 and Table 2.

Several studies on Arabic script variants were carried out by [3] to identify Urdu script, this study also discussed the shape of the main parts of the script. For classification, this research uses Convolutional Neural Network (CNN). Another research on Urdu script was

Accepted: 19-10-2022 | Received in revised: 16-01-2023 | Published: 03-02-2023

carried out by [4] who also used CNN with the transfer learning method. Another study on Urdu script was conducted by [5] using the Multi-Dimensional Long Short-Term Memory Recurrent Neural Network (MDLSTM RNN). In addition to research on Urdu script, research on Farsi script which is also a variant of Arabic script was carried out by [6]-[8].

Table 1. Arabic Script for Writing Jawi

No	Script	Jawi	Form			
			Isolated	Initial	Middle	End
1	Alif	ا	ا			ا
2	Ba	ب	ب	بـ	ـبـ	ـب
3	Ta	ت	ت	تـ	ـتـ	ـت
4	Tsa	ث	ث	ثـ	ـثـ	ـث
5	Jim	ج	ج	جـ	ـجـ	ـج
6	Hha	ح	ح	حـ	ـحـ	ـح
7	Kha	خ	خ	خـ	ـخـ	ـخ
8	Dal	د	د			ـد
9	dzal	ذ	ذ			ـذ
10	ra	ر	ر			ـر
11	zai	ز	ز			ـز
12	sin	س	س	سـ	ـسـ	ـس
13	syin	ش	ش	شـ	ـشـ	ـش
14	shad	ص	ص	صـ	ـصـ	ـص
15	dhad	ض	ض	ضـ	ـضـ	ـض
16	tho	ط	ط	طـ	ـطـ	ـط
17	zho	ظ	ظ	ظـ	ـظـ	ـظ
18	ain	ع	ع	عـ	ـعـ	ـع
19	ghain	غ	غ	غـ	ـغـ	ـغ
20	fa	ف	ف	فـ	ـفـ	ـف
21	qaf	ق	ق	قـ	ـقـ	ـق
22	kaf	ک	ک	کـ	ـکـ	ـک
23	lam	ل	ل	لـ	ـلـ	ـل
24	mim	م	م	مـ	ـمـ	ـم
25	nun	ن	ن	نـ	ـنـ	ـن
26	wau	و	و			ـو
27	ha	ه	ه	هـ	ـهـ	ـه
28	hamzah	ء	ء	ء	ء	ء
29	ya	ي	ي	يـ	ـيـ	ـي

Table 2. Additional Jawi Script

No	Huruf	Jawi	Form			
			Isolated	Initial	Middle	End
1	nya	ڽ	ڽ	ڽـ	ـڽـ	ـڽ
2	ca	چ	چ	چـ	ـچـ	ـچ
3	nga	ڠ	ڠ	ڠـ	ـڠـ	ـڠ
4	pa	ڤ	ڤ	ڤـ	ـڤـ	ـڤ
5	ga	گ	گ	گـ	ـگـ	ـگ
6	va	ۏ	ۏ			ـۏ

One of the studies on Jawi script OCR was carried out by [9] to classify Jawi script using the Freeman Chain Code (FCC) as a feature extraction method and SVM and decision rules as a classification method. The results of this study were able to classify Jawi script with an accuracy of 80%. Another research on Jawi script is [10] by using the angular features of Jawi script and SVM as a classifier. In an effort to improve the accuracy of Jawi script recognition, it is necessary to conduct studies that support the improvement of Jawi script recognition performance so that it is hoped that the legacy of Jawi writing in the archipelago can be maintained both as history and as a culture for future generations.

In this study, the characters used were the same as those used in the study [9] in a separate writing. The script was written by 10 authors with different educational backgrounds. In this study, not all of these characters were used because some of them have the same shape in the main part of the script. Based on research [9] we found several characters classified by SVM into unsuitable character classes. In this study we perform to increase in the accuracy of Jawi script recognition by grouping Jawi characters with the same main part or having similarities into a certain class. For example, the characters "ت", "ب", and "ٹ" have the main part form "ـب" but some of them are classified by SVM into classes with the main part form "ف". By grouping based on the similarity of "ب" and "ف" into the same class, the number of classes classified by SVM is reduced. In addition, several character classes were combined into a certain class by taking into account the tendency of misclassification by SVM. Classification into the appropriate Jawi script, not into character class, is done through decision-making rules. Grouping into certain classes based on similarities in the shape of the main part of the script in the classification process using SVM was not carried out in research conducted by [9]. In addition, combining several classes based on the tendency of misclassification by SVM into a group and then classified by decision-making rules to in the appropriate class was also not carried out in the research conducted by [9].

2. Research Methods

The methods used in this study are pre-processing, feature extraction, SVM classification, and classification using decision-making rules.

2.1 Research Flow

As seen in Figure 1, the Jawi script used in this study consists of 350 total characters. From these characters, 80 percent (8 characters from each script) are used for training characters while the remaining 20 percent are used for testing characters. Initially, the training characters were grouped into 19 classes based on the differences in the shape of the main parts as shown in table 4. From each group, 80 percent (8 characters with the same main part) were randomly selected, while the testing characters were 2×35 = 70 test characters. The grouping of training characters was then changed due to the reduction of groups, which resulted in a reduction in the number of classes in the SVM classification.

Reducing the number of classes in the SVM classification is done by combining characters that have the same main body form or have a similar main part form into a class. For example, combining the Jawi character class with the main form of the "ک" script and

DOI: https://doi.org/10.29207/resti.v7i1.4600

another class with the main part of the "ى" character. The merging of the two classes caused the number of classes to decrease to 18. In addition to reducing the number of classes based on the similarity in the shape of the main parts of the script, the reduction in the number of classes was also carried out by taking into account the similarity in the shape of the main parts of the characters, so that the number of classes continued to decrease to 15 classes. Reducing the number of classes based on the similarity of the main parts was not carried out in the research conducted by [9].

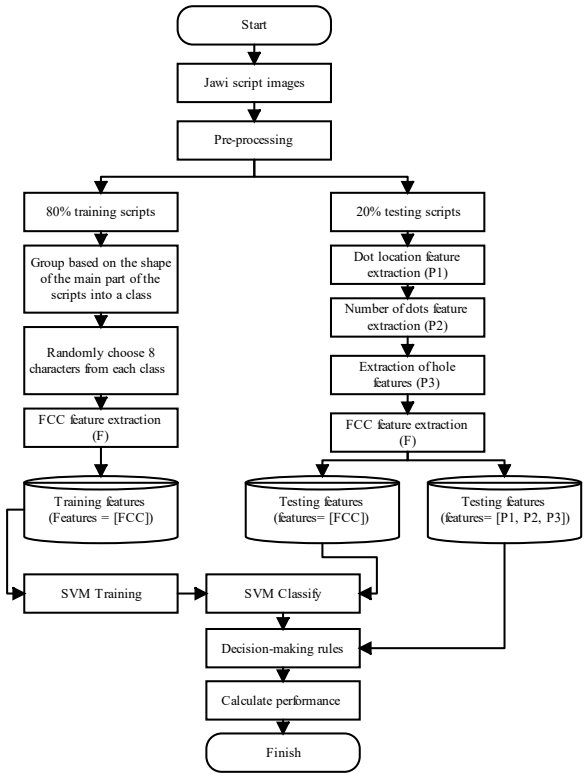

Figure 1. Research Flow

2.2 Pre-processing

Pre-processing in this study refers to research conducted by [9]. In the pre-processing stage, binarization and thinning processes are carried out to obtain a character framework. Binarization is carried out by global thresholding using the Otsu method [11] while thinning to obtain the character framework is used by the Improved Zhang-Suen Thinning Algorithm method [12].

2.3 Extraction of FCC Feature

FCC as feature extraction is used to read the direction of change per pixel from the origin coordinates to the destination coordinates with the FCC value as shown in Figure 2. Figure 3 and Table 3 show examples of how to read FCC in the form of the main part of the characters “ain”, “ghain” and “nga” at 8 neighboring pixels. This feature extraction method is also applied to research [9]. FCC features are given the FCC notation.

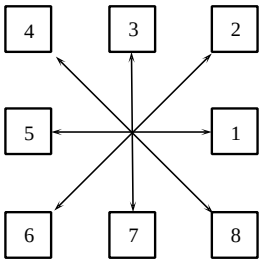

Figure 2. 8-Neighborhood FCC [13]

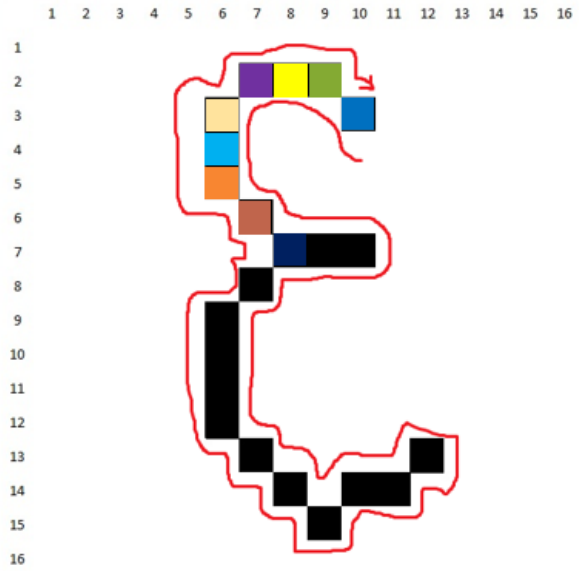

Figure 3. FCC Reading Directions

Table 3. Process of Reading FCC of Figure 3

Origin Coordinates		Destination Coordinates		$(x_1 - x_2)$	$(y_1 - y_2)$	FCC Value
x_1	y_1	x_2	y_2	x	y	
3	10	2	9	–1	–1	4
2	9	2	8	0	–1	5
2	8	2	7	0	–1	5
2	7	3	6	1	–1	6
3	6	4	6	1	0	7
4	6	5	6	1	0	7
5	6	6	7	1	1	8
6	7	7	8	1	1	8

2.4 Extraction of Dot Location Feature

Dot is identified as an area between 5 to 30 pixels that are connected. To determine whether the pixels are connected or not, a connectivity trace is carried out on the script framework with a 3x3 kernel. If the results of the connectivity trace show that the connected pixels exceed 30 pixels, it is considered the main part of the character. It is considered imperfect in pre-processing if it is less than 5 pixels. In this study, the dot location feature is given the notation P1. This feature extraction method is also applied to research [9].

Features the location of the dots is divided into three parts, namely the top, middle and bottom. The top dot

DOI: https://doi.org/10.29207/resti.v7i1.4600

is the location of the pixel set of dots between the topmost part of the image and below the top 6 pixels of the main character, above the yellow line in Figure 4. If the dot is at the top then P1 = 1. The bottom dot is the location of dot pixels under the main part of the character, under the blue line in Figure 4. If the dot is at the bottom, then P1 = 2. The dot in the middle is the pixel location between the upper dot pixels and the lower dot pixels, between the yellow and blue lines in Figure 4. If the dots are in the middle, then P1 = 3. If the processed character does not have a dot, then P1 = 4.

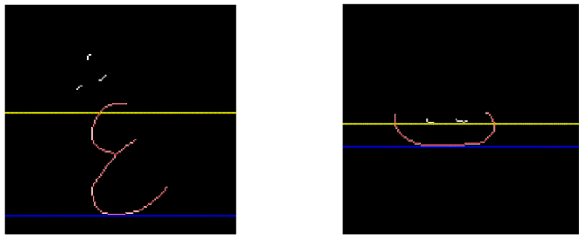

Figure 4. Dot Location Features [9]

2.5 Feature Extraction Number of Dots

The definition of the dot is the same as the feature extraction of the location of the dots, which is an area between 5 to 30 pixels that are connected. The number of dots features is given the notation *P2.* This feature extraction method is also applied in research [9].

The number of dots features is divided into four categories. If the input image does not find dots, then *P2* = 1. If the number of dots on the character is one, then *P2* = 2. If the number of dots on the character is two, then *P2* = 3. If the number of dots on the character is three, then *P2* = 4.

2.6 Extraction Holes Features

In addition to the location of the dots and the number of dots, the character of the hole is also a characteristic of the Jawi script. Characters with the main form of "ب" have a similar shape to characters with the main form of "ف". By detecting the presence of holes, the difference between the two forms of the main part of the character can provide additional information for the classifier. In this study, the kernel [0 1 0; 1 1 1; 0 1 0] to detect the presence of holes. By using the kernel, the inverse image that is being processed will produce 0 pixels in all parts of the input image except for the part of the image that is limited by the edge of the pixel with a value of 0 [13]. If the result of the operation with the kernel produces a pixel that is not 0 then there is a hole in the input image. In this study, it will be considered a hole if the number of pixels is not 0 and the area is more than 5 pixels, this is done to avoid imperfect pre-processing results.

If there is a hole in the input character then P3 = 1, if there is no hole then P3 = 2. For example, the characters "ج" and "ن" do not have holes, then the value of the hole in the character is P3 = 2. The characters "ڤ" and "ز "has a hole then the value of the hole is P3 = 1.

2.6 Classification

In the study for the classification of SVM using the one-against-one, the voting results determine the selected character class. Decision-making rules are applied to the class resulting from the SVM classification so as to produce a classification in the form of Jawi script as shown in Figure 5.

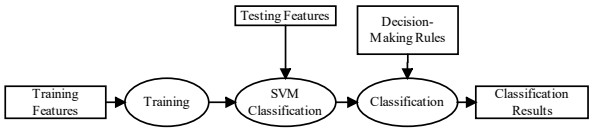

Figure 5. Classification Method

2.7 Testing Procedures

To measure the results of the classification by SVM and decision-making rules, a testing procedure is carried out using equation (1).

$$Accuracy = \frac{Total\ of\ classified\ correctly}{Total\ amount\ of\ testing\ data} \times 100\% \quad (1)$$

3. Results and Discussions

Implementation of the results of this study carried out using Matlab software. In the implementation, several stages are carried out including pre-processing in the form of binarization and thinning, feature extraction, and classification testing using SVM and classification testing using decision-making rules.

3.1 SVM Classification Testing Into 19 Classes

Grouping into 19 classes refers to the shape of the main part of the character without looking at the diacritical marks in the form of the location of dots, the number of dots or the presence of holes. By referring to the shape of the main part of the script, the Jawi script can be grouped into 19 classes, the results of the grouping into 19 classes can be seen in table 4. The number of models built for classification is 19*(19-1)/2 = 171 model. The testing script used is 2×19 = 38 classes of the form of the main part of the character.

In testing the grouping into 19 classes, SVM performs FCC feature classification of the testing characters in 19 classes of training characters. The results of the test grouping into 19 classes can be seen in Table 4. In the table it can be seen that the percentage of success of Jawi script classification by grouping into 19 classes is best obtained at vector lengths of 350 and 500 which is 81.58% with an average the success of classification is 76.32%. The results of this test resulted in the percentage of classification success which was slightly different from the research conducted by [9]. In research [9] the best percentage of success occurred at

DOI: https://doi.org/10.29207/resti.v7i1.4600

400 vector length, which was 80.00%. This is caused by the selection of different training and testing data.

Table 4. Grouping into 19 classes

No	Class	Main Part of Body	Jawi			
			1	2	3	4
1	Alif	ا	ا	-	-	-
2	BaTaTsa	ب	ب	ت	ٹ	-
3	JimHhaKhaCa	ح	ج	ح	خ	چ
4	DalDzal	د	د	ذ	-	-
5	RaZai	ر	ر	ز	-	-
6	SinSyin	س	س	ش	-	-
7	ShadDhad	ص	ص	ض	-	-
8	ThoZho	ط	ط	ظ	-	-
9	AinGhainNga	ع	ع	غ	ڠ	
10	FaPa	ف	ف	ڤ	-	-
11	Qaf	ق	ق	-	-	-
12	KafGa	ک	ک	ڬ	-	-
13	Lam	ل	ل	-	-	-
14	Mim	م	م	-	-	-
15	NunNya	ن	ن	ڽ	-	-
16	WauVa	و	و	ۏ	-	-
17	Ha	ه	ه	-	-	-
18	Hamzah	ء	ء	-	-	-
19	Ya	ى	ي	-	-	-

Table 5. Test Results of SVM Classification Into 19 Class

No.	FCC Vector Length	Percentage of Success
1	100	68,42
2	150	71,05
3	200	71,05
4	250	76,32
5	300	78,95
6	350	81,58
7	400	78,95
8	450	78,95
9	500	81,58

In this SVM classification test, some characters are not classified correctly by SVM. For example, the character class "ن" ("NunNya") is classified by SVM into a class with the basic character form "ب" ("BaTaTsa group"). Another example is the character class "ف" ("FaPa)" which is classified into a class with the basic character form "ن" (class "NunNya") can be seen in Table 5.

3.2 SVM Classification Testing Into 18 Classes

Grouping into 18 classes is done because some characters with the main part form "ک" (class "KafGa"), characters "ک" and "ڬ", handwritten results have the same writing as the part form. main letter "ي", can be seen in Figure 6. So, the three characters (" ,"ک ڬ"" and "ي") are grouped into one class (the "KafGaYa" class). The grouping into 18 classes can be seen in Table 6.

(1) "kaf" (2) "ga" (3) "ya"

Figure 6. Examples of similarities in the shape of the main parts of the characters "ڬ", "ک" and "ي"

Table 6. Grouping Into 18 Classes

No	Class	Main Part of Body	Jawi			
			1	2	3	4
1	Alif	ا	ا	-	-	-
2	BaTaTsa	ب	ب	ت	ٹ	-
3	JimHhaKhaCa	ح	ج	ح	خ	چ
4	DalDzal	د	د	ذ	-	-
5	RaZai	ر	ر	ز	-	-
6	SinSyin	س	س	ش	-	-
7	ShadDhad	ص	ص	ض	-	-
8	ThoZho	ط	ط	ظ	-	-
9	AinGhainNga	ع	ع	غ	ڠ	
10	FaPa	ف	ف	ڤ	-	-
11	Qaf	ق	ق	-	-	-
12	KafGaYa	ک &ى	ک	ڬ	ي	-
13	Lam	ل	ل	-	-	-
14	Mim	م	م	-	-	-
15	NunNya	ن	ن	ڽ	-	-
16	WauVa	و	و	ۏ	-	-
17	Ha	ه	ه	-	-	-
18	Hamzah	ء	ء	-	-	-

In order to be classified by SVM, the binary models built are 18*(18-1) = 153 models. In the character class "KafGaYa" using 4-character training with the main part "ک" and 4-character testing with the main part "ى". While the script testing used is still 2 × 19 = 38 forms of the main part of the script.

Table 7. Test Results of SVM Classification Into 18 class

No.	FCC Vector Length	Percentage of Success
1	100	73,68
2	150	65,79
3	200	71,05
4	250	78,95
5	300	78,95
6	350	81,58
7	400	76,32
8	450	81,58
9	500	81,58

In Table 7 above, it can be seen the results of the introduction of Jawi characters which are grouped into 18 classes, namely by grouping the characters "ڬ", "ک" and "ي" into one class "KafGaYa".

In this test, the highest percentage of classification success was obtained at vector lengths of 350, 450 and 500, which was 81.58% with an average classification success of 76.61%. There was a slight increase in the percentage of recognition compared to introduction into the 19 classes.

3.3 SVM Classification Testing Into 17 Classes

The grouping into 17 classes was carried out because in the 19-class class test there was a misclassification where the character group with the main part "ف" was classified into a character class with the main part "ب" (BaTaTsa class).

In this test there is also a classification error where the group of characters with the main part "ف" is classified into the "Qaf" class and into the "NunNya" class.

DOI: https://doi.org/10.29207/resti.v7i1.4600

In addition, the group of characters with the main form of "ق" (class "Qaf"), is recognized as a class of characters with the main form of "ف" (class "FaPa") and "ں" (class "NunNya").

By considering the shape of the main part and ignoring the holes in the characters, the group of characters with the main part "ب" and "ف" are grouped into the same class, namely the "BaTaTsaFaPa" class, as well as the "ق" and "ں" character groups. into the same class, namely the "QafNunNya" class. The grouping into 17 classes can be seen in Table 8.

Table 8. Grouping into 17 Classes

No	Class	Main part of Body	Jawi				
			1	2	3	4	5
1	Alif	ا	ا	-	-	-	-
2	BaTaTsaFaPa	ب &ف	ب	ت	ٹ	ف	ڤ
3	JimHhaKhaCa	ح	ج	ح	خ	چ	-
4	DalDzal	د	د	ذ	-	-	-
5	RaZai	ر	ر	ز	-	-	-
6	SinSyin	س	س	ش	-	-	-
7	ShadDhad	ص	ص	ض	-	-	-
8	ThoZho	ط	ط	ظ	-	-	-
9	AinGhainNga	ع	ع	غ	ڠ		
10	QafNunNya	ق& ں	ق	ن	ڽ	-	-
11	KafGa	ک	ک	ڬ	-	-	-
12	Lam	ل	ل	-	-	-	-
13	Mim	م	م	-	-	-	-
14	WauVa	و	و	ۏ	-	-	-
15	Ha	ه	ه	-	-	-	-
16	Hamzah	ء	ء	-	-	-	-
17	Ya	ى	ي	-	-	-	-

In order to be classified by SVM, the binary model built is 17*(17-1)/2 = 136. In the "BaTaTsaFaPa" class, 4 training characters are used with the main part "ب" and 4 training characters with the main part. "ف".

In the "QafNunNya" script class, 4 training characters are used with the main part "ق" and 4 training characters with the main part "ں". The testing script used is 2 × 19 = 38 forms of the main part of the character.

From Table 9, it can be seen that the highest classification success is at the vector length of 500, which is 86.84% with an average SVM classification success of 79.24%. There was an increase in recognition compared to the introduction in 18 classes and the introduction in 19 classes.

Table 9. Results of SVM Classification Tests Into 17 Class

No.	Vector Length FCC	Percentage of Success
1	100	71,05
2	150	76,32
3	200	73,68
4	250	81,58
5	300	81,58
6	350	78,95
7	400	81,58
8	450	81,58
9	500	86,84

3.4 SVM Classification Tests Into 16 Classes

The grouping into 16 classes is carried out because in the grouping test into 17 script classes, some characters with the main part form "س", "س" and "ش" characters, are recognized as characters with the part shape. main "ص" (class "ShadDhad") or vice versa. So that groups of characters with the main form of "س" and "ص" are grouped into the same class (class "SinSyinShadDhad").

In order to be classified by SVM, the binary model built is 16*(16-1)/2 = 120. In the "BaTaTsaFaPa" class, 4 training characters are used with the main part "ب" and 4 training characters with the main part. "ف". In the "QafNunNya" script class, 4 training characters are used with the main part "ق" and 4 training characters with the main part "ں". In the script class "SinSyinShadDhad" used 4 training characters with the main part "س" and 4 training characters with the main part "ص". The testing script used is 2 × 19 = 38 forms of the main part of the character.

Grouping into 16 classes is a simplification of grouping into 17 classes, the grouping can be seen in Table 10. In the vector length test of 100, 2 groups of characters with the main parts "س" and "ص" are classified into "Yes" class. " In other vector lengths, the group of characters with the main part form "س" and "ص" are classified into the classes "Ya" and "KafGa". Likewise, when a testing character with the main part "ى" is given as SVM input, several times SVM classifies the "ى" character class into the "SinSyinShadDhad" class. Merging into one character class "ض", "ص", "ش", "س" and "ي" is possible because there are differences in the location of the dots (P1), the number of dots (P2) and the presence of holes (P3) among the five the characters, while combining "ض", "ص", "ش", "س" with the characters "ک" and "ڬ" (class "KafGa") cannot be done because the characters "س" and "ک" have a dot position. (P1), the number of dots (P2) and the presence of holes (P3) are the same.

In this test, there was also a significant misclassification of characters with the main part form "د", these characters are often classified into the "RaZai" class, or vice versa. Scripts with the main form of "ر" are classified into the "DalDzal" class. However, all the characters from the two classes cannot be combined into one group because the characters "د" and "ر" or the characters "ذ" and "ز" because they have the location of dots (P1), the number of dots (P2) and the presence of holes (P3) the same one.

From Table 11, it can be seen that the results of the introduction of Jawi characters which are grouped into 16 classes are at the vector length of 350 which is 84.21 % with an average success of SVM classification of 78.95%. There was a decrease in recognition compared to the introduction of 17 classes. This happens because

DOI: https://doi.org/10.29207/resti.v7i1.4600

the group of characters with the main form of "س" and "ص" is classified into the "Ya" class.

Table 10. Grouping into 16 Class

No	Class	Main part of Body	Jawi				
			1	2	3	4	5
1	Alif	ا	ا	-	-	-	-
2	BaTaTsaFaPa	ب &ف	ب	ت	ث	ف	ڤ
3	JimHhaKhaCa	ح	ج	ح	خ	چ	-
4	DalDzal	د	د	ذ	-	-	-
5	RaZai	ر	ر	ز	-	-	-
6	SinSyin	س	س	ش	-	-	-
7	ShadDhad	ص	ص	ض	-	-	-
8	ThoZho	ط	ط	ظ	-	-	-
9	AinGhainNga	ع	ع	غ	ڠ		
10	QafNunNya	ق & ن	ق	ن	ڽ	-	-
11	KafGa	ک	ک	ک	-	-	-
12	Lam	ل	ل	-	-	-	-
13	Mim	م	م	-	-	-	-
14	WauVa	و	و	ۏ	-	-	-
15	Ha	ه	ه	-	-	-	-
16	Hamzah	ء	ء	-	-	-	-

Table 11. Results of SVM Classification Tests Into 16 Class

No.	FCC Vector Length	Percentage of Success
1	100	68,42
2	150	78,95
3	200	76,32
4	250	81,58
5	300	76,32
6	350	84,21
7	400	78,95
8	450	81,58
9	500	84,21

3.5 Testing SVM Classification Into 15 Classes

Grouping into 15 classes is carried out because in testing the grouping into 16 classes, several characters with the main part form "ى", are classified into the "SinShadShadDhad" class or vice versa. So that in this test the characters "ض", "ص", "ش", "س" and "ي" are grouped into the same class, namely the "SinShadShadDhadYa" class.

In order to be classified by SVM, the binary model built is 15 *(15-1)/2 = 105. In the “BaTaTsaFaPa” class, 4 training characters are used with the main part “ب” and 4 training characters with the main part “ف”. In the "QafNunNya" script class, 4 training characters are used with the main part "ق" and 4 training characters with the main part "ن". In the script class "SinSyinShadDhadYa" used 3 training characters with the main part "3", س training characters with the main part "ص" and 2 training characters with the main part "ى".

Grouping in 15 classes is a simplification of grouping into 16 classes, the grouping can be seen in Table 12. In this test, the characters with the main part form "ن " are very dominant misclassified by SVM, SVM dominant classifies the two characters into classes "BaTaTsaFaPa". The testing script used is 2 × 19 = 38 forms of the main part of the character.

Table 12. Grouping into 15 Class

No	Class	Main Part of Body	Jawi				
			1	2	3	4	5
1	Alif	ا	ا	-	-	-	-
2	BaTaTsaFaPa	ب &ف	ب	ت	ث	ف	ڤ
3	JimHhaKhaCa	ح	ج	ح	خ	چ	-
4	DalDzal	د	د	ذ	-	-	-
5	RaZai	ر	ر	ز	-	-	-
6	SinSyinShadDhadYa	س, ص, &ى	س	ش	ص	ض	ي
7	ThoZho	ط	ط	ظ	-	-	-
8	AinGhainNga	ع	ع	غ	ڠ		
9	Qaf NunNya	ق & ن	ق	ن	ڽ	-	-
10	KafGa	ک	ک	ک	-	-	-
11	Lam	ل	ل	-	-	-	-
12	Mim	م	م	-	-	-	-
13	WauVa	و	و	ۏ	-	-	-
14	Ha	ه	ه	-	-	-	-
15	Hamzah	ء	ء	-	-	-	-

Table 13. Test Results of SVM Classification Into 15 Class

No.	***FCC Vector Length***	***Percentage of Success***
1	100	68,42
2	150	78,95
3	200	76,32
4	250	81,58
5	300	76,32
6	350	84,21
7	400	78,95
8	450	81,58
9	500	84,21

From Table 13, it can be seen that the highest percentage of classification success is at vector lengths of 350 and 500, which is 84.21% with an average classification success of 78.95%. There was an increase in recognition compared to the introduction of 16 classes, 17 classes, 18 classes and recognition in 19 classes.

3.6 Testing the Classification of Decision-Making Rules on the Results of SVM Classification

This test was conducted to determine the Jawi script based on the results of the SVM classification on the SVM grouping into 15 classes. The SVM grouping into 15 classes was chosen to apply the classification of decision-making rules because it has a better percentage of success when compared to other groupings. In the SVM classification test, several characters with the same main part or similar in shape to the main character are grouped in the same class. For example, the characters “ض”, “ص, “”ش”, “س” and “ي” are grouped into the same class, namely the “SinSyinShadDhadYa” class. Another example is the letters “ٹ”, “ت”, “ب”, “ف” and “ڤ” which are also grouped into the same class, namely the “BaTaTsaFaPa” class. In this test, the Jawi script can be defined if the validation results according to Table 14 all meet. In this test the average percentage of successful classification is better than grouping into 15 classes. In the grouping of 15 classes classified as SVM the average success reached 78.95%,

DOI: https://doi.org/10.29207/resti.v7i1.4600

while after applying the decision-making rules the average success increased to 80.48%. These results indicate that the extraction of the location of the dot, the number of dots and the presence of holes can make a difference in increasing the percentage of classification success. The results of the classification with the decision-making rules on the results of the SVM classification can be seen in Table 15.

Table 14. Decision-Making Rules Based on SVM Classification Class

No	Script	Jawi	Rules		
			Class	P 1	P 2	P 3
1	Alif	ا	Alif	4	1	2
2	Ba	ب	BaTaTsaFaPa	2	2	2
3	Ta	ت	BaTaTsaFaPa	1	3	2
4	Tsa	ث	BaTaTsaFaPa	1	4	2
5	Jim	ج	JimHhaKhaCa	3	2	2
6	Hha	ح	JimHhaKhaCa	4	1	2
7	Kha	خ	JimHhaKhaCa	1	2	2
8	Dal	د	DalDzal	4	1	2
9	Dzal	ذ	DalDzal	1	2	2
10	Ra	ر	RaZai	4	1	2
11	Zai	ز	RaZai	1	2	2
12	Sin	س	SinSyinShadDhadYa	4	1	2
13	Syin	ش	SinSyinShadDhadYa	1	4	2
14	Shad	ص	SinSyinShadDhadYa	4	1	1
15	dhad	ض	SinSyinShadDhadYa	1	2	1
16	Tho	ط	ThoZho	4	1	1
17	Zho	ظ	ThoZho	3	2	1
18	Ain	ع	AinGhainNga	4	1	2
19	ghain	غ	AinGhainNga	1	2	2
20	Fa	ف	BaTaTsaFaPa	1	2	1
21	qaf	ق	QafNunNya	1	3	1
22	kaf	ک	KafGa	4	1	2
23	lam	ل	Lam	4	1	2
24	mim	م	Mim	4	1	1
25	nun	ن	QafNunNya	1	2	2
26	wau	و	WauVa	4	1	1
27	ha	ه	Ha	4	1	1
28	hamzah	ء	Hamzah	4	1	2
29	ya	ي	SinSyinShadDhadYa	2	3	2
30	nya	ڽ	QafNunNya	1	4	2
31	ca	چ	JimHhaKhaCa	3	4	2
32	nga	ڠ	AinGhainNga	1	4	2
33	pa	ڤ	BaTaTsaFaPa	1	4	1
34	ga	گ	KafGa	1	2	2
35	va	ۏ	WauVa	1	2	1

In this test, there are several shortcomings, including in the "ن" testing script. In the script test, the class resulting from the SVM classification is “BaTaTsaFaPa”, but of all the characters that are included in the “BaTaTsaFaPa” class, there is no character that has one dot on the top and does not have a hole. The character "ب" has a dot but does not have a hole, but the location of the dot of the letter "ب" is below not above. While the character “ف” has a dot at the top, the character “ف” has a hole. In this case, the character “ن” cannot be classified as a character because it does not meet all the criteria in Table 14.

Likewise, the character "ف" which is classified by SVM into the "QafNunNya" character class. In that class, there are no characters that have a dot and are located at the top and have holes. The character "ق" has two dots on the top while the character "ن" has one dot on the top but the character "ن" has no holes. Similar to the case of the character “ف”, the character “ڤ” also cannot be classified as a character because it does not meet all the criteria in Table 15.

Table 15. Results of Testing Decision-making Rules Based on Class SVM Classification Results

No.	FCC Vector Length	Percent Success
1	100	78,57
2	150	77,14
3	200	77,14
4	250	84,29
5	300	82,86
6	350	85,71
7	400	80,00
8	450	80,00
9	500	78,57

3.7 Testing the Classification of Decision-Making Rules Against the Combination of Several Classes from the SVM Classification Results

To overcome the problems that occur in Sub-section 3.6, some of the testing characters will be classified into the appropriate characters if selected from the appropriate character class or a combination of several character classes using the "||" operator. (operator “or”). In addition to a class or a combination of several classes, to define the testing script into Jawi script, three extractions are also needed, namely the location of the dot (P1), the number of dots (P2) and the presence of a hole (P3). The combined decision-making rules for several classes resulting from the SVM classification can be seen in Table 16.

In testing using the decision-making rules in Table 16, there was an increase in the success of the classification. This is because some of the misclassified characters in Sub-section 3.6 can be classified correctly. For example, the character “ن” which is classified by SVM into the class “BaTaTsaFaPa”. If you only use one class “BaTaTsaFaPa” then the testing script “ن” cannot be classified into Jawi script because none of them meet the criteria of Table 14. However, if the testing script “ن” is defined as a combination of classes “QafNunNya || BaTaTsaFaPa", and other criteria which have the location of the dot above (P1), the number of dots is one (P2) and does not have a hole (P3), the results of the classification of decision-making rules will be correct, namely classified into the letter "ن".

Another example is the character “ڤ” which is classified by SVM into the “QafNunNya” class. If you only use one class "QafNunNya" then the testing script "ڤ" cannot be classified into Jawi script because none of them meet the criteria of Table 14. However, if the testing script "ڤ" is defined as a combination of classes "BaTaTsaFaPa || QafNunNya”, and other criteria which have the location of the dot above (P1), the number of

DOI: https://doi.org/10.29207/resti.v7i1.4600

three dots (P2) and have a hole (P3), the results of the classification of decision-making rules will be correct, which is classified into the letter "ڤ".

Table 16. Decision-making Rules with A Combination of Several Classes of SVM Classification Results.

No	Script	Jawi	Rules			
			Class	P 1	P 2	P 3
1	alif	ا	Alif	4	1	2
2	ba	ب	BaTaTsaFaPa || QafNunNya	2	2	2
3	ta	ت	BaTaTsaFaPa || QafNunNya	1	3	2
4	tsa	ث	BaTaTsaFaPa	1	4	2
5	jim	ج	JimHhaKhaCa	3	2	2
6	hha	ح	JimHhaKhaCa	4	1	2
7	kha	خ	JimHhaKhaCa	1	2	2
8	dal	د	DalDzal || WauVa	4	1	2
9	dzal	ذ	DalDzal || WauVa	1	2	2
10	ra	ر	RaZai	4	1	2
11	zai	ز	RaZai	1	2	2
12	sin	س	SinSyinShadDhadYa	4	1	2
13	syin	ش	SinSyinShadDhadYa	1	4	2
14	shad	ص	SinSyinShadDhadYa	4	1	1
15	dhad	ض	SinSyinShadDhadYa	1	2	1
16	tho	ط	ThoZho	4	1	1
17	zho	ظ	ThoZho	3	2	1
18	ain	ع	AinGhainNga	4	1	2
19	ghain	غ	AinGhainNga	1	2	2
20	fa	ف	BaTaTsaFaPa || QafNunNya	1	2	1
21	qaf	ق	QafNunNya || BaTaTsaFaPa	1	3	1
22	kaf	ک	KafGa	4	1	2
23	lam	ل	Lam	4	1	2
24	mim	م	Mim	4	1	1
25	nun	ن	QafNunNya || BaTaTsaFaPa	1	2	2
26	wau	و	WauVa || DalDzal || RaZai	4	1	1
27	ha	ه	Ha	4	1	1
28	hamzah	ء	Hamzah	4	1	2
29	ya	ي	SinSyinShadDhadYa	2	3	2
30	nya	ڽ	QafNunNya	1	4	2
31	ca	چ	JimHhaKhaCa	3	4	2
32	nga	ڠ	AinGhainNga	1	4	2
33	pa	ڤ	BaTaTsaFaPa || QafNunNya	1	4	1
34	ga	ک	KafGa	1	2	2
35	va	ۏ	WauVa	1	2	1

In other cases, as seen in Sub-section 3.1, some characters "ت" and "ٹ" are classified into the "FaPa" class. The feature of presence or absence of holes in a character cannot be maximally classified by SVM. To maximize character recognition, the incorporation of several groups of characters into a class is very important in the introduction of Jawi characters.

From the Table 17, it can be seen the modifications to the decision-making rules. Modification of the rules is carried out by observing that there are no dot location features (P1), number of dots (P2) and the same hole (P3) from each character in a decision-making rule validation class. The results of testing the application of the modified decision-making rules can be seen in Table 17. From the table it can be seen that the percentage of successful classification based on a combination of several classes resulting from the SVM classification is at vector lengths of 300 and 350 with a success rate of 88.57% with an average success classification of 84.13%. When compared with the test results based on the SVM classification results class in Sub-section 3.6, the test by combining several SVM classification results obtained better results.

Table 17. Test Results Based on the Combination of Several Classes of SVM Classification Results

No.	FCC Vector Length	Percent Success
1	100	85,71
2	150	78,57
3	200	81,43
4	250	87,14
5	300	88,57
6	350	88,57
7	400	82,86
8	450	82,86
9	500	81,43

3.7 Testing the Classification of Decision-Making Rules for the Combination of Several Classes by Observing the Tendency of the SVM Classification Results

This test is carried out because there is a tendency for some characters to be classified by SVM into different classes. For example, the character "ق " which is classified by SVM into Classes "QafNunNya" and "SinSyinShadDhadYa".

There is a tendency for SVM to classify into different classes, for example from two testing characters "ق", there is a tendency for one testing character to be classified into the "QafNunNya" class and another testing character to be classified into the "SinSyinShadDhadYa" class. This occurs at vector lengths of 100, 200, 250, 400 and 450. To overcome the problem of the tendency of a character to be classified by SVM into different character classes, a combination of several broader classes can be applied but by taking into account the differences in dot location features (P1), the number of dots (P2) and the presence of holes (P3) from each character in each class combined with the "||" operator. For example, to define the character "ق", the character class can be derived from the class "QafNunNya || BaTaTsaFaPa || SinSyinShadDhadYa" because none of the characters from the entire class have two dots above and have holes.

In Table 18, it can be seen modifications to the decision-making rules by taking into account that there are no dot location features (P1), the number of dots (P2) and the presence of the same hole (P3) for each character in a class so that the validation of the decision-making rules is not wrong. The results of testing the application of the modified decision-making rules can be seen in Table 19. From the table it can be seen that the percentage of

DOI: https://doi.org/10.29207/resti.v7i1.4600

successful classification based on a combination of several classes taking into account the tendency of SVM classification is at the vector length of 350 which is 92.86% with an average success of 87 .14%.

Table 18. Decision-Making Rules by Combining Several Classes by Paying Attention to The Trend of SVM Classification Results.

No	Script	Jawi	Rules			
			Class	P 1	P 2	P 3
1	alif	ا	Alif	4	1	2
2	ba	ب	BaTaTsaFaPa || QafNunNya	2	2	2
3	ta	ت	BaTaTsaFaPa || QafNunNya	1	3	2
4	tsa	ث	BaTaTsaFaPa	1	4	2
5	jim	ج	JimHhaKhaCa	3	2	2
6	hha	ح	JimHhaKhaCa	4	1	2
7	kha	خ	JimHhaKhaCa	1	2	2
8	dal	د	DalDzal || WauVa || Lam	4	1	2
9	dzal	ذ	DalDzal || WauVa || Lam	1	2	2
10	ra	ر	RaZai	4	1	2
11	zai	ز	RaZai	1	2	2
12	sin	س	SinSyinShadDhadYa || QafNunNya	4	1	2
13	syin	ش	SinSyinShadDhadYa || QafNunNya	1	4	2
14	shad	ص	SinSyinShadDhadYa	4	1	1
15	dhad	ض	SinSyinShadDhadYa	1	2	1
16	tho	ط	ThoZho	4	1	1
17	zho	ظ	ThoZho	3	2	1
18	ain	ع	AinGhainNga	4	1	2
19	ghain	غ	AinGhainNga	1	2	2
20	fa	ف	BaTaTsaFaPa || QafNunNya	1	2	1
21	qaf	ق	QafNunNya || BaTaTsaFaPa || SinSyinShadDhadYa	1	3	1
22	kaf	ک	KafGa	4	1	2
23	lam	ل	Lam	4	1	2
24	mim	م	Mim	4	1	1
25	nun	ن	QafNunNya || BaTaTsaFaPa	1	2	2
26	wau	و	WauVa || DalDzal || RaZai	4	1	1
27	ha	ه	Ha	4	1	1
28	hamzah	ء	Hamzah	4	1	2
29	ya	ي	SinSyinShadDhadYa || KafGa	2	3	2
30	nya	ڽ	QafNunNya	1	4	2
31	ca	چ	JimHhaKhaCa	3	4	2
32	nga	ڠ	AinGhainNga	1	4	2
33	pa	ڤ	BaTaTsaFaPa || QafNunNya	1	4	1
34	ga	ک	KafGa	1	2	2
35	va	ۏ	WauVa || DalDzal || RaZai	1	2	1

In Table 19 it can be seen the level of success of the classification of the application of the decision-making rules for the introduction of Jawi characters with a combination of several classes by taking into account the tendency of the SVM classification results in different lengths of FCC feature vectors. When compared with the test results based on the class of SVM classification results in sub-section 3.6. then the test by combining several classes by paying attention to the tendency of the SVM classification results to get better recognition results.

Table 19. Test Results Based on the Combination of Several Classes of SVM Classification Results

No.	FCC Vector Length	Percent Success
1	100	87,14
2	150	80,00
3	200	84,29
4	250	90,00
5	300	91,43
6	350	92,86
7	400	84,29
8	450	87,14
9	500	87,14

Errors in prediction are generally caused by the shape of the main part of the character which has similarities and has the same location and number of dots. For example, the character "ر" is classified into the character "د" or vice versa, as shown in Figure 7(a). In addition, the character "ت" is also mostly classified into the character "ڽ" or vice versa as shown in Figure 7(b). This misclassification occurs at all feature lengths.

(a) (b)

Figure 7. Most misclassification (a) characters "ر" and "د", and (b) characters "ت" and "ڽ"

The test results above show that classification results can be improved by combining several Jawi script classes that have similarities in the shape of the main parts. Another test result is that decision making rules can improve classification accuracy by combining several classes that are often misclassified by SVM.

However, the application of the Jawi script classification as a result of this research is not yet ready to be applied. A complete Jawi script dataset is needed other than in a separate form, namely in a connected form at the initial, in the middle and at the end so that the classification process can be carried out. The collection of this dataset will be carried out in further research.

4. Conclusion

Character recognition research using Freeman Chain Code (FCC) feature extraction and Support Vector Machine (SVM) as a classifier can be one solution in recognizing handwritten Jawi characters. The results of this study indicate that the application of feature extraction of Jawi script using FCC into 19 classes which are distinguished based on the shape of the main part of the character can be classified by SVM with a success percentage of 81.58%. Re-simulation is done by reducing the number of classes by combining several

DOI: https://doi.org/10.29207/resti.v7i1.4600

forms of characters into one class. Regrouping into 15 classes is able to produce a better SVM recognition rate, reaching 84.21%. In addition, by applying the decision-making rules to the class of SVM classification results into Jawi script by taking into account the tendency of the SVM classification results to increase the percentage of success of Jawi script classification with a success rate of 92.86%. Subsequent research to classify the Jawi script using OCR technology is to acquire a dataset of the Jawi script in a connected form at the initial, in the middle and at the end and calculate the effect of the vector length of the point location and the vector length of the number of points on the main parts of the script.

Reference

[1] K. Saddami, K. Munadi, and F. Arnia, "A database of printed Jawi character image," in *2015 Third International Conference on Image Information Processing (ICIIP)*, Dec. 2015, pp. 56–59. doi: 10.1109/ICIIP.2015.7414740.

[2] M. F. Nasrudin, K. Omar, M. S. Zakaria, and L. C. Yeun, "Handwritten Cursive Jawi Character Recognition : A Survey," pp. 247–256, 2008, doi: 10.1109/CGIV.2008.36.

[3] F. Mushtaq, M. M. Misgar, M. Kumar, and S. S. Khurana, "UrduDeepNet: offline handwritten Urdu character recognition using deep neural network," *Neural Comput Appl*, vol. 33, no. 22, pp. 15229–15252, Nov. 2021, doi: 10.1007/s00521-021-06144-x.

[4] M. A. K.O and S. Poruran, "OCR-Nets: Variants of Pre-trained CNN for Urdu Handwritten Character Recognition via Transfer Learning," *Procedia Comput Sci*, vol. 171, pp. 2294–2301, 2020, doi: 10.1016/j.procs.2020.04.248.

[5] S. Naz *et al.*, "Offline cursive Urdu-Nastaliq script recognition using multidimensional recurrent neural networks," *Neurocomputing*, vol. 177, pp. 228–241, Feb. 2016, doi: 10.1016/j.neucom.2015.11.030.

[6] G. A. Montazer, H. Q. Saremi, and V. Khatibi, "A neuro-fuzzy inference engine for Farsi numeral characters recognition," *Expert Syst Appl*, vol. 37, no. 9, pp. 6327–6337, Sep. 2010, doi: 10.1016/j.eswa.2010.02.088.

[7] A. Broumandnia and J. Shanbehzadeh, "Fast Zernike wavelet moments for Farsi character recognition," *Image Vis Comput*, vol. 25, no. 5, pp. 717–726, May 2007, doi: 10.1016/j.imavis.2006.05.014.

[8] M. Namazi, Ms. Student, and K. Faez, "Application of a Neural Network for multifont Farsi character recognition using fuzzified Pseudo - Zernike moments," in *Proceedings IWISP '96*, Elsevier, 1996, pp. 361–364. doi: 10.1016/B978-044482587-2/50080-X.

[9] . S., F. Arnia, and R. Muharar, "Pengenalan Aksara Jawi Tulisan Tangan Menggunakan Freemen Chain Code (FCC), Support Vector Machine (SVM) dan Aturan Pengambilan Keputusan," *Jurnal Nasional Teknik Elektro*, vol. 5, no. 1, p. 45, 2016, doi: 10.25077/jnte.v5n1.185.2016.

[10] S. Safrizal, "Pengenalan Karakter Jawi Tulisan Tangan Menggunakan Fitur Sudut," *VOCATECH: Vocational Education and Technology Journal*, vol. 1, no. 1, pp. 1–4, 2019, doi: 10.38038/vocatech.v1i0.1.

[11] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans Syst Man Cybern*, 2008, doi: 10.1109/tsmc.1979.4310076.

[12] W. Chen, L. Sui, Z. Xu, and Y. Lang, "Improved Zhang-Suen thinning algorithm in binary line drawing applications," in *2012 International Conferesnce on Systems and Informatics (ICSAI2012)*, May 2012, no. Icsai, pp. 1947–1950. doi: 10.1109/ICSAI.2012.6223430.

[13] D. Nasien, H. Haron, and S. S. Yuhaniz, "Support Vector Machine (SVM) for English Handwritten Character Recognition," in *2010 Second International Conference on Computer Engineering and Applications*, 2010, vol. 1, no. January, pp. 249–252. doi: 10.1109/ICCEA.2010.56.

DOI: https://doi.org/10.29207/resti.v7i1.4600