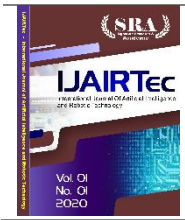


International Journal of Artificial Intelligence and Robotic Technology (IJAIRTec)

PISSN: 2775-5800 EISSN: 2775-5797
Volume: 1 Number: 3 pp. 90-100, December 2021
http://publishing.sracademic.org/index.php/IJAIRTec

KOREAN LETTER HANDWRITING RECOGNITION USING CONVOLUTIONAL NEURAL NETWORK METHOD VGG-16 ARSITEKTUR ARCHITECTURE

Ery Hartati[1*], Derry Alamsyah[2], Nataliatara[3]
[1,2,3]Multi Data Palembang University, South Sumatra, Indonesia
[1,2,3]Rajawali Street No.14, 9 Ilir, Kec. Ilir Tim. II, Palembang City, South Sumatra 30113, Indonesia
E-Mail: ery_hartati@mdp.ac.id[1], derry@mdp.ac.id[2], nataliaantara@mhs.mdp.ac.id[3]
Correspondence: ery_hartati@mdp.ac.id*

Article Info
Article history:
Received October 28, 2021
Revised November, 2021
Accepted December 5, 2021

Keywords:
Handwritten;
Korean Alphabet;
VGG-16.

Abstract
Handwritten is a unique characteristic because each people has different handwriting. Handwritten can be an object to recognition of someone. In research on handwritten Korean alphabet recognition using the Convolutional Neural Network method with VGG-16 architecture. Data is scanned from 24 Korean handwritten alphabets with 14 kinds of consonants and 10 kinds of vocals on paper with black ink. Data there are two scenarios namely research using original data without binarization and data with binarization which for both scenarios are previously data has been resized. This research uses k-fold cross-validation with a value for k=5 and a confusion matrix. The result showed that both of scenarios are can be recognized with 99,52% accuracy, 95,56% precision, 94,11% recall for first scenario and 99,42% accuracy, 95,94% precision, 93,11% recall for second scenario.

I. INTRODUCTION

Korean culture is growing rapidly in many countries. In Indonesia, Korean culture has a strong appeal among women, especially teenagers. One of Korean culture is Korean music called kpop which has also successfully collaborated with world musicians and won international awards such as boyband BTS, boyband EXO and Wendy RedVelvet who collaborated with John Legend [1]. As the number of fans of Korean culture increases, more and more people learn Korean. Korean characters are different like languages with Indonesian. Learning Korean has its own difficulties. This is because the characters in Korean cannot stand alone and must be combined with other characters so that they have their own meaning. People who want to learn are difficult to understand on their own [2].

In Korean character recognition research, data can be obtained in various ways, namely public data and collecting data independently. Public data on handwriting Korean characters for example SERI95a and PE92 [3]. SERI95a has 520 character classes with each class having 1000 samples whereas, PE92 has 2350 classes with each class having 100

samples. Collecting data that was taken independently without using public datasets, one of which was using tools such as canvas [4][5] or manually using handwritten paper [6].

Deep Learning is one of the artificial neural network models that are now widely used, including Korean letter recognition, for example the Convolutional Neural Network (CNN) method which is the development of the MultiLayer Peceptron (MLP) and is part of the Deep Neural Network [7]. Research using other handwritten paper was carried out in the introduction of Javanese characters using the CNN method scanned into images using a scanner. The results obtained stated that CNN was able to recognize characters with the highest level of accuracy of 73.60% [8]. Another Javanese language recognition study that uses the CNN method produces an accuracy rate of 97.52% for the custom model and 97.558% for the VGG-16 model [9] and Tamil produces an accuracy rate of 90% [10].

Unlike alphabet characters, Korean characters are formed by more than one character unit (grapheme). The placement of each grapheme on a syllable depends on the type of letter included in the consonant or vowel. This makes writing Korean characters difficult to recognize [11]. After searching for existing literature studies, many have made Korean characters the object of research aimed at recognizing Korean characters because Korean characters are difficult to recognize. There is a study that performs Korean character recognition yields relatively low results of 22.083% [6]. The author wants to try to improve the accuracy of Korean letter recognition with the same dataset. The dataset has colors other than black and white, so this research is divided into 2 scenarios. The first is a scenario with data that is not binary and the second is a scenario with data that is binary. Binerization is done to change the color intensity value to 0 (black) and 1 (white). The Convolutional Neural Network method that has previously been used has yielded quite high results. Thus, the author uses the CNN or Convolutional Neural Network method to perform handwriting recognition of Korean letters.

II. RESEARCH METHOD

2.1 Problem Identification

This stage begins with a topic regarding the development of research on letter recognition. Many studies on letter recognition have been carried out with different objectives and results.

2.2 Literature Study

At this stage, literature study is carried out from journals and books as well as internet references related to the topic of this research, namely research related to the introduction of Korean letters and the CNN or Convolutional Neural Network method. This stage aims to collect relevant journals and books as well as internet references so that the research has a strong basis.

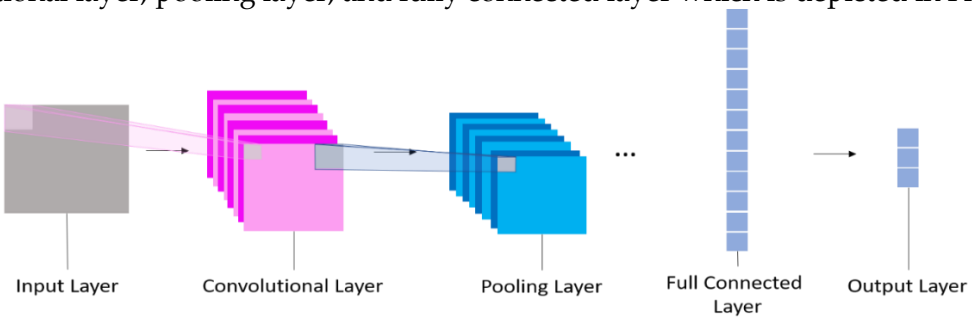
2.3 Korean

Korean has the official name Hangeul (한글). In the Chosun dynasty in 1443, Korean characters were created by the Great King Sejong which were used by the Korean people [12]. Korean as the official language used by two countries, namely South Korea and North Korea, has 70% of the vocabulary formed from Hanja or taken from Mandarin. The number of true Hangul letters consists of 24 letters. The letter is divided into 2, namely 10 vowels and 14 consonants [4]. Hangul letters also consist of extended vowels which are a combination of basic vowels. There are 11 expansion vowels, such as and, as well as, and which are not too different. So, the vowel is not too much attention. Consonant letters also have

2.4 Convolutional Neural Network (CNN)

The Convolutional Neural Network, also known as CNN, is a development of the MultiLayer Peceptron (MLP) and is part of the Deep Neural Network. MLP is used to

process two-dimensional data. In image classification, MultiLayer Peceptron considers each pixel as an independent feature so that the results are not good so it is not suitable for storing spatial image data. With the development of deep learning, the lack of MLP in complex data requires that the data transform function is easier to understand. Thus, deep learning develops where the model is given several layers. This layer aims to perform data transformation. This triggers the number of layers in the neural network model to increase. Finally, the number of layers in the neuron network is considered as a hyperparameter using a searching approach [7]. CNN is a type of neural network consisting of a convolutional layer, pooling layer, and fully connected layer which is depicted in Figure 1.



Picture 1 Ilustrasi Convolutional Neural Network

Convolutional layer is a layer that is useful for performing convolution operations on the previous layer's output. The main process in CNN is carried out at the convolution layer with the aim of extracting features from the input image data. The result of the convolution is a linear transformation of the input data according to the spatial information in the data. Pooling layer is a layer that reduces the size of the matrix by performing pooling operations. Pooling consists of average pooling and max pooling. Average pooling uses the average value. Max Pooling uses the average value. Full Connected Layer is used to transform dimensions into one-dimensional data. This is done with the aim of the data can clarify linearly. Before the neuron is entered into the fully connected layer, the neuron is transformed first [7].

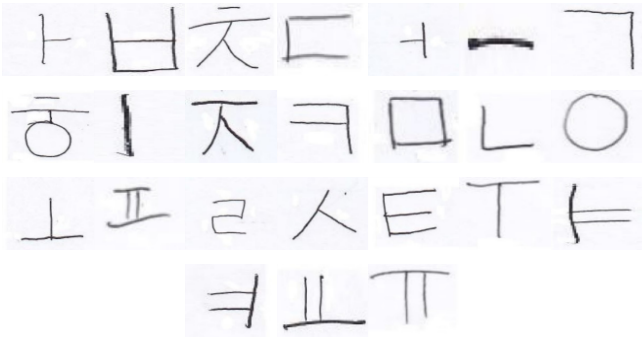
2.5. Visual Geometry Group (VGG-16)

The VGG network is an architectural type designed by Karen Simonyan and Andrew Zisserman of the Visual Geometry Group, Department of Engineering Science, University of Oxford. Architecture with input size of 224 × 224 pixels. In the 2014 ImageNet Challenge competition, this architecture won the top ranking for localization and classification [14]. VGG is also divided into several layers, namely 11 layers, 13 layers, 16 layers and 19 layers. VGG-16 which is a version of the VGG architecture which has as many as 16 layers. In previous studies, the error value shown by VGG-16 has the smallest value compared to the other VGG error values [15]. Comparison of multi-scale training and single-scale training methods was carried out. Multi-scale training is to change the image size to 1 size with a certain distance then, the image is cut and then adjusted to the input layer. Single-scale training is to resize the image to a size and crop it and adjust it to the input layer. Single-scale training has a greater chance of losing objects than multi-scale training. VGG-19 or VGG with 19 layers produces an error value of 9% in single-scale training and 8% in multi-scale training, while VGG-16 produces an error value of 8.8% in single-scale training and 8.1% in multi-scale training.

2.6. Data collection and preprocessing

At this stage, data collection is carried out in the form of a dataset of Korean fonts. The dataset was obtained from previous research [6]. The dataset contains 14 types of Korean

consonants and 10 vowels. An example of an image can be seen in Figure 2. The dataset has a variable amount of data so that in this study 45 photos were used for each type. So, the total data is 1080 photos. The dataset was resized to a size of 224 x 224 and carried out 2 scenarios, namely data that was not binary and binary was performed.



Picture 2. Sample Dataset of Each Letter

2.7. System Design

At this stage, data collection is carried out in the form of a dataset of Korean fonts. The dataset was obtained from previous research [6]. The dataset contains 14 Korean consonants and 10 vowels. The dataset has a variety of data, so in this study 45 photos were used for each type. So, the total data is 1080 photos. The dataset was resized to a size of 224 x 224 and carried out 2 scenarios, namely data that was not binary and binary was performed.

2.8. Implementation

At this stage, the implementation of Korean letter object recognition is carried out based on system design and using existing datasets. The introduction implementation uses the Python programming language and the Tensorflow library and Keras Convolutional Neural Network (CNN) method with the VGG-16 architecture. Transfer learning is carried out in the implementation of CNN, which transfers previous training knowledge. The VGG-16 architecture is divided into a feature extraction layer and a fully connected layer. In this study, the VGG-16 architecture used is feature extraction and fully connected as many as 2 consisting of 4096 neurons and 24 neurons for output. The neurons in the output represent 24 classes i.e. 24 Korean typefaces

2.9. System Evaluation

This stage evaluates the system that has been designed previously. The purpose of this stage is to ensure that the research process runs smoothly and minimizes errors. Evaluation of model performance using k-fold cross validation. Where the value used is the value of k = 5, meaning that the data will be divided into 5 folds and 5 iterations. In 5 folds in each iteration will be divided into training data, validation data and test data. Evaluation of the results is carried out using the Confusion Matrix method which will calculate precision, recall and accuracy

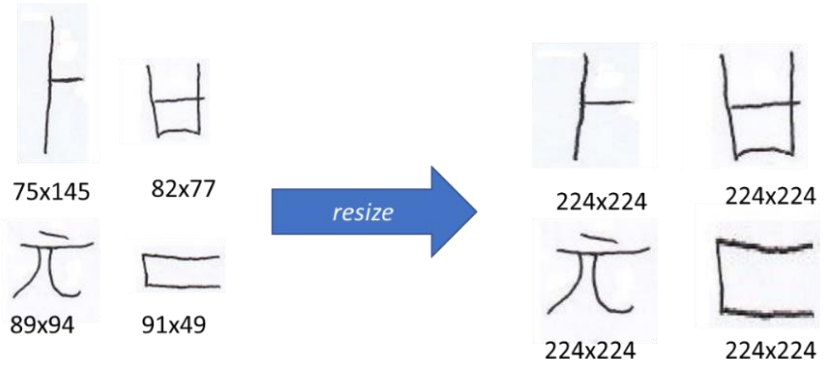
III. RESULTS AND DISCUSSION

3.1 Results

The CNN (Convolutional Neural Network) method performs recognition with Korean letter objects to classify 24 types of Korean letters consisting of 14 types of basic consonants and 10 types of basic vowels. The dataset comes from previous research in the form of handwritten scans on paper. The datasets have varying amounts so that in this study 45 pieces were used for each Korean typeface. The initial images with various sizes were

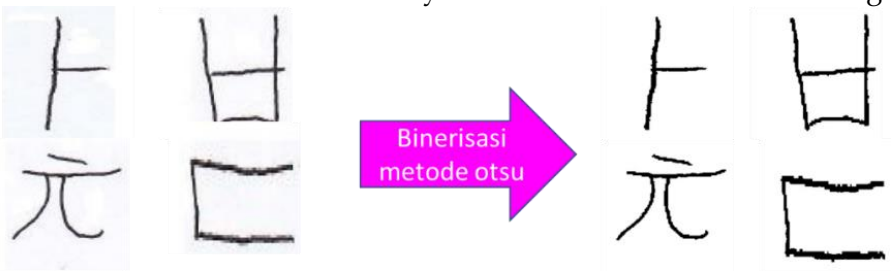
resized to 224 × 224 pixels and labeled manually according to the classification. The 1080 dataset is divided into training data and test data. The training data is subdivided into training data and validation data.

In this study, the author uses 2 preprocessing scenarios, namely resizing the original data and resizing and binaryizing data using the Otsu method. The resizing process is carried out from the original data that has different sizes, the resizing process is carried out to the same size, which is 224x224. In this study, resizing was carried out for scenario 1 and scenario 2. The illustration of resizing is shown in Figure 3



Picture 3 Resize illustration

After the data has the same size that is 224x224. In scenario 2, the data will be binary so that the data will be black and white. Binaryization illustration is shown in Figure 4



Picture 4 Binaryization Illustration

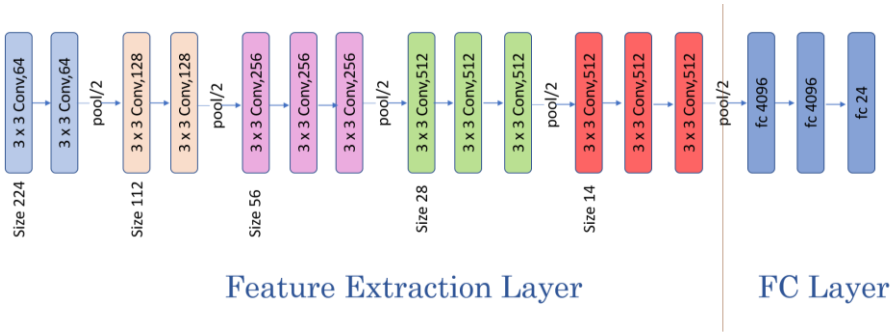
The data that has been preprocessed will be carried out in the training process using the CNN (Convolutional Neural Network) method with the VGG-16 architecture. For illustration, the model design is shown in Figure 5

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
vgg16 (Functional)           (None, 7, 7, 512)         14714688
_________________________________________________________________
flatten (Flatten)            (None, 25088)             0
_________________________________________________________________
dense (Dense)                (None, 4096)              102764544
_________________________________________________________________
dense_1 (Dense)              (None, 4096)              16781312
_________________________________________________________________
dense_2 (Dense)              (None, 24)                98328
=================================================================
Total params: 134,358,872
Trainable params: 119,644,184
Non-trainable params: 14,714,688
```

Picture 5 Illustration of VGG-16 Model

The model used is VGG-16 which is weighted from the previous training. The VGG-16 model in detail can be seen in Figure 6. This study uses 11 VGG-16 feature extraction layers and the addition of 2 full connected layers with a weight of 4096 and 1 full connected layer

with a weight of 24 as output according to the classification class using the function Softmax activation.



Picture 6. Model CNN Arsitektur VGG-16

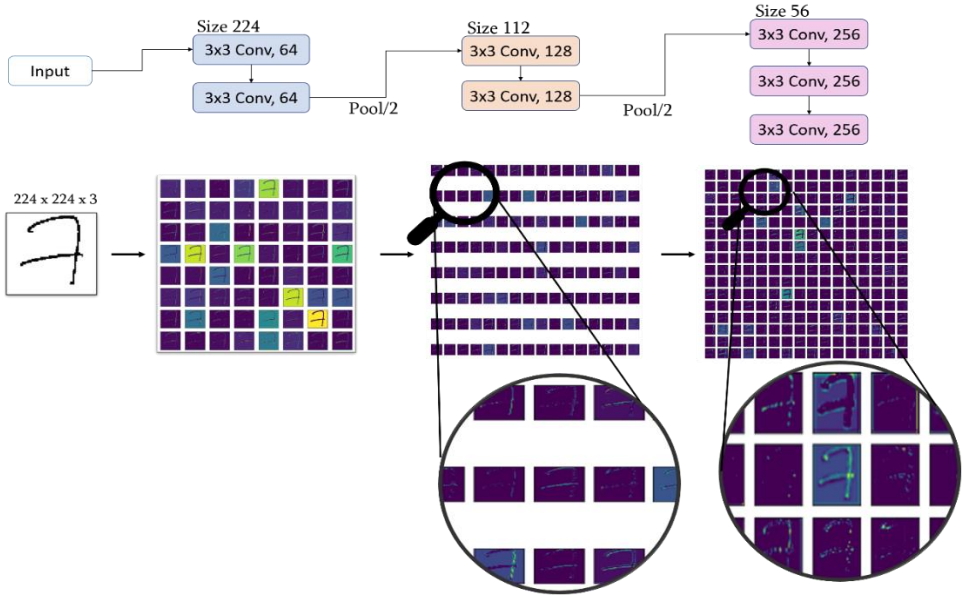


Figure 7 Results of Implementation of the CNN Architecture VGG-16 Method (1)

This research begins with the input size of 224 x 224, the first convolution is carried out using feature maps which amount to 64 channels with a kernel filter size of 3x3. Convolution aims to extract features from an image which can be seen in Figure 4.5. The size of the kernel filter on the VGG-16 architecture is always the same, namely 3x3 which differs only in the number of channels used. The first convolution produces an image with a size of 224x224x64. For the second convolution, use feature maps and the same filter size as the first layer, namely feature maps of 64 channels and a kernel filter size of 3x3 resulting in an image size of 224x224x64. After the second convolution is completed, the image is pooled in a layer where the image size will be reduced (downsampling). The pooling layer used is max pooling. Max pooling used kernel filter size 2x2 with stride 2. The result of this operation becomes 112x112x64.

After the first pooling is done, it is followed by the third and fourth convolutions, both of which use a kernel filter size of 3x3 with feature maps totaling 128 channels. The result is an image measuring 112x112x128. Performed a second pooling and resized to 56x56x128. The fifth, sixth and seventh convolutions use a kernel filter size of 3x3 and feature maps of 512 channels. Then the third pooling is done which changes the size to 28x28x512.

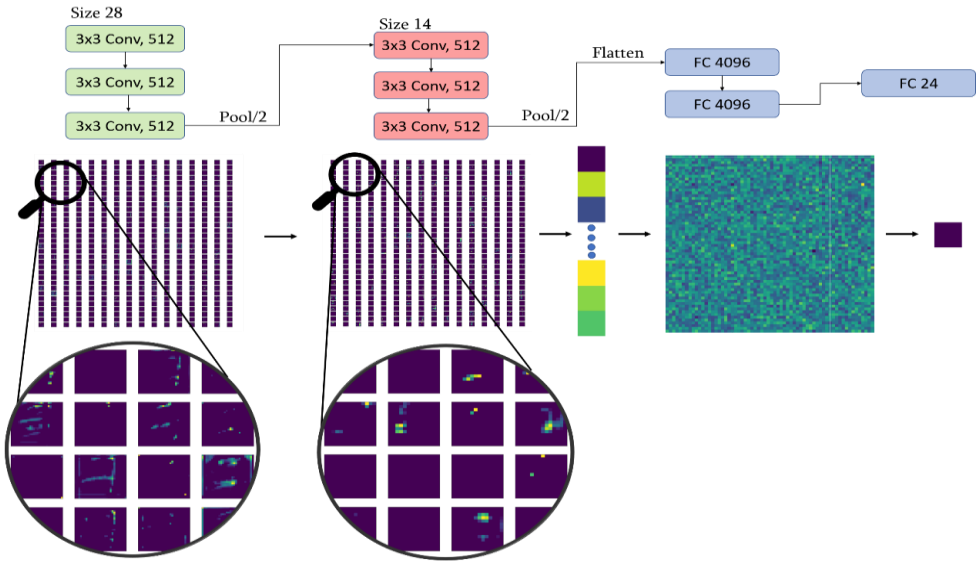


Figure 3 Results of the Implementation of the VGG-16 Architecture CNN Method (2)

The 8th to 13th convolutions use feature maps totaling 512 channels and a kernel filter size of 3x3 which is used to produce feature extraction as shown in Figure 7. For the 8th to 10th convolutions using an image size of 28x28x512 and doing pooling fourth so that the 9th to 13th convolutions use an image size of 14x14x512. After all the convolutions have been carried out, the image is reduced in size, namely the fifth pooling to become the result of feature extraction to 7x7x512. After all the feature extraction layers have been done, then the 3 layers are fully connected.

Before entering fully connected, it performs flattening for data transformation into 1-dimensional data. The first Fully connected layer has a size of 25088. This number is obtained from the last output of the feature extraction, which is 7x7x512. The first and second fully connected layers have the same number of neurons, namely 4096. In this layer, the value of 1-dimensional data (neurons) measuring 25088 will be connected to 4096 neurons. Starting from the first convolution stage to the fully connected/dense layer stage, use the relu activation function to get the value. The last fully connected or dense layer stage uses a softmax activation function of 24 which produces 24 as the output of 24 Korean fonts.

3.2. Discussion

In the discussion phase, the results of testing were carried out on 24 Korean fonts. The test uses epoch values of 30 and batchsize of 36 with 2 test scenarios, namely scenarios with resizing data and scenarios with resizing data and then binaryization. The test is carried out as a whole and uses k-fold cross validation. The results of the test will be calculated using a confusion matrix to get the values of accuracy, recall and precision.

3.2.1. Scenario Discussion 1

Scenario 1 is a scenario where the original data is resized to a size of 224 × 224 pixels. Data validation uses the K-fold cross validation method which uses a value of k = 5 so that the data becomes 5 folds and 5 iterations. Scenario 1 produces an average accuracy per class for 5 iterations and then an average of 99.52% for the whole. In addition to accuracy performance, it also produces precision of 95.56% and recall of 94.11%. For more detail described in Table 1.

Table 1 Results of Scenario 1

Letter	*Accuracy* (%)	Standard Deviation	*Precision* (%)	*Recall* (%)
A	99,07	0,93	96	83,33
B	99,91	0,21	100	98
CH	99,72	0,41	96	94,18
D	99,44	0,6	86,9	100
EO	99,07	0,65	86,76	95,14
EU	99,72	0,41	97,78	95,28
G	99,72	0,41	96,36	97,14
H	99,35	1,01	100	86
I	99,63	0,51	95	91,36
J	99,54	0,8	94,83	97,33
K	99,63	0,39	98,46	93,14
M	99,81	0,41	100	96
N	99,91	0,21	100	98,18
NG	100	0	100	100
O	99,54	0,33	96,18	92,7
P	99,17	1,37	90,67	96,9
R/L	99,91	0,21	98,33	100
S	99,91	0,21	100	97,14
T	99,54	0,46	98	92,36
U	99,35	1,45	100	86
YA	99,44	0,6	96,36	90
YEO	98,8	0,62	85	89,49
YO	99,07	0,8	90,81	88,81
YU	99,17	1,86	90	100
Average	99,52	-	95,56	94,11

Table 1 is the level of accuracy, precision and recall per class for the first scenario. Korean typeface has 24 types so that the class it has is 24 classes. The average performance of scenario 1 is 99.52% accuracy, 95.56% precision and 94.11% recall. Table 4.1 shows that the letter NG ( ) has the highest level of performance, which is 100% with a 100% precision level and a 100% recall rate.

The letters that have the lowest performance level, namely YEO letters, have 98.8% accuracy, 85% precision and 89.59% recall. The letter YEO ( ) also has a lower level of precision than the other letters. Judging from the standard deviation, the letter YEO tends to be stable at 0.62. The lowest recognition rate of 24 types of letters is owned by letter A ( ) which is 83.33%. Letter A has an accuracy level of 99.07% with a standard deviation of 0.93 and a precision level of 96%.

3.2.2. Scenario 2 Discussion

Scenario 1 is a scenario where the original data is resized to a size of 224 × 224 pixels. After resizing, the data is binaryized using the Otsu method. Data validation uses the K-fold cross validation method which uses a value of k = 5 so that the data becomes 5 fold and 5 iterations. Scenario 2 has an average for all letters in the form of an

accuracy of 99.42%. In addition to accuracy performance, it also produces 93.94% precision and 93.11% recall. For more details, see Table 2.

Table 2 Results of Scenario 2

Letter	*Accuracy* (%)	Standard Deviation	*Precision* (%)	*Recall* (%)
A	99,17	0,89	88,51	88,33
B	99,72	0,62	94	100
CH	99,54	0,8	91,11	97,14
D	99,44	0,39	93,46	92,91
EO	98,89	1,01	91,33	79,43
EU	99,91	0,21	100	97,78
G	99,63	0,51	94,29	96,36
H	98,89	1,16	100	71,71
I	99,81	0,41	95	95
J	99,26	1,07	90,24	97,33
K	99,63	0,39	93,5	98
M	98,98	2,03	90,91	88,41
N	99,81	0,41	100	96
NG	100	0	100	100
O	99,54	0,57	91,18	100
P	98,98	0,69	85,67	92,98
R/L	99,81	0,25	96,67	97,78
S	100	0	100	100
T	99,35	0,53	92,05	94,36
U	99,44	0,6	96,67	93,71
YA	99,07	1,18	94,44	82
YEO	98,61	1,64	84,93	94,35
YO	99,07	0,93	96,67	80,87
YU	99,72	0,41	93,86	100
Average	99,42	-	93,94	93,11

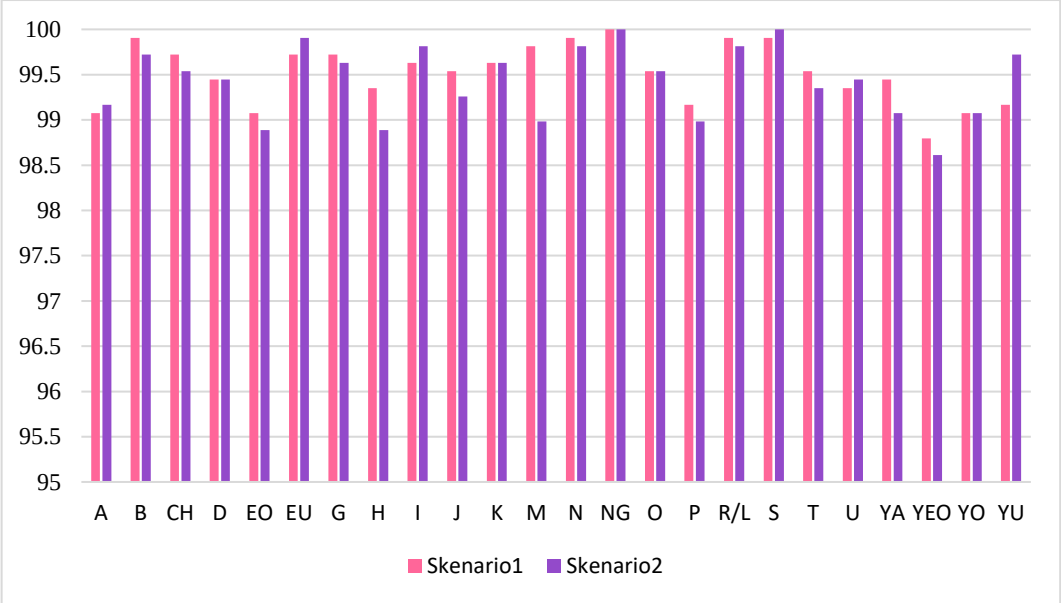
Based on Table 2, 24 types of Korean fonts have the highest level of performance owned by the letters NG () and the letter S (). The average performance of scenario 2 is 99.42% accuracy, 93.94% precision and 93.11% recall. The letters NG and S can be recognized by all letters correctly so that they have the same level of accuracy, precision and recall, which is 100%.

The Korean typeface which has the lowest level of performance is owned by the YEO letter which has a standard deviation of 1.64 resulting in an average accuracy rate of 98.61%. If you look at the precision and recall of the letter YEO, it is known that the letter YEO also has the lowest precision value among other letters, namely 84.93%. There is the letter H which has the lowest recall value, which is 71.71%. The precision level of the letter H is 100% and the accuracy is 98.89% and the standard deviation is 1.16.

3.2.3. Comparison of accuracy scenario 1 and scenario 2

The average Accuracy of Scenario 1 and Scenario 2 has a fairly low difference of 0.1. The average accuracy obtained in scenario 1 is 99.52% and the average for scenario 2

is 99.42. Figure 4.15 describes the comparison of the accuracy of each class in scenario 1 and scenario 2



Picture 8 Comparison of Accuracy Scenario 1 and Scenario 2

Figure 8 shows that the letters D, K, NG, O and YO have the same level of accuracy between scenario 1 and scenario 2. The letter M has a higher level of accuracy in scenario 1 which shows that the letter M is more recognized in scenario 1 while the letter YU is more recognized in scenario 1 scenario 2. The letter NG has high accuracy in scenario 1 and scenario 2 but the letter YEO has the lowest accuracy among other letters in scenarios 1 and 2

IV. CONCLUSION

Based on the research on handwriting recognition of Korean letters using the Convolutional Neural Network (CNN) VGG-16 Architecture method that has been carried out, the following conclusions are obtained:

1. The results of scenario 1 performance with an average accuracy of all letters of 99.52%. Precision performance is 95.56% and recall performance is 94.11%. The best known letters are NG letters with 100% accuracy, precision, and recall performance. The lowest letter to be recognized is YEO with 98.8% accuracy, 85% precision, 89.49% recall.
2. The performance results of scenario 2 with an average accuracy of all letters of 99.42%. Precision performance is 93.94% and recall performance is 93.11%. The best recognized letters are the letters NG and S with 100% accuracy, precision, and recall performance. The lowest letter to be recognized is the letter YEO with 98.61% accuracy, 84.93% precision, 94.35% recall.

REFERENCES

[1] Rizka Aulia Ramadhani, "Perilaku konsumsi dan produksi budaya penggemar pada komunitas army BTS jogja," *Univ. Islam Indones.*, 2019.

[2] K. P. Sari, I. I. Tritoasmoro, and E. Susatio, "Pengalih citra karakter korea-indonesia menggunakan klasifikasi k-nearest neighbour dan template matching," *Fak. Tek. Elektro*, 2012.

[3] I. J. Kim and X. Xie, "Handwritten Hangul recognition using deep convolutional neural networks," *Int. J. Doc. Anal. Recognit.*, vol. 18, no. 1, pp. 1–13, 2014.

[4] M. Velentina Y.C, E. Erin, R. Angreni, and D. Alamsyah, "Penerapan algoritma hough transform untuk pengenalan goresan tangan huruf korea," *Stmik Gi Mdp*, no. x, pp. 1–10, 2016.
[5] G. U. S. Lumanauw, "Pengenalan kata tulisan tangan huruf korea menggunakan deep convolutional neural network berbasis android," *J. Phys. Conf. Ser.*, vol. 978, no. 1, p. 46, 2018.
[6] J. Sandywan and C. A. Chandra, "Pengenalan tulisan tangan menggunakan metode hidden markov model," *Stmik Gi Mdp*, no. x, pp. 1–8, 2019.
[7] I. W. Suartika, A. Y. Wijaya, and R. Soelaiman, "Klasifikasi citra menggunakan convolutional neural network (CNN) pada caltech 101," *J. Tek. ITS*, vol. 5, no. 1, 2016.
[8] S. C. A. Pradhana, "Pengenalan aksara jawa dengan menggunakan metode convolutional neural network," *e-Proceeding Eng.*, vol. 7, no. 1, pp. 2558–2567, 2020.
[9] A. T. Christopher Albert Lorentius, Rudy Adipranata, "Pengenalan aksara jawa dengan menggunakan metode convolutional neural network," *e-Proceeding Eng.*, vol. 7, no. 1, pp. 2558–2567, 2019.
[10] M. A. Pragathi, K. Priyadarshini, S. Saveetha, A. S. Banu, and K. O. Mohammed Aarif, "Handwritten tamil character recognition using deep learning," *Proc. - Int. Conf. Vis. Towar. Emerg. Trends Commun. Networking, ViTECoN 2019*, pp. 1–5, 2019.
[11] D. Maharani, "Pengenalan karakter Korea (hangeul) tulisan tangan menggunakan zone centroid and zone (ZCZ) dan backpropagation," *Fak. Ilmu Komput. Univ. Sriwij.*, 2016.
[12] A. N. Septiawan and T. Setiadi, "Aplikasi pengenalan huruf hangeul berbasis multimedia interaktif," *J. Sarj. Tek. Inform.*, vol. 1, no. 1, pp. 347–357, 2013.
[13] A. K. Hwa, C. H. Yong, R. N. Adinda, S. Agung, and F. Hutagualung, "Buku korea terpadu untuk orang indonesia." Seoul, 2013.
[14] M. M. Taslim, K. Gunadi, and A. N. Tjondrowiguno, "Deteksi rumus matematika pada halaman dokumen digital dengan metode convolutional neural network," *Univ. Kristen Petra*, vol. Vol 7, 2019.
[15] Y. Achmad, R. C. Wihandika, and C. Dewi, "Klasifikasi emosi berdasarkan ciri wajah wenggunakan convolutional neural network," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 11, pp. 10595–10604, 2019.