Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 8, No. 2, May, Pp. 545-554

Comparison of CNN's architecture GoogleNet, AlexNet, VGG-16, Lenet -5, Resnet-50 in Arabic handwriting pattern recognition

Gibran Satya Nugraha*[1], Muhammad Ilham Darmawan[1], Ramaditia Dwiyansaputra[1]
Department of Informatics Engineering, Faculty of Engineering, Universitas Mataram, Mataram, Indonesia [1]

Article Info

Keywords:
Arabic, Pattern Recognition, Handwriting, CNN, Deep Learning

Article history:
Received: January 18, 2023
Accepted: April 01, 2023
Published: May 31, 2023

Cite:
G. S. Nugraha, M. I. Darmawan, and R. Dwiyansaputra, "Comparison of CNN's Architecture GoogleNet, AlexNet, VGG-16, Lenet -5, Resnet-50 in Arabic Handwriting Pattern Recognition", KINETIK, vol. 8, no. 2, May 2023.
https://doi.org/10.22219/kinetik.v8i2.1667

*Corresponding author.
Gibran Satya Nugraha
E-mail address:
gibransn@unram.ac.id

Abstract

The Arabic script is written from right to left and consists of 28 characters, with no capital or lowercase letters. The Arabic script has several orthographic and morphological properties that make handwriting recognition of the Arabic script challenging. In addition, one of the biggest challenges in recognizing Arabic script patterns is the different handwriting styles and characters of each person's writing. The authors propose a study to compare the accuracy of handwriting pattern recognition in Arabic script which has been done previously by comparing five CNN architectures, namely GoogleNet, AlexNet, VGG-16, LeNet-5, and ResNet-50. Considering that previous research has not obtained excellent accuracy. The number of datasets used is 8400 image data and the most optimal comparison of testing and training data is 80:20. Based on the research that has been done, there are several things that the author can conclude. The model is made using 64 filters for each convolution layer because the optimal size is used for 5 architectures, kernel size is 3x3, neurons is 128, dropout weight is 50% to reduce overfitting, learning rate is 0.001, image size is 64x64, the normalization method with the ReLU activation function, and 1-dimensional input image (grayscale), and with a comparison of testing and training data of 80:20. The VGG-16 architectural model is the architecture that gets the highest score, namely 83.99%. This can have good potential to be developed as a medium for learning Arabic script.

1. Introduction

Arabic and the Qur'an are a unity that cannot be separated from one another. Learning the Qur'an, Arabic is an absolute requirement that must be mastered. Arabic is one of the most widely used languages in the world; because many people use it, Arabic has become an international language and is recognized by the world. So, it is not excessive if learning Arabic needs to get emphasis and attention starting from the elementary level (Elementary School) to higher education institutions, both public, private, and general, where religion is taught and developed according to the abilities and development of students. However, it is not easy matter to be able to understand Arabic because it is not the language of native speakers that is commonly used [1].

Arabic script is a language that is contained in the holy book of Islam and Indonesia is a country with the most people who are Muslims. The Arabic script is written from right to left and consists of 28 characters, with no capital or lowercase letters. Dots play an essential role in Arabic characters. Certain characters' forms and dot counts are similar, such as. (ب, ت, ث), which can occur above or below characters with different pronunciations. Arabic characters consist of 28 main characters, and the forms of some characters are similar but can be different with the number and position of dots [2]. Therefore, learning Arabic letters can be done by recognizing Arabic handwriting patterns, one of which is by utilizing machine learning algorithms which will then be used in learning applications such as Android, IOS, or other learning media [3].

Artificial Neural Network (ANN) is a machine learning algorithm that can recognize the shape and pattern of handwriting [4]. ANN is widely applied to solve problems regarding pattern recognition, voice recognition, character recognition for document reading, signal recognition, determining nutritional patterns, image processing and other issues. This type of ANN model, which consists of many layers, is called a Multi-Layer Perceptron (MLP) which fully functions as a link between the neural networks [5]. The ability of this MLP can accurately classify data that has been previously studied, and the amount of data that is classified is not too large. When the input is an image, however, the classification method employing MLP has a flaw. Images to be classified must be pre-processed, segmented, and features from images must be extracted to obtain optimal performance. Another development of MLP that can overcome this problem is the Convolutional Neural Network (CNN) [6].

Convolutional Neural Network (CNN) is a deep learning technique that can find and identify objects in a variety of digital images datasets with different patterns and shape. CNN is widely used to solve several problems in image classification, object detection, object localization, and image segmentation. Deep learning is an implementation of the basic concepts of machine learning that implements ANN algorithms with more layers. With the number of hidden layers between the input layer and the output layer, this network can be said to be a deep neural net. In the last few years, CNN has shown extraordinary performance. This is primarily due to more decisive computational factors, large datasets, and techniques to train deeper networks. CNN capability is claimed to be the best model for solving object detection and object recognition problems [7]. CNN is also used to recognize handwriting patterns, one of which is Arabic script handwriting. The Arabic script has several orthographic and morphological properties that make handwriting recognition of the Arabic script challenging. In addition, one of the biggest challenges in recognizing Arabic script patterns is the different handwriting styles and characters of each person's writing [8].

Based on the explanation above, previous studies that have been carried out by [9], using the basic CNN architecture and the same dataset, obtained accuracy, precision, and recall that were not very good, namely 78.10%, 79.68%, 77.82% respectively. Therefore, this research will compare the basic CNN architectural development methods, namely GoogleNet, AlexNet, VGG-16, LeNet-5, and ResNet-50. These architectures get good results for studies on handwriting pattern recognition. The number of datasets used is 8400 image data and the most optimal comparison of testing and training data is 80:20 because the amount of data is vast and widely used in research [10]. From previous research, the five improvised CNN architectures should get better accuracy.

2. Related Works

Several previous studies have implemented CNN architectures such as VGG-16, ResNet-50, LeNet5, GoogleNet, and Alexnet to recognize patterns from an image. Research on Arabic handwriting was conducted by Al Jarrah, *et al.* [11] this study proposes the CNN model for Arabic Handwriting Character Recognition. The model has been trained on a large image dataset called AHCD (Arabic Handwriting Character Dataset) which contains approximately 16,800 Arabic handwritten characters in various styles and variations. The work results show excellent recognition performance on the test dataset compared to other models - where the proposed model achieves an accuracy of 97.2%. The model's performance has been enhanced by the addition of new data, which has raised accuracy by 97.7%. Nayef, *et al*. [12] suggested a leaky ReLU optimized to retain more negative vectors utilizing a CNN architecture with a batch normalization layer. The National Institute of Standards and Modification Technology (MNIST), the AHCD, self-collected data, and the AlexU Isolated Alphabet were the four data sets utilized to assess the suggested technique (AIA9K). The proposed method performs noticeably better in terms of accuracy, precision, and recall action when compared to state-of-the-art techniques. The outcomes demonstrate notable advancements over the known leaked ReLUs: 90% for the HIJJA dataset, 99% for MNIST Digits, 99% for AHCD, and 95.4% for self-collected data. Between the training, validation, and testing phases, the proposed leak optimization, and the suggested CNN design ReLU performs consistently in terms of accuracy and error rate. This suggests that most of the samples received acceptable training and classification.

A new dataset of Arabic letters named Hijja, created only by youngsters between the ages of 7 and 12, was presented in another study. 47,434 characters total, written by 591 participants, make up the data set. Additionally, they suggest a convolutional neural network-based approach for automatic handwriting detection. They used the AHCD data set and Hijja to train the model. The results demonstrate the model's promising performance, surpassing previous models in the literature, and obtaining accuracy of 97% and 88% on the AHCD data set and Hijja data set, respectively. [13]. A neural technique is suggested by Moumen, *et al*. [14] to address challenges with Arabic real-time scene text identification. There are two processes in the detection of Arabic handwriting. A version of VGG-16 called TextBlockDetector FCN is used in the first stage to eliminate non-textual elements, localize wide-scale text, and offer text-scale estimation. To achieve the best performance, the second phase determines a limited set of text scales. The VGG-16 approach was used to evaluate the system. A manually processed dataset of 575 photos is used for training and testing, and data augmentation is used to enhance the training process. 71% of the findings for recollection were obtained.

The convolutional neural network structure proposed by Yanmei, *et al*. [15] greatly decreases the scaling of the network's training parameters. To strengthen the LeNet-5 network, it substitutes the global average pooling technique for the complete connection algorithm. While the number of subsampling layers is decreased to the ideal number, the number of convolution kernels is raised. The results demonstrate that the training parameters of the modified network are only 34.8% of the original, and the recognition accuracy may reach 99.3% after verification with the MINST handwritten Arabic numerals data set. Rasheed, *et al*. [16] use modified Convolution Neural Networks and transfer learning theory to offer a classification framework for the automatic recognition of handwritten Urdu letters and numerals with greater recognition accuracy. A trained AlexNet CNN model with an SVM classifier and a finely tuned AlexNet for feature extraction and classification are used to evaluate transfer learning performance. The author enhanced the data to prevent over-fitting and improved the AlexNet hyperparameters to get higher accuracy. The usefulness of the suggested research for handwritten letter and number recognition using tuned AlexNet is demonstrated by experimental results and quantitative comparisons. For Urdu characters, digits, and hybrid data sets, the proposed study based on AlexNet outperformed comparable advanced research, obtaining classification accuracy of 97.08%, 98.21%, and 94.92%, respectively. The techniques described can be used for study on Urdu characters and in many different fields, including handwritten text capture, postal address reading, bank check processing, and preserving and scanning old manuscripts.

Almisreb, *et al*. [17] assess and validate the usage of deep transfer learning models. Hench, the best model for categorizing handwritten photographs written by native or non-native, author is determined using seven different deep learning transfer models, including AlexNet, GoogleNet, ResNet18, ResNet50, ResNet101, VGG16, and VGG19. A recently created deep learning model was evaluated and validated using two data sets made up of images of Arabic handwriting. Each model's output was then classified as being written by foreign (non-native) authors or native authors. Using the original and supplemental datasets, the training set and validation were carried out. The outcomes demonstrated that the GoogleNet deep learning model for both regular and augmented data sets had the highest accuracy. In classifying original handwriting, the best accuracy was attained at 93.2% using average data and 95.5% using augmented data.

Ibrahim, *et al*. [18] suggest a brand-new textural characteristic for text-based authorship identification. Based on the orientation histogram for various text angles, to extract important characteristics, the feature histogram of the oriented gradient (HOG) is modified. Convolutional neural networks (CNN) and these features combined yield a fantastic vector of sophisticated features. Then, they use a genetic algorithm to decrease the components by picking the best candidates. The characteristics are normalized using the normalization procedure before being added to the neural network classifier. The experimental results demonstrate that as the amount of data rises, the proposed augmenter produces better results for the proposed model and the HOG and ResNet50 features. Thanks to the enormous amount of data, the system can learn a lot about the nature of writing patterns. The proposed model's real results were obtained using a variety of models, and they were then compared to CNN and ResNet50 for all paragraphs, lines, and subwords. Whole sections get the best results across all models since they contain abundant information and models can use a range of attributes for different words. For all paragraphs with augmentation, the accuracy rates for the HOG and CNN features are 94.2%, 83.2% for lines, and 78% for sub-words. This work gives a system that can identify writers based on their handwriting and creates a solid model that can allow author identification based on sentences, words, and sub-words.

From the research that has been discussed previously, it shows that several CNN architectures such as GoogleNet, AlexNet, VGG-16, LeNet-5, ResNet-50 provide different results in conducting classification. Research on the comparison of these architectures is essential to do to find out the most appropriate architecture to be implemented in the Arabic handwriting dataset that the author uses.

3. Research Method

3.1 Data training and data testing

The data used for this study came from the handwriting of the Arabic script, which amounted to 8400 Arabic script letters, of which there were 28 Arabic letters, of which there were still 6720 letters for training data and 1680 for test data. The 6720-training data have details of each Arabic letter having 240 images. Meanwhile, for the test data, each Arabic letter has 60 images.

3.2 Pre-processing

Pre-processing is done to make images from existing datasets so that they are easier to process during the training process. In the first pre-processing, the existing image will be equalized in size first by resizing the image size. The sizes used in this study is 64 x 64. After resizing the image, the color dimension will be changed to a grayscale image. This color dimension change is intended so that the image has only 1 color dimension so that it can be continued when processing on the convolution layer.

3.3 CNN Model

Classification with the convolution neural network method is carried out by conducting training first according to the architecture of the CNN in Figure 1. The input for this model is in the form of training data images with size 64 x 64 which have been grayscale. The input image will enter the feature learning stage first which consists of a convolution layer and a pooling layer. Furthermore, the stage that is carried out before introducing the Arabic script is the classification stage which consists of a fully connected layer with dropouts. In this study, we will try to use a research scenario, namely 80% of the dataset for training and 20% for testing.

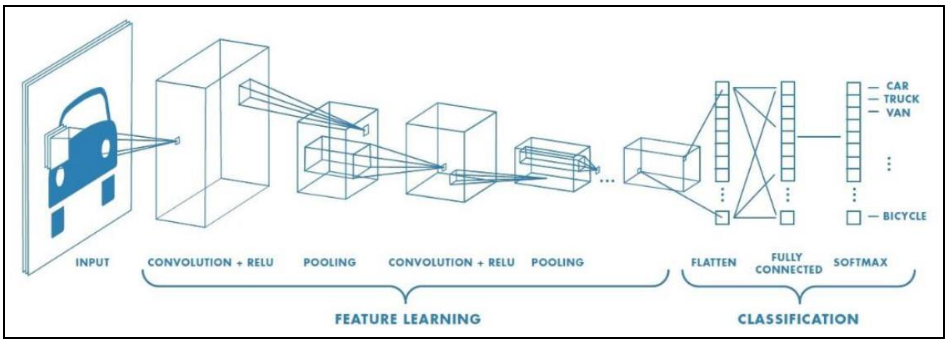

Figure 1. CNN Architecture [19]

a. Image acquisition

The inputted image will be represented in a 2-dimensional image matrix so that the image can enter the feature learning stage. Figure 2 is an example of writing Arabic letters in black and white format where the pixel values are represented in matrix form.

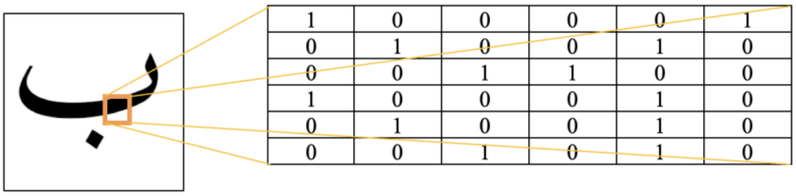

Figure 2. Example of Writing Arabic Letters (Matrix Format)

b. Convolution process

In the convolution layer, a convolution process is carried out which aims to filter the input matrix. Then so that all matrices can be convolved, and the size is fixed, zero padding is done. Convolution is carried out using matrices of size 3 x 3 and 5 x 5. The output from this convolution layer will be input to the pooling layer.

The results of calculations on the convolution layer with negative values will be carried out with additional calculations to remove negative values in the image matrix. In this study the calculation approach that will be used is the ReLU activation function approach. The ReLU activation function will change matrix values with negative values to 0. The ReLU activation function has been widely used in previous studies using the convolution neural network method and got good performance. Figure 3 illustrates a convolution process example.

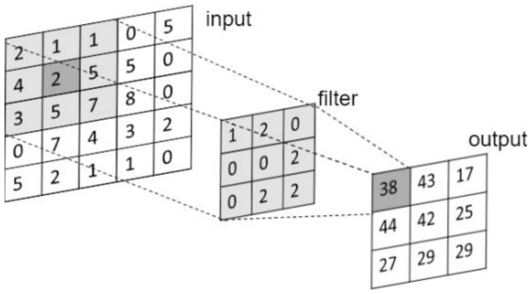

Figure 3. Convolution Process [20]

c. Max pooling

In max pooling, the process is carried out to reduce the size of the image to make the feature map process faster. This study uses pooling with a 2x2 matrix with a stride of 2. So, pooling will shift by 2 indices and look for the most significant value from the pooling or can be referred to as max pooling

d. Flatten

The results of the convolution process and max pooling from the previous process will be flattened. Flatten is a process for converting a matrix into a vector. The faltten process is carried out to adjust the input format to match the input format on the neural network. The output from the faltten will then enter the fully connected layer for classification.

e. Fully Connected Layer with dropout

This study will use three different numbers of hidden layers, namely one hidden layer, two hidden layers and three hidden layers to determine the effect of the number of hidden layers in the study. Each hidden layer has 64 neurons. To reduce the possibility of overfitting, dropouts will be applied to the model (overfitting is a process where training performance is better than during testing). This study will apply a dropout with a weight of 20%, which means that it will randomly select neurons in each layer as much as 20% to be deactivated

f. Parameters and architecture

The parameters used for each architecture, especially the kernel size and stride, can be seen in Table 1. While for each method, the architecture can be seen in Figure 5, Figure 6, Figure 7, Figure 8, and Figure 4

Table 1. Parameter of Each Architecture

No	Arsitektur	Size and Feature	Kernel Size	Stride
1	*GoogleNet*	112x112x64	3x3	2
2	*AlexNet*	55x55x96	11x11	4
3	*VGG-16*	112x112x64	3x3	2
4	*LeNet-5*	28x28x6	5x5	1
5	*ResNet-50*	112x112x64	7x7	2

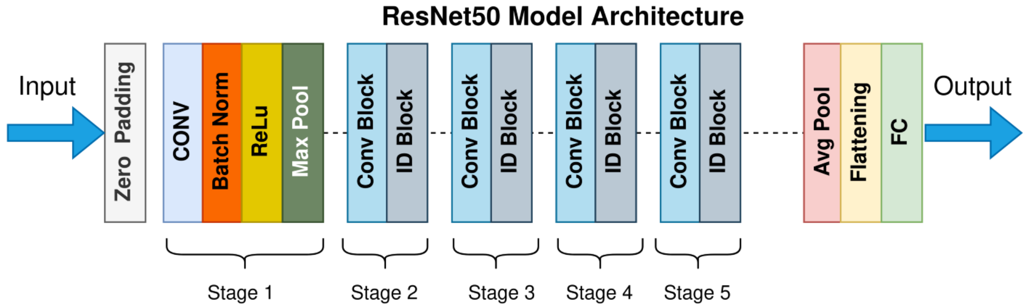

Figure 4. ResNet50 Architecture [21]

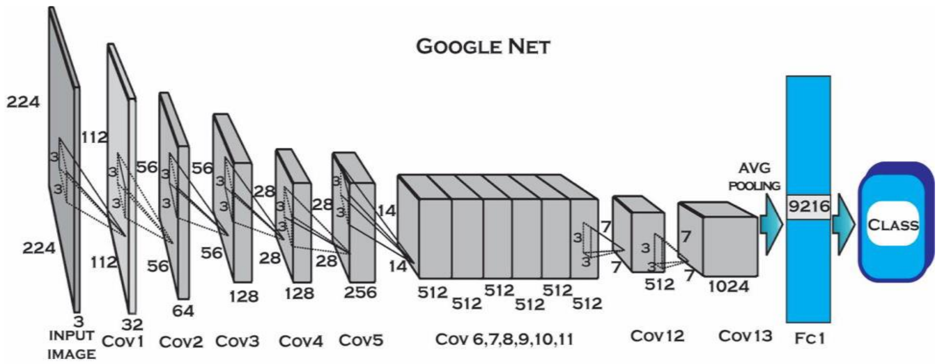

Figure 5. GoogleNet Architecture [22]

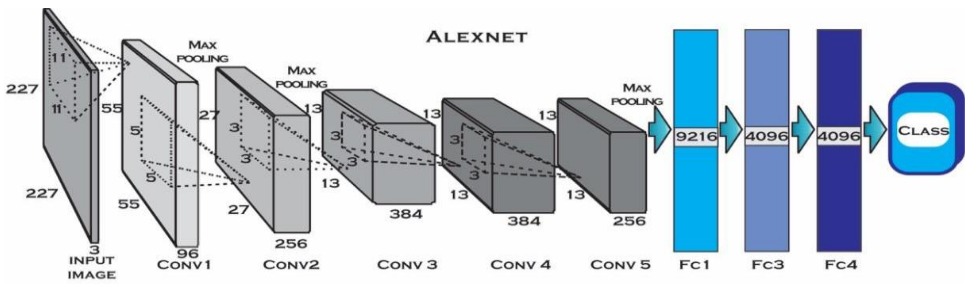

Figure 6. AlexNet Architecture [23]

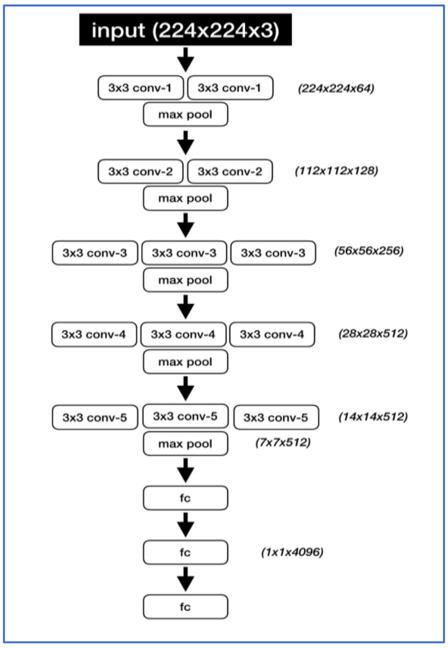

Figure 7. VGG-16 Architecture [24]

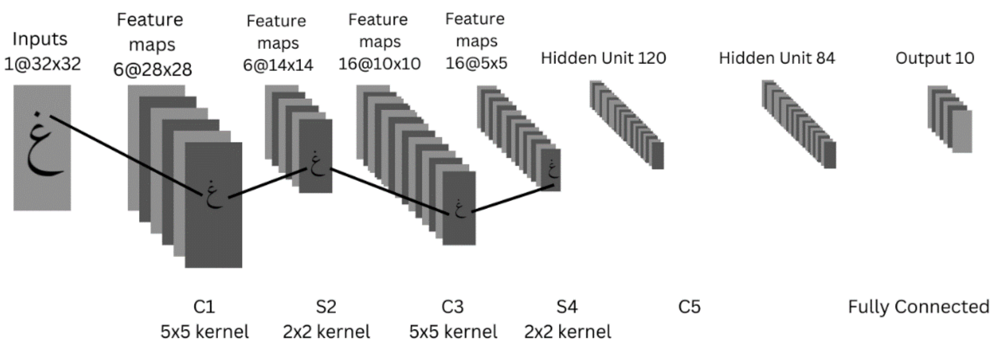

Figure 8. LeNet-5 Architecture [25]

4. Result and Discussion

In this study, a machine learning model is created using sources from previous research. The sampled Arabic script image dataset is from the age of 5 to 20 years and over; both those who have studied and those who have not studied Arabic script are written using a black marker so that the collected dataset is 8400 image data, the dataset used is a private dataset. The dataset can be seen in Figure 9.

Comparative research was carried out with several trials to get the best value. The first step is to compare the number of datasets to perform architectural comparisons. Researchers take the dataset with as many as 8400 data images. Next, the architectural model was tested using the previously mentioned datasets. This test is carried out with various parameters with the following test sequence.

1. Using each architectural method whose parameters have been modified
2. Separation of training and testing data is done automatically (the library does separation) and manually (separation by the author)

The best architectural model was obtained from the testing results with the previous research dataset totalling 8400 to compare it to the five architectures and look for the best accuracy value. Tests were carried out for comparison of 5 architectures to find values for accuracy, precision, and recall using an image size of 64 x 64. The author modified the size of the convolution layer to be different from that in Table 1. Changes were made because when using parameter configurations as in Table 1 the authors experienced problems or errors during training. For example, when using the convolution layer size as shown in Table 1, there is no learning where the accuracy from epoch 1 to 50 obtains a stagnant accuracy of 3%. Because the type of image is uniform, so it does not require a variety of convolution layer sizes. The comparison results of each architecture produce the best accuracy values with the VGG-16 architecture, namely 83%, 85% precision and 83% recall. The following (Table 2) are the results obtained from each architecture

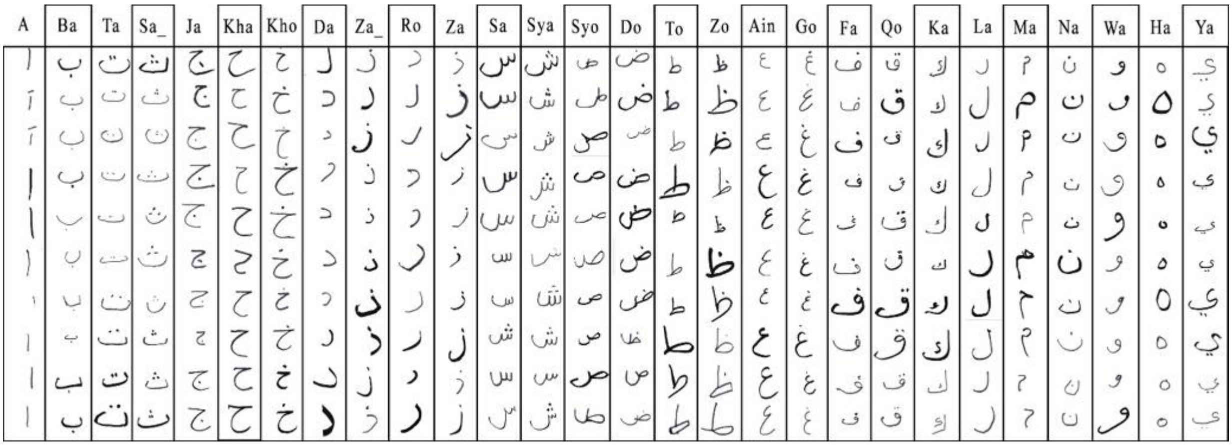

Figure 9. Arabic Handwriting Dataset

Table 2. Comparison of Each Architecture Based on Accuracy, Precision, and Recall

No	Architecture	*Accuracy*	*Precision*	*Recall*
1	*AlexNet*	0.8107	0.8270	0.8024
2	*GoogleNet*	0.7780	0.7858	0.7750
3	*LeNet-5*	0.6435	0.8489	0.6381
4	*ResNet-50*	0.7238	0.7432	0.7012
5	*VGG-16*	0.8399	0.8551	0.8327

Figure 10 shows a graphical display of the results of the VGG-16 architectural model which received the highest score of the 5 architectures with an accuracy of 83%, a precision of 85% and a recall of 83% at the 50th epoch. The learning carried out by VGG-16 is relatively fast in obtaining optimal accuracy, namely in the 22nd epoch. This fast learning is due to the compact VGG-16 architecture because it only has two main parts, namely feedforward and backpropagation. The VGG-16 architecture uses 3x3 filters uniformly which makes VGG-16 produce the best accuracy. The filter size can overcome the problem of datasets that have varying sizes.

In **Error! Reference source not found.** is a graphic display of the results of the LeNet-5 architectural model which obtains values with an accuracy of 64%, a precision of 64% and a recall of 63% at the 50th epoch. LeNet-5 obtains accuracy that is not as good as AlexNet and GoogleNet because LeNet-5 has a sensitive weakness for datasets that vary in terms of differences in image size. In this study, the original size of the image varies from 113x113 to 1200x1200.

In **Error! Reference source not found.** is a graphical display of the results of the AlexNet architectural model which gets the 2nd highest score with an accuracy of 81%, a precision of 82% and a recall of 80% in the 50th epoch. The AlexNet architecture is relatively slow in obtaining the highest accuracy in the 40th epoch. The slowness of this architecture in obtaining the highest accuracy is due to the parameters used in AlexNet, as many as 61 million (details can be seen in Table 1 and Figure 6). Following the characteristics of AlexNet, which will be optimal when classifying objects in very many classes.

In **Error! Reference source not found.** is a graphical display of the results of the GoogleNet architectural model which obtains values with an accuracy of 77%, a precision of 78% and a recall of 77% in the 50th epoch. GoogleNet is an architectural development of AlexNet so that the accuracy obtained is not too different from AlexNet. However, the fundamental difference between the two architectures is that GoogleNet has many convolution layers. If AlexNet stops at five layers and becomes one with fully connected, then GoogleNet goes through the inception phase first.

Figure 14 shows a graphical display of the results of the ResNet-50 architectural model which received the highest score of the 5 architectures with an accuracy of 72%, a precision of 74% and a recall of 70% at the 50th epoch. ResNet-50 consists of 5 stages of the convolution process which is then continued by average pooling and ends with a fully connected layer as a prediction layer. One stage of the convolution process consists of convolution, normalization, activation, and max pooling. This large number of processes requires many epochs to achieve optimal accuracy.

To see the extent to which the contribution of the research results that have been carried out is complete. We compared the results of this study with other studies on Arabic handwriting recognition, with different datasets and methods. The table comparison can be seen in Table 3. From Table 3, we can see that even though the number of datasets used is small and the variation is high. but the VGG 16, LeNet-5, AlexNet, GoogleNet, ResNet50 algorithms are able to provide quite good results. Threfore from the personal dataset that we use, additional data and augmentation will be carried out so that the use of this method can be optimal.

Table 3. Result Comparison with Another Research

Researcher's name	Year	*Dataset (amount)*	*Algorithm*	*Results (accuracy)*
Al Jarrah, *et al*	*2021*	AHCD (16800)	CNN Base	97.2%
Nayef, *et al*.	*2022*	HIJJA, MNIST, AHCD	CNN Base with modification	90%, 99%, 95.4%
Altwaijry, et al	*2023*	AHCD, HIJJA (47434)	CNN Base with modification	97%, 88%
Yanmei, *et al*	*2020*	Private dataset	LeNet-5	99.3%
Rasheed, et al	*2022*	Urdu	AlexNet	98.2%
Almisreb, *et al*.	*2022*	Private dataset	GoogleNet	95.5%
Ibrahim, *et al*.	*2021*	Private dataset	HOG and CNN	94.2%
Gibran, *et al.*	***2023***	**Private dataset (8400)**	**VGG 16, LeNet-5, AlexNet, GoogleNet, ResNet50**	**83%, 65%, 81%, 77%, 72%**

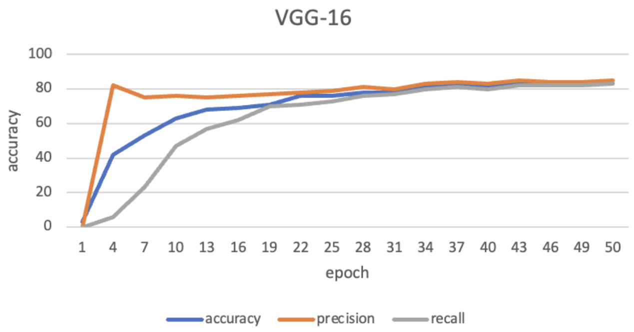

Figure 10. Graph of Accuracy, Precision, and Recall of VGG-16

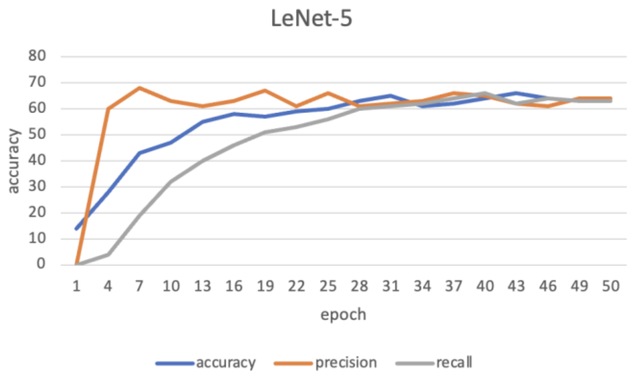

Figure 11. Graph of Accuracy, Precision, and Recall of LeNet-5

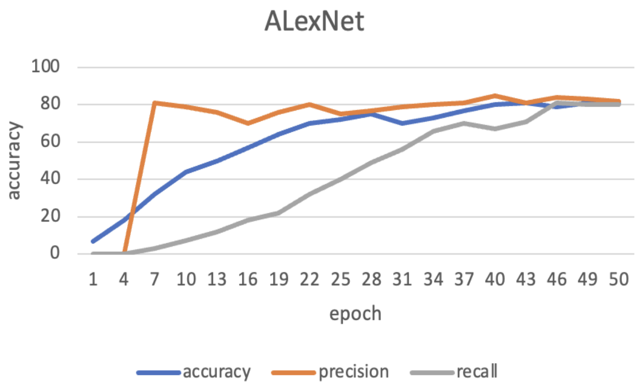

Figure 12. Graph of Accuracy, Precision, and Recall of ALexNet

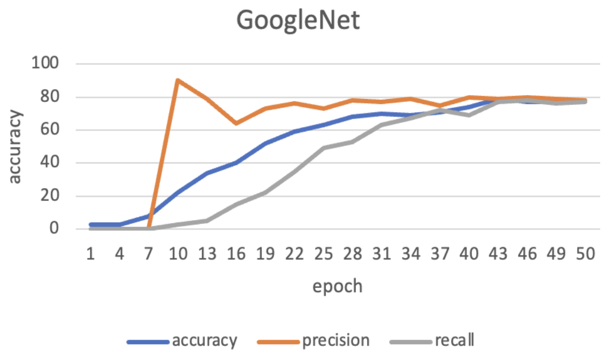

Figure 13. Graph of Accuracy, Precision, and Recall of GoogleNet

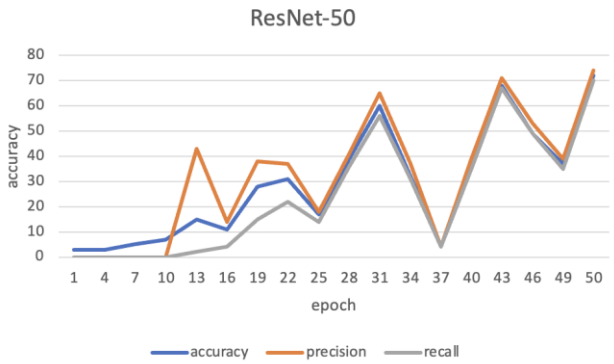

Figure 14. Graph of Accuracy, Precision, and Recall of ResNet50

5. Conclusion

On the basis of the completed research, there are several things that the author can conclude. The model is made using 64 filters for each convolution layer because the optimal size is used for 5 architectures, kernel size is 3x3, neurons is 128, dropout weight is 50% to reduce overfitting, learning rate is 0.001, image size is 64x64, the normalization method with the ReLU activation function, and 1-dimensional input image (grayscale), and with a comparison of testing and training data of 80:20. The VGG-16 architectural model is the architecture that gets the highest score, namely 83.99%. This can have good potential to be developed as a medium for learning Arabic script

References

[1] M. Athoillah and R. K. Putri, "Handwritten arabic numeral character recognition using multi kernel support vector machine," *KINETIK: Game technology, information system, computer network, computing, electronics, and control*, pp. 99–106, 2019, Accessed: Jan. 15, 2023. https://doi.org/10.22219/kinetik.v4i2.724

[2] R. Ahmed *et al.*, "Offline arabic handwriting recognition using deep machine learning: A review of recent advances," in *International conference on brain inspired cognitive systems*, Springer, 2020, pp. 457–468. Accessed: Jan. 15, 2023. https://doi.org/10.1007/978-3-030-39431-8_44

[3] S. Dargan, M. Kumar, M. R. Ayyagari, and G. Kumar, "A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning," *Archives of Computational Methods in Engineering*, vol. 27, no. 4, pp. 1071–1092, Sep. 2020. https://doi.org/10.1007/s11831-019-09344-w

[4] A. F. Hidayatullah, S. Cahyaningtyas, and R. D. Pamungkas, "Attention-based cnn-bilstm for dialect identification on javanese text," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, pp. 317–324, 2020, Accessed: Jan. 15, 2023. https://doi.org/10.22219/kinetik.v5i4.1121

[5] W. Wang and Y. Yang, "Development of convolutional neural network and its application in image classification: a survey," *Optical Engineering*, vol. 58, no. 04, p. 1, Apr. 2019. https://doi.org/10.1117/1.OE.58.4.040901

[6] D. Sutaji and H. Rosyid, "Convolutional Neural Network (CNN) Models for Crop Diseases Classification," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, vol. 7, no. 2, May 2022. https://doi.org/10.22219/kinetik.v7i2.1443

[7] W. Setiawan, A. Ghofur, F. Hastarita Rachman, and R. Rulaningtyas, "Deep Convolutional Neural Network AlexNet and Squeezenet for Maize Leaf Diseases Image Classification," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, vol. 6, no. 4, Nov. 2021. https://doi.org/10.22219/kinetik.v6i4.1335

[8] M. A. Mostafa, M. Al-Qurishi, and H. I. Mathkour, "Towards personality classification through Arabic handwriting analysis," in *The International Research & Innovation Forum*, Springer, 2019, pp. 557–565. Accessed: Jan. 15, 2023. https://doi.org/10.1007/978-3-030-30809-4_51

[9] 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, Accessed: Jan. 15, 2023. https://doi.org/10.29303/jtika.v3i1.136

[10] R. Ahmed *et al.*, "Novel deep convolutional neural network-based contextual recognition of Arabic handwritten scripts," *Entropy*, vol. 23, no. 3, p. 340, 2021, Accessed: Jan. 15, 2023. https://doi.org/10.3390/e23030340

[11] M. N. AlJarrah, M. Z. Mo'ath, and R. Duwairi, "Arabic handwritten characters recognition using convolutional neural network," in *2021 12th International Conference on Information and Communication Systems (ICICS)*, IEEE, 2021, pp. 182–188. Accessed: Jan. 15, 2023. https://doi.org/10.1109/ICICS52457.2021.9464596

[12] B. H. Nayef, S. N. H. S. Abdullah, R. Sulaiman, and Z. A. A. Alyasseri, "Optimized leaky ReLU for handwritten Arabic character recognition using convolution neural networks," *Multimed Tools Appl*, vol. 81, no. 2, pp. 2065–2094, 2022, Accessed: Jan. 15, 2023. https://doi.org/10.1007/s11042-021-11593-6

[13] N. Altwaijry and I. Al-Turaiki, "Arabic handwriting recognition system using convolutional neural network," *Neural Comput Appl*, vol. 33, no. 7, pp. 2249–2261, 2021, Accessed: Jan. 14, 2023. https://doi.org/10.1007/s00521-020-05070-8

[14] R. Moumen, R. Chiheb, and R. Faizi, "Real-time Arabic scene text detection using fully convolutional neural networks," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 2, p. 1634, Apr. 2021. http://doi.org/10.11591/ijece.v11i2.pp1634-1640

[15] H. Yanmei, W. Bo, and Z. Zhaomin, "An improved LeNet-5 model for Image Recognition," in *Proceedings of the 2020 4th International Conference on Electronic Information Technology and Computer Engineering*, New York, NY, USA: ACM, Nov. 2020, pp. 444–448. https://doi.org/10.1145/3443467.3443797

[16] A. Rasheed, N. Ali, B. Zafar, A. Shabbir, M. Sajid, and M. T. Mahmood, "Handwritten Urdu Characters and Digits Recognition Using Transfer Learning and Augmentation With AlexNet," *IEEE Access*, vol. 10, pp. 102629–102645, 2022. https://doi.org/10.1109/ACCESS.2022.3208959

[17] A. A. Almisreb, N. Md Tahir, S. Turaev, M. A. Saleh, and S. A. M. al Junid, "Arabic Handwriting Classification using Deep Transfer Learning Techniques," *Pertanika J Sci Technol*, vol. 30, no. 1, pp. 641–654, Jan. 2022. https://doi.org/10.47836/pjst.30.1.35

[18] S. Ibraheem Saleem and A. Mohsin Abdulazeez, "Hybrid Trainable System for Writer Identification of Arabic Handwriting," *Computers, Materials & Continua*, vol. 68, no. 3, pp. 3353–3372, 2021. https://doi.org/10.32604/cmc.2021.016342

[19] M. Hacibeyoglu, "Human Gender Prediction on Facial Mobil Images using Convolutional Neural Networks," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 3, no. 6, pp. 203–208, Sep. 2018. https://doi.org/10.18201/ijisae.2018644778

[20] S. A. Bello, S. Yu, C. Wang, J. M. Adam, and J. Li, "Review: Deep Learning on 3D Point Clouds," *Remote Sens (Basel)*, vol. 12, no. 11, p. 1729, May 2020. https://doi.org/10.3390/rs12111729

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 770–778. https://doi.ieeecomputersociety.org/10.1109/CVPR.2016.90

[22] Z. Guo, Q. Chen, G. Wu, Y. Xu, R. Shibasaki, and X. Shao, "Village Building Identification Based on Ensemble Convolutional Neural Networks," *Sensors*, vol. 17, no. 11, p. 2487, Oct. 2017. https://doi.org/10.3390/s17112487

[23] J. Fan, J. Lee, and Y. Lee, "A Transfer Learning Architecture Based on a Support Vector Machine for Histopathology Image Classification," *Applied Sciences*, vol. 11, no. 14, p. 6380, Jul. 2021. https://doi.org/10.3390/app11146380

[24] N. M. Blauch, M. Behrmann, and D. C. Plaut, "Computational insights into human perceptual expertise for familiar and unfamiliar face recognition," *Cognition*, vol. 208, p. 104341, Mar. 2021. https://doi.org/10.1016/j.cognition.2020.104341

[25] A. El-Sawy, H. EL-Bakry, and M. Loey, "CNN for Handwritten Arabic Digits Recognition Based on LeNet-5," 2017, pp. 566–575. https://doi.org/10.1007/978-3-319-48308-5_54