JOIV : Int. J. Inform. Visualization, 7(4) - December 2023 2183-2188

INTERNATIONAL JOURNAL
ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Arabic Character Recognition Using CNN LeNet-5

Gibran Satya Nugraha [a,*], I Gede Pasek Suta Wijaya [a], Fitri Bimantoro [a], Ario Yudo Husodo [a], Faqih Hamami [b]

[a] *Informatics Department, Mataram University, Selaparang, Mataram, 83125, Indonesia*
[b] *School of Industrial and System Engineering, Telkom University, Dayeuhkolot, Bandung, 40257, Indonesia*
*Corresponding author: *gibransn@unram.ac.id*

Abstract— **The human handwriting pattern is one of the research areas of pattern recognition; it is very complex. Therefore, research in this field has become quite popular. Moreover, human handwriting pattern recognition is needed for several things, one of them being character recognition. Recognition of Arabic handwriting is complex because everyone has different characteristics in writing and Arabic characters have quite abstract shapes and patterns. From previous research, Convolutional Neural Network (CNN), a deep learning-based algorithm, has a fairly high accuracy value when used for public datasets such as AHDB and private datasets. In this study, private datasets are used with a fairly high level of complexity because the respondents appointed to write Arabic letters come from different age categories. The CNN architecture used in this research is the architecture developed by Yan LeCun known as LeNet-5. The local dataset used was 8400 images, with details of 6720 for training data (each letter has 240 images) and 1680 for testing data (each letter has 60 images). The total respondents who wrote Arabic script were 30 people, and each person wrote each letter ten times. The accuracy obtained is 81% higher than in previous studies. The following study will test a number of additional CNN architectures to increase the accuracy of the results. In addition to accuracy, this study will also calculate the misclassification rate, root mean square error, and mean absolute error.**

Keywords— **Arabic; handwriting; pattern; deep learning; convolutional neural network.**

Manuscript received 15 Dec. 2021; revised 29 Feb. 2022; accepted 27 Aug. 2023. Date of publication 31 Dec. 2023.

I. Introduction

Pattern recognition is currently something very necessary because the development of the digital world is growing. All patterns are anything the computer must recognize to facilitate all human activities [1], [2]. The human handwriting pattern is one of the research areas of pattern recognition; it is very complex. Therefore, research in this field has become quite popular. Moreover, human handwriting pattern recognition is needed for several things, such as character recognition, numbers, important document analysis, bank checks, signature verification, etc. [3].

Deep learning has become a popular technique in recent years due to the rising demand for pattern identification and the growing volume of data [4]. Deep learning is widely used in the recognition or classification of certain animal characteristics [5], limb movement identification [6], disaster early warning systems [7], signal and voice recognition [8], medical image analysis [9], and also character recognition based on human handwriting [10], [11], [12]. Based on implementing deep learning in these various fields, it is certain that deep learning will greatly develop in the future. Researchers will also compete to improve research accuracy results and reduce errors when implementing deep learning to recognize patterns quickly and accurately.

Convolutional Neural Network is one of the most well-liked deep learning methods (CNN). The Artificial Neural Network (ANN) has evolved into CNN, which has the benefit of classifying enormous amounts of data combined with a variety of intricate features [13]. CNN has the characteristic of extracting features automatically; the features taken are very detailed. The processing performed on these features is carried out in each layer. This is done so that the machine can learn optimally to obtain sufficient accuracy when implementing CNN [14]. Nowadays, CNN is the main gateway to develop deep learning methods [15].

Feature extraction is a very important part of the classification process. Feature is a characteristic that can distinguish one class from another, thus the classified data is spread out accurately according to each class [16]. ANN requires a manual feature extraction process, which is very inefficient because it will spend much time and should be coupled with a feature selection process to determine the most relevant features to use [17]. CNN can overcome this; with its many layers, CNN can extract and study the features from the most general to the most detailed in each of its hundreds of layers [18]. This is the author's consideration in researching this Arabic handwriting recognition because human handwriting on Arabic letters, with various shapes and patterns, is something very complex. It will be very difficult to find if distinctive features are sought. Therefore, the CNN algorithm needs a deep-learning approach to overcome this [19].

Many studies have been conducted for handwriting recognition both offline and online. Rajalakshmi et al. [3] found quite complex difficulties in recognizing handwriting patterns because everyone has unique writing characteristics. The recognition system must recognize each writing pattern effectively and efficiently, producing high accuracy and with low error.

Agrawal et al. [10] look for an adaptive model to recognize handwritten numbers using machine learning and deep learning. The dataset utilized had information on 70.000 photos and was from MNIST. Several deep learning and machine learning techniques, including Support Vector Machine (SVM), RFC, K-nearest Neighbor (K-NN), Multilayer Perceptron (MLP), and Convolutional Neural Network, were tested in this study (CNN). When compared to other approaches, CNN's results had the highest accuracy: 99.06% for training accuracy and 98.80% for testing accuracy with ten epochs. RMSprop was utilized in this work as the model optimizer.

Darapaneni et al. [11] extracted handwriting on a form. The model was created using CNN. The process began by scanning the handwritten form, then removing the existing noise. The dataset used came from the EMINST dataset, which consisted of numbers and capital & non-capital letters. The accuracy obtained was 91%. Raundale et al. [12] explain that handwriting recognition is very necessary today and in the future. The system will store each input in the form of handwriting for important purposes in the future. Of course, it takes a system that can recognize handwriting accurately. However, there is found a difficulty, namely the various sizes of each person's handwriting. Several methods tried in this research were Simple (Artificial) Neural Network, Convolutional Neural Network, and Recurrent Neural Network (RNN).

Eltay et al. [18] researched Arabic handwriting, which is known to be very complex because of its cursive form, then very sensitive to the meaning of the word, non-uniform spaces between words, placement of words, periods, and non-uniform diacritics. This study looked for the best deep-learning algorithm in recognizing Arabic handwriting. The algorithms tried included CNN, RNN, long short-term memory (LSTM), and bi-directional long short-term memory (Bi-LSTM). The dataset used came from the Arabic handwritten database (AHDB). This study also proposed an adaptive method for augmenting imbalanced datasets that had been successfully applied to the AHDB dataset.

Arlobah et al. [20] combined the CNN method with the support vector machine (SVM) – eXtreme Gradient Boosting (XGBoost) classifiers and used the Hijaa dataset. CNN was used to extract features from Arabic handwriting, then passed on to the classification phase by machine learning. The accuracy obtained was 96.3% for 29 classes or letters. Alsaeedi et al. [21] explained that Arabic handwriting recognition is quite complex. This study used CNN for character recognition and Transparent Neural Network (TNN) for word reading. The accuracy obtained was 98%.

Al-Barhamtoshy et al. [22] used CNN to recognize Arabic letters obtained through handwritten, printed, and calligraphy media. The training data used were 13,440, obtained from each letter with 280 Arabic handwritten images, while the training data, used as many as 3360, obtained from each letter with 120 Arabic handwritten images. The accuracy obtained was 100%. Almansari et al. [23] compared the performance of CNN with multi-layer perceptron (MLP) using the Arabic Handwritten Characters Dataset (AHCD). The results obtained by CNN got an accuracy of 95.27%, while the MLP was 72.08%

II. Materials and Method

Recognition of Arabic handwriting is complex because everyone has different characteristics in writing and Arabic characters have quite abstract shapes and patterns. From previous research, CNN, a deep learning-based algorithm, has a fairly high accuracy value when used for public datasets such as AHDB and AHCD as well as private datasets, and CNN's performance will be optimal when working on large datasets. In this study, private datasets are used with a fairly high level of complexity because the respondents appointed to write Arabic letters come from different age categories. The CNN architecture used in this research is the architecture developed by Yan LeCun known as LeNet-5 [24].

A. Materials

The local dataset used was 8400 images, with details of 6720 for training data (each letter has 240 images) and 1680 for testing data (each letter has 60 images). The total respondents who wrote Arabic script were 30 people, and each person wrote each letter ten times. The dataset had previously been used in research [25], using the usual CNN algorithm, and obtaining an accuracy of 78.10%. The dataset can be seen in Fig 4.

B. Method

Convolutional Neural Network (CNN) is a deep learning algorithm. LeNet-5 is one of the CNN architectural models found by [24]. LeNet-5 has four main layers [26]

1) Convolution layer: The convolution layer is the main part of a CNN process. Convolution can be defined as repeatedly applying a function to another function's output. An example can be seen in Fig 1. Kernel is a function that will be applied to all parts of the pixel image that allow it to be convoluted, and the result of the convolution is at the bottom of the "convolved feature" image. The movement of the kernel begins from the top left and moves to the bottom right. This convolution method aims to identify features in the input image.

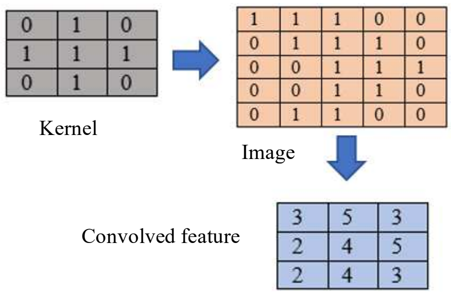

Fig. 1 Convolution process

2) Subsampling layer: This process aims to reduce the size of image data. The most used subsampling technique on CNN is max pooling [27]. As illustrated in Fig. 2, max pooling reduces the output of the convoluted picture into a smaller image matrix by taking the maximum value from each of the small grids.

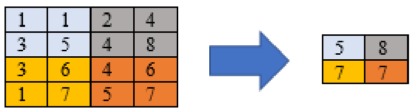

Fig. 2 Max pooling process

3) Fully connected layer: The data dimensions are prepared for linear classification in this layer. Each neuron in the convolution layer needs to be turned into one-dimensional data before moving on to a fully linked layer [28].

4) Activation Function: An ANN may transform input data into higher dimensions using the nonlinear activation function, allowing for straightforward hyper plane cuts that enable classification. CNN employs the sigmoid function as an activation function [29].

The distribution function of the sigmoid function is depicted in Fig 3. It changes the input x values range to between 0 and 1. The sigmoid function's primary flaw is that its output value range is not centered on zero. As a result, it is no longer often employed in practice.

LeNet-5 is a CNN architecture used in this study (Fig 5). LeNet-5 comprises eight layers: one input layer, one output layer, three convolutional layers, two feature extraction subsampling layers, and one fully connected layer. The second Layer, the first convolution layer, has 6 map features, with a total of 4704 neurons, 122304 connections, and a kernel with 5x5 size.

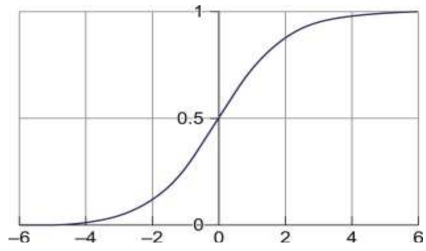

Fig. 3 Sigmoid Function [30]

The next part of LeNet-5 is the third layer which contains first subsampling layer, with a total of 6 map features, 1176 neurons, and 5580 connections, the number of kernels owned is 2x2. Next, the second convolution layer (fourth layer), which is larger than the first, has 1600 neurons, 151600 connections, 16 maps features, and a kernel with 5x5 size.

The fifth layer is a subsampling layer with 16 maps features, 400 neurons, 2000 connections, and a 2x2 kernel. The sixth layer is a convolution layer that have features maps, neurons, and connections, respectively 120, 120, 48120. The seventh layer, the last layer, is fully connected and has a feature map with 10 neurons and 1210 connections. Table 1 contains information on the LeNet-5 layers [31].

TABLE I
LENET-5 LAYERS [31]

Layers	**Description (number of)**		
	Feature Maps	**Neuron**	**Connection**
1	1	0	0
2	6	4704	122304
3	6	1176	5880
4	16	1600	151600
5	16	400	2000
6	120	120	48120
7	10	10	1210

III. RESULTS AND DISCUSSION

This section will explain the implementation of the CNN method with the LeNet-5 architecture. The programming languages used were Python and the Tensorflow library. The number of datasets used was as described in sub-chapter III.A; an example of the dataset was in Fig. 4, while for the LeNet-5 architecture as in Fig. 5

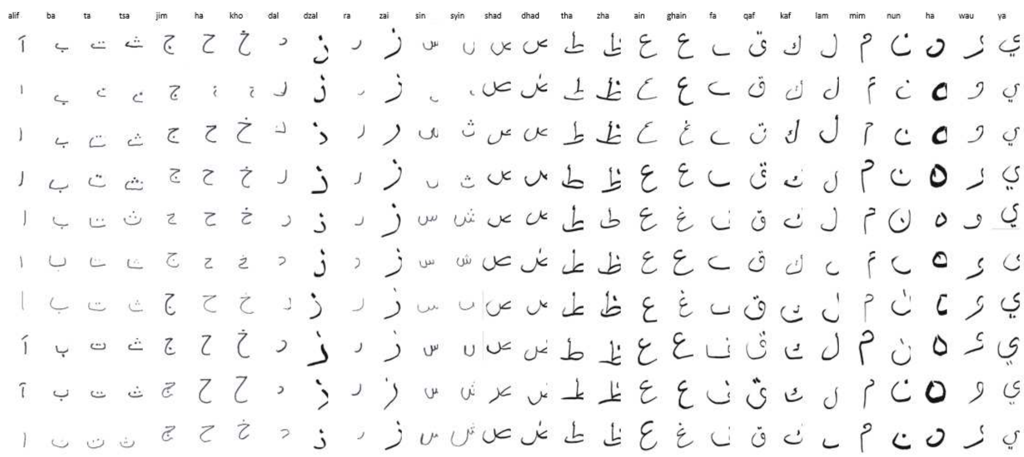

Fig. 4 Arabic handwriting dataset

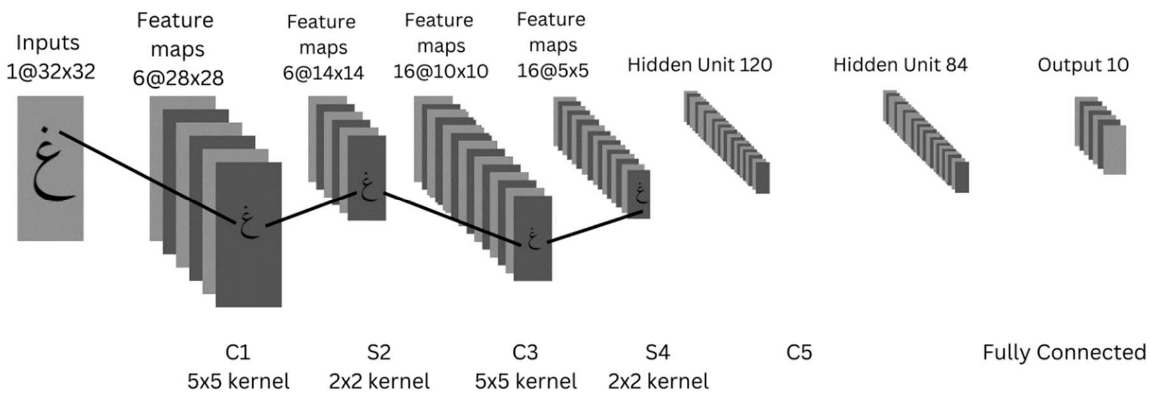

Fig. 5 LeNet-5 architecture [31]

The parameters of accuracy rate, root mean square error, mean absolute error, and misclassification error were used to test the classification findings on all Arabic handwriting images. In addition, the confusion matrix was also used to calculate how many images of each letter were successful or failed to be classified. In Fig 6, the accuracy continues to increase as the epoch increases. This shows that the LeNet-5 architecture requires many learning phases to achieve optimal accuracy. There are two graph lines; the blue one is the accuracy for training (in the legend it is written as "accuracy"), while the yellow one is for the validation phase (in the legend it is written as "val_accuracy"). As shown in Fig 1, the optimal accuracy in the validation phase obtained at 81% is not much different from that during training.

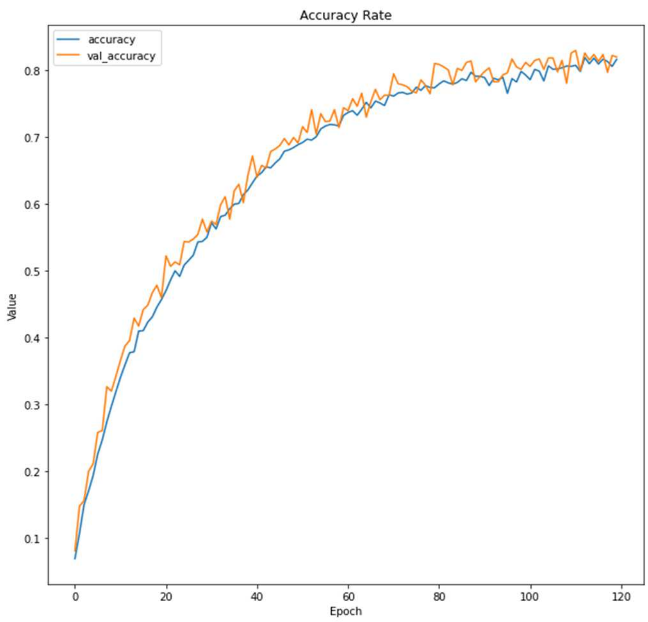

Fig. 6 Accuracy rate

The root means square error calculates the error level in the prediction results where the smaller (closer to 0) the RMSE value, the more accurate the prediction results will be. The RMSE results can be seen in Fig 7. At the 120th epoch, the RMSE value is very small, below 0.10. The RMSE values for the train (in the legend it is written as "root_means_squared_error") and validation (in the legend it is written as "val_root_means_squared_error") data do not significantly differ. Mean absolute error (MAE) is one of the methods used to measure the accuracy of the prediction model. MAE results show the mean value of the absolute error of the actual value with the predicted value. In Fig 8, the MAE value decreases as the number of epochs increases. The MAE results for the train (in the legend it is written as "mae") and validation (in the legend it is written as "val_mae") processes are almost similar.

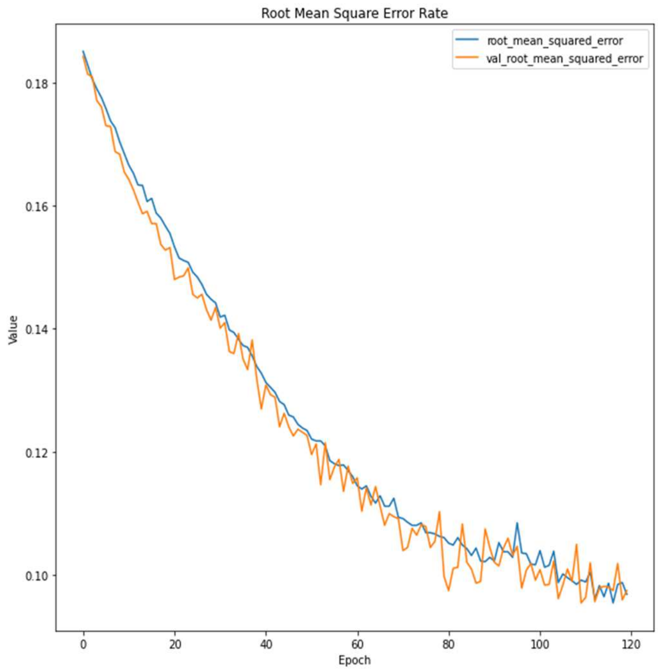

Fig. 7 Root means square error rate

The misclassification rate graph is a graph that shows the progress of the classification errors made. In Fig 9, the error value continues to decrease as the epoch increases. The error values during train (in the legend it is written as "loss") and validation (in the legend it is written as "val_loss") have almost the same value. The confusion matrix in Fig 10, shows that most Arabic letters can be classified correctly. In the confusion matrix, the description of the order of Arabic letters from the left-right and top-bottom columns is alif, ba, ta, tsa, jim, ha, kho, dal, dzal, ra, zai, sin, syin, shad, dhad, tha, zha, ain, ghain, fa, qaf, kaf, lam, mim, nun, ha, wau, ya. When compared with previous studies [25] which used basic CNN, the accuracy was 78.1%, meaning that the implementation of the LeNet-5 architecture in this study got better results.

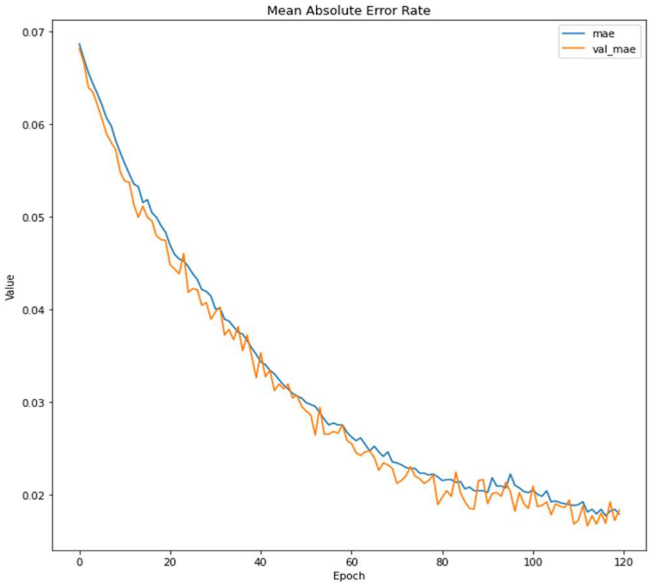

Fig. 8 Mean absolute error rate

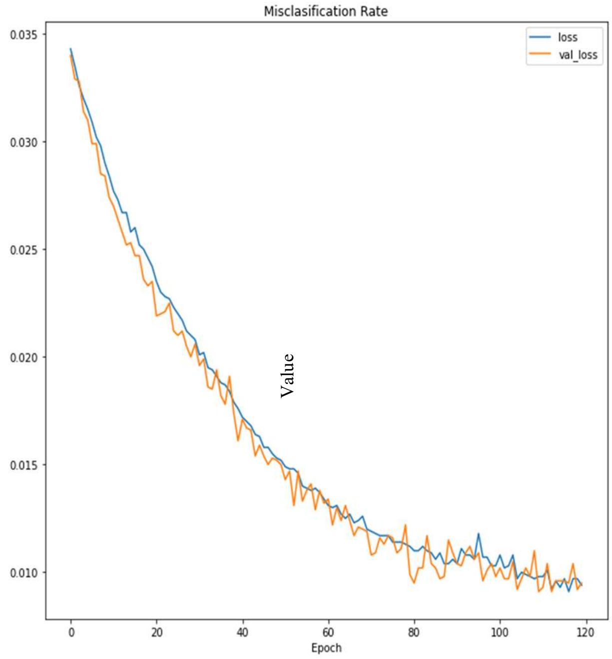

Fig. 9 Misclassification rate

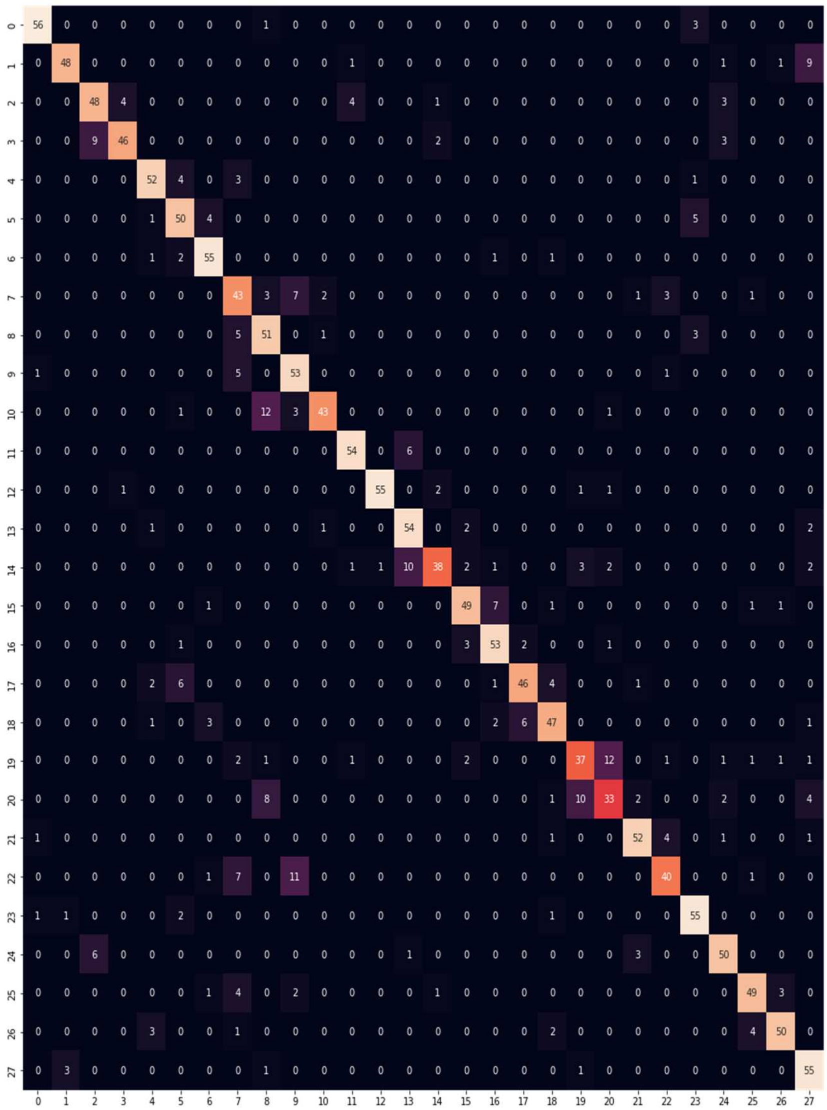

Fig. 10 Confusion matrix

IV. Conclusion

Recognition of Arabic handwriting is an area of research that continues to grow because of the complexity of each person's patterns when writing. In this study, the CNN algorithm was implemented using the LeNet-5 architecture to recognize Arabic handwriting patterns using a local dataset. The accuracy obtained in this study is 81%, and it is better than in previous studies (78.10%) [25]. The upcoming research can attempt several other CNN architectures to improve the accuracy of the classification results.

References

[1] M. Yu. Mikheev, Yu. S. Gusynina, and T. A. Shornikova, "Building Neural Network for Pattern Recognition," 2020 International Russian Automation Conference (RusAutoCon), Sep. 2020, doi:10.1109/rusautocon49822.2020.9208207.

[2] X. Bai *et al.*, "Explainable deep learning for efficient and robust pattern recognition: A survey of recent developments," *Pattern Recognit*, vol. 120, p. 108102, 2021.

[3] M. Rajalakshmi, P. Saranya, and P. Shanmugavadivu, "Pattern Recognition-Recognition of Handwritten Document Using Convolutional Neural Networks," 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Apr. 2019, doi:10.1109/incos45849.2019.8951342.

[4] S. I. Ali, S. S. Ebrahimi, M. Khurram, and S. I. Qadri, "Real-Time Face Mask Detection in Deep Learning using Convolution Neural Network," 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT), Jun. 2021, doi: 10.1109/csnt51715.2021.9509704.

[5] A. Varshney, A. Katiyar, A. K. Singh, and S. S. Chauhan, "Dog Breed Classification Using Deep Learning," 2021 International Conference on Intelligent Technologies (CONIT), Jun. 2021, doi:10.1109/conit51480.2021.9498338.

[6] K. Neeraja, K. Srinivas Rao, and G. Praneeth, "Deep Learning based Lip Movement Technique for Mute," 2021 6th International Conference on Communication and Electronics Systems (ICCES), Jul. 2021, doi: 10.1109/icces51350.2021.9489122.

[7] J. Han, G. Kim, C. Lee, Y. Han, U. Hwang, and S. Kim, "Predictive Models of Fire via Deep learning Exploiting Colorific Variation," 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Feb. 2019, doi:10.1109/icaiic.2019.8669042.

[8] T. Singh, S. R. Kudavelly, and K. Venkata Suryanarayana, "Deep Learning Based Fetal Face Detection And Visualization In Prenatal Ultrasound," 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Apr. 2021, doi: 10.1109/isbi48211.2021.9433915.

[9] F. Syuhada, R. Indraswari, A. Z. Arifin, and D. A. Navastara, "Multi-Projection Segmentation on Dental Cone Beam Computed Tomography Images Using Level Set Method," *Journal of Computer Science and Informatics Engineering (J-Cosine)*, vol. 5, no. 2, pp. 130–139, 2021.

[10] A. K. Agrawal, A. K. Shrivas, and V. kumar Awasthi, "A Robust Model for Handwritten Digit Recognition using Machine and Deep Learning Technique," 2021 2nd International Conference for Emerging Technology (INCET), May 2021, doi:10.1109/incet51464.2021.9456118.

[11] N. Darapaneni et al., "Handwritten Form Recognition Using Artificial Neural Network," 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS), Nov. 2020, doi:10.1109/iciis51140.2020.9342638.

[12] P. Raundale and H. Maredia, "Analytical Study of Handwritten Character Recognition: A Deep Learning Way," 2021 International Conference on Intelligent Technologies (CONIT), Jun. 2021, doi:10.1109/conit51480.2021.9498347.

[13] S.-Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros, "CNN-Generated Images Are Surprisingly Easy to Spot… for Now," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2020, doi: 10.1109/cvpr42600.2020.00872.

[14] I. Ahmed, M. Ahmad, F. A. Khan, and M. Asif, "Comparison of Deep-Learning-Based Segmentation Models: Using Top View Person Images," IEEE Access, vol. 8, pp. 136361–136373, 2020, doi:10.1109/access.2020.3011406.

[15] S. Dong, P. Wang, and K. Abbas, "A survey on deep learning and its applications," *Comput Sci Rev*, vol. 40, p. 100379, 2021.

[16] D. Kollias and S. Zafeiriou, "Exploiting Multi-CNN Features in CNN-RNN Based Dimensional Emotion Recognition on the OMG in-the-Wild Dataset," IEEE Transactions on Affective Computing, vol. 12, no. 3, pp. 595–606, Jul. 2021, doi: 10.1109/taffc.2020.3014171.

[17] M. Sreenivasulu and M. Sridevi, "Comparative study of statistical features to detect the target event during disaster," Big Data Mining and Analytics, vol. 3, no. 2, pp. 121–130, Jun. 2020, doi:10.26599/bdma.2019.9020021.

[18] M. Eltay, A. Zidouri, and I. Ahmad, "Exploring Deep Learning Approaches to Recognize Handwritten Arabic Texts," IEEE Access, vol. 8, pp. 89882–89898, 2020, doi: 10.1109/access.2020.2994248.

[19] Y. Hamdi, H. Boubaker, and A. M. Alimi, "Online Arabic handwriting recognition using graphemes segmentation and deep learning recurrent neural networks," in *Enabling Machine Learning Applications in Data Science*, Springer, 2021, pp. 281–297.

[20] N. Alrobah and S. Albahli, "A Hybrid Deep Model for Recognizing Arabic Handwritten Characters," IEEE Access, vol. 9, pp. 87058–87069, 2021, doi: 10.1109/access.2021.3087647.

[21] A. Alsaeedi, H. A. Mutawa, S. Snoussi, S. Natheer, K. Omri, and W. A. Subhi, "Arabic words Recognition using CNN and TNN on a Smartphone," 2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR), Mar. 2018, doi:10.1109/asar.2018.8480267.

[22] H. M. Al-Barhamtoshy, K. M. Jambi, S. M. Abdou, and M. A. Rashwan, "Arabic Documents Information Retrieval for Printed, Handwritten, and Calligraphy Image," IEEE Access, vol. 9, pp. 51242–51257, 2021, doi: 10.1109/access.2021.3066477.

[23] O. A. Almansari and N. N. W. N. Hashim, "Recognition of Isolated Handwritten Arabic Characters," 2019 7th International Conference on Mechatronics Engineering (ICOM), Oct. 2019, doi:10.1109/icom47790.2019.8952035.

[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[25] 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.

[26] M. Kayed, A. Anter, and H. Mohamed, "Classification of Garments from Fashion MNIST Dataset Using CNN LeNet-5 Architecture," 2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE), Feb. 2020, doi: 10.1109/itce48509.2020.9047776.

[27] J. D. Bodapati, U. Srilakshmi, and N. Veeranjaneyulu, "FERNet: A Deep CNN Architecture for Facial Expression Recognition in the Wild," *Journal of The Institution of Engineers (India): Series B*, pp. 1–10, 2021.

[28] Z. A. Sejuti and M. S. Islam, "An efficient method to classify brain tumor using CNN and SVM," in *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, 2021, pp. 644–648.

[29] M. Bhuvaneshwari and E. G. M. Kanaga, "Convolutional Neural Network for Addiction Detection using Improved Activation Function," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021, pp. 996–1000.

[30] K. Chakraborty, S. Bhattacharyya, R. Bag, and A. A. Hassanien, "Sentiment Analysis on a Set of Movie Reviews Using Deep Learning Techniques," Social Network Analytics, pp. 127–147, 2019, doi:10.1016/b978-0-12-815458-8.00007-4.

[31] A. El-Sawy, H. EL-Bakry, and M. Loey, "CNN for Handwritten Arabic Digits Recognition Based on LeNet-5," Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016, pp. 566–575, Oct. 2016, doi: 10.1007/978-3-319-48308-5_54.