Implementation of Convolutional Neural Network with VGG-16 Architecture in Digital Hiragana Handwriting Image Recognition

Fitri Mintarsih[1], **Rheditia Ferdiansyah**[2], **Victor Amrizal**[3], **Hendra Bayu Suseno**[4] **Tjut Awaliyah Zuraiyah.**[5*]

[1,2,3,4] Departement of Informatics Engineering, Faculty of Science and Technology, State Islamic University Syarif Hidayatullah, Jakarta

[5] Department of Computer Science, Faculty of Mathematics and Natural Science, Pakuan University, Bogor, West Java, 16143, Indonesia

Abstract

The number of Japanese language learners in Indonesia ranks second at 711,732 people. Hiragana is the first letter to be learned, especially at the beginner level and is usually learned before Katakana and Kanji. Some characters in Hiragana have similar main forms such as nu (ぬ) and me (め), ne (ね) and wa (わ), thus adding complexity to the recognition process. Like previous research that created a Hiragana pronunciation learning application and previous research that was an English writing learning application, allowing people to learn on their own, by applying CNN (Convolutional Neural Network) to recognize written characters, researchers were inspired to apply this in learning to write Hiragana letters. Therefore, researchers created a digital Hiragana handwriting recognition model using the VGG-16 CNN Architecture method so that the model created can later be used in a Hiragana learning application for writing. This study used a dataset in the form of digital Hiragana handwriting images totaling 1518 data with 33 data for each label (46 types of letters). The hyperparameters used in this study to train the model were 5 epochs, a batch size of 32, the Adam Optimizer, and a Learning rate of 0.001. Based on the test results with the aforementioned parameters, the Accuracy value was 98.55%, Precision was 98.91%, Recall was 98.55%, and the F1-Score was 98.51%.

Keywords: *Japanese Language, Hiragana, CNN, VGG-16, Image Recognition*

1) Introduction

Japanese is a foreign language highly sought after by Indonesians. A significant number of learners choose Japanese as their foreign language. In 2021, Indonesia ranked second in the number of Japanese learners after China, with 711,732 learners [24], representing 18.8% of the world's total. Of these 711,732 Japanese learners, 642,605 (90.28%) were at the high school/vocational high school (SMA) or equivalent level. There are four language skills that must be mastered when learning a foreign language: reading, writing, listening, and speaking [19]. Learning a foreign language, especially Japanese, presents some challenges for Indonesians due to the distinctive characteristics of the Japanese alphabet. In Japanese, there are three types of Letters that must be learned: Hiragana, Katakana, and Kanji. Hiragana is the first to be learned, especially for beginners, and is usually learned before Katakana and Kanji. Hiragana letters only represent a specific sound and do not have any meaning like the Indonesian alphabet, although there are also words in Japanese that consist of one syllable, such as me/め (eye), ki/き (tree), ni/に (two). This is different from Kanji, where each character represents a specific meaning [23]. Japanese does not have characters like spaces that separate words. There are two punctuation marks called kuten (。) which functions as a period and toten (、) which functions as a comma. Some characters in hiragana also have similar main forms such as nu (ぬ) and me (め), ne (ね) and wa (わ), thus adding complexity to the recognition process [19].

In research [11], which analyzed the difficulties of Japanese language learning among students at SMK Bagimu Negeriku Semarang, it was found that 62.8% of students experienced difficulties in writing and reading hiragana and katakana due to the similarity of some characters. In research [19] to determine the

*Corresponding author. *E-mail address: fitri.mintarsih@uinjkt.ac.id*
Received: xx xxxxx 20xx, Accepted: xx xxxxx 20xx and available online XX Januari 2026
DOI: https://doi.org/10.33751/komputasi.v19i2.5260

mastery of hiragana among 10th-grade students at SMAN 9 Pekanbaru, the students' mastery of hiragana was still in the sufficient category, as seen from the average student test score of 68.8. Many students still found it difficult to master hiragana and seemed to have difficulty distinguishing similar hiragana letters. In research [16], which examined the factors of difficulty in learning to write hiragana letters among 10th grade students at SMA Labschool Surabaya in the 2019/2020 academic year, it was shown that there were several difficulties in learning to write hiragana letters, namely difficulty in writing hiragana letters due to an inability to read the letters, difficulty remembering because hiragana letters are considered complicated and students are not used to writing hiragana letters, and inability to distinguish similar hiragana letters. [3] In his research, he successfully created an application for learning Japanese speaking using mobile learning. This study designed a Japanese hiragana learning application for sentences using speech recognition. In further development, the researchers hope the application will not only serve as a learning medium for pronunciation, but also for writing and listening.

[30] In his research, he created an application for learning English writing that allows learners to learn independently without a teacher. This application uses a deep learning algorithm, namely a Convolutional Neural Network, to recognize characters written by the user. Based on research conducted by [23], [2], and [25], it was found that the CNN algorithm achieves the best performance when applied to image data compared to other machine learning algorithms. Based on research [21] on a comparison of Convolutional Neural Network architectures for Fundus Classification, the architectures tested were AlexNet, Visual Geometry Group (VGG) 16, VGG 19, Residual Network (ResNet) 50, Resnet101, GoogleNet, Inception V3, InceptionResNetV2, and Squeezenet. The test results showed the best architecture was VGG16, with an accuracy of up to 92.31%. Based on the author's explanation above, the author was motivated to conduct research to develop a Japanese handwriting recognition model, specifically Hiragana, that could be applied to Japanese language learning. The researcher titled the research "Implementation of the Convolutional Neural Network with VGG-16 Architecture in Digital Hiragana Handwriting Image Recognition."

2) Methods

A. OSEMN Methods

1. Obtain

At this stage, the author collects the input data needed for system modeling. The collected data consists of digital handwritten images of hiragana letters, with the author limiting the categories in this study to 46 basic hiragana letters. The author collects data from internet sources and supplements it with data created himself using a drawing tablet. The collected data is tested using several data testing methods, such as data consistency testing and data accuracy testing, to improve data quality and integrity.

The collected dataset consisted of 1,518 images, from which 1,380 images were taken for model training and 138 images for model evaluation. The dataset was divided into 1,104 training data (24 images for each letter) and 276 validation data (6 images for each letter), with 138 testing data (3 images for each letter) for evaluation.

2. Scrub

Once the entire image dataset has been successfully collected, the data will be processed to ensure it meets the model's requirements. This stage involves resizing, splitting, and normalizing the data.

a) Resize

The authors resized the data to 224x224 pixels using the Pillow library. This was done because the VGG-16 architecture, trained on the ImageNet dataset, only accepts input with a pixel size of 224x224 and three color channels, RGB.

b) Dataset Splitting

The data was prepared for the training process. The 1,380 images collected for model training were divided into 80% training data and 20% validation data. The testing data was separated beforehand because it was not used in the model training process.

c) Normalize

The final step in the scrubbing process is normalization. If the image data is used as is in the model, calculating the values per pixel can increase the training complexity. Rescaling the pixel values in the data to within the [0-1] range will reduce the number of values and speed up the computation.

3. Explore

At this stage, the authors explored the data to determine the amount of available data and ensure a balance of data for each label before proceeding to the model creation process. They used visualizations to examine the data distribution per label.

a) Data Distribution per Label

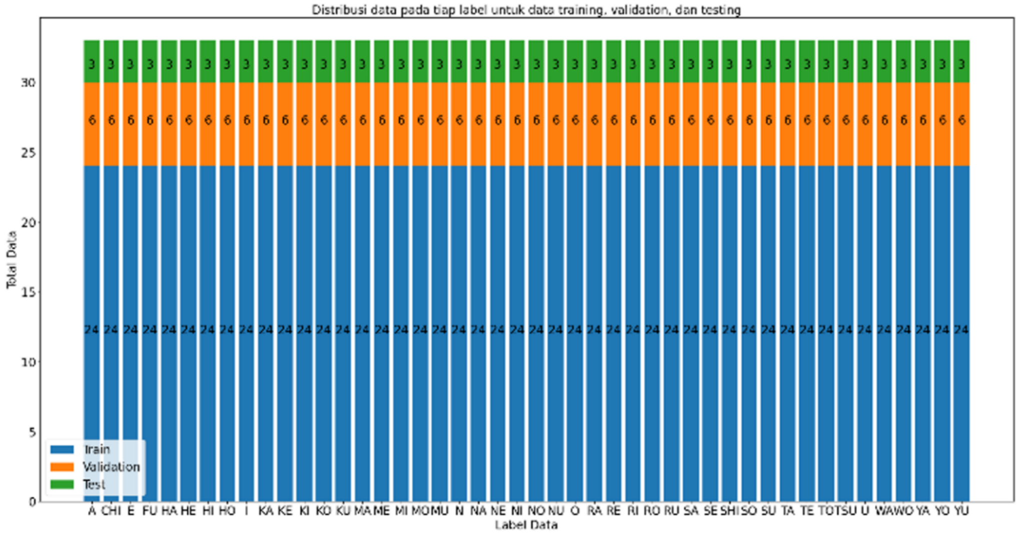

Figure 1. Data Distribution Bar Chart

Data distribution is evaluated to ensure a balanced amount of data in each label. A balanced distribution of data in each label is important to avoid bias during training. In Figure above, the bar chart visualization shows that each label's data has the same height, indicating that the data in each label is evenly distributed.

4. Model

The author designed and built a model using the CNN method with the VGG-16 architecture using an existing dataset to train the model so that it can recognize digital handwritten hiragana images. The model that the author created applies transfer learning by using the VGG-16 model that has been trained on the Imagenet dataset and reused to train the hiragana image model. At this stage, the author also prepared several parameters that will be tested to determine the best parameters for the model to be created.

5. Interpret

After the model has been successfully built, the author will conduct an interpretation of the modeling results, which includes evaluating the metrics through manual testing using a confusion matrix.

3) Result and Discussion

A. Model Design

At this stage the author creates a model architecture to be used in model training.

Table 1. Layer Model Structure

Model Functional

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 224, 224, 3)	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1.792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36.928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73.856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147.584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295.168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590.080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590.080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1.180.160

block4_conv2 (Conv2D)	(None, 28, 28, 512)	2.359.808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2.359.808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2.359.808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2.359.808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2.359.808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 256)	6.422.784
dense_1 (Dense)	(None, 256)	65.972
dense_2 (Dense)	(None, 46)	11.882

Total params :21,215,086 (80.93 MB)
Trainable params : 6,500,398 (24,80 MB)
Non-trainable params : 14,714,688 (56,13 MB)

The table above shows the layer structure of the model to be used. The layers used consist of input, convolution, pooling, flatten, and dense layers.

B. Parameter Tuning

In creating the model, the author performed parameter tuning to select the best parameters for the model to be used

Table 2. Parameter Tuning

Optimizer	Learning rate	Accuracy	Loss
SGD	0,01	0,2609	2,339
SGD	0,001	0,2826	3,4143
SGD	0,0001	0,0435	3,8094
Adam	0,01	0,0435	3,7325
Adam	0,001	0,9855	0,0764
Adam	0,0001	0,9855	0,3191
RMSprop	0,01	0,0217	3,8288
RMSprop	0,001	0,9348	0,1627
RMSprop	0,0001	0,9565	0,3244
Adagrad	0,01	0,6014	1,4983
Adagrad	0,001	0,6522	2,6331
Adagrad	0,0001	0,1087	3,6872
Adadelta	0,01	0,8043	2,7031
Adadelta	0,001	0,0362	3,7733
Adadelta	0,0001	0,0217	3,9099

The table above shows that the best parameter results were obtained using the Adam optimizer with a learning rate of 0.001, which achieved an accuracy of 98.55%.

C. Model Training

```
Epoch 1/5
35/35 - 16s - 447ms/step - accuracy: 0.3605 - loss: 2.9145 - val_accuracy: 0.6159 - val_loss: 1.6445
Epoch 2/5
35/35 - 7s - 212ms/step - accuracy: 0.8877 - loss: 0.5753 - val_accuracy: 0.9239 - val_loss: 0.4295
Epoch 3/5
35/35 - 8s - 216ms/step - accuracy: 0.9801 - loss: 0.1537 - val_accuracy: 0.9674 - val_loss: 0.2319
Epoch 4/5
35/35 - 8s - 221ms/step - accuracy: 0.9937 - loss: 0.0624 - val_accuracy: 0.9964 - val_loss: 0.1148
Epoch 5/5
35/35 - 8s - 227ms/step - accuracy: 0.9964 - loss: 0.0366 - val_accuracy: 0.9928 - val_loss: 0.0753
```

Figure 2. Model Training Result

The image above shows the results of model training. The model was trained with hyperparameters of 5 epochs, a batch size of 32, Adam as the optimizer, and a learning rate of 0.001. The trained model achieved an accuracy of 99.64%.

D. Interpret

At this stage, the author evaluates the model created by testing it with previously prepared testing data. The author evaluates this by examining the recognition results for each letter in the testing dataset and comparing the original labels to see which letters were correctly and incorrectly recognized.

1. Confusion Matrix

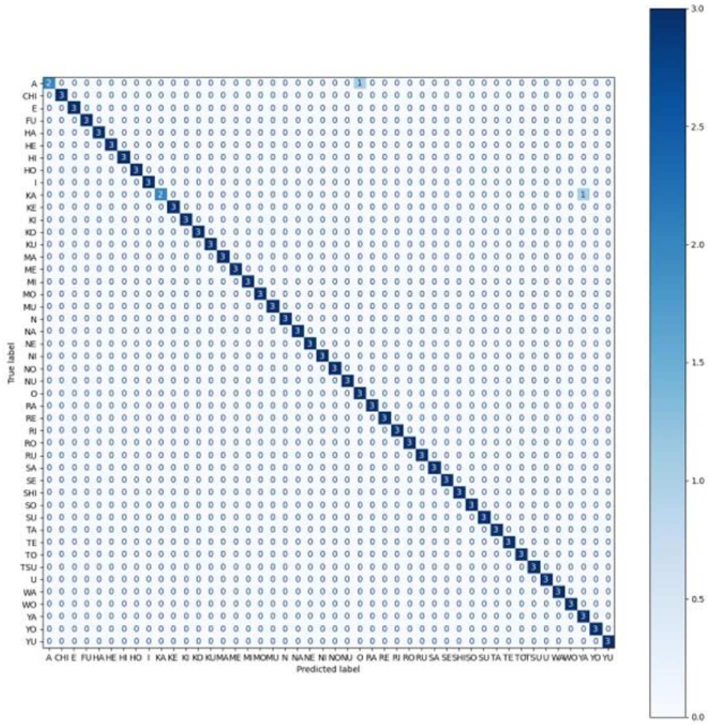

Figure 3. Confusion Matrix

In the picture above, it is shown that there are still errors in recognizing the letters A and the letters KA. Meanwhile, the other letters are recognized correctly. Of the 138 testing data, there were 2 data that were incorrectly recognized.

2. Model Evaluation

Table 3. Evaluation Metrics Table

Testing Set			
Accuracy	**Precision**	**Recall**	**F1-Score**
98.55%	98.91%	98.55%	98.51%

The table above is the result of the evaluation of the model that has been created, it can be seen that after being evaluated with testing data, the model with Adam as the Optimizer and a Learning rate of 0.001 has an accuracy of 98.55%, Precision of 98.91%, Recall of 98.55% and F1 Score of 98.51%.

4) Conclusion

Based on the research results described by the author, the following conclusions can be drawn: In developing a model for recognizing digital handwritten Hiragana images, the author prepared a dataset of 1,518 digital handwritten Hiragana images representing 46 letters. The convolutional neural network architecture (VGG-16) was used. The model consisted of a 224x224 pixel input with three color channels (RGB), 13 convolution layers, five pooling layers, one flattening layer, and three dense layers. The final layer had 46 neurons, corresponding to the number of recognized Hiragana letters. The hyperparameters were 5 epochs, a batch size of 32, Adam as the optimizer, and a learning rate of

0.001. The model was able to recognize digital handwritten Hiragana images, achieving an accuracy of 98.55%, a precision of 98.91%, a recall of 98.55%, and an F1-score of 98.51% when tested on previously unseen data.

References

[1] Ali, A. H., Yaseen, M. G., Aljanabi, M., Abed, S. A. (2023). Transfer Learning: A New Promising Techniques. Mesopotamian journal of Big Data, Vol. (2023), pp 31-32

[2] Amrustian, M. A., Muliati, V. F., & Awal, E. E. (2021). Studi Komparasi Metode Machine Learning untuk Klasifikasi Citra Huruf Vokal Hiragana. Jurnal Media Informatika Budidarma, 5(3), 905. https://doi.org/10.30865/mib.v5i3.3083

[3] Anjasmara, R., Lestari, I., & Dewi, D. M. (2019). Android Menggunakan Speech Recognition. Jurnal Komputer Terapan, 5(2). https://jurnal.pcr.ac.id/index.php/jkt/

[4] Atanasova, T., & Dineva, K. (2018). Osemn Process for Working Over Data Acquired By Iot Devices Mounted in Beehives. Current Trends in Natural Sciences (on-Line), 7(13), 47-53. http://www.natsci.upit.ro

[5] Byrne, C. (2017). Development workflows for O'Reilly, pp.28. Data Scientists.

[6] Colaboratory – Google.(n.d.).

[7] Damayanti, F., Rafika, A., Rismayanti (2023). Analisis Fungsi Aktivasi Sigmoid Biner dan Bipolar pada Backpropagation dalam Mengklasifikasikan Pasien Kanker Pasca Operasi. Jurnal unitek, 16(2), 237-249.

[8] DeepAI. (2020). Weight (Artificial Neural Network). DeepAI. https://deepai.org/machine-learning glossary-and-terms/weight-artificial neural-network

[9] Efrian, M. R., & Latifa, U. (2022). Image Recognition Berbasis Convolutional Neural Network (CNN) untuk Mendeteksi Penyakit Kulit pada Manusia. Jurnal POLEKTRO: Jurnal Power Elektronik, 11(1), 276-282.

[10] Fernandi, A., Magdalena, R., & Saidah, S. (2024). International Journal on Informatics Visualization journal homepage : www.joiv.org/index.php/joiv International Journal on Informatics Visualization Handwritten Hiragana Letter Detection Using CNN. www.joiv.org/index.php/joiv

[11] Istiqomah, D., Diner, L., & Wardhana, C. K. (2015), Analisis Kesulitan Belajar Bahasa Jepang Siswa SMK Bagimu Negeriku Semarang. Journal of Japanese Learning and Teaching, 4(1). http://journal.unnes.ac.id/sju/index.php /chie

[12] Jasin, R. A., Santoso, V. L., Tanuwijaya, E., & Sugianto, N. (2023). Pengembangan Arsitektur VGG16 dan DCNN7 pada Convolutional Neural Network dalam Melakukan Klasifikasi Pose Yoga. Jurnal Sistem dan Teknologi Informasi, 11(2), 314-322.

[13] Karthikeyan, J., Hie, T. S., & Jin, N. Y. (2021). Learning Outcomes of Classroom Research. L Ordine Nuovo Publication.

[14] Kunert, H. (2020). Contemporary Loanwords in hiragana: An Analysis of Typical Traits and Contexts, Japanese Studies, 40(1), 21-40.

[15] Lensun, S. F. (2016). Peningkatan Penguasaan Kanji dengan Metode Nemonik Melalui Multimedia.

[16] Mulyana, T. O. E. (2020). Faktor Kesulitan Belajar MenulisS Huruf Hiragana pada Siswa Kelas X SMA LabsSchool Surabaya Tahun Ajaran 2019/2020. Hikari, 4(1), 61-67.

[17] Nugroho, N. E. W., & Harjoko, A. (2021). Transliteration of Hiragana and Katakana Handwritten Characters Using CNN-SVM. Indonesian Journal of Computing and Cybernetics System, 15(3), 221-232. https://doi.org/10.22146/ijccs.66062

[18] Omori, Y., & Shima, Y. (2020). Image Augmentation for Eye Contact Detection Based on Combination of Pre-trained Alex-Net CNN and SVM. Journal of Computers, 15(3), 85–97. https://doi.org/10.17706/jcp.15.3

[19] Panjaitan, S. H., Basri, M. S., & Budiani, D. (2022). The Mastery of Hiragana Letters in Japanese Language Learning for Class X Students of SMAN 9 Pekan Baru.

[20] Renariah (2002). Bahasa Jepang dan karakteristiknya. Jurnal Sastra Jepang Fakultas Sastra Universitas Kristen Maranatha, 1(2).

[21] Setiawan, W. (2019). Perbandingan Arsitektur Convolutional Neural Network untuk Klasifikasi Fundus. Jurnal Simantec, 7(2), 48 53. https://doi.org/10.21107/simantec.v7i2. 6551

[22] Sugihartadi, B. E. (2015). Nakayoku Nihongo X. Bandung: RIZQI Press

[23] Susilo, M. M., Wonohadidjojo, D. M., & Sugianto, N. (2017). Pengenalan Pola Karakter Bahasa Jepang Hiragana Menggunakan 2D Convolutional Neural Network. https://www.researchgate.net/publicati on/331889051

[24] The Japan Foundation. (2021). Survey Report on Japanese Language Education Abroad 2021.

[25] Wardhani, P. R. C., & Florestiyanto, M. Y. (2024). Pengenalan Karakter Tulisan Tangan Bahasa Jepang Hiragana menggunakan Support Vector Machine dan Scale Invariant Feature Transform Recognition Of Hiragana Japanese Handwriting Characters Using Support Vector Machine And Scale Invariant Feature Transform. Jurnal Informatika Dan Teknologi Informasi, 21(2), 135–144. https://doi.org/10.31515/telematika.v21 i2.12042

[26] What is Python? Summary | Python.org. (n.d.)

[27] Wijoyo, A., Saputra, A. Y., Ristanti, S., Sya'Ban, S. R., Amalia, M., & Febriansyah, R. (2024). Pembelajaran Machine Learning. Oktal: Jurnal Ilmu Komputer dan Science, 3(2), 375-380.

[28] Willyanto, A., Alamsyah, D., & Irsyad, H. (2021). Identifikasi Tulisan Tangan Aksara Jepang Hiragana Menggunakan Metode CNN Arsitektur VGG-16. Jurnal Algoritme, 2(1), 1–11.

[29] Woods, R. E., & Gonzalez, R. C. (2008). Digital Image Processing Third Edition Pearson International Edition prepared by Pearson Education

[30] Zin, T. T., Thant, S., Pwint, M. Z., & Ogino, T. (2021). Handwritten character recognition on android for basic education using convolutional neural network. (Switzerland), Electronics 10(8). https://doi.org/10.3390/electronics1008 0904