CICES (*Cyberpreneurship Innovative and Creative Exact and Social Science*)
DOI: 10.33050/cices.v11i1.3505 ISSN: **2356-5209**
pp. 70~81 *Online* ISSN: **2655-3058**

Recognition of Handwritten Hangeul Characters Using Convolutional Neural Network

Kezia[1], Anggraini Puspita Sari*[2], Hendra Maulana[3]
[1,2,3]Informatics, Faculty of Computer Science, Veteran National Development University of East Java
E-mail:[1]tankezia44@gmail.com ,***[2]anggraini.puspita.if@upnjatim.ac.id** ,
[3]hendra.maulana.if@upnjatim.ac.id

Abstrak

Peningkatan jumlah wisatawan yang berkunjung ke Korea Selatan, didorong oleh pengaruh global dari "Korean Wave," telah memicu minat yang semakin besar terhadap budaya dan bahasa Korea. Lonjakan minat ini menekankan perlunya pendekatan digital yang efisien untuk mempermudah proses pengenalan karakter tulisan tangan Korea, khususnya Hangeul. Penelitian ini bertujuan untuk mengeksplorasi efektivitas penggunaan Convolutional Neural Network (CNN) dalam mengklasifikasikan karakter Hangeul tulisan tangan. Dataset yang digunakan terdiri dari 2.400 gambar yang dikategorikan ke dalam konsonan dan vokal, digunakan untuk proses pelatihan, validasi, dan pengujian model CNN. Studi ini menggunakan tiga rasio pembagian data yang berbeda (60:20:20, 70:15:15, 80:10:10) dengan dua tingkat pembelajaran (0,001 dan 0.0001) selama 10 epoch untuk menentukan konfigurasi yang paling efektif. Hasil penelitian menunjukkan bahwa model CNN mencapai akurasi tertinggi sebesar 95% dengan rasio pembagian data 80:10:10 dan tingkat pembelajaran 0,001. Penelitian ini menunjukkan bahwa CNN adalah metode yang menjanjikan dalam pengenalan karakter tulisan tangan Hangeul, memberikan kontribusi yang signifikan bagi bidang linguistik dan teknologi budaya.

Kata Kunci*—Artificial Neural Network, Convolutional Neural Network, Deep Learning, Pengenalan Karakter Tulisan Tangan, Hangeul.*

Abstract

The rapid increase in tourists visiting South Korea, driven by the global influence of the "Korean Wave," has sparked a growing interest in Korean culture and language. This surge in interest has highlighted the necessity for an efficient digital approach to simplify the process of recognizing handwritten Korean script characters, specifically Hangeul. This research aims to explore the effectiveness of using Convolutional Neural Network (CNN) in classifying handwritten Hangeul characters. A dataset consisting of 2,400 images, categorized into consonants and vowels, was utilized to train, validate, and test the CNN model. The study employed two different data split ratios (60:20:20, 70:15:15, 80:10:10) and two learning rates (0.001 and 0.0001) across 10 epochs to determine the most effective configuration. The results indicate that the CNN model achieved the highest accuracy of 95% with a data split ratio of 80:10:10 and a learning rate of 0.001. This research shows that CNN is a promising method for the recognition of handwritten Hangeul characters, making a significant contribution to the fields of linguistics and cultural technology.

Keywords*—Artificial Neural Network, Convolutional Neural Network, Deep Learning, Handwritten Character Recognition, Hangeul.*

DOI: 10.33050/cices.v11i1.3505 ISSN: **2356-5209**
pp. 70~81 *Online* ISSN: **2655-3058**

1. INTRODUCTION

The global influence of the "Korean Wave," known as Hallyu, has further fueled this interest, leading to a growing number of international tourists visiting South Korea [1]. This cultural phenomenon, driven by the popularity of Korean pop culture, has not only enhanced interest in South Korean entertainment but also sparked a growing curiosity about the Korean language and its script, Hangeul [2]. As more individuals engage with Korean culture, the ability to understand and use Hangeul has become increasingly important, particularly in the context of language learning and cultural immersion.

Hangeul, the Korean alphabet, is a unique and systematic alphabet, designed to be written horizontally from left to right or vertically from top to bottom [3]. However, one of the primary challenges in learning Hangeul, especially for non-native speakers, is recognizing handwritten characters due to variations in writing style, character size, and visual similarities that complicate identification [4]. Traditional methods of character recognition often require extensive manual effort and may not be efficient or accurate enough to handle the diversity found in handwritten characters. This issue highlights the necessity for a more effective and automated approach to recognize handwritten Hangeul characters, thereby facilitating language learning and cultural exchange. In recent years, advancements in machine learning, particularly in the field of deep learning, have provided promising solutions to complex problems such as image and character recognition. Convolutional Neural Network are widely used in artificial intelligence for image processing and pattern recognition, which show high accuracy in handwritten character recognition [5].

Researchers have demonstrated the impressive capability of Convolutional Neural Network (CNN) in achieving high recognition accuracy for handwritten character recognition. For example, an experiment using CNN with the NIST (National Institute of Standards and Technology) dataset has yielded remarkable results. In this research, CNN models were trained and tested using a subset of the NIST dataset, which included 1,000 training images and 200 test images. The CNN achieved an outstanding accuracy of 92.91% on the test set [6]. This high level of performance highlights the effectiveness of CNN in accurately recognizing handwritten characters, as they can learn and generalize from complex features in the data. In addition to the study involving the NIST dataset, other research has further demonstrated the effectiveness of convolutional neural network in handwritten character recognition. For instance, a study conducted by Vinh et al. (2020) focused on Vietnamese characters to evaluate CNN performance. The dataset is divided into three parts (58,000 examples for training, 7,250 examples for validation, and 7,250 examples for testing) [7]. The Vietnamese handwritten character dataset used consists of 72,500 labeled images representing 29 different Vietnamese handwritten characters. The dataset was created by combining 57,500 patterns from the NIST dataset with 15,000 unique Vietnamese character patterns. The researchers reported an impressive recognition accuracy of 97%, demonstrating the CNN's ability to effectively handle the variability inherent in handwritten characters. This research emphasizes that CNN excels in learning and extracting features from complex handwriting data, which significantly proves its superior performance in character recognition tasks, underscores the growing body of evidence supporting CNN as a powerful algorithm for handwritten character recognition.

While previous research has shown the potential of CNN in recognizing handwritten characters from different languages, there has been limited exploration of their application to Hangeul character recognition. This paper aims to address this gap by evaluating the effectiveness of CNN in classifying handwritten Hangeul characters. Using a dataset of 2,400 images categorized into consonants and vowels, the study investigates various data split ratios and learning rates to identify the most effective CNN configuration. The primary objective is to develop a robust and accurate method for recognizing handwritten Hangeul characters, thereby contributing to both linguistic research and cultural technology, and supporting the global interest in Korean language and culture.

DOI: 10.33050/cices.v11i1.3505 ISSN: **2356-5209**
pp. 70~81 *Online* ISSN: **2655-3058**

2. RESEARCH METHODS

The research process, as depicted in Figure 1, is structured into three distinct yet interconnected stages, each crucial for the successful development and assessment of the proposed model.

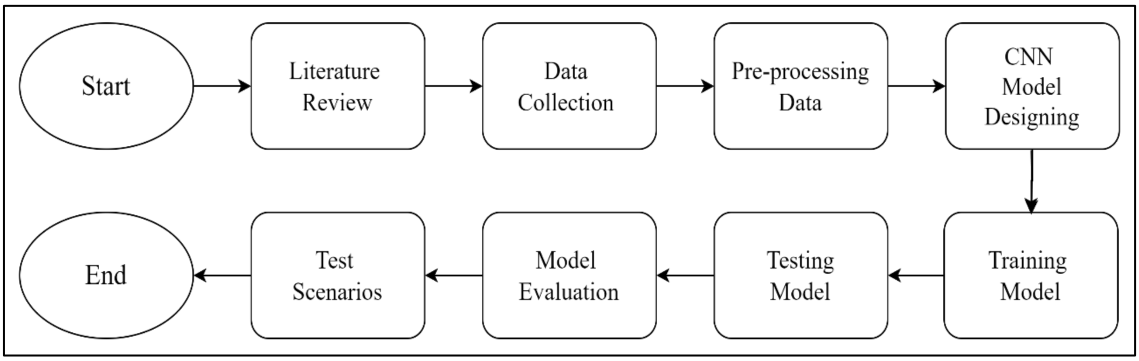

Figure 1. Stages Of The Research

The initial stage establishes the foundation for the research through a comprehensive literature review and data collection. This involves an extensive examination of existing studies and methodologies related to handwritten character recognition, with a specific emphasis on Convolutional Neural Network (CNN). The literature review identifies gaps in current research, guides the selection of appropriate techniques, and provides insights into best practices. Concurrently, a comprehensive dataset of handwritten Hangeul characters is assembled to ensure that the data is representative and relevant for subsequent stages of the research.

The main stage consists of two important tasks, which is data pre-processing and convolutional neural network model design. Data pre-processing involves techniques such as normalization and organization to prepare the raw data for effective model training, improve data quality, and enhance model performance. Simultaneously, CNN model design leverages its feature extraction capabilities to develop a robust and efficient model for recognizing handwritten Hangeul characters.

The final stage focuses on the practical application and assessment of the developed model. Training involves using the pre-processed data to train the CNN model, adjusting parameters, and optimizing performance. Following training, the model is tested on a separate validation dataset to evaluate its accuracy and generalizability. This testing phase is essential for identifying potential issues and assessing model performance with new, unseen data. The evaluation of the model’s performance is conducted using various metrics, including accuracy, precision, recall, and F1-score. This comprehensive evaluation ensures that the model meets the research objectives and provides insights into its effectiveness and potential areas for improvement.

By structuring the research process into these stages, this paper systematically addresses each aspect of model development and performance evaluation, ensuring a thorough and methodical approach to achieving the research goals.

2.1 Data Collection

This paper uses a dataset obtained based on secondary data sources collected through the official Kaggle website with the title “Handwritten Hangul Characters”. This dataset is publicly available, allowing anyone to access and utilize it. The dataset was uploaded by James Casia in 2021 [8]. The dataset consists of 2,400 images of handwritten Hangeul characters. In this study, the dataset will be downloaded from Kaggle and manually organized into 30 folders, each folder representing 80 Hangeul characters. This organization made it easy to identify and access the characters during the training process.

DOI: 10.33050/cices.v11i1.3505 ISSN: **2356-5209**
pp. 70~81 *Online* ISSN: **2655-3058**

2.2 Pre-processing Data

The preprocessing pipeline for the handwritten Hangeul character images initiates with the input of each image, originally sized at 28 × 28 pixels, which is then resized to 128 × 128 pixels. This resizing step is crucial as it enlarges the image dimensions, allowing the model to capture more intricate details and finer features of the characters, which in turn enhances the model's detection accuracy and overall performance.

Following the resizing, the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique is employed to improve the contrast of the image. CLAHE is particularly effective in enhancing image details in regions with low brightness, making subtle features more discernible. This contrast adjustment is vital for ensuring that important character details are not lost or obscured, which can be crucial for accurate recognition.

Once contrast enhancement is completed, normalization is carried out to standardize the pixel values across the image. This involves scaling the pixel intensity values from their original range (0 to 255) to a normalized range of 0 to 1. Normalization is performed by dividing each pixel value by 255. This process helps in improving the model's training efficiency and accuracy by ensuring that all input data is on a consistent scale, which facilitates better convergence during the training phase. After normalization, the image undergoes reshaping to match the input dimensions required by the model. This step ensures that the preprocessed images are formatted correctly for the model's architecture, aligning with its input layer specifications.

The final preprocessing step involves splitting the dataset into three distinct subsets which is training, validation, and testing sets. This partitioning is essential for optimizing the model's performance and mitigating the risk of overfitting. To achieve this, the dataset is divided using three different ratios: 60:20:20, 70:15:15, and 80:10:10. Each ratio represents the distribution of the 2,400 images across the training, validation, and testing sets. This stratified splitting ensures a balanced evaluation of the model's performance and aids in fine-tuning the model parameters effectively. The entire preprocessing workflow is visually summarized in Figure 2.

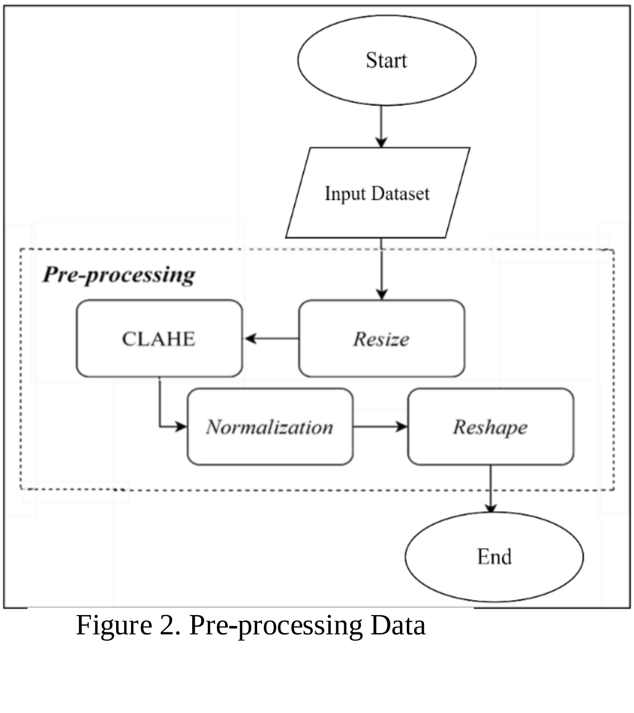

Figure 2. Pre-processing Data

2.3 Model Designing

This section presents the design of the Convolutional Neural Network (CNN) model employed for recognizing handwritten Hangeul characters. The architecture of the CNN was selected due to its demonstrated effectiveness in image recognition tasks, particularly its ability to extract spatial hierarchies and localized information, which is essential for detailed pattern

DOI: 10.33050/cices.v11i1.3505 ISSN: **2356-5209**
pp. 70~81 *Online* ISSN: **2655-3058**

recognition. The research methodology begins with the design of a CNN model, focusing on capturing and analyzing local features of the images [9]. The model consists of several layers, including convolutional layers for feature extraction and max-pooling layers for dimensionality reduction. Max pooling is a downsampling technique that reduces the spatial dimensions of an input by selecting the maximum value from a defined region, typically a small window, to retain the most prominent features [10]. These layers are crucial for extracting and condensing important features from input data [11]. Additionally, fully connected layers are included for classification. Each layer is configured with specific parameters, such as the number of filters and filter sizes, to achieve optimal performance. The decision to use a CNN model is further supported by its capability to automatically learn spatial hierarchies of features [12], making it particularly well-suited for this application.

2.4 Training and Testing

The system will be subjected to an extensive evaluation process involving training and testing across six distinct scenarios. Each of these scenarios will implement three different data split ratios: 60:20:20, 70:15:15, and 80:10:10. These ratios represent the division of the dataset into training, validation, and testing subsets, respectively. In the 60:20:20 split, 60% of the data is allocated for training, 20% for validation, and the remaining 20% for testing. Similarly, the 70:15:15 and 80:10:10 splits adjust the proportions accordingly to investigate how varying the amount of data used for training and validation impacts the system's performance.

The training data is primarily used to allow the model to learn and identify patterns within the dataset [13]. This process involves the model iteratively adjusting its internal parameters to minimize errors and improve its predictive accuracy. The validation data, on the other hand, plays a crucial role during the training process by providing a benchmark to monitor the model's performance. It helps in tuning the model and ensuring that it is neither overfitting, where the model becomes too tailored to the training data and performs poorly on new data nor underfitting, where the model fails to capture the underlying patterns in the data.

Once the model is trained and validated, the testing data is used to evaluate the final performance of the model. This step is critical as it provides an unbiased assessment of how well the model is likely to perform on unseen data, reflecting its true generalization capability. In addition to varying the data split ratios, the system will also experiment with two different learning rates 0.001 and 0.0001. The learning rate is a key hyperparameter that controls how much the model's weights are adjusted with respect to the loss gradient during each iteration of training [14]. By testing both a higher and a lower learning rate across all scenarios, the reseach aims to understand the impact of learning rate selection on the model's convergence speed and accuracy.

Each scenario will be run for a consistent duration of 10 epochs, where an epoch represents one complete pass through the entire training dataset. By keeping the number of epochs constant across all scenarios, the analysis is able to isolate and accurately evaluate the effects of different data split ratios and learning rates on the model's overall performance. This systematic approach ensures that any observed differences in performance can be attributed to the variations in data distribution and learning rate rather than other factors.

2.5 Model Evaluation

The final step is to analyze the evaluation metrics from the classification report and the confusion matrix to assess the model's performance, particularly in terms of accuracy.

DOI: 10.33050/cices.v11i1.3505 ISSN: **2356-5209**
pp. 70~81 *Online* ISSN: **2655-3058**

3. RESULTS

2.1 Data Collection

This dataset comprises a comprehensive collection of 2,400 images, each depicting handwritten Hangeul characters, which are an integral part of the Korean writing system. The characters in this dataset are categorized into a total of 30 distinct Hangeul characters, reflecting the complexity and diversity of the Hangeul script. It is actually supposed to include 40 Hangeul characters in total [15]. These characters are systematically divided into two main groups, consonants and vowels, which are foundational elements in Hangeul phonetics.

In the consonant category, there are 17 unique characters. Out of these, 14 are single consonants, representing the basic building blocks of Korean syllables. The remaining 3 consonants are classified as double consonants, which are more complex in their formation and pronunciation, and often used to convey different sounds or emphases in the Korean language. This differentiation allows for a nuanced understanding of how consonant sounds are represented in handwritten form. Similarly, the vowel group consists of 13 characters. Among these, 9 are single vowels, each representing a simple vocal sound essential for constructing syllables. The other 4 vowels are double vowels, which combine two vowel sounds to create more complex phonetic combinations. This classification enables a detailed analysis of the various vowel sounds and their handwritten representations.

By organizing the dataset into these specific categories, it supports an in-depth examination of both simple and intricate forms of Hangeul characters. This structured classification facilitates various analyses, including the recognition and differentiation of individual character forms, and the evaluation of handwriting variations for both consonants and vowels. The Hangeul or Korean alphabet used in this research is 30 characters which is ㄱ (g, k), ㄴ (n), ㄷ (d, t), ㄹ (r, l), ㅁ (m), ㅂ (b, p), ㅅ (s), ㅇ (ng), ㅈ (j), ㅊ (ch), ㅋ (k), ㅌ (t), ㅍ (p), ㅎ (h) as single consonants (자음 - Jaeum), ㄲ (kk), ㅃ (pp), ㅆ (ss) as double consonants (쌍 자음 - Ssang Jaeum), ㅏ (a), ㅑ (yes), ㅓ (eo), ㅗ (o), ㅛ (yo), ㅜ (u), ㅠ (yu), ㅡ (eu), ㅣ (i) as single (모음 - Moeum) vowels, and ㅐ (ae), ㅒ (yae), ㅔ (e), ㅖ (ye) as double (모음 - Moeum) vowels.

2.2 Pre-processing Data

The next phase of the research involves transitioning the dataset into the pre-processing stage, a critical step that is meticulously guided by the sequence of steps outlined in the flowchart presented in Figure 2. This stage is pivotal in preparing the image data for subsequent analysis and model training by ensuring that it is thoroughly cleaned, standardized, and formatted. Pre-processing serves several essential functions. Initially, it addresses issues related to data quality, such as removing noise, correcting distortions, and handling missing or incomplete data. Standardization is another key aspect, which involves resizing images to a consistent format, normalizing pixel values, and ensuring uniformity across the dataset. Proper formatting ensures that the data is compatible with the requirements of the machine learning models and algorithms that will be used in later stages.

These steps are crucial for optimizing the data's quality, ensuring that the dataset is fully prepared for model training. Effective pre-processing not only refines the data but also significantly enhances the efficiency and accuracy of the training process, ultimately leading to more reliable model performance. By systematically addressing potential issues such as noise, inconsistencies, and variability within the dataset, the pre-processing stage plays a pivotal role in setting a strong foundation for superior results in both analysis and model evaluation. This preparation is essential, as it allows the model to focus on the most relevant features, reducing the likelihood of errors during training. Figure 3 provides a visual representation of the raw,

DOI: 10.33050/cices.v11i1.3505 ISSN: **2356-5209**
pp. 70~81 *Online* ISSN: **2655-3058**

unprocessed data, emphasizing the initial state of the images before any pre-processing techniques were applied. This highlights the transformation that the data undergoes during the pre-processing stage, which is integral to achieving high-quality model outcomes.

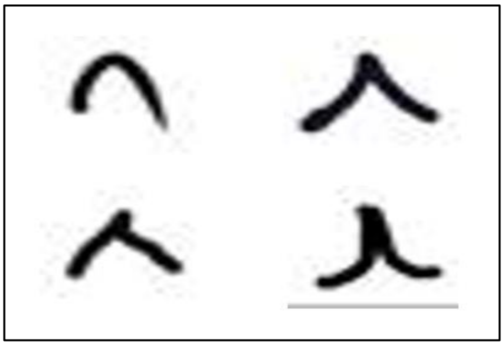

Figure 3. Real Data Before Pre-processing

In contrast, Figure 4 showcases an image that has undergone pre-processing, demonstrating the improvements in quality and consistency achieved through this critical stage. This visual comparison underscores the importance of pre-processing in preparing the dataset for effective and accurate model training.

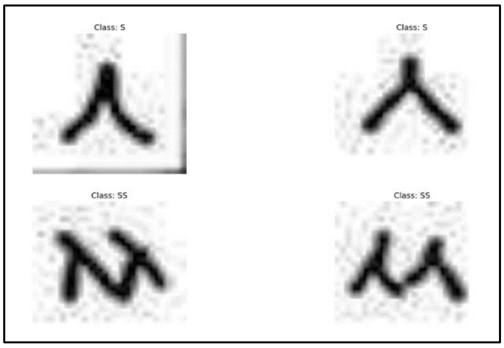

Figure 4. Data After Pre-processing

2.3 Model Designing

The diagram provided in Figure 5 visually represents the architecture of a Convolutional Neural Network (CNN) designed to enhance image recognition capabilities, which is closely aligned with the theoretical discussion in this paper. CNN model begins with a series of convolutional layers, starting with 16 Conv2D filters of size 3 × 3, followed by max pooling layers with a pooling size of 2 × 2. This sequence of convolutional and max-pooling layers is progressively stacked, with the number of filters increasing to 32, 64, and finally 128 Conv2D filters, each paired with max-pooling layers. This hierarchical arrangement allows the network to gradually learn and extract increasingly complex features from the input images, starting from basic edges and textures to more intricate patterns and structures. After the convolutional layers, the model includes a flattening layer that converts the 2D feature maps into a 1D vector, preparing the data for the fully connected layers. The architecture proceeds with a dense layer of 256 units, followed by another dense layer with 30 units, which utilizes a softmax classifier to perform the final classification. This structure showcases the CNN's capability to process visual data, enabling it to handle complex image recognition tasks efficiently.

Max-pooling plays a crucial role in this architecture by reducing the spatial dimensions of the feature maps, which not only reduces the computational load but also helps to achieve translation invariance, making the network more robust to variations in the input images. The combination of convolutional, pooling, and fully connected layers forms a powerful model capable of achieving high accuracy in image recognition tasks, as this paper seeks to

DOI: 10.33050/cices.v11i1.3505 ISSN: **2356-5209**
pp. 70~81 *Online* ISSN: **2655-3058**

demonstrate. This paper further explores the effectiveness of CNN like the one depicted, focusing on their application in various scenarios to develop an efficient and accurate image recognition system. The architecture shown in Figure 5 of the paper highlights how these CNN models are structured to maximize their performance, emphasizing their potential to advance the field of image recognition technology.

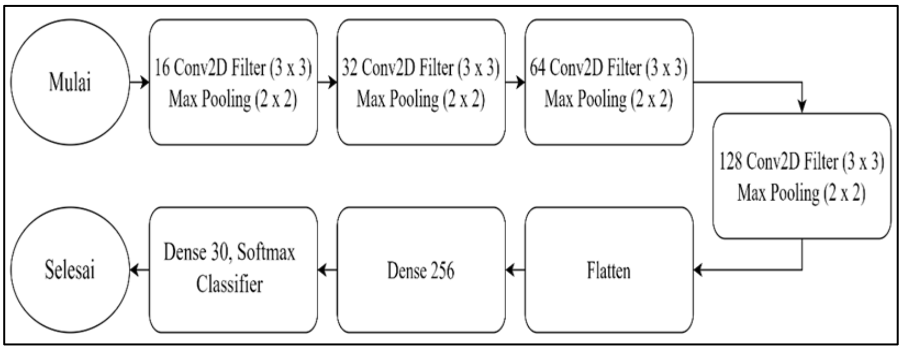

Figure 5. CNN Model

2.4 Training and Testing

The Convolutional Neural Network (CNN) model demonstrated its optimal performance during the fifth testing scenario, utilizing a dataset split of 80:10:10. This data allocation meant that 80% of the images were used for training, 10% for validation, and the remaining 10% for testing. Such a strategy provided a robust framework for the model to learn essential features, as the large training set allowed the CNN to effectively capture and generalize patterns. Simultaneously, the validation set was adequately sized to fine-tune the model's parameters and minimize the risk of overfitting, ensuring that the model remained both accurate and generalizable across unseen data.

In this scenario, the CNN was trained with a learning rate of 0.001, a critical hyperparameter that controls the pace at which the model updates its weights during the learning process. This specific learning rate was chosen to strike a balance between rapid convergence and the need for stability, enabling the model to gradually refine its predictions without overshooting the optimal solution. The gradual adjustment of weights and biases under this learning rate facilitated a stable learning process, where the model could steadily improve its accuracy over time.

As a result of this carefully calibrated setup, the CNN model achieved an impressive accuracy of 95%. This high level of accuracy indicates that the model was highly effective in recognizing patterns and features within the dataset, making it a reliable tool for the task at hand. The high accuracy achieved in this scenario, as detailed in Table 1, highlights the model's superior performance and reliability for the task of image recognition

Table 1. CNN Results

Scenarios	**Model**	**Data Split**	**Learning Rate**	**Accuracy**
1.	Convolutional Neural Network (CNN)	60:20:20	0.001	93%
2.			0.0001	88%
3.		70:15:15	0.001	91%
4.			0.0001	92%
5.		**80:10:10**	**0.001**	**95%**
6.			0.0001	94%

Confusion matrix provides an overview to see how the classification model works by comparing the model predictions with the tested data values. In Table 2, True Positive (TP) is data that is classified correctly and the result is true, True Negative (TN) is data that is classified

DOI: 10.33050/cices.v11i1.3505 ISSN: **2356-5209**
pp. 70~81 *Online* ISSN: **2655-3058**

incorrectly and the result is false, False Positive (FP) is data that is classified incorrectly, but the result is true, False Negative (FN) is data that is classified correctly, but the result is false.

Table 2. Confusion Matrix

Confusion Matrix		***Prediction Results***	
		Positive (P)	*Negative* (N)
Actual Observations	*Positive* (P)	*True Positive* (TP)	*False Negative* (FN)
	Negative (N)	*False Positive* (FP)	*True Negative* (TN)

In the confusion matrix, there are several evaluation indicators that can be calculated to evaluate the performance based on the classification model. Some of the main indicators that are commonly used in the calculation process of the confusion matrix are accuracy, precision, recall, and F1-score.

$$Accuracy = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FP} + \text{FN} + \text{TN})} = \frac{\text{Correct } Predictions}{\text{Total Predictions}} \quad (1)$$

$$Precision = \frac{\text{TP}}{(\text{TP} + \text{FP})} = \frac{\text{P}redictions\ Actually\ Positive}{\text{Total Predicted Positive}} \quad (2)$$

$$Recall = \frac{\text{TP}}{(\text{TP} + \text{FN})} = \frac{\text{P}redictions\ Actually\ Positive}{\text{Total Actually Positive}} \quad (3)$$

$$F1 - Score = \frac{(2 \times Precision \times Recall)}{(Precision + Recall)} \quad (4)$$

The following section details the confusion matrix results for the most optimal scenario derived from the Convolutional Neural Network (CNN) model, as illustrated in Figure 6. The dataset, comprising a total of 2,400 images of handwritten Hangeul characters, was divided using an 80:10:10 split ratio. This division resulted in three subsets 1,920 images for training, 240 images for validation, and the remaining 240 images for testing. The training set was utilized to train the CNN model in recognizing and classifying Hangeul characters. This phase involved adjusting the model's parameters to minimize classification errors through iterative learning. The validation set, consisting of 240 images, was employed to tune the model's hyperparameters and prevent overfitting by providing a separate dataset for evaluation during training. The test set, also comprising 240 images, served as an unbiased evaluation tool to assess the model's final performance after training.

Figure 6 presents the confusion matrix resulting from the evaluation of the CNN model on the test set. The confusion matrix is a critical tool for understanding model performance by comparing predictions to the actual labels of the test images. The analysis revealed a total of 12 classification errors out of the 240 images assessed. These errors are categorized into true positives, true negatives, false positives, and false negatives, providing a comprehensive overview of where the model succeeded and where improvements are needed. The confusion matrix offers valuable insights into the effectiveness of the CNN model in classifying handwritten Hangeul characters. By analyzing the distribution of errors, specific areas where the model's performance may be lacking can be identified. For instance, a high number of false positives or false negatives might indicate issues with certain character classes or variations

DOI: 10.33050/cices.v11i1.3505 ISSN: **2356-5209**
pp. 70~81 *Online* ISSN: **2655-3058**

within the dataset. These insights are essential for guiding future refinements of the model, such as adjusting its architecture, retraining with additional data, or employing advanced techniques to enhance accuracy.

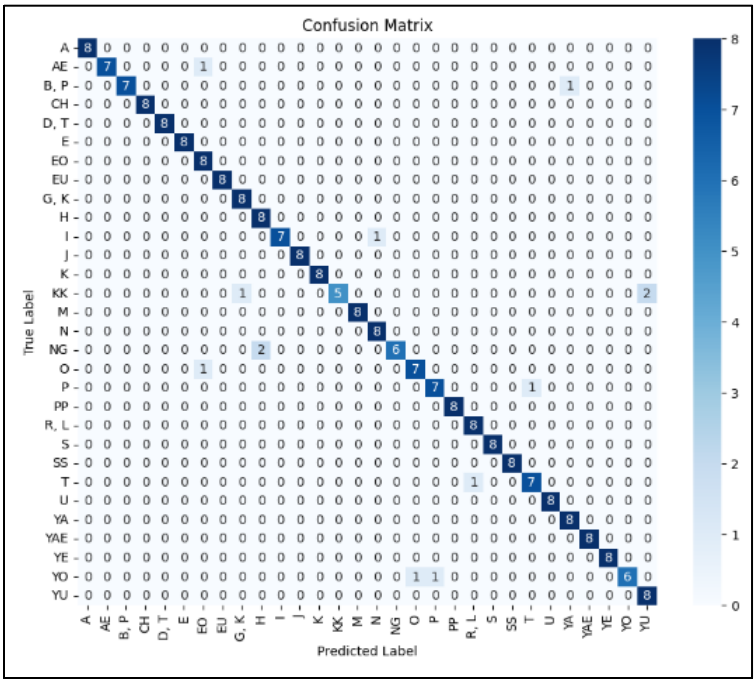

Figure 6. Confusion Matrix

Figure 7 shows the classification report for the optimal model, evaluating its performance across 30 classes of handwritten Hangeul characters, with each class having 8 images in a total of 240 test images.

Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	8
1	1.00	0.88	0.93	8
2	1.00	0.88	0.93	8
3	1.00	1.00	1.00	8
4	1.00	1.00	1.00	8
5	1.00	1.00	1.00	8
6	0.80	1.00	0.89	8
7	1.00	1.00	1.00	8
8	0.89	1.00	0.94	8
9	0.80	1.00	0.89	8
10	1.00	0.88	0.93	8
11	1.00	1.00	1.00	8
12	1.00	1.00	1.00	8
13	1.00	0.62	0.77	8
14	1.00	1.00	1.00	8
15	0.89	1.00	0.94	8
16	1.00	0.75	0.86	8
17	0.88	0.88	0.88	8
18	0.88	0.88	0.88	8
19	1.00	1.00	1.00	8
20	0.89	1.00	0.94	8
21	1.00	1.00	1.00	8
22	1.00	1.00	1.00	8
23	0.88	0.88	0.88	8
24	1.00	1.00	1.00	8
25	0.89	1.00	0.94	8
26	1.00	1.00	1.00	8
27	1.00	1.00	1.00	8
28	1.00	0.75	0.86	8
29	0.80	1.00	0.89	8
accuracy			0.95	240
macro avg	0.95	0.95	0.94	240
weighted avg	0.95	0.95	0.94	240

Figure 7. Classification Report

DOI: 10.33050/cices.v11i1.3505 ISSN: **2356-5209**
pp. 70~81 *Online* ISSN: **2655-3058**

4. CONCLUSION

In this paper, the Convolutional Neural Network (CNN) is utilized to extract features from input images by applying convolutional operations, which are particularly effective at identifying local patterns such as edges, textures, and small details within the images. The study demonstrates the effectiveness of CNN in recognizing handwritten Hangeul characters by systematically evaluating different configurations of data split ratios and learning rates using a dataset of 2,400 images. The CNN model achieved a peak accuracy of 95% with a data split ratio of 80:10:10 and a learning rate of 0.001, affirming its capability in accurately classifying complex handwritten characters. These results underscore the potential of CNNs as a robust solution for Hangeul character recognition, contributing to advancements in both linguistic research and cultural technology. The findings lay a foundation for further studies, with implications for enhancing digital tools in Korean language learning and supporting the growing global interest in Korean culture.

5. SUGGESTION

This research has identified several aspects that require further improvement, particularly concerning the dataset. One significant limitation is the insufficiency of the dataset, which may affect the generalization and robustness of the model. The current dataset, consisting of 2,400 images representing 30 handwritten Hangeul characters, is relatively small, and it does not include all 40 Hangeul characters, missing some critical elements necessary for comprehensive character recognition. To address this limitation, future research should focus on expanding the dataset to include a more diverse and complete set of Hangeul characters. This expansion could involve collecting additional handwritten samples from a wider range of individuals to capture variations in writing styles, which would improve the model's ability to generalize across different handwriting patterns. By addressing these issues, future research can create a more robust and accurate system for recognizing handwritten Hangeul characters, ultimately leading to better real-world applications and broader usability.

REFERENCE

[1] S. Lim and E. Giouvris, "Tourist arrivals in Korea: Hallyu as a pull factor," in *Current Issues in Asian Tourism: Volume II,* Routledge, 2021, pp. 9-40.

[2] Y. Kim, Ed., Introducing Korean popular culture. London and New York: Routledge, 2023.

[3] A. N. Septiawan and T. Setiadi, "Aplikasi pengenalan huruf Hangeul berbasis multimedia interaktif," Jurnal Sarjana Teknik Informatika, vol. 1, no. 1, pp. 347-357, 2013.

[4] O. S. Kwon, M. Kim, M. Park, and Y. B. Kwon, "A Cursive On-line Hangul Recognition System Based on the Combination of Line Segments" in Proc. 2nd Int. Conf. Document Analysis and Recognition (ICDAR'93), Oct. 1993, pp. 200-203.

[5] S. N. Amartama, A. N. Hidayah, P. K. Sari, and R. A. Ramadhani, "Implementasi Convolutional Neural Network (CNN) dalam Pengenalan Pola Penulisan Tangan," Prosiding Seminar Nasional Teknologi dan Sains, vol. 3, no. 1, pp. 133-138, 2024.

[6] I. Khandokar, M. Hasan, F. Ernawan, S. Islam, and M. N. Kabir, "Handwritten character recognition using convolutional neural network," in Journal of Physics: Conference Series, vol. 1918, no. 4, p. 042152, June 2021, IOP Publishing.

DOI: 10.33050/cices.v11i1.3505 ISSN: **2356-5209**
pp. 70~81 *Online* ISSN: **2655-3058**

[7] Q. V. Truong, H. D. Le, and N. T. Nhan, "Vietnamese handwritten character recognition using convolutional neural network," IAES International Journal of Artificial Intelligence, vol. 9, no. 2, p. 276, 2020.

[8] J. Casia, "Handwritten Hangul Characters," Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/wayperwayp/hangulkorean-characters/data. [Accessed: Aug. 20, 2024].

[9] V. Zhou, "CNNs, Part 1: An Introduction to Convolutional Neural Networks," Nov. 2019. [Online]. Available: https://victorzhou.com/blog/intro-to-cnns-part-1/. [Accessed: Aug. 20, 2024].

[10] D. Yu, H. Wang, P. Chen, and Z. Wei, "Mixed pooling for convolutional neural networks," in *Proc. 9th Int. Conf. Rough Sets and Knowledge Technology (RSKT'14)*, Oct. 2014, pp. 364-375.

[11] A. P. Sari, H. Suzuki, T. Kitajima, T. Yasuno, D. A. Prasetya, and N. Nachrowie, "Prediction model of wind speed and direction using convolutional neural network-long short term memory," in 2020 IEEE International Conference on Power and Energy (PECon), Dec. 2020, pp. 356-361.

[12] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo, "Exploring hierarchical convolutional features for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 11, pp. 6712-6722, 2018.

[13] A. Paullada, I. D. Raji, E. M. Bender, E. Denton, and A. Hanna, "Data and its (dis)contents: A survey of dataset development and use in machine learning research," *Patterns,* vol. 2, no. 11, 2021.

[14] L. N. Smith and N. Topin, "Super-convergence: Very fast training of neural networks using large learning rates," in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, vol. 11006, pp. 369-386, May 2019, SPIE.

[15] K. Swawikanti, "Macam-Macam Huruf Korea (Hangul), Pelafalan, Cara Membaca dan Menulisnya," 2023. [Online]. Available: https://www.ruangguru.com/blog/belajar-huruf-hangul-korea. [Accessed: Aug. 20, 2024].