SCIENCE AND INFORMATION TECHNOLOGY
SINTECH
JOURNAL

SINTECH JOURNAL | ISSN 2598-7305 | E-ISSN 2598-9642
Vol. 7 No 3 – Desember 2024 | https://s.id/sintechjournal
DOI: https://doi.org/10.31598
Publishing: Prahasta Publisher

Hangul Character Recognition of A New Hangul Dataset with Vision Transformers Model

Shana Aurelia[1], Ni Putu Sutramiani[2], Desy Purnami Singgih Putri[3]

[1,2,3]Department of Information Technology, Engineering Faculty, Udayana University
Udayana University Jimbaran Street, Bali, Indonesia

e-mail: aurelia_na037@student.unud.ac.id[1], sutramiani@unud.ac.id[2], desysinggihputri@unud.ac.id[3]

Received : November, 2024	Accepted : December, 2024	Published : December, 2024

Abstract

This study aims to develop a Vision Transformers (ViT) model for recognizing Korean characters (Hangeul) in response to the growing interest in learning the Korean language and Korean culture in Indonesia. The research methodology involves training the ViT model using a comprehensive dataset of 29,636 base Korean characters. The ViT model has achieved a significant level of accuracy with the score of 93% in recognizing base Korean characters. By integrating deep learning, this study is expected to make a positive contribution to the development of language learning tools for Korean character recognition, unlocking its potential for applications and systems based on the Korean language.

Keywords: *ViT, hangul, character, recognition, dataset*

Abstrak

Penelitian ini bertujuan untuk mengembangkan model Vision Transformer (ViT) untuk mengenali karakter Korea (Hangul) sebagai respon terhadap meningkatnya minat dalam mempelajari bahasa dan budaya Korea di Indonesia. Metodologi penelitian adalah pelatihan terhadap model ViT menggunakan dataset komprehensif sebanyak 29.636 karakter dasar Korea. Model ViT telah mencapai tingkat akurasi yang signifikan sebesar 93% dalam mengenali karakter dasar Korea. Dengan mengintegrasikan deep learning, penelitian ini diharapkan dapat memberikan kontribusi positif terhadap pengembangan alat pembelajaran bahasa untuk pengenalan karakter bahasa Korea, meningkatkan potensi pengaplikasiannya dalam sistem-sistem berbasis bahasa Korea.

Kata Kunci: *ViT, hangul, karakter, pengenalan, dataset*

1. INTRODUCTION

Human communication occurs through spoken and written means. Written communication across countries employs specific characters that reflect cultural heritage, serving as unique identifiers for each nation. Korea, for instance, uses distinctive characters called Hangeul.

A 2018 survey revealed that approximately 44% of test-takers cited studying in Korea as their primary reason for taking the Test of Proficiency in Korea (TOPIK), followed by 25% who wished to assess their Korean language skills [1]. This data indicates a growing interest in learning Korean and studying in Korea globally, including in Indonesia. With the establishment of Korean language institutes worldwide, including in Indonesia, more people have formal access to learning Korean. This underscores the importance of developing technology to support Korean language learning, particularly in recognizing Korean characters (Hangeul).

Moreover, Indonesia is known for its significant number of Korean Wave (Hallyu) enthusiasts. In 2021, the Korean Foundation for International Cultural Exchange (KOFICE) ranked Indonesia fourth globally for Korean Wave fans [2]. The Korean Wave, encompassing music, drama, film, and fashion, has shaped popular culture trends in Indonesia. Korean culture, including K-Pop and K-Drama, has gained substantial popularity in Indonesia, especially among the youth. This is evident from the rising sales of Korean music albums and concert tickets, as well as the increasing number of Korean language courses in Indonesia. This trend reflects the growing Indonesian interest in Korean culture.

In this context, advancements in information technology (IT) and artificial intelligence (AI) have enabled the development of increasingly sophisticated and accurate recognition systems. Deep learning, a notable AI technique, has proven effective in various recognition and classification tasks. Applying deep learning to recognize Korean characters (Hangeul) is both relevant and urgent. Deep learning can identify the unique features of Korean characters by learning complex patterns from data. Previous research has extensively explored Korean character recognition [3]. The goal is to significantly contribute to applications and systems that support Korean language learning, both within and outside Korea.

Vision Transformers (ViT), is a deep learning architecture, that have shown success in image processing tasks. Regarding deep learning, research on utilizing deep learning methods for character recognition across various scripts has been conducted by several authors. [4] applied deep learning to recognize Devanagari characters. [5] used ensemble machine learning methods for alphabetical and numeric character recognition. Other character recognition studies include those by [6], [7], [8], [9], and [10], focusing on Javanese, Chinese, and Bangla scripts.

Other recent advancements in character recognition have employed Transformer models too. [11] developed Transformer-based Optical Character Recognition (TrOCR), [12] introduced a Naïve Bayes Classifier Optical Character Recognition (OCR), [13] created a Hybrid Convolutional Recurrent Neural Network (CRNN), and [14] proposed Decoder-only Transformer for Optical Character Recognition (DTrOCR). These models have significantly improved text recognition in various contexts, including Chinese character recognition. These studies provide a strong foundation for developing Korean character recognition methods using deep learning approaches such as Vision Transformers (ViT) or Convolutional Neural Networks (CNN).

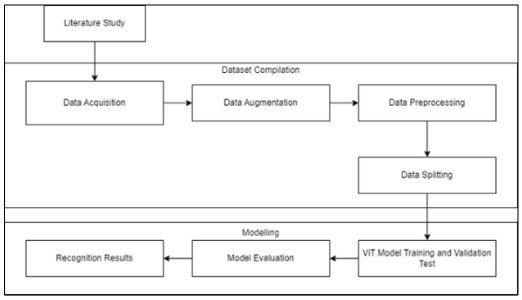
While ViTs have demonstrated impressive results in various image recognition tasks, their application to Korean character recognition, especially using a newly compiled dataset of base Korean characters, remains relatively unexplored. This research aims to bridge this gap by fine-tuning ViT models on this dataset to improve the accuracy and efficiency of Korean character recognition systems. Unlike previous studies that focused on various languages or character recognition scenarios, this study's goal is to enhance Korean character recognition's accuracy using the latest method of deep learning.

By leveraging ViT's strengths in character recognition, this research intends to fill gaps in the literature on Korean character recognition specifically base Korean characters. The method involves training ViT models using a new

compiled dataset of base Korean characters and employing strategic experimental approaches by the method of fine tuning. By combining Vision Transformers technology with specialized development strategies, this research aims to achieve high accuracy results in recognizing base Korean characters. Consequently, this study hopes to contribute in developing innovative models and systems that support Korean language learning and enhance the Korean Wave experience, particularly in Indonesia.

2. RESEARCH METHOD

A structured and planned research process is essential for conducting an effective study. The research flow applied in the study is illustrated in the diagram below.

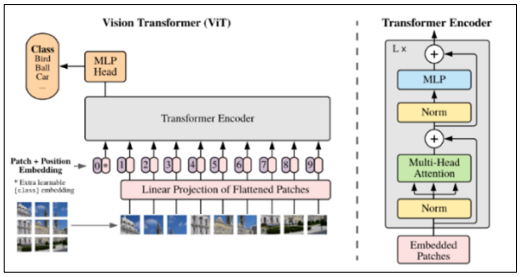


Picture 1. Research Flow Diagram
[Source: Private]

Figure 1 shows the research flow diagram implemented in this study. The research begins with literature studies, followed by the data compilation process for Korean characters (Hangul). The dataset is then preprocessed, augmented, and lastly divided (training and testing data). The next steps involve training and developing the ViT model, which includes the hyperparameters model fine-tuning. Finally, the last steps are model evaluation and the character recognition results obtained for the learning process.

2.1 Vision Transformers

The Vision Transformers (ViT) architecture leverages a self-attention mechanism to analyze and process images [15]. The core architecture comprises multiple transformer blocks, each featuring two sub-layers: a multi-head self-attention layer and a feed-forward layer.



Picture 2. ViT General Architecture Based on The Reference Journal
[Source: Dosovitskiy]

The Vision Transformer (ViT) architecture shown in Figure 2 was introduced in the research paper "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale" by [16] and has achieved competitive performance in a variety of computer vision tasks [17], including image classification, semantic image segmentation, and object detection.

In this architecture, the process begins with the patch and position embedding stage. The input image is separated into small pieces called patches, which are then flattened into vectors and projected into higher-dimensional representations using linear projection. Each patch's position is embedded to help the model recognize the sequence and location of each patch in the image. Additionally, a special class embedding is added to assist in the classification process. In the transformer encoder stage, the projected and embedded patches are processed through the transformer encoder. The transformer encoder is consisted of several layers (denoted by Lx, indicating L repetitions). Each layer has key components:

1) Multi-Head Attention: Measures the attention each patch should receive, determining the importance of each patch relative to others.
2) Norm: Normalization is performed after the attention process to maintain training stability and efficiency.
3) MLP (Multi-Layer Perceptron): The vectors are further transformed through the MLP, followed by another normalization.

This process is repeated L times to maximize the understanding of the input image.

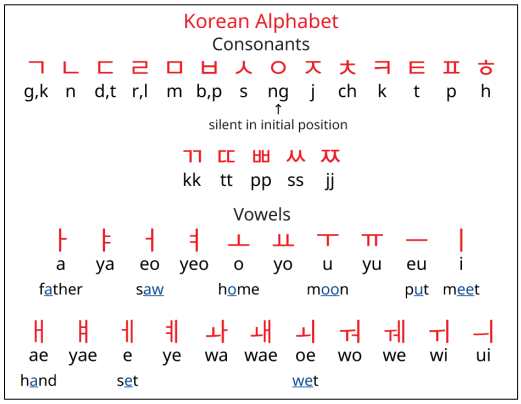
The final stage is the MLP Head. After the transformer encoder, the output is a representation of the processed image. This representation is then fed into the MLP Head,

responsible for the final classification. The MLP Head produces the class prediction for the image, such as the type of bird, type of ball, car, etc.

The model's effectiveness depends on choices such as dataset-specific hyperparameters, network depth, and the optimizer. Generally, compared to ViT models, CNN models are more simple being optimized for specific tasks. The referenced paper suggests combining ViT with a CNN front end to address this gap. In this study, the focus will be on applying the ViT model itself to the collected dataset to test the model's configuration against the author's dataset.

2.2 Base Korean Characters

Korean, or Hangeul (한글), is the writing system used for the Korean language. This script was developed by King Sejong in the 15th century and officially announced in 1443.



Picture 3. Base Korean Characters (Hangul) [Source: thinkzone.wlonk.com]

Hangeul consists of 40 alphabetic characters known as "hangeul" or "hangeul jamo," including 19 consonants and 21 vowels. Each hangeul character representing a vowel or consonant sound is combined to form words and sentences. Recognizing Korean characters is crucial in various applications, such as handwriting recognition, text recognition in images, and automatic translation [18].

This research used a comprehensive number of 29,636 base Korean characters curated and compiled both through online and offline methods. To ensure the dataset quality and diversity, the selection criteria focused on well-cropped base Korean characters images. The aim was mostly to get a minimum of 10 consonants and 10 vowels each dataset type, with a maximum of 40 classes in total, so a more diverse classes distribution is allowed.

2.3 ImageDataGenerator

ImageDataGenerator is a class in the Keras library used for real-time image data augmentation. Data augmentation is a method to increase dataset diversity by creating new variations from existing data [19]. This technique helps trained models generalize better (not just perform well on training data) and prevents overfitting.

The general steps in the augmentation process using this technique begin with creating an instance of the ImageDataGenerator object and configuring augmentation parameters to generate data variations such as rotation, scaling, zooming, skewing, flipping, and more [20]. These parameter configurations must align with the data characteristics, ensuring the dataset's meaning is preserved. For instance, Korean characters should not be excessively rotated to maintain their integrity. Besides parameter configuration, the batch size of the dataset can also be specified during the data augmentation process. The final step is training with the augmented data to evaluate whether the results meet the objectives. If not, the data augmentation process can be repeated.

2.4 Fine Tuning

This process is a method of implementing a pre-trained model to a specific task to enhance its performance or adjust it to similar tasks [21]. A simple example of a fine-tuning process that can be performed is changing hyperparameters.

Hyperparameters are parameters that cannot be learned by the model during training, but they can be set before or during the training process. Common hyperparameter examples in deep learning models include number of epochs, learning rate, number of layers and units in each layer, type of activation function, and others.

The strategy for changing hyperparameters is done through experimentation with different values to see their impact on model performance. This process can help improve accuracy and generalize the model on a specific dataset or in the context of tasks that have been done before or more specifically.

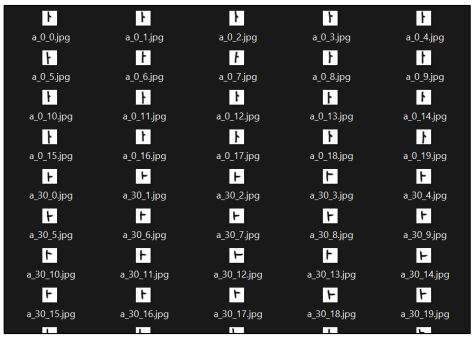
3. RESULT AND DISCUSSION

3.1 Data Collection

First initial data was collected online through a literature review conducted on the Kaggle platform. The Hangul character dataset obtained was published by James Casia and comprises 30 classes, each containing 80 variations of handwritten characters [22].

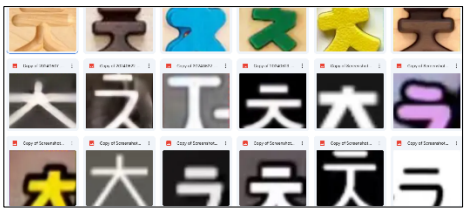
The other primary data was collected offline, written on paper using various colored pens or markers by fellow Udayana University's students. These handwritten Hangul characters were written on clean white HVS paper and subsequently scanned using a smartphone scanner app. The scanned images were then categorized and stored in folders based on the corresponding character class.

Additional primary data acquired online, was from diverse digital platforms, including social media and educational websites such as YouTube and Korean character education websites. These data sources were accessed through search engines like Google. Examples of the collected data used in this research are as shown in the figures below.



(a)

(b)



(c)

Picture 4. Example of Online Character Data (a), Example of Offline Handwritten Character Data Scan Results, and (c) Example of Character Data Results on Other Media Obtained Online
[Source: Private]

Figure 4 shows the initial data collection results. Image (a) was obtained online, comprising a Hangul character dataset with 80 variations per class (30 classes in total). Image (b) was collected offline through handwritten data from 15 individuals (1 Korean, 14 Indonesian), aiming for 15 variations per class (40 classes in total). Image (c) was compiled online from various media, including images and videos, with 100-104 variations per class (24 classes in total).

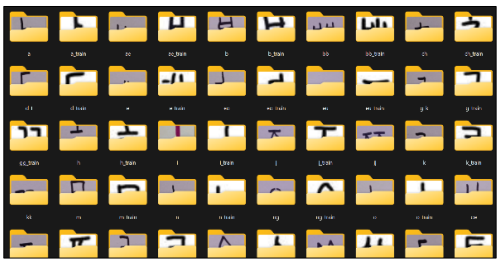
Image (a) is a sample of one character class, "A" (ㅏ), from the online dataset. Images (b) and (c) represent individual variations of data collected offline (handwritten on paper) and online (from various digital media) respectively.

Table 1: Total Image Datasets Per Alphabets Collected

Total Alphabet Images			
Alphabets	Images		
	Hangul Character	Multi-Media (Papers)	Multi-Media (Others)
a	80	15	61
ae	80	15	-
b	80	15	61
bb	80	15	-
ch	80	15	61
d	80	15	61
e	80	15	-
eo	80	15	61
eu	80	15	61
g	80	15	61
h	80	15	61
i	80	15	61
j	80	15	61
jj	-	15	-
k	80	15	61
kk	-	15	-
m	80	15	61
n	80	15	61
ng	80	15	61
o	80	15	61

Total Alphabet Images			
Alphabets	Images		
	Hangul Character	Multi-Media (Papers)	Multi-Media (Others)
oe	-	15	-
p	80	15	61
r	80	15	61
s	80	15	61
ss	80	15	-
t	80	15	61
tt	-	15	-
u	80	15	61
ui	-	15	-
wa	-	15	-
wae	-	15	-
we	-	15	-
wi	-	15	-
wo	-	15	-
ya	80	15	61
yae	80	15	-
ye	80	15	-
yeo	-	15	61
yo	80	15	61
yu	80	15	61

Table 1 above shows the specific detail count of alphabet images in the dataset collected through both offline and online methods. These alphabet images in the table above have not been preprocessed through the image augmentation process. The image data (alphabets) is the result of a combination of the three types of datasets where there they are differed into different classes count, resulting in a total of 4,464 initial alphabets images.



Picture 5. Data Folder Creation
[Source: Private]

The collected data was then organized into individual folders, one for each character class, as illustrated in Figure 5. Each class folder contains various image variations acquired from both online and offline sources. To facilitate the augmentation process, these images are further categorized based on data type: Hangul character dataset and multi-media dataset (including paper-based and other media).

3.2 Data Preprocessing

During the online or offline data collection process, the images obtained were initially captured as scanned or screenshoted pictures from various devices. These screenshots often contained unnecessary elements or background noise that might hinder the character recognition process. Therefore, a cropping procedure was applied to isolate the desired characters and remove irrelevant parts of the image. The following is an example of the cropping process carried out during data collection.

(a)

(b)

Picture 6. Before Image Cropping (a) and After Image Cropping to Focus on The Korean Character (b)
[Source: Private]

Figure 6 shows the image cropping process. Image (a) shows the screenshot before cropping, while image (b) displays the cropped image. In this example, the cropping is performed to isolate the character “EO” (ㅓ) for data collection purposes. This process is repeated numerous times until the desired amount of data is acquired.

3.3 Data Augmentation

Data augmentation is performed to achieve two primary objectives; to increase the variety of classes that have fewer samples compared to other classes, and to expand the overall dataset (both Hangul character data and multi-media data) to enhance the accuracy of the model. This process aims to balance the distribution of Korean character images within the dataset. The tables below shows the result of each augmented datasets total numbers.

Table 2: Hangul Character Dataset Details
[Source: Private]

Before Augmentation		After Augmentation	
Classes	Images	Classes	Images
30	2.400	30	16.483

Table 2 above shows that the image augmentation process on the Hangul character dataset has successfully increased the overall count of images in the dataset. The image data per class has grown to 549 images per class, resulting in a total of 16,483 images for all 30 classes after data augmentation.

Table 3: Multi-Media Dataset on Papers Details
[Source: Private]

Before Augmentation		After Augmentation	
Classes	**Images**	**Classes**	**Images**
40	600	40	3.000

Table 3 above shows that the image augmentation process on the paper-based multi-media dataset has successfully increased the overall count of images in the dataset. The image data per class has grown to 75 images per class, resulting in a total of 3,000 images after data augmentation.

Table 4: Multi-Media Dataset on Other Media Details
[Source: Private]

Before Augmentation		After Augmentation	
Classes	**Images**	**Classes**	**Images**
24	1.464	24	10.153

Table 4 above shows that the image augmentation process on the 'other media' portion of the multi-media dataset has successfully increased the overall count of images in the dataset. The image data per class has grown to 420 images per class, resulting in a total of 10,153 images after data augmentation.

3.4 Modelling

A) Hyperparameter Scenarios

A total of 36 experiments were conducted. The focus was on 3 types of datasets: a public dataset from Kaggle as the Hangul Character Dataset, a private dataset combining paper-based media and other media as the Multi-Media Dataset, and a combined dataset from the Hangul Character Dataset and the private paper-based and other media from the Multi-Media Dataset, as the New Hangul Dataset. Each experiment batch was conducted with 15 and 25 epochs to observe the impact of different values, for every 6 numbers of different hyperparameters as the example shown in Table 4.

Table 5: Experiment Batch Scenarios
[Source: Private]

No	Part	Dataset	Epochs
1	**I, III, V**	**Hangul Character Dataset**	**15 Epochs**
2			
3			
4			
5			
6			
1	**II, IV, VI**		**25 Epochs**
2			
3			
4			
5			
6			

The table 5 above presents the 6 different values of corresponding hyperparameters, which were determined based on trials and errors to achieve the desired results for the research objectives. Each corresponding hyperparameters is shown in table 5 below.

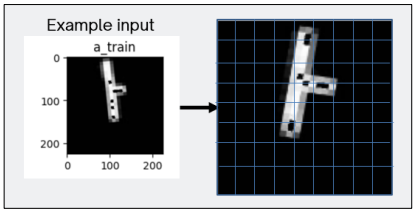
Table 6: Hyperparameter Scenario Experiments
[Source: Private]

Batch Size	Learning Rate	Dropout Rate	Weight Decay	Early Stopping Patience
16	0.01	0.2	1e-3	5
16	0.01	0.3	1e-5	10
32	0.01	0.2	1e-3	5
32	0.01	0.3	1e-5	10
64	0.01	0.2	1e-3	5
64	0.01	0.3	1e-5	10

Table 6 above shows the differences in the values of each hyperparameters used, which are epochs, batch size, dropout rate, weight decay, and early stopping patience. Hyperparameters such as learning rate were fixed at 0.01 because this value yielded the best results (accuracy above 50%) in all previous experiments. Batch size was limited to 3 values: 16, 32, and 64.

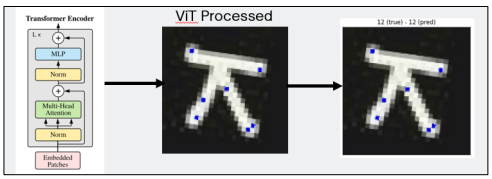
B) ViT Modelling

To understand the ViT model's operation on base Korean character recognition, please see the visualization below.



Picture 7. Input Image Divided into Smaller Patches of Smaller Images

In Figure 7 above, the input Korean character image (ㅏ), is initially divided into smaller patches. These patches are then embedded into a high-dimensional space, capturing their visual features.

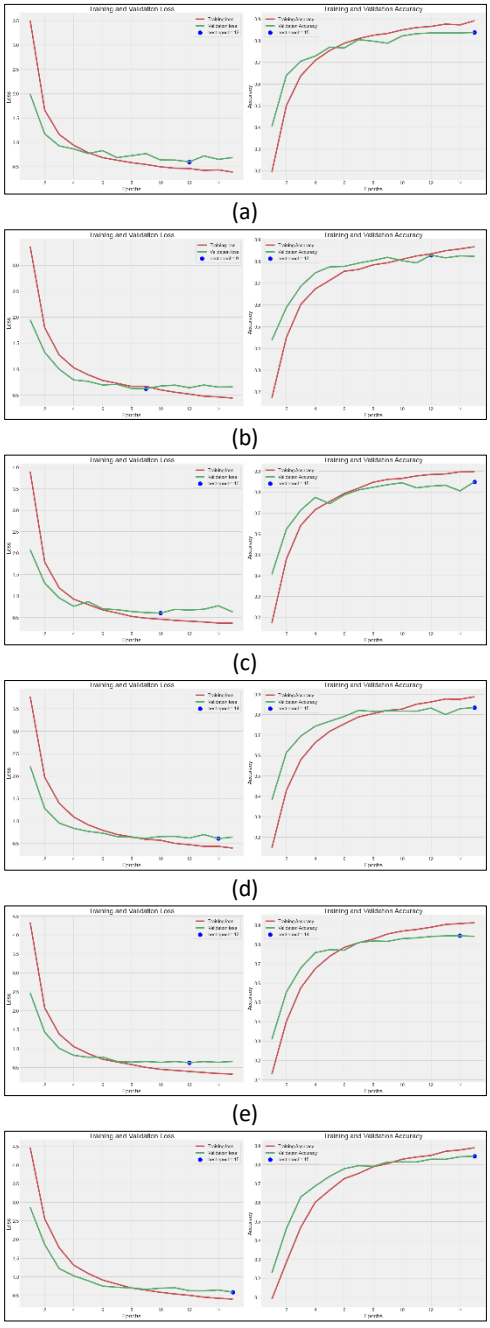


Picture 8. Other Input Image Example, Processed with ViT Algorithm and Resulted in Prediction

Then with the Figure 8 above, with another input image example, during the ViT modelling process, positional encoding is added to provide spatial information about the relative positions of the patches. The transformer encoder processes these embedded patches, leveraging self-attention to weigh the importance of different parts of the image. This attention mechanism allows the model to focus on the most relevant features for accurate character recognition. Finally, a classification layer predicts the character class, outputting the recognized Korean character.

C) Training and Validation Test Results

The results for training and validation experiment tests, was visualized for every experiment batches, from part I until part VI. Every part was shown in the graph of history training for the increase and decrease of the training and validation test to compare the accuracy and loss results. The accuracy values for part I until part IV are in the ranges between 52% to 76%. Below are the graphs result examples for part V and VI only because the accuracy values are over than 80%.



(a)

(b)

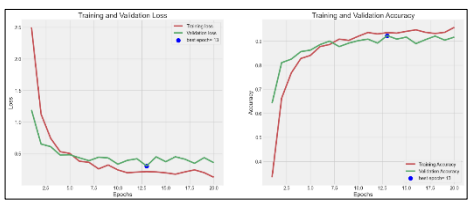
(c)

(d)

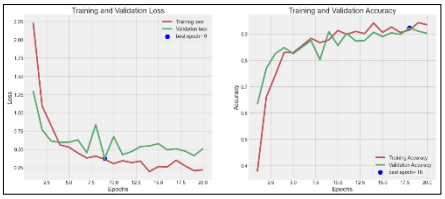
(e)

(f)

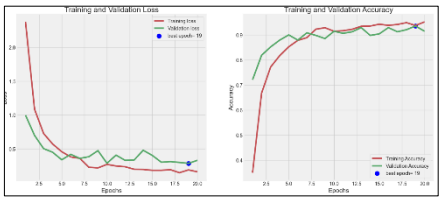
Picture 9. Results of Training and Validation of Research Experiment Part V
[Source: Private]



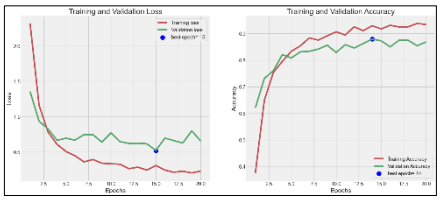
(a)



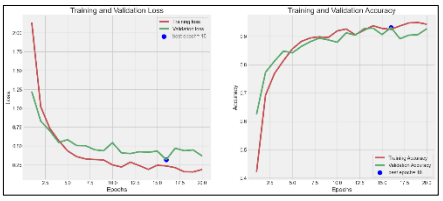
(b)



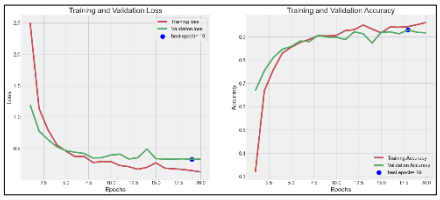
(c)



(d)



(e)



(f)

Picture 10. Results of Training and Validation of Research Experiment Part VI
[Source: Private]

Figure 7 and 8 above presents the training and validation test results for the experimental scenarios in Part V and Part VI respectively. Here, each model in Part V achieved an accuracy ranging from 82% to 85%. Each model in Part VI achieved an accuracy ranging from 86% to 93%.

Training loss consistently decreased, indicating successful learning. Validation loss also steadily decreased and at a certain epoch became stable. Training accuracy improved steadily, while validation accuracy continued to increase to show the successful learning and generalization.

Based on the training history of each model with the specified hyperparameters, it can be concluded that each model in Part V and VI successfully learned from the training data, as shown by the increase in accuracy and decrease in loss on the training data. When comparing the results of Part I with Part II and Part III with Part IV, there is a clear increase in accuracy of 20% to 40%. This suggests that the type of data used, whether it's variations in data objects or the results of data augmentation, has a significant impact.

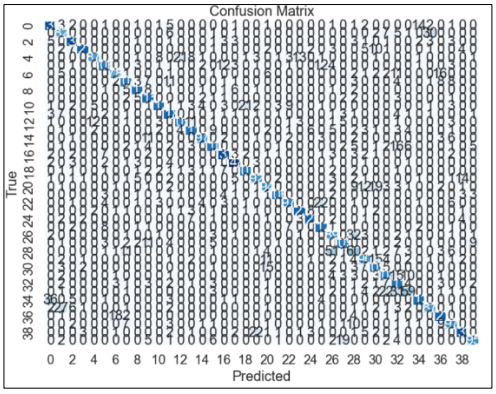
D) Model Evaluation Results

Based on the history results presented before, only the results of Part V and VI will be shown in this section, as these parts achieved an accuracy of over 80%.

While the training history provides an analysis of the learning and generalization of the data, the classification report and confusion matrix offer a more in-depth look at the performance of the research model for each Korean character class. A detailed explanation has been provided in Chapter 2, so the results of Parts V and VI, as the best results in the experimental scenarios, will be presented. The following is the classification report and confusion matrix for Part V.

	precision	recall	f1-score	support
a_train	0.78	0.88	0.83	77
ae_train	0.71	0.93	0.81	75
b_train	0.95	0.86	0.91	72
bb_train	0.95	0.91	0.93	82
ch_train	0.83	0.83	0.83	76
d_train	0.83	0.77	0.80	82
e_train	0.69	0.84	0.76	81
eo_train	0.75	0.92	0.82	72
eu_train	0.87	0.91	0.89	86
g_train	0.81	0.87	0.83	67
gg_train	0.76	0.84	0.80	75
h_train	0.86	0.77	0.81	82
i_train	0.91	0.90	0.91	81
j_train	0.91	0.87	0.89	92
k_train	0.88	0.96	0.92	77
m_train	0.76	0.74	0.75	86
n_train	0.89	0.92	0.90	76
ng_train	0.93	0.86	0.89	79
o_train	0.81	0.81	0.81	79
p_train	0.81	0.76	0.78	83
r_train	0.91	0.84	0.87	82
s_train	0.87	0.96	0.91	69
ss_train	0.96	0.91	0.94	80
...				
accuracy			0.85	2400
macro avg	0.85	0.85	0.85	2400
weighted avg	0.85	0.85	0.85	2400

(a)



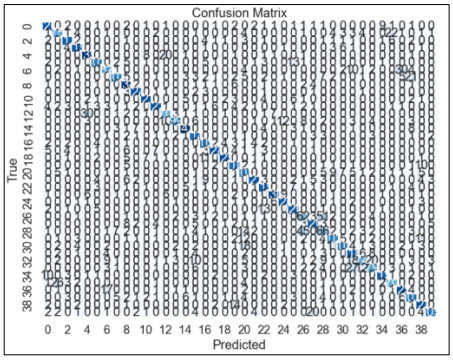
(b)

Picture 11. Classification Model Evaluation Results Report; Accuracy, Precision, Recall, F1-score (a) and Confusion Matrix (b), Part V
[Source: Private]

Figure 9 presents the evaluation results of one of the models in the conducted research, including accuracy, precision, recall, F1-score, and confusion matrix. Both evaluation metrics show that the results obtained have an accuracy of up to 85%, meaning that the model is quite good.

	precision	recall	f1-score	support
a_train	0.77	0.94	0.85	18
ae_train	0.80	1.00	0.89	16
b_train	1.00	0.93	0.97	15
bb_train	1.00	0.94	0.97	16
ch_train	0.92	0.85	0.88	13
d_train	0.71	0.77	0.74	13
e_train	0.86	1.00	0.93	19
eo_train	1.00	1.00	1.00	15
eu_train	0.85	1.00	0.92	17
g_train	0.96	0.96	0.96	25
gg_train	1.00	0.94	0.97	16
h_train	1.00	0.94	0.97	16
i_train	0.89	0.89	0.89	18
j_train	0.86	0.86	0.86	14
k_train	0.93	1.00	0.97	14
m_train	0.95	0.90	0.93	21
n_train	0.92	1.00	0.96	12
ng_train	0.93	0.93	0.93	14
o_train	1.00	0.85	0.92	20
p_train	1.00	0.88	0.94	17
r_train	1.00	1.00	1.00	13
s_train	1.00	1.00	1.00	20
ss_train	1.00	0.87	0.93	15
...				
accuracy			0.92	480
macro avg	0.92	0.92	0.91	480
weighted avg	0.92	0.92	0.92	480

(a)



(b)

Picture 12. Classification Model Evaluation Results Report; Accuracy, Precision, Recall, F1-score (a) and Confusion Matrix (b), Part VI
[Source: Private]

Figure 10 presents the evaluation results of one of the models in the conducted research, including accuracy, precision, recall, F1-score, and confusion matrix. Both evaluation metrics show that the results obtained have an accuracy of up to 93%, meaning that the model is already very good for the model training results expected.

E) Recognition Results

The obtained results from the model training and validation are character recognition results displayed with the predicted character. An example of the recognition results from one of the experiments is as follows.

Picture 13. Hangeul Character Recognition Results with Fine-Tuned Model ViT
[Source: Private]

Figure 11 demonstrates that the model has successfully recognized Hangeul characters according to the trained classes. With an accuracy of 90%, there are inevitably 1 to 2 errors in the trials, thus requiring further research to improve accuracy.

While the previous research on Devanagari characters uses the idea of fine-tuning the model [4], this research could contribute in the process of fine-tuning hyperparameters for the ViT model, using specific new comprehensive dataset of varied specifics, with a total of 29,636 images. As stated, the aim of this research was for training the model to achieve high accuracy in recognizing only base Korean characters and

present new potential ways of leveraging ViT model.

However, with the experiments conducted in this research, the obtained results are sufficient to find answers to the research problems. The following table of experimental results includes the training and accuracy values of each focus hyperparameter.

Table 7: Detailed Table of Training and Validation Test Results with The Experiment Scenarios [Source: Private]

No	Part	Dataset	Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	V	New Hangul Dataset	15 Epochs	89%	0.3714	84%	0.7492
2				87%	0.4441	82,25%	0.8210
3				90%	0.3655	85%	0.7897
4				88,43%	0.3923	83,42%	0.7320
5				90,69%	0.3327	84,13%	0.6503
6				89%	0.4004	84,46%	0.6525
7	VI		25 Epochs	94,14%	0.2771	92%	0.6849
8				92,61%	0.1912	90%	0.6565
9				95,33%	0.1598	91,23%	0.3303
10				93%	0.2295	87%	0.6394
11				94%	0.1866	93%	0.3672
12				95%	0.1237	91,60%	0.3381

Table 7 above shows the highest accuracy achieved was 93%. This result was obtained from experiment scenario VI with 25 epochs, using the dataset of a combination between privately collected dataset and the publicly available Kaggle dataset, yielding the best outcome among all trials.

4. CONCLUSION

As the results shown in the tables above, specifically on table 6, it can be concluded that the experiments conducted using the New Hangul dataset compiled, is compatible with the proposed hyperparameters and the ViT configuration model. For the New Hangul Dataset in total of 29,636 images, using the ViT base model, it is shown that the accuracy values are able to reach 90%, to give the results that is successful in learning and generalizing new data for the Base Korean characters.

However, it is important to acknowledge certain limitations of this research. Firstly, the dataset, while comprehensive, may not fully capture the diversity of more handwriting styles and variations encountered in real-world scenarios. Secondly, the ViT model, though powerful, may be computationally expensive for larger datasets and complex tasks. Future research could be done to explore more efficient architectures or techniques to address these limitations.

Based on the experiments from this research, it is suggested that there is improvement for another Base Korean dataset to be used on the training and validation test experiment, using the ViT model, to make a character recognition project for educational purposes. There is also the possibility of using another method using the proposed dataset compilation idea in this research for other related character recognition systems.

ACKNOWLEDGEMENTS

The author would like to thank Udayana University students, other private respondents, and the online sources, for their invaluable contribution in providing both handwritten and cropped images data for this research. We also express our gratitude to Udayana University for providing the necessary resources and facilities to complete this research.

REFERENCES

[1] L. Yoon, “Leading reasons for taking the Test of Proficiency in Korean (TOPIK) in 2018.” Accessed: Jul. 10, 2023. [Online]. Available:

https://www.statista.com/statistics/1057989/south-korea-reasons-for-taking-the-korean-language-test/

[2] Henry and Dinny Mutiah, "Indonesia Tempati Urutan ke-4 Penggemar Korean Wave Terbesar di Dunia," Liputan6.com. Accessed: Jul. 10, 2023. [Online]. Available: https://www.liputan6.com/lifestyle/read/4678671/indonesia-tempati-urutan-ke-4-penggemar-korean-wave-terbesar-di-dunia

[3] Radikto and Rasiban, "Pengenalan Pola Huruf Hangeul Korea Menggunakan Jaringan Syaraf Tiruan Metode Backpropagation dan Deteksi Tepi Canny," *Jurnal Pendidikan dan Konseling*, vol. 4, no. 5, pp. 1–10, 2022.

[4] B. Yadav, A. Indian, and G. Meena, "HDevChaRNet: A deep learning-based model for recognizing offline handwritten devanagari characters," *Journal of Autonomous Intelligence*, vol. 6, no. 2, 2023, doi: 10.32629/jai.v6i2.679.

[5] S. R. Zanwar, Y. H. Bhosale, D. L. Bhuyar, Z. Ahmed, U. B. Shinde, and S. P. Narote, "English Handwritten Character Recognition Based on Ensembled Machine Learning," *Journal of The Institution of Engineers (India): Series B*, Oct. 2023, doi: 10.1007/s40031-023-00917-9.

[6] Irham Ferdiansyah Katili, Mochamad Arief Soeleman, and Ricardus Anggi Pramunendar, "Character Recognition of Handwriting of Javanese Character Image using Information Gain Based on the Comparison of Classification Method," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 1, pp. 193–200, Feb. 2023, doi: 10.29207/resti.v7i1.4488.

[7] M. B. Bora, D. Daimary, K. Amitab, and D. Kandar, "Handwritten Character Recognition from Images using CNN-ECOC," in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 2403–2409. doi: 10.1016/j.procs.2020.03.293.

[8] V. Pomazan, I. Tvoroshenko, and V. Gorokhovatskyi, "Handwritten Character Recognition Models Based on Convolutional Neural Networks," 2023.

[9] M. M. Khan, M. S. Uddin, M. Z. Parvez, and L. Nahar, "A squeeze and excitation ResNeXt-based deep learning model for Bangla handwritten compound character recognition," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 3356–3364, Jun. 2022, doi: 10.1016/j.jksuci.2021.01.021.

[10] D. Gui, K. Chen, H. Ding, and Q. Huo, "Zero-shot Generation of Training Data with Denoising Diffusion Probabilistic Model for Handwritten Chinese Character Recognition," *Computer Vision and Pattern Recognition (cs.CV)*, May 2023, [Online]. Available: http://arxiv.org/abs/2305.15660

[11] M. Li *et al.*, "TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models," Sep. 2021, [Online]. Available: http://arxiv.org/abs/2109.10282

[12] Hubert, P. Phoenix, R. Sudaryono, and D. Suhartono, "Classifying Promotion Images Using Optical Character Recognition and Naïve Bayes Classifier," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 498–506. doi: 10.1016/j.procs.2021.01.033.

[13] A. Sharma, S. Kaur, S. Vyas, and A. Nayyar, "Optical Character Recognition Using Hybrid CRNN Based Lexicon-Free Approach with Grey Wolf Hyperparameter Optimization," 2023, pp. 475–489. doi: 10.1007/978-981-99-2730-2_47.

[14] M. Fujitake, "DTrOCR: Decoder-only Transformer for Optical Character Recognition," Aug. 2023, [Online]. Available: http://arxiv.org/abs/2308.15996

[15] Y. Li, D. Chen, T. Tang, and X. Shen, "HTR-VT: Handwritten text recognition with vision transformer," *Pattern Recognit*, vol. 158, p. 110967, Feb. 2024, doi: 10.1016/J.PATCOG.2024.110967.

[16] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," Oct. 2020, [Online]. Available: http://arxiv.org/abs/2010.11929

[17] V. Agrawal, J. Jagtap, and M. P. Kantipudi, "Decoded-ViT: A Vision Transformer Framework for Handwritten Digit String Recognition," *Revue d'Intelligence Artificielle*, vol. 38, no. 2, pp. 523–529, Apr. 2024, doi: 10.18280/ria.380215.

[18] K.-M. Lee and S. R. Ramsey, *A History of The Korean Language*. 2011.
[19] F. Chollet, *Deep Learning with Python*. Manning Publications, 2017.
[20] Keras Team, "Keras ImageDataGenerator Documentation," TensorFlow. Accessed: Jul. 20, 2024. [Online]. Available: https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/ImageDataGenerator
[21] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
[22] J. Casia, "Handwritten Hangul Characters," Kaggle. Accessed: Dec. 14, 2023. [Online]. Available: https://www.kaggle.com/datasets/wayperwayp/hangulkorean-characters/data