J. Korean Soc. Ind. Appl. Math. Vol.27, No.2, 135–145, 2023 https://doi.org/10.12941/jksiam.2023.27.135

HANDWRITTEN HANGUL RECOGNITION MODEL USING MULTI-LABEL CLASSIFICATION

HANA CHOI[1,†]

[1]DEPARTMENT OF INNOVATION CENTER FOR INDUSTRIAL MATHEMATICS, NATIONAL INSTITUTE FOR MATHEMATICAL SCIENCES, SOUTH KOREA
Email address: †hanachoi@nims.re.kr

ABSTRACT. Recently, as deep learning technology has developed, various deep learning technologies have been introduced in handwritten recognition, greatly contributing to performance improvement. The recognition accuracy of handwritten Hangeul recognition has also improved significantly, but prior research has focused on recognizing 520 Hangul characters or 2,350 Hangul characters using SERI95 data or PE92 data. In the past, most of the expressions were possible with 2,350 Hangul characters, but as globalization progresses and information and communication technology develops, there are many cases where various foreign words need to be expressed in Hangul. In this paper, we propose a model that recognizes and combines the consonants, medial vowels, and final consonants of a Korean syllable using a multi-label classification model, and achieves a high recognition accuracy of 98.38% as a result of learning with the public data of Korean handwritten characters, PE92. In addition, this model learned only 2,350 Hangul characters, but can recognize the characters which is not included in the 2,350 Hangul characters

1. INTRODUCTION

Recently, with the development of IT technology, existing non-digitized data has been converted into a database. In particular, there is a high demand for databases of documents written by hand over a long period of time. Optical character recognition (OCR) is a representative technology for recognizing characters in a scanned document. OCR refers to a technology for finding the region of characters in an image and recognizing which character it is. Recently, NAVER and Google have begun commercializing OCR services.

Character recognition technology has a high recognition rate for printed characters but a low recognition rate for handwritten characters, especially the Korean alphabet, known as Hangul. Although Hangul is structurally similar to the alphabet, there are many different characters, and the handwritten character style is complex and diverse, making it difficult to recognize. In

Received November 30 2022; Revised December 20 2022; Accepted in revised form April 2 2023; Published online March 25 2023.

2000 *Mathematics Subject Classification.* 93B05.

Key words and phrases. Handwritten Hangul Recognition, Multi-label classification, Deep learning, Optical Character Recognition(OCR).

† Corresponding author.

addition, the alphabet has 52 uppercase and lowercase letters, but Hangul has 19 initial consonants, 21 medial vowels, and 28 final consonants (including those without final consonants), and the number of possible combinations is 11,172. This is more difficult and complex.

With the recent development of deep learning technology, the recognition rate has improved a lot in handwritten Hangul recognition, but most studies are limited to recognizing 2,350 Korean characters, which are frequently used. However, since the 2,350 Hangul characters are selected to be assigned to the computer code regardless of the structure of Hangul, all Korean words cannot be entered. For example, in the case of '뙹경모치(doeng-gyeongmochi)', a fish endemic to Korea, it cannot be written because there is no '뙹(doeng)' in the 2,350 Hangul characters, and similarly, the standard word 'Doryanjikhada' cannot be expressed. In addition, with globalization and the development of information and communication technology, there are many cases in which various foreign words or foreign words must be written in Hangul, but there is a limit to expressing them only with the 2,350 Hangul characters. For example, the Japanese "이와이 슌지(Iwai Shunji)", who is famous for directing the movie Love Letter, is also known as "슈운지(Shunji)" because there is no "슌(syun)" in the 2,350 Hangul characters.

NAVER's 'Clova OCR', which shows the best performance in recognizing Hangul, also does not recognize characters such as "쓩(ssyung), 뙹(doeng), 럄(lyam), 믜(mui), 슌(syun)" that are not in the 2,350 Hangul characters, as shown in Figure 1.

In this paper, we used a multi-label classification model that recognizes Hangul by combining initial consonant, medial vowel, and final consonants, as well as the 2,350 Hangul characters, to achieve a high accuracy of 98.38% in PE92 without using a data augmentation technique or a learning rate change technique. In addition, the proposed model can recognize the characters which is not included in the 2,350 Hangul characters by learning 2,350 Hangul character data.

2. RELATED WORK

2.1. **Public datasets of handwritten Hangul.** PE92 [1] and SERI95 are public datasets that are widely used in handwritten Hangul recognition. PE92 is the dataset collected by POSTECH of Science and Technology, and it is supported by ETRI in 1992. PE92 dataset consists of 100 sets based on 2,350 Hangul characters. SERI95 is data created in 1997 with support from the System Engineering Research Institute (SERI). It consists of 1,000 sets based on 520 frequently used characters among 2,350 Hangul characters. Table 1 shows the comparison of PE92 and SERI95.

TABLE 1. Public datasets of handwritten Hangul.

	PE92	SERI95
# of Classes	2350	520
# of Instances per class	100	1000

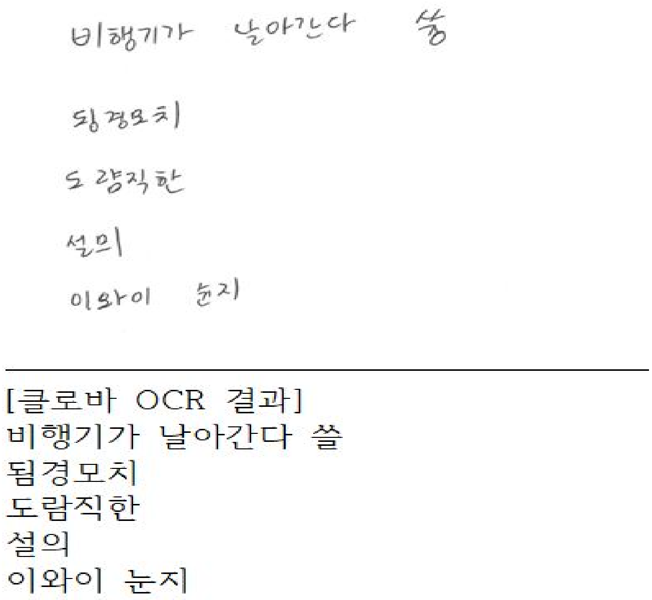

FIGURE 1. CLOVA OCR recognition result: Characters that are not included in the 2,350 Hangul characters, such as “쓩(ssyung), 됭(doeng), 럄(lyam), 믜 (mui), 슌(syun)”, are not recognized[2].

2.2. **Handwritten Hangul recognition.** There are two types of handwritten character recognition models: the online method and the offline model. The online model recognizes characters by receiving input data from a touch interface, such as a tablet with a pen, and is commonly used in smartphones and tablet PCs. In the online model, the input data is the coordinates and moving direction of the pen. On the other hand, the offline model refers to a model that receives a scanned image as input. This model recognizes the text region in the image and then classifies each character in the text region. OCR is a representative example of an offline model.

Deep learning significantly improved the recognition rate of handwritten Hangul. In 2015, I. Kim et al.[3]. introduced a deep learning model for the first time and used a CNN (Convolutional Neural Network) model consisting of 10 layers. In order to enhance the recognition rate, the number of data was increased using Elastic distortion, which randomly adds +1 or -1 to each pixel of the image. Through this, 92.92% in PE92 and 95.96% in SERI95 were obtained. These are 5.22% and 5.25% higher than the previous results, respectively. Later, in 2016, I. Kim et al.[4] used hybrid learning, which combines two networks, one that classifies similar classes of Hangul and the other that recognizes Hangul. These two networks share the CNN network. The model that classifies similar classes finds the difference between 10 similar classes, and the recognition network recognizes which character it is. Through the hybrid

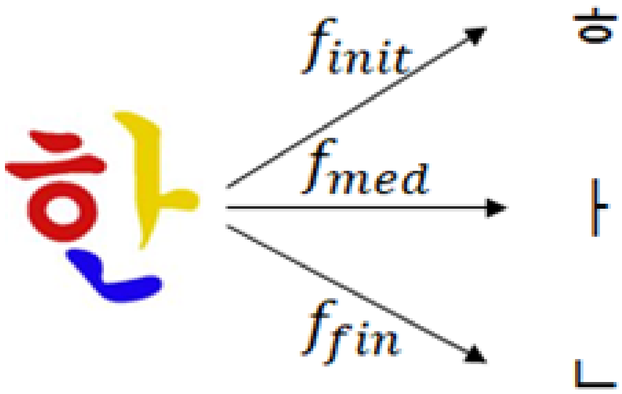

FIGURE 2. Handwritten Hangul recognition flow[6].

learning, 96.34% in PE92 and 97.67% in SERI95 were obtained. Later, in 2018, H. Kim and Y. Chung[5] modified and used GoogLeNet, one of the deep CNN algorithms, to fit the handwritten Hangul data, and learned the model by changing the learning rate to prevent convergence to the local optimization during learning. As a result, they obtained 98.64% in PE92.

The method of recognizing Hangul by dividing it into an initial consonant, a medial vowel, and a final consonant was introduced by H. Choi in a doctoral dissertation in 2020[6]. In this paper, three VGG19 models recognize the initial consonant, the medial vowel, and the final consonant of each Hangul and combine the results of each model as shown in the Figure 2. This model was trained using 35 sets of the 2,350 Hangul characters collected directly without using public data, and the training data was augmented by rotating the image left and right. Using this data, the accuracy of each model was 99.14% for the initial consonant, 98.24% for the medial vowel, and 98.35% for the final consonant, and finally, the accuracy for each character was 96.39%. If the model recognizes each character, there is one image for each class in one set of the 2,350 handwritten Hangul dataset. On the other hand, if the model recognizes the character by dividing it into an initial consonant, a medial vowel, and a final consonant, multiple images for each class can be obtained in one set, as shown in Figure 5. Therefore, the model performs better with a smaller amount of data than the PE92 or SERI95 datasets used in previous studies. However, since this method recognizes one image as three models, this causes many disadvantages. One of the biggest problems is that the number of learning parameters is too large and the computation is inefficient. In this paper, we propose the handwritten Hangul recognition model, which reduces learning parameters and unnecessary computation by recognizing initial consonants, medial vowels, and final consonants with the multi-label classification model instead of three VGG models.

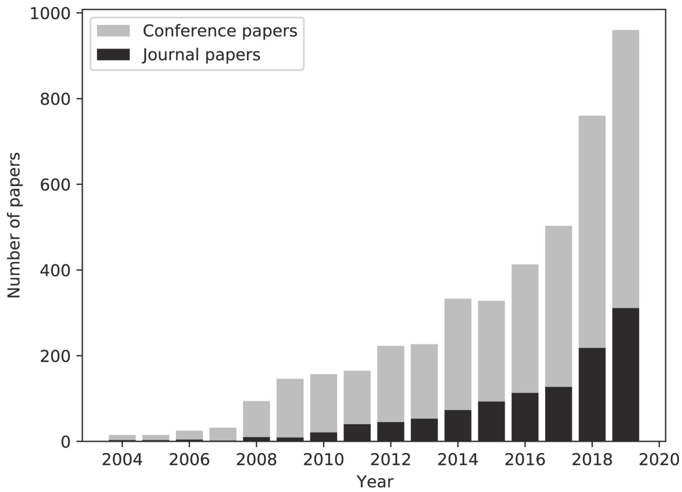

FIGURE 3. Graph of the number of papers in the SCOPUS database related to multi-label classification[7].

2.3. **Multi-label classification.** Multi-label classification refers to predicting all possible labels for one object. This is distinguished from multi-class classification. Multi-class classification means that there are multiple classes a given object can belong to, but it can only belong to one class. Therefore, multi-class classification can be regarded as a kind of multi-label classification. According to a study by J. Bogatinovski et al. in 2022[7], interest in various methodologies and datasets is also increasing as research on machine learning increases in multi-label classification. Figure 3 shows that the number of papers published in conferences and journals related to multi-label classification topics continues to increase every year in the SCOPUS database (https://www.scopus.com). Multi-label classification can be used to solve a variety of real-world problems. In 2016, J. Xu et al.[8] studied to predict the region of subcellular according to sequence, and F. Bruggs et al.[9] used audio signal processing to predict the type of bird included in a given recording. In addition, multi-label research is being actively conducted in natural language processing and image processing. In natural language processing, there is a research of finding all the subjects in a text when a given text has more than one subject.

3. PROPOSED METHOD AND LEARNING

3.1. **Architecture.** The handwritten Hangul recognition model proposed in this paper is based on ResNet [10] and adds a network that recognizes initial consonants, medial vowels, and final consonants at the end of the network. ResNet is a model that won the 2018 ILSVRC (ImagNet large scale visual recognition challenge), developed by Microsoft. A neural network model can perform complex modeling as the depth of the network deepens, but as the depth increases, it becomes difficult to learn due to the gradient vanishing problem. The gradient vanishing problem refers to a problem in which the derivative of the activation function is multiplied several times so that the gradient of the resulting value approaches 0, so that the weights are not updated when the gradient descent method is used. To avoid this problem, ResNet introduces a residual connection (or skip connection) into the network that adds the input value of the previous layer to the output value of the next layer or the next layer. Using this, the vanishing gradient problem was solved, and the layers of the network were stacked deeper to 34, 50, 101, and 152, enabling complex models to be implemented.

As shown in Table 2, the proposed model receives an $225 \times 255 \times 3$ image containing exactly one Korean letter as input and discriminate initial consonants, medial vowels, and final consonants by connecting three fully connected networks in parallel at the bottom of the ResNet34 model. Finally, the three values are combined to recognize the character.

3.2. **Training data.** PE92 dataset consists of a total of 100 sets of the 2,350 Hangul characters. That is, if the model classifies into 2,350 classes, there are 100 training data for each class. Figure 4 shows the sample of PE92 training data. This small amount of data can cause overfitting. However, the proposed model recognizes initial consonants, medial vowels, and final

TABLE 2. Proposed model network.

	Initial consonant	Medial vowel	Final consonant
Input	$225 \times 225 \times 3$		
conv1	7×7, 64, stride 2		
conv2x	3×3, max pool, stride 2		
	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$, stride 1		
conv3x	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times$, stride 2		
conv4x	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$, stride 2		
conv5x	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$, stride 2		
	average pool, 512-d fc		
fc1	19-d fc	21-d fc	29-d fc
# of parameters	45.6×10^6		

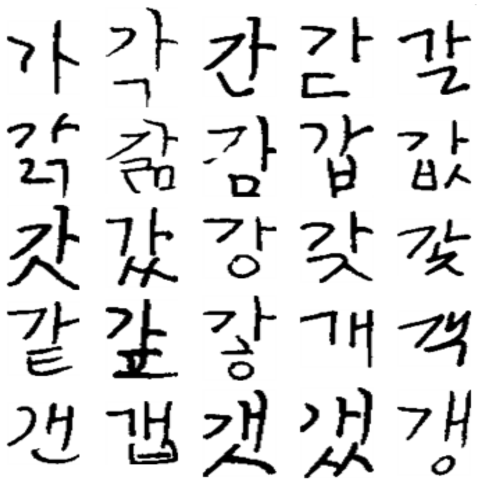

FIGURE 4. Sample of PE92 training data.

consonants, the number of classes decreases, and the number of data per class also increases as shown in Figure 5 in a set of the 2,350 Hangul characters. The public dataset of PE92 is divided into 90% of training dataset and 10% of test dataset.

4. EXPERIMENTS

The proposed algorithm was learned from the following system environment.

- Intel Xeon CPU E5-2046(v5)
- NVIDIA TITAN V
- PyTorch 1.8.1

After randomly shuffling the PE92 training dataset, 80% was used as training data and 20% as validation data. We trained the model using the Adam optimizer for optimization with the fixed learning 0.0001 and using pretrained model weights of ResNet34. The cross entropy loss is used for loss function. As a result, the recognition rate for the initial consonants, the medial vowels, and the final consonant was 99.45%, 99.44%, and 99.56%, respectively, and the final recognition rate for one character was 98.38%. The accuracy of each element of the initial consonants, the medial vowels, and the final consonant is shown in the Figure 6.

5. CONCLUSION

In the past, it was possible to express most of the words with the 2,350 Hangul characters, but as globalization progresses and information and communication technology develops, there are many cases in which foreign words need to be written in Hangul. In this situation, the importance of recognizing the characters not included in the 2,350 Hangul characters is gradually increasing. However, it is difficult to recognize all characters because there is not

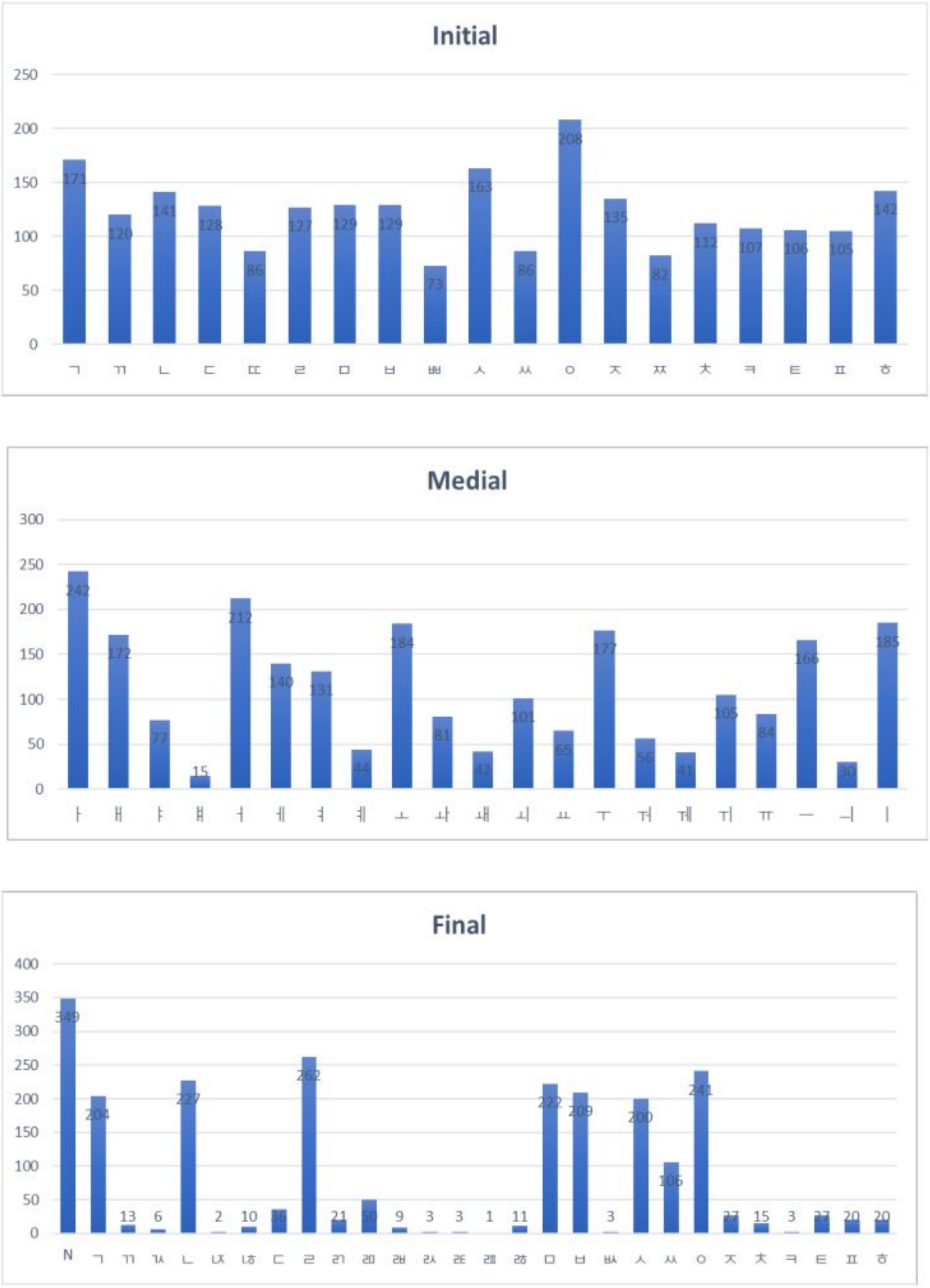

FIGURE 5. Distribution of initial consonants, medial vowels, and final consonants included in the 2,350 Hangul characters. The 'N' in the final graph indicates that the character has no final consonant.

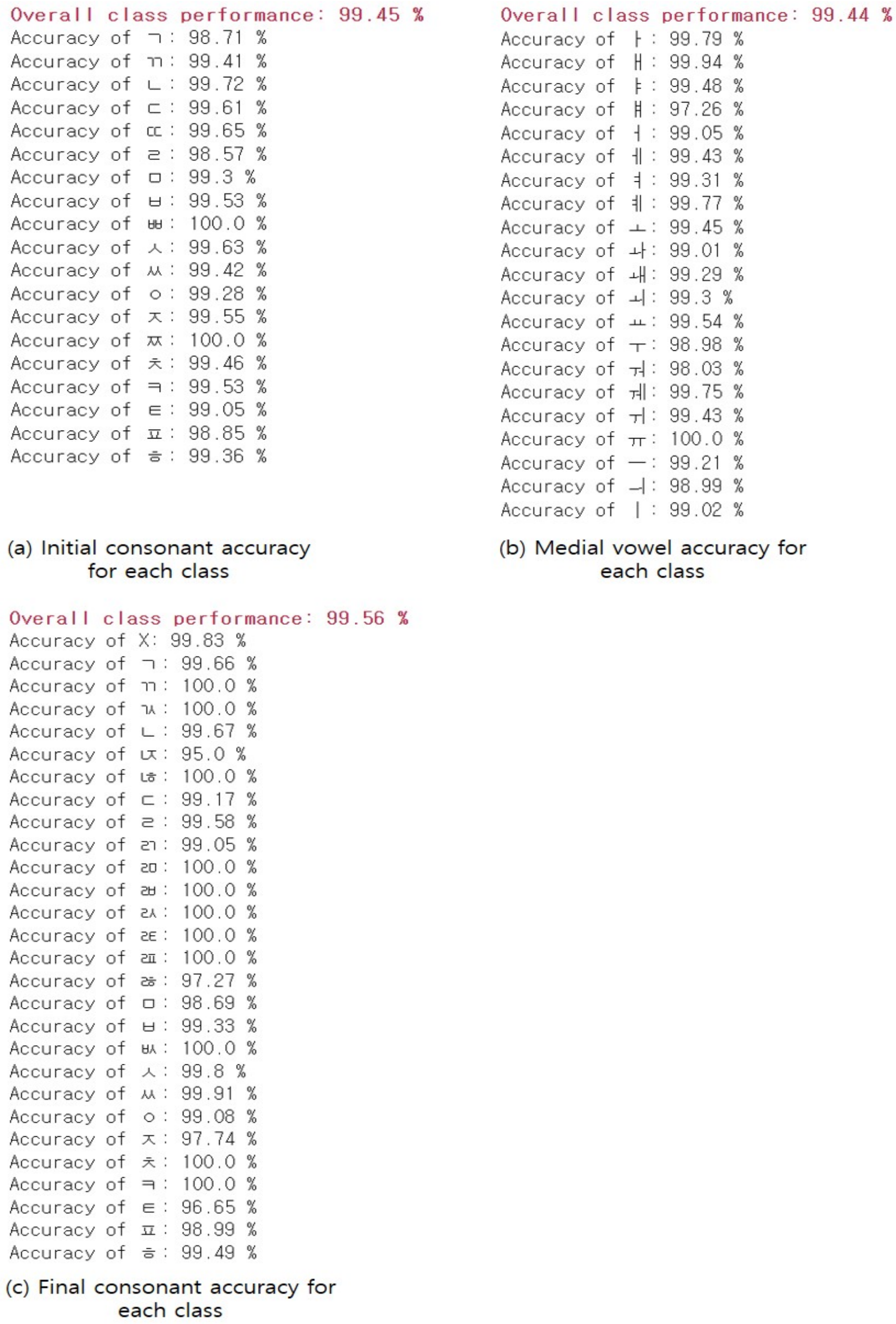

FIGURE 6. Accuracy for each class of initial consonants, medial vowels, and final consonants.

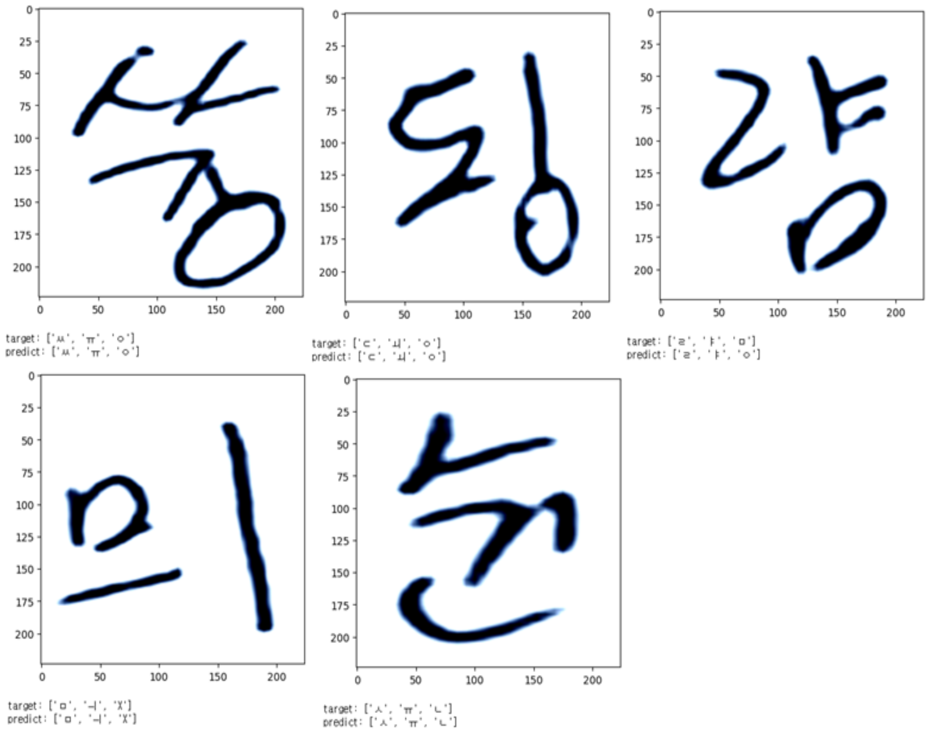

FIGURE 7. The results of recognizing characters not in the PE92 dataset with the proposed model.

enough handwritten Hangul data. But, as you can see from the Figure 7, the proposed model can recognize not only 2,350 Hangul characters but also characters not included in the 2,350 Hangul characters by training with a dataset covering only the 2,350 Hangul characters.

ACKNOWLEDGMENTS

The work of H. Choi was supported by National Institute for Mathematical Sciences(NIMS) grant funded by the Korea government(MSIT) No.B22810000.

REFERENCES

[1] D. H. Kim and S. Y. Bang, *An Overview of Hangul Handwritten Image Database PE92*, Proceedings of Annual Conference on Human and Language Technology, 1992.
[2] *CLOVA OCR*, NAVER Cloud Platform, accessed July 19, 2022, https://clova.ai/ocr.
[3] I. Kim and X. Xie, *Handwritten Hangul recognition using deep convolutional neural networks*, International Journal on Document Analysis and Recognition (IJDAR), **19** (2015), 1–13.
[4] I. Kim, C. Choi and S. Lee, *Improving discrimination ability of convolutional neural networks by hybrid learning*, International Journal on Document Analysis and Recognition (IJDAR), **19** (2016), 1–9.

[5] H. Kim and Y. Chung, *Improved Handwritten Hangeul Recognition using Deep Learning based on GoogLenet*, Journal of the Korea Contents Association, **18** (2018), 495–502.

[6] H. Choi, *Applications of Deep Convolutional Neural Networks: Enhanced Handwritten Hangul Recognition Model*, Diss. Sungkyunkwan Univ, (2020), Print.

[7] J. Bogatinovski, L. Todorovski, S. Dzeroski and D. Kocev, *Comprehensive comparative study of multi-label classification methods*, Expert Systems with Applications, **203** (2022).

[8] J. Xu, J. Liu, J. Yin and C. Sun, *A multi-label feature extraction algorithm via maximizing feature variance and feature-label dependence simultaneously*, Knowledge-Based Systems, **19** (2016), 192–84.

[9] F. Briggs, B. Lakshminarayanan, L. Neal, X. Z. Fern, R. Raich, S. J. K. Hadley, A.S. Hadley and M. G. Betts, *Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach*, Journal of the Acoustical Society of America, **131** (2012), 4640–4650.

[10] K. He, X. Zhang, S. Ren and J. Sun, *Deep Residual Learning for Image Recognition*, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR), NV, USA 2016.