

Article

Improving the Performance of Frequently Used Korean Handwritten Character Verification Based on Artificial Intelligence through Multimodal Fusion

Kyung Won Jin ¹ and Eui Chul Lee ^{2,*} ¹ Department of Artificial Intelligence & Informatics, Graduate School, Sangmyung University, Seoul 03016, Korea; 202031030@sangmyung.kr² Department of Human-Centered Artificial Intelligence, Sangmyung University, Seoul 03016, Korea

* Correspondence: elee@smu.ac.kr

Featured Application: This study proposes a handwriting verification method for frequently used Korean characters, and experimentally investigates the discrimination power according to the number of strokes and the recognition rate according to multi-character fusion. The significance of this study is to show the feasibility of off-line handwriting verification based on artificial intelligence and to calculate the reliability of handwriting verification results according to the likelihood ratio.

Abstract: Handwriting verification is a biometric recognition field that identifies individuals' unique characteristics contained in their handwriting. A single written character shows subtle differences depending on habits accumulated over time or the manner of writing. Based on this, it is often adopted in forensic investigations and as evidence in court. Existing handwriting verification is conducted by an expert, and is affected by the expert's ability or subjectivity, causing different results to arise depending on the expert. Therefore, we propose a handwriting verification method that excludes human subjectivity and has objectivity. Using computer vision and artificial intelligence (AI), we derived results that excluded human subjectivity, and the judgment strength was expressed through a likelihood ratio. To improve the existing method's accuracy, we performed a more accurate verification through multimodal use from the biometric field. Multimodal handwriting verification is conducted using up to four characters (not just one) because individual handwriting in each character is different. For learning, n -fold tests were conducted to maintain test objectivity, and the average performance of single character-based verification was 80.14% and the multimodal method averaged 88.96%. Here, we proposed the objectivity of handwriting verification through learning using AI, and show that performance improved through multimodal fusion.

Keywords: handwriting verification; multimodal; behavioral biometric; Siamese network; ResNet; artificial intelligence



Citation: Jin, K.W.; Lee, E.C. Improving the Performance of Frequently Used Korean Handwritten Character Verification Based on Artificial Intelligence through Multimodal Fusion. *Appl. Sci.* **2021**, *11*, 8413. <https://doi.org/10.3390/app11188413>

Academic Editor: Larbi Boubchir

Received: 26 August 2021

Accepted: 9 September 2021

Published: 10 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Handwriting verification is a valuable biometric because each person's handwriting style is different. Even in a single character, differences such as stroke length or curve are used to represent an individual [1]. In existing research, handwriting verification has often been conducted on English sentences or signatures. Here, we show that the verification accuracy can be used for Korean characters and improved by using a multimodal method that combines the verification of individual characters.

First, there is a method for verification using the Siamese network [2,3]. Second, Korean characters can be verified using the geometric features of handwriting [4]. Among the aforementioned verifications, a signature is used only by individuals, and many personal characteristics are reflected. Since sentence verification consists of several characters, it

is vulnerable to individual character verification. Therefore, we conduct a verification based on frequently used Korean characters that do not contain personal characteristics. Existing studies used multimodal by adding other biometric features such as voice, electrocardiogram, and fingerprinting to handwriting [5]. Ross and Jain [6] also identified better performance through the decision tree, linear descriptive function, and sum rule. In addition to the method of performing multimodal according to the characteristics of biometrics, there are also studies showing that it differs according to the level of fusion of multimodal data [7]. This study extracts the verification accuracy by applying multimodal at the predicted value level using only the features of several characters without additional bio-signals.

The deep learning model used for handwriting verification uses a ResNet-based Siamese network. To train the Siamese network, data are composed of a pair of handwriting images from the same person and a pair of handwriting images composed from the others. An image consisting of a pair is input as two-ResNet learning networks. The two ResNet learning networks have the same structure, and the weights are updated identically. In the network, a pair of images pass through the ResNet, extracting a feature vector, calculating the difference between the feature vectors of the two images, and conducting training to reduce the difference. If you place a pair of images that have not been used for training with the learned weights, you can predict whether they belong to the same person. The amount of data used for learning is increased using the data augmentation technique.

Multimodal technology is widely used in the field of biometric recognition to increase recognition accuracy. We proceed by applying multimodal handwriting during biometric recognition. If the accuracy was previously measured through one character, it tries to improve the accuracy of handwriting recognition by using two or more characters through multimodal. In one character, the threshold of the predicted value is indicated by a dot. When using multimodal, two or more characters are combined, and a threshold is set for dividing the same pair and another pair. For example, if two characters are used, a two-dimensional (2D) plane can be used to separate the threshold with a line. If three characters are used, you can use a three-dimensional (3D) space to separate the threshold with a plane.

The multimodal matching method uses a method combining the predicted values of the pairs obtained through learning. For accuracy, the predicted value of the handwriting pair of the same person is combined with the predicted value of the handwriting pair of the same person in different characters. The predicted value of the other handwriting pair must be combined with the predicted value of the other handwritten pair, which is equally different in other characters. For example, the predicted value of the handwriting pair of A and B for the character '없' should also be combined with the predicted value of the handwriting pair of A and B for the character '김'.

2. Implementation

2.1. Data Set

2.1.1. Training Data

Figure 1 shows a will that contains the four characters used in the experiment. For the data, the four characters were selected based on the complexity according to the number of strokes among frequently used Korean characters. The characters used were '없', '김', '다' and '이', and the complexity decreased in order from the left.

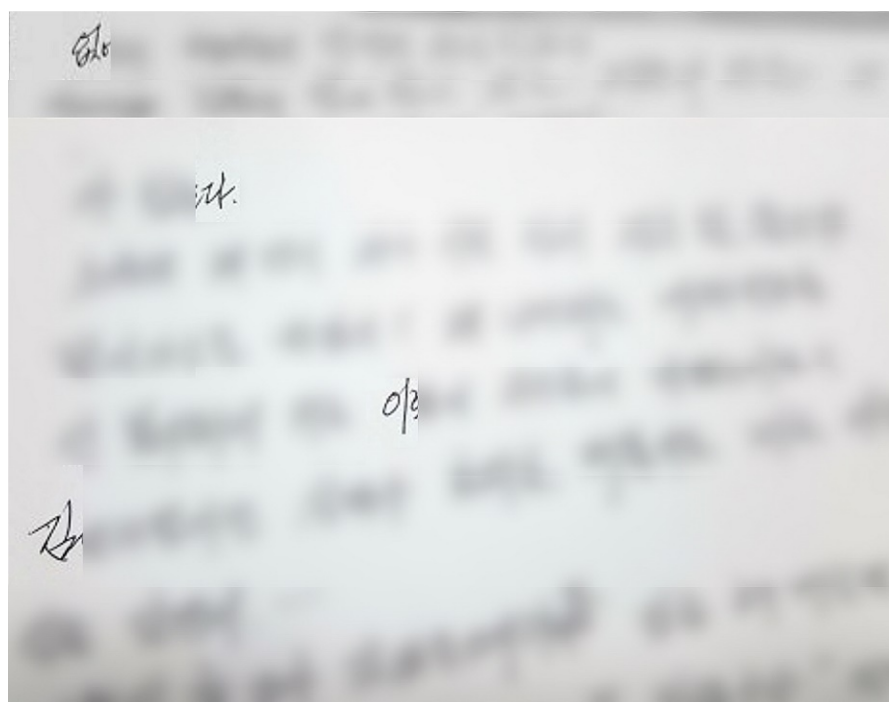


Figure 1. Example of a will containing four characters.

The experimenters who participated in the handwriting verification were 20 men and women, and each experimenter wrote the characters ‘없’, ‘김’, ‘다’, and ‘이’ 10 times each. Since 20 people wrote it down 10 times, there were 200 images for each character, and the data for 800 characters were collected. In the process of cutting out the character written in the manuscript, the size was normalized to 112×112 pixels so that the aspect ratio of the original character was not affected.

Through the data augmentation technique, 10% of the image size was applied for enlargement and reduction, $\pm 5^\circ$ for rotation, and 3% for horizontal and vertical movements. This increased the number of images by nine times for the training and validation set and five times for the test set. Therefore, the number of images in the training, validation, and test sets consisted of 61, 29, and 50 images per person, respectively. Training was conducted using the Siamese network, and due to the nature of this network, data were input in pairs. The data were composed of combinations of pulling two of the images in an augmented set. The number of training, verification, and test sets (ratio) comprised 54,900 (70%), 12,180 (15%), and 12,250 (15%), respectively. The order of the combinations was decided at random, and because the number of combinations to be drawn from the handwriting of others was much larger than the number of combinations from the same person, the number of combinations to be drawn from the handwriting of others was matched to the number of combinations to be drawn with the same person.

2.1.2. Multimodal Data

Multimodal analysis was performed by combining the handwriting pair predicted values. Through the network, the predicted values for the test set came out to each of ‘없’, ‘김’, ‘다’, and ‘이’, and this predicted value was grouped in a multimodal.

Figure 2a was written by person A, and Figure 2b was another text written by person A. Figure 2c is a copy of Figure 2b written by person B. All blue and orange lines are pairs of a single predictive model. It also had a predicted value for each pair of lines. The multimodal model combined the predicted values of each of these characters to obtain a new predicted threshold and prediction accuracy. In this process, the combination of the predicted values of the characters had a certain rule rather than a random rule, which is explained in the figure below.

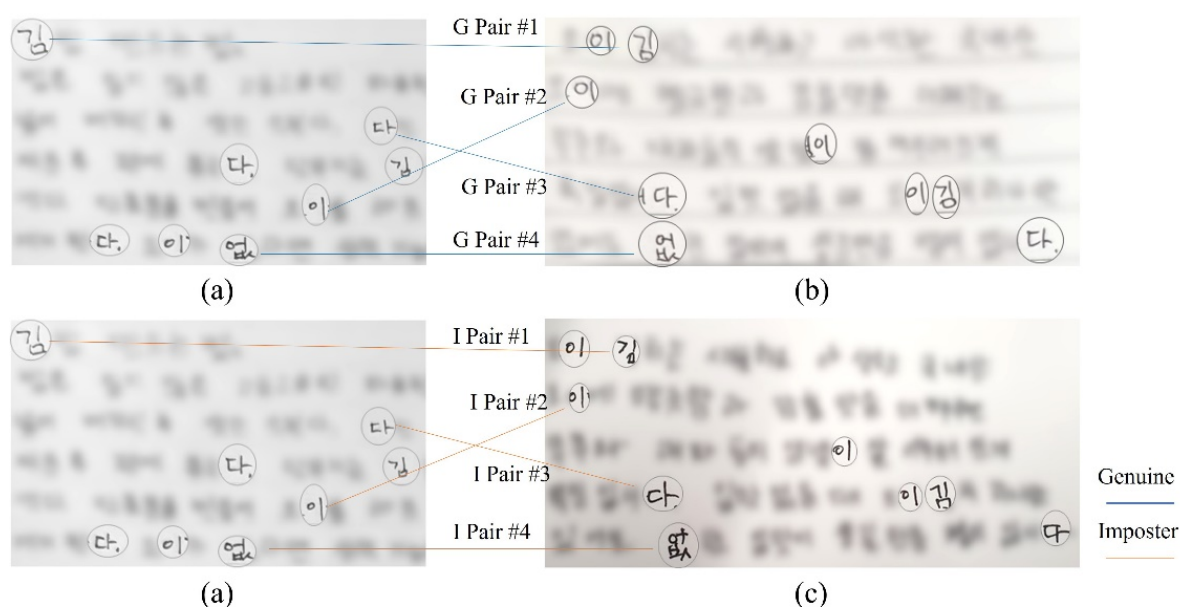


Figure 2. Combination example of each character of genuine and imposter pair. (a): Text #1 written by person A, (b): Text #2 written by person A, (c): Text #2 written by person B.

The first was the creation of a genuine multimodal pair. The blue lines between Figure 2a,b connect the same characters written by the same person. When combining the two characters, we used G Pair #1 and G Pair #2 together. In the case of the three characters, G Pair #3 was combined. The important point was that when combining the predicted values of a character pair such as this, different characters written by the same person had to be combined.

The following was a method for creating a multimodal imposter pair. A predicted value was extracted between the character written by A and the same character written by B and combined with the predicted value of another character pair. Then, I Pair #1 was the predicted value between the characters written by A and the same character written by B. For multi-modal, the predicted value to be combined had to be connected with I Pair #2, which was the predicted value of the characters written by A and the same character written by B. That is, when combining, it had to be joined by a person of the same configuration (e.g., A-B with A-B, C-D with C-D).

When creating an imposter pair, the number of pairs formed of different persons was different because the number of combinations were randomly created. Therefore, the number was unified according to the smallest number, and the number of genuine pairs were matched with the number of imposter pairs. After this process, 23,618 predicted values were used, with 11,809 genuine and imposter pairs each.

2.2. Network

Here, a Siamese network composed of ResNet was used. ResNet is an abbreviation for a residual neural network. It uses skip connections and residual blocks to maintain input information for each block [8]. In Figure 3, input x was added to the value that passed through the weight layer to prevent loss of information during learning. This solved the problem of a vanishing gradient as the network deepened.

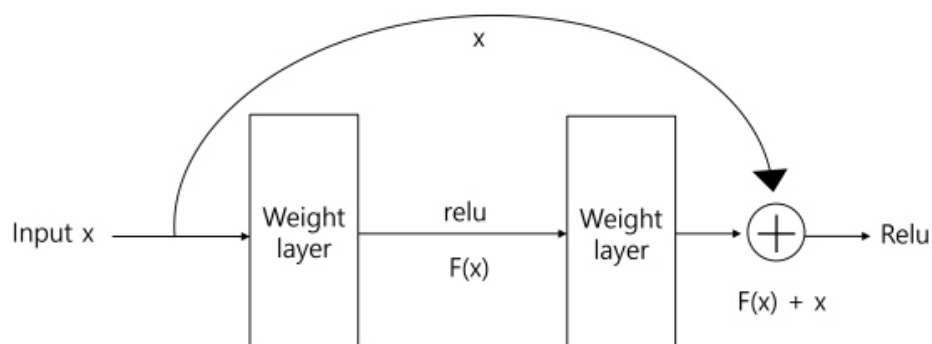


Figure 3. Residual block structure in ResNet.

The Siamese network is a representative method of one-shot learning that enables accurate prediction with little data [9]. We sent two characters in pairs through the Siamese network and learned how similar they were. In Figure 4, looking at the structure of the Siamese network, the handwriting of the same person or the handwriting of another person were given as input values. Each image passed through each ResNet network that shared a weight, and the difference between the feature vectors of each image obtained through this was obtained as L1_norm [10]. This value was changed to a value between 0 and 1 through the sigmoid activation function. If the two pictures belonged to the same person, the result was close to 1, and if the picture was from another person, the result was close to 0. The SGD optimizer was used for training, and dropouts were not used [11].



Figure 4. Siamese network structure.

2.3. Multimodal

As explained in Section 2.2, when passing the sigmoid, the predicted value was between 0 and 1, but because the value was compressed, it could not be used to obtain an accurate threshold or to calculate the likelihood ratio. Therefore, the predicted values, excluding the sigmoid layer, were saved as comma separated value files.

Multimodal approaches seek to improve accuracy by combining two or more features. Since multimodal in the predicted value part was performed, the predicted values of the four characters, '없', '김', '다' and '이' had to be combined. As the complexity decreased, the prediction accuracy tended to decrease, so from the characters with high accuracy, two ('없' and '김'), three ('없', '김' and '다'), and four characters ('없', '김', '다' and '이') were combined. Through this, it was confirmed whether the accuracy improved even when performing multimodal with data with low accuracy.

Before combining, Figure 5 shows the distribution plot of each letter '없', '김', '다', and '이'.

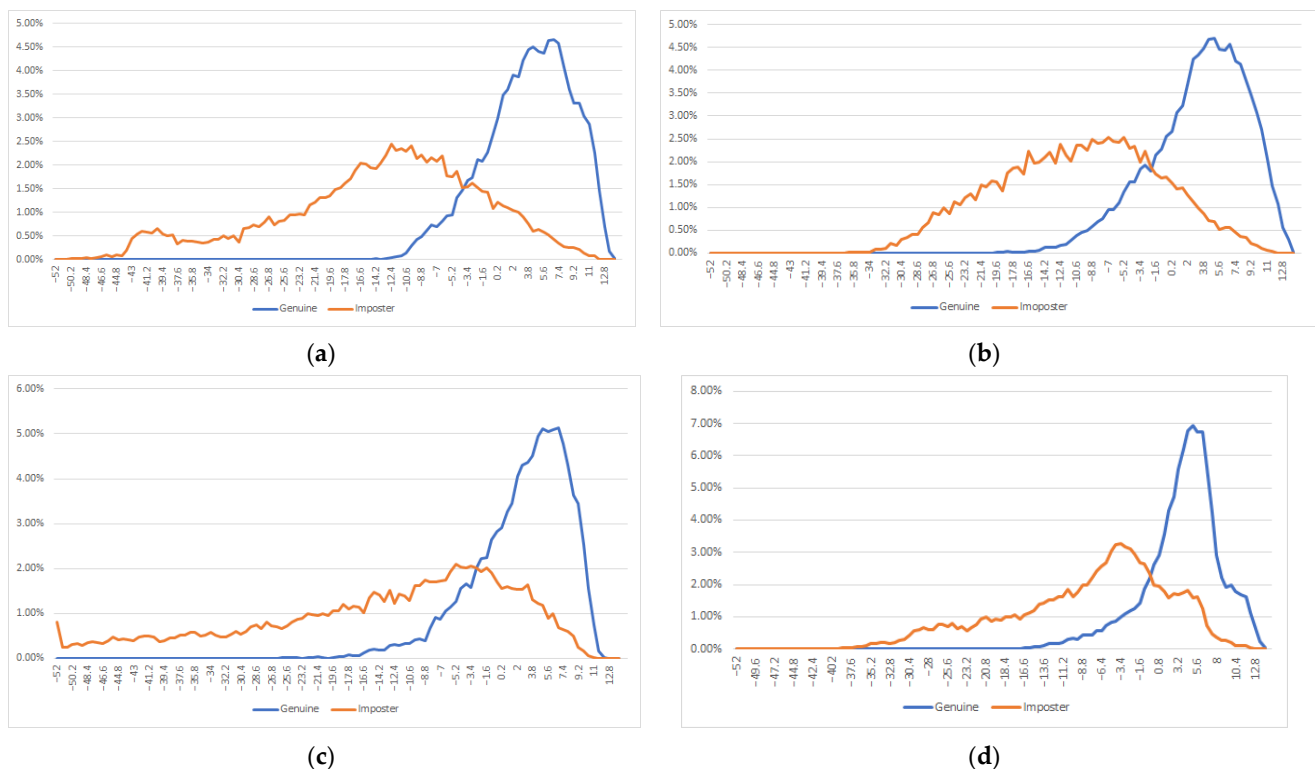


Figure 5. Probability density functions for genuine/imposter matching of four characters: (a) ‘없’; (b) ‘김’; (c) ‘다’; (d) ‘이’.

The combined predicted values were verified using a support vector machine (SVM) [12]. Through the SVM model, a boundary value that divided the area of the training set was derived, and the accuracy of the distribution area of the test set was expressed through this boundary value. The kernel of the SVM model was experimented with linear and radial basis function (RBF), and the linear kernel represented the boundary of the region with a straight line or plane, and the RBF kernel denoted the boundary of the region with a curve or a curved surface [13]. The parameter C used in the SVM model determined the influence of each data point in the scatter plot. The larger the value, the greater the influence of each point on the model, resulting in a more accurate classification by bending the boundary. Here, $C = 100$ was used because the difference in accuracy according to parameter C was not large.

Figure 6a combined the two-character predictions and Figure 6b combined the three-character predictions. The blue and orange dots represented the genuine and imposter pairs, respectively. Combinations of more than four characters were possible, but because visualization was difficult, only the distribution of combinations of two and three characters is shown.

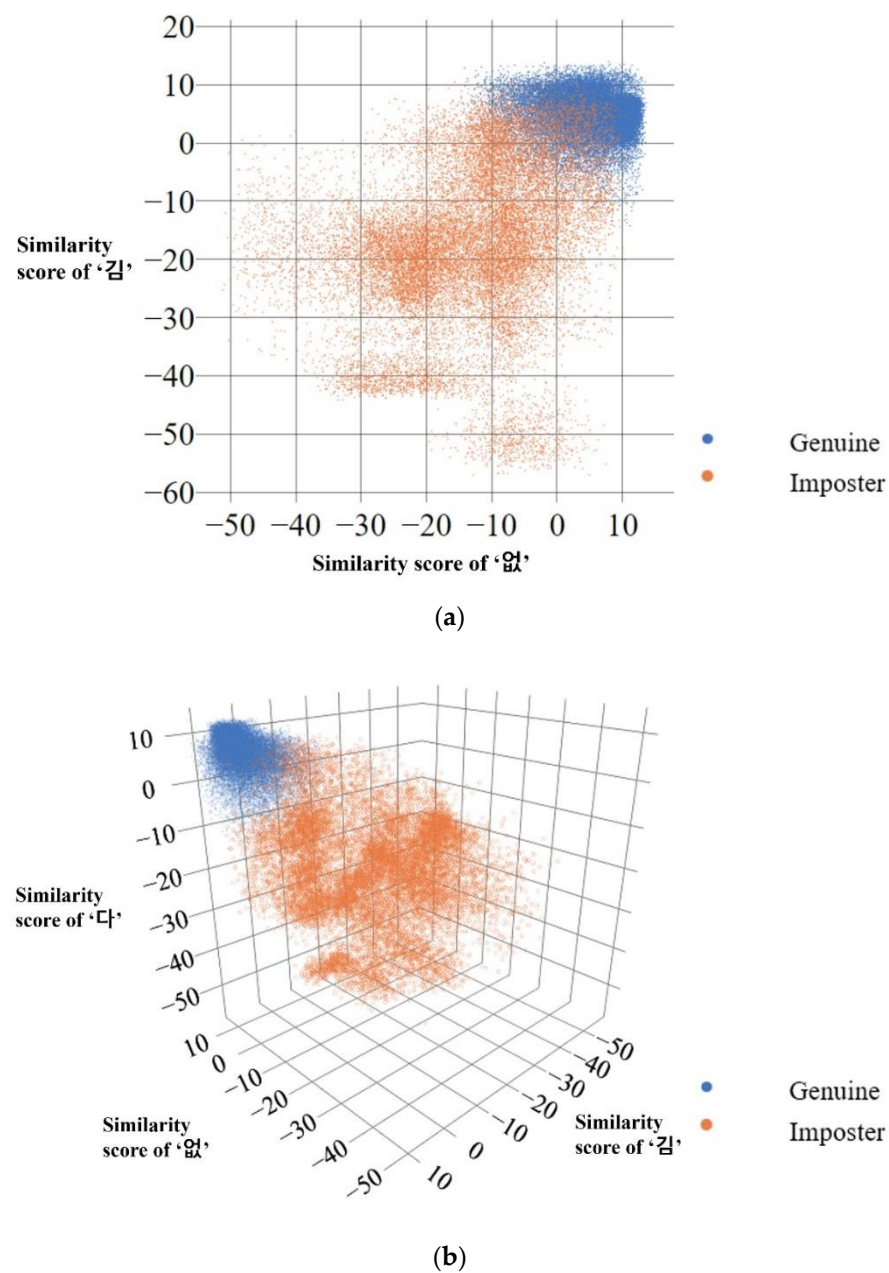


Figure 6. Scatter plots of multimodal character pairs: (a) multimodal of the '없' and '김' pairs; (b) multimodal of the '없', '김' and '다' pairs.

The test method used the n -fold method to cross-validate the training and test sets so that all data sets were used for training and testing [14]. Thus, the test set prevented bias in the model evaluation index owing to coincidence, and increased the reliability of the performance evaluation.

3. Results and Discussion

Table 1 shows the accuracy of each character used in the multimodal model. We trained each of the four characters through the Siamese network. The complexity of the characters was high in the order of '없', '김', '다', and '이'. Looking at the table, the lower the complexity of the character, the lower the accuracy, because the character had fewer characteristics. Therefore, we proceeded with the combination according to the complexity of the characters when combining the characters that would be used in the multimodal.

When combining more characters, we wanted to prove the effectiveness of the verification by combining the characteristics of lower complexity.

Table 1. Verification accuracy of single character.

Accuracy (%)	없	김	다	이
	82.66%	81.36%	78.54%	78.01%

We experimented with two types of kernels of the SVM model, linear and RBF, and the difference between the two kernels was not large because the distribution of data was divided in a balanced manner. If the distribution was lower, the RBF kernel performed better because the curve or surface provided a difference to the finer details in denoting the boundary. Figures 7 and 8 show the data classified using linear and RBF kernels, and visually show the classification boundaries.

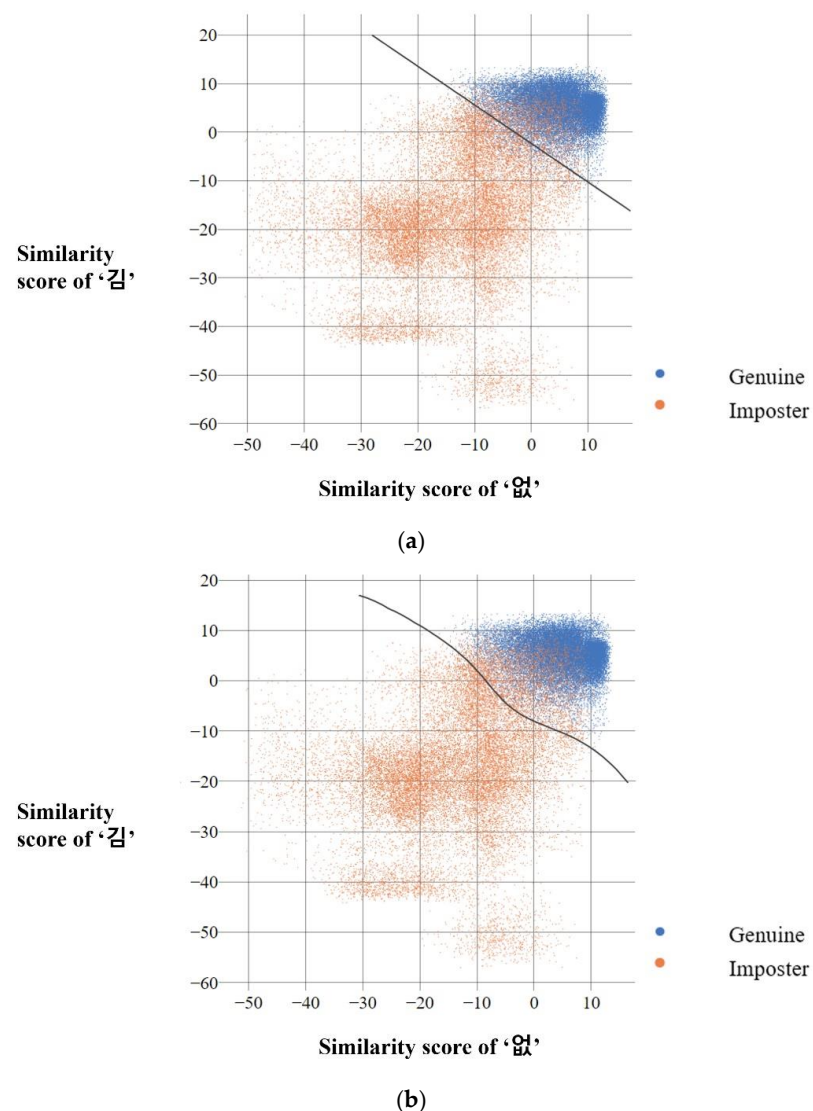


Figure 7. Scatter plots of multimodal of two characters with a classifier: (a) linear kernel boundary; (b) RBF kernel boundary.

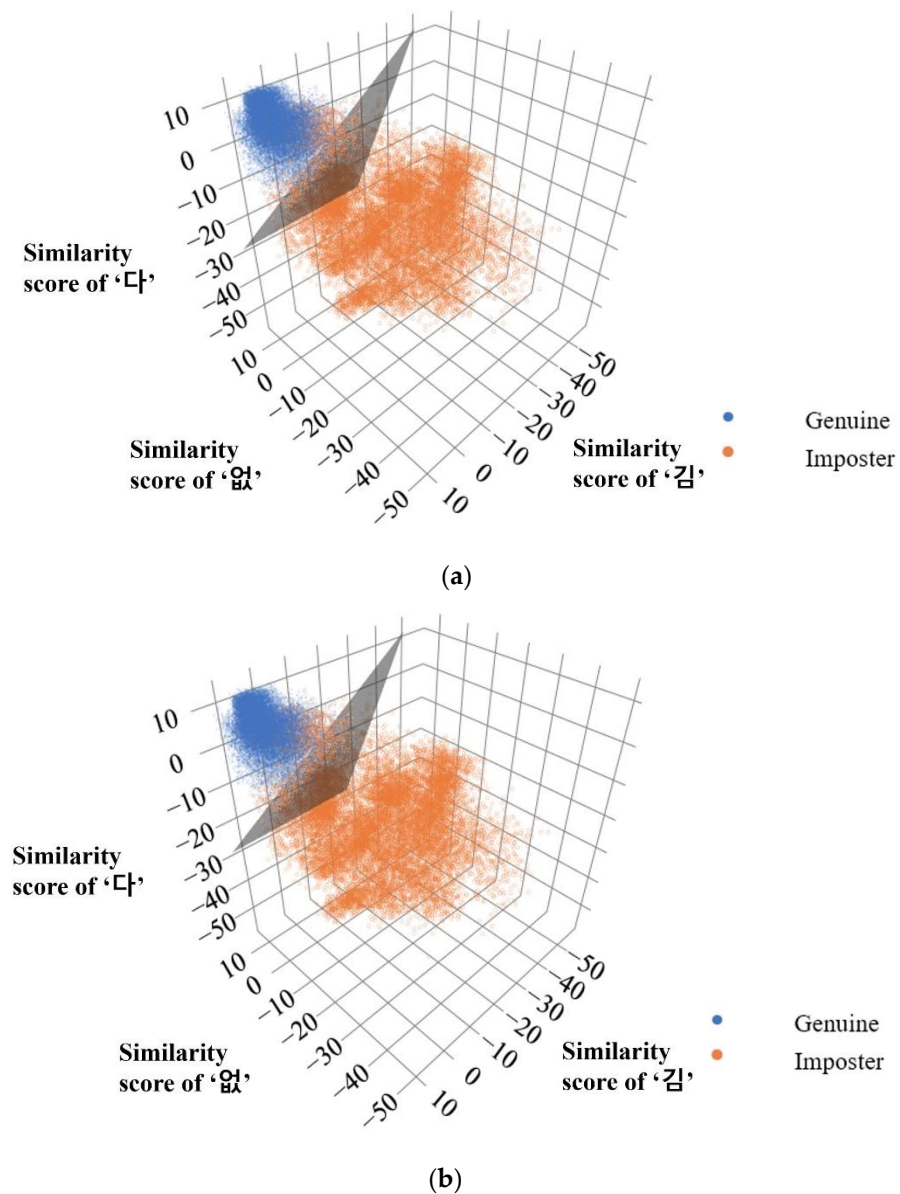


Figure 8. Scatter plots of multimodal of three characters with a classifier: (a) linear kernel boundary; (b) RBF kernel boundary.

Table 2 shows the accuracy compared using the RBF kernel according to the parameter C , which was 100, and the number of characters. The average multimodal accuracy was confirmed to be over 88%. This meant that the multimodal accuracy was higher than the single-character accuracy. For more than two characters, the reason for the high accuracy was that the predicted values, previously classified as one threshold, were classified in more detail through nonlinear thresholds.

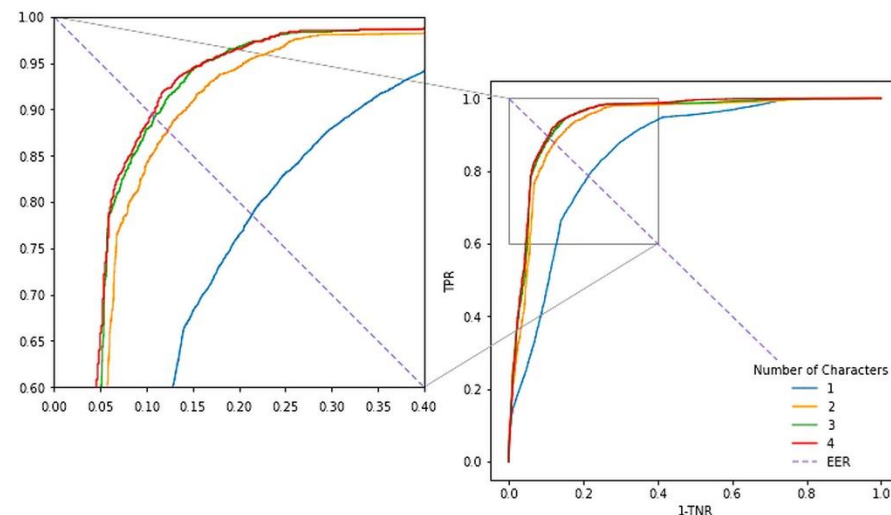
Table 2. Verification accuracy through multimodal.

Accuracy (%)	Number of Characters			
	One ¹	Two ²	Three ³	Four ⁴
	80.14%	87.65%	89.22%	90.01%

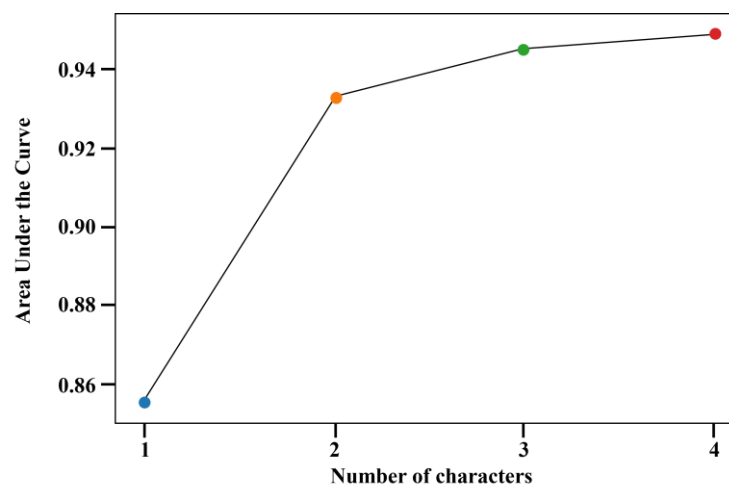
¹ Accuracy mean for each of the four characters ('없', '김', '다', '이'). ² Multimodal accuracy of '없', '김' characters.

³ Multimodal accuracy of '없', '김' and '다' characters. ⁴ Multimodal accuracy of '없', '김', '다' and '이' characters.

Figure 9 shows the receiver operating characteristics (a) and the area under the curve (AUC) of the test results (b) [15]. Looking at (a), we saw that the larger the number of characters, the greater the AUC, and (b) shows the AUC value according to the number of characters.



(a)



(b)

Figure 9. An example of receiver operating characteristic curve (ROC) and AUC: (a) ROC curve; (b) AUC by number of characters.

Figure 10 shows an example of the multimodal combination and the results that this paper intended to show. Figure 10a shows an example of combining the two characters. The character ‘없’ on the left had a predicted value of -3.16 and a predicted label of 0 was incorrectly predicted. However, combining the correctly predicted ‘김’ characters on the right through multimodal showed that the two characters were correctly predicted. In Figure 10b, as in Figure 10a, we saw that the incorrectly predicted ‘없’ and ‘다’ characters combined with the correctly predicted ‘김’ characters were correctly predicted. Figure 10c shows the same results. The key point was that multimodal approaches can be used to correctly predict previous incorrectly predicted outcomes.

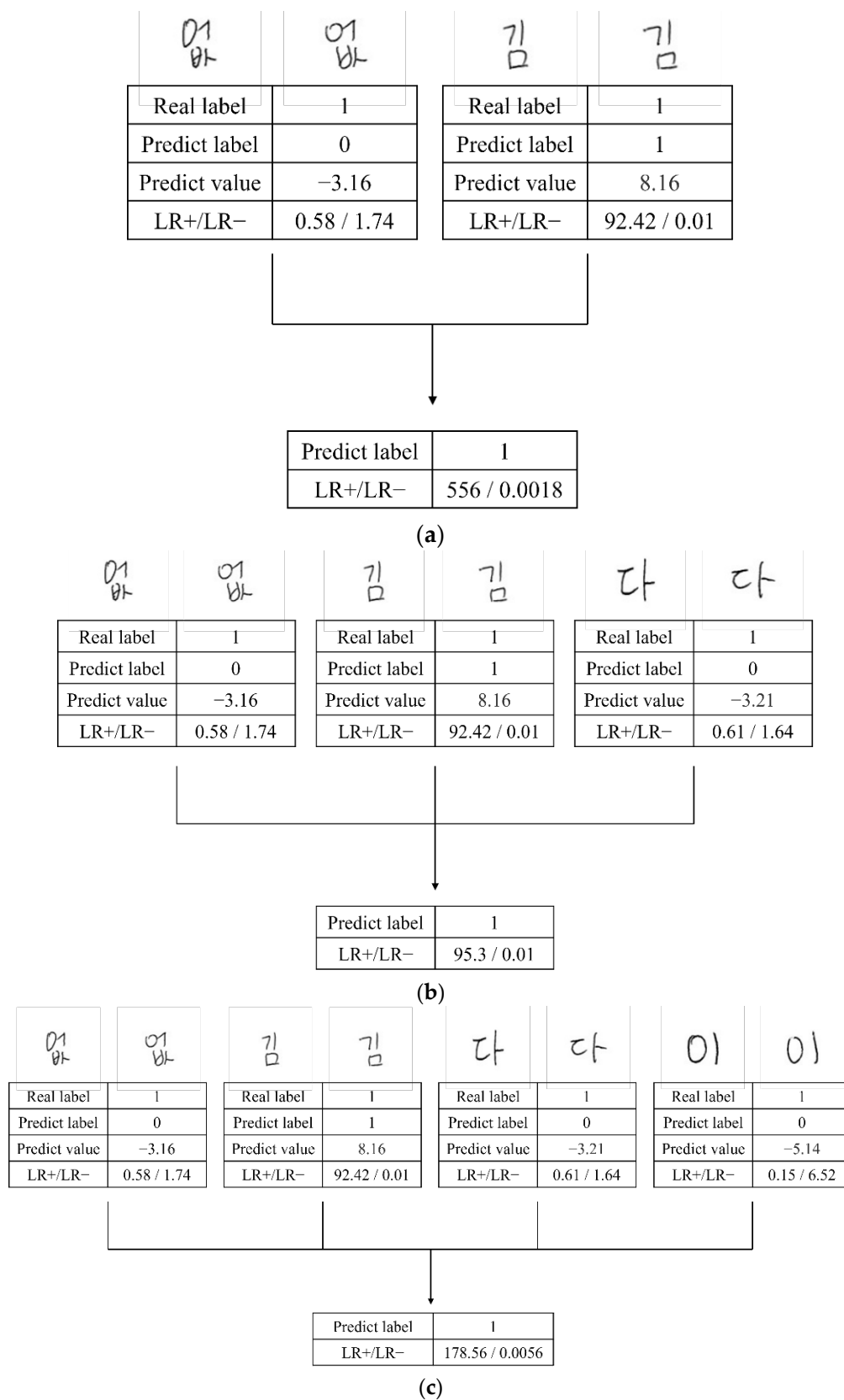


Figure 10. Example of multimodal and likelihood ratio according to the number of combined characters: (a) two characters; (b) three characters; (c) four characters.

Also, likelihood ratios were used to evaluate these computer-based verifications in biometrics, particularly in forensic evidence evaluations [16]. The figure below shows the

likelihood ratio for each character and the likelihood ratio value when combined through multimodal. In general, we saw that the value of the likelihood ratio increased when multimodal was used instead of the verification of a single character. Further, we evaluated the accuracy through a multimodal analysis using calculated likelihood ratios.

Confirmation through the experimental results can be summarized as follows. First, in the case of a single character, it was confirmed that the more strokes, the higher the verification accuracy. Second, the verification accuracy increased as more characters were fused multimodally. It also means cases of false acceptance or false rejection in a single-character comparison can be correctly classified by multimodal fusion. In the above case, the small LR value in the single-character comparison was increased enough to support the decision after multimodal fusion.

4. Conclusions

Here, a multimodal method was shown to improve the performance of handwriting verification. The accuracy of verification was improved by combining the predicted values of two or more characters through a multimodal method. The advantage of this method was that characters misclassified by a single predictive model were properly classified through curves or curved spaces in 2D and 3D spaces. It also had the advantage of reducing the false acceptance rate or false rejection rate used as a measure of biometric accuracy. As for the verification accuracy, the average accuracy of the conventional single-character verification was 80.14%, and the average verification accuracies of two, three, and four characters were 88.96%, showing a remarkable performance improvement. This point can contribute to improving the performance of the existing handwriting technique to the next level.

Also, we indicate the judgment strength through the likelihood ratio. This is important for handwriting verification using AI to be scientifically recognized in forensic science.

By comparing the accuracy of two, three, and four characters, it was possible to see the improvement in accuracy as the number of comparison dimensions increased. In the future, higher accuracy in higher dimensions can be expected by using additional character data. By applying a multimodal method to improve the performance of existing handwriting verification networks, the additional performance can be improved.

In the future, forged fake handwriting will be dealt with through counterfeit biometric detection. Recognition of handwriting that imitates other people's handwriting should be dealt with in terms of identification of fake biometrics, not biometrics authentication. Since handwriting data are not biometrics that utilize the characteristics of human body parts, the difficulty is thought to be much higher than that of physiological-based biometrics such as face, iris, and fingerprints.

Author Contributions: Conceptualization, E.C.L.; methodology, E.C.L. and K.W.J.; software, K.W.J.; validation, E.C.L. and K.W.J.; investigation, K.W.J.; data curation, K.W.J.; writing—original draft preparation, K.W.J.; writing—review and editing, E.C.L.; visualization, E.C.L. and K.W.J.; supervision, E.C.L.; project administration, E.C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The obtained data can be shared by contacting to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Srihari, S.N.; Cha, S.H.; Arora, H.; Lee, S. Individuality of handwriting. *J. Forensic Sci.* **2002**, *47*, 856–872. [[CrossRef](#)] [[PubMed](#)]
2. Dey, S.; Dutta, A.; Toledo, J.I.; Ghosh, S.K.; Lladós, J.; Pal, U. Signet: Convolutional Siamese network for writer independent offline signature verification. *arXiv* **2017**, arXiv:1707.02131.
3. Dlamini, N.; Terence, Z. Author identification from handwritten characters using Siamese CNN. In *2019 International Multidisciplinary Information Technology and Engineering Conference, Vanderbijlpark, South Africa, 21–22 November 2019*; IEEE Publications: Vanderbijlpark, Gauteng, South Africa, 2019; Volume 2019, pp. 1–6.
4. Jang, W.; Kim, S.; Kim, Y.; Lee, E.C. Automated verification method of Korean word handwriting using geometric feature. In *Lecture Notes in Electrical Engineering*; Springer: Singapore, 2018; pp. 1340–1345.
5. Ross, A.; Jain, A.K. Multimodal biometrics: An overview. In *12th European Signal Processing Conference, Vienna, Austria, 6–10 September 2004*; IEEE Publications: Piscataway, NJ, USA, 2004; Volume 2004, pp. 1221–1224.
6. Ross, A.; Jain, A.K. Information fusion in biometrics. *Pattern Recognit. Lett.* **2003**, *24*, 2115–2125. [[CrossRef](#)]
7. Garg, S.N.; Vig, R.; Gupta, S. A survey on different levels of fusion in multimodal biometrics. *Indian J. Sci. Technol.* **2017**, *10*, 1–11. [[CrossRef](#)]
8. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016*; pp. 770–778.
9. Koch, G.; Zemel, R.; Salakhutdinov, R. Siamese neural networks for one-shot image recognition. In *Proceedings of the International Conference on Machine Learning Deep Learning Workshop, Lille, France, 6–11 July 2015*.
10. Wang, H.; Nie, F.; Huang, H. Robust distance metric learning via simultaneous l1-norm minimization and maximization. In *Proceedings of the International Conference on Machine Learning, Beijing, China, 21–26 June 2014*; pp. 1836–1844.
11. Bottou, L. Stochastic gradient learning in neural networks. *Proc. Neuro-Nîmes* **1991**, *91*, 12.
12. Noble, W.S. What is a support vector machine? *Nat. Biotechnol.* **2006**, *24*, 1565–1567. [[CrossRef](#)] [[PubMed](#)]
13. Patle, A.; Chouhan, D.S. SVM kernel functions for classification. In *Proceedings of the International Conference on Advances in Technology and Engineering (ICATE), Mumbai, India, 23–25 January 2013*; IEEE Publications: Mumbai, India, 2013; pp. 1–9.
14. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI* **1995**, *14*, 1137–1145.
15. Hanley, J.A.; McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*, 29–36. [[CrossRef](#)] [[PubMed](#)]
16. Meuwly, D.; Ramos, D.; Haraksim, R. A guideline for the validation of likelihood ratio methods used for forensic evidence evaluation. *Forensic Sci. Int.* **2017**, *276*, 142–153. [[CrossRef](#)] [[PubMed](#)]