Study of Handwriting Recognition Implementation in Data Entry of *Survei Angkatan Kerja Nasional (SAKERNAS)* using CNN

Y F Mustafa[1], **F Ridho**[1], **S Mariyah**[1]

[1]Computational Statistics Department, Politeknik Statistika STIS, Indonesia

*Corresponding author's e-mail: 221710082@stis.ac.id

Abstract. The use of Paper and Pencil Interviewing (PAPI) at BPS requires manual data entry that cannot be separated from the human ability to recognize handwriting. For computers, handwriting recognition is complex work that requires complex algorithms. Convolutional Neural Network (CNN) is an algorithm that can accommodate the complexity of handwriting recognition. This research intends to conduct a study on the implementation of the handwriting recognition model using CNN in recognizing handwriting on PAPI questionnaire in data entry activities. Handwriting recognition model was built using the EMNIST dataset separately according to its character type and provides 89% accuracy for characters in the form of letters and numbers, 95% for characters in the form of letters, and 99% for characters in the form of numbers. Implementation of the handwriting recognition on the questionnaire image shows good results with 83.33% accuracy. However, there are problems found in the process of character segmentation where characters are not segmented correctly because the line of writing that continues on the character that should be separated and disconnected characters when they should be joined. The result obtained in this study is expected to be a consideration regarding the entry method data used by BPS later.

1. Introduction

In this digital era, the use of digital devices has changed how work related to texts and documents is done [1]. However, the use of paper-based media will still be an option for some people or organizations [2]. One example that cannot be separated from the use of paper media is the Central Statistics Agency (BPS).

BPS is a non-ministerial government agency whose role is to provide data needs for the government and the community. In data collection activities, the dominant method used by BPS is PAPI (Paper and Pencil Interviewing) [3]. PAPI is an interview method in which the enumerator holds a paper questionnaire, reads the questions to the respondent, then fills in the respondent's answers to the questionnaire by hand. Unlike CAPI (Computer-Assisted Personal Interviewing) or computer-assisted interviews, PAPI requires almost no technical skills to be applied and it is easier for enumerators or respondents to provide open or qualitative answers [4]. The consequence of the implementation of PAPI is that manual data entry activities are required from the handwritten data on the PAPI questionnaire into the database using human labor.

The average speed of a professional typist is 50 to 80 words per minute, so that would be approximately 200 pages per hour for a form with 15 one-word blocks and not including reading and sorting time [5]. In terms of costs, Regulation of the Head of BPS RI No. 51 of 2013 states that the average cost of data entry activities is 10.000 IDR per questionnaire. Through this description, the idea emerged to implement a technology that involves handwriting recognition and image processing to automate data entry activities. Furthermore, a comparative study was conducted on the effectiveness and efficiency between the data entry method using this technology and the conventional or manual data entry method.

Humans can recognize an object easily. The more often someone sees an object, the more familiar with the object. In recognizing objects, humans process visual information in semantic space by extracting meaningful features such as line segments, boundaries, shapes, and so on [6]. For computers, the ability to recognize objects has its challenges compared to humans who consider it as an easy thing [7].

Handwriting recognition in particular has been the subject of many studies [8]. Some of the problems encountered in handwriting recognition are high data uncertainty due to the different characteristics of each person's handwriting, some characters have similar shapes, there are broken or distorted characters, differences in the thickness of the characters written, and the use of various scanners [9]. Handwriting recognition cannot be separated from image processing which is identical to pattern recognition.

Pattern or object recognition is usually done by feature extraction and classification. The feature extraction process typically uses various methods to obtain a representation of the data. After that, the classifier is used to process the data classification. These processes are still done manually and separately. Recently, feature extraction and classification are integrated automatically into a single method. This method is commonly used to model data with a high level of abstraction and is often known as a deep learning technique [10] .Deep learning is an implementation of basic machine learning concepts that apply neural network algorithms, but with more layers. From the large number of hidden layers that are used between the input and output layers, a neural network can be said to be a deep neural network.

One of the deep learning architectures that can be used to recognize an object in a digital image is the Convolutional Neural Network (CNN) [11]. CNN is a type of Neural Network that is specifically for processing data that has a grid-shaped topology. CNN is claimed to be the best model in handling object detection and object recognition. In digital image recognition, the recognition accuracy can rival humans in the case of certain datasets [12].

As with other deep learning models, CNN has a challenge in the length of the model training process. But along with the development of increasingly sophisticated hardware, this can be overcome using a Graphical Processing Unit (GPU), one of which is on the Tensorflow platform. The system from Tensorflow can accommodate a wide variety of algorithms, including training and inference algorithms for deep neural network models. Tensorflow has been used to conduct research and apply machine learning systems to production in many fields of computer science and other fields, such as speech recognition, computer vision, robotics, information retrieval, natural language processing, geographic information extraction, and computational drug discovery [13].

This study takes a case study of the SAKERNAS questionnaire because the Analysis of the Results of the Survei Kebutuhan Data (SKD) 2019 shows that when viewed from the data source, the most microdata obtained from the Central BPS data provider is SAKERNAS microdata. In addition, SAKERNAS is a survey conducted in a fairly frequent frequency, namely twice a year. Therefore, efforts to streamline the existing processes in the implementation of SAKERNAS will be required by BPS. In addition, the characteristics of the contents of the questionnaire in the SAKERNAS questionnaire can represent the characteristics of the contents in other BPS questionnaires because they include open and closed entries, letter entries, number entries, letters and numbers entries, and checkmark entries.

Based on the description of the problems above, this study will build a handwriting recognition model to be implemented in the SAKERNAS August 2020 questionnaire and then conduct a comparative study between the data entry method using the handwriting recognition model and the manual data entry method. The handwriting recognition model that is built will use the Convolutional

Neural Network algorithm with the Tensorflow library. The results of the study of data entry methods are expected to be used as considerations or even recommendations in determining data entry methods at BPS in the future. Furthermore, the handwriting recognition model obtained is expected to be applied to other BPS activity questionnaires.

2. Related Works

Several previous studies related to the topic of handwriting recognition and used as references for this research.

Trnovszky et al. [14] compared CNN with several other classification methods in animal object recognition. The classification methods compared to CNN are Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Local Binary Patterns Histograms (LBPH), and Support Vector Machine (SVM). The dataset used in that research is animal dataset of 500 subjects which is divided equally into 5 classes. The results showed that of all the methods compared in classifying, the use of the CNN method gave the best results with an accuracy rate of 98%. It shows that the CNN method is very good to be implemented in the classification of an image.

Bhagat et al. [15] use handwriting recognition that was carried out using the Optical Character Recognition (OCR) method. The working principle is divided into six stages: obtaining images, preprocessing, segmentation, feature extraction, classification, and post-processing. OCR offers processing with high speed and accuracy.

Nath et al. [16] state that the efficiency of handwriting recognition depends on the feature extraction and classification method. That study reviews the comparison of methods and accuracy of several studies related to handwriting recognition from various languages. The result shows that the best classification method is feed-forward neural network and counterlet extract.

Another study on handwriting recognition in [17] implemented a Neural Network which was developed into CNN to classify the MNIST Digits dataset. The results show that the CNN model architecture can improve performance in solving handwriting recognition problems in the form of digits.

Research [18] constructed a handwriting recognition system on form documents that implements CNN in the feature extraction process and SVM in the classification process. The system built in this research includes preprocessing, segmentation, and character recognition. The preprocessing method used in that research is Median Filtering technique.

In research [3], a handwriting recognition model has been successfully built at the level of character prediction using CNN which is implemented on the Sensus Penduduk 2020 PAPI questionnaire in predicting the contents of the questionnaire. In the image processing stage, a bounding box template is used which does not yet accommodate the type of open field box where each character field is not limited by a vertical line. As a result, the prediction scheme has an accuracy of 78.35% in recognizing handwriting on the Sensus Penduduk 2020 questionnaire.

3. Proposed Method

This section contains explanation of research methods to achieve research objectives.

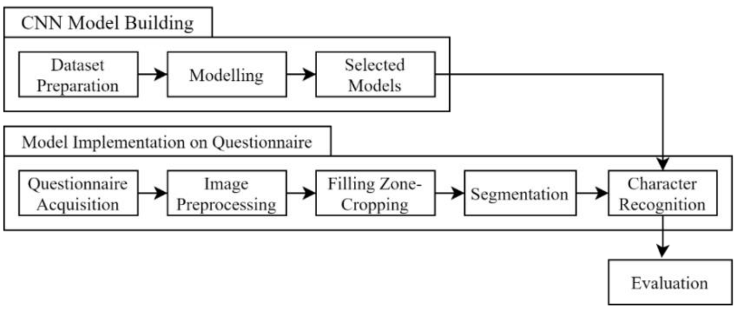

Figure 1. Research Method.

3.1. CNN Model Building

3.1.1. Dataset Preparation. The data used as training data is the Extended Modified NIST (National Institute of Standards and Technology) dataset as known as EMNIST. The EMNIST dataset is a development of the NIST Special Database 19 dataset which contains a collection of handwritten letters and numbers from more than 500 people [19]. The use of this dataset is because it is free and has been widely used as training data in several studies on handwriting recognition described in the related work section.

The EMNIST dataset is divided into six sub-datasets: EMNIST Balanced, EMNIST by Merge, EMNIST by Class, EMNIST Letters, EMNIST Digits, and EMNIST MNIST. From the six EMNIST sub-datasets, this study only uses three sub-datasets, namely EMNIST Balanced--consists of both letters and digits, EMNIST Letters--consists of only letters, and EMNIST MNIST--consists of only digits characters. Therefore, three models with different characteristics will be constructed to be implemented according to the type of questionnaire filling zone. The resulting model from the EMNIST Balanced training data will be implemented in a box consisting of both letters and numbers, the resulting model from the EMNIST Letters training data will be implemented in a box consisting of only letters, and the resulting model from the EMNIST MNIST training data will be implemented in the box consisting of numbers only.

Each of the sub-datasets will be divided into training data and test data. EMNIST Balanced will be divided into 112,800 train data and 18,800 test data. EMNIST Letters will be divided into 88,800 train data and 14,800 test data. EMNIST MNIST will be divided into 60,000 train data and 10,000 test data. the training data will be used for the model learning stage while the test data will be used to measure model's testing accuracy. While testing the implementation of the model will use 10 images of the second page of the questionnaire. Page 2 was chosen because it already contains all types of entries in the questionnaire and the total characters it contains is the most compared to other pages.

3.1.2. Modelling. Handwriting recognition models built in this study use CNN. CNN architecture accommodates feature extraction and classification processes, CNN extracts feature maps from 2-dimensional images and then uses them in the classification process.

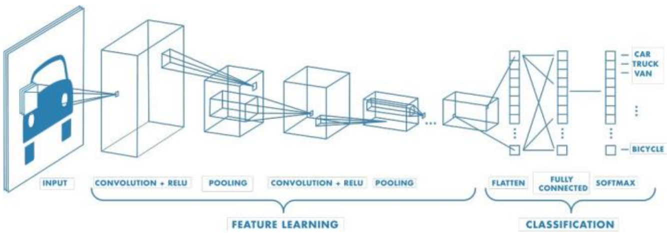

Figure 2. CNN Architecture

Almost all CNN architectures have the same general design principles, applying a convolution layer to the input data, performing downsampling (Max pooling) on the spatial dimensions along with increasing the number of feature maps regularly. There are also fully connected layers, activation functions, and loss functions. However, of all the operations in the CNN, the core of the CNN are convolutional layers, pooling layers, and fully connected layers [20].

In convolutional layers, convolution process is carried out on the input data using filters/kernels to generate a feature map. The filter takes adjacent pixels in the data image and then extracts their position relationship. Convolution is done by shifting the filter in all parts of the input where the "dot" operation is performed to produce a feature map. After getting the feature map, downsampling is applied to reduce the dimensions of the feature map so that it can speed up the computing process and prevent overfitting. The pooling technique used here is Max pooling which takes the largest value from the set of values in the filter. Furthermore, the feature map which is still in the form of a multidimensional array is transformed into a column vector as input for the fully connected layer. The fully connected layer is a layer where all activation neurons from the previous layer are connected to be further processed for classification. The activation function is a function that describes the relationship between the level of internal activity and can be either linear or non-linear [21]. It is the function that determines whether the neuron is activated or not. ReLU is a nonlinear activation function in the neural network that returns only positive values and otherwise is 0 [22]. The ReLu function can be described by equation 1 below:

$$f(x) = max(0, x) \tag{1}$$

Softmax activation function is used in classifying. Softmax is a function that calculates the probability value of each category/label. The calculation can be described in equation 2 below:

$$softmax(x)_i = \frac{e^{xi}}{\sum_{j=1}^{n} e^{xj}} \tag{2}$$

In order to prevent overfitting and speed up the learning process, dropout technique is used. Dropout is a neural network regularization technique in which some neurons will be randomly selected and not used during training.

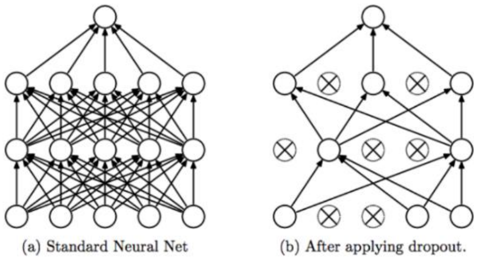

Figure 3. Dropout Regularization

The learning model phase is equipped with a loss function, optimizer, and metrics. The loss function is used to calculate the model's performance loss in predicting the target. The loss value can be minimized by using the optimizer. In this study, the loss function used is a categorical loss entropy with the Adam optimizer. The metric used is accuracy--to describe how precisely the model can predict correctly.

Experimental approach is used in order to determine the architecture of the models. First is determining how many pairs of convolutional-pooling that will be used. With an image dimension of 28x28 pixels, it will only be tested whether one, two, or three pairs show optimal results. In addition, it is not possible to apply 4 pairs because the image dimensions will be too small. Next, a similar approach is taken in determining the size of the feature maps, dense layers, and dropouts. Finally, an experiment was carried out by applying several advanced features that were claimed to improve the performance of the model that had been made [23] [24]. The advanced features that will be tested are changing the 5x5 convolution with two 3x3 convolutions, replacing the pooling layer with 5x5 convolution with strides 2, applying batch normalization, and applying data augmentation.

The experiment to determine the model architecture will use EMNIST MNIST to build the Digit model by considering the smaller size and number of classes so that the length of the model training experiment will be relatively faster. Next, the selected architecture will be applied to the Letter model using EMNIST Letters and the Mixed model using EMNIST Balanced.

3.1.3. Selected Models. The models that have been built will later be implemented according to their respective characteristics. The Letter model is used to recognize letter entries, the Digit model is used to identify numeric entries, and the Mixed model is used to identify entries that contain both letters and numbers. In addition, the accuracy of each model in recognizing handwriting will be calculated. Accuracy is defined as the percentage of data classified to the correct class and is calculated using the following equation 3:

$$Accuracy = \frac{TP+TN}{P+N} \tag{3}$$

3.2. Model Implementation on Questionnaire

3.2.1. Questionnaire Acquisition. At this step, scanning is carried out on the SAKERNAS 2020 questionnaire which has been filled out by the enumerator. Furthermore, the scanned image will be used to implement the handwriting recognition model that has been built.

3.2.2. *Image Preprocessing.* The pre-processing stage of the image is very important in character recognition which includes noise removal, binarization, skeletonization, and normalization [25]. Noise is a level of unwanted value in the image that does not give a significant value to the output [15]. The noise removal technique used is the median filter. The median filter sorts all the pixels in a given area and replaces the centre pixels with the median of the sorted values. Binarization is the process of converting a colour or grayscale image into a bi-level image. The adaptive thresholding method is the method used in this binarization process. This method calculates the threshold value for a small part of the image so that it is possible to obtain different values for each part of the same image, Skeletonization is a morphological process that transforms an image into a one-pixel wide representation without affecting its connectivity. The thinning process is needed to get the image frame by removing pixels that have more than two neighbours. Normalization includes normalizing the size and slope of the image. Size normalization is done by changing the image size to A4 paper size. Slope normalization is done by matching the input image to the image that has been used as a template.

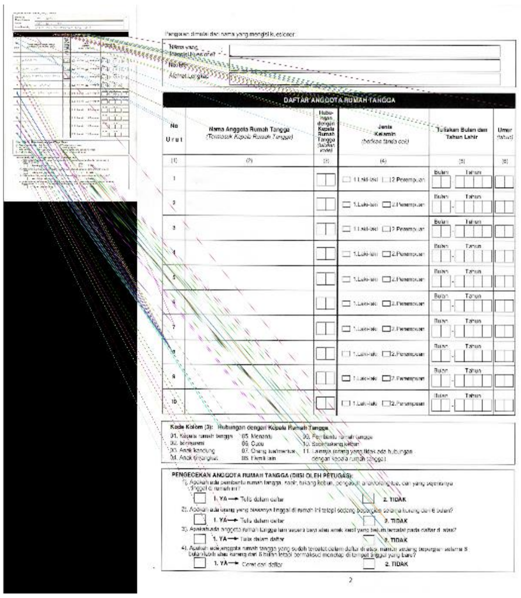

Figure 4. Image Matching

3.2.3. Filling Zone-Cropping. After the image has gone through the pre-processing stage, cropping is carried out for each entry zone in the questionnaire using a bounding box template that has been created by marking the coordinates of each field.

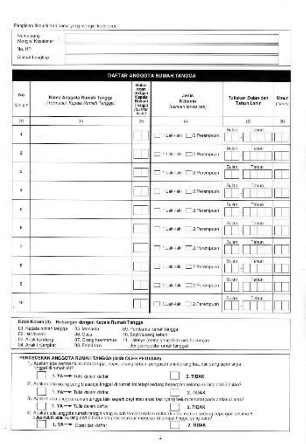

Figure 5. Questionnaire Page 2

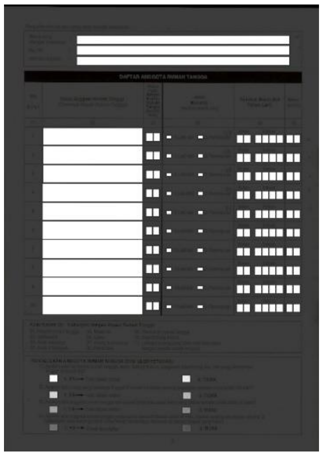

Figure 6. Filling Zone Page 2

3.2.4. Segmentation. The image of the questionnaire that has been pre-processed will be cropped in its contents section. For each cropping result, segmentation is carried out to the character level to be predicted using the model.

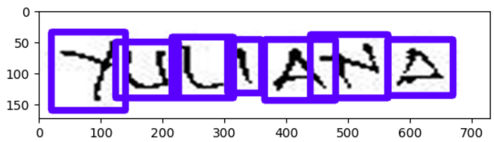

Figure 7. Image Segmentation

3.2.5. Character Recognition. The segmented character image is used as input data in the handwriting recognition model according to the type of content to identify the character. For the checkbox type, the method used to identify the character input is by counting the number of black pixels in the cropped image of the checkbox type.

4. Results and Discussion

The search for the CNN model architecture to be used begins with determining how many pairs of convolutional-pooling are optimal. The following are the accuracy results obtained on running models of 1 pair, 2 pairs, and 3 pairs.

Table 1. Experimental Results in Determining How Many Pairs of Convolutional-Pooling Layers

Convolutional-Pooling Pair	Train Accuracy	Test Accuracy
1	0.99996	0.99167
2	0.99996	0.99325
3	0.99981	0.99442

From the above experiment, it can be seen that 3 pairs of convolutional-pooling are slightly better than 2 pairs. However, in order to streamline additional computational costs, 2 pairs of convolutional-pooling are chosen as the first architectural parameter. Next, an experiment was conducted to determine the size of the feature map to be included in the 2 pairs of convolutional-pooling layers.

Table 2. Experimental Results in Determining Feature Maps Size

Maps	Train Accuracy	Test Accuracy
8 – 16	0.99981	0.99150
16 – 32	0.99998	0.99383
24 – 48	0.99998	0.99408
32 – 64	0.99998	0.99408
48 – 96	0.99998	0.99442
64 – 128	0.99998	0.99392

From the table above, it can be seen that the 24 feature maps in the first convolution layer and 48 feature maps in the second convolution layer are the best. The 32-64 feature map combination yields the same accuracy as 24-48 but was not selected due to the larger number. In addition, the combination of 48-96 feature map pairs with double the size only performs slightly better and is not worth the additional computational cost. Furthermore, an experiment was conducted to find the size of the dense layer by applying a pair of feature maps as much as 24 and 48.

Table 3. Experimental Results in Determining Dense Layer Size

Dense Layer Size	Train Accuracy	Test Accuracy
0	0.99994	0.99233
32	0.99996	0.99350
64	0.99996	0.99367
128	0.99996	0.99458
256	0.99998	0.99517
512	0.99996	0.99483
1024	0.99990	0.99475
2048	0.99998	0.99342

It can be seen that the best size is 256 units of layer. The next step is to apply it to the experiment to find the optimal dropout size.

Table 4. Experimental Results in Determining Dropout Size

Dropout	Train Accuracy	Test Accuracy
0	0.99998	0.99325
0.1	0.99985	0.99442
0.2	0.99944	0.99442
0.3	0.99881	0.99433
0.4	0.99710	0.99417
0.5	0.99475	0.99475
0.6	0.98971	0.99392
0.7	0.98258	0.99133

Dropout of 0.3 is the best measure from the experimental result. With this, we have obtained a basic CNN architecture (without implementing advanced features) which shows the best results. However, a final experiment will be carried out by applying advanced features to find out whether there is a significant difference in the accuracy of the model.

Table 5. Experimental Results in Applying Advanced Features

Advanced Features	Train Accuracy	Test Accuracy
CNN basic	0.99779	0.99517
32C3-32C3	0.99854	0.99550
32C5S2	0.99925	0.99492
both+BN	0.99919	0.99583
both+BN+DA	0.99616	0.99675

The application of advanced features in the form of changing a 5x5 convolution into two 3x3 convolutions, replacing the pooling layer with a 5x5 convolution with strides 2, applying batch normalization, and applying data augmentation has proven to improve the accuracy of the basic CNN model. Therefore, the architecture of the model chosen to be implemented in the CNN model development in this study is in the form of a convolutional layer with 24 feature maps followed by a convolutional layer with 48 feature maps, a dense layer with a size of 256 units, a dropout of 30% and the application of advanced features.

With the selected architecture that has been implemented for the construction of the three models, the accuracy for each model is obtained in table 6.

Table 6. CNN Models Test Accuracy

Model	Dataset	Test Accuracy
Digits	EMNIST Balanced	99%
Letter	EMNIST Letters	95%
Mixed	EMNIST MNIST	89%

Next, the model is implemented on the second page of the questionnaire image to identify the characters in the contents. From the implementation of the model that has been done, the accuracy of the character segmentation process and the accuracy of the model in recognizing the character input will be calculated. With the correct segmentation process, exactly one complete character will be obtained and from there, the accuracy of the model in recognizing handwritten characters will be calculated.

The following results from the calculation of the implementation are presented in table 7.

Table 7. Model Implementation Accuracy on the Questionnaire

Variables	Total of Process	True Process	Accuracy
Overall Character Segmentation	1021	846	82.85%
Character Segmentation on Open Fill Box	685	519	75.76%
Character Segmentation on Closed Fill Box	336	327	97.32%
Character Recognition Using CNN	846	705	83.33%
Character Recognition on Checkbox	68	68	100%

Since the built model is still limited to character level recognition, when the cropped image is not segmented properly, the character recognition results are inaccurate. This is because the model is not capable of recognizing two or more characters in one recognition process. Here's an example of the recognition result on open fill box in Figure 10.

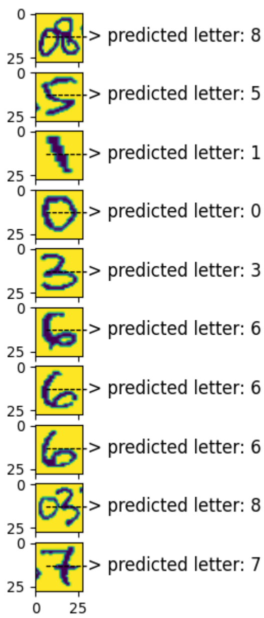

Figure 8. Example 1 of Open Fill Segmentation Result

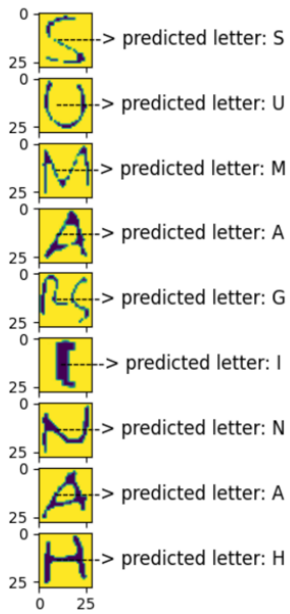

Figure 9. Example 2 of Open Fill Segmentation Result

In addition, from closed fields, there are also character segmentation problems found. It is because the character that was supposed to be whole is truncated. The following is an example of the introduction of closed fields in Figure 11.

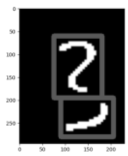

Figure 10. Example of Closed Fill Segmentation Results whose Characters are Disconnected

Apart from the character segmentation error, the accuracy of the implementation of character recognition by the CNN model is 83.33% which can be said to be good. However, there are findings of letter character recognition errors when the implementation of the letter model is carried out on letter characters that have similar shapes. This is one of the factors that cause the accuracy of the CNN model implementation to decrease. The following is an example of an error in recognizing letters between "O" and "D".

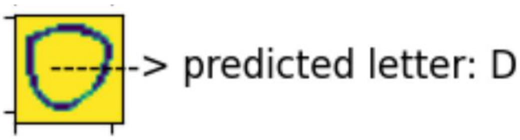

Figure 11. Character "O" recognized as "D"

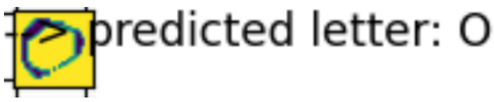

Figure 12. Character "O" recognized Correctly

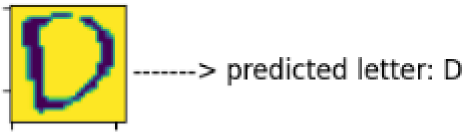

Figure 13. Character "D" recognized Correctly

The following is an example of an error in recognizing letters between "N" and "H".

Figure 14. Character "N" recognized as "H"

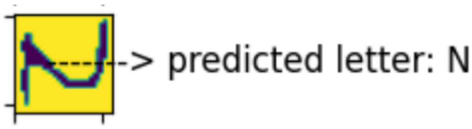

Figure 15. Character "N" recognized Correctly

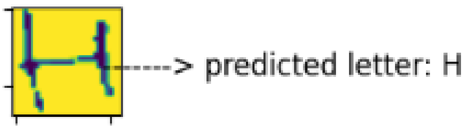

Figure 16. Character "H" recognized Correctly

To overcome errors like above, adjustments can be made from the model training dataset side to the writing characteristics of the enumerators or from the writing side of the enumerators whose characters match the writing characteristics of the dataset. This is because the CNN model learns a handwriting characteristic based on the characteristics of the training data used.

5. Conclusions

Three models of handwriting recognition using CNN have been successfully built to recognize entries in the SAKERNAS 2020 questionnaire. When the accuracy of the model is tested using EMNIST test data, the model for recognizing letter entries has an accuracy of 95%, the model for recognizing numeric entries has an accuracy of 99%, and the model for recognizing both letters and numbers has an accuracy of 89%.

When the model is implemented in the image questionnaire, the results of character recognition produced are quite accurate, namely 83.33%. In the character segmentation process, the overall accuracy is 82.85%. In the open field, there are findings in the segmentation process where several characters that should be separated are combined into one character because the writing lines between these characters are unified or do not have empty spaces as separators so that the accuracy obtained is 75.76%. While in the closed entry, the segmentation process is constrained by disconnected characters so that an accuracy of 97.32% is obtained. In addition, it can be due to poor image pre-processing parameters or poor input image quality. Therefore, it is also necessary to pay attention to the device used for the questionnaire scanning process and the condition of the questionnaire used.

References

[1] J. Harber, M. A. Nacenta and S. Carpendale, "Paper vs. Tablets: The Effect of Document Media in Co-located Collaborative Work," AVI '14: Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces, pp. 89-96, 2014.

[2] A. Sellen and R. Harper, The Myth of The Paperless Office, MIT Press, 2003.
[3] A. L. Irfiansyah, "Pembangunan Model Tulisan Tangan Pada Form Kuesioner PAPI (Studi Kasus Kuesioner Sensus Penduduk 2020)," 2020.
[4] The World Bank Group, "DIME Wiki," 2021. [Online]. Available: https://dimewiki.worldbank.org/wiki/Pen-and-Paper_Personal_Interviews_(PAPI).
[5] T. R. Ostrach, "Typing Speed: How Fast is Average: 4,000 typing scores statistically analyzed and," 2012.
[6] B. Zhang, "Computer Vision vs. Human Vision," in *Proc. 9th IEEE Int. Conf. on Cognitive Informatics (ICCI'10)*, 2010.
[7] N. Pinto, D. D. Cox and J. J. D. Carlo, "Why is Real-World Visual Object Recognition Hard?," *PLoS Computational Biology*, pp. 151-156, 2008.
[8] K. A. Hamad and M. Kaya, "A Detailed Analysis of Optical Character Recognition Technology," *International Journal of Applied Mathematics, Electronics and Computers*, pp. 244-249., 2016.
[9] N. Sharma, T. Patnaik and B. Kumar, "Recognition for Handwritten English Letters: A Review," Vols. 2, no. 7, pp. 318-321, 2013.
[10] Janalta Interactive Inc., "definition," 7 March 2021. [Online]. Available: https://www.techopedia.com/definition/30325/deep-learning.
[11] I. Goodfellow, Y. Bengio and A. Courville, Deep Learning (Adaptive Computation and Machine Learning Series), The IMT Press, 2016.
[12] A. Coates, H. Lee and A. Y. Ng, "An Analysis of Single-Layer Networks in Unsupervised Feature learning.".
[13] Google Research, "Tensorflow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," *Preliminary White Paper*, 2015.
[14] T. Trnovszky, P. Kamencay, R. Orjesek, M. Benco and P. Sykora, "Animal Recognition System Base On Convolutional Neural Network," *Digital Image Processing And Computer Graphics*, pp. 517-525, 2017.
[15] M. S. S. Bhagat, M. A. R. Joshi, M. V. S. Gajbhiye, M. S. R. Nandanwar and P. M.M.Ingle, "Handwritten Character Detection Using Optical Character Recognition Method," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 6, pp. 4724-4726, 2018.
[16] S. N. R. S and A. S, "Handwritten Character Recognition – A Review," International Journal of Scientific and Research Publications, vol. 5, no. 3, 2015.
[17] W. Zhu, "Classification of MNIST Handwritten Digit Database using Neural Network," 2018.
[18] Darmatasia and M. I. Fanany, "Handwriting Recognition on Form Document Using Convolutional Neural Network and Support Vector Machines (CNN-SVM)," *2017 Fifth International Conference on Information and Communication Technology (ICoICT)*, p. 1, 2017.
[19] G. A. S. T. J. &. v. S. A. Cohen, "EMNIST: An extension of MNIST to handwritten letters," *ArXiv.*, 2017.
[20] M. A. Hossain and M. M. Ali, "Recognition of Handwritten Digit using Convolutional Neural Network (CNN)," *Global Journal of Computer Science and Technology: D Neural & Artificial Intelligence*, vol. 19, no. 2, pp. 27-33, 2019.
[21] T. Nurhikmat, Implementasi Deep Learning Untuk Image Classification Menggunakan Algoritma Convolutional Neural Networks (CNN) Pada Citra Wayang Golek, Yogyakarta: Universitas Islam Indonesia, 2018.
[22] T. Wood, "Activation Function," DeepAI, [Online]. Available: https://deepai.org/machine-learning-glossary-and-terms/activation-function#:~:text=What%20is%20an%20Activation%20Function,%22%2C%20otherwise%20it%20does%20nothing. [Accessed 27 3 2021].

[23] J. T. Springenberg, A. Dosovitskiy, T. Brox and M. Riedmiller, "Striving for Simplicity: The All Convolutional Net," in *International Conference on Learning Representations*, San Diego, 2015.

[24] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," 2015.

[25] A. &. S. D. Pal, "Handwritten English Character Recognition Using Neural Network," vol. 1, no. 2, p. 141–144, 2010.