JURNAL RISET INFORMATIKA
Vol. 4, No. 4. September 2022
P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v4i3.435
Accredited rank 3 (SINTA 3), excerpts from the decision of the Minister of RISTEK-BRIN No. 200/M/KPT/2020

ALEXNET ARCHITECTURE AND FUZZY ANALYSIS ON TALENT JUDGE DECISION PREDICTION BASED ON FACIAL EXPRESSION

Muhammad Zaki[-1*)], Anggunmeka Luhur Prasasti [-2], Marisa W. Paryasto [-3]

Department of Computer Engineering, School of Electrical Engineering
Telkom University
https://telkomuniversity.ac.id/
Bandung, Indonesia
mhdzaki@student.telkomuniversity.ac.id[1*)], anggunmeka@telkomuniversity.ac.id[2],
marisaparyasto@telkomuniversity.ac.id[3]
(*) Corresponding Author

Abstract

The expression on the human face is a means of non-verbal communication. In the talent search event, the facial expressions shown by the judges when watching the participants' performances became one of the components to see whether the contestant who was performing could qualify for the next round or he would fail. Haar cascade is used to provide the location of the face in the frame and to classify the expressions on the face, a CNN model with modified AlexNet architecture is used which increases the accuracy by 5% from the original alexnet. A fuzzy Algorithm is used to predict the judge's decision based on how many facial expressions appear during the participant's appearance. The decision prediction system for talent search judges based on facial expressions using fuzzy is considered effective in predicting decisions, after being tested the system can predict decisions with an accuracy rate of 83%.

Keywords: AlexNet; CNN; Facial Expression Recognition; Fuzzy

Abstrak

Ekspresi pada wajah manusia merupakan salah satu sarana komunikasi non verbal. Pada ajang pencarian bakat, ekspresi wajah yang diperlihatkan para juri saat menyaksikan penampilan peserta menjadi salah satu komponen untuk melihat apakah peserta yang sedang tampil itu bisa lolos ke babak selanjutnya ataukah dia akan gagal. Haar cascade digunakan untuk memberikan lokasi wajah didalam frame dan untuk mengklasifikasi ekspresi yang ada pada wajah tersebut digunakan model CNN dengan arsitektur AlexNet modifikasi yang meningkatkan akurasi 5% dari alexnet original. Algoritma Fuzzy digunakan untuk memprediksi keputusan juri berdasarkan seberapa banyak ekspresi wajah yang muncul dalam saat penampilan peserta. Sistem prediksi keputusan juri pencarian bakat berdasarkan ekspresi wajah menggunakan fuzzy dinilai efektif dalam melakukan prediksi keputusan, setelah diuji sistem dapat memprediksi keputusqan dengan tingkat akurasi sebanyak 83%.

Kata kunci: AlexNet, CNN, Fuzzy, Pengenalan Ekspresi Wajah.

INTRODUCTION

The need for technology is increasing daily, along with the human condition to use electronic goods like smartphones. To maintain the security of smartphones, users usually enter a password, and along with changes in time, the use of a password in the form of a combination of letters and numbers has changed to a facial recognition security system. The facial security system has the concept of facial recognition, which is to match facial input data with predetermined facial data. The concept of facial recognition can recognize a person's face but cannot determine what the person's expression is.

In the talent search event, the facial expressions of the judges when watching the participants' performances become one of the important components in determining whether the contestants who appear can pass to the next round or fail. Human facial expressions can produce a form of communication without words and also play an important role in the process of interaction between humans (Nafis, Navastara, & Yuniarti, 2020).In previous studies, machine learning was applied to recognizing expressions in videos (Isman, Prasasti, &

Nugrahaeni, 2021), (Afriansyah, Nugrahaeni, & Prasasti, 2021). If only using facial expression recognition followed by expression detection on video is still not enough to make predictions on human decisions. Then we need an algorithm that can predict decisions based on facial expressions.

The development of previous research (Isman et al., 2021)(Afriansyah et al., 2021) was carried out with the development of the convolution method on CNN and the addition of a fuzzy algorithm to predict the jury's decision based on the jury's expressions displayed when the contestants performed their performances. Implementing machine learning with a modified CNN AlexNet architecture performs expression recognition on the jury and designs a fuzzy algorithm to predict the jury's decision.

RESEARCH METHODS

The research was carried out in stages to achieve the objectives. The steps of the method carried out in this study are presented in Figure 1.

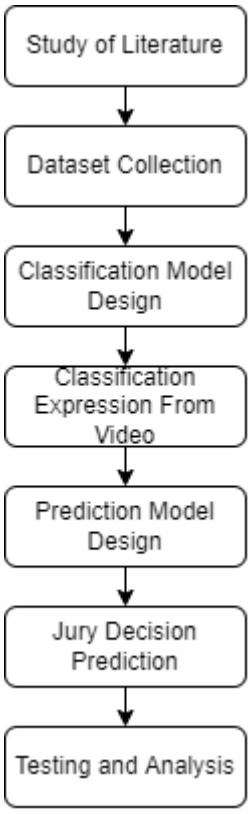

Figure 1. Research Method

Study of Literature

The first step in doing research is looking for references related to the research. References can be scientific journals, books, or related software documentation. After that, the related references will be studied, and the part suitable for the research will be used.

Dataset Collection

The dataset acquisition is divided into two parts: the first dataset for classification and the second video dataset for prediction. These data have their respective functions and uses.

The classification dataset is taken from Kaggle, an online provider of datasets. Expression data is divided into 7 basic expressions (Meng, Liu, Cai, Han, & Tong, 2017),(Zahara, Musa, Prasetyo Wibowo, Karim, & Bahri Musa, 2020). Datasets are used for the training and validation process on the machine learning model in classifying emotions. Five expression data are used for the jury's expression classification model: Angry, Disgust, Happy, Neutral, and Surprise. The data in FER-2013 was modified by deleting some images that were not expressions that existed in the dataset, as seen in picture 2, and some images that were deemed not to match the expressions they should have.

Figure 2. Non-Expression Dataset Fer2013

The Video datasets are obtained from the YouTube video platform, data from the appearance of one of the participants is downloaded, and then cuts are made on the video to take part when the camera highlights the jury. Then the pieces are put together into one video. In one video, there is only one judge. The length of the video pieces is equal to 5 seconds.

Classification Model Design

In this research, the expression classification model uses CNN. Expression classification using the CNN model with alexnet architecture(Krizhevsky et al., 2017), alexnet architecture will be modified so that the model's accuracy in classifying expressions increases. The modified Alexnet CNN layer has several layers: a convolution layer, batch normalization layer, max pool layer, drop-out layer, and fully connected layer.

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v4i3.435
Accredited rank 3 (SINTA 3), excerpts from the decision of the Minister of RISTEK-BRIN No. 200/M/KPT/2020

Convolutional Layer is one part of the CNN architecture that extracts images into several parts to be used as feature maps by multiplying the input value by a filter that produces feature maps(Chang & Sha, 2017). An illustration of the convolution process can be seen in Figure 3. in addition to convolution kernel size and filters, the variable that forms the value of feature maps is padding and strides(Z. Li, Liu, Yang, Peng, & Zhou, 2021). Padding is a value of 0 at the edge of the input matrix, and stride is the number of pixels passed when shifting the kernel(Sun, Li, Zhou, & He, 2016).

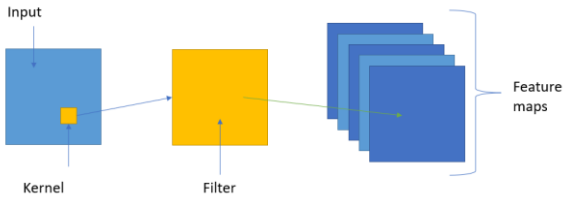

Figure 3. Convolution Process Overview

$$outLayer = \frac{i+2P-k}{S} + 1 \quad \text{.....(1)}$$

From equation 1, we can get the output of the width and height of the matrix after passing through the convolution layer. Where i is the height or width of the input matrix, P is the padding, k is the size of the convolution kernel, and S is the stride.

Training process on modified alexnet architecture such as In the input layer, which contains 48x48x1, adjusts to the size and dimensions of the image. Then the convolution layer uses an 11x11 kernel to retrieve more data on the image, then proceeds with the normalization bach layer to start the training process and improve implementation. Then, with the max-pooling layer taking the largest value from the feature map in the pool_size 2x2, the drop-out layer of 0.3 is used to control the number of dimensions and avoid overfitting.

Using three convolution layers with a 3x3 kernel with a batch normalization layer sequentially increases features, followed by a max pool layer with a pool_size of 2x2. The second convolution layer uses a kernel size of 5 x 5, followed by a batch normalization layer, a max-pooling layer, and a drop-out layer of 0.3. The flattened layer converts all m x n shapes to m x 1 shapes. Then it is connected to a fully connected layer with a dense of 4096 for two layers, and each fully connected layer is given a drop-out layer of 0.3 to avoid overfitting.

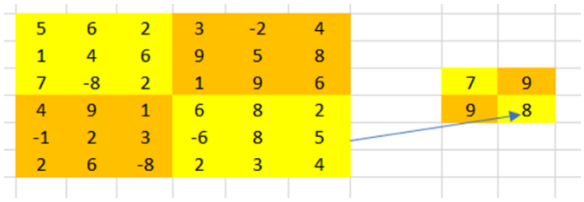

Figure 4. Max Pooling Illustration

Max pooling is a process to reduce data with the largest value from a cell(Wang et al., 2018). For example, in Figure 4, a 6 x 6 matrix with a pool size of 3 x 3 will produce a max pooling matrix of 2 x 2.

Batch Normalization is a layer on the convolutional neural network added after the convolution layer so that different elements on feature maps with other locations can be normalized in the same way to speed up the training process and increase the accuracy of the model (Chen, Yang, Wang, & Zou, 2017).

A drop-out layer added to the convolution layer can increase accuracy(Sungheon Park, 2017). The drop-out layer is a regularization process that aims to speed up the training process and prevent overfitting. It is also very effective for use on fully-connected layers.

In the classification process, the fully connected layer is the layer in which all the layers are connected. Before connecting, all arrays or feature maps are still in the form of two-dimensional arrays, first processed in the flattened layer to convert the two-dimensional array into a one-dimensional one so that classification can be carried out. (S. Li & Deng, 2018).

ReLU stands for Rectified Linear unit used to change the negative value in the matrix to 0. Softmax is a function used in the output layer, used to return the probability value of each class and if the sum of the probabilities of all classes will be one(Ramdhani, Djamal, & Ilyas, 2018), (Wang et al., 2018)

The difference between the alexnet architecture and the modified alexnet architecture is adding a batch normalization layer after the max pooling layer and a drop-out layer after the max-pooling layer.

Classification Expression from Video

The expression reading from the video starts with the video that is entered into the video capture method in the cv2 library. Then each frame will be converted to grayscale. After that, call the classifier variable containing the haar case cascade, which has entered the cascade classifier method to detect the position of the face in the frame. The model that has been trained will be used for the face classification process in the image, developed to be able to classify expressions directly from the video(Isman et al., 2021).

The bounding box containing the face will be resized following the input size in the CNN model. After resizing part of the face frame, it will be predicted using the CNN model, which was previously loaded using load_model from the Keras library.

Prediction Model Design

A fuzzy Is a logic with a vague or ambiguous meaning developed in 1965 by Lotfi A. Zadeh. Normal human logic only recognizes two values, namely one true and 0 false, while fuzzy has a value between 0 and 1(Saleha, 2020). Fuzzy algorithms can also be interpreted as logic that lacks clarity and contains uncertain elements.

In the system, the fuzzy algorithm was created using the python skfuzzy library designed using the membership value. The membership value is obtained by calculating the value of each expression in the video and making this value the limit between the membership of each expression.

The skfuzzy antecedent method begins by entering a variable value for each expression and decision. Followed by entering the membership value of each expression with two classes, with labels high and low, and the membership of the jury's decision with two classes, with labels yes and no, using the trapmf function from the skfuzzy library.

Rules are made with the skfuzzy function. To find the number of regulations, you can use the number of members to the power of the number of expressions. In this program, $2^{5,}$ which is 32. After all the rules are created, they will be initialized with the control system function of the skfuzzy control. The data derived from the calculation variable of the previous expression will be entered into the control system simulation function to be calculated and then used in the compute function to determine the decision.

Jury Decision Prediction

A jury decision prediction program combines a program for classifying expressions from videos with several additional methods, such as array processing and a method to count every expression in the video. The expression is calculated and entered into the fuzzy algorithm in the design part of the program prediction model.

Testing and Analysis

The expression classification model with the best value is obtained by combining each parameter. The parameters used are split data, batch size, and learning rate. The confusion matrix is used to find the accuracy, precision, recall, and f1 score value.

Fuzzy testing to predict talent judge decision by testing the fuzzy algorithm with video data other than video, which is used to determine the membership value. Video is entered into the jury's decision prediction program, and the program is a combination of classification expressions from video with a fuzzy algorithm.

RESULTS AND DISCUSSION

The best model for expression classification search combines parameters such as data splitting, batch size, and learning speed. Table 1 shows the results of the combination of each parameter.

Table 1. The value of accuracy, precision, recall, and f1 score

split	bs	lr	acc	prc	rcl	f1
9,1	64	1E-03	0.78	0.75	0.71	0.72
		1E-04	0.81	0.8	0.77	0.78
		1E-05	0.71	0.62	0.68	0.63
	32	1E-03	0.78	0.75	0.69	0.71
		1E-04	0.81	0.78	0.75	0.76
		1E-05	0.7	0.66	0.63	0.63
	16	1E-03	0.78	0.75	0.69	0.71
		1E-04	0.81	0.76	0.76	0.75
		1E-05	0.76	0.68	0.72	0.69
	8	1E-03	0.77	0.61	0.62	0.62
		1E-04	0.82	0.79	0.79	0.79
		1E-05	0.76	0.67	0.72	0.68
8,2	64	1E-03	0.79	0.75	0.73	0.74
		1E-04	0.79	0.74	0.74	0.74
		1E-05	0.68	0.6	0.63	0.61
	32	1E-03	0.76	0.71	0.71	0.7
		1E-04	0.75	0.73	0.68	0.7
		1E-05	0.71	0.63	0.65	0.64
	16	1E-03	0.77	0.76	0.66	0.67
		1E-04	0.79	0.75	0.75	0.75
		1E-05	0.73	0.64	0.68	0.65
	8	1E-03	0.75	0.6	0.59	0.59
		1E-04	0.8	0.75	0.75	0.75
		1E-05	0.75	0.66	0.69	0.67
7,3	64	1E-03	0.73	0.68	0.65	0.66
		1E-04	0.77	0.71	0.73	0.71
		1E-05	0.67	0.58	0.59	0.59
	32	1E-03	0.76	0.73	0.67	0.69
		1E-04	0.79	0.75	0.73	0.74
		1E-05	0.7	0.62	0.65	0.63
	16	1E-03	0.72	0.72	0.59	0.62
		1E-04	0.77	0.71	0.73	0.71
		1E-05	0.68	0.6	0.63	0.61
	8	1E-03	0.73	0.59	0.57	0.57
		1E-04	0.77	0.74	0.71	0.72
		1E-05	0.74	0.66	0.68	0.67
best 9,1	8	1E-04	0.82	0.79	0.79	0.79

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v4i3.435
Accredited rank 3 (SINTA 3), excerpts from the decision of the Minister of RISTEK-BRIN No. 200/M/KPT/2020

Accuracy, precision, recall, and f1 scores were obtained using the report_classification function in the sklearn library. In the combination in table 1, the highest value is in split 9.1, batch size 8, and learning rate 1E-04 or 0.0001.

Table 2 CNN Architecture Comparison

Architecture	ACC	PRC	RCL	F1-S
Alexnet Modification	0.82	0.79	0.79	0.79
Alexnet	0.77	0.78	0.68	0.71

With the same combination, the training process is carried out on the alexnet architecture, in table 2 shows differences in the values of accuracy, precision, recall, and f1 score. The Alexnet architecture modification has a higher score than the original alexnet architecture.

They were predicting the jury's decision with video outside of video data to find membership values. The fuzzy algorithm with the classification model can predict the jury's decision. Table 3 shows that 6 data videos used for the system test successfully predicted five videos correctly.

Table 3 Total Jury Expression Calculation and Jury Decision Prediction

Data	Angry	Disgust	Happy	Neutral	Surprise	Prediction
1	0	0	0	93	0	TRUE
2	2	0	47	50	17	TRUE
3	0	0	0	84	0	TRUE
4	19	0	0	78	0	TRUE
5	0	0	77	15	0	TRUE
6	73	1	12	3	0	FALSE

The judge's decision on data 6 in table 3 is 'yes,' but the system predicts the jury's decision is 'no.' It happens due to the lack of accuracy of the expression classification model in the classification process.

CONCLUSIONS AND SUGGESTIONS

Conclusion

The modified Alexnet architecture gets 5 percent higher accuracy than the original alexnet architecture using the same parameters. The Modified Alexnet Model is also effective in classifying expressions, with the highest accuracy being 81%, with an average f1-score for five expressions of 79%. Overall, the jury's decision prediction system is considered effective in classifying the jury's expressions and making correct predictions by 83%.

Suggestion

The dataset used should be changed to AffectNet to get a better expression classification. Add the use of feature extraction methods before the data goes to the convolution layer. This method can be the next development to get better results.

REFERENCES

Afriansyah, Y., Nugrahaeni, R. A., & Prasasti, A. L. (2021). Facial Expression Classification for User Experience Testing Using K-Nearest Neighbor. In *Proceedings - 2021 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology, IAICT 2021*. Retrieved from https://doi.org/10.1109/IAICT52856.2021.9532535

Chang, J., & Sha, J. (2017). An efficient implementation of 2D convolution in CNN. *IEICE Electronics Express*, 14(1), 20161134–20161134. Retrieved from https://doi.org/10.1587/elex.13.20161134

Chen, X., Yang, X., Wang, M., & Zou, J. (2017). Convolution neural network for automatic facial expression recognition. In *2017 International Conference on Applied System Innovation (ICASI)* (pp. 814–817). IEEE. Retrieved from https://doi.org/10.1109/ICASI.2017.7988558

Isman, F. A., Prasasti, A. L., & Nugrahaeni, R. A. (2021). Expression Classification for User Experience Testing Using Convolutional Neural Network. In *AIMS 2021 - International Conference on Artificial Intelligence and Mechatronics Systems*. Retrieved from https://doi.org/10.1109/AIMS52415.2021.9466088

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. Retrieved from https://doi.org/10.1145/3065386

Li, S., & Deng, W. (2018). Deep Facial Expression Recognition: A Survey. Retrieved from https://doi.org/10.1109/TAFFC.2020.2981446

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 1–21. Retrieved from https://doi.org/10.1109/TNNLS.2021.3084827

Meng, Z., Liu, P., Cai, J., Han, S., & Tong, Y. (2017). Identity-Aware Convolutional Neural Network

for Facial Expression Recognition. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)* (pp. 558–565). IEEE. Retrieved from https://doi.org/10.1109/FG.2017.140

Nafis, A. F., Navastara, D. A., & Yuniarti, A. (2020). Facial Expression Recognition on Video Data with Various Face Poses Using Deep Learning. *ICITEE 2020 - Proceedings of the 12th International Conference on Information Technology and Electrical Engineering*, 362–367. Retrieved from https://doi.org/10.1109/ICITEE49829.2020.9271740

Ramdhani, B., Djamal, E. C., & Ilyas, R. (2018). Convolutional Neural Networks Models for Facial Expression Recognition. *2018 International Symposium on Advanced Intelligent Informatics (SAIN)*, 96–101. Retrieved from https://doi.org/10.1109/SAIN.2018.8673352

Saleha, B., Nasution, S. M., & Prasasti, A. L. (2020). Design of IoT-based smart laundry applications using fuzzy algorithms. In *2020 International Conference on Information Technology Systems and Innovation, ICITSI 2020 - Proceedings* (pp. 393–397). Institute of Electrical and Electronics Engineers Inc. Retrieved from https://doi.org/10.1109/ICITSI50517.2020.9264936

Sun, B., Li, L., Zhou, G., & He, J. (2016). Facial expression recognition in the wild based on multimodal texture features. *Journal of Electronic Imaging*, 25(6), 061407. Retrieved from https://doi.org/10.1117/1.JEI.25.6.061407

Sungheon Park, N. K. (2017). Analysis on the Dropout Effect in Convolutional Neural Networks, 10112. Retrieved 19 July 2022 from https://doi.org/10.1007/978-3-319-54184-6_12

Wang, S.-H., Phillips, P., Sui, Y., Liu, B., Yang, M., & Cheng, H. (2018). Classification of Alzheimer's Disease Based on Eight-Layer Convolutional Neural Network with Leaky Rectified Linear Unit and Max Pooling. *Journal of Medical Systems*, 42(5), 85. Retrieved from https://doi.org/10.1007/s10916-018-0932-7

Zahara, L., Musa, P., Prasetyo Wibowo, E., Karim, I., & Bahri Musa, S. (2020). The Facial Emotion Recognition (FER-2013) Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi. In *2020 Fifth International Conference on Informatics and Computing (ICIC)* (pp. 1–9). IEEE. Retrieved from https://doi.org/10.1109/ICIC50835.2020.9288560

.