JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2765-2772

ANALYSIS OF DIGITAL IMAGE RECOGNITION OF INDONESIAN SIGN LANGUAGE USING THE DEEP LEARNING CNN ARCHITECTURE VGG19 METHOD

Dimas Prayoga*[1], Ema Utami[2], Dhani Ariatmanto[3]

1. Universitas AMIKOM Yogyakarta, Indonesia
2. Universitas AMIKOM Yogyakarta, Indonesia
3. Universitas AMIKOM Yogyakarta, Indonesia

Article Info

Keywords: Artificial Intelligence, Machine Learning, Deep Learning, Sign language

Article history:
Received 14 January 2025
Revised 20 February 2025
Accepted 8 March 2025
Available online 1 September 2025

DOI :
https://doi.org/10.29100/jipi.v10i3.7353

* Corresponding author.
Corresponding Author
E-mail address:
dimas32113@students.amikom.ac.id

ABSTRACT

This study examines the application of the CNN method with the VGG19 architecture for digital image analysis in recognizing Indonesian sign language. The data used in this study is the BISINDO data set type, with 8,814 samples divided into 26 alphabetical categories. Implementing sign language recognition using the VGG19 architecture produces good accuracy results, reaching 93.24% with epoch 25 (without hyper-parameters tuning).These results confirm the model's extraordinary ability in image recognition and performing precise analysis. However, the results of this study can be improved again by performing Hyper parameters tuning on the architecture used, namely VGG19, by changing certain variables that affect increasing accuracy. Other aspects can be improved to achieve optimal performance, considering the excellent results. By integrating modern hyper-parameter tuning approaches and incorporating a variety of additional data, the model generalization is expected to be improved, leading to higher accuracy in many real-world settings.

I. INTRODUCTION

Effective communication is vital in an individual's life, enabling them to interact with their surroundings. By engaging in communication, individuals acquire valuable information that aids in their adaptation to the environment. Typically, humans communicate through spoken language, utilizing sound as a medium to exchange information [1]. However, individuals with hearing impairments face challenges in using verbal language to communicate. Alternative methods, such as sign language or non-verbal communication, can address this issue. These forms of communication were explicitly developed to facilitate interaction with individuals with disabilities [2]. In some areas, sign language is very helpful in providing visual language that people studying the knowledge can understand. For example, special schools for students who are deaf and mute can understand the learning material taught by special teachers using sign language. Moreover, if combined with technology, students can learn sign language with sign language recognition technology based on data training that has been carried out using deep learning [3].It is important to note that there is no universal sign language globally, as different countries may have their unique sign languages. This diversity stems from sign language's linguistic characteristics akin to spoken languages [4]. Consequently, various types of non-verbal communication exist, including facial expressions, mouth movements, hand gestures, and body language. Among these, finger-based communication is particularly popular for effectively communicating with individuals with limited hearing abilities [5].

Indonesian Sign Language is the primary means of communication for deaf and hard-of-hearing individuals in Indonesia. Despite its importance, BISINDO recognition remains challenging due to the complexity and variability of hand gestures and the need for real-time and accurate interpretation [6]. This research is urgently needed to investigate and evaluate the effectiveness of deep learning approaches, specifically Convolutional Neural Networks (CNNs), in the context of Indonesian Sign Language recognition. Contribution to this research could improve communication accessibility for the deaf and hard-of-hearing community.

Research on communicating using sign language with others and oneself. Using media pipe and deep learning implementations can help optimize sign language recognition that incorporates Computer Vision for its detection media [7]. Overall, this research produced a high level of accuracy using two types of data, namely 2D and 3D images. Both data were processed using deep learning methods, and the Media Pipe approach was used with a feedforward neural network. It is expected that the research that the researchers will do will be done with deep learning methods with different architectural models and BISINDO datasets.

Research related to the use of technology used for sign language communication needs. The importance of communication for the good of each other, good communication makes the information received absorbed better which is beneficial to society including for people who are deaf and mute. This study discusses identifying the performance of deep CNN and VGG 16 methods against sign language datasets [8]. Getting an average accuracy result of 95% in each class and each class has static data, for real-time data or video is a shortcoming in this study. The research will be conducted using the VGG19 architecture and BISINDO dataset taken in realtime which is expected to provide better or optimal performance than previous studies.

Further research has been conducted on sign language, a unique form of communication involving various movements and postures used when interacting with individuals who are mute or deaf. These movements, predominantly performed by hand, can be comprehended and have potential applications in areas like Augmented Reality, gaming, robotics (utilizing sensors or cameras), and more [9]. Optimized for American Sign Language, the Deep CNN method achieved higher accuracy and validation levels than the deep CNN model tested, with a maximum accuracy rate of 94.34%, surpassing other advanced classification techniques. Future studies utilizing the deep learning method and BISINDO datasets in real-time aim to enhance the model's performance.

The next presents a novel approach for Korean Sign Language (KSL) recognition using a Transformer-based deep neural network. Sign language recognition plays a crucial role in bridging the communication between the deaf and hearing communities. Traditional methods often face challenges in accurately capturing the intricate movements and gestures of sign language [10]. Following a combination of local and long-term dependency features, classification modules are used to classify data. From the proposed model was assessed using our KSL benchmark dataset and lab dataset, resulting in an accuracy of 89.00% for the 77-label KSL dataset and 98.30% for the lab dataset. Superior performance indicates that the proposed model has common properties while requiring significantly fewer computing resources. In the upcoming research, a comprehensive investigation will be conducted utilizing the advanced CNN model integrated with the state-of-the-art VGG19 architecture. The research will be augmented by utilizing a meticulously curated dataset of the Indonesian Sign Language. By leveraging these cutting-edge resources, the study aims to culminate in the development of a high-performing system, capable of delivering exceptionally accurate results and paving the way for significant advancements in the field.

II. REASEARCH METHOD

In this study, there are several stages to get the final result of a model made. From collecting data from clear sources and already determined datasets that will be used when testing and training data, then pre-processing the data, after that data training, softmax and models [11], can be seen in figure 1.

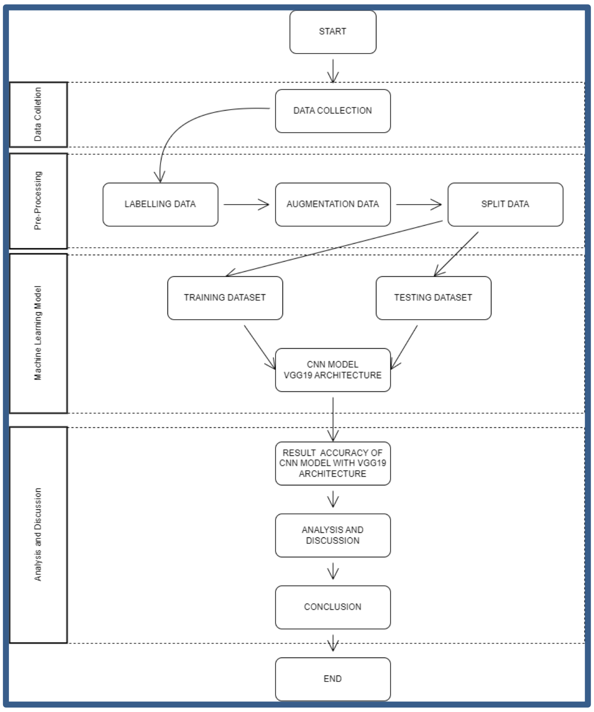

Figure 1. Diagram Workflow

A. Data Collection

The data used in this study was obtained from the Kaggle web dataset source, which provided a comprehensive dataset comprising 26 alphabet classes. Each class consisted of 140-150 photo files, and all photos had 640x640 pixels dimensions [12]. These images will become datasets used in research, and then raw data will be processed again in the data pre-processing section to make the dataset into good data or clean data can be seen in figure 2.

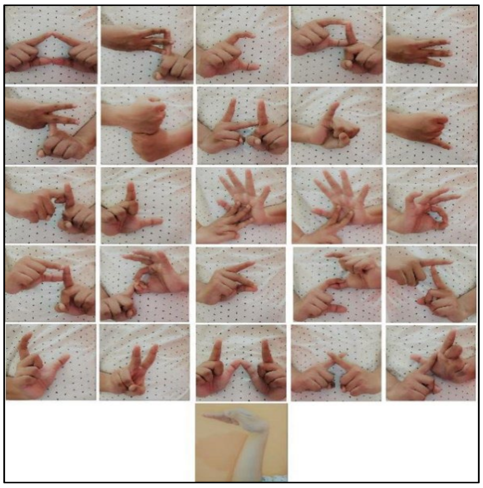

Figure 2. Images Dataset

B. Data Preprocessing

The initial stage of sign language recognition involves data preprocessing. This critical process meticulously converts and sets up raw sign language data, optimizing it for a machine learning model to analyze accurately and efficiently. Various techniques and methodologies in data preprocessing aim to enhance the overall quality of input data by reducing noise interference and disturbances and identifying relevant features that aid the model in learning and providing exact predictions [13]. This fundamental preprocessing stage forms the bedrock for the next steps in sign language recognition, laying the groundwork for successful and dependable outcomes from machine learning.

C. VGG 19

This comprehensive research employed a Convolutional Neural Neural (CNN) model, specifically harnessing the strength of the VGG19 architecture, to train a diverse range of image data extracted from a meticulously compiled dataset. The VGG19, an abbreviation for Visual Geometry Group with 19 layers, is a highly respected deep learning model. Renowned for its simplicity yet powerful capacity, VGG19 stands as a state-of-the-art model that enables high-level feature extraction from images, making it an excellent choice for rigorous image processing tasks in numerous fields of study [14]. the VGG19 architectural model can be seen in figure 3.

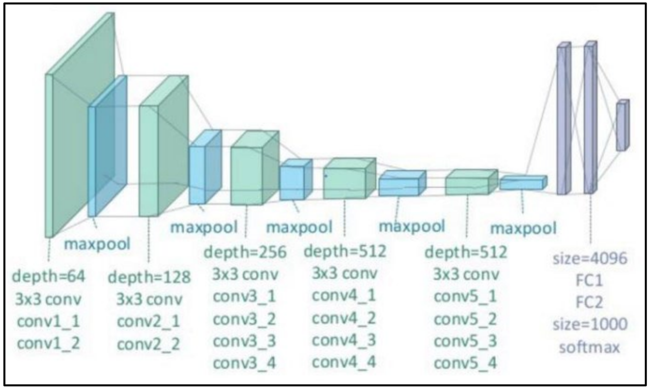

Figure 3. VGG19 Architecture

An established, intricately designed algorithm is subsequently employed to methodically train the existing, rich dataset, utilizing a meticulously predefined ratio to censure optimal learning and predictive performance [15].

D. Model Evaluation

The aim of conducting research using evaluation models like accuracy (1), recall (2), precision (3), and F1 score (4) is to assess classification models' performance in data processing thoroughly [16][17]. Precision indicates the ratio of correct predictions to the total number of predictions made. However, this metric may only partially represent the scenario when working with imbalanced data. Accuracy becomes critical when the repercussions of false optimistic predictions are substantial, making precision an essential consideration. Recall reflects the model's ability to accurately recognize all cases that should be classified as positive, which is especially important in situations where false negatives can have serious consequences. The F1 score, derived as the harmonic mean of precision and recall, offers a more balanced view of the two metrics, especially in class imbalance in data. By determining and balancing between precision, recall, and F1 score, we can ensure that the model performs well in relevant contexts and meets the application's needs [18]. The evaluation process focuses on measurement and assessment.

$$Accuracy = \frac{(TN+TP)}{(TP+TN+FP+FN)} \times 100\% \tag{1}$$

$$Recall = \frac{TP}{(TP+FN)} \times 100\% \tag{2}$$

$$Precision = \frac{TP}{(TP+FP)} \times 100\% \tag{3}$$

$$F1\ Score = \frac{2\times(recall\times precision)}{(recall\times precision)} \times 100\% \tag{4}$$

III. Result and Discussion

The dataset used in this study amounted to 8.814 data and was divided into 26 classes; each class contained approximately 150 data images. This data has been cleaned, and we can proceed to the next stage. In this study, the CNN algorithm with VGG19 architecture was implemented to recognize the data subjects used. Specifically in this study, the dataset was divided into 70% training data and 30% testing data. The results of the experiment can be seen in table 1.

TABLE 1
COMPARISON OF ACCURACY AND LOSS

Metode arsitektur	Epoch	Training		Testing	
		Accuracy	Loss	Accuracy	Loss
VGG19	25	96,00%	00,30%	93,239%	01,40%

From the research experiments conducted, it is evident that the findings are of significant importance. There is a difference in accuracy values between training and testing. Some datasets used in the research are not recognized or have low accuracy values when the research is carried out using the specified method. The loss value in training is lower than the loss value in testing, but the accuracy value in training is already above 90%. In this study, we are conducting another experiment to increase the accuracy value with the Hyper-parameter Tuning method. Several variables will be changed in their values, which were initially typical values in tuning, in order to get the expected results, namely an increase in the accuracy value and a decrease in the loss value. Can be seen in figure 4.

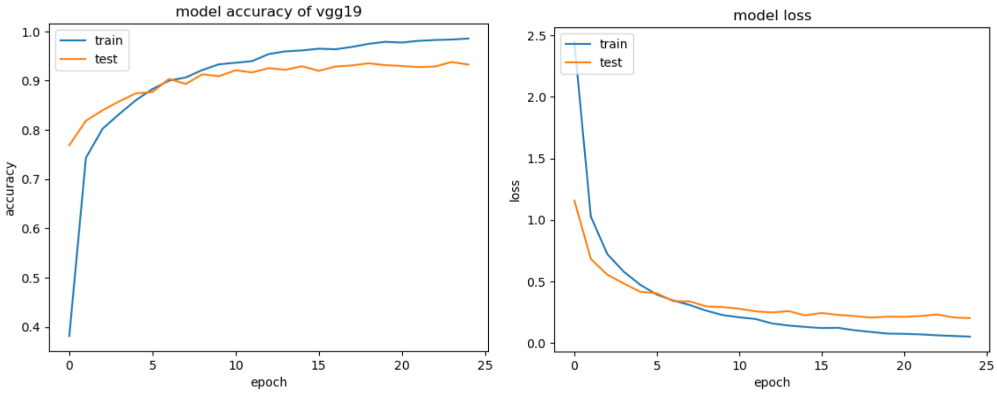

Figure 4. Accuracy and Loss VGG19

Comparison with the methods used in previous research, namely the VGG16 and VGG19 architectures, before hyper-parameter tuning was performed. can be seen in table 2.

TABLE 2
COMPARISON OF PROPOSED MODELS WITH
EXISTING MODEL

Metode arsitektur	Dataset	Result
VGG16 [19]	10,000 sign language samples	Accuracy: 91%
VGG16 [20]	8,500 sign language images	Accuracy: 92%
VGG19 [21]	8,000 sign language instances	Accuracy: 93%
VGG19	8.814 sign language image	Accuracy: 94,602%

In the research conducted, using the CNN method with the VGG19 architecture that has been done, hyper-parameter tuning obtained better results than previous research and better results than other architectures in previous research. VGG19 excels in data training and data testing where the results obtained exceed 90% and above and can be said to be outstanding results

The data testing phase, a critical stage that follows the training process, is primarily aimed at evaluating the model's performance across various scenarios. Each data point is meticulously tested during this phase to assess key performance metrics, including precision, recall, F1-score, dan accuracy. In ongoing research, three critical metrics must be considered, namely precision, recall and F1-score, because these three metrics provide a comprehensive view of the model's working system in recognizing positive and negative samples [22]. The results of this evaluation are presented using a confusion matrix, which provides a comprehensive overview of the model's effectiveness in accurately categorizing data, can be seen in table 3.

TABLE 3
COMPARISON OF ACCURACY AND LOSS

Metode Arsitektur	Parameter			Training		Testing	
	Epoch	learning rate	Batch size	Accuracy	Loss	Accuracy	Loss
VGG19	25	0,01	64	96,00%	00,30%	93,239%	01,40%
	50	0,01	64	97,00%	00,10%	94,057%	00,25%
	50	0,001	256	98,00%	00,10%	94,602%	00,25%

Table 2 demonstrates the performance of CNN Model with VGG19 architecture method in recognizing Indonesian sign language. Several parameters were changed to improve the method's performance, including three important parameters: Epoch (Iteration in training the model), Learning Rate (Setting the size of the step in updating the model weight during training), and Batch Size (Number of training samples used in 1 iteration). There are 3 scenarios in this study, the first scenario: is VGG19 architecture for the first experiment, the epoch value is 25, the learning rate is 0.01 and the batch size is 64. Getting a value for accuracy of 96% and a loss value of 00.30% for training data, and for testing data getting an accuracy value of 93.239% and a loss value of 01.40%. In the second scenario, the epoch value is 50 for the same learning rate and batch size getting an accuracy value of 97.00% and a loss value of 00.10% for training data, then for testing data getting an accuracy value of 94.057% and a loss value of 00.25. In the third scenario the accuracy value gets a result of 98.00% and a loss of 00.10% for training data, then for testing data getting an accuracy result of 94.602% and a loss value of 0.25%.

In the architectural model used in this experiment which is divided into 3 experimental scenarios, the third scenario gets a value of 98.00% training and 94.602% testing. From each scenario that is run, the scenario experiences an increase from the experiment scenario 1 and scenario 2, because the hyper-parameter tuning is carried out. Several parameters are changed in the third scenario, there is a learning rate that was previously 0.01 to 0.001, then the batch size from 64 to 256, with epoch 50. The precision, recall, f1-score, and support values produce good, stable, and increasing results in the recognition of Indonesian sign language which is influenced by parameter changes.

The core of the research conducted is to obtain the accuracy and consistency of VGG19 in recognizing Indonesian sign language. The performance of VGG19 shows strong results and good overall performance, by performing tunning parameters it is able to improve and obtain better results.

There are several common errors, but there are several solutions to overcome them to avoid errors or inequalities in the results obtained in further research. The dataset needs to be normalized to reduce redundancy and improve the integrity of the data used so that the data becomes organized. When loading, the data must be the same as the input form; there are no differences in variables that cause errors. To prevent overfitting in further research, which indeed uses more data that has been normalized and can split the data with a 70/30 or 80/20 comparison of training data and testing data [23].

The following image is an accuracy graph and a loss graph from scenario 1 which had the best results of the three scenarios, can be seen in figure 5.

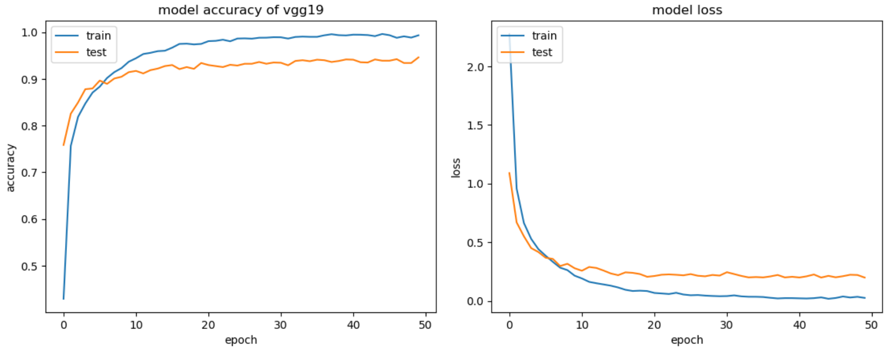

Figure 4. Accuracy and Loss Scenario 1

Figure 6 is the confusion matrix of Improved VGG19, after training the model with the given images [24s]. The cross-section shows the amount of image data that has been correctly detected, Can be seen in figure 6.

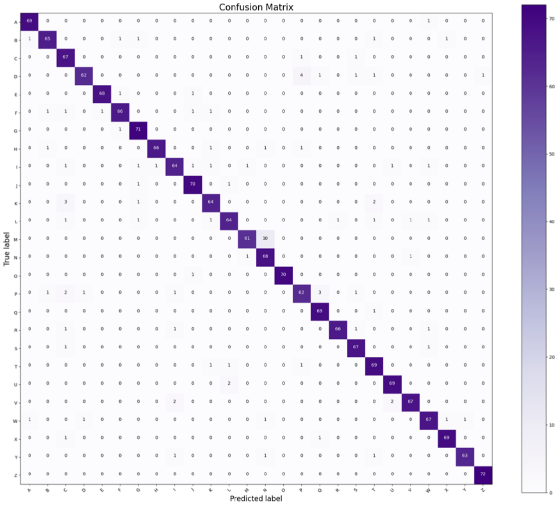

Figure 5. Confusion matrix of VGG19 with Hyper-Parameter Tuning (Scenario 1)

IV. Conclusion

In this study, the data used in this study is the BISINDO data set, with a sample size of 8,814 divided into 26 alphabetical categories. The implementation of sign language recognition uses the VGG19 architecture and the presence of hyper-parameters tuning that produce better accuracy and consistency results than before with a training value of 98.00% and testing of 94.602%. The results obtained come from changes in parameters that affect performance, namely Epoch, Learning Rate, and batch size in this study, which can produce better results than their normal values. It is hoped that in the future this research will be useful in the development of continued research, by implementing the methods in the study then a larger and more balanced and consistent dataset. In the future, it can be implemented in other methods or architectures as well and can improve performance in recognizing or categorizing variables.

In prospective studies, it would be worthwhile for researchers to broaden the scope of datasets. They could achieve this through augmentation techniques, which would amplify the volume of the existing datasets. Moreover, researchers could consider refining the VGG19 model or blending it with other models, in a bid to attain even more optimum outcomes.

References

[1] Nurul Khotimah, W., Andika Saputra, R., Suciati, N., & Rahman Hariadi, R. (n.d.). Khotimah, Saputra, Suciati, and Hariadi-Alphabet Sign Language Recognition Using Leap Motion Technology and Rule Based Backpropagation-Genetic Algorithm Neural Network (RBBPGANN) ALPHABET SIGN LANGUAGE RECOGNITION USING LEAP MOTION TECHNOLOGY AND RULE BASED BACKPROPAGATION-GENETIC ALGORITHM NEURAL NETWORK (RBBPGANN).

[2] Kothadiya, D., Bhatt, C., Sapariya, K., Patel, K., Gil-González, A. B., & Corchado, J. M. (2022). Deepsign: Sign Language Detection and Recognition Using Deep Learning. Electronics (Switzerland), 11(11). https://doi.org/10.3390/electronics11111780

[3] Papadimitriou, K., Potamianos, G., Sapountzaki, G., Goulas, T., Efthimiou, E., Fotinea, S. E., & Maragos, P. (2023). Greek sign language recognition for an education platform. Universal Access in the Information Society, 0123456789. https://doi.org/10.1007/s10209-023-01017-7

[4] Bora, J., Dehingia, S., Chetia, A. B. A. A., & Gogoi, D. (2023). Real-time Assamese Sign Language Recognition using MediaPipe and Deep Learning. Procedia Computer Science, 2018, 1384–1393. https://doi.org/10.1016/j.procs.2023.01.117

[5] Badiger, R. M., & Lamani, D. (2022). DEEP LEARNING BASED SOUTH INDIAN SIGN LANGUAGE RECOGNITION BY STACKED AUTOENCODER MODEL AND ENSEMBLE CLASSIFIER ON STILL IMAGES AND VIDEOS. Journal of Theoretical and Applied Information Technology, 100(21), 6587-6597. https://doi.org/10.1016/j.compeleceng.2020.106898

[6] Mia, K., Islam, T., Assaduzzaman, M., Shaha, S. P., Saha, A., Razzak, M. A., Dhar, A., Sarker, T., & Alvy, A. I. (2023). Isolated sign language recognition using hidden transfer learning. International Journal of Computer Science and Information Technology Research, 11(1), 28-38. https://doi.org/10.5281/zenodo.7615847

[7] Das, P., Ahmed, T., & Ali, M. F. (2020). Static Hand Gesture Recognition for American Sign Language using Deep Convolutional Neural Network. ResearchGate Conference Paper.

[8] Shin, J., Musa Miah, A. S., Hasan, M. A. M., Hirooka, K., Suzuki, K., Lee, H. S., & Jang, S. W. (2023). Korean Sign Language Recognition Using Transformer-Based Deep Neural Network. Applied Sciences, 13(5), 3029.

[9] Perdana, I., Darma Putra, I., & Arya Dharmaadi, I. (2021). Classification of Sign Language Numbers Using the CNN Method. JITTER : Jurnal Ilmiah Teknologi Dan Komputer, 2(3), 485-493. Retrieved from https://ojs.unud.ac.id/index.php/jitter/article/view/78264

[10] Rexona. (n.d.). Mengenal Dua Jenis Bahasa Isyarat di Indonesia | Rexona ® ID. Retrieved July 19, 2023, from https://www.rexona.com/id/zona-keringat/mengenal-dua-jenis-bahasa-isyarat-di-indonesia/

[11] Kembuan, O., Caren Rorimpandey, G., & Milian Tompunu Tengker, S. (2020, October 27). Convolutional Neural Network (CNN) for Image Classification of Indonesia Sign Language Using Tensorflow. 2020 2nd International Conference on Cybernetics and Intelligent System, ICORIS 2020. https://doi.org/10.1109/ICORIS50180.2020.9320810

[12] Ma'aruf, A. (2024). Indonesian Sign Language - BISINDO. Kaggle. Retrieved January 13, 2025, from https://www.kaggle.com/datasets/agungmrf/indonesian-sign-language-bisindo

[13] geeksforgeeks. (2023, May 6). Data Preprocessing in Data Mining. GeeksforGeeks. Retrieved July 19, 2023, from https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/

[14] Kaushik, A., & Castrejon, L. (n.d.). Understanding the VGG19 Architecture. OpenGenus IQ. Retrieved July 19, 2023, from https://iq.opengenus.org/vgg19-architecture/

[15] Tanwar, K. (n.d.). What is the VGG-19 neural network? Quora. Retrieved July 19, 2023, from https://www.quora.com/What-is-the-VGG-19-neural-network

[16] Erickson, B. J., & Kitamura, F. (2021). Magician's corner: 9. performance metrics for machine learning models. In Radi-ology: Artificial Intelligence (Vol. 3, Issue 3). Radiological Society of North America Inc. https://doi.org/10.1148/ryai.2021200126

[17] Yacouby Amazon Alexa, R., & Axman Amazon Alexa, D. (n.d.). Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models.

[18] Woodman, R. J., & Mangoni, A. A. (2023). A comprehensive review of machine learning algorithms and their application in geriatric medicine: present and future. Aging clinical and experimental research, 35(11), 2363–2397. https://doi.org/10.1007/s40520-023-02552-2

[19] Jones (2020, Descember).Efficacy of Pretrained VGG-16 in Sign Language Recognition," in IEEE Transactions on Image Processing, vol. 28, no. 12, pp. 5909-5916.

[20] Smith and Garcia. (2019, October), "VGG-16 for Gesture Recognition in Sign Language," in IEEE Signal Processing Letters, vol. 26, no. 10, pp. 1501-1505.

[21] Zhang. (2017, June), "Transfer Learning for Sign Language Recognition with VGG-19," in IEEE Access, vol. 5, pp. 12143-12152.

[22] Yacouby, R., & Axman, D. (2020, November). Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models. In Proceedings of the first workshop on evaluation and comparison of NLP systems (pp. 79-91).

[23] Gupta, R., Gupta, K., Pandit, C., & Singh, P. (2024, February). Sign Language Recognition using VGG16 and ResNet50. In 2024 11th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 996-1001). IEEE.

[24] Trivusi. (2022, July 16). Metriks Evaluasi Sistem Menggunakan Confusion Matrix. Trivusi. Retrieved July 19, 2023, from https://www.trivusi.web.id/2022/04/evaluasi-sistem-dengan-confusion-matrix.html