JURNAL EMACS
e-ISSN: 2686-2573
[Engineering, MAthematics and Computer Science] Vol.5 No.3 September 2023: 175-180 DOI: 10.21512/emacsjournal.v5i3.10621

Deep Transfer Learning for Sign Language Image Classification: A Bisindo Dataset Study

Ika Dyah Agustia Rachmawati[1*], Rezki Yunanda[2], Muhammad Fadlan Hidayat[3], Pandu Wicaksono[4]

[1] Cyber Security Program, Computer Science Department, School of Computer Science,
[2,4] Software Engineering Program, Computer Science Department, School of Computer Science,
[3] Computer Science Department, School of Computer Science,
Bina Nusantara University,
Jakarta, Indonesia 11480
ika.rachmawati001@binus.ac.id; rezki.yunanda@binus.ac.id; muhammad.hidayat003@binus.ac.id;
pandu.wicaksono005@binus.ac.id

*Correspondence: ika.rachmawati001@binus.ac.id

Abstract – *This study aims to identify and categorize the BISINDO sign language dataset, primarily consisting of image data. Deep learning techniques are used, with three pre-trained models: ResNet50 for training, MobileNetV4 for validation, and InceptionV3 for testing. The primary objective is to evaluate and compare the performance of each model based on the loss function derived during training. The training success rate provides a rough idea of the ResNet50 model's understanding of the BISINDO dataset, while MobileNetV4 measures validation loss to understand the model's generalization abilities. The InceptionV3-evaluated test loss serves as the ultimate litmus test for the model's performance, evaluating its ability to classify unobserved sign language images. The results of these exhaustive experiments will determine the most effective model and achieve the highest performance in sign language recognition using the BISINDO dataset.*

Keywords: *Bisindo; Sign Language; MobileNetv4; EffisienNetB1, Resnet50, InceptionV3*

I. INTRODUCTION

Communication provides an essential part in social interactions, as it enables individuals or collectives to transmit information and establish connections with their surroundings and other people. Communication often involves the utilization of linguistic expressions, either orally or in written form. However, those who possess physical impairments may resort to nonverbal means of communication, such as sign language, to convey their messages. In March 2022, the number of individuals who are deaf and speech impaired in Indonesia will reach 19,392 people, which is equivalent to 9.14% of the total number of people with disabilities in Indonesia (Arisandi & Satya, 2022).

Sign language is used to overcome limitations associated with being deaf and speech impairment. In the context of Indonesia, it is common to observe the utilization of two distinct sign languages, specifically Indonesian Sign Language (BISINDO) and Indonesian Sign Language System (SIBI) (Arisandi & Satya, 2022). BISINDO, the Indonesian Sign Language, was established by the deaf community and exhibits regional variances that reflect the many origins of the community. Consequently, BISINDO is regarded as a versatile sign language system (Pusbisindo, 2023).

The usage of SIBI is challenging for individuals who are deaf or have speech impairments due to its adherence to grammatical rules and taken from American Sign Language (ASL). SIBI is only used to communicate in formal settings, such as school activities (Handhika et al., 2018)namely Indonesian Signal System (SIBI. BISINDO consists of an alphabet with 26 characters consisting of the letters A to Z, formed by one hand for the characters C, E, I, J, L, O, R, U, V, and Z, while the letter characters formed with two hands are A, B, D, F, G, H, K, M, N, P, Q, S, T, W, X, and Y, which are seen in Figure 1 (Bestari, 2018).

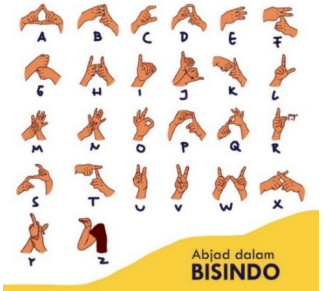

Figure 1. BISINDO Sign Language

In the comparison of responses to the use of BISINDO and SIBI in Indonesia as sign languages (Mursita, 2015), out of 100 respondents, 91% of the deaf prefer to use BISINDO rather than SIBI because they find it difficult to use SIBI rather than BISINDO, and only 9% use SIBI. BISINDO is able to enrich expressions so that it can liven up the atmosphere, make it easier to connect with lots of friends, and there are no barriers to communication (Pusbisindo, 2023).

Artificial Intelligence (AI) involves the utilization of computational methods to replicate human cognitive processes like interpretation, reasoning, decision-making, estimation, and categorization. This multidisciplinary field incorporates knowledge from diverse scientific domains such as mathematics, biology, genetics, engineering, and computer science. The primary aim of AI research is to empower computers to swiftly, effectively, accurately, and efficiently execute tasks that are typically associated with human cognition. Unlike traditional software, AI techniques can manage incomplete and uncertain data by establishing meaningful connections among data points, allowing for inferences about past occurrences and predictions concerning future outcomes. Besides its applications in design and engineering disciplines, AI is now extensively employed in various sectors, including engineering education, healthcare, transportation, economics, law, and manufacturing (Khaleel et al., 2023).

Machine Learning is a subset of AI that emphasizes the development of algorithms and statistical models that enable computers to learn and improve their performance on specific tasks through experience and data input (Fauzi et al., 2023). Deep Learning is a specialized branch of machine learning that involves artificial neural networks, specifically deep neural networks with multiple layers, allowing it to automatically discover intricate patterns and representations in data, making it particularly effective for tasks like image and speech recognition (Fadlilah et al., 2021).

Transfer learning is a technique for machine learning in which a model trained on one task or dataset is repurposed or fine-tuned to perform a different but related task or work with a different dataset. Transfer learning is a technique that capitalizes on the knowledge and features acquired during the initial training of a neural network or model, thereby enabling a substantial decrease in training time and resource demands. This methodology is especially advantageous in situations where there is a scarcity of data accessible for the specific objective, as it allows the model to leverage the insights acquired from a more extensive, interconnected dataset, hence enhancing its efficacy in the novel task (Susanty et al., 2021).

Deep transfer learning is a machine learning method that combines the capabilities of deep neural networks with transfer learning. It involves modifying a pre-existing deep learning model to fit a specific task, utilizing the hierarchical characteristics and representations from the original training. This approach is particularly useful in situations with limited data for the target task, enhancing model performance by applying knowledge from a source task to the target task. It has gained significant popularity in computer vision and natural language processing (Toengi, 2018).

The initial approach centers on the identification of sign language signs presented through visual media such as photographs or videos. This methodology involves the utilization of deep learning models, namely convolutional neural networks (CNNs), to discern visual patterns within images or video frames that correspond to sign language signs. The model has undergone training using a dataset comprising photographs of sign language signs that are suitable for multiple sign languages. Consequently, it possesses the ability to accurately identify and interpret these signs, generating corresponding textual or auditory representations (Triwijoyo et al., 2023).

The second approach places greater emphasis on the recognition of hand motions, which serve as the primary element in sign language. Deep learning is employed for the purpose of discerning and comprehending the gestures and postures of the hands that are utilized in the act of communicating through sign language. Deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have the capability to see and evaluate hand movements within video or picture data, then convert them into suitable textual or auditory representations (Li et al., 2019).

Both methodologies exhibit significant potential for enhancing accessibility for those with impairments who rely on sign language as their primary mode of communication. Deep learning enables systems to autonomously acquire knowledge from data, hence enhancing the model's proficiency in accurately detecting and understanding sign language as the volume of available data increases (Indra et al., 2019).

The sign language dataset image examines (Wadhawan & Kumar, 2020) the use of deep learning-based convolutional neural networks (CNN) for the robust modelling of static signals in sign language recognition. The study accumulates 35,000 sign images from a variety of users and assesses the performance of the proposed system on 50 CNN models. The system obtains the highest training accuracy of 99.72% and 99.90% on colored and grayscale images, demonstrating its effectiveness over earlier works.

Other research seeks (Bantupalli & Xie, 2019) to develop a vision-based application that provides sign language to text translation, thereby facilitating communication between signers and non-signers. The model derives temporal and spatial features from video sequences, utilizing Inception for spatial recognition and a recurrent neural network for temporal training.

InceptionV3 excels at efficiently capturing multi-scale features due to its inception modules with multiple filter sizes (1x1, 3x3, and 5x5). It is computationally efficient, reduces parameters with global average pooling, and provides pre-trained weights for effective transfer learning, which makes it ideal for image classification tasks. InceptionV3 The study utilizes a dataset of 36 English characters and digits and achieves 90% accuracy using American Sign Language data. The modified inceptionV3 architecture outperforms earlier research by 99.81%. (Hasan et al., 2020).

The Resnet50 model effectively addresses the issue of vanishing gradients, enabling the training of deeper models and consequently achieving superior accuracy in tasks related to image recognition. The Resnet50 model demonstrated recognition accuracy ranging from 88.38% to 93.88% for illnesses and from 95.38% to 98.42% for pests in the context of disease and pest identification studies (Yin et al., 2020).

EfficientNet was used to recognize facial expressions using transfer learning, and the results obtained were high facial expression recognition accuracy, namely 99.24% for CK+in from the 10% sampling used (Alam et al., 2022).

The main aim of this study is to perform a comparative analysis using the BISINDO dataset and various algorithms such as MobileNetV4, EfficientNetB1, ResNet50, and InceptionV3 is to identify the pre-trained model that achieves the highest accuracy and fastest performance in the classification of sign language images.

II. METHODS

The Figure 2 are several method steps that can be used in research that focuses on sign language image classification by utilizing the BISINDO dataset and the transfer learning approach.

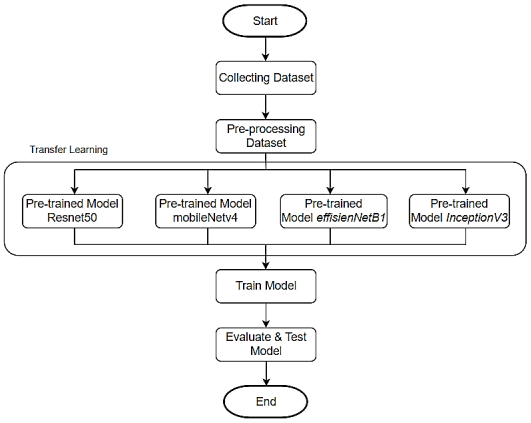

Figure 2. The Propose of Methods

2.1 Collecting Dataset

The dataset used is a dataset obtained from the Kaggle Public dataset (Noer, 2021), which contains portraits of letters A–Z at a scale of 1:1 with three different backgrounds (plain white shirt, white wall, and white shirt with dots). The photo was taken from a front-view perspective with a distance of 70 cm between the object and the camera lens. Each background was photographed with 4 photos for each letter, so there were a total of 1,727 photos of letters A–Z. The example of sign language dataset seen in Figure 3.

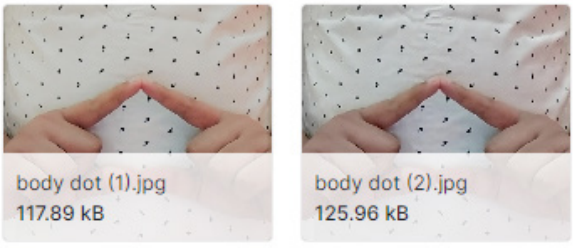

Figure 3. The Example of Sign Language Dataset

2.2 Pre-Processing Dataset

Preprocessing improved the dataset before training, validation, and testing. Rescaling pixel values to [0, 1], rotating images to introduce variations, horizontally flipping images for data enhancement, applying shear transformations to account for geometric distortions, and adjusting the fill mode for transformed regions were part of this preprocessing. The dataset was then divided into three subsets: the training set, which contains 70% of the data and trains the deep learning model; the validation set, which contains 20% and monitors model performance during training and hyperparameter tuning; and the testing set, which contains 10% and evaluates the model's generalization and final performance. These preprocessing and dataset division methods create a stable and well-structured experimental design for sign language image classification, deep learning model construction, and evaluation.

2.3 Pre-Trained Model

This research utilizes three pre-trained deep learning models MobileNetV4, EfficientNetB1, and ResNet50, as well as InceptionV3, in order to leverage their extensive prior knowledge and feature extraction capabilities. These models, which were initially trained on vast datasets such as ImageNet, are proficient at recognizing intricate patterns and high-level features in images. By fine-tuning these pre-trained models on our pre-processed sign language image dataset, we hope to exploit their transfer learning potential, enabling our system to recognize and classify sign language with greater accuracy and efficiency, even when working with limited data specific to our intended task.

2.4 Trained Model

The trained models, which include MobileNetV4, EfficientNetB1, ResNet50, and InceptionV3, represent the culmination of our deep learning strategy, wherein these models underwent fine-tuning on our pre-processed sign language image dataset. By adapting these pre-trained models to the complexities of sign language recognition, we have exploited their immense knowledge and feature extraction capabilities, resulting in potent classification tools for sign language BISINDO. These models have effectively learned to recognize and interpret the distinctive visual signals within images of sign language, demonstrating their skill at capturing subtle details and achieving impressive classification performance.

2.5 Evaluation & Test Model

The evaluation and testing phase involves a rigorous assessment of the trained models' performance in sign language image classification. These models, including MobileNetV4, EfficientNetB1, ResNet50, and InceptionV3, are subjected to an exhaustive battery of tests utilizing our dedicated testing subset. The purpose of these tests is to evaluate the models' ability to generalize and accurately classify unseen sign language BISINDO, thereby providing a thorough assessment of their real-world applicability by meticulously analyze accuracy, and loss metrics.

The measure of accuracy is determined by dividing the total number of correct predictions (Developer Google, 2022) (i.e., accurate classifications) by the total number of data samples that were examined. The categorical cross

entropy loss function is a loss function used in image classification tasks to measure the difference between predicted probability distribution and actual class labels. It minimizes the loss by taking the negative logarithm of the predicted probability of correct class assignment, promoting more likely correct class labels and is particularly effective in multi-class classification problems.

III. RESULTS AND DISCUSSION

In this section, we present a detailed analysis of the outcomes obtained through our experiments, shedding light on the performance metrics, including accuracy and loss, across a range of pretrained. These findings will be critically examined within the broader context of our research, allowing us to draw meaningful insights into the effectiveness of these models for sign language image classification.

To provide context for our analysis, we initially divided the dataset into training, validation, and testing sets, followed by preprocessing. Subsequently, pretrained models, namely ResNet50, MobileNetV4, EfficientNetB1, and InceptionV3, were downloaded with the 'include_top=False' option, omitting the top classification layers. These models were then compiled using the Adam optimizer and categorical cross entropy loss function, employing accuracy as the evaluation metric. Finally, we executed the training process for 25 epochs to fine-tune the pretrained models specifically for the sign language image classification task.

In Table I and Figure 3, we can observe the results of the model training process. ResNet50 demonstrated exceptional performance, achieving an impressive training accuracy of 99% and maintaining a high validation accuracy of 99%, with a slightly lower but still robust test accuracy of 98%.

In contrast, EfficientNetB1 encountered challenges throughout the training process, resulting in a low training accuracy of 8%, an even lower validation accuracy of 3%, and the lowest test accuracy at 1%. On the other hand, MobileNetV4 delivered outstanding results with a remarkable training accuracy of 98%, a flawless validation accuracy of 100%, and a strong test accuracy of 99%. InceptionV3 also displayed strong performance, achieving a training accuracy of 96%, closely resembling ResNet50 with a 99% validation accuracy and a solid test accuracy of 99%.

These findings underscore the varying effectiveness of pretrained models within the context of the classification task, with ResNet50 and MobileNetV4 emerging as the top-performing models.

Table I. Training Acc Results Table

No	Pretrained Model	Train Acc	Validasi Acc	Test Acc
1.	ResNet50	0.99	0.99	0.98
2.	EffecientNetB1	0.08	0.03	0.01
3.	MobileNetV4	0.98	1.00	0.99
4.	InceptionV3	0.96	0.99	0.99

In Table II and Figure 4 , we can observe the performance of different pretrained models throughout the training and testing phases. Notably, ResNet50 achieved the lowest loss values, with a training loss of approximately 0.03, a validation loss of around 0.06, and a test loss of 0.06. Conversely, EfficientNetB1 struggled during training, exhibiting notably high loss values with a training loss of about 3.40, a validation loss of approximately 3.34, and a test loss of 3.33.

MobileNetV4 demonstrated robust performance with low loss values, boasting a training loss of around 0.06, a validation loss of about 0.03, and a test loss of 0.02. InceptionV3 also displayed strong performance, with a training loss of roughly 0.12, a validation loss of approximately 0.01, and a test loss of 0.01. These loss values offer critical insights into how well these pretrained models generalize from the training data to the testing data, where lower loss values indicate better classification performance.

Table II. Training Loss Results Table

No	Pretrained Model	Train Loss	Validation Loss	Test Loss
1.	ResNet50	**0.03**	0.06	0.06
2.	EffecientNetB1	3.40	3.34	3.33
3.	MobileNetV4	0.06	**0.03**	0.02
4.	InceptionV3	0.12	0.01	**0.01**

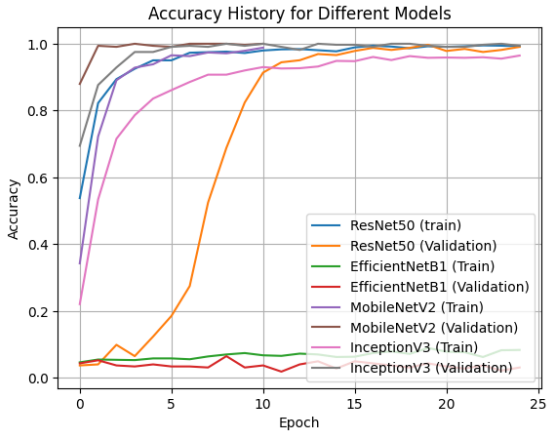

Figure 3. Accuracy History of Model Training

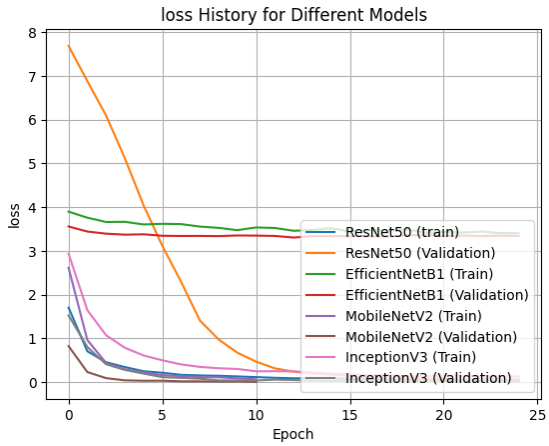

Figure 4. Loss History of Model Training

IV. CONCLUSION

Based on the training results of the models on the BISINDO image classification dataset, it is evident that ResNet50 and MobileNetV4 have exhibited superior performance compared to EfficientNetB1 and InceptionV3. ResNet50 achieved the highest accuracy levels across all phases of training, with remarkable training, validation, and test accuracy scores on BISINDO image classification data. MobileNetV4 closely followed suit, boasting consistently high accuracy across these phases.

Additionally, both ResNet50 and MobileNetV4 showcased low loss values, indicating their robust generalization capabilities for image classification in the BISINDO dataset. In contrast, EfficientNetB1 encountered significant challenges throughout the training process, yielding remarkably low accuracy and high loss values. InceptionV3 demonstrated commendable performance but fell slightly behind ResNet50 and MobileNetV4.

Therefore, in terms of overall performance in BISINDO image classification, ResNet50 and MobileNetV4 emerge as the top-performing models.challenges throughout the training process, yielding remarkably low accuracy and high loss values.

REFERENCES

Alam, I. N., Kartowisastro, I. H., & Wicaksono, P. (2022). Transfer Learning Technique with EfficientNet for Facial Expression Recognition System. *Revue d'Intelligence Artificielle*, *36*(4), 543–552. https://doi.org/10.18280/ria.360405

Arisandi, L., & Satya, B. (2022). Sistem Klarifikasi Bahasa Isyarat Indonesia (Bisindo) Dengan Menggunakan Algoritma Convolutional Neural Network. *Jurnal Sistem Cerdas*, *5*(3), 135–146. https://doi.org/10.37396/jsc.v5i3.262

Bantupalli, K., & Xie, Y. (2019). American Sign Language Recognition using Deep Learning and Computer Vision. *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, 4896–4899. https://doi.org/10.1109/BigData.2018.8622141

Bestari, H. (2018). *Mengenal Bahasa Isyarat*. Website. https://www.ypedulikasihabk.org/2018/11/09/mengenal-bahasa-isyarat/

Developer Google. (2022). *Classification: Accuracy*. Website. https://developers.google.com/machine-learning/crash-course/classification/accuracy

Fadlilah, U., Mahamad, A. K., & Handaga, B. (2021). The Development of Android for Indonesian Sign Language Using Tensorflow Lite and CNN: An Initial Study. *Journal of Physics: Conference Series*, *1858*(1). https://doi.org/10.1088/1742-6596/1858/1/012085

Fauzi, M. Z., Sarno, R., & Hidayati, S. C. (2023). Recognition of Real-Time BISINDO Sign Language-to-Speech using Machine Learning Methods. *International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*. https://doi.org/10.1109/ICCoSITE57641.2023.10127743

Handhika, T., Zen, R. I. M., Murni, Lestari, D. P., & Sari, I. (2018). Gesture recognition for Indonesian Sign Language (BISINDO). *Journal of Physics: Conference Series*, *1028*, 012173. https://doi.org/10.1088/1742-6596/1028/1/012173

Hasan, M. M., Srizon, A. Y., Sayeed, A., & Hasan, M. A. M. (2020). Classification of American Sign Language by Applying a Transfer Learned Deep Convolutional Neural Network. *ICCIT 2020 - 23rd International Conference on Computer and Information Technology, Proceedings*, 19–21. https://doi.org/10.1109/ICCIT51783.2020.9392703

Indra, D., Purnawansyah, Madenda, S., & Wibowo, E. P. (2019). Indonesian Sign Language Recognition Based on Shape of Hand Gesture. *Procedia Computer Science*, *161*, 74–81. https://doi.org/10.1016/j.procs.2019.11.101

Khaleel, M., Ahmed, A. A., & Alsharif, A. (2023). Artificial Intelligence in Engineering. *Brilliance: Research of Artificial Intelligence*, *3*(1), 32–42. https://doi.org/10.47709/brilliance.v3i1.2170

Li, G., Tang, H., Sun, Y., Kong, J., Jiang, G., Jiang, D., Tao, B., Xu, S., & Liu, H. (2019). Hand Gesture Recognition Based on Convolution Neural Network. *Cluster Computing*, *22*, 2719–2729. https://doi.org/10.1007/s10586-017-1435-x

Mursita, R. A. (2015). Respon Tunarungu Terhadap Penggunaan Sistem Bahasa Isyarat Indonesa (Sibi) Dan Bahasa Isyarat Indonesia (Bisindo) Dalam Komunikasi. *Inklusi*, *2*(2), 221. https://doi.org/10.14421/ijds.2202

Noer, A. (2021). *Bahasa Isyarat Indonesia (BISINDO) Alphabets*. Kaggle. https://www.kaggle.com/datasets/achmadnoer/alfabet-bisindo/data

Pusbisindo. (2023). *Mengapa Belajar BISINDO?* Website. https://www.pusbisindo.org/#mengapa

Susanty, M., Fadillah, R. Z., & Irawan, A. (2021). Model Penerjemah Bahasa Isyarat Indonesia (BISINDO) Menggunakan Pendekatan Transfer Learning. *Petir*, *15*(1), 1–9. https://doi.org/10.33322/petir.v15i1.1289

Toengi, R. (2018). *Application of Transfer Learning to Sign Language Recognition Using an Inflated 3D Deep Convolutional Neural Network*.

Triwijoyo, B. K., Karnaen, L. Y. R., & Adil, A. (2023). *An Approach for Sign Language Recognition with Deep Learning Algorithm*. *9*(1), 1–10. https://doi.org/10.1007/978-981-99-1435-7_1

Wadhawan, A., & Kumar, P. (2020). Deep Learning-Based Sign Language Recognition System for Static Signs. *Neural Computing and Applications*, *32*(12), 7957–7968. https://doi.org/10.1007/s00521-019-04691-y

Yin, H., Gu, Y. H., Park, C. J., Park, J. H., & Yoo, S. J. (2020). Transfer Learning-Based Search Model for Hot Pepper Diseases and Pests. *Agriculture (Switzerland)*, *10*(10), 1–16. https://doi.org/10.3390/agriculture10100439